Harnessing AI for Creating Definitions of Georgian Military Terminological Neologisms in the context of the Russia-Ukraine War

Ketevan Mchedlishvili
Ilia State University
E-mail: ketevani.mchedlishvili.3@iliauni.edu.ge

Abstract

The outbreak of the Russia-Ukraine war was followed by a rapid increase in drone-related terminology in the Georgian language. However, these terms remain absent from existing bilingual and encyclopedic Georgian military dictionaries to date. This paper explores whether Generative AI tools can be integrated into the terminological workflow, particularly to streamline the process of crafting draft definitions for Georgian military neoterms. For this, 10 drone-related multiword terminological neologisms were selected from Georgian comparable corpora, MiliCorpKa, consisting of analytical blog materials regarding the Russia-Ukraine war. Next, draft definitions were generated with five LLMs (ChatGPT, Grok, Claude, Gemini, DeepSeek) using two prompting strategies: a so-called basic prompt and an ISO-based prompt. The outputs were assessed against a self-developed evaluation scheme. Preliminary results of this study emphasize the importance of fine-tuning prompts and the differences in effectiveness among 5 models for creating definitions in Georgian. The results of this exploratory, qualitative research offer a glimpse of the extent to which Generative AI Chatbots can be a helpful tool for post-editing terminological work and for creating draft definitions for low-resourced languages, such as Georgian.

Keywords: LLMs; GenAI; Georgian military neoterms; draft definitions; Russia-Ukraine war

1 Introduction

Given the central role of technological innovation in contemporary armed conflicts, it is not surprising that these developments are accompanied by linguistic change, manifested in the creation of new terms and the revival of previously unrecorded ones (Georgieva 2015: 42; Struk, Semiginovskaya & Sitko 2022). The outbreak of the Russia-Ukraine war, which has already been assessed as new generation warfare by military experts (Sorge 2023), caused a rapid influx of drone-related terminology into the Georgian military language.

Although drone-related terminology is well-established and recorded in English terminological resources, this is not the case for Georgian specialized dictionaries. The preliminary

analysis of the Georgian military term დრონი / *droni* ('drone') in existing Georgian military dictionaries (2009, 2017) and in Georgian general-language corpora revealed its absence in lexicographic resources and its recent emergence in the latter[1]. Based on these facts, drone-related military terms can be considered candidates of Georgian Neoterms, new terminological units "specifically coined for a given general concept" (Cabre 1999: 205; ISO 1087 2019).

Given the absence of drone-related terms in Georgian military terminology and the importance of the ongoing conflict for Georgia within the framework of NATO partnership, it is crucial to analyze, define, and document these terms in updated versions of military dictionaries. One of the essential components of terminological entries is definitions. However, the process of their crafting can be laborious and time-consuming (San Martin 2024: 1).

Recently, Generative AI (GenAI), with ChatGPT and other large language models (LLMs) under the spotlight of public and academic interest, has been assessed as a potential game-changer for many professionals, including linguists. A growing body of literature focusing on employing GenAI tools, such as Chatbots, for various lexicographic and terminological tasks (Lew 2023; Trap-Jensen 2024; Sabanés & Cunha 2025) has opened the possibility of integrating LLMs into lexicographic/terminological workflows. According to Trap-Jensen (2024: 30), the majority of researchers agree that "ChatGPT is doing a fine job as a defining lexicographer". The following statement is further supported by San Martin (2024), who proposes ways to use ChatGPT to craft draft terminological definitions. According to him, GenAI tools especially those powered by LLMs, can be linguistically promising and with varying degrees of reliability can reduce time and effort needed for creating terminological definitions (San Martin 2024: 1). Therefore, they have the potential of changing the existing terminological workflow and heralding the beginning of new age of post-editing or AI-assisted terminography (San Martin 2024: 4).

Despite a growing body of scientific papers mainly dedicated to applying LLMs, particularly ChatGPT, to lexicographic/terminological tasks in English (de Schryver 2023; Nunzio, Gallina, & Vezzani 2024; Heinisch 2025; de Schryver 2025), research on applying GenAI tools to Georgian is thin on the ground. This topic has only recently begun to attract scholarly interest, namely, 3 papers were presented at the II International Conference *Lexicography in the XXI Century* (Lomidze 2025; Chighladze 2025; Mchedlishvili 2025). This paper expands on the talk presented during the above-mentioned conference (Mchedlishvili 2025). It aims to evaluate the effectiveness of large language models in crafting draft definitions for Georgian drone-related military neologisms by comparing outputs from 5 freely available LLMs (ChatGPT 5.0, Gemini 2.5 Flesh, DeepSeek V3-1, Grok 3.0 Fast, Claude Sonnet 4.0) and 2 prompting strategies (basic prompt vs. ISO-based prompt). Additionally, their potential as a support tool for crafting draft definitions and evaluating the role of the human editors will be further explored.

Chapter 2 presents a brief overview of AI-assisted lexicography and terminography, with a key focus on crafting definitions. In the 3rd chapter, the methodology adopted for this

1 Available at: http://aranea.juls.savba.sk/aranea/run.cgi/rst?corpname=AranGeor&reload=1[accessed 17/02/2026]

study is discussed, while chapter 4 presents results and interpretations per-prompt and per-model. Finally, the article concludes by summarizing the findings, discussing the limitations of the current study, and suggesting future directions.

2 AI in Lexicography and Terminography: a Brief Overview

As de Schryver (2023: 358) states: "Lexicographers have always been at the forefront of dreaming up ways to harness the latest technology in order to compile better dictionaries". Starting from building a corpus to applying various corpus-based methods for linguistic analysis, lexicographers contemplated whether the full automation of dictionary compilation without further human assistance is possible (Rundell & Kilgariff 2011). It should be noted that the field of artificial intelligence, alongside its subfield natural language processing, has been well-established and is not a novelty for the field of lexicography. However, recent developments, particularly the public availability of AI chatbots, so-called LLMs, have led to a renewed interest in exploring ways to use new technology to the benefit of scientific disciplines.

OpenAI launched ChatGPT in November 2022, and since then, other companies, including xAI, Anthropic, and Google, have rapidly entered the field with their own generative AI models. Nowadays, various models, such as Gemini (Google), GPT (OpenAI), Claude (Anthropic), DeepSeek (High-Flyer), Grok (xAI), Qwen (Alibaba), Gemma (Google), and others, have become widely available not only for domain experts but also for laypeople. After the lecture given by Gilles-Maurice de Schryver (de Schryver and Joffe 2023) in Tokyo, the Center for Open Data in the Humanities (CODH) in February 2023 titled "The End of Lexicography, Welcome to the Machine", the study of GenAI has gained momentum.

The first scientific articles dedicated to exploring the application of LLMs, particularly focused on the usage of ChatGPT as an innovative linguistic tool (de Schryver 2023; Lew 2023). In his review paper, de Schryver (2023) critically overviews the existing 10 papers on the usage of ChatGPT. The author points out that although views vary, and concerns towards this technology are present, "one should jump on the wagon and use the technology, to free up time for us humans to do more useful things" (de Schryver 2023: 361). He further claims that in-the-flesh lexicographers are still crucial for the lexicographic process, though, human intervention is better to be saved for the so-called vetting stage (de Schryver 2023: 361). This view is supported by follow-up studies (Barret 2023, as cited by de Schryver 2023: 361; San Martin 2024; Heinisch 2025) requiring human involvement both in the beginning to create an optimal prompt and at the end, to review and refine output.

Notwithstanding varying views and uncertainties around LLMs, they have been successfully applied to different lexicographic tasks, including but not limited to generating headword lists, illustrative examples, equivalents in target languages, and sense division (Jakubíček & Rundell 2023; Lew 2023; Trap Jensen 2024; de Schryver 2025). Another topic that has been in the limelight of academic interest is the creation of definitions (McKean & Fitzgerald 2023, as cited by de Schryver 2023: 369; Lew 2023; Rundell 2023; Trap-Jensen 2024). Although results vary, Trap Jensen (2024: 30) states that most researchers agree that ChatGPT is capable of crafting concise, eloquent definitions that require minimal lexicographic scru-

tiny at the post-editing stage. The results can be further enhanced by feeding specific illustrative examples, whole corpora, or specifying and refining prompts (Barret 2023, as cited by de Schryver 2023: 362; McKean & Fitzgerald 2023, as cited by de Schryver 2023: 369).

Although the advantages of using LLMs, including user-friendliness, quick access, and potentially lightening the load of lexicographic work, are hard to overlook, scholars indicate important limitations to applying new technology. Among these disadvantages, one of the most important ones is the lack of reliability and replicability (Trap-Jensen 2024: 35). The absence of information regarding sources and constant evolution of the existing models, which in turn, results in different outputs, can be deemed as a deal-breaker for scientific research (Rundell 2023; Trap-Jensen 2024). Additionally, the same prompts can generate completely different answers that decrease inter-agreement and comparability of scientific results (Trap-Jensen 2024). Moreover, the enthusiasm for applying LLMs to lexicographic tasks in English cannot be carried over to other languages, especially low-resourced or endangered ones (Trap-Jensen 2024: 35; de Schryver 2025: 397). The fact that over 90% of the training data is in English creates bias, and all the prompts given in any other language are answered via English (de Schryver 2023: 370). Thus, it is too soon or overly enthusiastic to predict "heralding the end of lexicography" (de Schryver 2023: 356).

Similar tendencies toward integrating AI in the workflow have been observed in terminology. Likewise, terminologists started to apply with different levels of efficacy GenAI tools to various terminological tasks, including term coinage, term extraction, generation of equivalents, terminological variants, definitions, building concept systems, etc. (Łukasik 2023; Nunzio, Gallina & Vezzani 2024; Sabanes & Da Cunha 2025; Heinisch 2025). Similar to lexicographic practice, the same views persist regarding human participation (Heinisch 2025; Sabanes & Da Cunha 2025). Henisch (2025: 68) points out the importance of terminologists' expertise in LLM literacy to "assess whether an LLM is suitable for a particular terminology task, domain, or language". Moreover, the restrictions in terms of languages are particularly evident when it comes to terminological work. As Henisch (2025: 76) states: "hallucinations are more likely in niche domains, rapidly developing domains, or newly emerging domains, as not all current data can be present during training of the model". Along the same line, Łukasik (2023) in his research on the low-resource Kashubian language notes that ChatGPT's performance is significantly worse compared to traditional, corpus-based methodology.

It is noteworthy that the creation of terminological definitions has been under the spotlight of terminologists, too (San Martin 2024; Nunzio, Gallina, & Vezzani 2024; Heinisch 2025). In his paper, San Martin (2024) presents various methods for the crafting and refinement of useful prompts for generating definitions. Some of his chosen techniques include simple, basic questions that yield encyclopedic definitions as well as more fine-tuned approaches. For instance, preparing definitional templates, specifying the type of terminological definitions needed, for example intensional definitions (Nunzio, Gallina & Vezzani 2024: 3232), feeding illustrative contexts from corpora that are especially helpful in new or emerging domains due to the lack of training data (San Martin 2024; Heinisch 2025: 71). Notwithstanding the existing challenges, the authors agree upon the inevitability of adopting LLMs into terminological tasks as the "distinction between human-created and AI-created content will become increasingly blurred" (San Martin 2024: 7). One thing remains unchanged,

the necessity of human editor involvement and assigning the role of assisting tool to GenAI (Łukasik 2023; San Martin 2024; Heinisch 2025).

The studies reviewed here so far indicate that "a new age, that of the successful application of generative AI in lexicography, has dawned" (de Schryver 2023: 355). Thus, in view of all that has been mentioned, it is crucial to scientifically analyze the potential of integrating GenAI tools into Georgian lexicographic/terminological work. In this paper, my main questions of interest are:

1. How do different freely available LLMs (ChatGPT, Grok, Claude, Gemini, DeepSeek) vary in their abilities to generate accurate, contextually appropriate draft definitions for drone-related Georgian military neologisms?

2. How do different prompts (basic, ISO-based) influence the quality, precision, and terminological adequacy of crafted definitions?

3. What are the advantages and disadvantages of using specific LLM models for creating draft definitions in Georgian?

4. What is the role of a human editor in assessing, refining, and validating generated definitions?

To answer these questions, I will discuss the methodology adopted for this study and further explain results per-prompt and per-model in the subsequent chapters.

3 Methodology

This exploratory, qualitative study uses a multistage methodology, including (1) using Georgian specialized comparable corpora (MiliCorpKa) to extract modern Georgian military multiword terms; (2) employing corpus-based methods (keywords, collocation and concordance analysis) for the semi-automatic extraction of the multiword terminology; (3) defining and adopting criteria for the concept of terminological neologism; (4) selecting 10 drone-related neoterms for the case study; (5) choosing 5 freely available LLMs for evaluating their effectiveness to generate draft definitions; (6) creating two prompts, namely a basic one, and ISO-based for further analysis; (7) developing evaluation scheme for the generated outputs.

The first stage of analysis was based on the English and Georgian military comparable corpora (MiliCorpEng, MiliCorpKa) previously compiled within the framework of PhD research (Mchedlishvili 2024: 247; Mchedlishvili 2025: 112). The Georgian corpus consisted of 426,362 tokens, 53,405 types, 9 sources related to military-analytical blog materials covering the Russia-Ukraine war. Within the framework of a PhD dissertation, using the #Lancsbox6.0 (Brezina et al. 2021) software tool and employing contrastive collocation analysis (Vicente 2012; Ďurčo 2020; Cabezas-García 2024) 27 drone-related English-Georgian multiword terms were extracted.

At the next stage, the Georgian terms were further analyzed to determine whether they could be considered Georgian terminological neologisms i.e. neoterms candidates. For this purpose, we followed Cabre's (1999: 205) approach, namely their absence from existing

Georgian military encyclopedic dictionary (2017) and English-Georgian military dictionaries (2009, 2017) coupled with the criteria of their recent emergence. In order to check the latter one, we used two Georgian general language corpora: KaWak[2], which includes texts up to 2013, and Aranea[3], a Georgian comparable general language corpus that consists of texts crawled from 2014 onwards, including the latest 2023-2024 crawl covering the Russia-Ukraine war. The absence from the dictionaries, combined with the presence of terminological units in Aranea, was considered a strong indicator of terminological neologisms. The analysis revealed that all Georgian drone-related terms emerged mostly after 2018. For this case study, 10 drone-related Georgian multiword terms were randomly selected (see Figure 1).

A	B	C
Georgian Term	**Transliteration**	**English Equivalent**
კამიკაძე დრონი	k'amik'adze droni	kamikaze drone
სადაზვერვო დრონი	sadazvervo droni	reconnaissance drone
თვითმკვლელი დრონი	tvitmk'vleli droni	suicide drone
FPV (ხედვა პირველი პირისგან	FPV (khedva p'irveli p'irisgan) (	FPV (First Person View) drone
დრონების ხროვა	dronebis khrova	swarm drone
"ლანცეტის" ტიპის დრონი	lantset'is t'ip'is droni	Lancet drone
დრონის ოპერატორი	dronis op'erat'ori	drone operator
წყალქვეშა დრონი	ts'qalkvesha droni	underwater drone
კვადროკოპტერის დრონი	k'vadrok'op't'eris droni	quadcopter drone
დამრტყმელი დრონი	damrt'qmeli droni	attack drone

Figure 1: selected terminological neologisms for the case study.

The next step involved selecting freely available LLM versions operating in the Georgian language. Given that underlying models evolve rapidly, it is important to mention that experiments related to this study were conducted between September 2025 and October 2025, particularly with these models: ChatGPT 5.0,[4] Gemini 2.5 Flesh,[5] DeepSeek V3-1,[6] Grok 3.0 Fast[7], Claude Sonnet 4.0.[8] Following San Martin (2024), two prompting strategies were selected. The first one was the so-called basic prompt which only involved asking to define a particular term. The second one was named ISO-based prompt in which it was specified that the terminological definition style should follow ISO-704 Standard (2022) for

2 Available at: https://cqpweb.lancs.ac.uk/kawac200mw [accessed: 20/02/2026]

3 Available at: http://aranea.juls.savba.sk/aranea [accessed 20/02/2026]

4 https://chatgpt.com [accessed 20/02/2026]

5 https://gemini.google.com/app [accessed 20/02/2026]

6 https://www.deepseek.com [accessed 20/02/2026]

7 https://grok.com [accessed 20/02/2026]

8 https://claude.ai [accessed 20/02/2026]

intensional definitions, and include a generic concept as well as differentiating criteria. All instructions were written and asked in Georgian. Accordingly, outputs were provided in the Georgian language. Definitions for both prompts were generated in multiple independent sessions in September 2025 using the web interface of the above-mentioned models. Table 1 below shows the exact formulation of the prompts in the original Georgian version, with an English translation provided.

Prompt type	Georgian Instruction	English translation
Basic prompt	შექმენი ქართული სამხედრო ტერმინის [კამიკაძე დრონი] ტერმინოლოგიური განმარტება	Generate terminological definition for the Georgian military term [კამიკაძე დრონი / kamikadze droni (Kamikaze drone)]
ISO-704 Based	ISO-704 სტანდარტის მიხედვით, შექმენი ქართული სამხედრო ტერმინისთვის [კამიკაძე დრონი] შინაარსის დამაზუსტებელი დეფინიცია (intensional definition). ტერმინოლოგიური დეფინიცია უნდა შეიცავდეს გვარ-სახეობით მიმართებას: უშუალო გვარეობით ცნებასა და გამმიჯნავ მახასიათებლებს	Craft an intensional definition of the Georgian military term [კამიკაძე დრონი / kamikadze droni (Kamikaze drone)] based on the ISO-704 standard. The terminological definition must include generic relation: a generic concept (superordinate) and delimiting characteristics

Table 1: Basic and ISO-704 based prompts in Georgian with English translation.

To evaluate the outputs, a self-developed scheme was designed. The process was carried out in a Microsoft Excel worksheet. In each case, a 3-point scale was used for both prompts for the 2 main criteria. The first one was named terminological and structural consistency (TSC) with the following specifications for assigning points: (3 points good) - fully structured, genus and differentiating features present; terminology precise and consistent; (2 points partial) - some structural elements present, minor encyclopedic information and redundancy; some terms correct but minor errors or inconsistencies remain; (1 point poor) – definition lacks ISO structure, genus, or differentiating features; terminology is incorrect or inconsistent. The second one was named content accuracy (CA) with the following scores: (3 points good) – fully accurate, facts correct, matches military context; (2 points – partial) - mostly accurate, minor factual/contextual errors; (1 point – poor) - incorrect facts or concept misinterpreted. In addition, special codes were designed to indicate error types in generated definitions: *O (omission of essential characteristics), I (inclusion of irrelevant information), G (genus term missing), V (vague/non-specific language), E (encyclopedic style definitions instead of terminological).* The average TSC and CA scores were calculated for all 10 terms under analysis, per-model and per-prompt (See Figure 2).

Kamikadze drone	ISO template based	Groc 3.0 fast (Free)	კამიკაძე დრონი არის უპილოტო საფრენი აპარატი (genus, superordinate), რომელიც განკუთვნილია ერთჯერადი გამოყენებისთვის თვითგანადგურებით, რათა მიაყენოს მაქსიმალური ზიანი სამიზნეს (deliminating characteristic).	A kamikaze drone is an unmanned aerial vehicle (genus, superordinate) designed for single-use self-destruction to inflict maximum damage on a target (deliminating characteristic).	2- language in georgian is clumsy	3	no error, a bit clumsy	Provides targeted, not encyclopedic information. All essential information is present.

Figure 2: extract of assessment from Microsoft Excel worksheet.

4 Results and Discussions

This chapter presents the results obtained from the outputs generated in response to individual prompts and offers an interpretation of the findings. It further examines the advantages and limitations of using these models in Georgian. Finally, it analyzes the role and extent of human editors' contribution to AI-assisted terminological work.

4.1 Basic Prompt – general assessment

As mentioned before, the first two questions in this study sought to determine how effective the selected 5 LLMs are at generating terminologically and linguistically accurate draft definitions of Georgian military terms. Additionally, my point of interest was whether differences in prompts can affect the accuracy and terminological adequacy of the generated draft definitions.

For the basic prompt, each model was asked to define the selected Georgian drone-related neologisms (see Table 1 above). Table 2 below presents the general assessment of the cumulative average TSC and CA scores across all 10 terms. Although in this subchapter only several representative examples are discussed per-model, the same procedure was applied to the analysis of all selected terms.

Model	Average TSC	Average CA	General errors
ChatGPT 5.0	2,1	3	E, I
Claude Sonnet 4.0	2,6	3	E
Gemini Flesh 2.5	2,1	3	E, I
Grok 3.0 Fast	2,1	3	E, I
DeepSeek V3-1	1,6	2,7	E, I, O

Table 2: Average scores per model with basic prompt.

As shown in Table 2 above, although content accuracy was generally acceptable across all models, the average terminological and structural consistency varied. Further analysis showed that, for a basic prompt, all LLMs tended to generate more encyclopedic than terminological definitions, resembling those of explanatory articles. Moreover, the main tendency observed in terms of typical errors was the inclusion of lots of irrelevant information, for instance, the detailed list of all types of suicide drones, or operations in which they were used (see Figure 3 below). What is striking about the data in this table is that the DeepSeek model in particular, occasionally (3 terms out of 10 terms) missed a generic concept and delimiting characteristics. Notwithstanding differences in average scores across LLMs, the syntactic and morphological errors in all outputs in Georgian were persistent. Figure 3 below illustrates examples of the generated draft definitions for the neoterm kamikaze drone of the Claude which had the highest average TSC score, and DeepSeek which had the worst result among others.

A	B	C	D
Model	**Definition in Georgian**	**Definition translated in English**	**Errors Indicated**
DeepSeek V.3-1	კამიკაძე-ტიპის დრონი, ასევე ცნობილი როგორც **"მოხმხუბარი საბრძოლო საშუალება" (Loitering Munition)**, არის ჰიბრიდული სისტემა, **რომელიც აერთიანებს უპილოტო საფრენ აპარატის (UAV) და მართვადი რაკეტის თვისებებს.**	A kamikaze-type drone, also known as a **"fighting munition" (Loitering Munition)**, is a hybrid system that combines the features of an unmanned aerial vehicle (UAV) and a guided missile.	Incorrect terminology, definition is general and vague. Does not provide essential characteristics of the concept
Claude Sonnet 4.0	**პილოტის გარეშე საჰაერო ხომალდი**, რომელიც განკუთვნილია ერთჯერადი გამოყენებისათვის და **მიზნის განადგურებისათვის თვითდაკოცნებით**. ეს უპილოტო საფრენი აპარატი ატარებს საბრძოლო ტვირთს და პროგრამირებულია ან **დისტანციურად მართულია მიზნისკენ მიფრენისა და მასთან პირდაპირი შეჯახებისათვის.**	**Pilotless aerial vehicle** intended for single use and for **destroying the target through self-kissing**. This unmanned flying apparatus carries a combat payload and is programmed or remotely controlled to fly toward the target and directly collide with it.	Clumsy sentence structure in Georgian, terminological errors in Georgian

Figure 3: illustrative examples of draft definitions generated by Claude and DeepSeek.

In these illustrative examples, even the best average scoring model made lots of mistakes in terms of using correct Georgian terminology, explaining kamikaze drone as *პილოტის გარეშე საჰაერო ხომალდი/p'ilot'is gareshe sahaero khomaldi* ('pilotless aerial vehicle'), or explaining that targets are being destroyed by self-kissing (*თვითდაკოცნით/ tvitdak'otsnit*), rather than self-destruction. As for the DeepSeek model, there is no information on the essential characteristic of a kamikaze drone, beyond the vague claim that it is a hybrid system that combines features of a guided missile and an unmanned aerial vehicle. In this case, the model used unusual and pretty comical equivalent for loitering munitions *მოხმხუბარი საბრძოლო საშუალება/ mochkhubari sabrdzolo sashualeba* ('fighting munitions'). However, it should be noted that none of the models could provide or coin a correct equivalent in Georgian for the loitering munitions, possibly because it was absent from the training data.

A closer analysis of illustrative examples revealed that the use of a basic prompt without further specifications yields very long, verbose, encyclopedic explanations. Thus, it increases the terminologists' post-editing burden. This finding is consistent with that of San Martin (2024: 5) where he highlights that giving specific instructions provides better results in terms of crafting terminological, rather than encyclopedic definitions. Additionally, limita-

tions of all these models in terms of handling the generation of the definitions in Georgian accords with earlier observations made by Trap-Jensen (2024: 35) and Henisch (2025: 76), the lack of training data in Georgian, or the fact that some terms are emerging cause terminological errors. For instance, DeepSeek's worst performance may be caused by less material available to the Chinese LLM for Georgian.

4.2 ISO-704 based prompt – general assessment

If we now turn to the ISO-704-based prompt, each model was instructed to provide an intensional terminological definition and it was further specified to include generic concept and essential characteristics of the concept. The most significant finding compared to basic prompt results was the generation of generally shorter targeted terminological definitions that better aligned with terminological expectations. However, the effect of using an ISO-based prompt was model-dependent. As shown in Table 3 below, the average TSC and CA scores improved significantly for ChatGPT and Grok.

Model	Average TSC	Average CA	General errors
ChatGPT 5.0	2,8	3	syntactic, morphological errors
Claude Sonnet 4.0	2,6	3	V
Gemini Flesh 2.5	2,1	3	O, V, syntactic errors
Grok 3.0 Fast	2,1	3	minor syntactic, morphological errors
DeepSeek V3-1	1,6	2,7	incorrect terminology, incorrect grammar

Table 3: Average scores per model with ISO-704 based prompt

A closer analysis of the generated outputs revealed that ChatGPT and Grok produced concise, accurate, terminologically acceptable definitions that had a clear structure consisting of a generic concept and delimiting characteristics. The results for Gemini and DeepSeek remained largely unchanged, still struggling with correct terminology and, occasionally the definitions crafted in Georgian were hard to grasp. Claude showed the most unexpected performance, in 4/10 terms definitions became vaguer and less clear than those produced by basic prompt. Claude's performance might be caused by the model's general knowledge of what an intensional definition is, and by suggesting broader phrasing for a generic concept and delimiting characteristics. This finding is in line with prior research by San Martin (2024: 5), suggesting that, at times, for greater clarity, models perform better when provided with context or definitional templates rather than simply being instructed to generate intensional definitions. Figure 4 below illustrates comparative examples: the ChatGPT model performed better, and the Claude model's definitions became vaguer.

Model	Term	Basic prompt	ISO-704 based prompt
ChatGPT 5.0	დრონების ხროვა (swarm drone)	სამხედრო ტექნოლოგიაში აღნიშნავს ერთობლივად მოქმედ უსაფრთხო აპარატთა (დრონების) ჯგუფს, რომლებიც კოორდინირებულად, როგორც ერთი სისტემა, ახორციელებენ დავალებებს. ასეთი ხროვა შეიძლება იყოს სხვადასხვა ზომის და დანიშნულების დრონებისგან შემდგარი და მისი მთავარი მახასიათებელია კოლექტიური მოქმედება, მათ შორის თვითმართული მანევრები, სამიზნეების ერთდროული აღმოჩენა, შეტევა ან თავდაცვითი ოპერაციები. **დრონების ხროვის გამოყენება ზრდის ოპერაციების ეფექტიანობას, ართულებს მოწინააღმდეგის თავდაცვით ზომებს და უზრუნველყოფს მოქნილ, დეცენტრალიზებულ სამხედრო მოქმედებებს.**	დრონების ხროვა არის **უპილოტო საფრენი აპარატების ჯგუფი,** რომელიც მოქმედებს **კოორდინირებულად საერთო ამოცანის** შესასრულებლად
ChatGPT 5.0	დრონების ხროვა (swarm drone)	English Translation - In military technology, it denotes a group of pilotless apparatuses (drones) operating jointly, which in a coordinated manner, as one system, carry out tasks. Such a swarm may consist of drones of different sizes and purposes, and its main characteristic is collective action, including self-directed maneuvers, simultaneous detection of targets, attack, or defensive operations. **The use of a drone swarm increases the effectiveness of operations, complicates the opponent's defensive measures, and ensures flexible, decentralized military actions.**	English Translation - Swarm Drone is a **group of unmanned aerial vehicles** that **operates in a coordinated manner** to **accomplish a common task.**
Claude Sonnet 4.0	თვითმკვლელი დრონი (suicide drone)	უპილოტო საბრძოლო საშუალება, რომელიც განკუთვნილია ერთჯერადი გამოყენებისთვის სამიზნის დასაზიანებლად ან **განსანადგურებლად** მისი ფიზიკური **შეჯახების ან აფეთქების შედეგად**	თვითმკვლელი დრონი არის უპილოტო საჰაერო ხომალდი, რომელიც განკუთვნილია **მიზნის გაუქმებისთვის საკუთარი განადგურების საშუალებით.**
Claude Sonnet 4.0	თვითმკვლელი დრონი (suicide drone)	English Translation - Unmanned munition, which is intended for one-time use **for damaging or destroying a target as a result of its physical collision or explosion.**	English Translation - Suicide drone is an unmanned aerial vehicle, which is intended for **elimination of a target by means of its own destruction.**

Figure 4: basic prompt vs. ISO-based prompt generated by ChatGPT and Claude.

These examples (Figure 4) indicate that when it comes to ChatGPT, a long encyclopedic-style definition that contains information about types of operations and models transforms into a terminological one with an ISO-based prompt. In this instance, for the term swarm drone we have a clearly defined generic concept – unmanned aerial vehicle, and a delimiting characteristic – coordinated work. As for the Claude Sonnet model, the basic prompt definition is more specified, explaining that the suicide drone is being destroyed after physical collision with the target. On the other hand, in the ISO-based definition we have a vaguer phrasing that a suicide drone is cancelling a target by self-destruction, without explaining how or what follows this self-destruction. Despite varying model performance, the main problem of morphological, syntactic and/or lexical errors persists to varying degrees across all models.

The results obtained from the use of the ISO-704-based prompt support the findings of other studies in the area of terminological definitions, which claim that specifying the type of definition leads to better, more targeted outputs (San Martin 2024; Heinisch 2025; Nunzio, Gallina & Vezzani 2024: 3232). On the one hand, the findings highlight the necessity of post-editing by a human lexicographer/terminologist as well as the importance of the in-flesh-terminologist to create and get optimal prompts to get satisfying output (Barret 2023, as cited by de Schryver 2023: 362; Heinisch 2025: 68). On the other hand, the sentence clumsiness and errors in the Georgian language indicate the bias towards English and the fact that all models answer via English, thus, resulting in unnatural sentence structures in Georgian (Jakubíček & Rundell 2023; Trap Jensen 2024).

Finally, if we return to the first two research questions, we can deduce that freely available large language models differ in their capacity to generate accurate, terminologically coherent definitions. Their performance is primarily dependent on the correct, detailed prompt

given. That is why an ISO-based prompt yields better results when crafting draft terminological definitions. In the following subchapter, I will briefly summarize the advantages and limitations of the models tested in this case study.

4.3 Assessment per-model: a brief summary

Following the cumulative results of a basic and ISO-704-based template, the tested LLMs were assigned the following places in terms of their effectiveness in handling the generation of draft definitions: 1. ChatGPT 5.0; 2. Grok 3.0 Fast; 3. Claude Sonnet 4.0; 4. Gemini Flesh 2.5; 5. DeepSeek V3-1.

ChatGPT and Grok shared most of the advantages and limitations. Their common strengths included considerably strong performance in Georgian, a good response to a more structured ISO-based prompt, a mainly consistent terminology and lexis. Notably, the most interesting advantage was the minimal human effort required for post-editing in the case of an ISO-based prompt. The distinguishing feature of Grok that deserves mention was the provision of English equivalents and some Georgian synonyms without specific request, which made this information particularly useful. However, this feature was not investigated systematically here. As for the common disadvantages, both of them still made occasional syntactic and morphological errors in Georgian.

As for the Claude Sonnet 4.0 model, one of the striking features of its output was the accuracy of its etymological information for Georgian terms. This information was provided right away without further request. For example, in the case of the equivalent for *swarm drone*, it was highlighted that the term ხროვა/*khrova* (‘pack’) alluded by analogy to the same semantic category as swarm. It should be noted that this feature was not explored in a systematic fashion within this paper. Moreover, this model performed worse with the ISO-based template, perhaps because it required more nuanced information or templates to craft definitions.

The worst performance was given by Gemini Flesh 2.5 and DeepSeek V3-1. The syntactic, contextual, and lexical errors were sometimes so abundant that it became hard to grasp the terms defined. In the majority of cases, one should resort to crafting definitions from scratch rather than putting effort into post-editing the definitions generated by these two models.

Based on this general overview and assessment of models, one could claim that ChatGPT and Grok are the first LLMs that should come to a lexicographer’s/terminologist’s mind for using its assistance. However, it should be mentioned that the main limitation of this study is the lack of reproducibility of the results. Because models are non-deterministic, the outputs given vary (Trap Jensen 2024: 36). In addition, models are frequently updated. Therefore, the results that are true for Gemini 2.5. Flesh are not valid when it comes to, for instance, Gemini 3.0 Thinking versions. As a result, findings cannot be generalized across systems without specifying prompts and model context.

Consequently, answering the fourth research question whether generated draft definitions can be readily used and where a human lexicographer/terminologist stands in this loop we should pay attention to the material we have. Even with the best models in this par-

ticular case study, ChatGPT 5.0 and Grok 3.0 Fast one still needs to validate and linguistically refine definitions when it comes to an ISO-based prompt. However, with the basic prompt, the work becomes time-consuming and laborious. Therefore, this research supports claims made previously by Barret (2023, de Schryver 2023: 361), San Martin (2024), Heinisch (2025) that in-flesh-linguist will still be necessary at both the prompt-creation and, especially, the post-editing phase.

5 Conclusions

This paper explores the effectiveness of using LLMs to assist in crafting draft definitions for Georgian drone-related terminological neologisms. The case study of the selected 10 terms, using two prompting strategies and comparing 5 different models, provided a practical view of how to integrate GenAI tools into terminological work. To the best of our knowledge, no previous study has conducted a comparative analysis of definitions generated by different LLMs, either in general or specifically for Georgian. This study seeks to address that gap by examining and comparing the definitions produced by different models in the context of Georgian terminography and lexicography.

The findings of this case study, on the one hand, indicate that LLMs can generate usable draft definitions for Georgian drone-related terminology, but with significant variation across models and with different prompts. On the other hand, the results demonstrated that all outputs should be checked and manually refined by terminologists, as Georgian syntax errors persist even in Grok and ChatGPT. Thus, GenAI tools should be assigned the role of co-terminologist or AI-assistant, rather than completely replacing human terminologists, especially in specialized domains, emerging terms, or low-resource languages.

While this study provides insights into AI-assisted terminological work, it is important to acknowledge its limitations. First of all, this case study was carried out on a limited set of drone-related terms. The evaluation relied on a self-developed scheme, which was assessed by only one person, potentially increasing bias. Additionally, only freely available LLMs were used. Paid versions or updated models may yield different results, thereby hindering the replicability of results.

Notwithstanding the above-mentioned limitations, this study provides two clear contributions. On the one hand, it makes an applied contribution by drafting terminological definitions for drone-related terminology that will be integrated into the existing multidomain English-Georgian dictionary[9]. On the other hand, it makes a methodological contribution, by assessing the effectiveness and potential integration of Generative AI for under-resourced languages, such as Georgian, and specialized terminology. This work also fosters the idea that chatbots should be used as a supporting tool and not as a replacement for human editors, since critical evaluation and validation of outputs are essential.

Future research should focus on a larger set of Georgian terms from different domains, particularly those formed using exclusively Georgian resources rather than borrowings or structural calques. In addition, it is advisable to have several terminologists to assess the

9 Available at: https://multiterm.iliauni.edu.ge (accessed 21.02.2026)

generated outputs against the existing evaluation scheme, as inter-annotator agreement can enhance the validity of the results. Further work may also focus on using different templates and prompt-generating strategies to yield better outputs.

Overall, we anticipate that research findings will contribute to the use of large language models in assisting with various terminological tasks, particularly in creating draft definitions in the Georgian language.

References

Anthropic Claude. 2025. Claude Response to Keti Mchedlishvili. 3 September. https://claude.ai/share/64cea118-86d3-4250-9015-3a5e708d69ef [accessed 21/02/2026]

Barrett, G. (2023). 'Defin-O-Bots: Challenging A.I. to Create Usable Dictionary Content.' Paper presented at the *24th Biennial Conference of the Dictionary Society of North America*. *Boulder, CO, USA, 31 May – 3 June 2023.*

Benko, V. (2024). The Aranea Corpora Family: Ten+ Years of Processing Web-Crawled Data. Nöth, E., A. Horák, and P. Sojka (Eds.). 2024. *Text, Speech, and Dialogue. TSD 2024. Lecture Notes in Computer Science:* 1-16. Cham: Springer. https://doi.org/10.1007/978-3-031-70563-2_5 [accessed 21/02/2026]

Brezina, V., P. Weill-Tessier & T. McEnery 2021. #LancsBox 6.x [software]. Available at: http://corpora.lancs.ac.uk/lancsbox. [accessed 21/02/2026]

Cabezas-García, M. (2024). *The Pragmatics of Multiword Terms. The Impact of Context*. New York: Routledge. DOI: https://doi.org/10.4324/9781003390091 [accessed 21/02/2026)

Cabré, M.T. (1999). Terminology: Theory, Methods and Applications. Amsterdam/ Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/tlrp.1. [accessed 21/02/2026]

Daraselia, S. (2015). *Issues of Compilation of New Georgian-European Learner's Dictionary Using the Corpus Methodology*. Unpublished Ph.D. thesis. Tbilisi: Ivane Javakhishvili Tbilisi State University. (In Georgian) https://cqpweb.lancs.ac.uk/kawac200mw/ [accessed 21/02/2026]

de Schryver, G.-M. & D. Joffe. (2023). 'The End of Lexicography, Welcome to the Machine: On How ChatGPT Can Already Take over All of the Dictionary Maker's Tasks.' Paper presented at the *20th CODH Seminar. Center for Open Data in the Humanities, Research Organization of Information and Systems, National Institute of Informatics, Tokyo, Japan, 27 February 2023*. https:// youtu.be/watch?v=mEorw0yefAs [accessed 21.02.2026]

de Schryver, G.-M. (2023). Generative AI and Lexicography: The Current State of the Art Us ing ChatGPT. In International Journal of Lexicography, 36(4), pp. 355-387. https://doi.org/10.1093/ijl/ecad021 [accessed 06/02/2026]

de Schryver, G.-M. (2025). Customised GPTs for Lexicography. In *Lexikos*, *35*(2), pp. 397-422. https://doi.org/10.5788/35-2-2104 [accessed 06/02/2026]

Ďurčo, P. Multiword Expressions in Comparable Corpora. Pastor G. C and J. Colson (Eds.). 2020. *Computational Phraseology*: 136-150. Amsterdam/Philadelphia: John Benjamin's Publishing Company. DOI: https://doi.org/10.1075/ivitra.24 [accessed 21/02/2026]

English-Georgian Military Dictionary. https://mil.dict.ge/en/ [accessed 21/02/2026]

Georgieva, V. (2015). *Military English: from Theory to Practice.* Sofia: Georgi Stoikov Rakovski National Defence College https://www.researchgate.net/publication/363582336_MILITARY_ENGLISH_FROM_THEORY_TO_PRACTICE [accessed 21/02/2026]

Google Gemini. 2025. Gemini Response to Keti Mchedlishvili. 3 September. https://gemini.google.com/app/4774657b6e333c9c [accessed 21/02/2026]

Heinisch, B. (2025). Next-gen Terminology: Transforming Terminology Work with Large Language Models. In *Across Languages and Cultures*, 26 (S), pp. 64-80. DOI: 10.1556/084.2025.01061. [accessed 11/02/2026]

High-Flyer. DeepSeek. 2025. DeepSeek Response to Keti Mchedlishvili, 3 September. https://chat.deepseek.com/share/8wi9r7dj36cs8lenx4 [accessed 21/02/2026]

ISO 1087: 2019. (2019). Terminology Work and Terminology Science – Vocabulary. Geneva: ISO

ISO 704:2022. Terminology work — Principles and methods. Geneva: ISO

Jakubíček, M. & M. Rundell. (2023). 'The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-Editing Lexicography?' In Medveď, Marek, Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas and Miloš Jakubíček (eds), *Proceedings of the eLex 2023 Conference: Electronic Lexicography in the 21st Century*. Brno: Lexical Computing, 508–523. https://elex.link/elex2023/wp-content/uploads/102.pdf [accessed 21/02/2026]

Lew, R. (2023). ChatGPT as a COBUILD Lexicographer. *Humanities and Social Science Communications, 10(1), 704*, pp 1-10. https://doi.org/10.1057/s41599-023- 02119-6 [accessed 21/02/2026]

Lomidze, K. (2025). AI in Lexicography for Low-Resource Languages: The Case of the Georgian Language. *In II International Conference Lexicography in the XXI century, book of abstracts*. Tbilisi: Ilia State University, pp. 97-101. https://lexicography21.iliauni.edu.ge/wp-content/uploads/2025/10/Book-of-Abstracts-2025.pdf [accessed 21/02/2026]

Łukasik, M. (2023). Corpus linguistics and Generative AI Tools in Term Extraction: a Case of Kashubian – a Low-resource Language. In *Applied Linguistics Papers* 27 (4), pp. 34-45. DOI: 10.32612/uw.25449354.2023.4.pp.34-45 [accessed 11/02/2026]

Mchedlishvili, K. (2024). Comparable Corpora: Compilation Methods and Areas of Application. *Kadmos. A Journal of the Humanities*, (16), 237–253. https://doi.org/10.32859/kadmos/16/237-253 (In Georgian) [accessed 21/02/2026]

Mchedlishvili, K. (2025). Harnessing AI for Creating Definitions of Georgian Military Terminological Neologisms in the context of the Russia-Ukraine War. *In II International Conference Lexicography in the XXI century, book of abstracts*. Tbilisi: Ilia State Univer-

sity, pp. 112-116. https://lexicography21.iliauni.edu.ge/wp-content/uploads/2025/10/Book-of-Abstracts-2025.pdf [accessed 21/02/2026]

McKean, E & W. Fitzgerald. (2023). 'The ROI of AI in Lexicography' *Proceedings of the 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users".* Seoul: Yonsei University, 10–20. DOI:10.1558/lexi.27569 [accessed 21/02/2026]

Medzmariashvili, E. (Ed.). 2017. *Georgian Military Encyclopedic Dictionary*. Tbilisi. (In Georgian)

Military-Explanatory Dictionary and Graphic Symbols. FM (Field Manual) 1-02. (2017). Tbilisi: e Ministry of Defense of Georgia, Command of Training and Military Education, Doctrine Development Center. (In Georgian)

Nunzio, G.M., Gallina, E., & Vezzani, F. (2024). UNIPD@SimpleText2024: A Semi-Manual Approach on Prompting ChatGPT for Extracting Terms and Write Terminological Definitions. In G. Faggioli, N. Ferro, P. Galuscáková, A. García Seco de Herrera (eds.) *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9–12 September 2024*. CEUR Workshop Proceedings 3740. CEUR-WS, pp. 3230–3237. DOI: 2-s2.0-85201605270 [accessed 21/02/2026]

OpenAI ChatGPT. 2025. ChatGPT Response to Keti Mchedlishvili, 3 September. https://chatgpt.com/share/e/6999bb7b-47c0-8011-a5d9-809019d9eced [accessed 21/02/2026]

Rundell, M. & A. Kilgarriff. (2011). 'Automating the Creation of Dictionaries. Where Will It All End?' In Meunier, Fanny, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot (eds), *A Taste for Corpora*. *In Honour of Sylviane Granger. In Studies in Corpus Linguistics 4,* pp. 257-281 Amsterdam: John Benjamins. https://doi.org/10.1075/SCL.45.15RUN [accessed 21/02/2026]

Rundell, M. (2023). 'Automating the Creation of Dictionaries: Are We Nearly There?' *Proceedings of the 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users"*. Seoul: Yonsei University, 1–9. https://www.hltmag.co.uk/feb24/automating-the-creation-of-dictionaries [accessed 21/02/2026]

Sabanés, L.V. & Da Cunha I. (2025). AI as a Resource for the Clarication of Medical Terminology. An Analysis of its Advantages and Limitations. In *Terminology, Computational terminology, 31(1)*, pp. 37-71. https://doi.org/10.1075/ term.00083.vid [accessed 21/02/2026]

San Martín, A. (2024). What Generative Artificial Intelligence Means for Terminological Definitions. In F. Vezzani, G. M. Di Nunzio, B. Sánchez-Cárdenas, P. Faber, M. Cabezas García, P. León Araúz, A. Reimerink & A. San Martín (eds.) In *Proceedings of the 3rd International Conference on Multilingual Digital Terminology Today (MDTT 2024), Granada, Spain, 27–28 June 2024*. Granada: CEUR-WS, pp. 1-37. Available at: https://ceur-ws.org/Vol-3703/paper1.pdf [accessed 21/02/2026]

Sorge, H. (2023). War in Ukraine: Unleashing Innovation and Unseen Frontiers. Policy Center of the New South. https://www.policycenter.ma/publications/war-ukraine-unleashing-innovation-and-unseen-frontiers [accessed 21/02/2026]

Struk I. V, T. H. Semyhinivska & A. V. Sitko. (2022). Formation and Translation of Military Terminology. *Scientific notes of V. I. Vernadsky Taurida National University, Series: Philology. Journalism Vol. 2. No. 2:* 29-33. http://www.philol.vernadskyjournals.in.ua/journals/2022/2_2022/part_2/5.pdf [accessed 21/02/2026]

Tchighladze, K. (2025). Integrating AI into the Term Creation Process: Case Study of English and Georgian Terms of Emerging Technologies. *In II International Conference Lexicography in the XXI century, book of abstracts*. Tbilisi: Ilia State University, pp. 149-155. https://lexicography21.iliauni.edu.ge/wp-content/uploads/2025/10/Book-of-Abstracts-2025.pdf [accessed 21/02/2026]

Trap-Jensen, L. (2024). The Best of Two Worlds: Exploring the Synergy between Human Expertise and AI in Lexicography. In T. Margalitadze (eds.) *Proceedings of the I International Conference Lexicography in the XXI Century, Tbilisi, 10-12 November 2023.* Tbilisi: Centre for Lexicography and Language Technologies Ilia State University, pp. 26-39. Available at: https://lexicography21. iliauni.edu.ge/wp-content/uploads/2024/06/03_Lars-Trap-Jensen.pdf. [accessed 21/02/2026]

Vicente, L. S. (2012). Approaching Secondary Term Formation Through the Analysis of Multiword Units: An English–Spanish Contrastive Study. Neology in Specialized CommunicationTerminology. *International Journal of Theoretical and Applied Issues in Specialized Communication*. *No. 18(1)*: 105-127, John Benjamins Publishing Company. DOI: 10.1075/term.18.1.06san [accessed 21/02/2026]

X Grok. 2025. Grok Response to Keti Mchedlishvili. 3 September. https://grok.com/share/c2hhcmQtNA_3060c25e-5479-44a0-be4a-59479efeffda [accessed 21/02/2026]