Using LLMs to Extract Instances of Schematic Constructions from Unannotated L2 Learner Corpora

Jelena Kallas[1], Ahto Kiil[1,2], Heete Sahkai[1], Geda Paulsen[1,3], Kertu Saul[1,2]
Institute of the Estonian Language[1], University of Tartu[2], Uppsala University[3]
jelena.kallas@eki.ee, ahto.kiil@gmail.com, heete.sahkai@eki.ee, geda.paulsen@eki.ee, kertu.saul@eki.ee

Abstract

Our previous study (Sahkai et al., 2025) found that generative LLMs can be successfully used to identify instances of schematic constructions (as defined in Construction Grammar) in unannotated L1 corpus data. This study tests the applicability of LLMs to also identify instances of constructions in unannotated L2 data. L2 learner corpora are notoriously difficult to annotate and query since they contain errors. Using LLMs can thus simplify the retrieval of construction data from L2 corpora. The identification of instances of constructions in L2 learner data has many potential uses in pedagogical applications of Construction Grammar and constructicography, such as identifying error-prone (properties of) constructions and the distribution of constructional instances across CEFR levels. Using the Estonian Nominal Quantifier Construction as the example construction and an Estonian CEFR-graded learner corpus as the source of L2 data, we tested several prompts and several models (OpenAI's o3-mini, o3, gpt-5-mini and gpt-5, Google DeepMind's Gemini Flash 2.5, Anthropic's Claude Sonnet 4.5 and Opus 4.1). We found that the best model, gpt-5, achieved F1-scores from 0.90 to 0.96, depending on the level of detail of the prompt.

Keywords: Large Language Models, Estonian, L2 learner corpora, Pedagogical Construction Grammar

1. Introduction

Pedagogical Construction Grammar (PCG) is a promising approach in L2 instruction, e.g., Littlemore (2009); De Knop and Mollica (2016); Herbst (2016); Boas (2022); Lyngfelt et al. (2026). According to PCG, what should be taught and learned are constructions as defined in Construction Grammar: lexically filled or (partially) schematic pairings of form and meaning (see, e.g., Goldberg (1995); Hoffmann (2022)).

In order to develop construction-based pedagogical resources, it is necessary to know at which level of the Common European Framework of Reference for Languages (CEFR) a particular construction or construction-collexeme pairing (in the sense of Stefanowitsch and Gries (2003)) should be taught and which (properties of) constructions present difficulties to learners. This information can be gained from the L2 learner corpora by identifying the instances of a construction to analyze their distribution across CEFR levels and the errors that learners make when using the construction.

However, before learner corpora can be effectively used for research, they require substantial manual and/or automatic annotation work and the development of dedicated search tools. Moreover, as Volodina (2024, pp. 84–86) notes, despite the relatively long history of Learner Corpus Research, there is still no agreement on what should be considered an ideal standard. This concerns decisions about which types of annotation to include, which standards and data formats to adopt, what metadata to provide, and which search tools to develop. Pérez-Paredes et al. (2025, pp. 100–116) illustrate that even with advanced tools that support complex corpus queries and regular expressions, the identification of grammatical patterns — including colligations, constructions, and collostructions — still requires considerable manual analysis.

Hence, an alternative approach is needed to facilitate the exploration of learner data and enable complex searches that can retrieve grammatical constructions from learner corpora, while simultaneously reducing the time and labour resources required for corpus annotation. In this study, we will examine whether LLMs can be used to automate the identification of instances of schematic constructions in L2 learner data, aiming to answer two research questions:

RQ 1: To what extent can LLMs identify instances of constructions in L2 learner data without morphosyntactic and error annotation, i.e., data that include orthographically and grammatically non-target-like instances?

RQ 2: What is the optimal prompt and the minimal amount of fine-tuning data needed for models to yield relevant results? By optimal prompt, we mean the shortest prompt that yields satisfactory results and is easily adaptable to different constructions.

As the test construction, we use the Estonian Nominal Quantifier Construction (NQC) (Metslang (2017); Pilvik et al. (2025)) – a schematic (pseudo-) partitive construction (Koptjevskaja-Tamm (2001)) consisting of two noun slots. The first noun functions as a quantifier, while the second noun denotes, in partitive use, the referent quantified over (e.g., *enamik selle rühma keeli* [majority.sg.nom this.gen group.sg.gen language.pl.par] 'the major-

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 10517–10524
11–16 May 2026.

ity of the languages of this group'), or, in pseudo-partitive use, the kind of entity quantified over (e.g., *hulk raha* [amount.sg.nom money.sg.par] 'a lot of money'). The construction shows complex inflectional behaviour, complicating corpus retrieval and presenting difficulties for L2 learners.

The data source used in the experiment is the Estonian L2 learner corpus (Institute of the Estonian Language, 2025), henceforth the EMMA_EKI. The corpus contains assessment test and exam materials administered by the Estonian Education and Youth Board — a government agency under the Ministry of Education and Research. The materials have been collected and prepared for research purposes at the Institute of the Estonian Language, including pseudonymisation and morphosyntactic parsing with the UDPipe parser with the UDPipe parser (estonian-edt-ud-2.15-241121)[1]. The corpus comprises 12,076 Estonian L2 texts (133,071 sentences), produced predominantly by learners whose first language is Russian. The materials cover the period 2016–2022 and are organized by school grades (3, 6, 7, 9, and 12), which, as established in the National Curriculum for Basic Schools (Government of the Republic of Estonia, 2013) and the National Curriculum for Upper Secondary Schools (Government of the Republic of Estonia, 2023), correspond to CEFR levels A2, B1, B2, and C1, respectively. Table 1 presents the division of sentences by class.

Class	Sentences
3	2,312
6	14,263
7	10,207
9	59,222
12	47,067
Total	**133,071**

Table 1: Distribution of sentences across school classes in the EMMA_EKI corpus.

The paper is structured as follows. Section 2 presents the research background. Section 3 describes the procedures undertaken. Section 4 reports and discusses the results, including the potential applications of the proposed method. Section 5 summarises the findings and outlines future directions.

2. Background

Identification of instances of schematic constructions is difficult not only in L2 data but also in L1 corpora, since constructions are pairings of form and meaning and usually cannot be identified by formal criteria alone. Traditionally, the corpus retrieval of constructional instances involves retrieving all instances corresponding to certain morphosyntactic structures, using Corpus Query Language or regular expressions, followed by manual inspection of the concordance lines (see Pérez-Paredes et al. (2025, pp. 100–116)). This approach is labour-intensive, especially when the corpus is large, the construction is frequent, or a large number of constructions need to be described, e.g., for a constructicographic resource. Therefore, in Sahkai et al., 2025, we tested the usability of LLMs to simplify this task, although a previous study suggested that LLMs do not acquire representations of fully schematic constructions, as they have access only to the lower levels of the constructional hierarchy (Bonial and Tayyar Madabushi (2024). We tested both BERT-based models and generative LLMs, using the NQC as the test set and varying the amount of fine-tuning data. The most promising result of the study was the ability of generative LLMs to identify constructional instances in unannotated L1 corpus data with few-shot fine-tuning: with only 10 positive and 5 negative example sentences, o3-mini achieved a recall of 0.89, which was comparable to the benchmark EstRoBERTa model[2] trained on 8,500 positive and 17,000 negative examples based on manually verified corpus data (recall = 0.92).

3. Prompt design and evaluation

To answer the research questions, we tested different language models using prompts with varying levels of detail. Here, we describe the process of prompt design. After testing each prompt, an evaluation was conducted, and a new prompt was created. The section starts by describing the test dataset (3.1), then proceeds to model selection (3.2), and finally describes the iterative prompt modification and evaluation process (3.3).

3.1. Creating the test dataset

To focus on the ability of the models to identify constructions as form–meaning pairings, we limited the test set to sentences where the corresponding morphological sequence already existed. We first extracted the sentences containing the morphological tag sequence corresponding to the NQC, i.e., a noun in singular or plural nominative, genitive, or partitive followed by a noun in singular or plural partitive.

The results of the preselection process are presented in Table 2.

[1] https://ufal.mff.cuni.cz/udpipe/2/models

[2] https://huggingface.co/tartuNLP/EstRoBERTa

Class	Candidates of NQC
3	–
6	144
7	170
9	1,940
12	1,418
Total	**3,672**

Table 2: Number of NQC candidates by school class.

We chose the Class 12 candidate sentences as the test dataset. Two annotators, both linguists from the Institute of the Estonian Language, independently reviewed 1,418 candidate sentences from the Grade 12 subcorpus and identified 146 valid instances of the NQC. The annotators agreed on 1,412 of 1,418 instances ($P_o = 0.9958$). Cohen's κ was 0.977 ($P_e = 0.8186$), indicating almost complete agreement (Cohen (1960)). Disagreement was limited to six cases involving the classification of certain lexical items as quantifiers within the NQC (e.g., *enamik* and *enamus*, 'the majority'). As we plan to use this data to evaluate the acquisition of NQC in future, we decided to exclude all misspelled and ungrammatical cases; still, we retained NQCs that were used ungrammatically in context but had a grammatically correct internal structure. The test set sentences were unannotated.

3.2. Selection of models

On our test dataset, we evaluated OpenAI's[3] o3-mini, o3, gpt-5-mini and gpt-5, Google DeepMind's Gemini Flash 2.5[4], and Anthropic's Claude Sonnet 4.5 and Opus 4.1[5], using the prompt from our previous study (see Section 3.3.1; henceforth, the baseline prompt), which performed successfully on L1 data. The models were selected based on their availability and ranking on LLM Arena[6]; o3-mini was used in our previous study and was retained as the reference model.

We tested the o3-mini model three times on three separate days using the same prompt[7] to demonstrate that, even if the results are not identical, performance in terms of recall does not differ significantly. The results are presented in Table 3.

We then discarded all other models except o3-mini (the most cost-effective) and gpt-5 (the best-performing).

[3]https://openai.com/

[4]https://deepmind.google/models/gemini/flash/

[5]https://www.anthropic.com/news/claude-opus-4-1

[6]https://lmarena.ai/leaderboard/text

[7]Note that the o3-mini reasoning model ignores the hyperparameter `temperature=0`.

Model	Precision	Recall	F1
sonnet	0.7226	0.6781	0.6996
opus	0.8686	0.8151	0.8410
gemini	0.8767	0.8767	0.8767
o3	0.9034	0.8973	0.9003
o3-mini-1	0.9167	0.9041	0.9101
o3-mini-3	0.8210	0.9110	0.8636
o3-mini-2	0.9160	0.9178	0.9147
gpt5-mini	0.8182	0.9247	0.8682
gpt-5	0.9034	0.9589	0.9302

Table 3: Performance of the seven models on the baseline prompt.

3.3. Iterative testing and evaluation of the two best models with increasingly shorter prompts

The prompting was based on the unannotated dataset consisting of 1,418 sentences (see Section 3.1) and the selected models, o3-mini and gpt-5 (see Section 3.2). The sentences were prompted consecutively via the respective APIs.

Below, we describe the development of the prompts and the evaluation stages.

3.3.1. Baseline prompt

The baseline prompt was compiled based on the following consideration: to identify as many instances of the NQC in the corpus as possible. The baseline prompt task, together with the nine definitions and rules, is provided below, excluding the example sentences ($N = 15$), which are presented in Appendix 1. The example sentences were of three types: adjacent (the quantifier noun and the noun denoting the quantified entity or referent are adjacent) ($N = 5$), intervening (modifiers of the second noun intervene between the two nouns) ($N = 5$), and no NQC ($N = 5$).

Task. In each input sentence, find the **first** quantifier construction that satisfies all rules 1–9 below and output it in the format:`<quantifier-phrase>;<space><original sentence>` If the sentence contains **no** valid quantifier construction, output: `NO_QUANT;<space><original sentence>`

Definitions and rules.

1. A quantifier construction consists of **exactly two words**.
2. **Word 1** is a noun in nominative, genitive, or partitive (sg./pl.) expressing a measurable quantity.
3. **Word 2** is a noun in **partitive** (sg./pl.) denoting the quantified entity.
4. The words may be adjacent or separated.
5. Word 2 is the syntactic **head** (parent) of Word 1.
6. Word 1 always **precedes** Word 2.
7. Exclude numeral-like words (e.g., *miljon*, *tuhat*). Units such as *gramm*, *kilo*, *meeter*, *liiter*, *tonn* are valid.

8. Pure numerals (e.g., 1, 12; *üks*, *kaksteist*) are **not** quantifiers.
9. Adjectives (e.g., *pikk*, *suur*) and adverbs (e.g., *mitu*, *palju*) are **not** quantifiers.

3.3.2. Extended prompt

In the extended prompt, the rules and stopwords were updated based on the findings from the baseline prompt responses. At this stage, we adjusted the rules to enhance the baseline prompt with the aim of identifying as many NQC cases as possible in the dataset. During the manual analysis of correct NQC instances, we discovered that the Estonian word *paar* ('pair, couple') requires a more careful treatment in both the training and test sets, as it can function both as a nominal quantifier and as a numeral.

We improved the baseline prompt by adding a role specification ('information-extraction system for Estonian quantifier expressions'), by adding problematic stopwords to the EXCLUDED list (*pool* 'half', *paar* 'a couple', *veerand* 'quarter'), and by shortening the few-shot examples and including edge cases (*paar tundi* 'couple of hours', *kaks kitse* 'two goats', *tüüpi inimesi* 'type of people').

We also added the word 'ONLY' to Rule 3 ('NOUN ONLY in Partitive (sg./pl.) that denotes the thing being quantified'), as the generative model tended to include other morphological forms (mostly nominative and genitive).

Consider the extended prompt below: the task with nine definitions and rules, excluding the example sentences.

Task. Given Estonian sentence(s), for each sentence find the **first** (leftmost) quantifier construction that satisfies all rules and return exactly ONE line: <two-word-quantifier>;<space><original sentence>. If no valid construction exists for a sentence, return: NO_QUANT;<space><original sentence>.

Definitions and rules.

1. Exactly two words in the construction.
2. Word 1 = NOUN in Nom/Gen/Par (sg./pl.) denoting a measurable quantity (time, amount, size, etc.).
3. Word 2 = NOUN ONLY in Partitive (sg./pl.) denoting the thing being quantified.
4. The words may be adjacent or separated by other words.
5. Morphosyntactically, Word 2 is the parent (head) of Word 1.
6. Word 1 precedes Word 2 in the sentence.
7. EXCLUDE: pure numerals (1, 12, ...), written numerals (*kolm*, *neliteist*, ...), numeral-like words that are not general quantifiers (*pool*, *paar*, *veerand*, *miljon*, *tuhat*, *sadakond*, *mustmiljon*, ...), adjectives (*pikk*, *suur*, ...), and adverbs (*mitu*, *palju*, *natuke*, ...).
8. INCLUDE: unit nouns such as *gramm*, *kilo*, *meeter*, *liiter*, *tonn*, *hektar*, *sekund*, *aasta* (and their inflected forms).
9. If a valid construction exists, do not output `NO_QUANT` for that sentence.

The example sentences are given in Appendix 2.

3.3.3. Reduced prompt

After achieving 99.32% recall and a 96.03% F1-score with OpenAI's gpt-5 using the extended prompt (cf. Table 4 in Section 4), we reduced the number of sample sentences, leaving only one example of each case (adjacent, intervening, no NQC), to measure the impact of the number of shots on the performance metrics. The task, definitions, and rules remained the same as in the previous prompts. The example sentences were as follows:

> *grupp tegelasi*; Uus *grupp tegelasi* haarab telekaesise oma valdusse.
> *hetk elu*; Ta oli valmis kõike tegema, et kinkida vaatajale *hetk tõelist elu*.
> `NO_QUANT`; Paar tundi hiljem jõudis lõpuks järg ka meie kätte.

3.3.4. Minimal prompt

The next step, after reducing the number of example sentences, was to pare the rules down to a bare minimum. With the best model's (gpt-5) performance at 96.58% recall and 92.16% F1-score, we removed all stopwords from the prompt, leaving only the rules and conditions of the construction and three examples. The prompt was reduced to the following:

Definitions and rules.

1. Exactly two words in the construction.
2. Word 1 = NOUN in Nom/Gen/Par (sg./pl.) denoting a measurable quantity (time, amount, size, etc.).
3. Word 2 = NOUN ONLY in Partitive (sg./pl.) denoting the thing being quantified.
4. The words may be adjacent or separated by other words.
5. Morphosyntactically, Word 2 is the parent (head) of Word 1.
6. Word 1 precedes Word 2 in the sentence.
7. EXCLUDE: pure numerals, written numerals, numeral-like words that are not general quantifiers.
8. INCLUDE: unit nouns (+ inflections).
9. If a valid construction exists, do not output `NO_QUANT` for that sentence.

3.3.5. No-shot prompt

To optimise the prompt even further, the final evaluation was conducted using only the NQC rule descriptions as in Section 3.3.4, and no sample sentences were provided to the model.

4. Results and discussion

Here we present the performance of the two selected models, o3-mini and gpt-5, across four prompts (extended, reduced, minimal, and no-shot; see Sections 3.3.2–3.3.5). The evaluated metrics are precision, recall, and F1-score.

The performance of the extended prompt on the test dataset (see Section 3.3.1) is shown in Table 4.

Model	Precision	Recall	F1
o3-mini	0.8933	0.9178	0.9054
gpt-5	0.9295	0.9932	0.9603

Table 4: Performance of o3-mini and gpt-5 on the extended prompt.

The results with the reduced prompt (three sample sentences) are shown in Table 5.

Model	Precision	Recall	F1
o3-mini	0.8322	0.8493	0.8407
gpt-5	0.8813	0.9658	0.9216

Table 5: Performance of o3-mini and gpt-5 on the reduced prompt.

After reducing the number of example sentences from 15 to 3, we expected lower performance; however, the recall for gpt-5 remains very high. The increase in false positives (from 11 to 19, reflected in the decrease in precision from 0.9295 to 0.8813) remains manageable for subsequent manual confirmation and decision-making.

In the next test, we stripped down the prompt to the bare minimum of rules (no stopwords), while keeping the same three examples. The performance metrics are shown in Table 6.

Model	Precision	Recall	F1
o3-mini	0.7987	0.8425	0.8200
gpt-5	0.8968	0.9521	0.9236

Table 6: Performance of o3-mini and gpt-5 on the minimal prompt.

The gpt-5 model identified two fewer NQC instances but was also able to reduce the number of false positives by three. As a result, the F1-score was marginally higher with the minimal prompt compared to the reduced prompt (0.9236 vs. 0.9216).

The no-shot prompt, consisting only of the bare rule descriptions, still demonstrated good recall in identifying NQC instances. The results are shown in Table 7.

Both o3-mini and gpt-5 successfully identified not only the most obvious cases of the NQC (e.g., *grupp inimesi* 'a group of people', *osa õpilasi* 'a part of the students'), but also more complex cases in which the first noun is not inherently a quantifier (e.g., *arv eriarvamusi* 'a number of differing opinions', *valik asju* 'a selection of things').

Model	Precision	Recall	F1
o3-mini	0.7959	0.8014	0.7986
gpt-5	0.9214	0.8836	0.9021

Table 7: Performance of o3-mini and gpt-5 on the no-shot prompt.

Of all tested prompts, the extended prompt achieved the highest results; however, creating such a prompt requires a relatively large set of example sentences and is therefore more costly. Overall, we consider the minimal prompt to be optimal, as it uses only three example sentences and a reduced set of rules while maintaining high performance.

Error analysis of the minimal prompt's results showed that most false positives (75% for gpt-5 and 35.5% for o3-mini) could likely be attributed to learner errors in the texts. These errors, such as dividing compounds into two separate words, or using the wrong case, caused the phrase's form to coincide with that of an NQC. Most false negatives of both models (71.4% for gpt-5 and 61.1% for o3-mini) occurred in cases where the first member of the NQC formed part of another construction, e.g., *neli päeva tegutsemist* 'four days of activity'.

Our first research question concerned the extent to which generative LLMs can identify instances of schematic constructions in unannotated L2 learner data. The results show that the number of NQCs extracted from our dataset, i.e., the highest recall, approaches the maximum (0.9932 for gpt-5 with the extended prompt). This value remained relatively high even for the lowest result (0.8014 for o3-mini with the no-shot prompt).

Our second research question was: what is the minimal prompt that still yields relevant results. The results showed that even the no-shot prompt without any example sentences and the rules reduced to a minimum, performs relatively satisfactorily. The minimal prompt with only three examples can therefore be considered optimal. The principle of laconic rules combined with a few-shot approach is easily adaptable to other (semi-)schematic constructions; however, performance should be tested on different types of constructions.

Our results do not support the claim by Bonial and Tayyar Madabushi (2024) that schematic constructions are not included in the linguistic knowledge of LLMs. The findings demonstrate that strong performance can be achieved using few-shot prompting, which aligns with the results reported by Brown et al. (2020).

4.1. Possible applications of the method

The identification of instances of a construction in L2 learner corpora, their distribution across CEFR levels, and the errors learners make when using the construction have several possible applications within the framework of Pedagogical Construction Grammar.

Pedagogically oriented constructicography. Our primary motivation for developing the proposed method was the needs of pedagogically oriented constructicography, which aims to cover a large number of constructions based on corpus data. Constructicography is an emerging field that aims to represent grammatical constructions in a lexicographic format. Its main output is a constructicon, a resource that describes various types of constructions associated with specific meanings or functions, ranging from fully schematic and semi-schematic constructions to specific lexical expressions (Borin and Lyngfelt (2025); Lyngfelt et al. (2018); Patel et al. (2023)). Some of these constructicons are explicitly designed for pedagogical purposes, aiming to support language learning, teaching, and assessment, e.g., Janda et al. (2020). A key type of information to be incorporated into pedagogically oriented constructicons concerns CEFR levels. This includes, for instance, the CEFR level at which a particular construction is first acquired or should be taught, the development of the construction across CEFR levels, as well as the distribution of particular construction–word pairings across CEFR levels. The proposed method facilitates the identification of such information in learner corpora. As a next step, this information makes it possible to develop a new type of educational resource, construxercises, which enable hands-on learning of constructions at different CEFR levels (see, e.g., Endresen et al. (2022)).

Compilation of construction-based L2 instructional materials. The distribution of constructional instances across CEFR levels, as well as information about learners' errors, is also necessary for the compilation of teaching materials and grammar profiles, and for data-driven learning.

Collostructional analysis. Corpus retrieval of constructional instances is also necessary for collostructional analysis (Stefanowitsch & Gries, 2003), which allows the identification of the most productive constructions in learner data and their most strongly associated collexemes.

5. Conclusion and next steps

This study showed that generative LLMs can be successfully used to identify instances of a schematic construction in unannotated L2 learner data. The optimal prompt, in terms of level of detail and number of examples, consisted of three rules (no stopwords) and three shots, and achieved a precision of 0.8968 and a recall of 0.9521 with the gpt-5 model.

We conclude that LLMs can successfully retrieve occurrences of a schematic construction and that the models can learn from a small number of examples—as our study shows, three examples may be sufficient. The rules and definitions can be laconic, as long as they include the central criteria and restrictions. The method can be applied, for example, in pedagogical constructicography and for analysing learner errors within a construction-based framework.

At the same time, the present study should be understood as a proof of concept rather than a generalized solution for construction extraction. Broader multilingual validation and more detailed error analysis are required to support stronger theoretical claims.

Our next steps will continue to explore the potential of large language models to retrieve corpus instances of constructions, focusing on different types of constructions, larger sets of corpus data, and the development of a user-friendly interface to integrate language models into a constructicographic workflow.

6. Acknowledgements

This work was supported by the Estonian Research Council grant (PRG 1978).

7. Bibliographical References

Hans C. Boas. 2022. From construction grammar(s) to pedagogical construction grammar. In Hans C. Boas, editor, *Directions for Pedagogical Construction Grammar*, pages 3–44. De Gruyter.

Claire Bonial and Harish Tayyar Madabushi. 2024. Constructing understanding: On the constructional information encoded in large language models. *Language Resources and Evaluation*.

Lars Borin and Benjamin Lyngfelt. 2025. Framenets and constructicons. In Mirjam Fried and Kiki Nikiforidou, editors, *The Cambridge Handbook of Construction Grammar*, pages 71–100. Cambridge University Press.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, et al. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901.

Jakob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(2):138–147.

Sabine De Knop and Fabio Mollica. 2016. A construction-based analysis of german ditransitive phraseologisms for language pedagogy. In Sabine De Knop and Gaëtanelle Gilquin, editors, *Applied Construction Grammar*, pages 53–88. De Gruyter, Berlin.

Anna Endresen, Valentina Zhukova, Elena Bjørgve, Daria Demidova, Natalia Kalanova, Zoia Butenko, George Lonshakov, and David Henrik Lavén. 2022. Construxercise! implementation of a construction-based approach to language pedagogy. *Russian Language Journal*, 72(1).

Adele E. Goldberg. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago and London.

Government of the Republic of Estonia. 2013. National curriculum for basic schools. Riigi Teataja. `https://www.riigiteataja.ee/aktilisa/1080/3202/3001/18m_pohi_lisa1.pdf` Accessed 25 February 2026.

Government of the Republic of Estonia. 2023. National curriculum for upper secondary schools. Riigi Teataja. `https://https://www.riigiteataja.ee/aktilisa/1250/6202/5013/18m_gym_lisa1.pdf` Accessed 25 February 2026.

Thomas Herbst. 2016. Foreign language learning is construction learning – what else? moving towards pedagogical construction grammar. In Sabine De Knop and Gaëtanelle Gilquin, editors, *Applied Construction Grammar*, pages 21–52. De Gruyter, Berlin.

Thomas Hoffmann. 2022. *Construction Grammar*. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge.

Laura A. Janda, Anna Endresen, Valentina Zhukova, Daria Mordashova, and Ekaterina Rakhilina. 2020. How to build a constructicon in five years: The russian example. *Belgian Journal of Linguistics*, 34(1):161–173.

Maria Koptjevskaja-Tamm. 2001. Partitive and pseudo-partitive nominal constructions in the circum-baltic languages. In Östen Dahl and Maria Koptjevskaja-Tamm, editors, *Circum-Baltic Languages: Volume 2: Grammar and Typology*, pages 523–568. John Benjamins Publishing Company.

Jeannette Littlemore. 2009. *Applying Cognitive Linguistics to Second Language Learning and Teaching*. Palgrave Macmillan, New York.

Benjamin Lyngfelt, Lars Borin, Kyoko Ohara, and Tiago Timponi Torrent, editors. 2018. *Constructicography: Constructicon Development across Languages*, volume 22. John Benjamins Publishing Company.

Benjamin Lyngfelt, Julia Prentice, and Azizah Lenté Degez. 2026. Constructicons as resources for language pedagogy – and vice versa. *Pedagogical Linguistics*, 7(1):38–66.

Helle Metslang. 2017. Kvantorifraas. In Mati Erelt and Helle Metslang, editors, *Eesti keele süntaks*, volume 3 of *Eesti keele varamu*, pages 463–478. Tartu Ülikooli Kirjastus, Tartu.

Malin Patel, Armine Garibyan, Elodie Winckel, and Stephanie Evert. 2023. A reference constructicon as a database. *Yearbook of the German Cognitive Linguistics Association*, 11(1):175–202.

Pascual Pérez-Paredes, Geraldine Mark, and Anne O'Keeffe. 2025. *Corpus Linguistics for Language Learning Research*, volume 12. John Benjamins Publishing Company.

Maarja-Liisa Pilvik, Liina Lindström, , and Helen Plado. 2025. Nimisõnafraasi ja hulgafraasi piirimail. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, 21:237–261.

Heete Sahkai, Jelena Kallas, Ahto Kiil, Geda Paulsen, and Kertu Saul. 2025. Using large language models to identify constructional instances in corpus data. In *Electronic Lexicography in the 21st Century (eLex 2025): Intelligent Lexicography*. Lexical Computing CZ s.r.o.

Anatol Stefanowitsch and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2):209–243.

Elena Volodina. 2024. On two swell learner corpora – swell-pilot and swell-gold. In *Proceedings of the Huminfra Conference (HiC 2024)*, volume 205 of *Linköping Electronic Conference Proceedings*, pages 83–94.

8. Language Resource References

Institute of the Estonian Language. 2025. *EMMA corpus: EKI subsection*. Institute of the Estonian Language (EKI). PID https://doi.org/10.15155/mg95-yr49. Accessed 25 February 2026.

9. Appendices

A. Baseline prompt examples

aasta aega; Juba peaaegu aasta aega olen pidanud kirjavahetust kloostri abti isa Jeaniga.

enamik inimesi; Kümme aastat tagasi ei vaielnud enamik inimesi oma arsti pandud diagnoosidele ja määratud retseptidele vastu, sest arst teadis alati paremini – arvatavasti seetõttu, et oli paremini informeeritud kui patsient.

grupp tegelasi; Uus grupp tegelasi haarab telekaesise oma valdusse.

hektarit maad; Agraarreformi ajal aastal 1920 kuulus Mālupe mõisale 1131 hektarit maad.

kamp riike; Aga ilma ühendavate väärtusteta oleks ju EL ainult kamp riike, mille tahtmine maailmapoliitikas midagi olla ületab kaugelt võimeid selle nimel ühiselt pingutada.

hulka rühmitusi; Algkristlus hõlmas suurt hulka erinevaid rühmitusi ja arusaamu.

rea kohtumisi; Hiljem korraldas ta rea mitteametlikke kohtumisi Briti ministrite ja Saksamaa diplomaatide vahel, et aidata kaasa Suurbritannia ja Saksamaa lähenemisele.

aastat iseseisvust; 100 aastat Soome iseseisvust.

grupp inimesi; Võrgufirma põhimureks on väike grupp pahatahtlikke reklaamivaenulikke inimesi, kes püüavad kõigiti takistada uut põnevat turustamismeetodit Internetis.

hetk elu; Ta oli valmis kõike tegema, teda tiivustas missioonitunne, olgu või lapsik – kinkida vaatajale hetk tõelist elu.

`NO_QUANT`; Loodan väga, et keegi neist ei kavatse elustada veritasu traditsioone.

`NO_QUANT`; Maailma kõige pisema kiriku ees rohtu krõmpsutav kits heidab meile sellise pilgu, nagu tahaks öelda: Ja mida teie siis ootasite?

`NO_QUANT`; Iga lõunaosariiklase hinge on sisse põletatud orjapidamise mälujälg.

`NO_QUANT`; WC automaat pakkus viskimaitselisi kondoome.

B. Extended prompt examples

tund aega; Juba peaaegu tund aega olen ma pidanud selgitama oma käitumist.

enamik inimesi; Kümme aastat tagasi ei vaielnud enamik inimesi oma arsti pandud diagnoosidele vastu.

grupp tegelasi; Uus grupp tegelasi haarab telekaesise oma valdusse.

hektarit maad; Agraarreformi ajal aastal 1920 kuulus Mālupe mõisale 1131 hektarit maad.

kamp riike; Aga ilma ühendavate väärtusteta oleks ju EL ainult tavapärane kamp riike.

hulka rühmitusi; Algkristlus hõlmas suurt hulka erinevaid rühmitusi ja arusaamu.

rea kohtumisi; Hiljem korraldas ta rea mitteametlikke kohtumisi Briti ministrite ja Saksamaa diplomaatide vahel.

aastat iseseisvust; 100 aastat Soome iseseisvust.

hetk elu; Ta oli valmis kõike tegema, et kinkida vaatajale hetk tõelist elu.

`NO_QUANT`; Maailm on täis igat tüüpi inimesi.

`NO_QUANT`; Kaks kiriku ees rohtu krõmpsutavat kitse heitsid meile põlgliku pilgu.

`NO_QUANT`; Me ei pane oma vaatajale vanuselisi piiranguid.

`NO_QUANT`; Paar tundi hiljem jõudis lõpuks järg ka meie kätte.

`NO_QUANT`; Meie lennuk maandus hilisel keskpäeval.