Università di Foggia

KING'S College LONDON

H²IOSC
Humanities and cultural Heritage Italian Open Science Cloud

CLARIN

Expanding PREMOVE

An LLM-Assisted Semantic Annotation of Preverbs Across Historical and Modern Languages

Computational Humanities Research group Seminar Series

Michele Ciletti - University of Foggia

Table of Contents

Challenge: Preverbs in Motion & Perception

- Preverbs and PREMOVE
- Verbs of perception
- Modern languages
- Corpora

Method: From Manual to LLM-Assisted

- Annotation schema
- State-of-the-art: semantic annotation and WSD
- LLMs and their affordances
- Annotating preverb semantics
- WordNet and its issues
- Techniques and practical tips
- Coming soon!

Tool: Interactive Training Resource

- Semantic annotation at scale
- Working with Colab
- Setup and settings
- Data preparation
- Let’s run it!
- Outcome and evaluation

Join at menti.com | use code 5500 7132

Mentimeter

Go to

www.menti.com

Enter the code

5500 7132

Or use QR code

Next slide

What are preverbs?

Preverbs are prefixes attached to verbal bases.

They modify meaning: compositional vs. lexicalized.

Example: Lat. eo (go) → ex-eo (go out - compositional).
Example: Lat. video (see) → in-video (envy/hate - lexicalized).

They drive semantic change in Indo-European languages (Booij & Van Kemenade, 2003).

The PREMOVE Dataset

Created by Andrea Farina (2025).

First cross-linguistic diachronic dataset of preverbs.

Focus: motion verbs in Ancient Greek and Latin.

2,800+ manually annotated occurrences across 20 layers.

Outcomes: A CLARIN resource; PrevNet; link to LiLa.

Take a look!

Expanding the Scope

Goal

Move beyond Motion and integrate Perception verbs (Class 30.1; Levin, 1995).

Targets

Lat. video / AGr. horáō ('see')
Lat. audio / AGr. akoúō ('hear')
Lat. prefixed compounds in -spiciō / AGr. blépō 'look'
Lat. sentiō / AGr. aisthánomai 'perceive'

New language

Italian (Motion verbs).

Upcoming! Today focuses on the Latin and Ancient Greek.

Methodology and the LLM Pipeline

Automatic annotation

Limitations

Manual annotations made by experts are reliable, but time-consuming, expensive and hard to scale.

Automatic pipelines

Natural Language Processing tools can process text at scale, if they are trained on high-quality data.

Scalability and open data

Efficient tools can effectively generate large amounts of useful data, which the community can then reuse.

Traditional approaches: embeddings, BERT

Promising results: Bamman & Burns, 2020; McGillivray, 2022; Lendvai and Wick, 2022; Ghinassi et al., 2024; Mercelis et al., 2025

Why? Enabling quantitative analyses such as tracking diachronic semantic change!

Word-Sense Disambiguation (WSD)

The computational task in Natural Language Processing (NLP) of identifying the correct meaning (sense) of an ambiguous word within a specific context.

Enter LLMs

Massive models with billions of parameters

General-purpose reasoning improved by prompt engineering, in-context learning (Brown et al., 2020), chain-of-thought (Wei et al., 2022), fine-tuning

Hypothesis: Can LLMs replace or assist expert annotators in low-resource settings?

Join at menti.com | use code 5500 7132

Mentimeter

Go to

www.menti.com

Enter the code

5500 7132

Or use QR code

Next slide

Task 1: Preverb Semantics

Objective

Having an LLM assign the correct meaning to a Latin or Ancient Greek preverb, choosing from a list of English words or expressions.

Ground truth

A ground truth is built by following precise guidelines (Farina, 2024) and computing inter-annotator agreement.

Model selection

A balance has to be found between size, openness, performance, and cost.

Prompt strategy

A complex, detailed system prompt is given to the model to instruct it in its reasoning and answering style.

Learning from examples

In-context learning is leveraged to let the models observe verified patterns, and fine-tuning is tested too.

Settings

Temperature and reasoning effort have to be balanced.

Evaluation

Standard metrics such as precision, recall and F1 are computed; then the data is manually inspected.

Archiving

Once it is verified and approved, the data is ready to be openly shared.

Task 2: Verb Semantics

(and why it's a lot more difficult)

Working with WordNet

A problem

Tagging meaning is complicated: which options should a model (or a human) be given?

WordNet

A large, electronic lexical database that groups words into sets of cognitive synonyms called synsets. Gives more options than vocabulary and is standardized, but it can be too granular.

Solutions?

Embeddings + retrieval, hierarchy, predefined sets, Latin and Ancient Greek WordNets

Join at menti.com | use code 5500 7132

Mentimeter

Go to

www.menti.com

Enter the code

5500 7132

Or use QR code

Next slide

Efficient System Prompting

Role and Goal

You are an expert in Ancient Greek and Latin Linguistics. You will be given an Ancient Greek or Latin sentence that includes a preverbed motion verb. Your primary task is to identify the semantics of the preverb, considering the compositionality of the verb.

Guidelines

* The semantics of the preverb must be annotated depending on the specific meaning that the preverb acquires in context.
* If a preverb does not possess a clear lexical meaning and only contributes to telicising the verbal base, tag it with 'completely'.
* In some occurrences, two meanings may overlap. Annotate both.

Output Requirements

* Respond with ONLY one or two English meanings from the list provided below. These can be adverbs, prepositions, or multi-word expressions.
* Provide nothing else — no explanations, no conversational text, and no punctuation beyond what is in the expression itself. If you need to annotate two meanings, separate them with a comma followed by a whitespace.

Allowed Meanings

* about
* across
* after [...]

Efficient System Prompting

You are an expert in historical linguistics, Latin, Ancient Greek, and lexical semantics.
Your task is to select the most appropriate WordNet synset that best captures the meaning of a Latin or Ancient Greek verb in context.

You will be given:
1. A verb token in its original form
2. The lemma (dictionary form) of the verb
3. The sentence context
4. A numbered list of possible synsets with their definitions
5. Some annotated examples for reference

Analyze the context carefully and choose the synset that best matches the semantic meaning of the verb.

IMPORTANT: Respond with ONLY a single number corresponding to your choice. Do not include any explanation or additional text.

In-Context Learning / Few-Shot Prompting

Examples can help distinguish between edge cases and understand requirements.

Too many examples can exhaust a model's context window. Five is usually a good middle ground, depending on the task.

Ground truth has to be reliable to avoid catastrophic misdirections.

Model Selection and Settings

There is a shortage of benchmarks: model families perform well on languages they have been trained on.

High parameter counts usually lead to better results, but small models can be more affordable, open, and privacy-focused (+ fine-tuned!)

Chain-of-thought enhances performance by a lot and can be used for evaluation - but must be kept under control.

Temperature should be as low as possible without impacting performance to ensure reproducibility.

Being clear, simple, and straight-forward always helps!

Great, but how do I know it works?

Metrics

Precision, recall and F1 are standard metrics to evaluate the performance of the model on the ground truth.

Confidence score

A model can be asked to output a confidence score for each decision. Even if it has been shown that they are consistenly too high, their relative value often predicts performance.

Verification

Close reading automatic annotations is always a good idea, and it can lead to discovering faults in the setup.

Fine-tuning and Low-Rank Adaptation (LoRA)

Fine-tuning is the process of taking a pre-trained AI model and training it further on a smaller, task-specific dataset to specialize its abilities

LoRA is a technique that entails training small, specialized “adapter” modules instead of the whole model.

Small models can drastically improve their performance and match closed ones, for a fraction of the cost.

Designing an interactive tool for LLM-based annotation

Google Colab is a platform that lets users execute Python code with high-end computing resources

The idea: creating a simple, accessible interface that can be mastered, remixed, and reused

Feedback will inform future developments!

A Google Colab notebook to make automatic annotation accessible

Open the notebook here:

https://colab.research.google.com/drive/18NR5nITwqYX9ZdN3f1aMdQNFEwDyAS36?usp=sharing

Selected References

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.
- Farina, A. (2024). Guidelines for a linguistic annotation of preverbed verbs of motion (Version 2). King's College London. https://doi.org/10.18742/25055573.v2
- Farina, A. (2025). PREMOVE – A diachronic dataset of Ancient Greek and Latin annotated PREverbed MOtion VErbs. Oxford Text Archive. https://doi.org/http://hdl.handle.net/20.500.14106/2579
- McGillivray, B. (2022). How to Use Word Embeddings for Natural Language Processing. SAGE Publications Ltd. https://doi.org/10.4135/9781529609578
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.

Thank You!

https://github.com/MikCil/semantic_annotation_training

michele.ciletti@gmail.com