Ancient Greek Parsing with AI

Unpacking Complex Word Composition with Agentic Systems

Braeden Singleton, class of 2025
IPHS 391 Frontiers in Generative AI (Fall 2024) Prof Jon Chun, Kenyon College

Abstract

This proof of concept shows that an agentic AI system, with the aid of grammatical rules and dictionaries, can effectively identify the form and definition of an Ancient Greek word, a task which current unassisted systems like ChatGPT tend to struggle with.

Introduction

Although translation and transcription technology has made great leaps with the advent of artificial intelligence, modern systems still rely on the large body of content translated by people for future translations. This can work exceedingly well for modern languages, but the approach breaks down if the volume of content is too small to effectively train on.

This is felt in the realm of ancient languages like Ancient Greek, where only about 90 million words survive, and no new works can be created. Contrast this with the millions of *documents* needed to train a large language model, and the difference is far too large to compensate for.

Agentic systems, however, have the ability to specialize and perform specific roles in the translation process, allowing them to use established grammatical rules and patterns to multiply their effectiveness. This project makes an agentic system that mirrors the real translation process by referencing declension and conjugation rules, guessing at the dictionary form, and searching for the translation in a dictionary. By this process, we greatly reduce errors caused by similar forms as well as hallucinations citing words that don't exist.

Methodology

We give a series of tools to our AI classicist as follows:

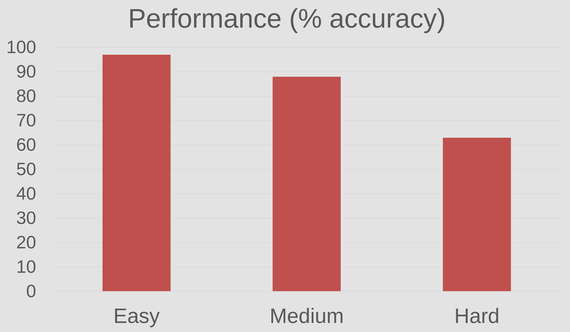
- Access to the LSJ Greek-English dictionary as JSON
- Access to Conjugation and Declension tables

With these tools, we instruct the Agent to follow this workflow:

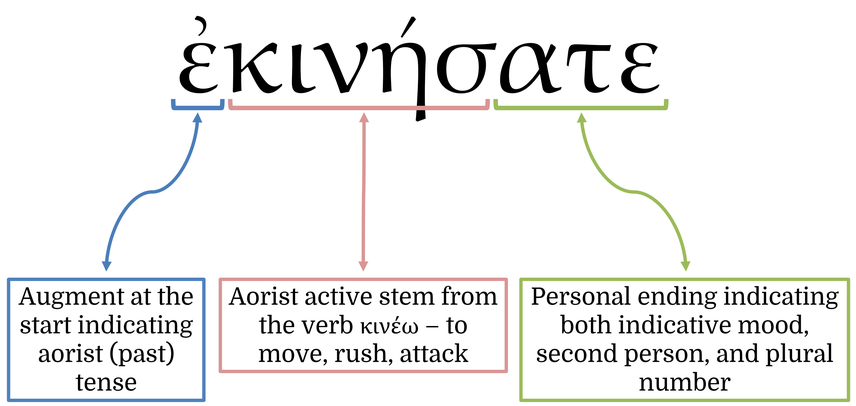
1. Check if the form exists in LSJ. If the form has a dictionary entry, it is either in dictionary form or the dictionary will link to the correct form
2. Make a guess about the part of speech
3. Based on this guess, identify the ending and generate the stem
4. Add the correct ending to the stem to find the dictionary entry
5. Present the dictionary entry to the user

Results

Testing was performed over a set of 12 easy, 9 medium, and 6 hard words. We then checked the parsing against the accepted one, allowing for variation where lack of accents made the form ambiguous. The scores in each were as follows:
Easy – 97%, Medium – 89%, Hard – 63%



The full testing data, analysis and code can be found in the github repository at github.com/Bubballoo3/Greek-Parser.git



This gives us the final parsing:

ἐκινήσατε
Second person
Plural
Aorist
Active
Indicative

When we combine this with the definition of κινέω and the grammatical rules of English, we find the translation:

"you rushed"

Conclusion

The translator agent was quite adaptable to a wide variety of forms and did an impressive job creating consistent formatting. Its performance over the testing categories shows that it is able to guess dictionary forms when they are close to the given word, and able to use the info from their unsuccessful dictionary searches to educate their guess. It does worse when

- The root word is rare
- The word is irregular, ie. the dictionary definition is very different from some forms
- The given word is a compound with a preposition. It had a difficult time separating prefixes from the root stem.

Future and Ethics Statement

This project showed that AI agents can accomplish much of what a human can do given the same tools, but is limited by narrow context. More resources for the agent to use would produce better results, although it may start to lose effectiveness when too much context is given. This system could also be combined with a translator model that uses the results of the parsing, as well as the full information of the dictionary entry, to understand full sentences

References/Acknowledgements

Code was started with the LangGraph Quick Start Tutorial at https://langchain-ai.github.io/langgraph/tutorials/introduction/

Resources for the agent were sourced from:

Liddel and Scott's *English-Greek Lexicon* (json sourced from https://github.com/perseids-project/lsj-js.git

Helma Dik's *Contract Verb Cheat Sheet* from https://classics.uchicago.edu/people/helma-dik/nifty-greek-handouts

In addition to some tables generated by ChatGPT