Decoding Morphosyntax: Can LLMs Handle Inflection and Derivation in English and Greek?

Authors: Athanasios Karasimos, Aristotle University
Vasiliki Makri, National and Kapodistrian University of Athens, Aristotle University
Abstract ID: 278
Event: Linguistics 2026 Athens
Subject: 17. Computational linguistics
Presenter Name: Athanasios Karasimos

Abstract

This study explores the extent to which Large Language Models (LLMs) can accurately analyze and compare inflectional morphosyntactic features and derivational patterns in English and Greek — two typologically divergent systems. While English is predominantly analytic, Greek exhibits a rich fusional inventory of morphosyntactic features encoded inflectionally, including tense, aspect, number, case, gender, and person. The central research question is whether LLMs can reliably identify, categorise, and disambiguate these features across the two systems and how their performance on inflectional paradigms interacts with their handling of derivational morphology. A secondary focus concerns the degree to which morphosyntactic feature encoding constrains or assists LLMs in recognizing derivational processes and their productivity. The study hypothesizes that the typological mismatch between English and Greek exposes systematic gaps in LLM morphological competence, particularly in the processing of inflectionally dense paradigms and derivationally complex lexemes.

Methodology

The study adopts a mixed-methods design integrating theoretical morphological analysis with computational modeling and empirical evaluation. Two annotated corpora — one for English and one for Greek — are constructed from diverse text sources, with words tagged for morphosyntactic feature values, derivational patterns (prefixation, suffixation), and morphological complexity ranging from transparent to opaque forms. Complex phenomena receiving special attention include syncretism, allomorphy, suppletion, and morphosemantic ambiguity — all of which pose well- documented challenges for both human parsers and computational models. State-of-the-art LLMs (GPT-5.1, Gemini 3, Claude 4.6, and Perplexity 4.5) are evaluated on the annotated datasets using standard metrics (precision, recall, F1-score), complemented by novel morphology-sensitive metrics developed specifically for LLM morphological evaluation: a Morphosyntactic Context Sensitivity metric, a Morphological Complexity Score, and a Morpheme Accuracy Metric, among others. Supervised fine-tuning and Reinforcement Learning from Human Feedback (RLHF) via prompt engineering are further explored as strategies for improving model performance.

Results

Evaluation results reveal that LLMs perform inconsistently across morphosyntactic feature categories, with the greatest difficulties emerging in Greek paradigms characterized by high degrees of syncretism and morphophonological alternation (e.g., inactive phonological phenomena, allomorphy). In derivational analysis, models tend to rely on surface analogical patterns rather than rule-governed morphological operations, leading to systematic errors in disambiguating derivation from inflection and in correctly identifying the base and affix structure of complex words. Cross- linguistic comparison further confirms that morphosyntactic typology significantly affects LLM generalization, with Greek consistently yielding lower accuracy scores than English across all evaluation metrics.

Conclusions

The findings demonstrate that current LLMs lack robust morphosyntactic feature representations and that their handling of inflection and derivation falls short of linguistically informed analysis. Crucially, the study shows that targeted fine-tuning on morphologically annotated data — particularly for feature-rich languages like Greek — can meaningfully improve performance. These results have direct implications for the design

of NLP tools in morphologically complex languages and call for evaluation frameworks that foreground morphosyntactic adequacy rather than surface fluency alone.