Linguistic Analysis of LLM-Generated Alsatian

Authors: Carine Graff, Colorado State University
Nikhil Krishnaswamy, Colorado State University
Abstract ID: 471
Event: Linguistics 2026 Paris
Subject: 11. Dialectology
Presenter Name: Carine Graff

Abstract

With about 7,000 languages in the world and approximately 43% of them endangered, researchers are looking for different methods to preserve them. Large language models (LLMs) could be a solution, but as of 2026, modern LLMs are not able to correctly recognize a low-resource dialect or generate it. To investigate if fine-tuning can address these limitations, we examined whether a small multilingual causal model (Qwen2-1.5B), fine-tuned on Alsatian theater data could reproduce dialect-specific linguistic features across different generation tasks and scenarios.

We compared generations from the base and fine-tuned (FT) models against Alsatian linguistic features using twenty prompts across five tasks and four prompt conditions (English, German, Alsatian, and Alsatian continuation). We examined eight lexical and morphological categories and seven grammatical constructions from the DIVITAL project using word pairs extracted from its Alsatian treebank. For each generation, we calculated the proportion of Alsatian versus German diagnostic words overall and by category, excluding shared forms but including contracted articles (e.g. *d'Fràu*). Constructions were classified as Alsatian, German, or absent. Finally, a dependency parser trained on DIVITAL and related German dialect treebanks evaluated verb-second order in main clauses and verb-final order in subordinate clauses.

The FT model's generations moved toward Alsatian, whereas the base model remained largely German, averaging fewer than one Alsatian function word per text. The mean diagnostic-lexicon share increased from 0.265 to 0.603, with shifts toward Alsatian in definite articles (0.054 → 0.708), negation (0 → 1.0), subject pronouns (0 → 0.45), subordinating conjunctions (0 → 0.231), and infinitive endings (0 → 1.0). However, personal pronouns remained equally divided between German and Alsatian forms, and apostrophe use exceeded Orthal conventions. Some base outputs were entirely in English. Verb-second word order was similar for both models (76.6% base; 74.8% FT), approaching the parser ceiling (81.1%). Five of seven DIVITAL constructions were unattested in either model. Possessives occurred at comparable rates, but the base model favored German and the FT model Alsatian forms. Null second-person subjects appeared only in FT (0 vs. 2).

The FT model shifted the measured categories towards Alsatian. Its gains were concentrated in lexis and morphological features including articles, negation, subject pronouns, subordinating conjunctions, and infinitive endings. For the same grammatical constructions, compared to the base, the fine-tuned model was more likely to use the Alsatian form rather than the German

equivalent. Improvements were less evident at the level of grammatical constructions, suggesting a need for more targeted data to improve grammatical coverage. The base model produced possessive constructions mainly by repeating instructions, whereas the FT model generated them independently. In conclusion, fine-tuning improves lexical and morphological accuracy, while more complex grammatical constructions remain a challenge.

The results show the importance of evaluating low-resource language models beyond fluency, as fluent output does not necessarily reflect the grammatical characteristics of the target variety.