When Vowels Change Meaning: Morphological Ambiguity and AI Interpretation of Unvowelled Arabic

Authors: Eyhab Bader Eddin, Dhofar University, Oman
Abstract ID: 464
Event: Linguistics 2026 Paris
Subject: 6. Morphology
Presenter Name: Eyhab bader Eddin

Abstract

In Arabic, there are no short-vowel marks, so that the same sequence of consonants can be used for different morphological analyses. This creates an issue for generative artificial intelligence (AI): a fluent translation of an unvowelled expression can mask unexposed grammatical alternatives. The present study investigates the impact of vowel omission on the identification of grammatical categories, verbal voice and participant roles in Arabic and how these ambiguities can be addressed in AI-generated English interpretations. It queries whether AI systems identify plausible alternative readings, what evidence in the context is used to make their choices, and how their answers express uncertainty. The investigation relates Arabic root-and-pattern morphology to the disambiguation of context and linguistic approaches to interpretability.

Methodology

The study is descriptive and comparative in nature, but no statistical scoring or frequency-based inferences are made. A set of 8 Arabic expressions will be created to represent different types of vowel-dependent ambiguity, and this set will be purposive. Every expression will be linguistically analysed, and alternative fully vowelled readings will be provided as well as grammatical explanations and contextually appropriate English renderings. For instance, فتح الباب can be read as an active statement (He opened the door) or a passive statement (The door was opened) depending on how it is vocalised and which case it is marked with. The expressions will then be fed to the generative AI system in two scenarios: without context and with disambiguating contexts. Responses will be analysed in detail with regard to morphosyntactic and semantic aspects, with emphasis on interpretation, contextual justification and acknowledgement of alternatives. Prompts and outputs will be recorded to ensure transparency and replicability of the analysis.

Results

A preliminary analysis of Arabic forms reveals three successive zones of ambiguity: alternative verbal voices, conflicting word-class assignments and varying assignments of grammatical agency. These are not just differences in pronunciation, but they can alter the subject of an action, the active or passive voice of a reported event and the lexical meaning that is available. The paired-context design allows for the study of whether the AI response provides a single unqualified English translation, or multiple readings or revises its interpretation when the contextual evidence changes. The actual choices of the system will be determined from the outputs recorded, not assumed beforehand.

Conclusions

The study considers Arabic vowel omission as an interpretative problem in the intersection of morphology, semantics and AI language processing. It suggests a repeatable procedure to evaluate the grammatical justification and uncertainty calibration of AI-generated interpretations. The analysis can be used to facilitate more transparent translation practices and better educational explanations of ambiguity and context-sensitive evaluation of AI-assisted reading in Arabic linguistics and language technology. These issues are illustrated through annotated contrasts and explanations in the poster.

Keywords: Arabic morphology; Morphological ambiguity; Unvowelled Arabic; Generative artificial intelligence; Contextual disambiguation.