Sing, Muse, of Small Data: Computational Approaches to Homer's *Odyssey.*

Dr. Annie K. Lamar, Department of Classics, UC Santa Barbara

In computational linguistics, low-resource contexts are conventionally defined by a scarcity of raw tokens or annotated corpora. However, data scarcity manifests along multiple distinct axes. First, ancient data is strictly limited; we cannot generate new texts, and there are no living speakers. Second, data can be constrained by its specific form, such as rigid metrical rules or unique dialects that do not align with the language present in larger datasets. Homeric epic, defined as the ancient Greek poetry tradition that produced the *Iliad* and the *Odyssey*, was composed orally and preserved in strict dactylic hexameter. Homeric epic thus combines both challenges. This talk introduces Homeric epic as a unique low-resource dataset for computational linguistics and natural language processing. I present quantitative methods and statistical metrics designed to model complex low-resource phenomena, focusing specifically on formulaic language—the recurring noun-epithet formulas, repeated half-lines, and structural patterns characteristic of oral composition. Using Homeric epic as a case study, this talk will show how we can design statistical metrics that take advantage of complex poetic structures and linguistic phenomena in low-resource contexts.