

How forms shape meaning: What distributional learning can and can't do in the era of LLMs

Alessandro Lenci | University of Pisa, Italy

Large Language Models (LLMs) are distributional learning systems: they acquire semantic competence from the statistical distribution of forms across massive text corpora. In this sense, LLMs represent the most advanced development of distributional semantics to date. Compared to earlier models, they appear capable of learning vastly larger amounts of knowledge, yet they still face significant limitations—many of which resemble those that characterized their predecessors. The distribution of linguistic forms is an extraordinarily rich source of information, far richer than we once imagined. At the same time, its limitations cannot be underestimated.