AI and Linguistics: a Critical Overview

Roberto Zamparelli
University of Trento
roberto.zamparelli@unitn.it

The explosion of AI is having a profound impact on society at many levels, including research. This is true in any field, but that fact that AI is dominated by language models (LMs), which consume and produce language and are capable of human-level fluency means that, in science, AI has affected linguistics the most; for the first time, language researchers have at their disposal *cheap, virtual non-human models of language* (comparable to animal models of other biological or cognitive functions), and the opportunity to assess their relevance in characterizing human language and the way humans learn it. This has created practical and sociological rifts in the theoretical linguistic community, with fundamental disagreement on several aspects, including how explanations of linguistic phenomena should be framed.

Part of the issue is, of course, that the Machine Learning core of LMs, coupled with the way context is implemented in transformer models, has pushed AI toward *end-to-end solutions*: language as input, language as output (give or take multimodality). Intermediate representations of linguistic phenomena (e.g. constituent structures), so dear to traditional theoretical linguistics, must now be pried from the inner states of the system, and can often be wielded equally well to confirm some aspects of traditional representations or to proclaim their irrelevance.

In this talk, I will argue that the key issue of much of current research at the border between theoretical and computational linguistics is that it has very limited 'shelf life': it applies to models that are constantly and rapidly evolving, whose input is unknown, whose structures make hidden assumptions that hinder the comparison between LMs and human learners (input size and presentation order, number of epochs, size and accessibility of context). In my view, our biggest challenge now is how to design comparisons that test *fundamental aspects*; in principle, first (e.g. does the Poverty of the Stimulus hypothesis hold given unlimited computational resources?), then by gradually restricting the capabilities of models to bring them in line with what is known about humans, including their intermediate learning stages and their patterns of error (e.g. overregularization). Since this goes against the grain of current NLP research, I hypothesize that we might end up using LMs are research colleagues before we can use them as models (crucially, 'them' has a sloppy reading here: the 'colleagues' are not going to be the same as the 'models').