Computational Linguistics for Studying Language in People: Principles, Applications and Research Problems

Invited talk

Bruce Hayes
Department of Linguistics, UCLA
bhayes@humnet.ucla.edu

One of the goals of computational linguistics is to create automated systems that can learn, generate, and understand language at all levels of structure (semantics, syntax, morphology, phonology, phonetics). This is a very demanding task whose complete fulfillment lies far in the future. Human beings can learn, generate, and understand language at all levels of structure, and the study of how they do it can be pursued by computational modeling. Indeed, one sort of "acid test" for theories in linguistics is whether they can serve as the basis for successful models of this kind. A research strategy for theoretical linguistics based on modeling thus invites close collaboration between "mainstream" linguists and their computational colleagues.

Such collaborations make the job of the computationalists, already very demanding, even harder. The collision between a computational model and human data arises when we apply the Turing test: the model ought to behave like a human, not just in generating a correct output, but in every conceivable sense: generating alternative outputs for a single input, (often with a nuanced sense of preferences among them), generating human-like mistakes, generating child-like mistakes when given incomplete information, and so on. I suggest that linguists could serve as good Turing testers, because it is their daily practice in professional life to interrogate their models in the most ingenious ways they can find, probing for deficiencies though comparison to complex human intuitions and behavior.

With this as backdrop, I offers a series of case studies, of three different kinds: (1) **Turing tests**: cases where interrogation by linguists revealed non-humanlike traits in computational models that performed well by traditional computational criteria (precision, recall, etc.); (2) **Success stories**: particular results of computational linguistics that have proven useful so far in modeling language in humans; (3) **Suggestions for future work**: proposals in linguistic theory that look promising and would benefit from computational analysis.

Proceedings of the Fifteenth Conference on Computational Natural Language Learning, page 48,
Portland, Oregon, USA, 23–24 June 2011. ©2011 Association for Computational Linguistics