PhD Research Summary

Daniel Quigley, Linguistics Department, University of Wisconsin - Milwaukee

Ellipsis is a linguistic phenomenon in which some parts of sentences are left unexpressed [2]. Ellipsis masks information from a machine that is otherwise available to a human; consequently, there is a need for NLP innovations to automatically detect and accurately interpret ellipsis. Developing methods for ellipsis resolution contributes to machines' achievement of human-level understanding of language, which integrates definitively with **Human Language Technologies**. My dream is to build machines that can navigate without human supervision the ambiguities of human language; to that end, my research focuses on ***learning, describing, and implementing the theoretical linguistics of ellipsis to develop frameworks for machines such that interpretation of intra- and extra-linguistic context to resolve ellipses with minimal human supervision can be achieved***. Ellipsis resolution — an example of **Human-Computer Interaction** — can aid in avoiding bias towards or against a speaker's production of spontaneous or deliberate speech in signed languages, dialectal variations, and personal idiolects; additionally, entwined with **Computer Vision**, resolution follows for extra-linguistic antecedents informed by real-world contexts and ellipsis constructions in signed languages. This interdisciplinary work contributes to an optimistic future of human-machine communication. Educational opportunities naturally follow for communicating ellipsis resolution's problems, processes, and results to the academic, industrial, and popular communities.

Applying practical NLP algorithms with tools in **anaphora resolution** [4] is a driving focus of my work; I investigate interpretation of ellipsis-as-anaphor with local intra-linguistic contexts prior to working with long-distance contexts; extra-linguistic contexts follow. Recasting ellipses into similar NLP problems (**machine reading comprehension** [1], **question answering** [6], **coreference resolution** [7]) is advantageous for the auxiliary data and computational methods therein. The main disadvantage of using **neural networks** is the clustering time, which is longer than in compared approaches; even considering this limiter, neural networks with **multi-pass sieve** architectures [5] show promise in resolving reference. Finally, it is a novel approach to apply key insights from **Construction Grammar** [3], which gains much of the representational flexibility of constructions while retaining existing NLP infrastructure. My hope is that the Construction Grammar and NLP communities will work together to define more flexible algorithms for NLP tasks.

References

[1] Rahul Aralikatte et al. "Ellipsis Resolution as Question Answering: An Evaluation". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*. Ed. by Paola Merlo, Jorg Tiedemann, and Reut Tsarfaty. Association for Computational Linguistics, 2021, pp. 810–817.

[2] Jeroen van Craenenbroeck and Jason Merchant. "Ellipsis phenomena". In: *The Cambridge Handbook of Generative Syntax*. Ed. by Marcel den Dikken. Cambridge Handbooks in Language and Linguistics. Cambridge University Press, 2013, pp. 701–745.

[3] Adele E. Goldberg and Florent Perek. "Ellipsis in Construction Grammar". In: *The Oxford Handbook of Ellipsis*. Ed. by Jeroen van Craenenbroeck and Tanja Temmerman. Oxford University Press, 2019. ISBN: 9780198712398.

[4] Zhengzhong Liu, Edgar Gonzàlez Pellicer, and Daniel Gillick. "Exploring the steps of Verb Phrase Ellipsis". In: *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016)*. San Diego, California: Association for Computational Linguistics, June 2016, pp. 32–40.

[5] Karthik Raghunathan et al. "A Multi-Pass Sieve for Coreference Resolution". In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA: Association for Computational Linguistics, Oct. 2010, pp. 492–501.

[6] Bo-Hsiang Tseng et al. "CREAD: Combined Resolution of Ellipses and Anaphora in Dialogues". In: *CoRR* (2021).

[7] Wei Wu et al. "CorefQA: Coreference Resolution as Query-based Span Prediction". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, July 2020, pp. 6953–6963.