

Annotated Bibliography

Dylan Orris

September 25, 2018

References

- [Cho03] Gobinda G Chowdhury. Natural language processing. *Annual review of information science and technology*, 37(1):51–89, 2003.
- [Lyt86] Steven L Lytinen. Dynamically combining syntax and semantics in natural language processing. In *AAAI*, volume 86, pages 574–587, 1986.

Lytinen’s article discusses issues with what were, at the time, the two main paradigms for natural language processing. One camp felt that syntactic structures should be examined first, with ambiguity later cleared up using semantics. The other felt that these concepts could not be split apart, so semantics and syntactics would be examined simultaneously by the processor. Lytinen explains that the semantics-second approach leads to a great deal of ambiguity and slow runtime. By ignoring semantics, which could easily clear up some ambiguity present in language, statements may rapidly become n-ambiguous, meaning that there are n points of ambiguity. This ambiguity rapidly multiplies, as two ambiguous points in a row mean there are four possible meanings for the sentence, assuming none are mutually exclusive. Lytinen then describes issues with the view of simultaneous semantic and syntactic analysis. Though this system clears many ambiguities away, making it less computationally expensive, it becomes space inefficient rapidly. This is because its rules are not generalizable, as the focus on semantics leads to many nearly identical rules being applied, but being programmed as though they are different. Finally, Lytinen introduces MOPTRANS, a parser which translates stories about terrorism. This parser uses semantics on statements within recent memory as a method to clear ambiguity early, while allowing for rules to be generalized. The parser works by checking how items in memory compare to rules that are

known by the parser, and groups them together early when high percentage likelihood matches are made. This system is storage efficient, while also retaining the benefits of considering semantics. While this may not be particularly applicable to my project, understanding the relationship of syntax with the semantics I work with may allow for more complex statements to be input and understood. Rather than requiring all claims to come in a certain form, it may be possible to program certain rules based on how words apply to their surroundings to better understand user input. The inclusion of multiple subjects or fact bases in a single claim would currently break my parser, so this article was useful in seeing an effective way of dealing with the ambiguity which may arise in these situations.

- [PB93] James Pustejovsky and Branimir Boguraev. Lexical knowledge representation and natural language processing. *Artificial Intelligence*, 63(1-2):193–223, 1993.

This article discusses the benefits of using processing power to determine what part of processed language a word is, rather than encoding every possible way it could be used. By attempting to enumerate every way to use a word or phrase, it is quite possible that the parser becomes confused when words are used in a new and novel way. By using a generative lexicon, we can eliminate this problem. The generative lexicon examines the text at a deeper level, attempting to use relative position and relation to other words to determine how each word behaves. Pustejovsky argues that this method not only captures similarities in meaning that enumerative parsers miss, but can be more storage-effective. By examining the difference between ‘forgot to’ and ‘forgot that’, Pustejovsky explains that while whatever was forgot did not occur in the first claim but did in the second, there is a clear underlying relation between the forgot. Rather than ignoring this and building in two separate rules for forgot into the parser, it is possible to look deeper into meaning.

- [Res99] Philip Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of artificial intelligence research*, 11:95–130, 1999.
- [Sme92] Alan F Smeaton. Progress in the application of natural language processing to information retrieval tasks. *The computer journal*, 35(3):268–278, 1992.