



Distinguished Fellowship Award Brief

Jointly modeling syntax and semantics in verb-learning (2015-2017)

Fellow: Aaron Steven White, PhD
Postdoctoral Fellow

Mentors:

Kyle Rawlins, PhD
Assistant Professor
Cognitive Science
Krieger School of Arts and Sciences

Benjamin Van Durme, PhD
Assistant Research Professor
Computer Science/CLSP
Whiting School of Engineering

Research Question: This project investigates how humans learn words that express relationships between different types of objects. Learning what types of objects are involved in a relationship and learning the relationship itself are complementary goals. This project aims to understand how humans use what they know about one to figure something out about the other.

Interdisciplinary Approach: This project brings insights from machine learning, cognitive science, and linguistic theory to bear on a long-standing question in language learning: how words are learned. It does this by constructing explicit computer models of what is going on in a speaker's mind when they are learning a word.

Potential Implications of Research: Humans are much better than machines at understanding human language. This research aims to construct explicit models of language learning that will make computers better at understanding human language.

Human experience burgeons with complex relationships between objects. Human languages allow us to label these relationships in extremely specific ways. For instance, the relationship expressed by the sentence *John knows that Mary is a nice person* is similar to the one expressed by *John believes that Mary is a nice person*. Both involve John's beliefs; but the first involves a relationship to a fact, whereas the second involves a relationship to a thought. How do humans learn such specific aspects of words' meanings? This project aims to answer this question by constructing explicit computer models of language learning.

At a high-level, computer models of human word-learning must serve two functions. On the one hand, they must show how human learners discover *semantic content*: the nature of the relationship that the word expresses. For example, belief is part of the semantic content of *know* and *believe*. On the other, they must show how human learners discover *ontology*: the types of objects that the word expresses a relationship between. For example, the ontology of *know* is that it expresses a relationship between individuals and facts, while the ontology of *believe* is that it expresses a relationship between individuals and thoughts.

The relationships expressed by language are intimately intertwined with the entities they relate. For example, a physical relationship like kicking seems to involve two individuals, but not, for example, an individual and a thought. This suggests that, if one knew the semantic content, this might help in figuring out the ontology, or conversely, if one knew the ontology, this might help in figuring out the semantic content. Yet many learning models attempt to learn either the semantic content or the ontology, with no knowledge of the other.

This project aims to study ways in which learning a word's semantic content at the same time as its ontology might make learning easier and thus help explain why humans are able to do it so well. It does this by directly mapping theoretical concepts of human mental architecture into explicit computer models. For example, the computer model might "know" that words have both semantic context and ontology and might use its current knowledge of one to update its understanding of the other, getting more accurate as it gets more experience. To test the cognitive plausibility of these models, this project will use large-scale psycholinguistic experiments made possible by crowd-sourcing platforms like Amazon Mechanical Turk.