

A central challenge in developing a theory of how the mind processes language is how the meanings of ambiguous words are resolved. Approximately 85% of the words in the world’s languages are ambiguous, yet in the vast majority of circumstances, humans resolve this ambiguity correctly and effortlessly (Klein and Murphy, 2001). For example, in the phrase “A prime minister wields significant power” a reader has no difficulty evoking a “political power” interpretation as opposed to an “electrical power” interpretation. However, how ambiguity is resolved remains an open question, despite having received extensive attention from interdisciplinary researchers interested in language (including philosophers such as myself, as well as linguists, and psychologists), and despite the game-changing role that a theory of ambiguity resolution could have for artificial intelligence and computational linguistics. Indeed, a firm grasp of the fundamental principles governing ambiguity resolution are likely to revolutionize applications of language technologies including building better “bots” to communicate with customers online, allowing search engines to better understand user queries, and better automated screening of user content for remarks that are inappropriate or discriminatory.

My proposal builds upon the latest insights from several disciplines to develop and test a novel account of ambiguity resolution. First, my philosophy background makes me appreciate the fundamental weakness of a cognitive system without grounded and referential information connected to the outer world (Harnad, 1990; Searle, 1980). Second, I will adopt the view in cognitive science that understanding language fundamentally requires language to be part of a broader system for understanding and communicating about situations, such as describing a route to work, or creating a mental image of a painting based on a friend’s verbal description of a trip to a museum. Third, from the psycholinguistics and neuroscience literatures (including work by my proposed advisor, Professor Blair Armstrong), I will draw on recent findings demonstrating how building in additional neurobiological plausibility into simulations (e.g., by having separate excitatory and inhibitory neurons in a neural network) are essential for simulating a number of aspects of cognition, including the time course of ambiguity resolution (Armstrong and Plaut, 2016).

RESEARCH PLAN: My research plan focuses on starting from a (relatively) simple neural network model and gradually increasing the complexity of this model to achieve better performance and to quantify exactly how each new level of complexity improves over simpler models. Such complexity is not achievable until the recent development in training methods and hardware. All of these models except the first “control” model will learn to see objects while simultaneously processing the verbal description of these objects, allowing information from the visual and verbal modalities to interact. This model will be trained on databases of images and their associated captions, such as the COCO (Common Objects in Context) dataset, and each model will be tested against published data in the psycholinguistics literature which will serve as the “gold standard” for successfully modeling human ambiguity resolution abilities. This will include both the speed and accuracy with which humans select a contextually appropriate interpretation of a word (Rice et al., 2019).

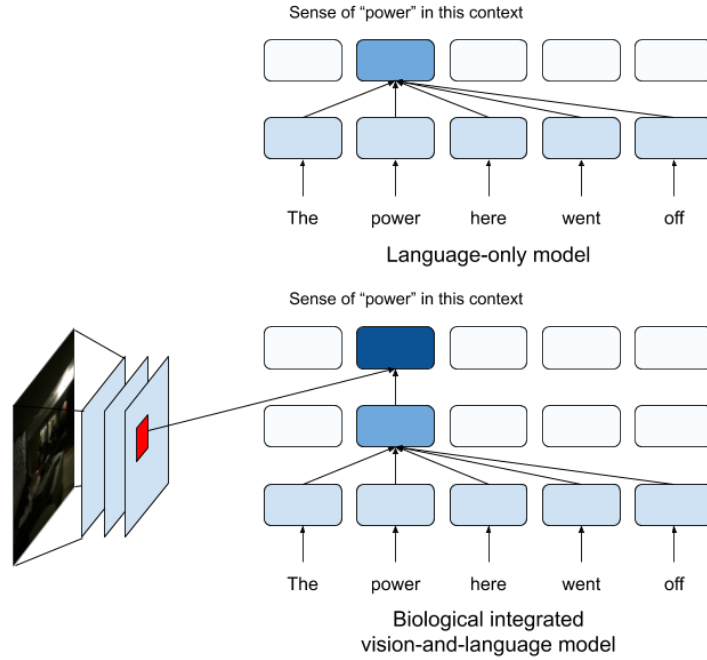


Figure 1: Computational models that are going to be build in this project.

1. Language-only model. The first model will be trained only on the verbal input in the COCO dataset, and provides a baseline for how much disambiguating information can be extracted from text alone. This approach is currently widespread in computational linguistics, and has been shown to contain substantial (although far from complete) disambiguating information. (Beekhuizen et al., 2018)

2. Integrated vision-and-language model. The second model will add a visual processing system to the model developed in (1), and will serve as a first validation for how visual information can provide additional constraint on word meaning than the prior model. A further key advantage of adding a visual system to this model is that it will be possible to omit presenting any visual information and allow the model to “hallucinate” what the visual representation would look like. This will provide an additional avenue for evaluating how well the model has successfully learned and fused multi-modal information.

3. More biologically plausible integrated model. The third model will focus on how several key principles from cognitive and systems neuroscience could further enhance performance in the model: 3(a) Addition of an attention mechanism. Humans do not process all of the information their eyes take in equally well, rather, they focus attention on one part of the visual field and extract substantially more information from that area than others. This “less-is-more” approach is expected to extract less “irrelevant” information and more “critical” visual information from the fields, which will enhance performance. 3(b) More realistic cortical connectivity. In artificial neural networks, artificial neurons are typically able to send both positive (excitatory) and negative (inhibitory) information to other neurons, and often every neuron in one simulated brain area is connected to every neuron in another brain area. In “real” neural networks, such as the human brain, there are very different constraints on

cortical connectivity. Neurons send excitatory OR inhibitory information, NOT both; there are vastly more ($\approx 85\%$) excitatory neurons than inhibitory neurons ($\approx 15\%$), and the size of the skull requires that there are relatively few connections between different brain regions, such as vision and language. Prior work (Laszlo and Armstrong, 2014; Armstrong and Plaut, 2016) has established that these principles critically shape what type of information shapes network performance, including word disambiguation. For example, the greater number of excitatory neurons means that it is easier to send information about what *was* read by the system rather than what *was not* read, and the relatively sparse connections between brain regions requires only the most important information to pass between brain regions (much as the visual attention mechanism limits what information is extracted from a visual scene).

EXPECTED OUTCOME AND IMPLICATIONS: I will revolutionize how ambiguity is thought about in many fields by merging insights from several fields to develop an account that would not emerge from any one discipline in isolation. Another interesting area to investigate is how abstract polysemous words are understood based on visual information. Political and ability sense of power are less connected to the physical world but people have been arguing that they share the same root as more physical words. This project will also serve as an attempt to test where the meanings of abstract polysemous words are rooted. Finally, the project can be potentially patented, if successful, to allow its free use to academics as well as its incorporation into industry, like Microsoft, Facebook, Google, etc.

References

- Armstrong, B. C. and Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, 31(7):940–966.
- Beekhuizen, B., Milic, S., Armstrong, B. C., and Stevenson, S. (2018). What Company Do Semantically Ambiguous Words Keep? Insights from Distributional Word Vectors. In *CogSci*.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346.
- Klein, D. E. and Murphy, G. L. (2001). The Representation of Polysemous Words. *Journal of Memory and Language*, 45(2):259–282.
- Laszlo, S. and Armstrong, B. C. (2014). PSPs and ERPs: Applying the dynamics of post-synaptic potentials to individual units in simulation of temporally extended Event-Related Potential reading data. *Brain and Language*, 132:22–27.
- Rice, C. A., Beekhuizen, B., Dubrovsky, V., Stevenson, S., and Armstrong, B. C. (2019). A comparison of homonym meaning frequency estimates derived from movie and television subtitles, free association, and explicit ratings. *Behavior Research Methods*, 51(3):1399–1425.
- Searle, J. R. (1980). Minds, brains, and programs. *ℳ ℳ*, pages 417–457.