

## **Vered Shwartz: Not a piece of cake: on lexical composition and implicit information**

Obtaining phrase representations by composing their constituent word vectors is an alternative to the low-quality distributional representations of infrequent phrases. Have contextualized word representations, like BERT, made lexical composition trivial? This talk will start by evaluating existing word representations for their capacity to address issues related to lexical composition. We found that contextualized word representations excel in detecting that the meaning of the phrase deviates from the meaning of its constituent words, such as for idiomatic noun compounds ("guilt trip") and verb-particle constructions ("carry on"). However, they still struggle with recovering implicit information introduced by the composition ("olive oil" is oil made of olives, while "baby oil" is oil made for babies).

I will then dive into revealing implicit meaning in compositional noun compounds. I will review various composition functions, discuss a method for noun compound relation classification to predefined relationships (e.g. source, purpose), and a model for free-text paraphrasing of noun compounds ("oil made of olives"). The talk will conclude by suggesting future research directions for obtaining meaningful phrase representations, and the many challenges on the way there.