

# A Long Hard Look at MWEs in the Age of Language Models

Vered Shwartz<sup>1,2,3</sup>

<sup>1</sup> Allen Institute for AI

<sup>2</sup> University of Washington

<sup>3</sup> University of British Columbia

vereds@allenai.org

## Abstract

In recent years, language models (LMs) have become almost synonymous with NLP. Pre-trained to “read” a large text corpus, such models are useful as both a representation layer as well as a source of world knowledge. But how well do they represent MWEs?

This talk will discuss various problems in representing MWEs, and the extent to which LMs address them:

- Do LMs capture the implicit relationship between constituents in compositional MWEs (from *baby oil* through *parsley cake* to *cheeseburger stabbing*)?
- Do LMs recognize when words are used non-literally in non-compositional MWEs (e.g. do they know whether there are fleas in the *flea market*)?
- Do LMs know idioms, and can they infer the meaning of new idioms from the context as humans often do?

## Bio

Vered Shwartz is a postdoctoral researcher at the Allen Institute for AI (AI2) and the University of Washington. She will join the Department of Computer Science at the University of British Columbia as an Assistant Professor in fall 2021. Her research interests include computational semantics and pragmatics, multiword expressions, and commonsense reasoning.