# Methods and metrics: Overview

Christopher Potts

Stanford Linguistics

CS224u: Natural language understanding

# Goal: Help you with your projects

- Managing data
- Establishing baselines
- Comparing models
- Optimizing models
- Navigating tricky situations

# Associated materials

- Evaluation metrics notebook: https://github.com/cgpotts/cs224u/blob/master/evaluation_metrics.ipynb
- scikit-learn guidance on model evaluation: http://scikit-learn.org/stable/modules/model_evaluation.html
- Evaluation methods notebook: https://github.com/cgpotts/cs224u/blob/master/evaluation_methods.ipynb
- Resnik and Lin 2010; Smith 2011, Appendix B

# Your projects

1. We will never evaluate a project based on how "good" the results are.
   - Publication venues do this, because they have additional constraints on space that lead them to favor positive evidence for new developments over negative results.
   - In CS224u, we are not subject to this constraint, so we can do the right and good thing of valuing positive results, negative results, and everything in between.
2. We will evaluate your project on:
   - The appropriateness of the metrics.
   - The strength of the methods.
   - The extent to which the paper is open and clear-sighted about the limits of its findings.

# References I

Philip Resnik and Jimmy Lin. 2010. Evaluation of NLP systems. In Alexander Clark, Chris Fox, and Shalom Lappin, editors, *The Handbook of Computational Linguistics and Natural Language Processing*, pages 271–295. Wiley-Blackwell.

Noah A. Smith. 2011. *Linguistic Structure Prediction*. Morgan & Claypool, San Rafael, CA.