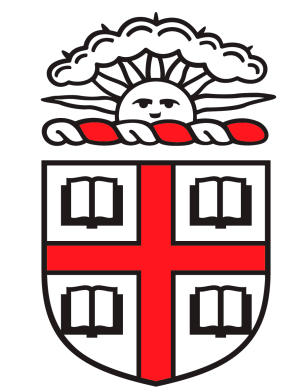


# Collecting Diverse Natural Language Inference Problems for Sentence Representation Evaluation

Adam Poliak, Aparajita Haldar, Rachel Rudinger, J. Edward Hu, Ellie Pavlick, Aaron Steven White, Benjamin Van Durme



## Natural Language Inference (NLI)

Task: Does a premise likely entail a hypothesis?

Premise: *The brown cat ran*

Hypothesis: *The animal moved*

## Motivation for the DNC

Unclear tests of models' reasoning capabilities

Common dataset construction methods lead to biases

Probing sets are small

## DNC: Diverse NLI Collection

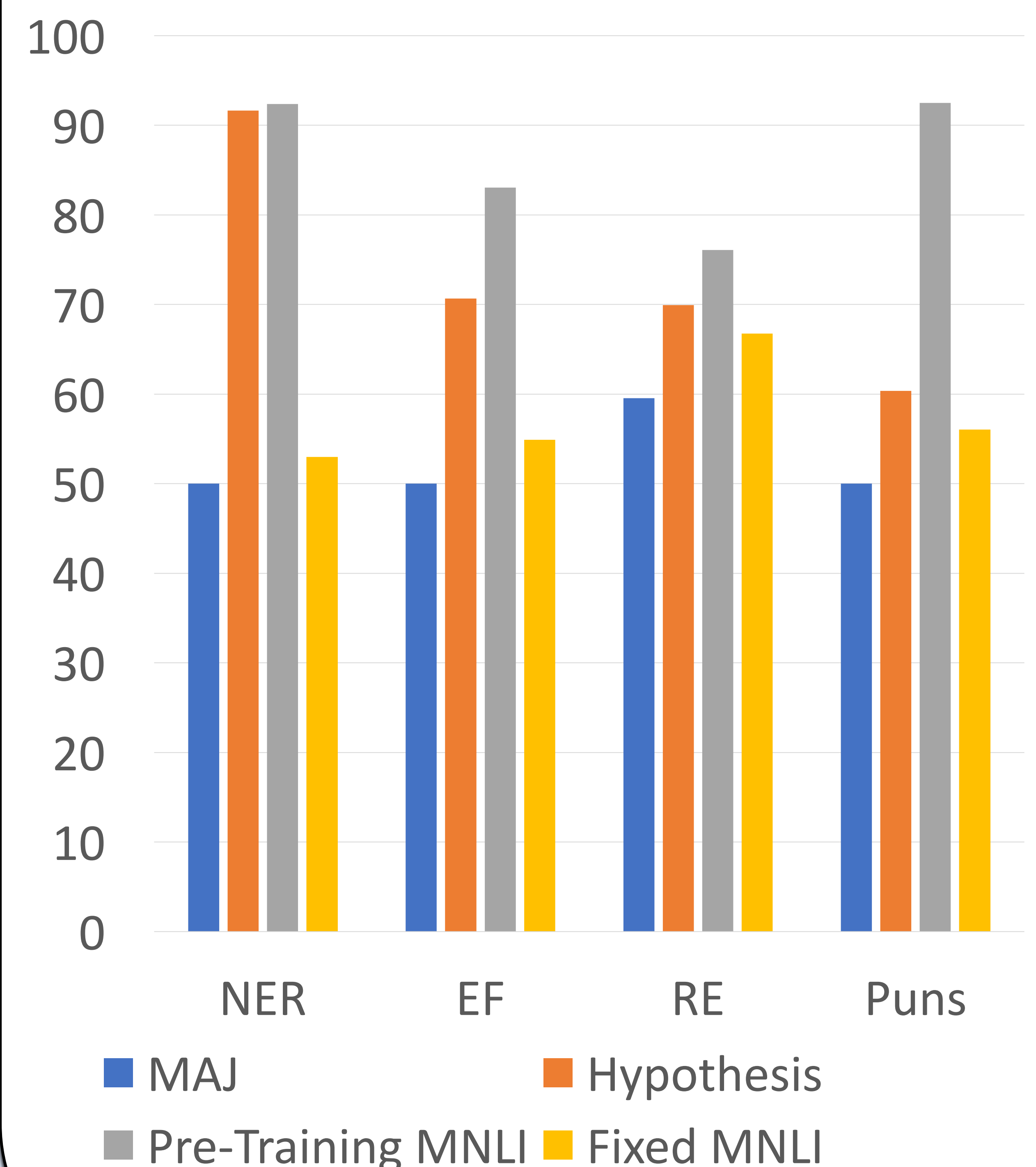
Convert 7 semantic phenomena into NLI from 13 public NLP datasets

Large Scale → 500K labeled examples

## DNC Examples

|                          |   |
|--------------------------|---|
| Event Factuality         | I would like to learn how<br>The learning did not happen  |
| Named Entity Recognition | Ms. Rice said the United States must work intensively<br>Ms. is a person's title                          |
| Gendered Anaphora        | The student met with the architect to view her blueprints for inspiration<br>The architect has blueprints |
| VerbNet                  | The Romans destroyed the city<br>The Romans caused the destroying   |
| VerbCorner               | Molly wheeled Lisa to Rachel<br>Someone moved from their location   |
| Relation Extraction      | At least 100,000 Chinese live in Lhasa, outnumbering Tibetans two to one<br>Tibetans live in Lhasa        |
| Puns                     | Jordan heard that my skiing skills are really going downhill<br>Jordan heard a pun                        |
| Sentiment Analysis       | When asked about the product, Liam said, "Don't waste your money"<br>Liam did not like the product        |

## Experiments & Results



Paper



Data

