

Attending to Entities for Better Text Understanding

Pengxiang Cheng (pxcheng@utexas.edu)

Katrin Erk (katrin.erk@utexas.edu)

Motivation

- Large-scale pre-trained language models are quickly approaching human performance on many NLP tasks.
- But still trailing on problems where complex and long-distance reasoning is necessary.
- Can **coreference knowledge** be helpful on such tasks, similar to syntactic knowledge helping SRL [1]?
- Experiments on LAMBADA [2], a word prediction task explicitly designed to require **broad discourse context**.

An Example from LAMBADA

Context: "By the way, Elizabeth asked if I'd seen you," Tony lied. He wanted Jon to leave so he could talk with Ezekiel alone. There was something that aunt Casey, Patella and Gabriella had said about Tom that had bothered him ever since meeting Ezekiel earlier that afternoon.

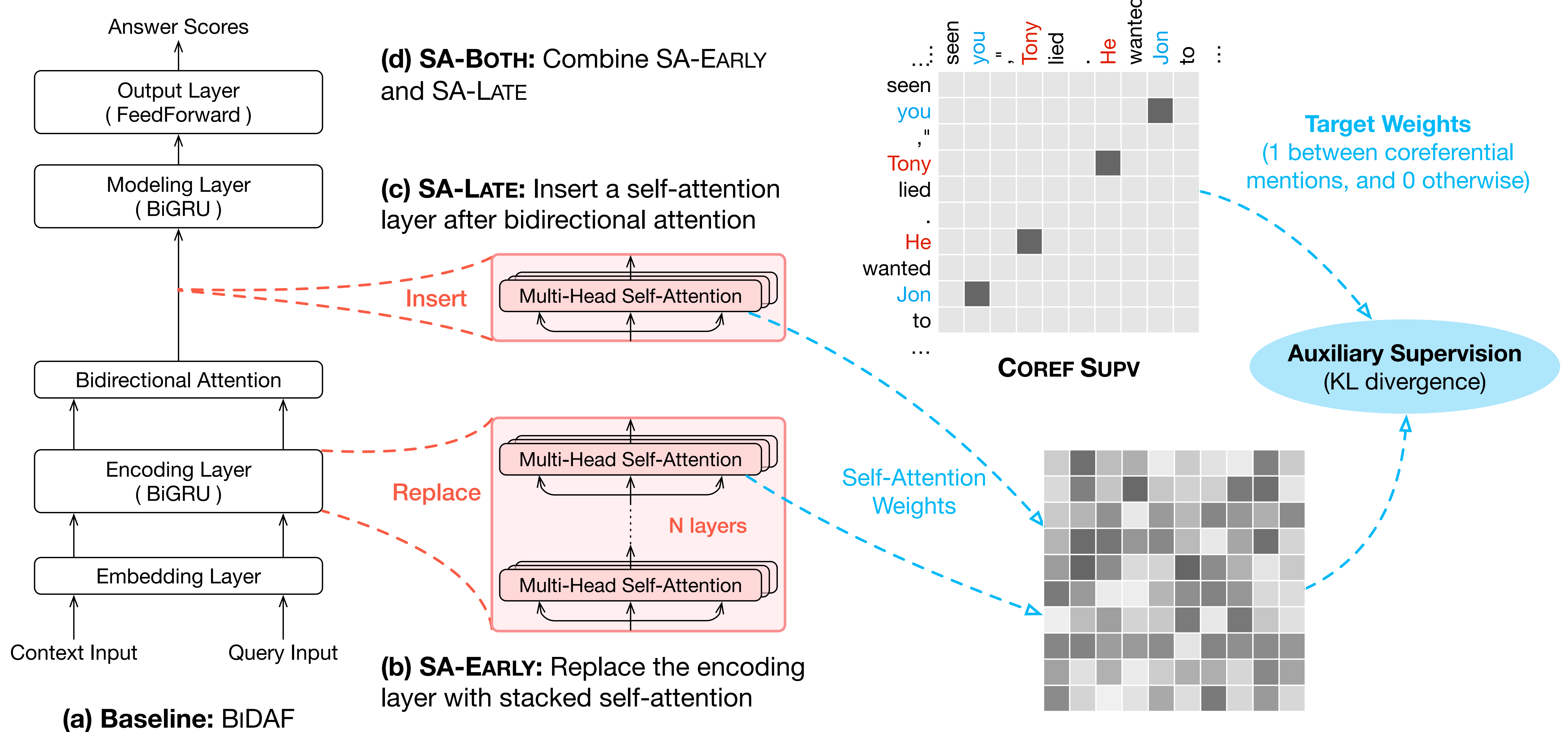
Target sentence: "I'm sure she'll find me," Jon remarked curtly, trying to cut short the conversation with __ .

Target word: Tony

Approach

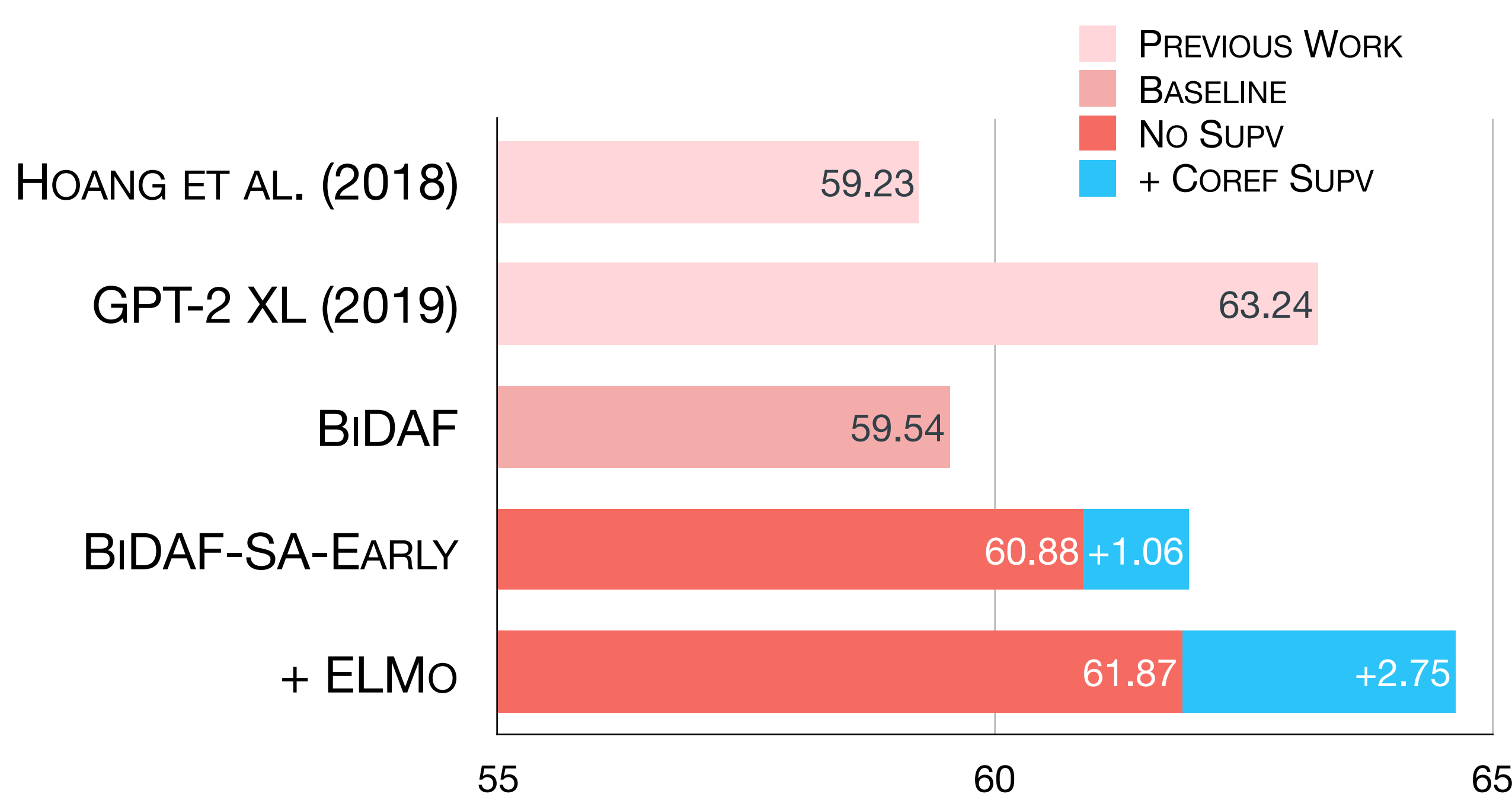
1. We fuse the stacked self-attention architecture into a standard reading comprehension model (BiDAF [3]).

2. We apply **auxiliary supervision** on self-attention weights based on **coreference chains**.



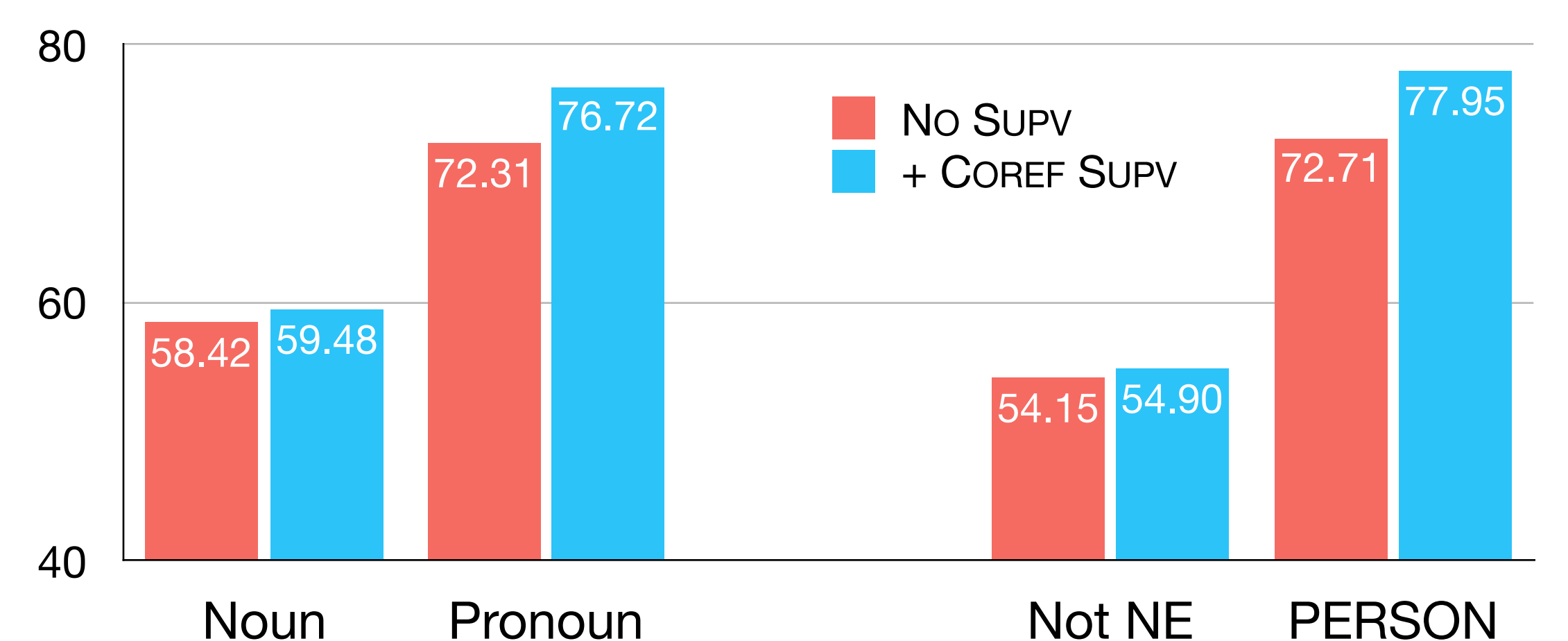
Results

- With coreference supervision, our model **outperforms** the largest GPT-2 model, while contains **a tiny fraction** of tunable parameters (2.6 million vs. 1.5 billion).

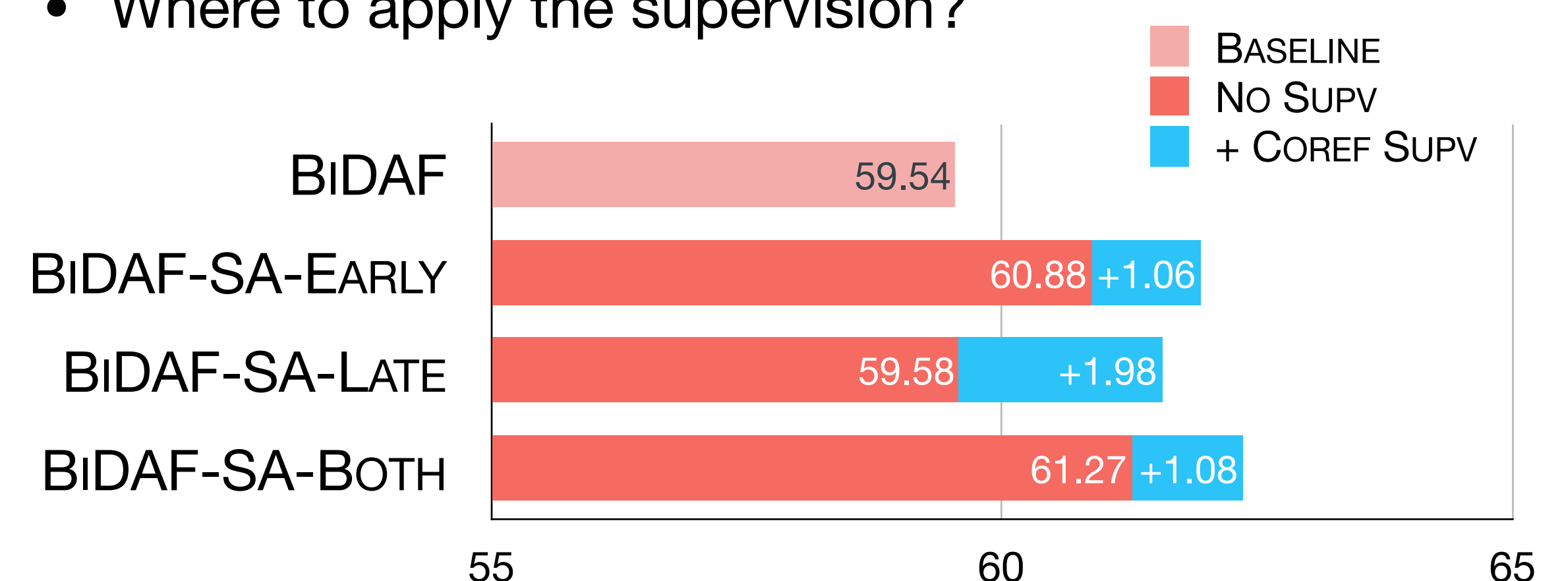


Analysis

- Does it really learn coreference knowledge?
 - Results breakdown by the POS / NER tag of the target word.
 - Significant improvements on pronouns and PERSON entities.



- Where to apply the supervision?



References

- [1] Strubell, Emma, et al. "Linguistically-informed self-attention for semantic role labeling." *EMNLP*. 2018.
- [2] Paperno, Denis, et al. "The LAMBADA dataset: Word prediction requiring a broad discourse context." *ACL*. 2016.
- [3] Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." *ICLR*. 2017.