Conjunction-Aware Word-level coreference resolution

CRAC 2023
(the Sixth Workshop on Computational Models of Reference, Anaphora and Coreference)

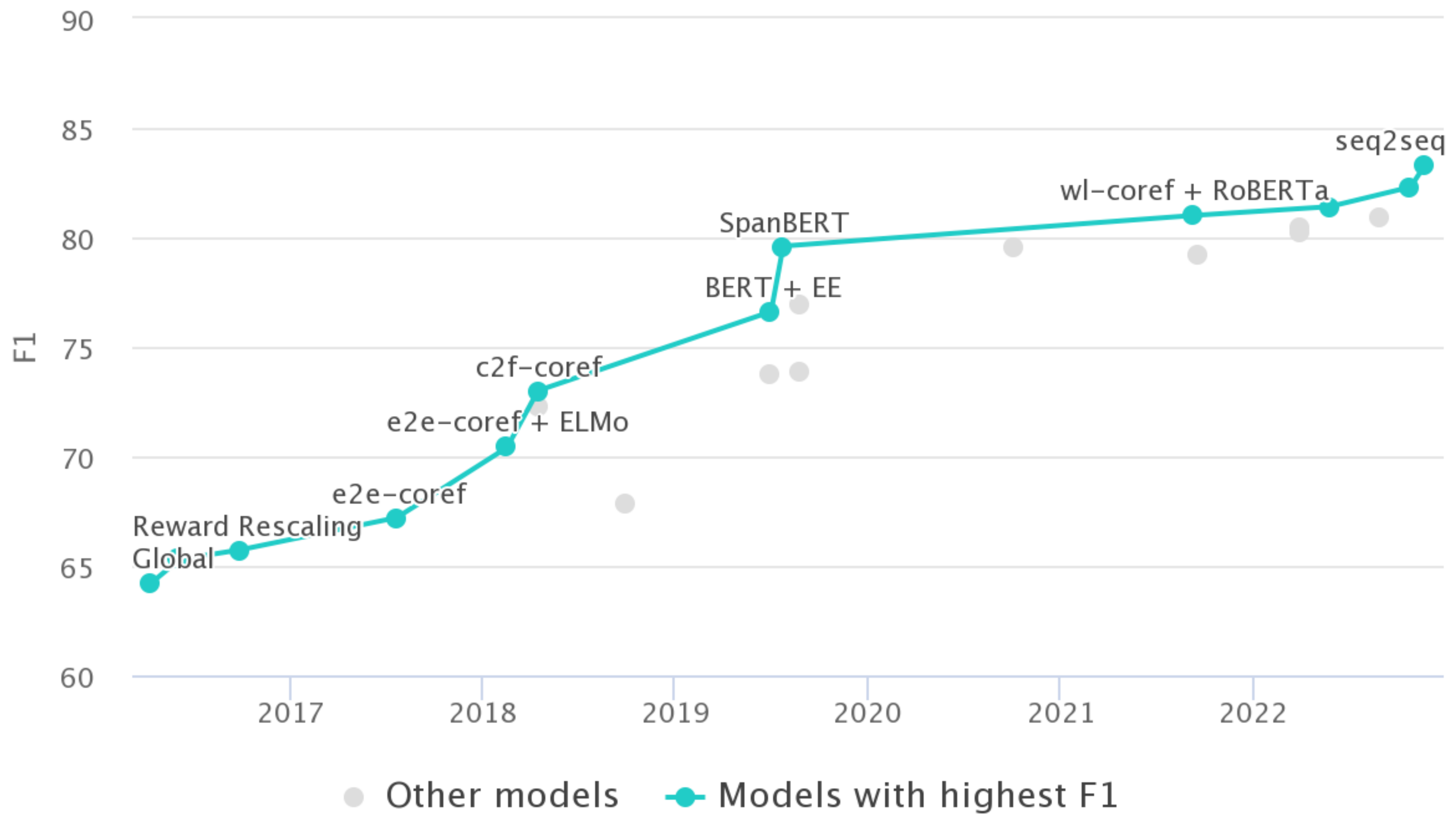
Karel D'Oosterlinck — 7th of December
3rd year PhD student, Ghent University
karel.doosterlinck@ugent.be

GHENT UNIVERSITY

Stanford University

Neural coreference resolution today

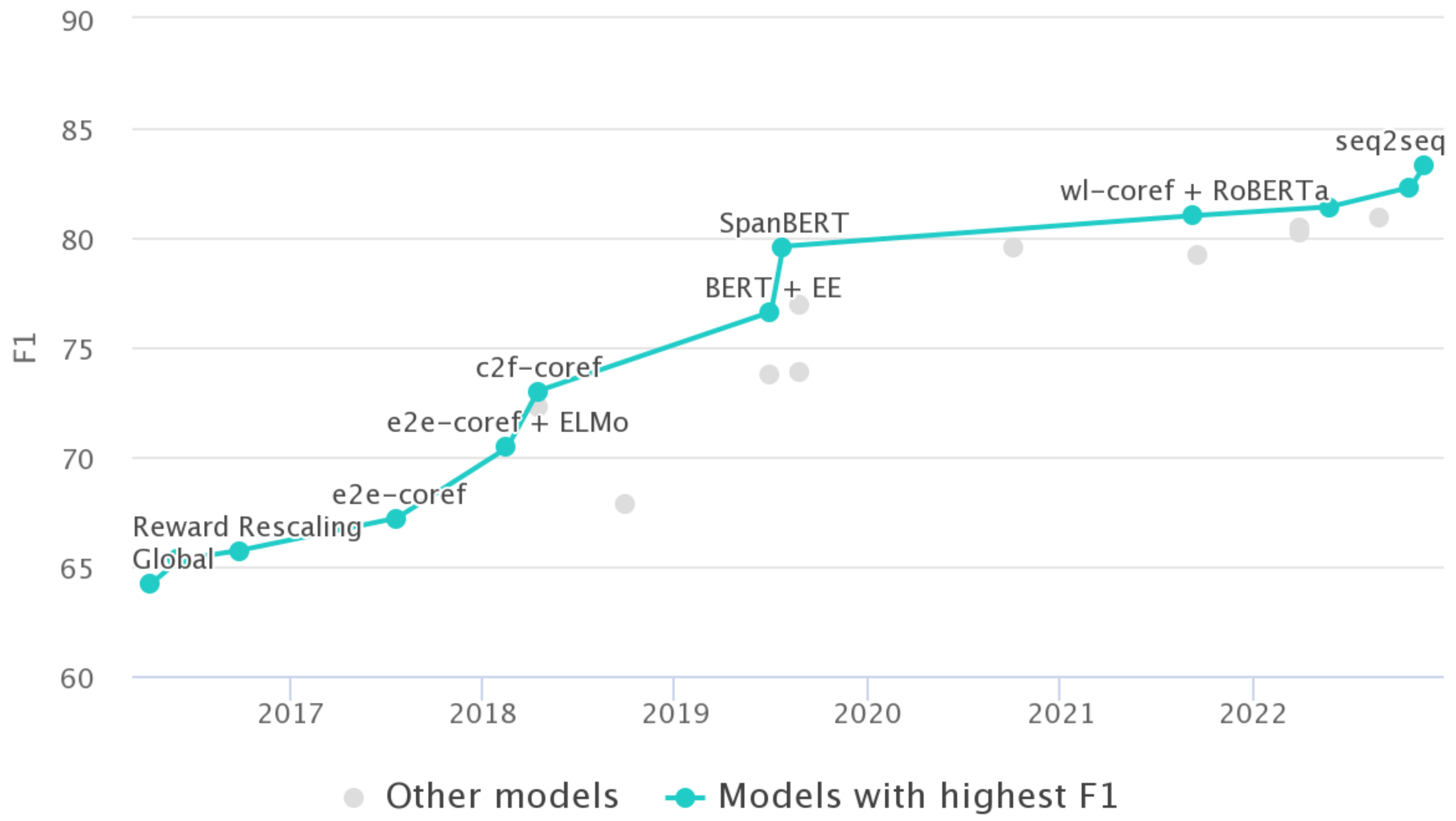
as measured on the OntoNotes English test split



https://paperswithcode.com/sota/coreference-resolution-on-ontonotes

Neural coreference resolution today

as measured on the OntoNotes English test split

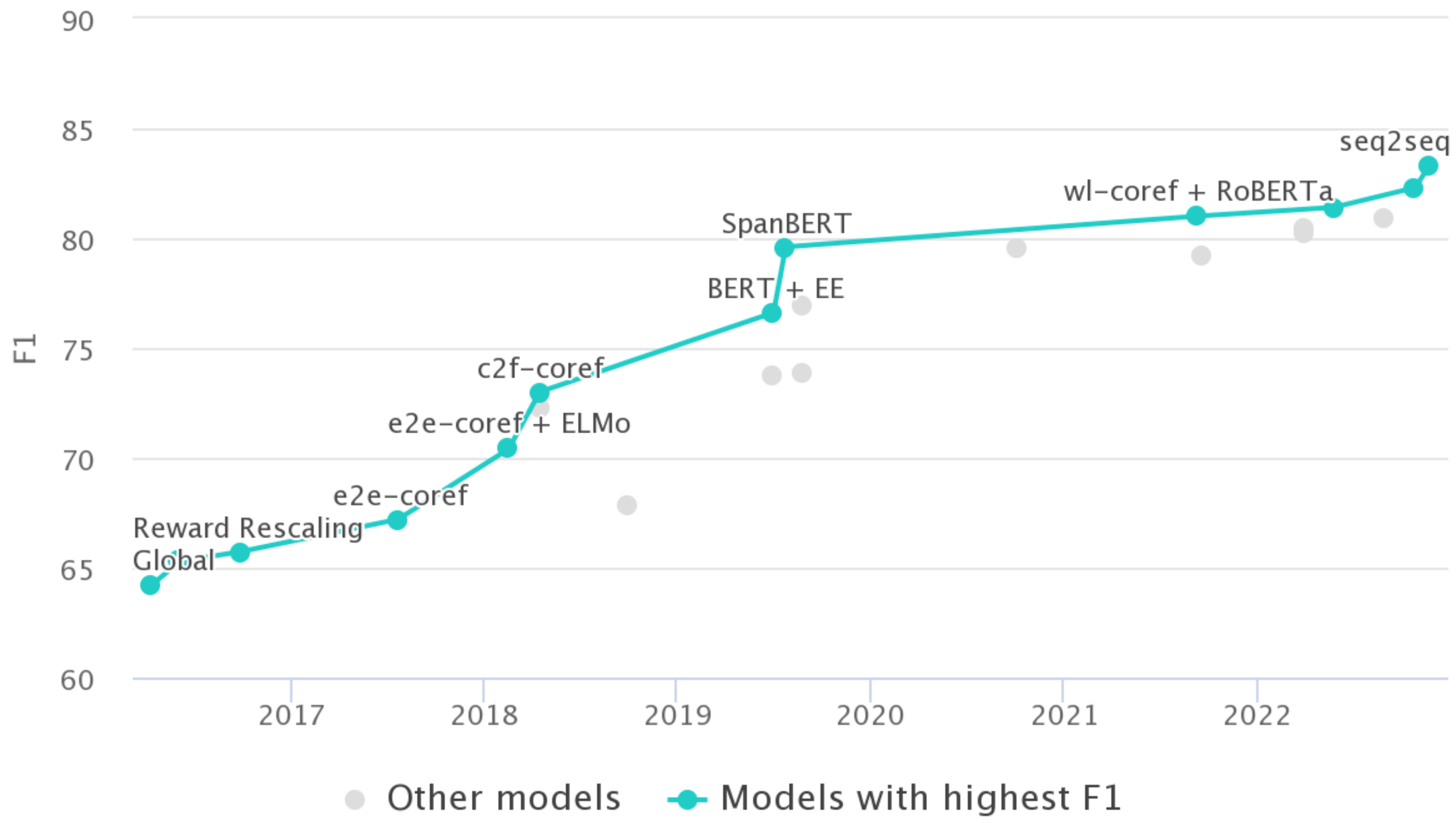


	LM calls	LM params.
link-append	$O(n)$	13B
corefqa	$O(n^2)$	340M
ASP	$O(n)$	11B

https://paperswithcode.com/sota/coreference-resolution-on-ontonotes

Neural coreference resolution today

as measured on the OntoNotes English test split



https://paperswithcode.com/sota/coreference-resolution-on-ontonotes

	LM calls	LM params.
link-append	$O(n)$	13B
corefqa	$O(n^2)$	340M
ASP	$O(n)$	11B

Some applications need more efficient methods

Conjunction-Aware Word-level (CAW) Coref

1. We revisit Word-Level coreference resolution (Dobrovolskii, 2021)

Conjunction-Aware Word-level (CAW) Coref

1. We revisit Word-Level coreference resolution (Dobrovolskii, 2021)
2. We identify two routine error cases of this model and their cause

Conjunction-Aware Word-level (CAW) Coref

1. We revisit Word-Level coreference resolution (Dobrovolskii, 2021)
2. We identify two routine error cases of this model and their cause
3. We update the training and close the relative gap with SOTA by ~35% (CONLL F1 on OntoNotes English; +0.9% absolute)

1. Word-Level coreference resolution

Dobrovolskii, Vladimir. "Word-Level Coreference Resolution." Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021.

Previous end-to-end systems (Lee et al., 2017):

1. Create $O(n^2)$ span representations
2. Compare $O(n^4)$ span pairs

1. Word-Level coreference resolution

Dobrovolskii, Vladimir. "Word-Level Coreference Resolution." Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021.

Previous end-to-end systems (Lee et al., 2017):

1. Create $O(n^2)$ span representations
2. Compare $O(n^4)$ span pairs

Word-Level coreference resolution:

1. Create $O(n)$ word representations
2. Compare $O(n^2)$ word pairs
3. Find spans associated with coreferent words

2. Routine WL-coref errors with conjunctions

Error type 1: WL-coref does not link Tom and Mary to They

2. Routine WL-coref errors with conjunctions

Error type 1: WL-coref does not link Tom and Mary to They

Tom and Mary are playing. He is 7 years old. They are siblings.

Error type 2: WL-coref links They to Tom, instead of Tom and Mary

Tom and Mary are talking. They are talking.

3. Update training data and retrain

To train WL-coref, you need:

1. Word-level coreference data
2. Word-to-span data

3. Update training data and retrain

To train WL-coref, you need:

1. Word-level coreference data
2. Word-to-span data

Dobrovolski (2021) selects the head of the span as the word that depends on a word outside of the span.
—> ***Tom*** *and Mary*

3. Update training data and retrain

To train WL-coref, you need:

1. Word-level coreference data
2. Word-to-span data

Dobrovolski (2021) selects the head of the span as the word that depends on a word outside of the span.
—> ***Tom*** *and Mary*

We define the coordinating conjunction as head when present.
—> *Tom* ***and*** *Mary*

3. Update training data and retrain

To train WL-coref, you need:

1. Word-level coreference data
2. Word-to-span data

Dobrovolski (2021) selects the head of the span as the word that depends on a word outside of the span.
—> ***Tom** and Mary*

We define the coordinating conjunction as head when present.
—> *Tom **and** Mary*

+0.9% absolute gain, ~35% relative gain with SOTA (CONLL F1)

Wrapping up

Still gains to be had for efficient coreference resolution.

My amazing collaborators:

- Semere Kiros Bitew, Brandon Papineau, Christopher Potts, Thomas Demeester, Chris Develder
- Houjun Liu, John Bauer

arXiv paper **GitHub** **Stanza**

Appendix: link-append system

Bohnet, Bernd, Chris Alberti, and Michael Collins. "Coreference Resolution through a seq2seq Transition-Based System." *Transactions of the Association for Computational Linguistics* 11 (2023): 212-226.

Input: *Speaker-A* I still have n't gone to that fresh French restaurant by your house
Prediction: SHIFT: next sentence

Input: *Speaker-A* I_2 still have n't gone to that fresh French restaurant by your house *Speaker-A* I_{17} 'm like dying to go there
Prediction:
A $I_{17} \rightarrow I_2$
B SHIFT: next sentence

Input: *Speaker-A* [1 I] still have n't gone to that fresh French restaurant by your house *Speaker-A* [1 I] 'm like dying to go there *Speaker-B* You mean the one right next to the apartment
Prediction:
A You → [1
B the apartment → your house
C the one right next to the apartment → that fresh French restaurant by your house
D SHIFT: next sentence

Input: *Speaker-A* [1 I] still have n't gone to [3 that fresh French restaurant by [2 your house]] *Speaker-A* [1 I] 'm like dying to go there *Speaker-B* [1 You] mean [3 the one right next to [2 the apartment]] *Speaker-B* yeah yeah yeah
Prediction: SHIFT: next sentence

Figure 1: Example of one of our transition-based coreference systems, the *Link-Append* system. The system processes a single sentence at a time, using an input encoding of the prior sentences annotated with coreference clusters, followed by the new sentence. As output, the system makes predictions that link mentions in the new sentence to either previously created coreference clusters (e.g., "You → [1") or when a new cluster is created, to previous mentions (e.g., "the apartment → your house"). The system predicts "SHIFT" when processing of the sentence is complete. Note in the figure we use the word indices 2 and 17 to distinguish the two incidences of "I" in the text.

Appendix: autoregressive-structure-prediction

Liu, T., Jiang, Y., Monath, N., Cotterell, R., & Sachan, M. (2022). Autoregressive structured prediction with language models. *arXiv preprint arXiv:2210.14698*.

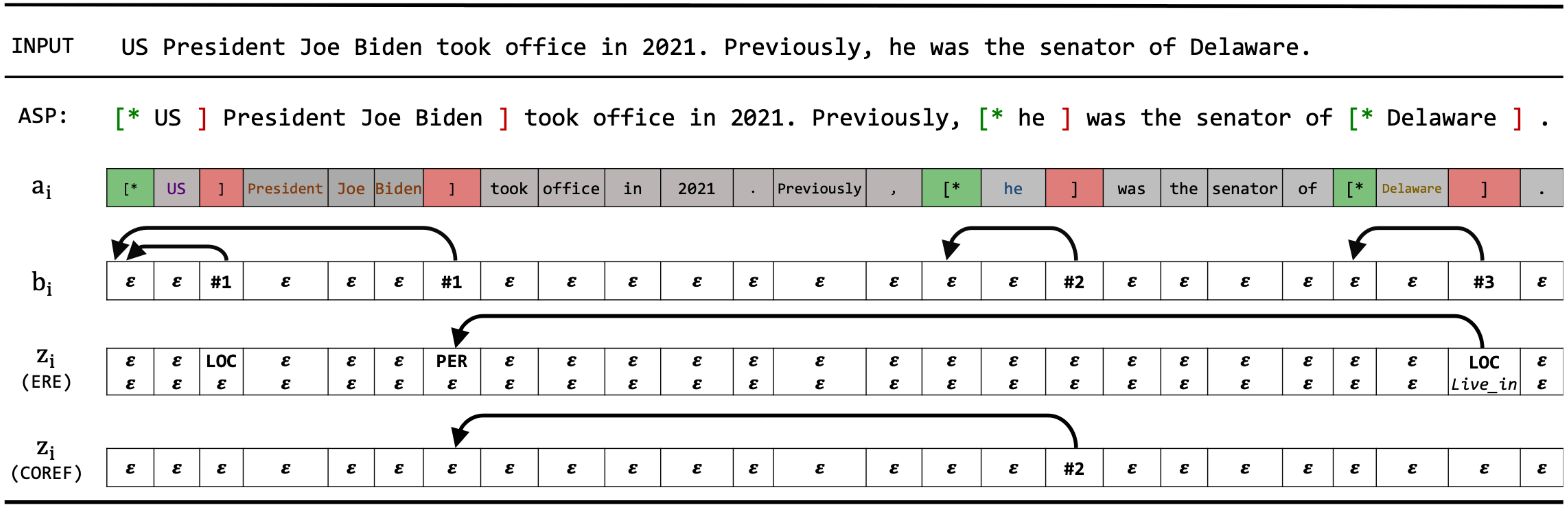


Figure 1: Illustration of the target outputs of our framework on coreference resolution (COREF) and end-to-end relation extraction (ERE). The lower part illustrates the decoding process of our model. The actions $\boldsymbol{y}_i$ are color-coded as], [* and copy. The structure random variables z_i are presented along with coreference links or relation links. We present words in the copy cells merely as an illustration.

Appendix: full results table

	LM calls	LM params.	Link compl.	MUC P	MUC R	MUC F1	B^3 P	B^3 R	B^3 F1	$CEAF_{\phi 4}$ P	$CEAF_{\phi 4}$ R	$CEAF_{\phi 4}$ F1	**Avg. F1**
link-append	$O(n)$	13B	/	**87.4**	88.3	**87.8**	81.8	**83.4**	**82.6**	79.1	**79.9**	**79.5**	**83.3**
corefqa	$O(n^2)$	340M	/	88.6	87.4	88.0	**82.4**	82.0	82.2	**79.9**	78.3	79.1	83.1
ASP	$O(n)$	11B	/	86.1	**88.4**	87.2	80.2	83.2	81.7	78.9	78.3	78.6	82.5
LingMess	1	355M	$O(n^4)$	85.1	88.1	**86.6**	**78.3**	**82.7**	**80.5**	76.1	78.5	77.3	81.4
s2e	1	355M	$O(n^4)$	**85.2**	86.6	85.9	77.9	80.3	79.1	75.4	76.8	76.1	80.3
CAW (ours)	1	355M	$O(n^2)$	85.1	**88.2**	**86.6**	77.0	78.0	77.5	**78.0**	**83.2**	**80.6**	**81.6**
WL†	1	355M	$O(n^2)$	84.8	87.5	86.1	76.1	76.7	76.6	77.1	82.1	79.5	80.7

Table 1: Results on the OntoNotes 5.0 English test set. Scores calculated with official scorer (Pradhan et al., 2014) or taken from original publication if available. **Avg. F1** is the main metric. We report the amount of LM calls and parameters of the LM used, as well as the coreference linking complexity if applicable. † Dobrovolskii (2021) reports an Avg. F1 of 81.0 as the best WL-coref run on the test set, while we report the result of our first run for both WL-coref and CAW-coref.

Appendix: some examples

Model	Step	Prediction	Correct
WL-coref	word	Tom and Anna are talking. They are talking.	Yes
WL-coref	span	Tom and Anna are talking. They are talking.	No
CAW-coref	word	Tom and Anna are talking. They are talking.	Yes
CAW-coref	span	Tom and Anna are talking. They are talking.	Yes
WL-coref	word	My friend David and my dad Bert are talking. They are talking.	No
WL-coref	span	My friend David and my dad Bert are talking. They are talking.	No
CAW-coref	word	My friend David and my dad Bert are talking. They are talking.	Yes
CAW-coref	span	My friend David and my dad Bert are talking. They are talking.	Yes
WL-coref	word	The Guardian and The Chronicle had a secret meeting . Both newspapers are on thin ice .	No
WL-coref	span	The Guardian and The Chronicle had a secret meeting . Both newspapers are on thin ice .	No
CAW-coref	word	The Guardian and The Chronicle had a secret meeting . Both newspapers are on thin ice .	Yes
CAW-coref	span	The Guardian and The Chronicle had a secret meeting . Both newspapers are on thin ice .	Yes

Table 2: Three hand-crafted examples and their word-level and span-level predictions for WL-coref and CAW-coref. Coreferent predictions are indicated with a colored box, where each unique entity has the same color. Predictions are considered correct or not correct for their respective step in the word-level pipeline.