Understanding Mention Detector-Linker Interaction in Neural Coreference Resolution

Zhaofeng Wu and Matt Gardner

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

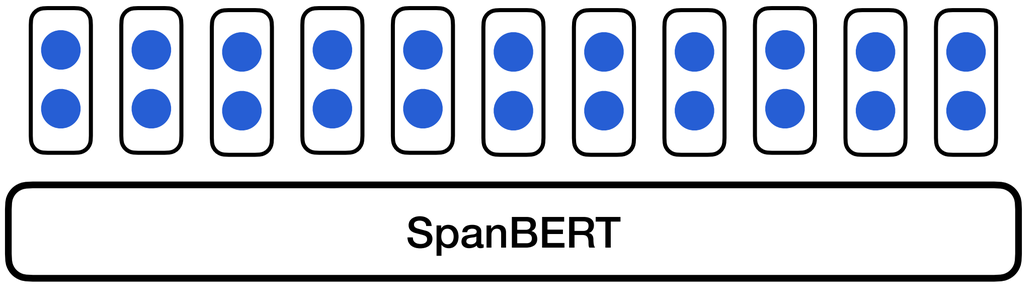
AI2

Neural Coreference Resolution

Neural Coreference Resolution

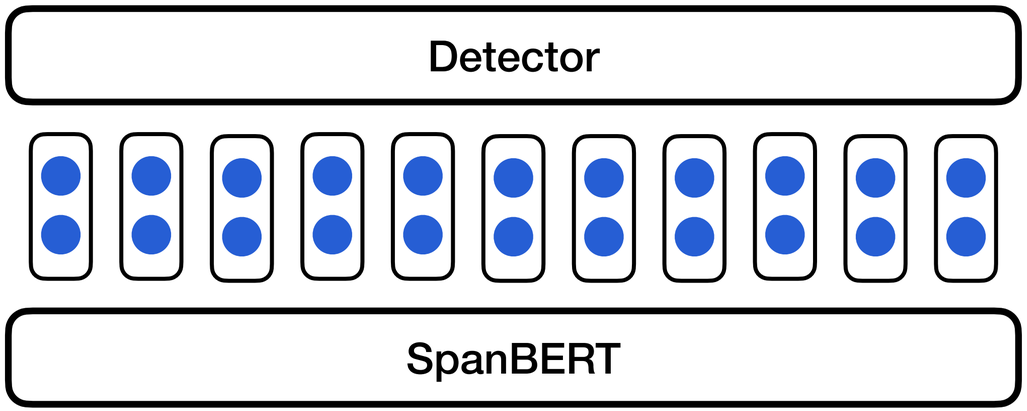
He and I laughed at my dog who bit himself.

Neural Coreference Resolution



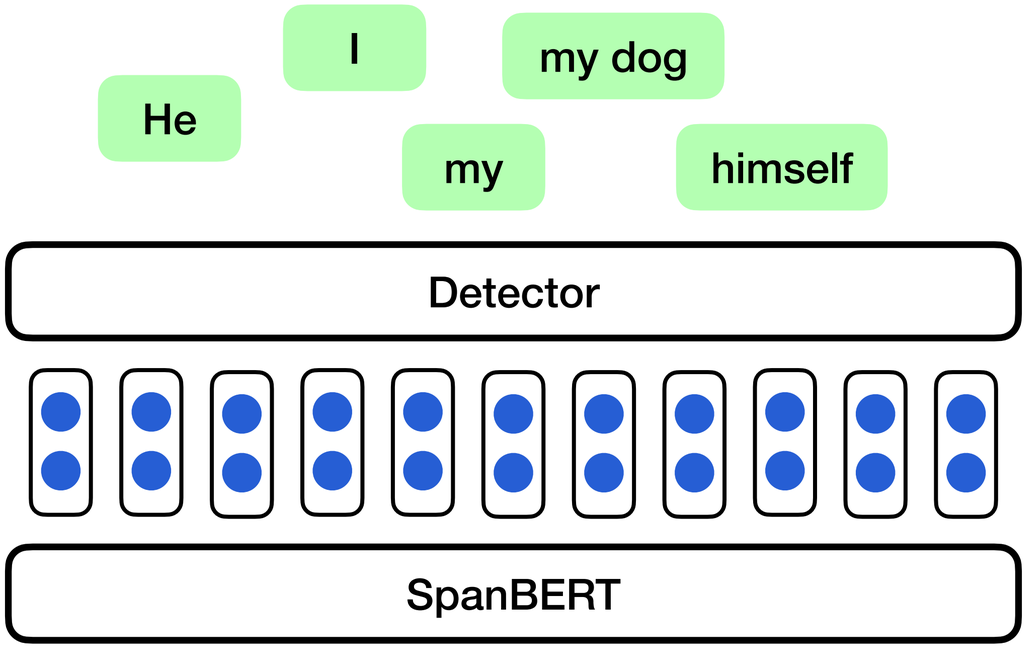
He and I laughed at my dog who bit himself.

Neural Coreference Resolution



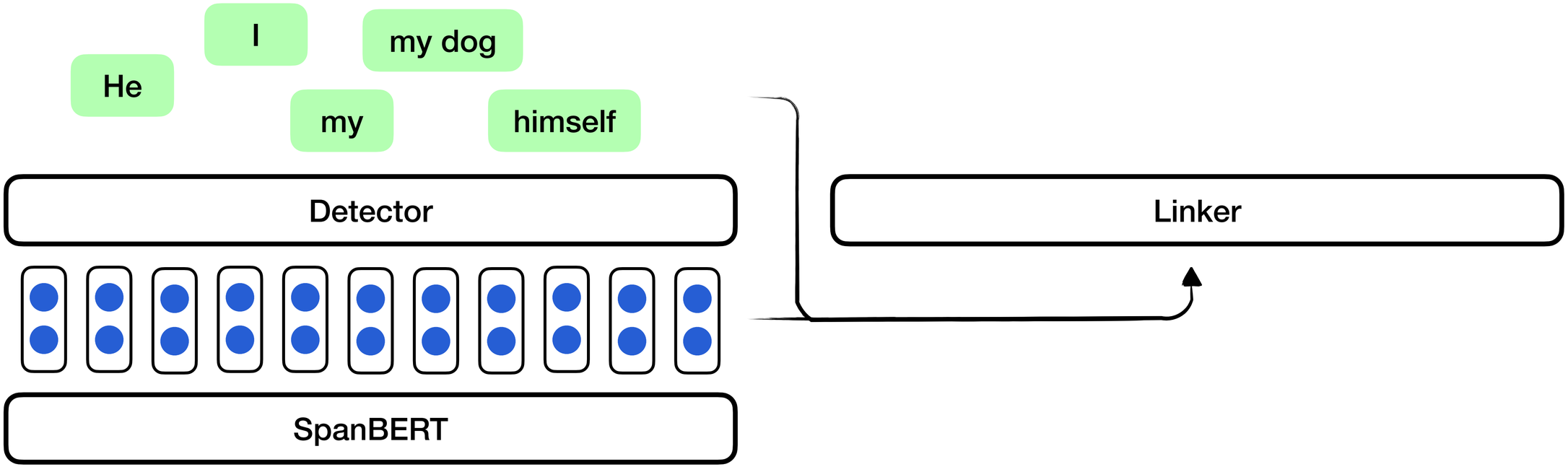
He and I laughed at my dog who bit himself.

Neural Coreference Resolution



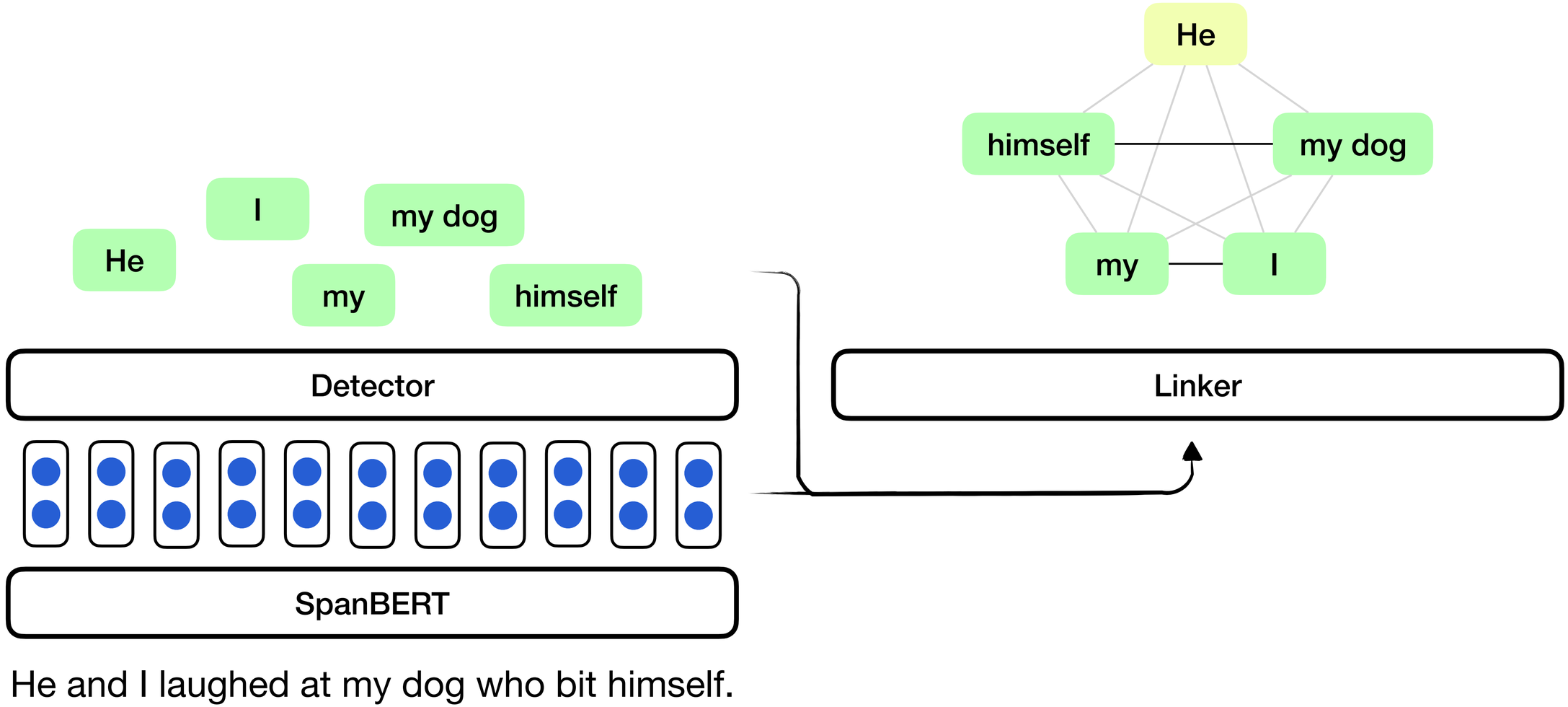
He and I laughed at my dog who bit himself.

Neural Coreference Resolution

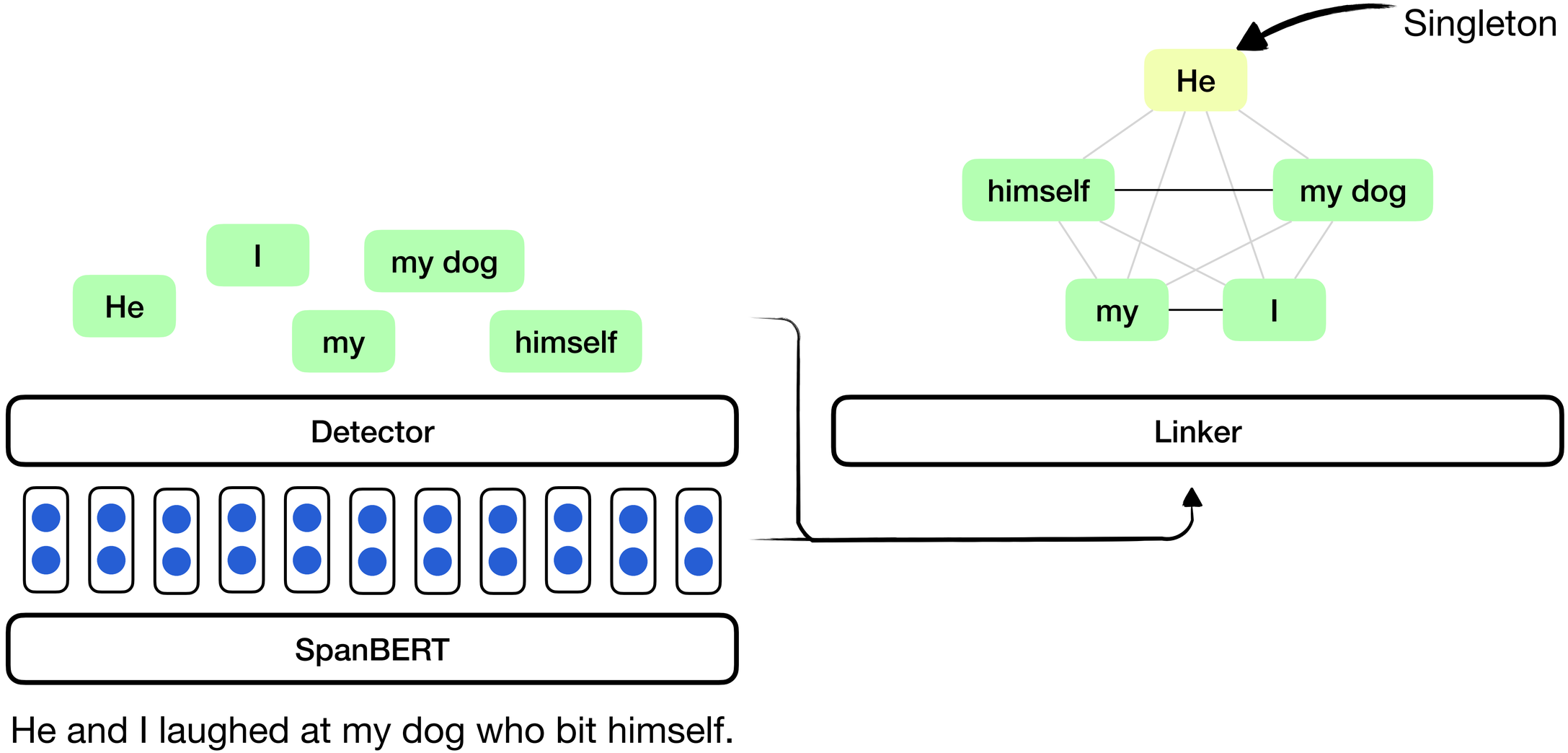


He and I laughed at my dog who bit himself.

Neural Coreference Resolution

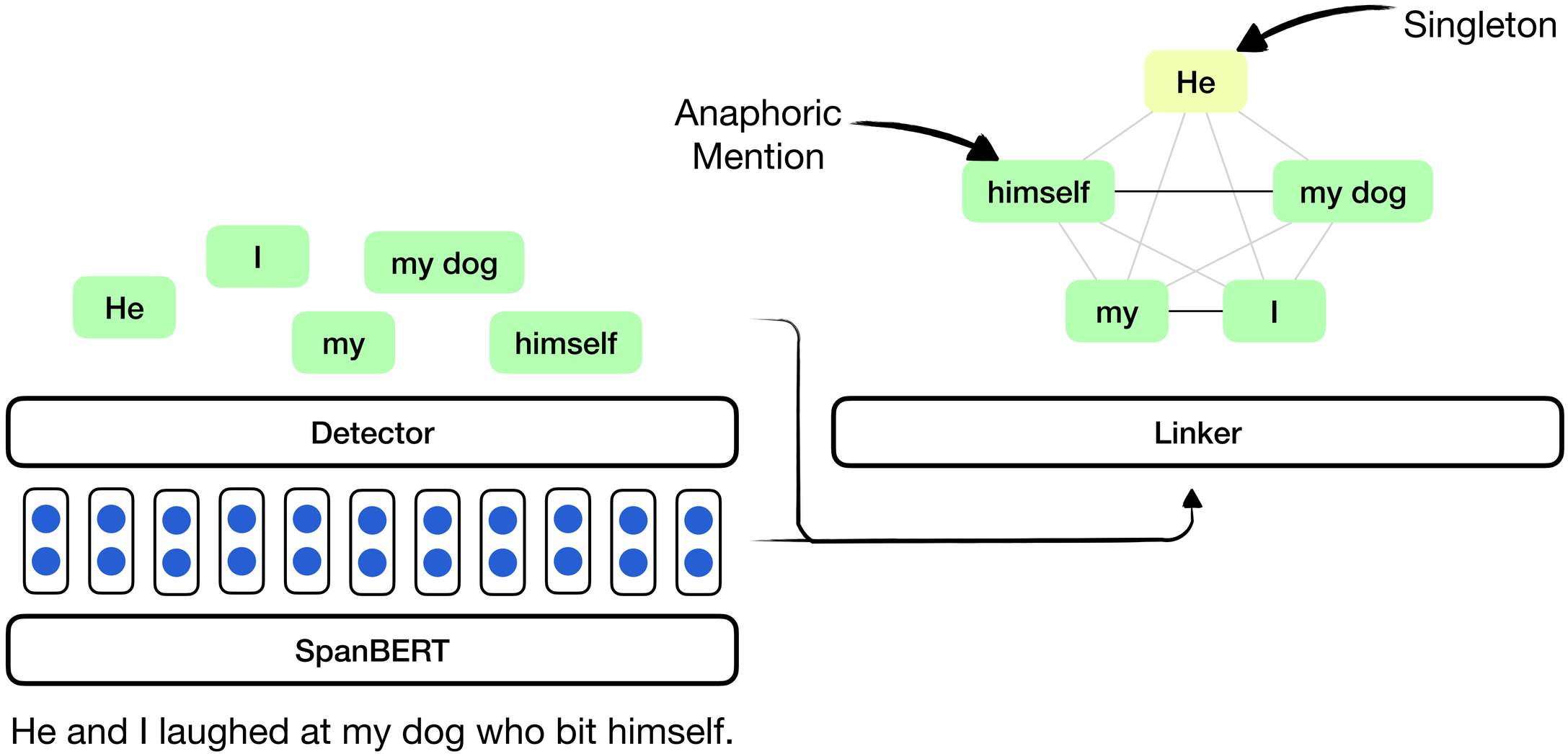


Neural Coreference Resolution

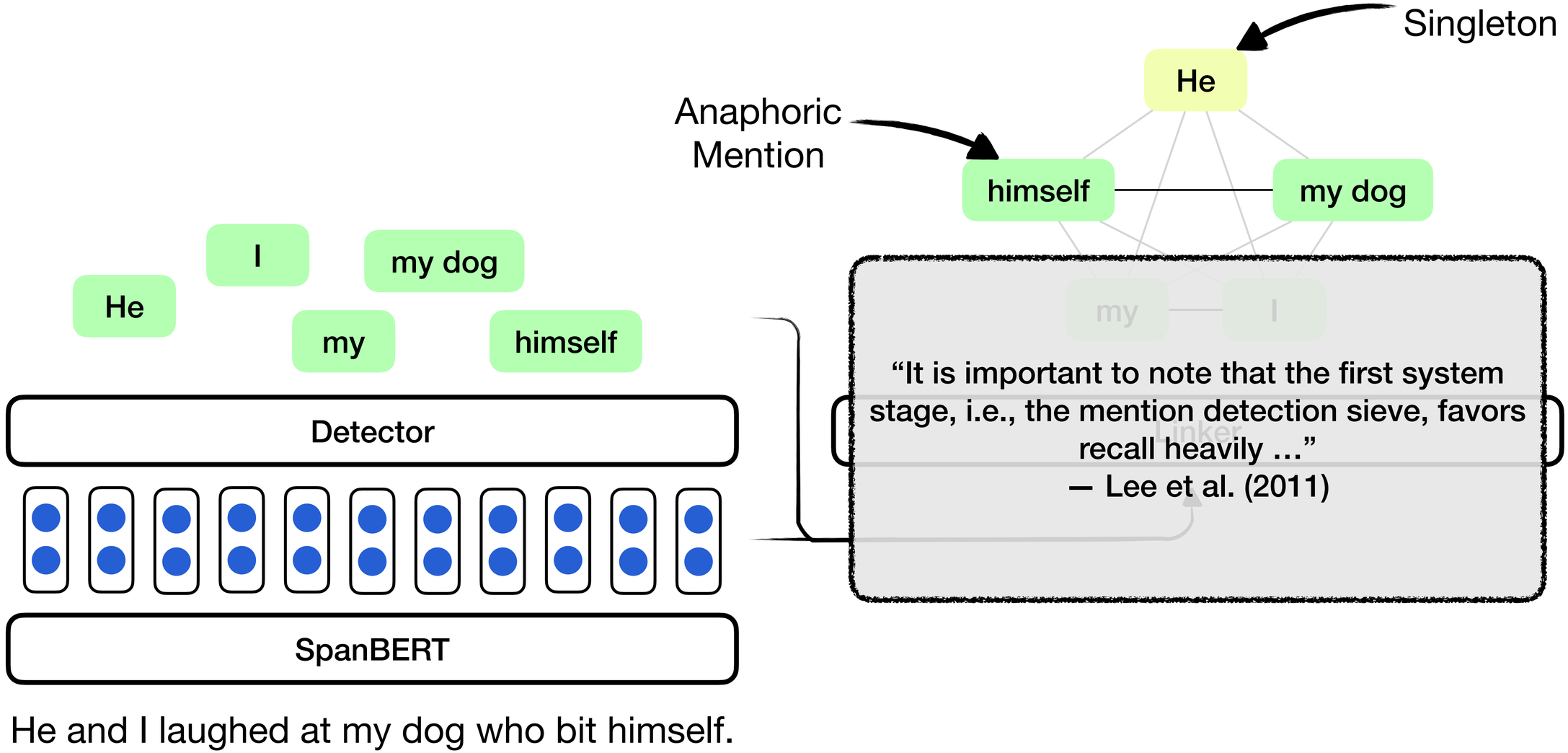


He and I laughed at my dog who bit himself.

Neural Coreference Resolution



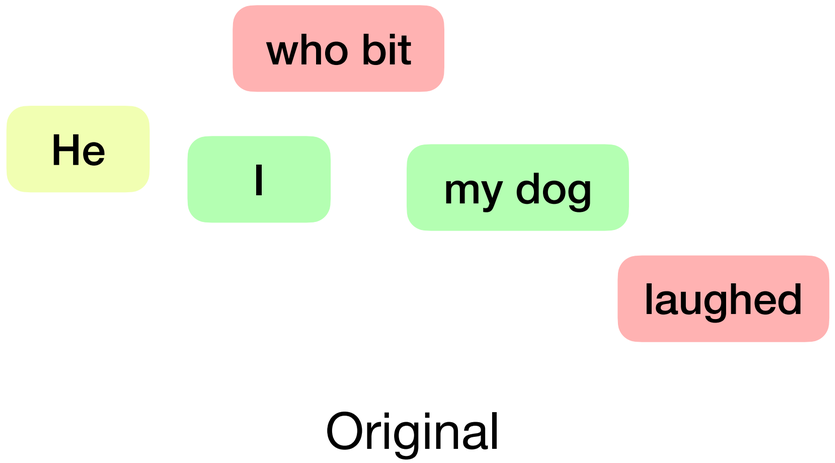
Neural Coreference Resolution



Is Detector Recall Always More Important Than Precision?

Oracle Detectors

Oracle Detectors



Original

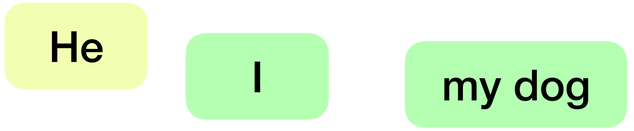
Oracle Detectors

Original

Perfect Recall

Oracle Detectors

Original

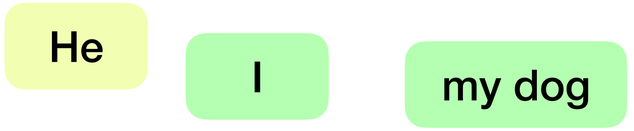


Perfect Precision

Perfect Recall

Oracle Detectors

Original



Perfect Precision

Perfect Recall

Perfect Recall & Precision

Datasets & Model Settings

Datasets & Model Settings

- Ontonotes (English)

Datasets & Model Settings

- Ontonotes (English)
- PreCo

Datasets & Model Settings

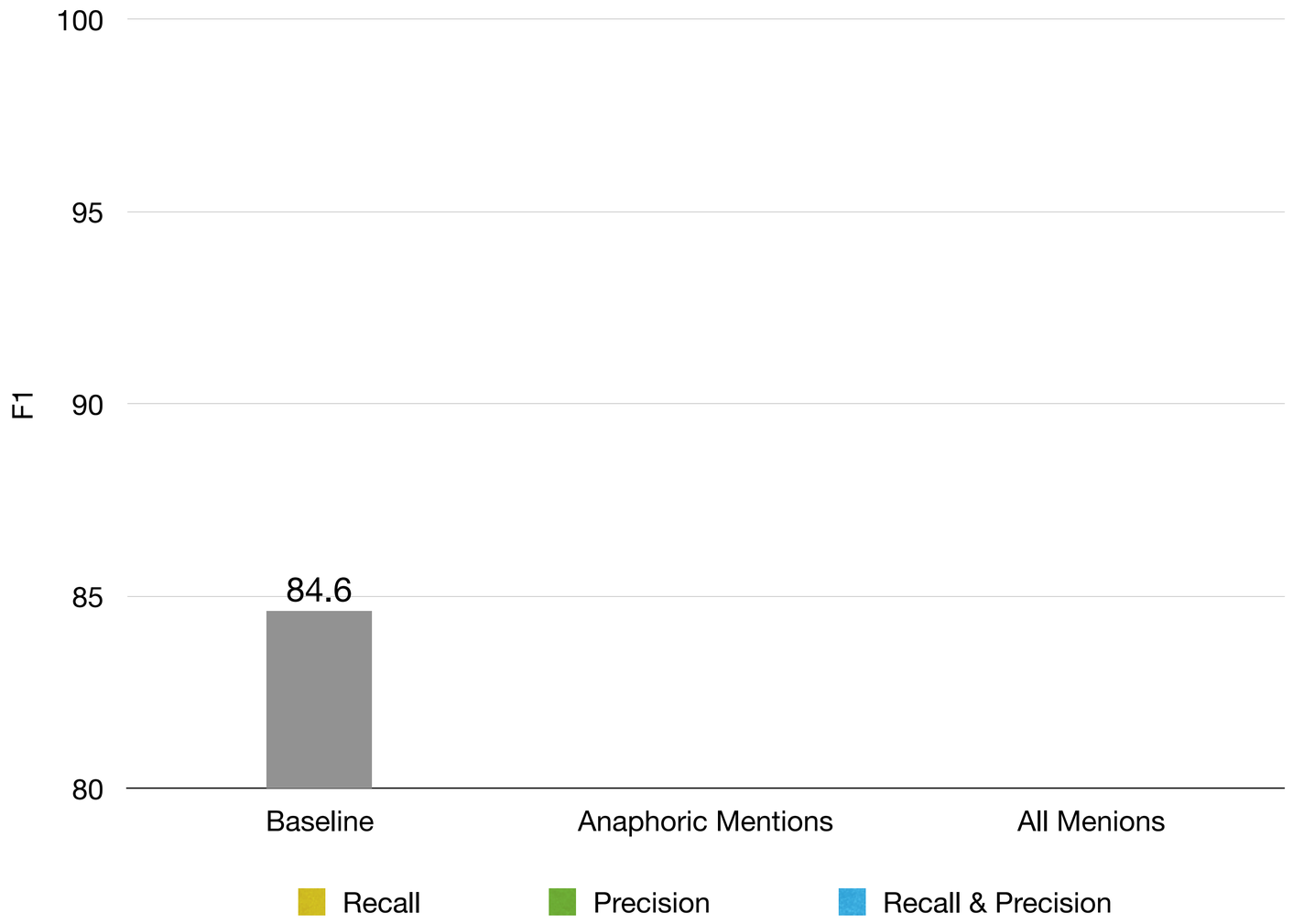
- Ontonotes (English)
- PreCo
- Embedder: SpanBERT-large

Datasets & Model Settings

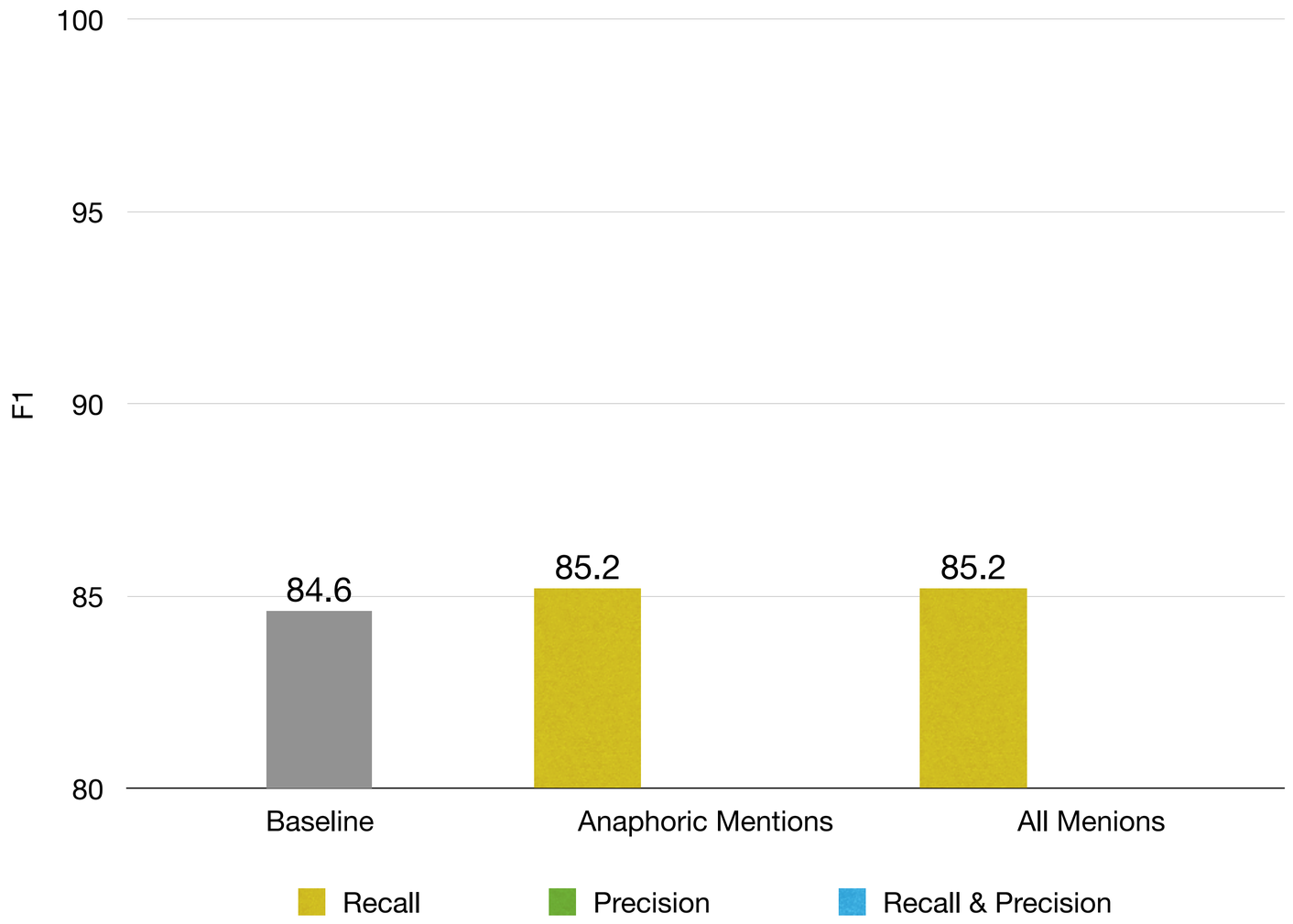
- Ontonotes (English)
- PreCo
- Embedder: SpanBERT-large
- Experimental configuration
  - https://github.com/allenai/allennlp-models/blob/main/training_config/coref/coref_spanbert_large.jsonnet

Importance of Precision

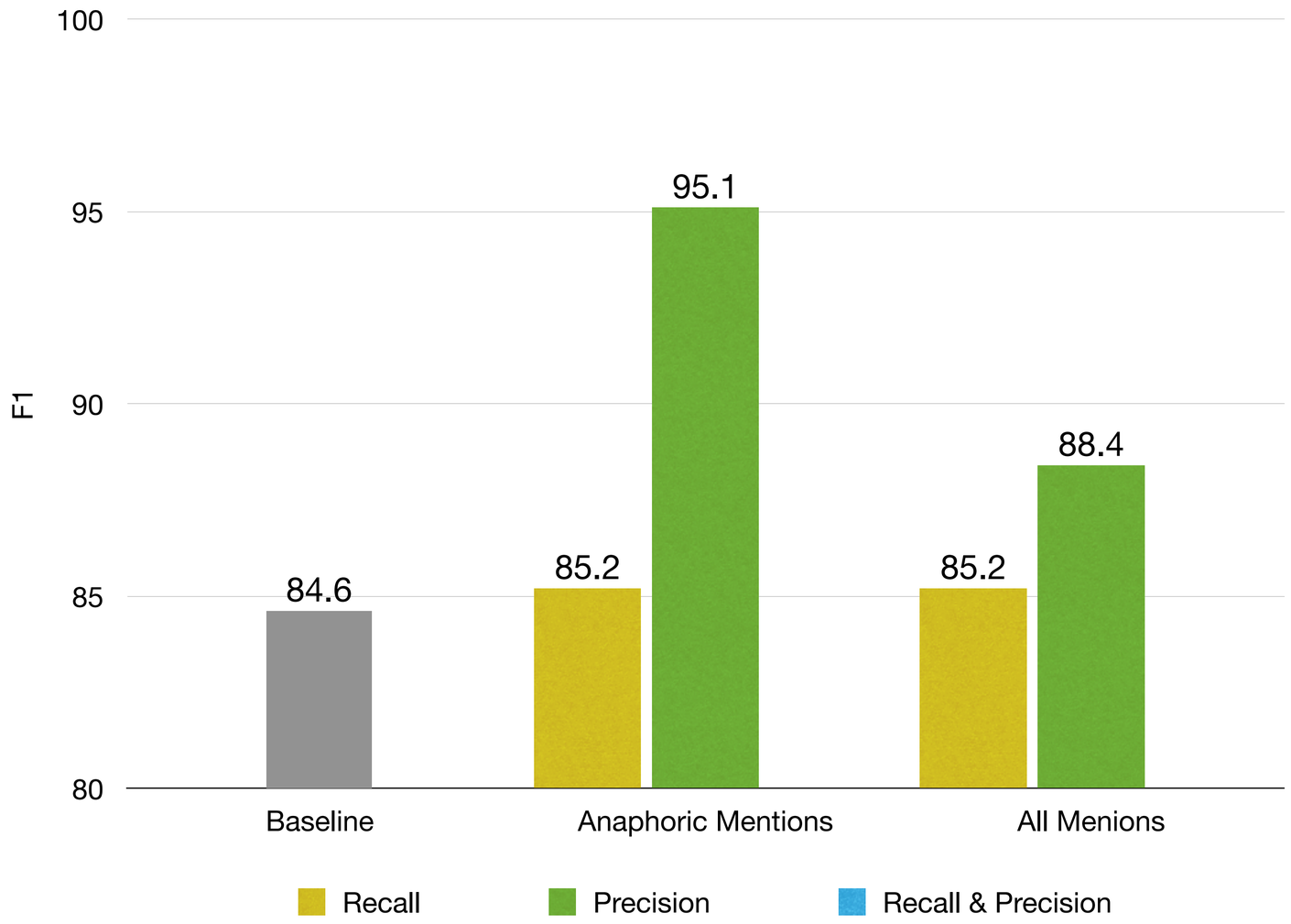
Importance of Precision



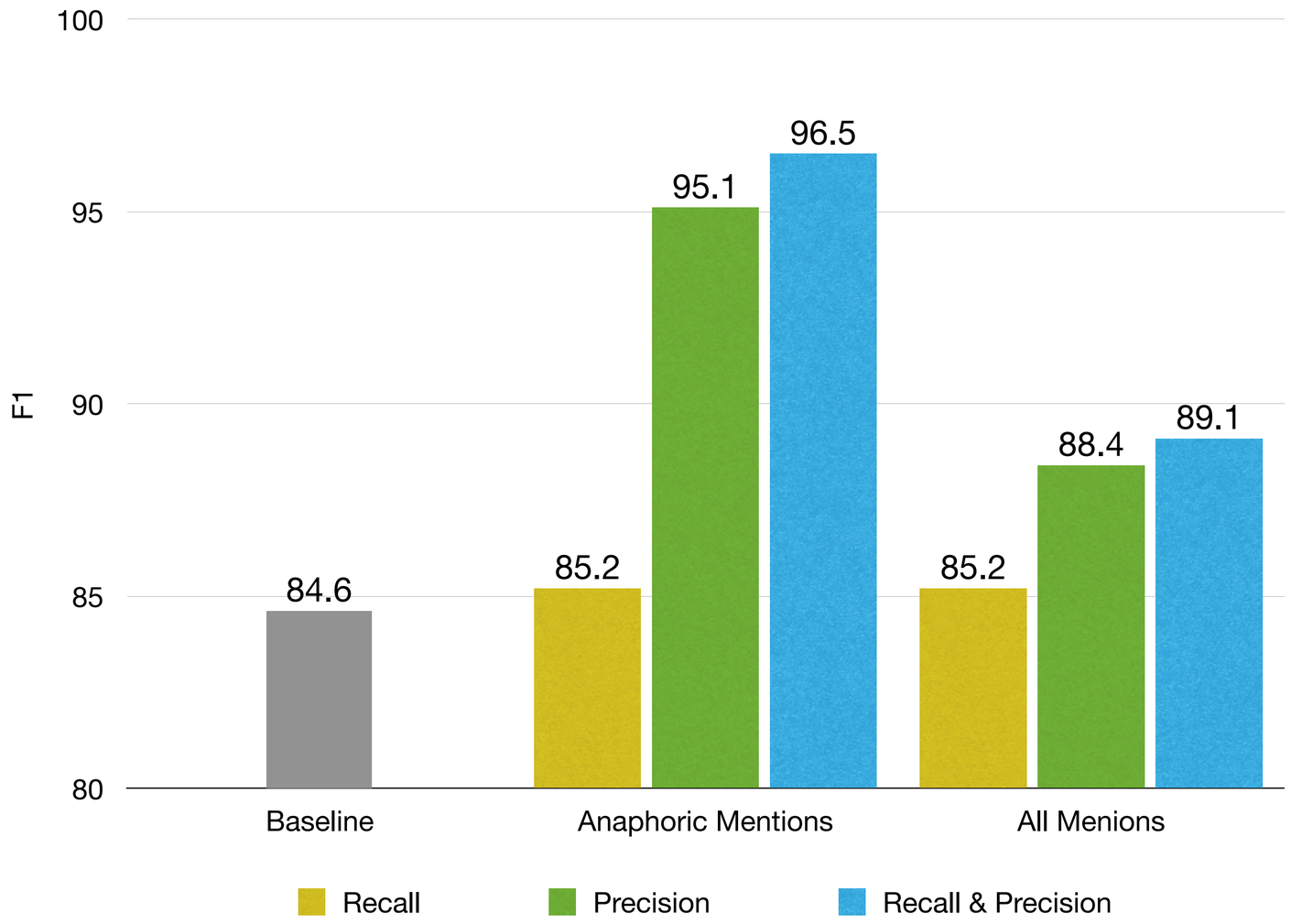
Importance of Precision



Importance of Precision

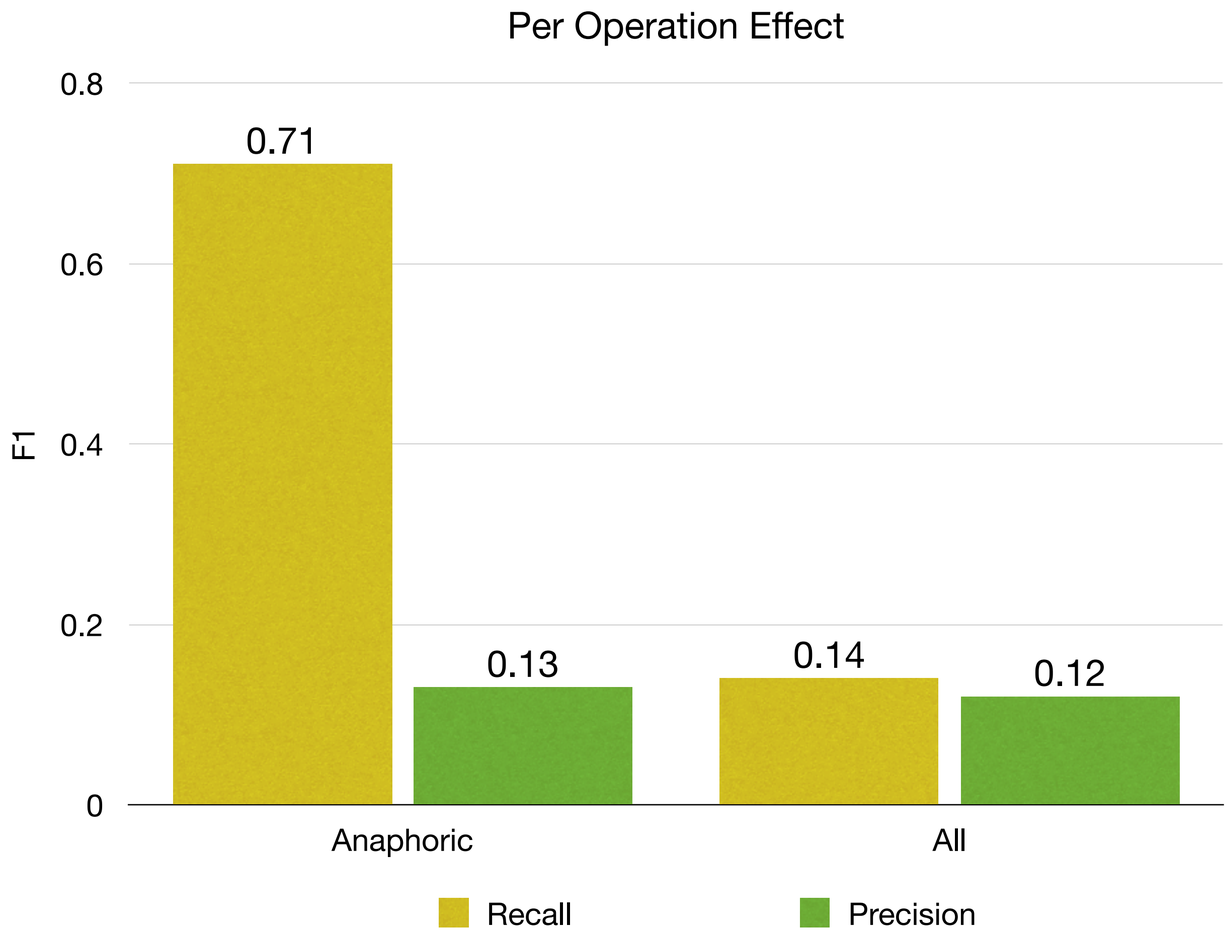


Importance of Precision

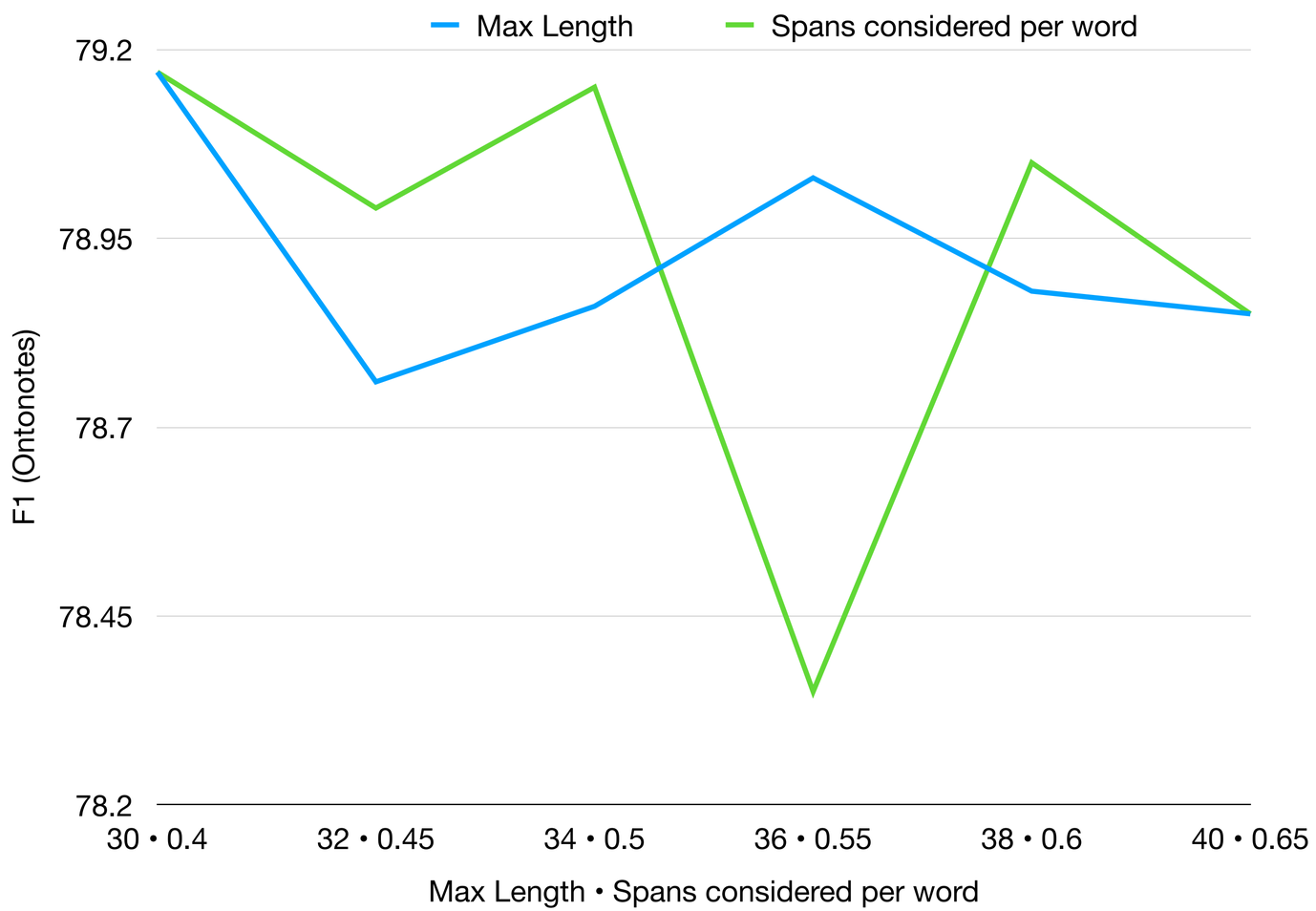


Importance of Recall

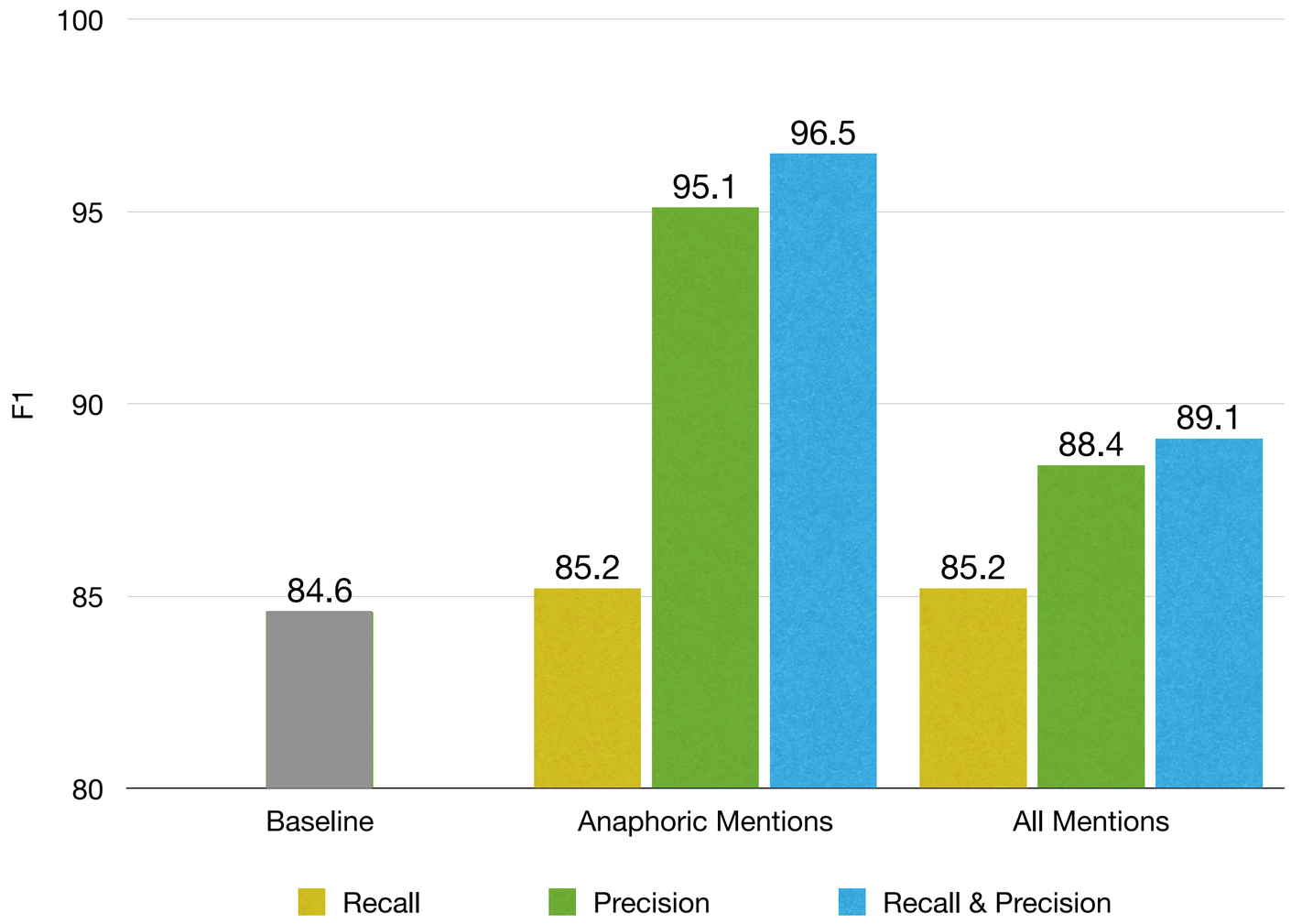
Importance of Recall



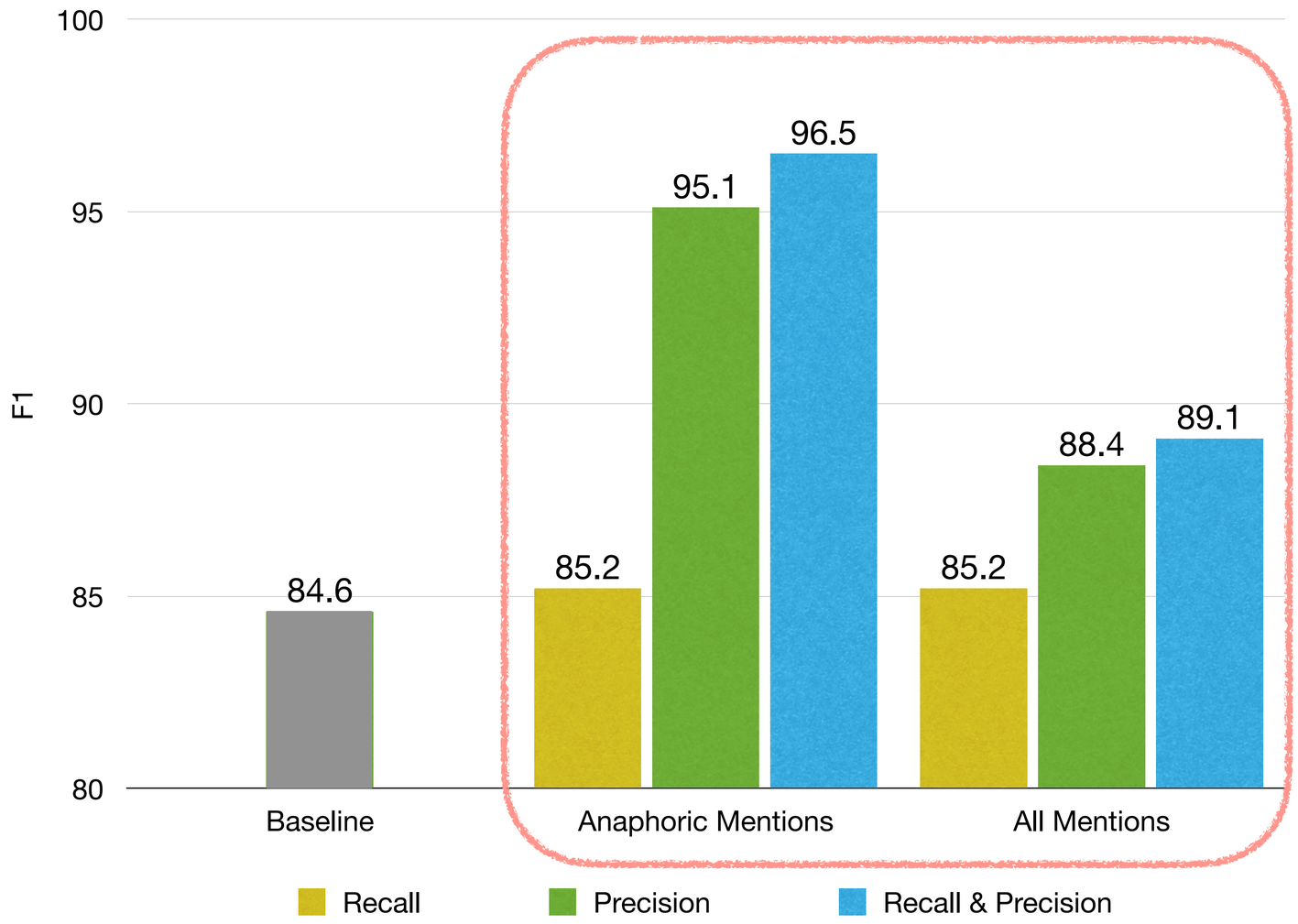
Precision-Recall Trade-Off



Previously …


100
95
90
85
80
F1
84.6
85.2
95.1
96.5
85.2
88.4
89.1
Baseline
Anaphoric Mentions
All Mentions
Recall
Precision
Recall & Precision

Previously …


100
95
90
85
80
F1
84.6
85.2
95.1
96.5
85.2
88.4
89.1
Baseline
Anaphoric Mentions
All Mentions
Recall
Precision
Recall & Precision

How Well Can the Detector Make Anaphoricity Decisions?

Anaphoricity Decisions

Anaphoricity Decisions

- Detector-like span classifiers that recognize:

Anaphoricity Decisions

- Detector-like span classifiers that recognize:
  - All mentions: 79.9 classification F1

Anaphoricity Decisions

- Detector-like span classifiers that recognize:
  - All mentions: 79.9 classification F1
  - Anaphoric mentions: 54.3 classification F1

Anaphoricity Decisions

- Detector-like span classifiers that recognize:
  - All mentions: 79.9 classification F1
  - Anaphoric mentions: 54.3 classification F1
- Confusion Index (CI): (singleton recall) / (anaphoric mention recall)

Anaphoricity Decisions

- Detector-like span classifiers that recognize:
  - All mentions: 79.9 classification F1
  - Anaphoric mentions: 54.3 classification F1
- Confusion Index (CI): (singleton recall) / (anaphoric mention recall)
  - A perfect detector has 0 CI; a random one has 1 CI

Anaphoricity Decisions

- Detector-like span classifiers that recognize:
  - All mentions: 79.9 classification F1
  - Anaphoric mentions: 54.3 classification F1
- Confusion Index (CI): (singleton recall) / (anaphoric mention recall)
  - A perfect detector has 0 CI; a random one has 1 CI
  - has 0.81 CI

Anaphoricity Decisions

- Detector-like span classifiers that recognize:
  - All mentions: 79.9 classification F1
  - Anaphoric mentions: 54.3 classification F1
- Confusion Index (CI): (singleton recall) / (anaphoric mention recall)
  - A perfect detector has 0 CI; a random one has 1 CI
  - has 0.81 CI
    - Degrades to 0.997 CI with text appearing as both mention types

What About the Linker?

Linker Errors

	Sentence A	Sentence B
Pronoun (109)	Tom lives in **Seattle**.	**It** was successfully launched.
Exact Match (6)	**Disney** is a global brand.	The subway to **Disney** has been constructed.
Head Match (11)	**Those landmark buildings** are tall.	He has not seen **these small buildings**.
Other Match (7)	**Dr. Henry** notices something else.	**Dr. Mann** is a successful researcher.
Semantic Proximity (12)	**Hong Kong cinema** has nurtured many directors.	It memorializes **Hong Kong's film history**.
Others (5)	**Paul Kelly** and **Steve Sodbury** have no idea.	

Summary

Summary

- Detector precision should not be overlooked

Summary

- Detector precision should not be overlooked
- Anaphoricity decisions are important, though they are difficult to make

Summary

- Detector precision should not be overlooked
- Anaphoricity decisions are important, though they are difficult to make
- Most linker errors require enhanced contextual reasoning

Summary

- Detector precision should not be overlooked
- Anaphoricity decisions are important, though they are difficult to make
- Most linker errors require enhanced contextual reasoning

arxiv.org/abs/2009.09363

Experimental Config