Adapting Coreference Resolution Models through Active Learning

ACL 2022

Michelle Yuan[1] Patrick Xia[2] Chandler May[2]
Benjamin Van Durme[2] Jordan Boyd-Graber[1]

[1]Department of Computer Science
University of Maryland

[2]Department of Computer Science
Johns Hopkins University

Coreference Resolution (CR)

The task of discovering spans of text that refer to the same entity

Traders said **municipals** were underpinned by influences, including the climb in Treasury issue prices. Also, **municipal bonds** lured buying because the stock market remains wobbly, traders contended. Mainly though, it was a favorable outlook for yesterday's new supply that propped up **municipals** , some traders said. Among the new issues was Massachusetts's $230 million of **general obligation bonds** . **The bonds** were won by a Goldman Sachs & Co. group with a true interest cost of 7.17%.

Cluster: { **municipals** , **municipal bonds** , **municipals** , **general obligation bonds** , **The bonds** }

Coreference Resolution (CR)

The task of discovering spans of text that refer to the same entity

Traders said municipals were underpinned by influences, including the climb in Treasury issue prices. Also, municipal bonds lured buying because the stock market remains wobbly, traders contended. Mainly though, it was a favorable outlook for yesterday's new supply that propped up municipals , some traders said. Among the new issues was Massachusetts's $230 million of general obligation bonds . The bonds were won by a Goldman Sachs & Co. group with a true interest cost of 7.17%.

Cluster: { municipals , municipal bonds , municipals , general obligation bonds , The bonds }

Neural, end-to-end models (Lee et al., 2018; Joshi et al., 2020) are SOTA for OntoNotes 5.0

Problem: Adapting CR Models

Models trained on OntoNotes may not immediately generalize to new domains

Traders said municipals were underpinned by influences, including the climb in Treasury issue prices. Also, municipal bonds lured buying because the stock market remains wobbly, traders contended. Mainly though, it was a favorable outlook for yesterday's new supply that propped up municipals , some traders said. Among the new issues was Massachusetts's $230 million of general obligation bonds . The bonds were won by a Goldman Sachs & Co. group with a true interest cost of 7.17%.

Cluster: { municipals , municipal bonds , municipals , general obligation bonds , The bonds }

A molecule is a group of atoms held together by chemical bonds . Imagine you and your friends standing in a circle holding hands. Each person stands for one atom, your hands represent the bonds , and the entire circle represents a molecule.

Problem: Adapting CR Models

Models trained on OntoNotes may not immediately generalize to new domains

Traders said municipals were underpinned by influences, including the climb in Treasury issue prices. Also, municipal bonds lured buying because the stock market remains wobbly, traders contended. Mainly though, it was a favorable outlook for yesterday's new supply that propped up municipals , some traders said. Among the new issues was Massachusetts's $230 million of general obligation bonds . The bonds were won by a Goldman Sachs & Co. group with a true interest cost of 7.17%.

Cluster: { municipals , municipal bonds , municipals , general obligation bonds , The bonds }

A molecule is a group of atoms held together by chemical bonds . Imagine you and your friends standing in a circle holding hands. Each person stands for one atom, your hands represent the bonds , and the entire circle represents a molecule.

Impedes immediate application for scenarios like distinguishing entities in scientific articles

Problem: Adapting CR Models

- (Xia and Van Durme, 2021) show the benefits of *continued training* where a model trained on OntoNotes is further trained on the target dataset

Problem: Adapting CR Models

- (Xia and Van Durme, 2021) show the benefits of *continued training* where a model trained on OntoNotes is further trained on the target dataset

- However, they assume labeled data already exist in the target domain

Problem: Adapting CR Models

- (Xia and Van Durme, 2021) show the benefits of *continued training* where a model trained on OntoNotes is further trained on the target dataset

- However, they assume labeled data already exist in the target domain

- How can we adapt CR models without requiring large amounts of newly annotated data?

Method: Active Learning

- Use active learning to find particular spans of text for users to label

Method: Active Learning

- Use active learning to find particular spans of text for users to label

- The goal is to adapt the model to the target domain by continue training it on spans labeled from active learning

What should we label?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician William Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

What should we label?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician William Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

Uncertainty in mention detection

What should we label?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician William Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

Uncertainty in mention clustering

What should we label?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician William Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

Uncertainty in mention clustering conditioned on mention detection

What should we label?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician William Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

Uncertainty in both mention detection and mention clustering

Experiments: Strategies

1. **ment-ent:** Mention detection entropy
2. **clust-ent:** Mention clustering entropy
3. **cond-ent:** Conditional entropy
4. **joint-ent** Joint entropy

Experiments: Strategies

1. **ment-ent:** Mention detection entropy
2. **clust-ent:** Mention clustering entropy
3. **cond-ent:** Conditional entropy
4. **joint-ent** Joint entropy
5. **random:** Randomly sample from all spans in the document

Experiments: Strategies

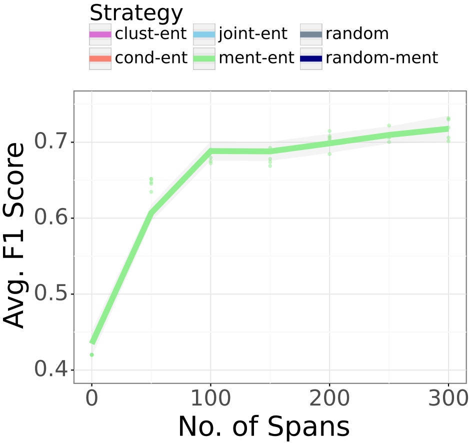
1. **ment-ent:** Mention detection entropy
2. **clust-ent:** Mention clustering entropy
3. **cond-ent:** Conditional entropy
4. **joint-ent** Joint entropy
5. **random:** Randomly sample from all spans in the document
6. **random-ment:** Randomly sample only from the pool of spans that are likely entity mentions

Experiments: Datasets

1. **OntoNotes 5.0 (source):**
   - Most common CR dataset (Pradhan et al., 2013)
   - News articles and telephone conversations
   - Only non-singletons are annotated

2. **PreCo (target):**
   - Large corpus of grade-school reading comprehension texts (Chen et al., 2018)
   - Singletons are annotated

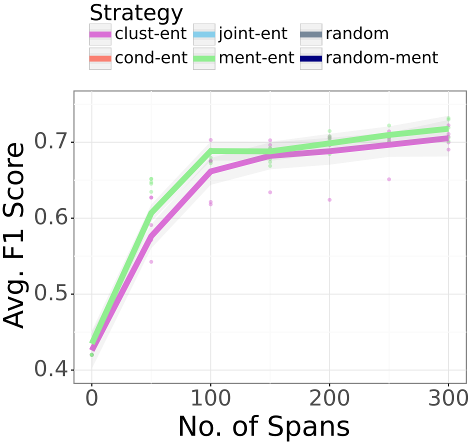
Comparing Strategies



Mention Detection Entropy

- Test Avg. F1 on PreCo
- For each cycle, we simulate labeling fifty spans from one document

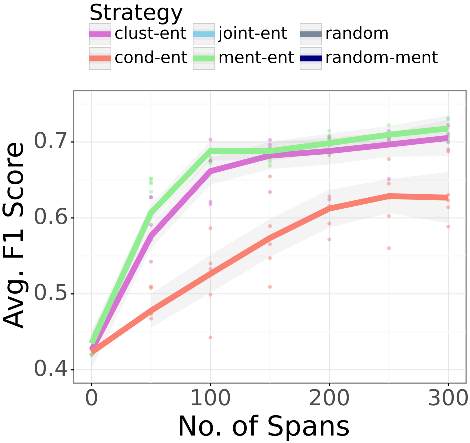
Comparing Strategies



Mention Clustering Entropy

- Test Avg. F1 on PreCo
- For each cycle, we simulate labeling fifty spans from one document

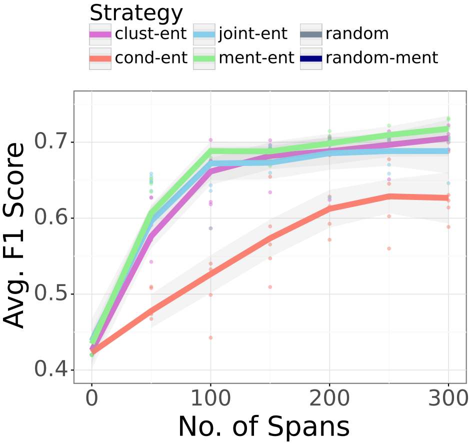
Comparing Strategies



Conditional Entropy

- Test Avg. F1 on PreCo
- For each cycle, we simulate labeling fifty spans from one document

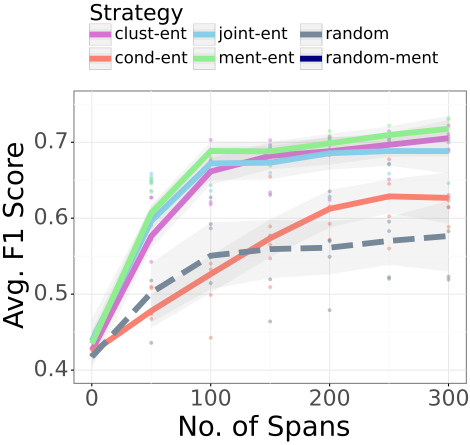
Comparing Strategies



Joint Entropy

- Test Avg. F1 on PreCo
- For each cycle, we simulate labeling fifty spans from one document

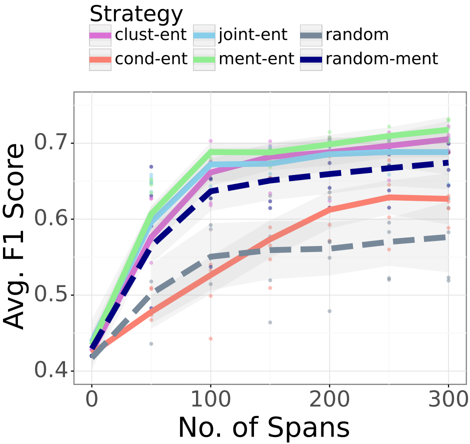
Comparing Strategies



Random

- Test Avg. F1 on PreCo
- For each cycle, we simulate labeling fifty spans from one document

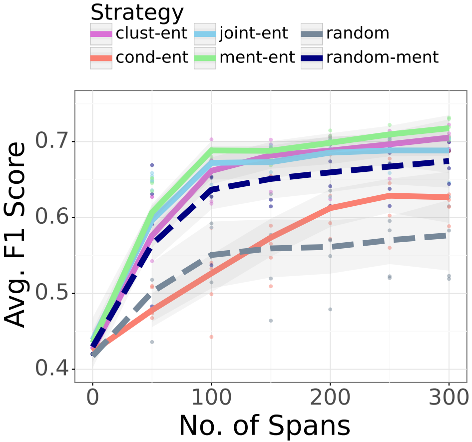
Comparing Strategies



Random Entity Mentions

- Test Avg. F1 on PreCo
- For each cycle, we simulate labeling fifty spans from one document

Ment-ent, Clust-ent, Joint-ent are Effective but Random Performs Worst!



Should we label spans within or across documents?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician Willam Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

Ha'penny Bridge might look like it's just another bridge. But if you read about the history, you will know how significant this bridge is. Apparently half a penny was the toll that had to be paid from 1816 until the year 1919 in order to cross the Liffey Bridge.

Lovely park. Easy to get to on public transport. I recommend getting the bikes for hire when you get there, it made getting around really easy and you can cut across the fields to go and see the deer more easily!

Should we label spans within or across documents?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician Willam Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

Ha'penny Bridge might look like it's just another bridge. But if you read about the history, you will know how significant this bridge is. Apparently half a penny was the toll that had to be paid from 1816 until the year 1919 in order to cross the Liffey Bridge.

Lovely park. Easy to get to on public transport. I recommend getting the bikes for hire when you get there, it made getting around really easy and you can cut across the fields to go and see the deer more easily!

Within

Should we label spans within or across documents?

A fantastic experience, very informative, very time consuming but enjoyable. So much information to take in about Guinness that you would've never known. For example, the brewery hired the statistician Willam Gosset in 1899. The "student" was known for developing the Student's t-test, a well-known method in statistical inference.

Ha'penny Bridge might look like it's just another bridge. But if you read about the history, you will know how significant this bridge is. Apparently half a penny was the toll that had to be paid from 1816 until the year 1919 in order to cross the Liffey Bridge.

Lovely park. Easy to get to on public transport. I recommend getting the bikes for hire when you get there, it made getting around really easy and you can cut across the fields to go and see the deer more easily!

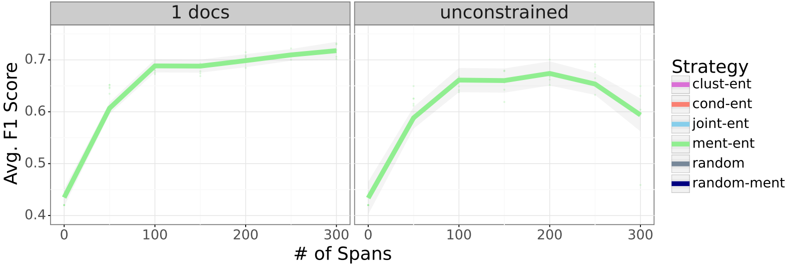
Across

Varying Number of Documents Read

- Test Avg. F1 on PreCo of each strategy
- On each cycle, we sample fifty spans, which each can belong to any document

Varying Number of Documents Read

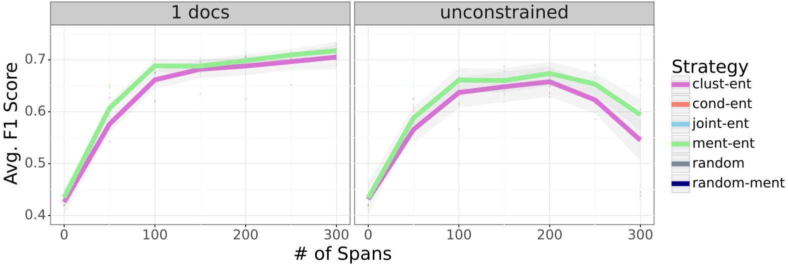
- Test Avg. F1 on PreCo of each strategy
- On each cycle, we sample fifty spans, which each can belong to any document



Mention Detection Entropy

Varying Number of Documents Read

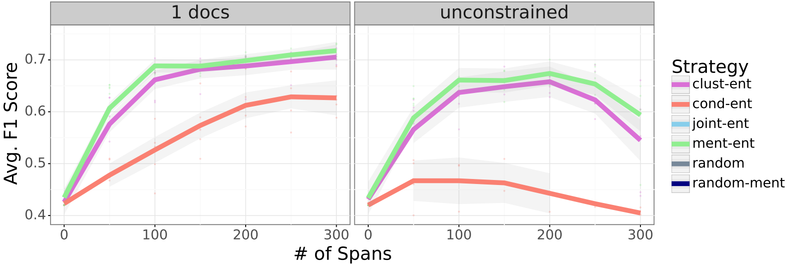
- Test Avg. F1 on PreCo of each strategy
- On each cycle, we sample fifty spans, which each can belong to any document



Mention Clustering Entropy

Varying Number of Documents Read

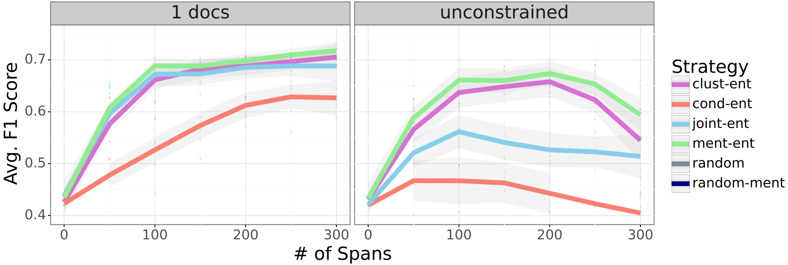
- Test Avg. F1 on PreCo of each strategy
- On each cycle, we sample fifty spans, which each can belong to any document



Conditional Entropy

Varying Number of Documents Read

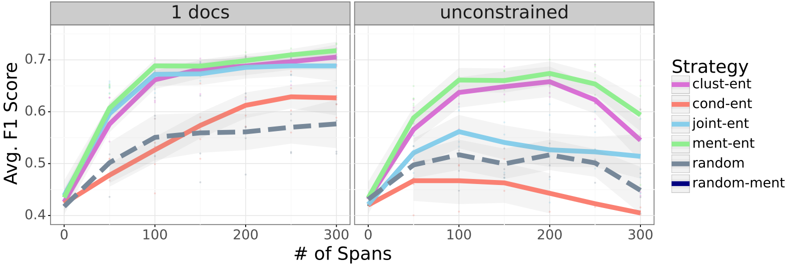
- Test Avg. F1 on PreCo of each strategy
- On each cycle, we sample fifty spans, which each can belong to any document



Joint Entropy

Varying Number of Documents Read

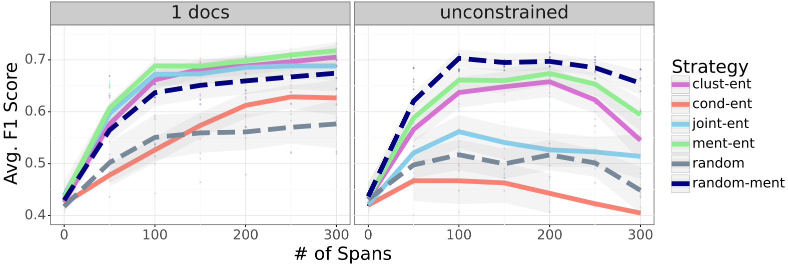
- Test Avg. F1 on PreCo of each strategy
- On each cycle, we sample fifty spans, which each can belong to any document



Random

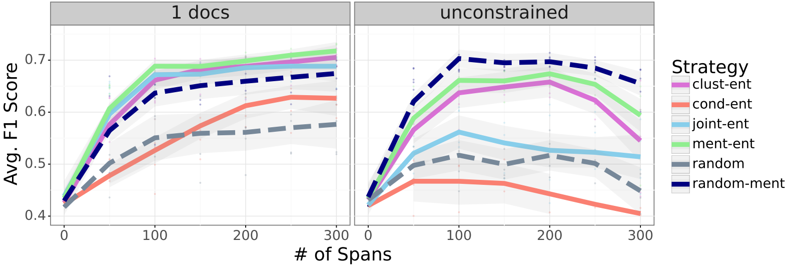
Varying Number of Documents Read

- Test Avg. F1 on PreCo of each strategy
- On each cycle, we sample fifty spans, which each can belong to any document



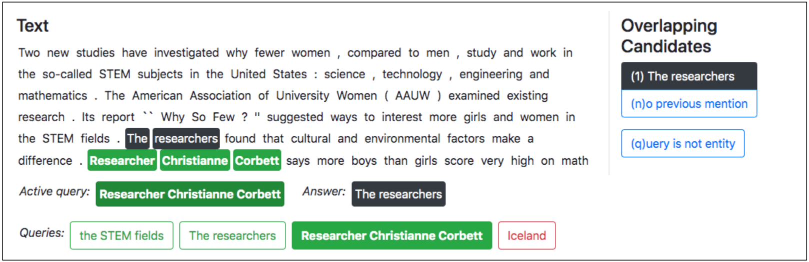
Random Entity Mentions

Model training is Unstable in Unconstrained Sampling



Humans Read and Label Coreference

Three users label spans sampled from PreCo

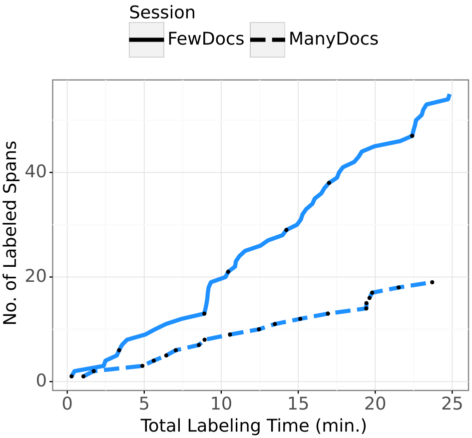


Humans Read and Label Coreference

Users complete two twenty-five minute sessions:

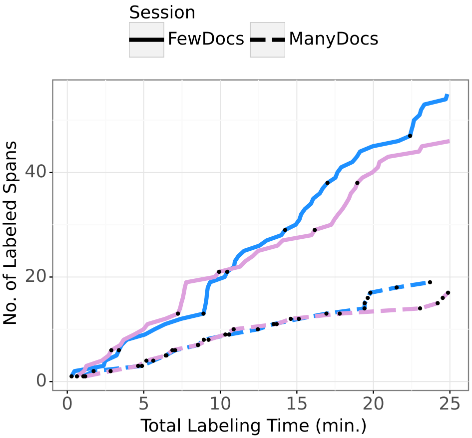
1. **FewDocs:** Read fewer documents and label multiple spans per document
2. **ManyDocs:** Read more documents and label one span per document

Labeling Throughput At Least Doubles in FewDocs



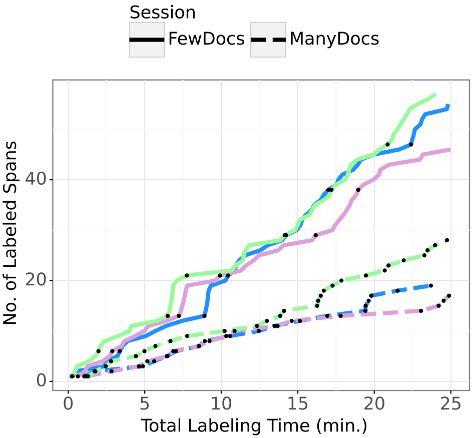
- Each color indicates one of three users and the linetype designates the session
- Black dots mark the first span labeled in a different document

Labeling Throughput At Least Doubles in FewDocs



- Each color indicates one of three users and the linetype designates the session
- Black dots mark the first span labeled in a different document

Labeling Throughput At Least Doubles in FewDocs



- Each color indicates one of three users and the linetype designates the session
- Black dots mark the first span labeled in a different document

Summary

1. To reduce annotation for transferring coreference resolution models, we use active learning to choose text spans for users to label

Summary

1. To reduce annotation for transferring coreference resolution models, we use active learning to choose text spans for users to label
2. We explore various aspects of active learning for coreference resolution, like sources of model uncertainty and the trade-off between reading and labeling

Summary

1. To reduce annotation for transferring coreference resolution models, we use active learning to choose text spans for users to label
2. We explore various aspects of active learning for coreference resolution, like sources of model uncertainty and the trade-off between reading and labeling
3. Surprisingly, sampling by mention detection entropy is more useful for domains like PreCo

Summary

1. To reduce annotation for transferring coreference resolution models, we use active learning to choose text spans for users to label
2. We explore various aspects of active learning for coreference resolution, like sources of model uncertainty and the trade-off between reading and labeling
3. Surprisingly, sampling by mention detection entropy is more useful for domains like PreCo
4. In both simulations and the user study, there are more benefits from sampling spans from the same document rather than different contexts

Thanks

Any Questions?

myuan@cs.umd.edu

DEPARTMENT OF COMPUTER SCIENCE

Slides customized from UFGTeX

References I

Xilun Chen, Yu Sun, Ben Athiwaratkun, Claire Cardie, and Kilian Weinberger. Adversarial deep averaging networks for cross-lingual sentiment classification. 6:557–570, 2018. doi: 10.1162/tacl_a_00039.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. SpanBERT: Improving pre-training by representing and predicting spans. 8:64–77, 2020. doi: 10.1162/tacl_a_00300.

Kenton Lee, Luheng He, and Luke Zettlemoyer. Higher-order coreference resolution with coarse-to-fine inference. 2018. doi: 10.18653/v1/N18-2108.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. Towards robust linguistic analysis using OntoNotes. 2013. URL `https://www.aclweb.org/anthology/W13-3516`.

Patrick Xia and Benjamin Van Durme. Moving on from OntoNotes: Coreference resolution model transfer. 2021. doi: 10.18653/v1/2021.emnlp-main.425.