ezCoref

Towards Unifying Annotation Guidelines for Coreference Resolution

Ankita Gupta Marzena Karpinska Wenlong Zhao Kalpesh Krishna
Jack Merullo Luke Yeh Mohit Iyyer Brendan O'Connor

UMassAmherst

Manning College of Information
& Computer Sciences

Coreference Resolution
Background

Determine which spans of text refer to the same entity

[John] doesn't like [Fred], but [he] still invited [him] to [the party].

Coreference Resolution
Background

Determine which spans of text refer to the same entity

[John] doesn't like [Fred], but [he] still invited [him] to [the party].

Antecedent ← Mention

Coreference Resolution
Background

Several datasets have been collected to facilitate modeling of coreferences.

OntoNotes, ARRAU

LitBank, GUM, Phrase Detectives

and many more…

Coreference Resolution

Motivation

However, application to other domains and languages requires new dataset curation.

Coreference Resolution
Motivation

To collect new datasets we need

1. Annotation guidelines

Coreference Resolution
Motivation

To collect new datasets we need

1. Annotation guidelines

2. Annotation Tool

Coreference Resolution
Motivation

To collect new datasets we need

1. Annotation guidelines

2. Annotation Tool

3. Annotation Workforce

Coreference Resolution
Motivation

To collect new datasets we need

1. Annotation guidelines

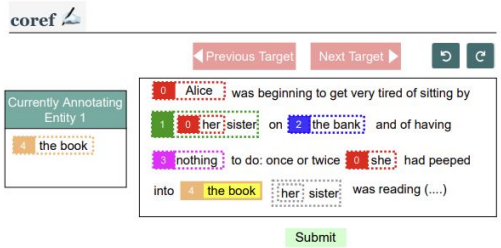
2. Annotation Tool

3. Annotation Workforce

Existing datasets differ widely in these aspects
making **new annotation efforts challenging**

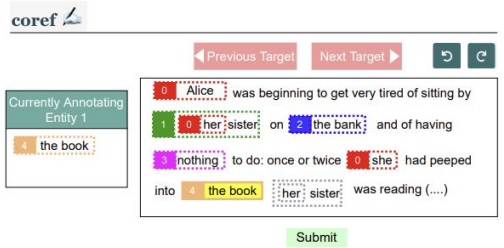
Key Contributions

1. **ezCoref annotation tool**
   a. Open-sourced
   b. Integration with crowdsourcing platforms
   c. Interactive Tutorial

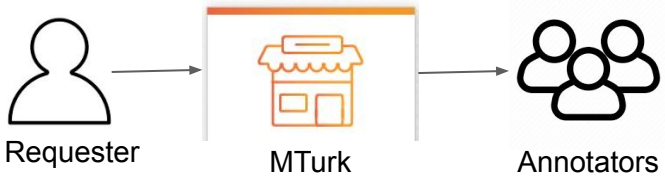


Key Contributions

1. ezCoref annotation tool
   a. Open-sourced
   b. Integration with crowdsourcing platforms
   c. Interactive Tutorial

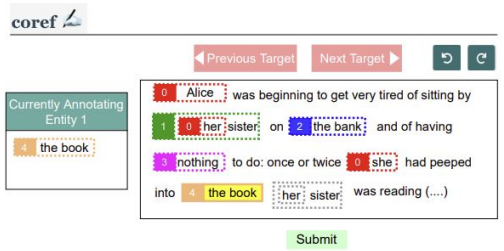


2. **Re-annotation study**
   a. seven coreference datasets
   b. seven domains

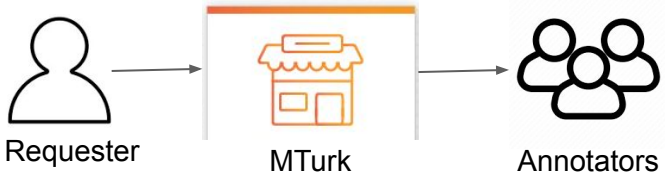


Key Contributions

1. ezCoref annotation tool
   a. Open-sourced
   b. Integration with crowdsourcing platforms
   c. Interactive Tutorial



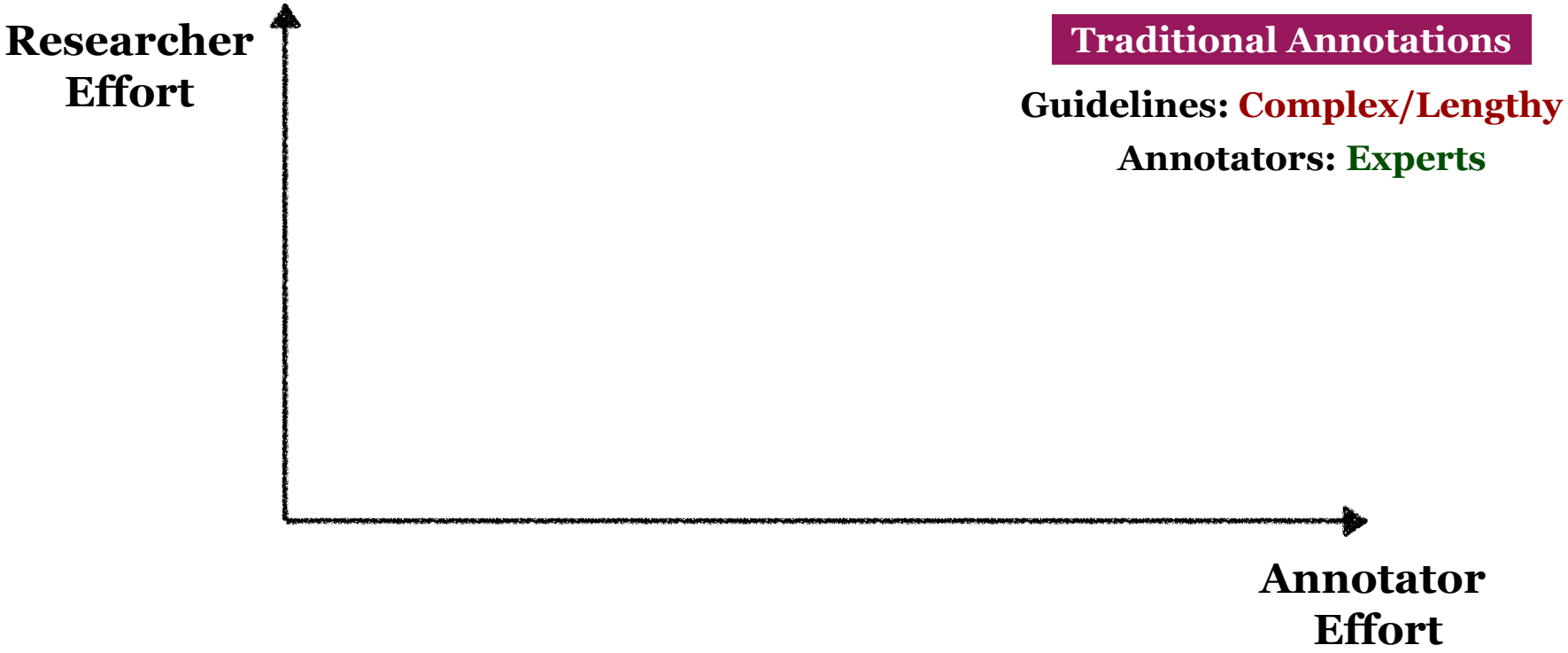
2. Re-annotation study
   a. seven coreference datasets
   b. seven domains



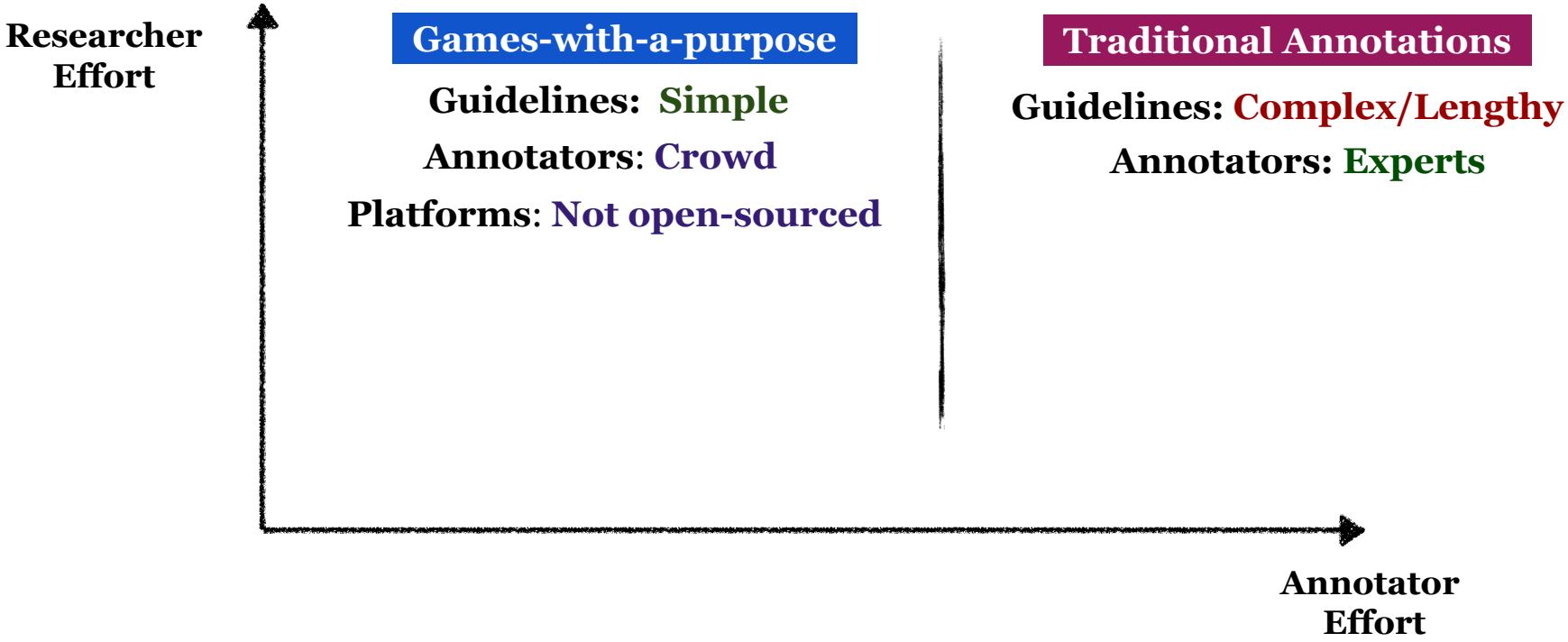
3. **Annotation analysis**
   a. agreements/disagreements among crowd annotators
   b. comparison with original annotations

1. ezCoref Annotation Tool

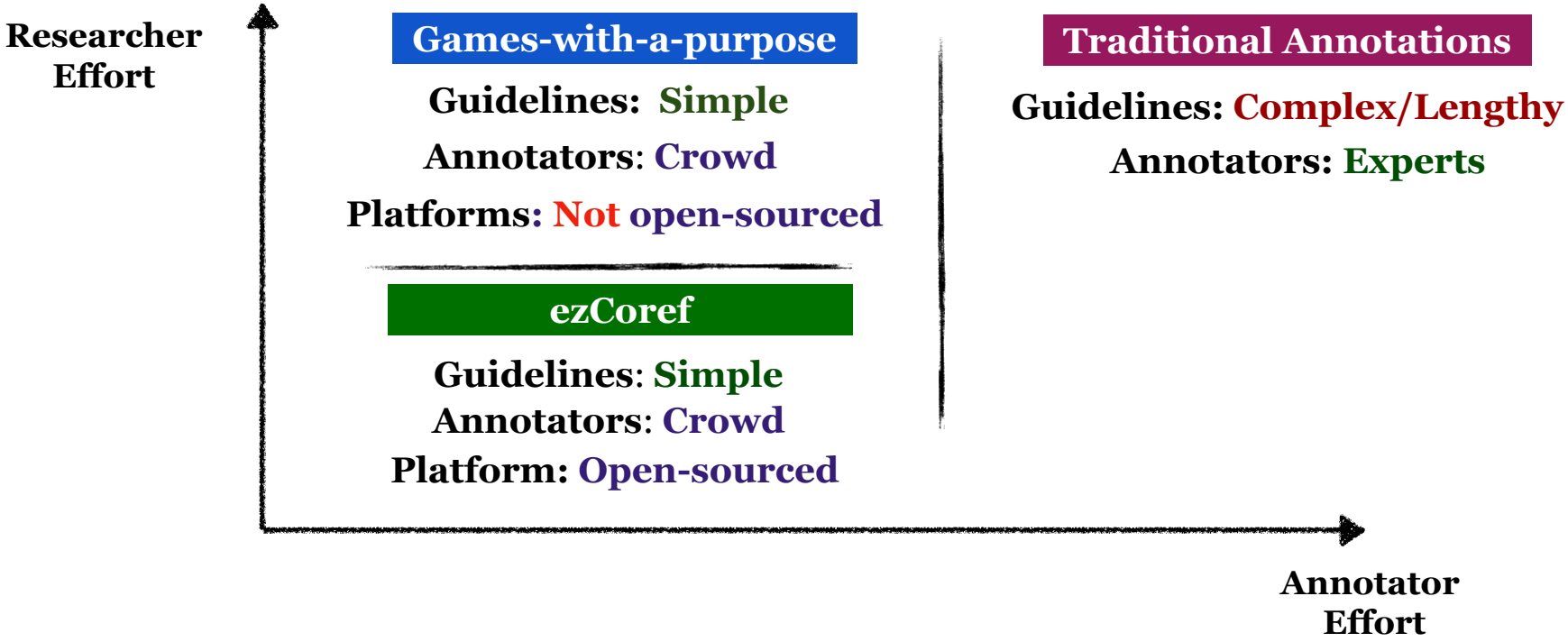
Annotation Approaches



Annotation Approaches



Annotation Approaches



ezCoref

Annotation Tool and Tutorial

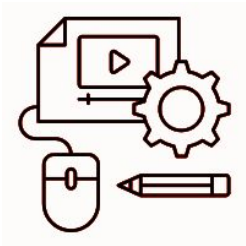
Open Sourced

Integration with crowdsourcing platforms (e.g., AMT, crowdflower)

Crowdsourcing Friendly UI

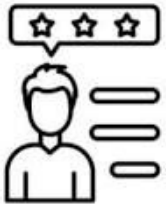
ezCoref

Annotation Tool and Tutorial



Tutorial

- ✓ Interactive
- ✓ Customizable
- ✓ Validate by 70+ AMT annotators



AMT annotator

"Absolutely beautiful, intuitive, and helpful. Legitimately the best one I've ever seen in my 2 years on AMT! Awesome job"

ezCoref
Annotation Tool and Tutorial

2. Reannotation Study

Crowdsourcing Setup

OntoNotes	GUM	ARRAU	LitBank	QuizBowl	Phrase Detectives	PreCo
Hovy et al. 2006	Zeldes 2017	Uryupina et al. 2019	Bamman et al. 2020	Guha et al. 2015	Chamberlain et al. 2016	Chen et al. 2018
Linguistic Experts				**Domain Experts**	**Crowd**	

Crowdsourcing Setup

OntoNotes	GUM	ARRAU	LitBank	QuizBowl	Phrase Detectives	PreCo
Hovy et al. 2006	Zeldes 2017	Uryupina et al. 2019	Bamman et al. 2020	Guha et al. 2015	Chamberlain et al. 2016	Chen et al. 2018

Linguistic Experts | **Domain Experts** | **Crowd**

News

Biographies

Weblogs

Fiction

Quizzes

Opinions

Wikipedia

Crowdsourcing Setup

OntoNotes	GUM	ARRAU	LitBank	QuizBowl	Phrase Detectives	PreCo
Hovy et al. 2006	Zeldes 2017	Uryupina et al. 2019	Bamman et al. 2020	Guha et al. 2015	Chamberlain et al. 2016	Chen et al. 2018

Linguistic Experts | **Domain Experts** | **Crowd**

News

Biographies

Weblogs

Fiction

Quizzes

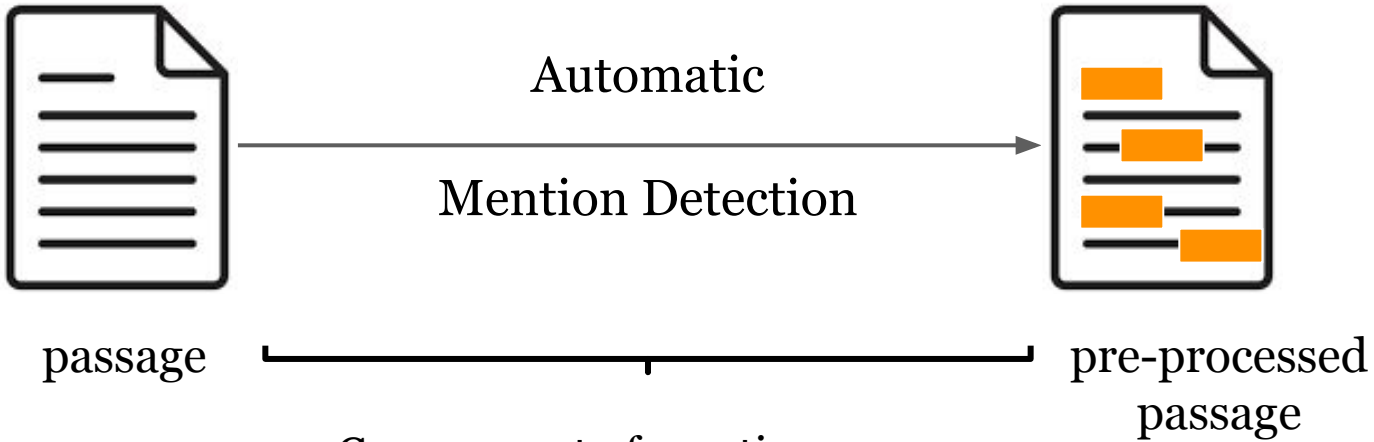
Opinions

Wikipedia

240 passages

Crowdsourcing Setup

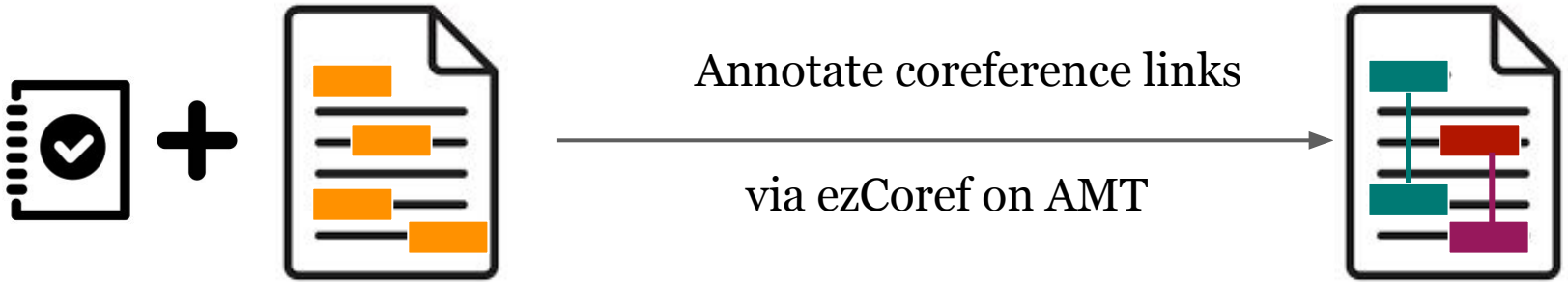
Preprocessing



- Common set of mentions
- High recall (>80%) across datasets

Crowdsourcing Setup

Annotating Coreferences



guidelines? pre-processed passage

annotated passage

Annotation Guidelines

Inconsistencies

> Maybe we need a CIA version of the Miranda warning: You have the right to conceal your coup intentions, because we may rat on you.

An example sentence drawn from Wall Street Journal

Annotation Guidelines

Inconsistencies

Maybe we need a CIA version of the Miranda warning: You have the right to conceal your coup intentions, because we may rat on you.

OntoNotes

Does not mark generic pronouns

Annotation Guidelines
Inconsistencies

Maybe we need a CIA version of the Miranda warning: You have the right to conceal your coup intentions, because we may rat on you.

OntoNotes

Does not mark generic pronouns

Maybe [we] need a CIA version of the Miranda warning: [You] have the right to conceal [your] coup intentions, because [we] may rat on [you].

ARRAU

Marks generic pronouns as *undef-reference* (not coreferent)

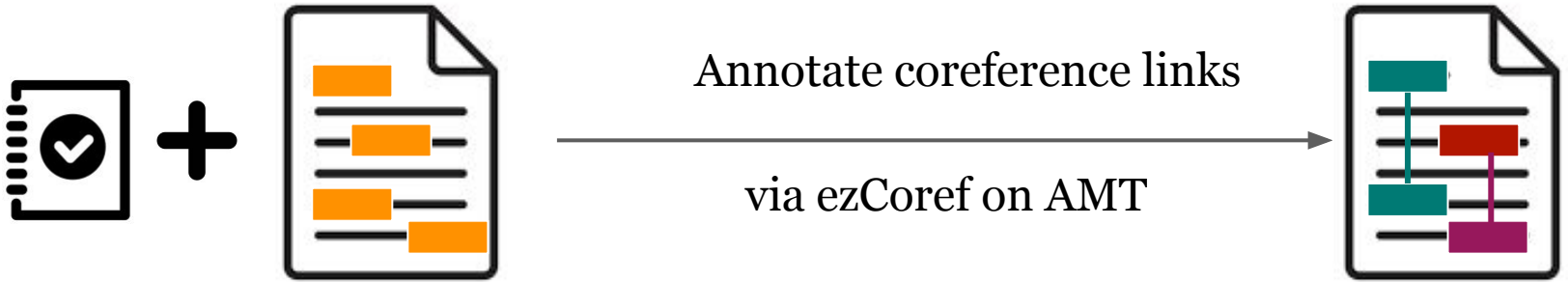
Annotation Guidelines

RQ: What **guidelines** to use when collecting new annotations?

coreference cases treated **uniformly** across all datasets as **simplified guidelines**

Crowdsourcing Setup

Annotating Coreferences



simplified guidelines

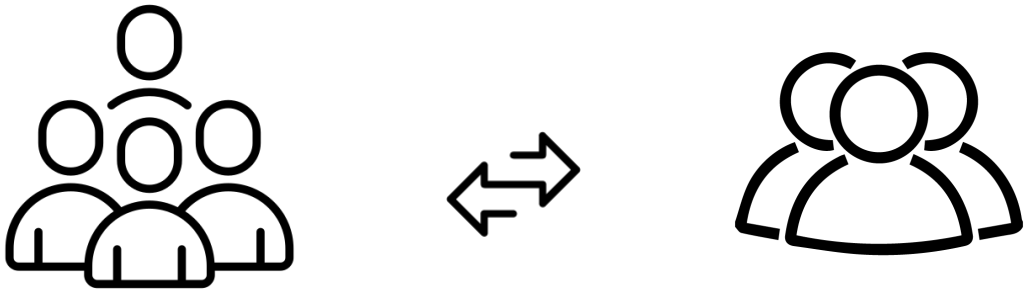
pre-processed passage

annotated passage

- **5 annotations**/passage
- **Majority voting**
- **12,200** mentions
- **42,108** tokens

3. Analysis of ezCoref annotations

Investigating Agreements

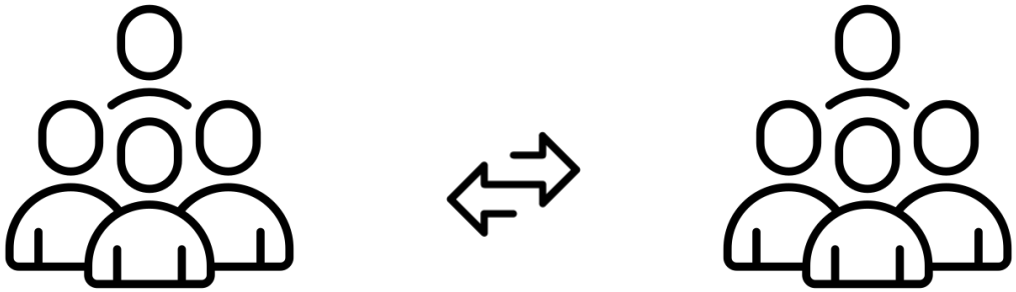


ezCoref Annotations

Original Annotations
(linguists, domain experts, crowd)

1. **Comparison** of ezCoref annotations with original annotations

Investigating Agreements

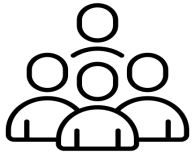


ezCoref Annotations

ezCoref Annotations

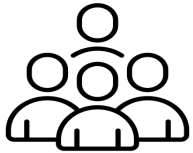
2. Agreements and disagreements among ezCoref annotators

Investigating Agreements

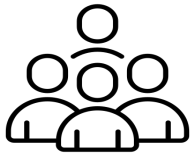


ezCoref Annotations

=



ezCoref Annotations



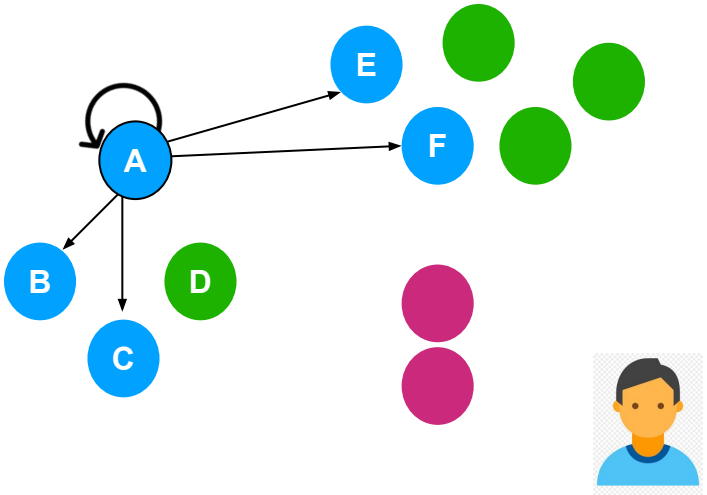
ezCoref Annotations

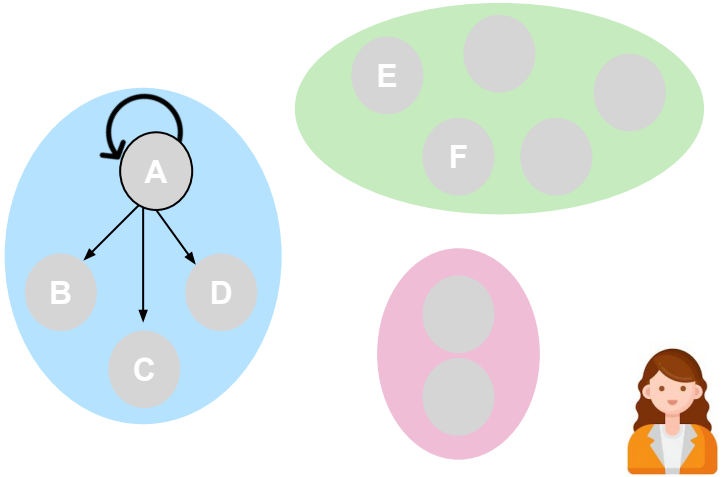
≠

Original Annotations

3. Deviations of ezCoref annotations from original annotations

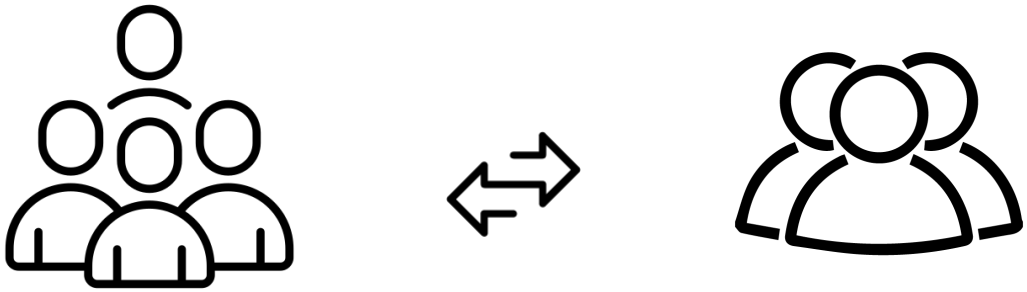
Measuring Agreements





Agreement: Fraction of mentions two annotators agree should be coreferent with a given mention, which is captured by **B3 precision and recall measure**

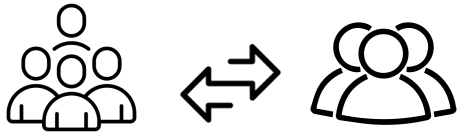
Investigating Agreements



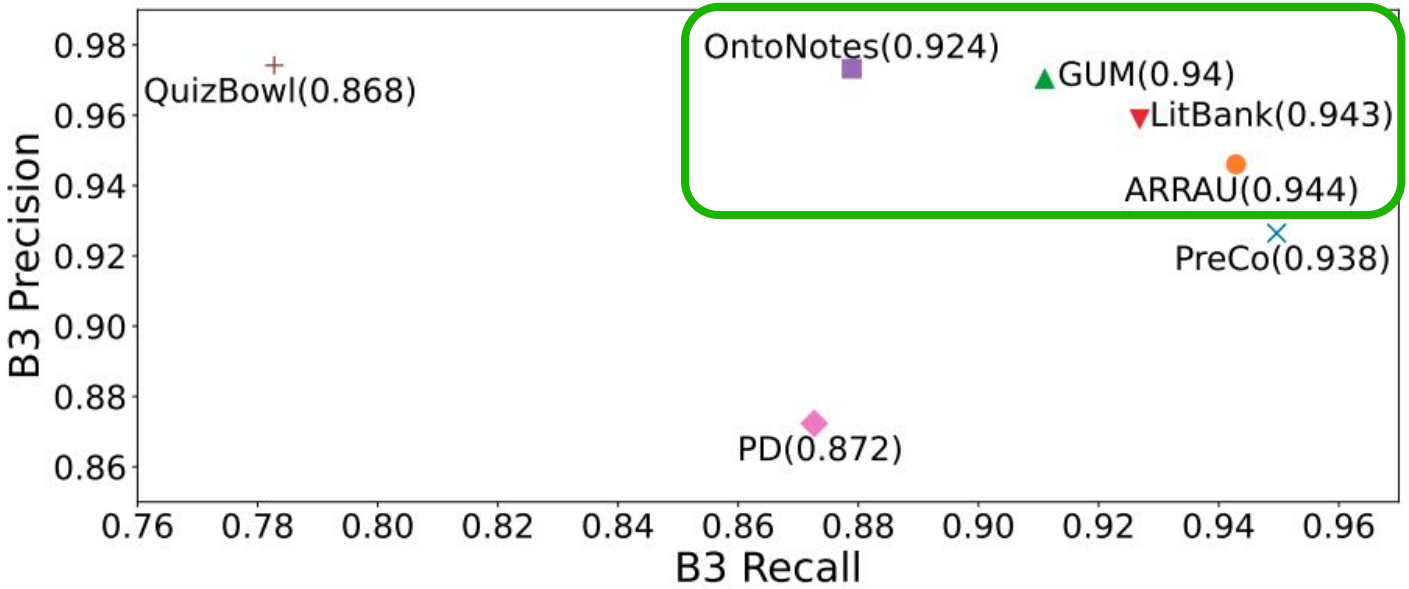
ezCoref Annotations **Original** Annotations

1. **Comparison** of ezCoref annotations with original annotations

Analysis

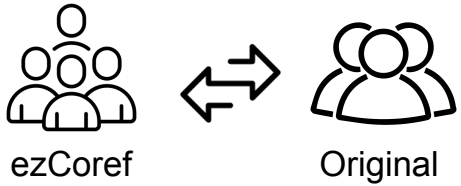


1. **Comparison** of ezCoref and original annotations



High agreement with expert annotated datasets

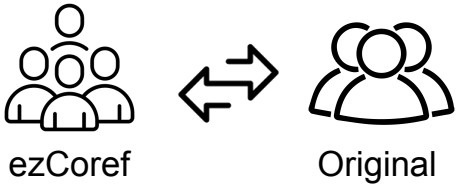
Analysis



High agreement with expert annotated datasets

- **Pronouns**
- **Named Entities**

Analysis



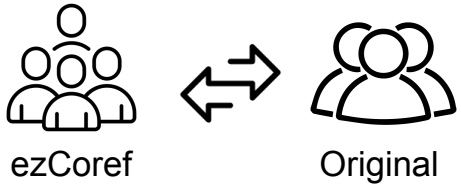
High agreement with expert annotated datasets

- Pronouns
- Named Entities
- **Appositives**

[Fuad Basya]$_{e1}$, [spokesman for the Indonesian military]$_{e1}$, said fisherman first noticed the people and a warship was deployed to retrieve them.

An example of **appositive construction** marked as **coreferent** by **ezCoref annotators, consistent** with the **GUM** guidelines

Analysis



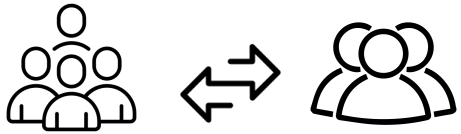
High agreement with expert annotated datasets

- Pronouns
- Named Entities
- Appositives
- **Generic mentions**

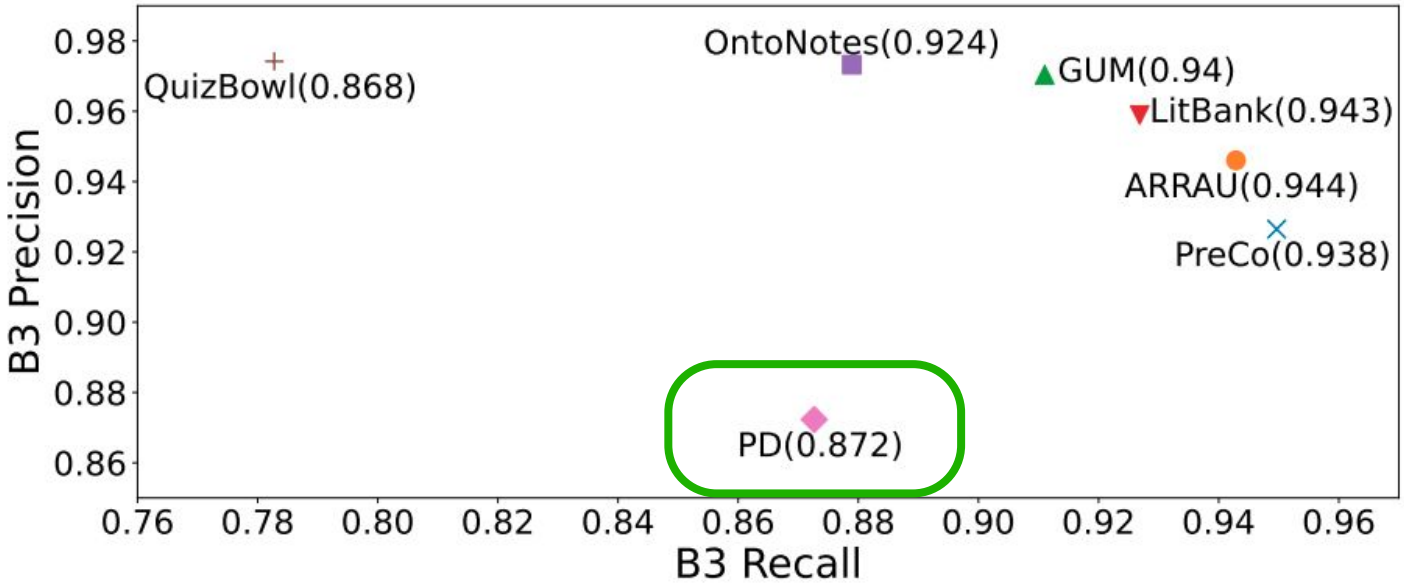
> Kidder is putting [brokers] through a 20 week training course (..) However, skeptics caution (..) [brokers] and investment bankers looks great on paper, but doesn't always happen.

An example of **generic mentions** annotated as **coreferent by ezCoref annotators, consistent** with the **ARRAU** guidelines

Analysis

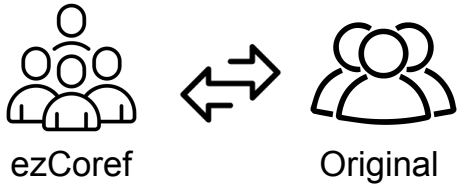


1. **Comparison** of ezCoref and original annotations



Low precision with Phrase Detectives suggesting ezCoref annotators merge original clusters

Analysis

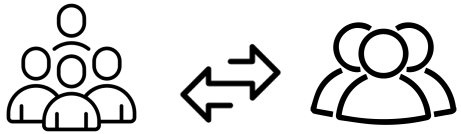


Low precision with Phrase Detectives suggesting ezCoref annotators merge original clusters

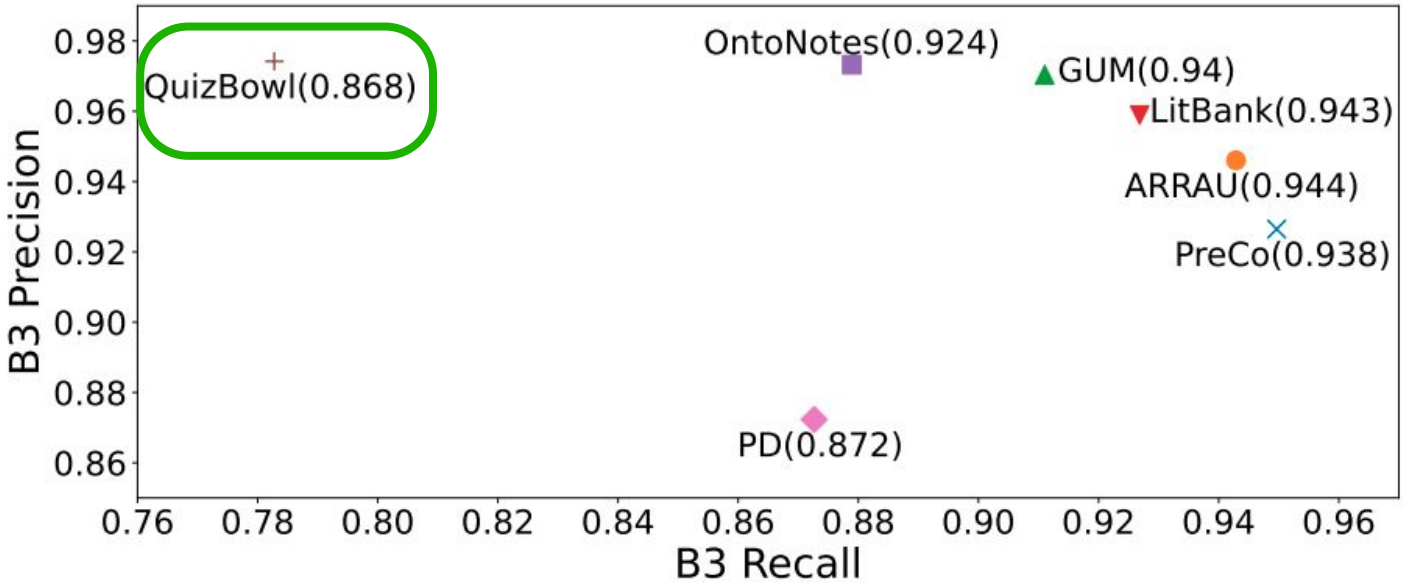
Not long after **[a suitor]** appeared (..) and the miller (..) betrothed his daughter to **[him]**. But the girl did not care for **[the man]**(...) she could not look at **[him]** nor think of **[him]**; without an inward shudder.

An example of **split-entities (missing links)** in Phrase Detectives' annotations. Instead, **ezCoref annotators** mark all mentions as **referring to the same entity**.

Analysis

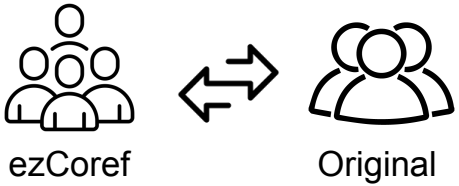


1. **Comparison** of ezCoref and original annotations



Low recall with QuizBowl suggesting ezCoref annotators miss mentions in the original clusters

Analysis

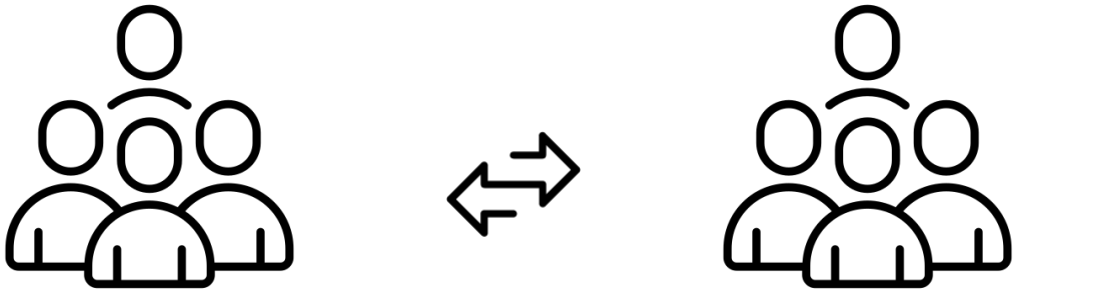


Low recall with QuizBowl suggesting ezCoref annotators miss mentions in the original clusters

[One character in this work]; is forgiven by [his] wife for an affair with a governess before beginning one with a ballerina.
In addition to [Stiva]; and [Nikolai], [another character in this work] (...) had earlier failed in [his] courtship of Ekaterina Shcherbatskaya.

An example from **QuizBowl** dataset with **cataphoric references missed** by **ezCoref annotators**

Investigating Agreements

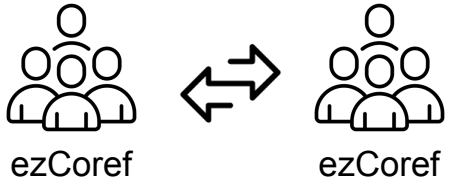


ezCoref Annotations

ezCoref Annotations

2. Agreements and disagreements among ezCoref annotators

Analysis

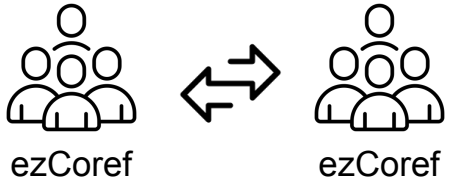


2. Agreements among ezCoref annotators

Which domains are suitable for crowdsourcing?

Fiction	Biographies	Opinion	Web	News	Wiki	Quiz

Analysis



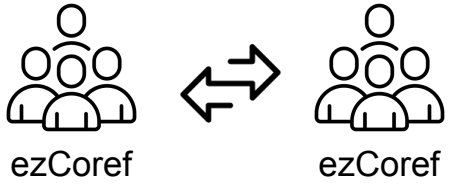
2. Agreements among ezCoref annotators

Which domains are suitable for crowdsourcing?

Fiction	Biographies	Opinion	Web	News	Wiki	Quiz
72.6	72.4	69.5	65.9	62.3	61.8	59.7

Domain-wise Inter Annotator **Agreement** using B3 (F1) scores

Analysis



2. Agreements among ezCoref annotators

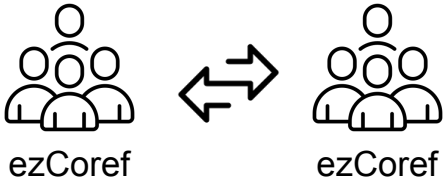
Which domains are suitable for crowdsourcing?

Fiction	Biographies	Opinion	Web	News	Wiki	Quiz
72.6	72.4	69.5	65.9	62.3	61.8	59.7

Domain-wise Inter Annotator **Agreement** using B3 (F1) scores

Fiction domain has highest inter-annotator agreement

Analysis

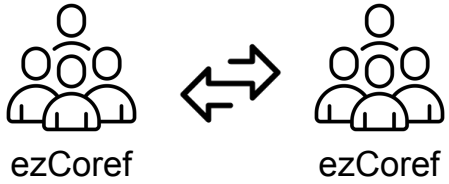


2. Agreements among ezCoref annotators

Fiction domain has highest inter-annotator agreement
abundant in **pronouns**, familiar **childhood stories**

A Wolf had been gorging on an animal **[he]** had killed, when suddenly a small bone in the meat stuck in **[his]** throat and **[he]** could not swallow it. **[He]** soon felt a terrible pain in **[his]** throat (...) **[He]** tried to induce everyone **[he]** met to remove the bone. “**[I]** would give anything," said **[he]**,” if [you]; would take it out.

Analysis



2. Agreements among ezCoref annotators

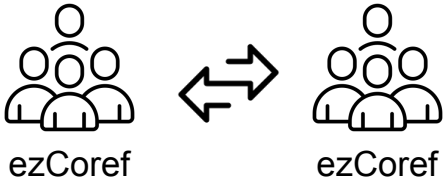
Which domains are suitable for crowdsourcing?

Fiction	Biographies	Opinion	Web	News	Wiki	Quiz
72.6	72.4	69.5	65.9	62.3	61.8	59.7

Domain-wise Inter Annotator **Agreement** using B3 (F1) scores

Quiz domain has lowest inter-annotator agreement

Analysis



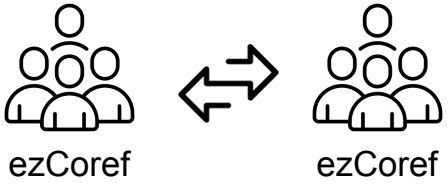
2. Agreements among ezCoref annotators

Quiz domain has lowest inter-annotator agreement

Abundant with **cataphora** and requires **factual knowledge**

[Another character in this work] rejects **[Ekaterina]** before (...) moving to St. Petersburg. For 10 points name this work in which **[Levin]** marries **[Kitty]**, (...) a novel by Leo Tolstoy.

Analysis



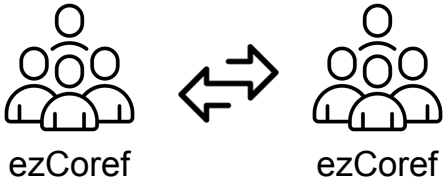
2. Disagreements among ezCoref annotators

Cases of genuine ambiguity in ezCoref annotations

[Fog] everywhere. **[Fog]** up **[the river]**, where **[it]** flows among green aits and meadows; **[fog]** down **[the river]**, where **[it]** rolls defiled among the tiers of shipping and the waterside pollutions of a great (and dirty) city.

—from *Bleak House* by Charles Dickens

Analysis



2. Disagreements among ezCoref annotators

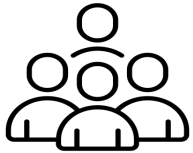
Cases of genuine ambiguity in ezCoref annotations

[Fog] everywhere. **[Fog]** up **[the river]**, where **[it]** flows among green aits and meadows; **[fog]** down **[the river]**, where **[it]** rolls defiled among the tiers of shipping and the waterside pollutions of a great (and dirty) city.

—from *Bleak House* by Charles Dickens

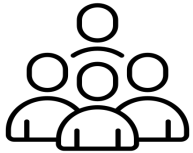
- Disagreement on mention **[it]** among ezCoref annotators
- Consistent with Szakolczai (2016)'s literary analysis

Investigating Agreements

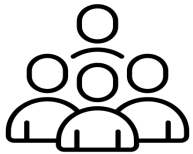


ezCoref Annotations

=



ezCoref Annotations



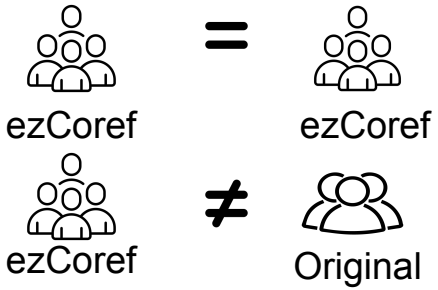
ezCoref Annotations

≠

Original Annotations

3. Deviations of ezCoref annotations from original annotations

Analysis



3. Deviations of ezCoref annotations from original annotations

Generic Pronouns

Maybe we need a CIA version of the Miranda warning: You have the right to conceal your coup intentions, because we may rat on you.

OntoNotes

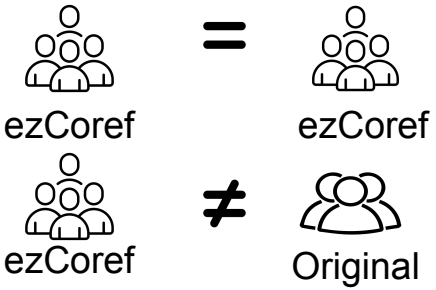
Not marked

Maybe [we] need a CIA version of the Miranda warning: [You] have the right to conceal [your] coup intentions, because [we] may rat on [you].

ARRAU

Marked as *undef-reference* (not coreferent)

Analysis



3. Deviations of ezCoref annotations from original annotations

Generic Pronouns

Maybe we need a CIA version of the Miranda warning: You have the right to conceal your coup intentions, because we may rat on you.

OntoNotes

Not marked

Maybe [we] need a CIA version of the Miranda warning: [You] have the right to conceal [your] coup intentions, because [we] may rat on [you].

ARRAU

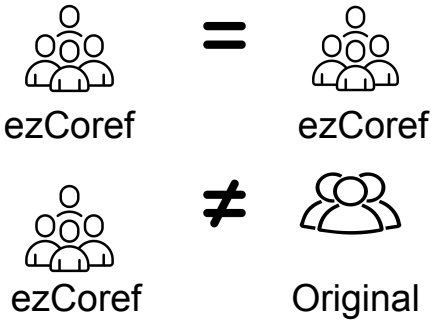
Marked as *undef-reference* (not coreferent)

Maybe [we] need a CIA version of the Miranda warning: [You] have the right to conceal [your] coup intentions, because [we] may rat on [you].

ezCoref annotations

Marked as coreferent

Analysis



3. Deviations of ezCoref annotations from original annotations

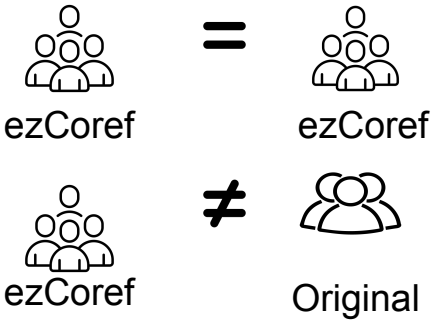
Appositives

Appos

[Fuad Basya], [spokesman for the Indonesian military], said (..) a warship was deployed to retrieve them.

LitBank

Analysis



3. Deviations of ezCoref annotations from original annotations

Appositives

Appos

[Fuad Basya], [spokesman for the Indonesian military], said (..) a warship was deployed to retrieve them.

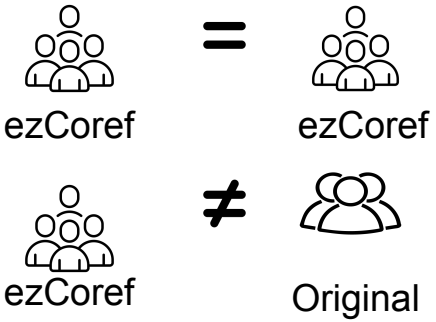
LitBank

Coref

[Fuad Basya], [spokesman for the Indonesian military], said (..) a warship was deployed to retrieve them.

GUM

Analysis



3. Deviations of ezCoref annotations from original annotations

Appositives

Appos

[Fuad Basya], [spokesman for the Indonesian military], said (..) a warship was deployed to retrieve them.

LitBank

Coref

[Fuad Basya], [spokesman for the Indonesian military], said (..) a warship was deployed to retrieve them.

GUM

Coref

[Fuad Basya], [spokesman for the Indonesian military], said (..) a warship was deployed to retrieve them.

ezCoref annotations

Takeaways

1. Open-source ezCoref annotation tool, conduct an annotation study

2. Coreference annotation remains a difficult task with **genuine ambiguities** (Plank, 2022)

3. **High-agreement with expert annotated datasets**, suggesting high quality annotations are achievable without extensive training.

4. Interesting **deviations** of ezCoref annotations from original annotations **(e.g., generic pronouns)**
   a. Research community should revisit these phenomena when curating future unified annotation guidelines

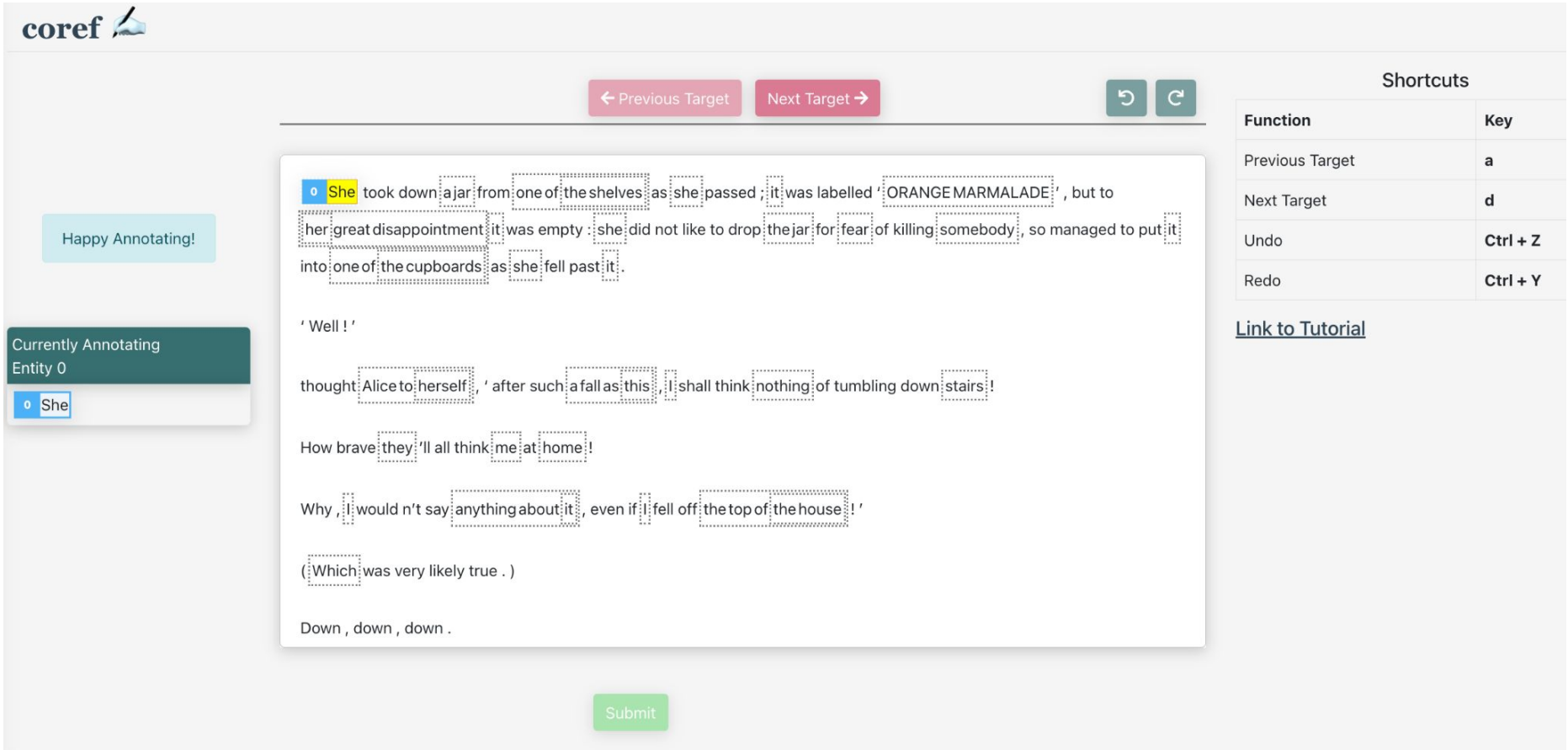
ezCoref

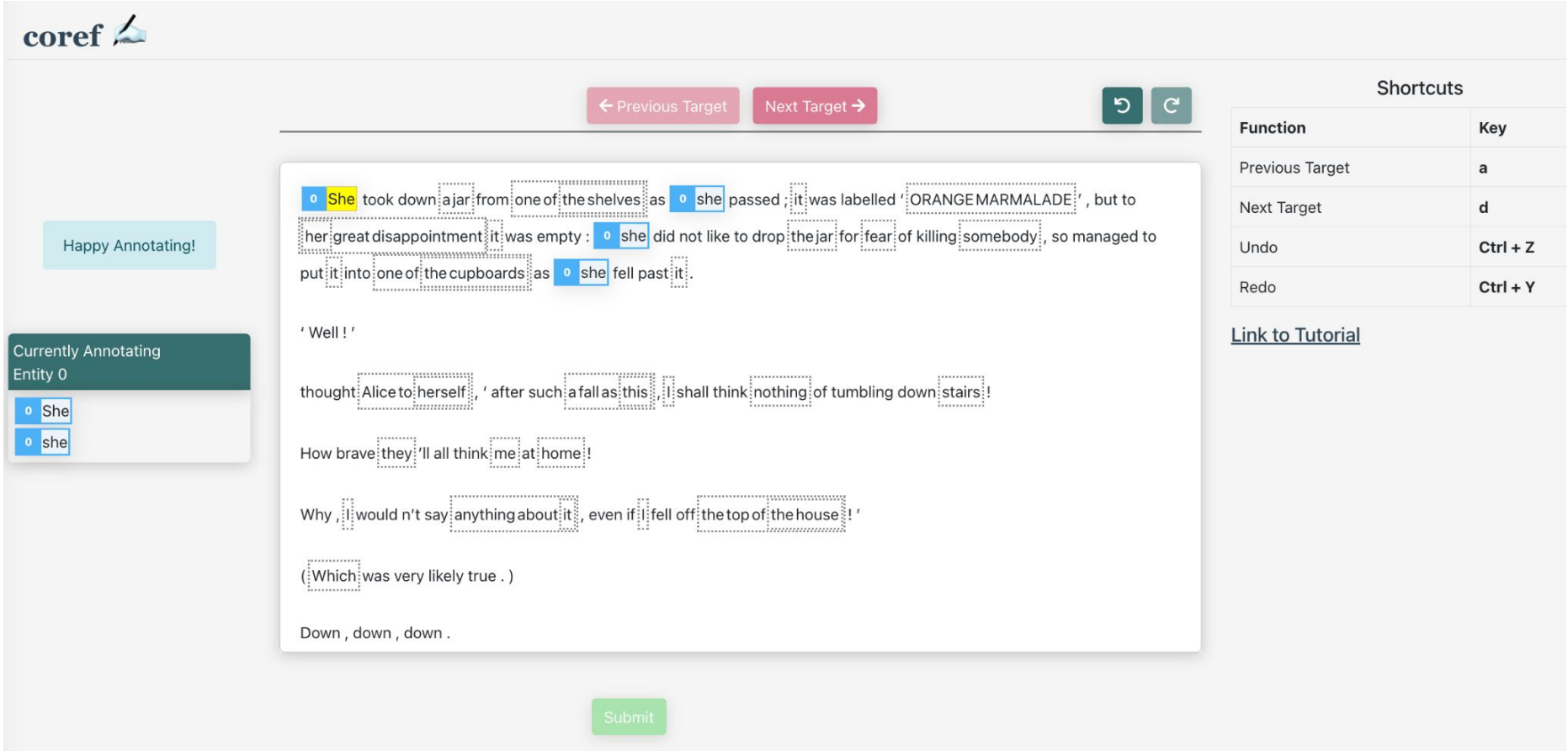
Towards Unifying Annotation Guidelines for Coreference Resolution

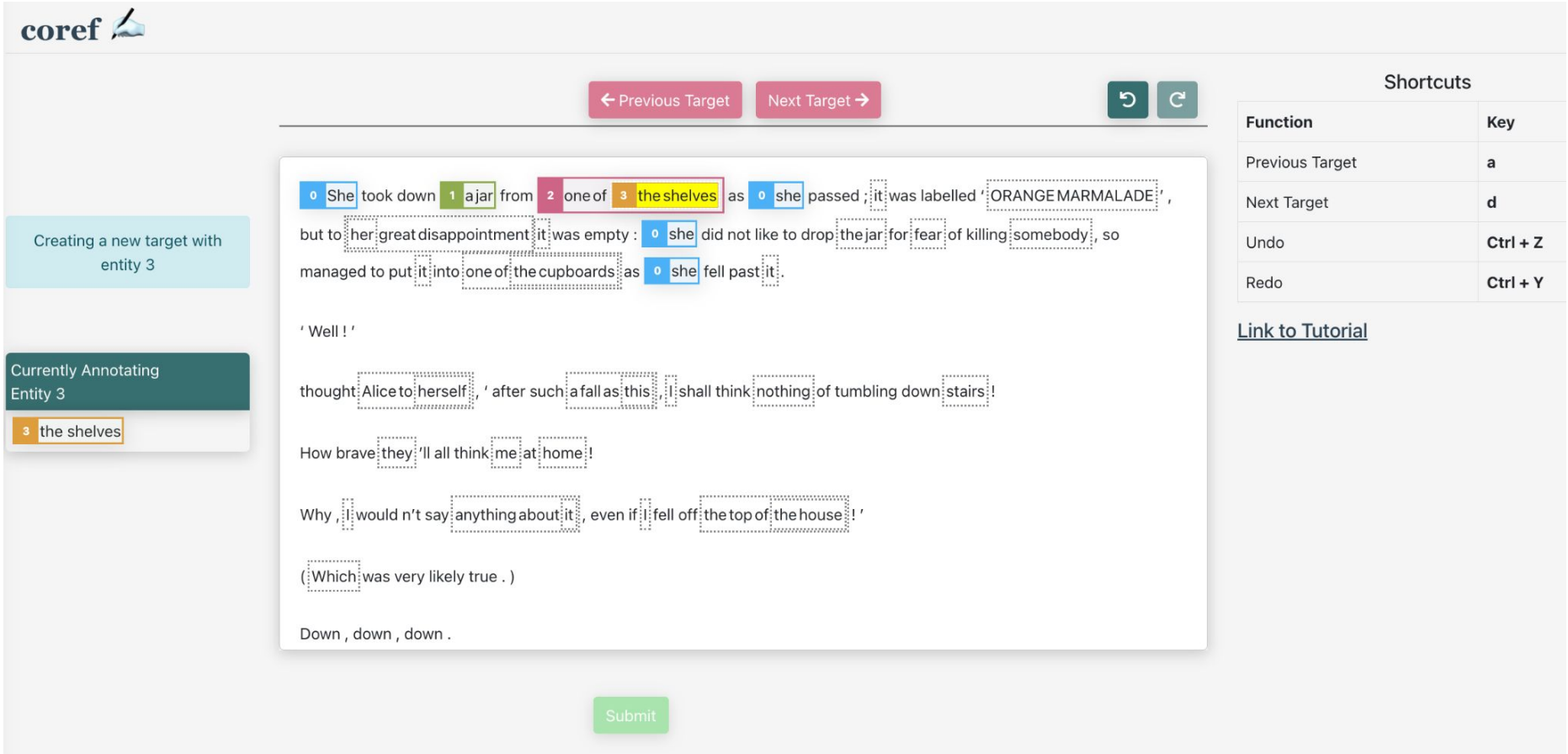
https://github.com/gnkitaa/ezCoref

Thank You. Questions?

Appendix


coref
Previous Target
Next Target
Shortcuts
Function
Key
Previous Target
a
Next Target
d
Undo
Ctrl + Z
Redo
Ctrl + Y
Link to Tutorial
Happy Annotating!
Currently Annotating
Entity 0
0 She
0 She took down a jar from one of the shelves as she passed ; it was labelled ' ORANGE MARMALADE ' , but to her great disappointment it was empty : she did not like to drop the jar for fear of killing somebody , so managed to put it into one of the cupboards as she fell past it .
' Well ! '
thought Alice to herself , ' after such a fall as this , I shall think nothing of tumbling down stairs !
How brave they 'll all think me at home !
Why , I would n't say anything about it , even if I fell off the top of the house ! '
(Which was very likely true .)
Down , down , down .
Submit


coref
← Previous Target
Next Target →
Happy Annotating!
Currently Annotating
Entity 0
0 She
0 she
0 She took down a jar from one of the shelves as 0 she passed ; it was labelled ' ORANGE MARMALADE ' , but to her great disappointment it was empty : 0 she did not like to drop the jar for fear of killing somebody , so managed to put it into one of the cupboards as 0 she fell past it .
' Well ! '
thought Alice to herself , ' after such a fall as this , I shall think nothing of tumbling down stairs !
How brave they 'll all think me at home !
Why , I would n't say anything about it , even if I fell off the top of the house ! '
(Which was very likely true .)
Down , down , down .
Submit
Shortcuts
Function
Key
Previous Target
a
Next Target
d
Undo
Ctrl + Z
Redo
Ctrl + Y
Link to Tutorial


coref
← Previous Target
Next Target →
Creating a new target with entity 3
Currently Annotating Entity 3
3 the shelves
0 She took down 1 a jar from 2 one of 3 the shelves as 0 she passed ; it was labelled ' ORANGE MARMALADE ' , but to her great disappointment it was empty : 0 she did not like to drop the jar for fear of killing somebody , so managed to put it into one of the cupboards as 0 she fell past it .
' Well ! '
thought Alice to herself , ' after such a fall as this , I shall think nothing of tumbling down stairs !
How brave they 'll all think me at home !
Why , I would n't say anything about it , even if I fell off the top of the house ! '
(Which was very likely true .)
Down , down , down .
Submit
Shortcuts
Function
Key
Previous Target
a
Next Target
d
Undo
Ctrl + Z
Redo
Ctrl + Y
Link to Tutorial

Simplified Annotation Guidelines

Example	Phenomena Taught
[John] doesn't like [Fred], but [he] still invited [him]to [the party].	(1) personal pronouns (2) singletons
[This dog] likes to play [catch].[It]'s better than other [dogs] at [this game]. [[Its] owner] is really proud.	(1) possessive pronouns (2) semantically similar expression which are not coreferring (3) non-person entities (animals)
[Director [Mackenzie]] spent [last two years] working on a ["Young Adam"]. During [this time] [he] often had to make [compromises] but [the movie] turned out to exceed expectations.	(1) nested spans (2) non-person entities (time, item)
[The office] wasn't exactly small either. [I]'m sure that 50, or maybe even 60, [people] could easily fit [there].	(1) non-person entities (place)