University of Massachusetts Amherst — BE REVOLUTIONARY

ezCoref: Towards Unifying Annotation Guidelines for Coreference Resolution

Introduction

- Several datasets have been collected facilitating the development of coreference resolution systems;
- Application to other domains and languages requires further data collection;

To collect new data we need: (1) **annotators**, (2) **guidelines**, (3) **annotation tool**.

Crowdsourcing opens new possibilities, **BUT**:
(1) there is **little agreement** on coreference guidelines;
(2) annotation tools are either **too complicated or not open-sourced**.

ezCoref Annotation Tool

- **Open sourced**
- **Integration with crowdsourcing platforms**
- **Crowdsourcing Friendly UI**

coref ✍

◀ Previous Target   Next Target ▶

Currently Annotating Entity 0

0 Alice was beginning to get very tired of sitting by
1 0 her sister on 2 the bank and of having
3 nothing to do: once or twice 0 she had peeped
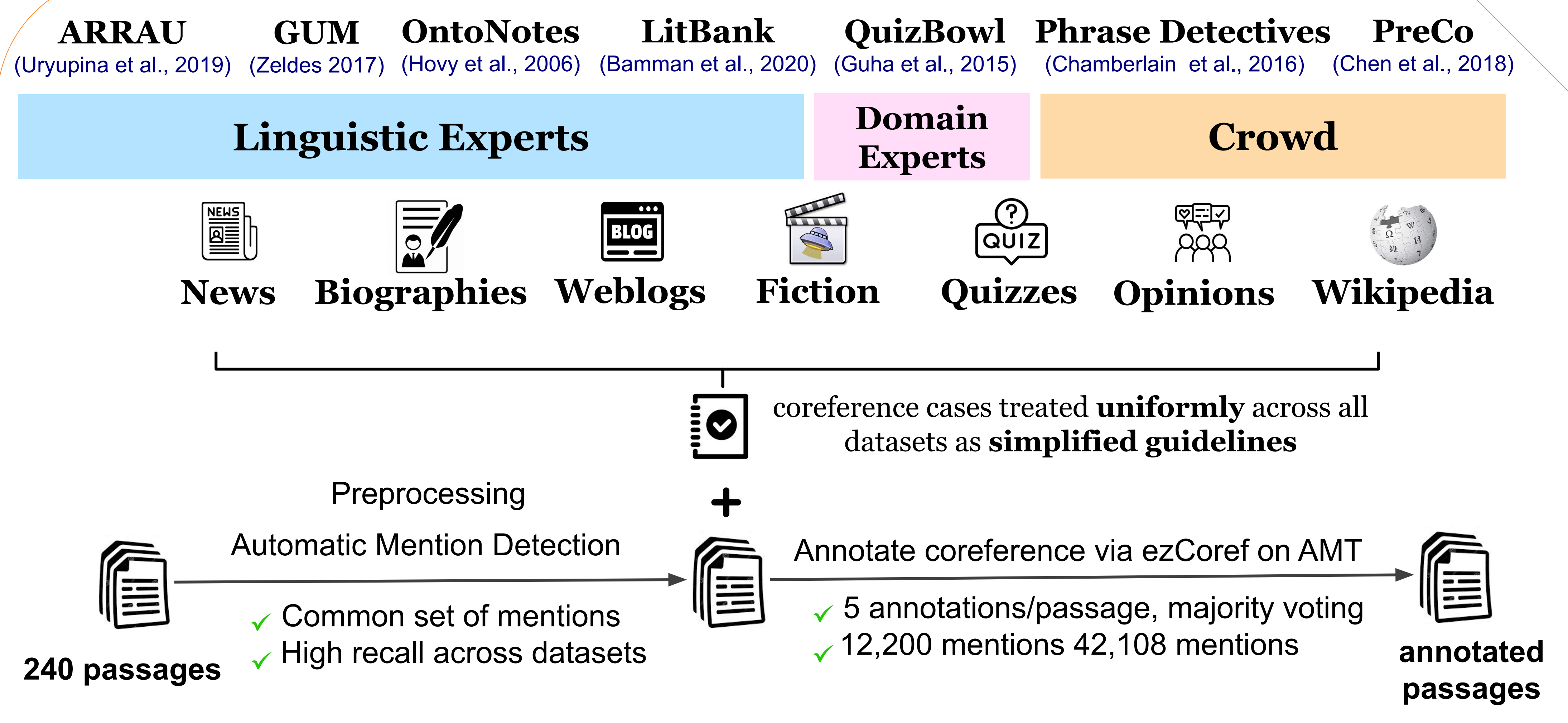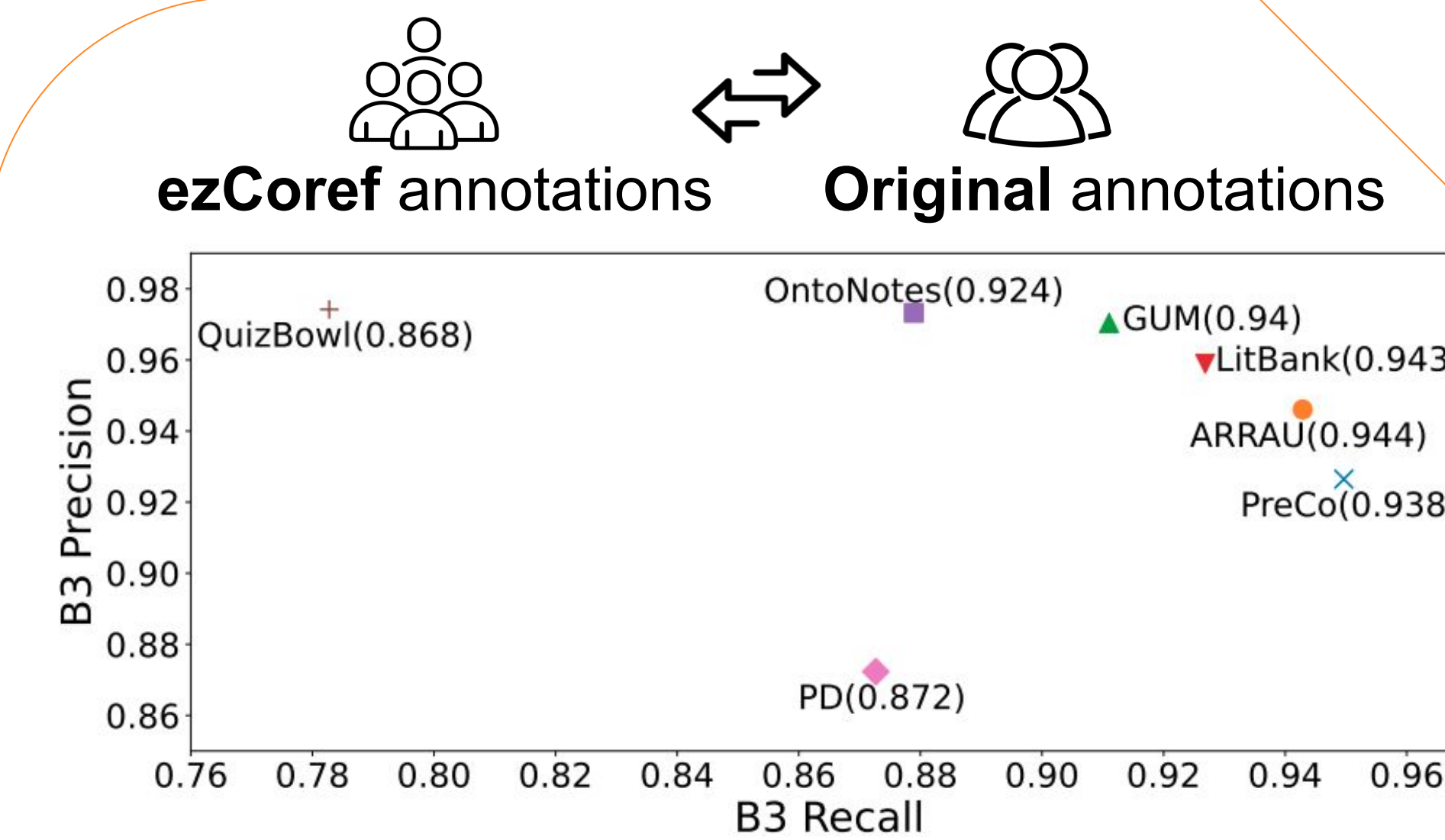into 4 the book her sister was reading (....)

Submit

Tutorial
- ✓ Interactive
- ✓ Customizable
- ✓ Validated by 70+ AMT annotators

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

Reannotation Study

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

Linguistic Experts	Domain Experts	Crowd

News   Biographies   Weblogs   Fiction   Quizzes   Opinions   Wikipedia

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

Preprocessing
Automatic Mention Detection
- ✓ Common set of mentions
- ✓ High recall across datasets

240 passages

Annotate coreference via ezCoref on AMT
- ✓ 5 annotations/passage, majority voting
- ✓ 12,200 mentions 42,108 mentions

annotated passages

Analysis

ezCoref annotations ⇄ **Original** annotations

QuizBowl(0.868)   OntoNotes(0.924)   ▲GUM(0.94)
                                        ▼LitBank(0.943)
                                        ARRAU(0.944)
                                        ✕PreCo(0.938)
                    PD(0.872)

(B3 Recall vs B3 Precision)

1. High Agreement with expert datasets
   a. Pronouns, Named Entities
   b. Appositives (GUM)

   [Fuad Basya], [spokesman for the Indonesian military]

   c. Generic Mentions (ARRAU and GUM)

   Kidder is putting [brokers] through a training course [..]. However, skeptics caution [...] [brokers] and investment bankers looks great only on paper.

2. Low precision with Phrase Detectives
   a. split-entities in PD annotations

   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].

3. Low recall with QuizBowl
   a. cataphora

   [One character in this work]; is forgiven by [his] wife (..). In addition to [Stiva] (..), [another character in this work] (...) had earlier failed in [his] courtship of Ekaterina Shcherbatskaya.

ezCoref annotations ⇄ **ezCoref** annotations

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

Domain-wise Inter Annotator Agreement (B3 scores)

1. High agreement on fiction
   a. abundant in pronouns, familiar childhood stories

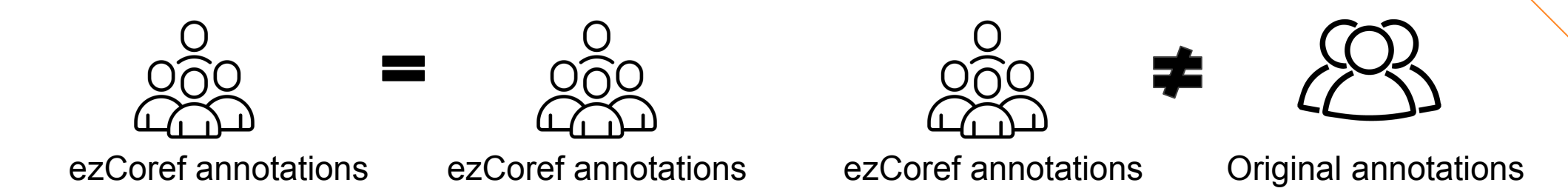   A Wolf had been gorging on an animal [he] had killed, when suddenly a small bone in the meat stuck in [his] throat and [he] could not swallow it. [He] soon felt a terrible pain in [his] throat (...) [He] tried to induce everyone (..)

2. Low agreement on QuizBowl
   a. cataphora
   b. world knowledge

   Another character in this work rejects [Ekaterina] before (...) moving to St. Petersburg. For 10 points name this work in which [Levin] marries [Kitty].

Deviations: Generic Pronouns

ezCoref annotations = ezCoref annotations   ezCoref annotations ≠ Original annotations

You have the right to conceal **your** coup intentions, because **we** may rat on **you**.

OntoNotes
[You] have the right to conceal [your] coup intentions, because [we] may rat on [you].

ARRAU
[You] have the right to conceal [your] coup intentions, because [we] may rat on [you].

ezCoref annotations

Takeaways

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

   [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. - Charles Dickens, Bleak House

   —from *Bleak House* by Charles Dickens

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

2. Reasonable quality annotations are already achievable even without extensive training.

3. Deviations of ezCoref annotations from original annotations suggest interesting cases which should be revisited by researchers when curating unified annotation guidelines.

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

NENLP 2023