

A Crowdsourcing Approach for Annotating Causal Relation Instances in Wikipedia

Kazuaki Hanawa, Akira Sasaki, Naoaki Okazaki✱, Kentaro Inui
(Tohoku University, ✱Tokyo Institute of Technology)

Abstract

Goal

- Annotate causal relation instances in Wikipedia

Approach

- Integrate a crowdsourcing service and brat

Contributions

- Collected 95,008 causal relation instances in 1,494 Wikipedia articles
(http://www.cl.ecei.tohoku.ac.jp/wikipedia_pro_sup/)
- The corpus can be used as supervision data for automatic recognition of causal relation instances
- Revealed valuable facts for improving the annotation process of this task

Annotation policy

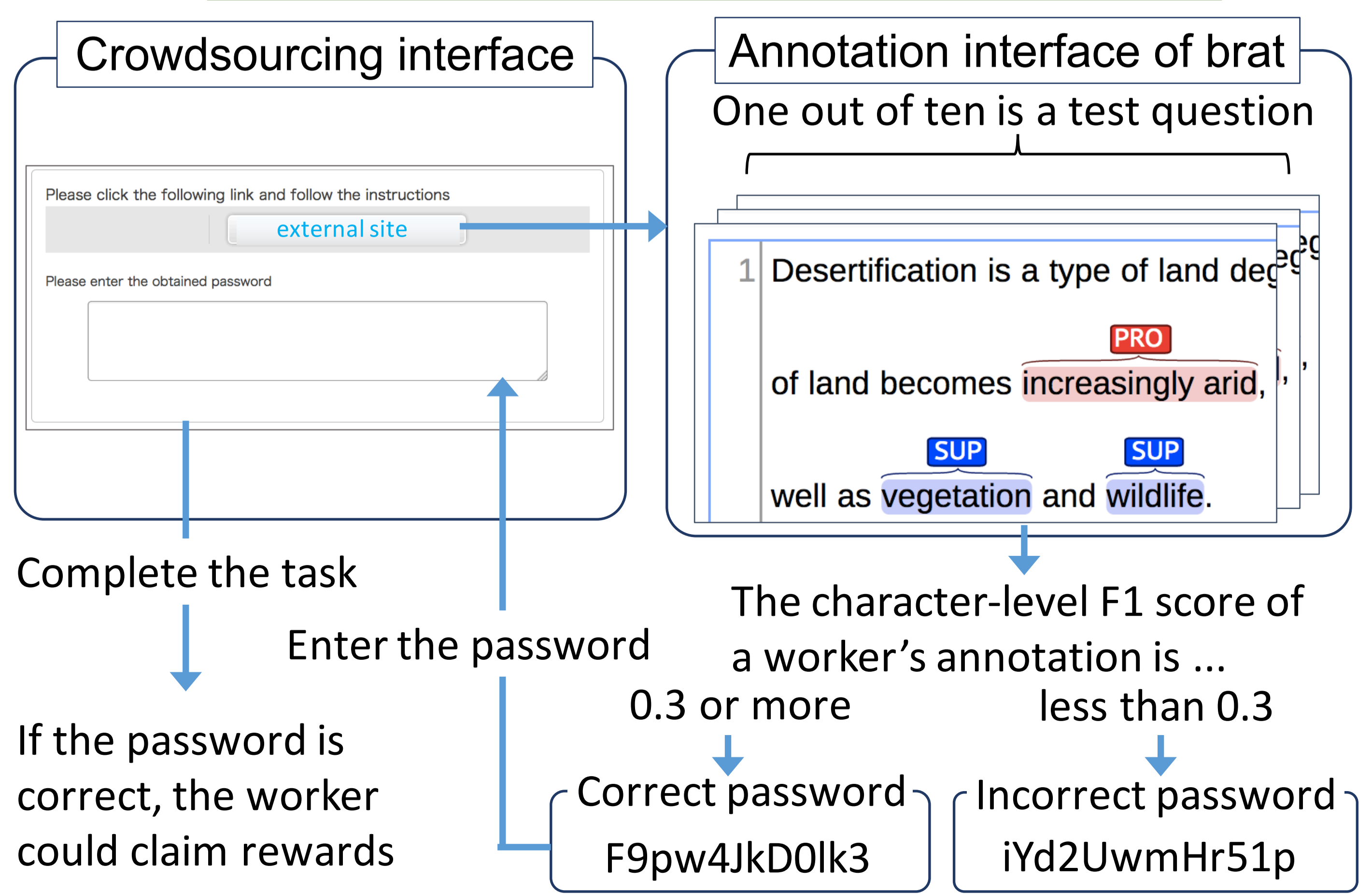
- X **promotes** Y
 - Y is activated when X is activated
- X **suppresses** Y
 - Y is inactivated when X is activated

Wikipedia article “Nyctalopia”

Nyctalopia, also called **night-blindness**, is a condition making it difficult to **see in relatively low light**. Nyctalopia may exist from birth, or be caused by **injury** or **severe malnutrition**.

- ⟨**PRO**, nyctalopia, night-blindness⟩
- ⟨**SUP**, nyctalopia, see in relatively low light⟩
- ⟨**PRO_BY**, nyctalopia, injury⟩ = ⟨**PRO**, injury, nyctalopia⟩
- ⟨**PRO_BY**, nyctalopia, severe malnutrition⟩

Using brat in crowdsourcing



Result

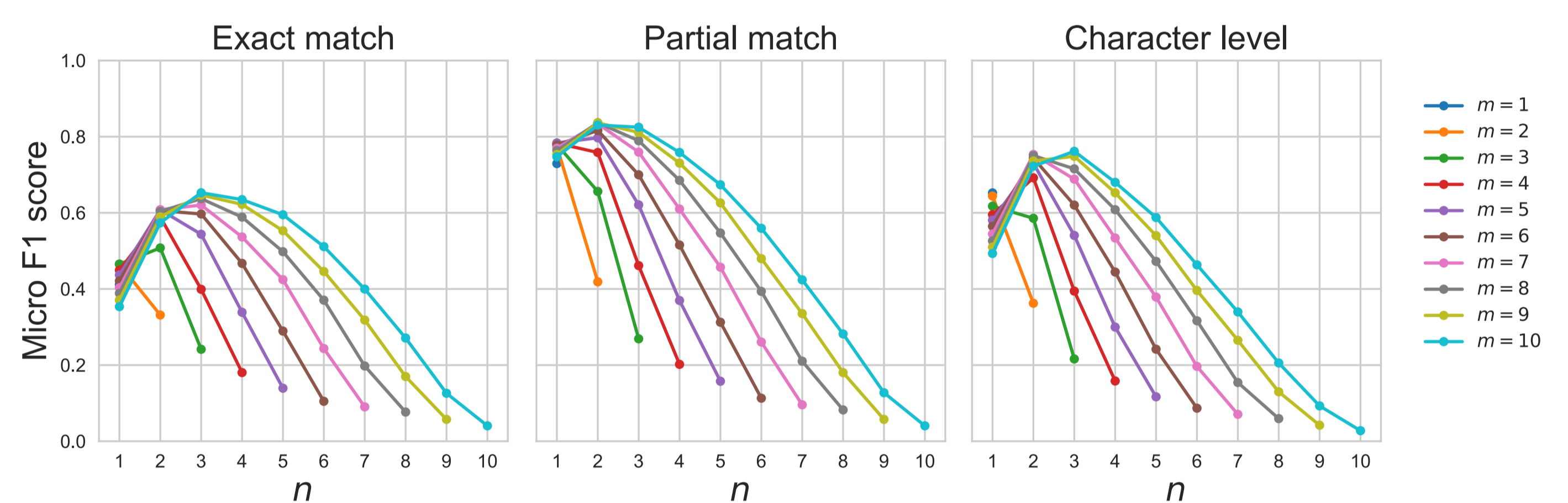
Collected ten annotations per an article

Example

PRO ... and result in high numbers of abnormal white blood cells.
 SUP Symptoms may include bleeding and bruising problems, ...
 PRO_BY the number of annotator Treatment may involve some combination of chemotherapy, ...
 SUP_BY

Micro-F1 between gold standard

- m : Number of annotators
- n : Adopt only spans with n or more exactly matched annotations



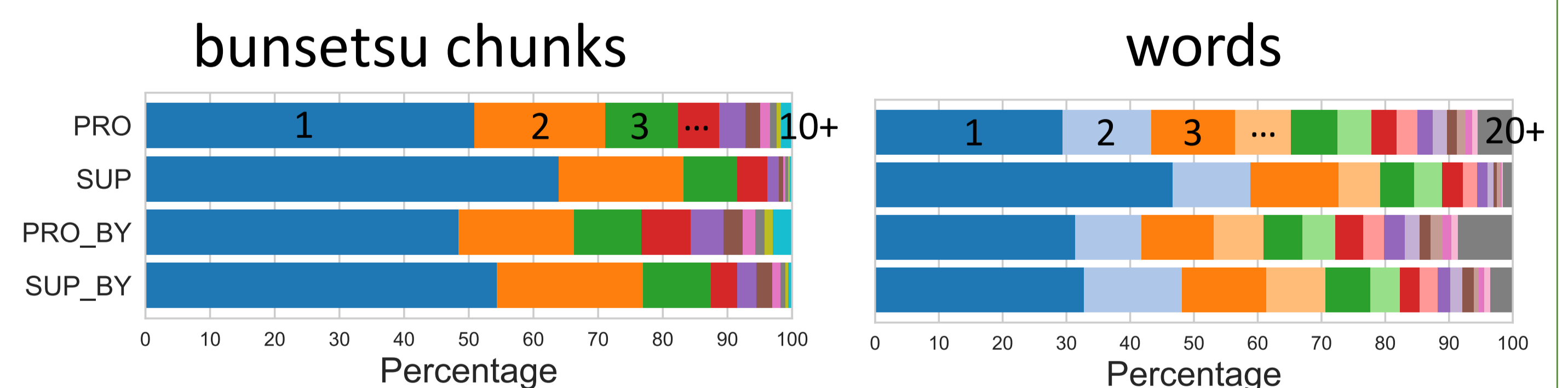
- Increasing the number of annotators improves the result
- Five annotations per article may be sufficient

Percentage of POS of head words

Noun	90.17	Mark	2.27
Verb	5.76	Particle	0.27
Auxiliary verb	1.09	Adverb	0.02
Adjective	0.41	Prefix	0.01

- It may be sufficient to limit annotation spans to noun phrases

Numbers of words and bunsetsu chunks



- Allowing crowd workers to choose their segment boundaries may be necessary

Automatic recognition

- Use $n = 2$ data as training and test data
- IOB2 notation was applied to the causal relations (e.g., B-PRO, I-PRO, B-SUP, I-SUP)
- Use one-layer bi-directional LSTM

Label	precision	recall	F1
PRO	0.507	0.364	0.424
SUP	0.354	0.275	0.310
PRO_BY	0.470	0.344	0.397
SUP_BY	0.259	0.178	0.211