COPA-SSE: Semi-Structured Explanations for Commonsense Reasoning

Ana Brassard[1,2] Benjamin Heinzerling[1,2] Pride Kavumba[2,1] Kentaro Inui[2,1]

[1]Riken AIP [2]Tohoku NLP Lab

https://github.com/a-brassard/copa-sse

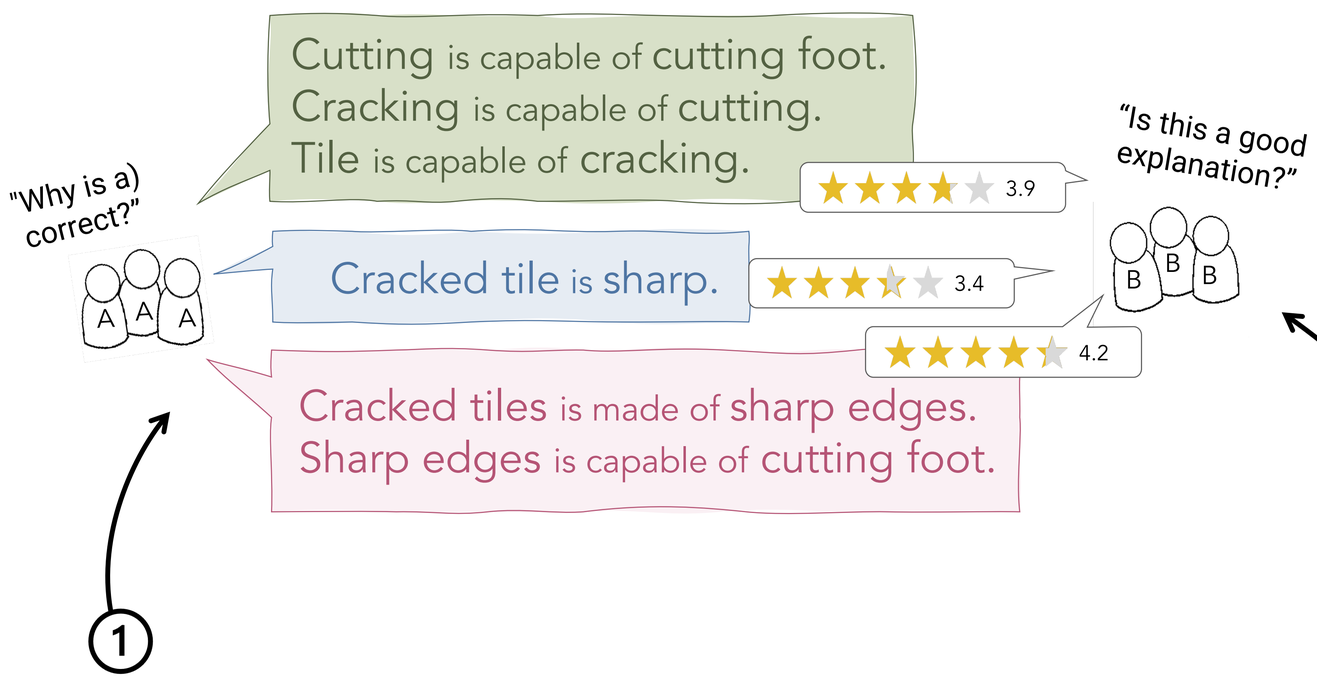
*Making explanation data is **hard**.*

Free-form crowdsourcing? ⟶ *Quality-control* issues

Extracting from KGs? ⟶ *Coverage/relevance* issues

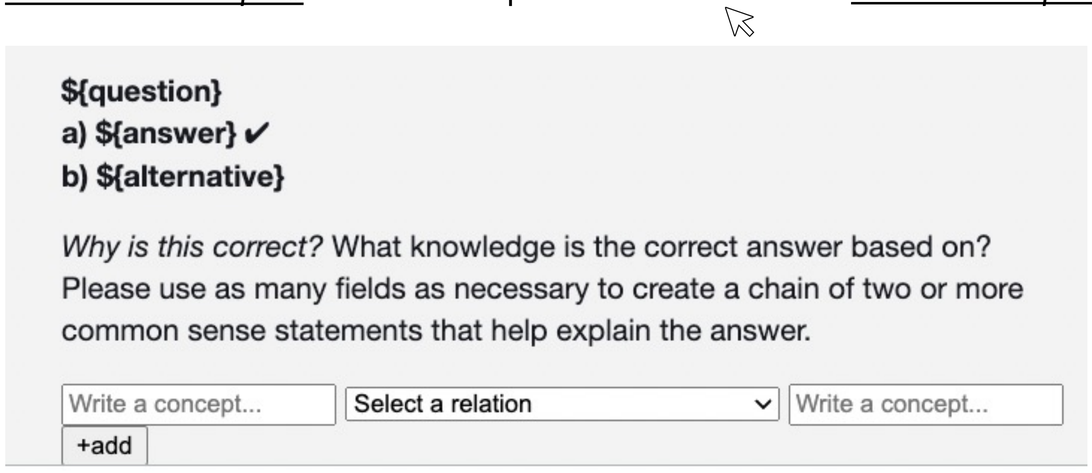
Our solution: **semi-structured** crowdsourcing!

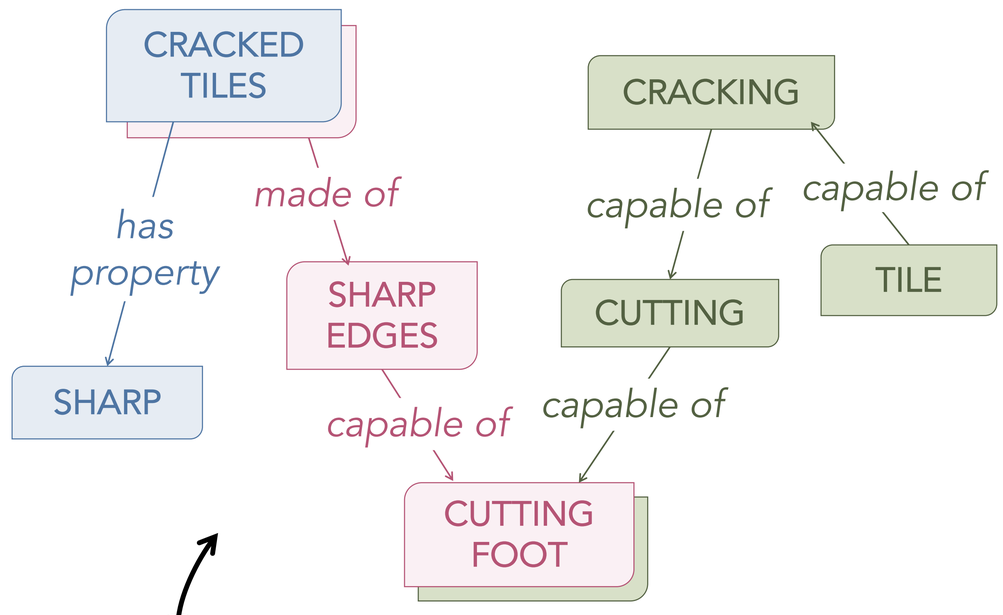
I cut my foot on the floor. What was the cause of this?
a) The tile was cracked. (correct)
b) The tile was wet.



Crowdworkers were shown COPA (*Choice of Plausible Alternatives*)[3] questions & asked to support the answer with **triples**:

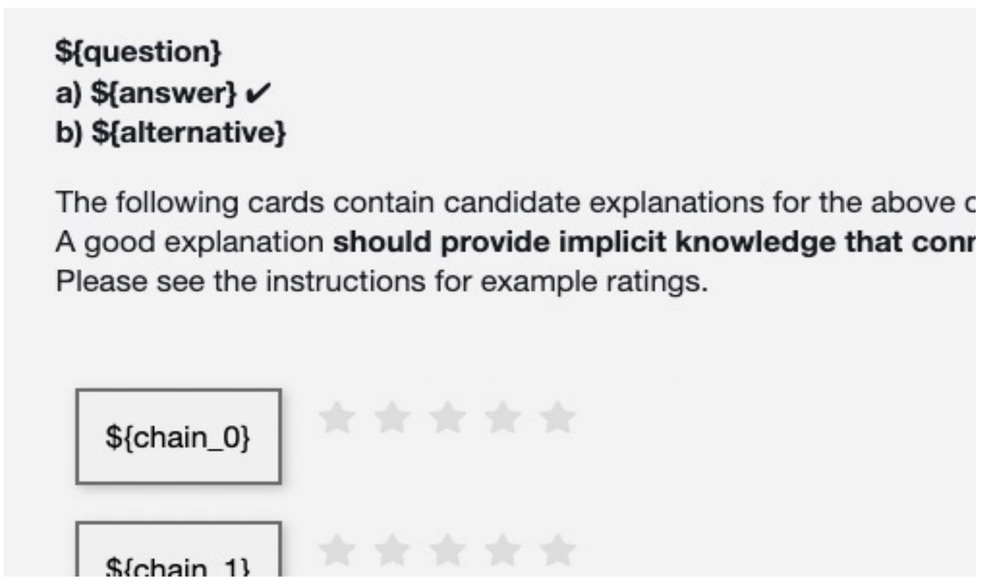
head concept <ConceptNet[4] relation> *tail concept*





Finally, the explanations can be aggregated into **graphs**, serving as examples of **relevant generated and/or extracted subgraphs.**

We then asked (different) crowdworkers to **rate** the explanations on a scale of 1 to 5 stars.

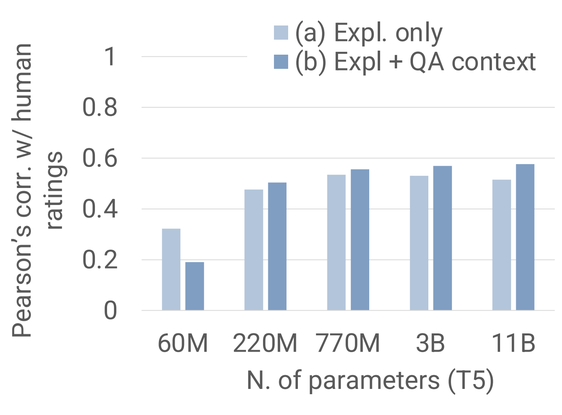


Quick stats

- **9,747** explanations for 1,500 Balanced COPA[5] questions
- 61% one-statement, 39% multi-statement
- 44% with 3.5 stars or more
- 98% of questions have at least one 3.5+ star explanation

Usage example | **Automatic explanation scoring**

- COPA-SSE has explanations w/ various quality ratings; can be used to **create automatic scoring** systems.
- Bigger T5 models[6] had higher correlation with human ratings, but still much room for improvement!



(a) Expl. only input:
Rate this explanation: Cracked tile is sharp.

(b) Expl. + QA context input:
Rate this explanation: I cut my foot on the floor. because The tile was cracked. Explanation: Cracked tile is sharp.

Expected output: **3.4**

[3] Roemmele, M., Bejan, C., and Gordon, A. (2011) Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning. AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, Stanford University, March 21-23, 2011.
[4] Speer, R., Chin, J., and Havasi, C. (2017) ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In proceedings of AAAI 31.
[5] Kavumba, P., Inoue, N., Heinzerling, B., Singh, K., Reisert, P., & Inui, K. (2019). When Choosing Plausible Alternatives, Clever Hans can be Clever. EMNLP 2019, 33.
[6] Roberts, A., Raffel, C., Lee, K., Matena, M., Shazeer, N., Liu, P. J., Narang, S., Li, W., and Zhou, Y. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. Technical report, Google.

RIKEN 東北大学