Collective Annotation: Applying Voting Theory to Computational Linguistics

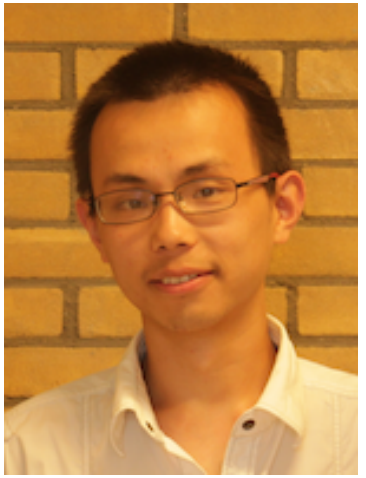
Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

[joint work with Raquel Fernández, Justin Kruger and Ciyang Qing]

Students Involved

Justin Kruger (Master of Logic 2014)

- Bachelor Philosophy, University of St Andrews, 2011
- Now: PhD Computer Science and Decision Analysis, Paris-Dauphine University



Ciyang Qing (Master of Logic 2014)

- Bachelor Computer Science, Peking University, 2012
- Now: PhD Linguistics & Cognitive Science, Stanford University

Challenge: Annotation for Linguistics

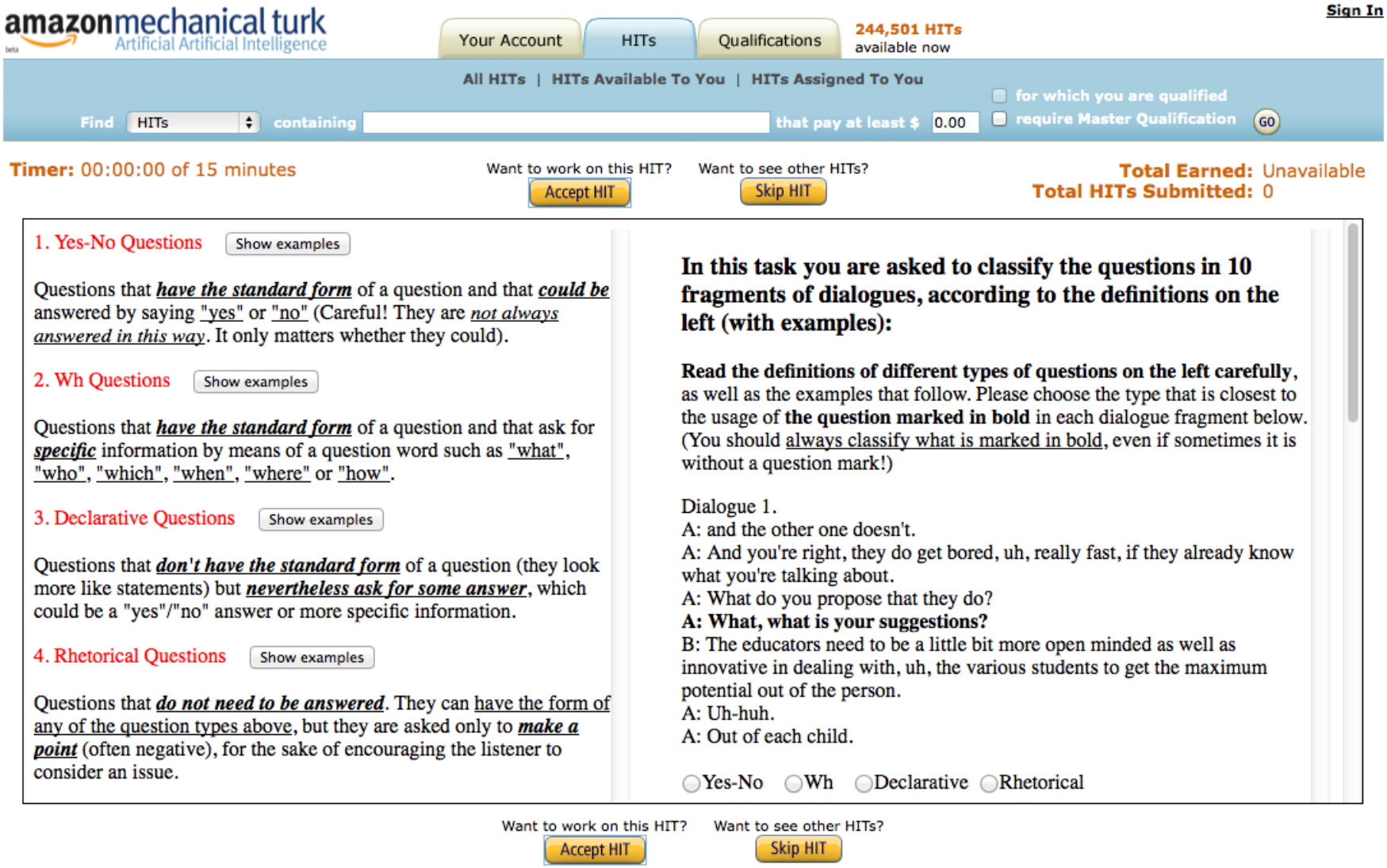
Imagine a researcher in computational linguistics, working on designing a new voice-controlled personal assistant, wants to understand what distinguishes *rhetorical questions* from other kinds of questions ...

They will need a lot of *annotated data*, like this:

B: [Noise] Yeah.
B: It, it's one of those necessities of life that we all have to, you know, pay taxes but, although it is kind of a pain sometimes though.
A: It's just scary though about, you know. —
A: How high are the taxes going to be when my children are my age?
B: Uh-huh.
A: You know, that, that's, that's scary too.

Yes-No ○ **Wh** ○ **Declarative** ○ **Rhetorical** ○

Collecting Raw Annotations: Crowdsourcing



Idea: Collective Annotation as Social Choice

Aggregating information from individuals is what *social choice theory* is all about. Classical case: aggregation of preferences in an election.

F: vector of individual preferences $\mapsto$ election winner

F: vector of individual annotations $\mapsto$ collective annotation

Example: Estimating Accuracy as Agreement

Naïve approach: *majority voting*. We have developed several more sophisticated aggregation rules. Here is one:

(1) Assume *annotator i* makes correct choice with *probability p_i*, and each of the wrong choices with equal probability $(1 - p_i)/(k - 1)$.

(2) Use *weighted majority voting*, giving more weight to annotators i with higher accuracy p_i. How much more? *Maximum likelihood* for:

$$weight_i \;=\; \log \frac{(k-1) \cdot p_i}{1 - p_i}$$

Great ... except that actually *we don't know* any of the p_i's!

(3) But we can try to *estimate* the *accuracy p_i* of annotator i as her observed *agreement with the simple majority rule*:

$$p_i \;\approx\; \frac{\#\,items\ where\ i\ and\ majority\ rule\ agree + 0.5}{\#\,items\ annotated\ by\ i + 1}$$

Results

Majority voting with *10 annotations* per item achieves *85% accuracy*, relative to an existing corpus annotated manually by experts.

Our rule achieves the *same accuracy* with just *6 annotations* per item.

For more rules, results, our papers, and our crowdsourced data, see:

`http://www.illc.uva.nl/Resources/CollectiveAnnotation/`

U. Endriss and R. Fernández. Collective Annotation of Linguistic Resources: Basic Principles and a Formal Model. Proc. ACL-2013.

J. Kruger, U. Endriss, R. Fernández, and C. Qing. Axiomatic Analysis of Aggregation Methods for Collective Annotation. Proc. AAMAS-2014.

C. Qing, U. Endriss, R. Fernández, and J. Kruger. Empirical Analysis of Aggregation Methods for Collective Annotation. Proc. COLING-2014.