

Introductory Course at ESLLI

*Bolzano, Italia
August 2016*



Crowdsourcing Linguistic Datasets LECTURE 5

Chris Biemann

biem@cs.tu-darmstadt.de

Lesson 5: Aggregation Strategies and Ethical Considerations

- Beyond Majority Vote: using knowledge about crowdworkers and tasks to improve the aggregation
- Gold Mine or Coal Mine: Working conditions and ethical considerations
- Voices of the Crowds

Aggregation of Multiple Signals

(materials taken from Alessandro Bozzon / Luca Galli)

- Majority Voting (MV) is a simple mechanism that does not take many factors into account
- Now: What signals can we use to improve MV?

What is truth?

- objective truth: exists freely and independently from subjects. E.g. number of birds in a picture
- cultural truth: shared belief of a group of people, often biased by their cultural background. E.g. how negative is “garbage” on a scale from 1 – 5?

Natural language is subjective and cultural by definition, thus we should not expect to have an objective truth available for most non-trivial tasks.

Latent Class Model

(materials taken from Alessandro Bozzon / Luca Galli)

- Observed: HIT outputs
- Latent: Truth, user experience, task difficulty, worker quality
- Desired: “Solution” that matches the truth

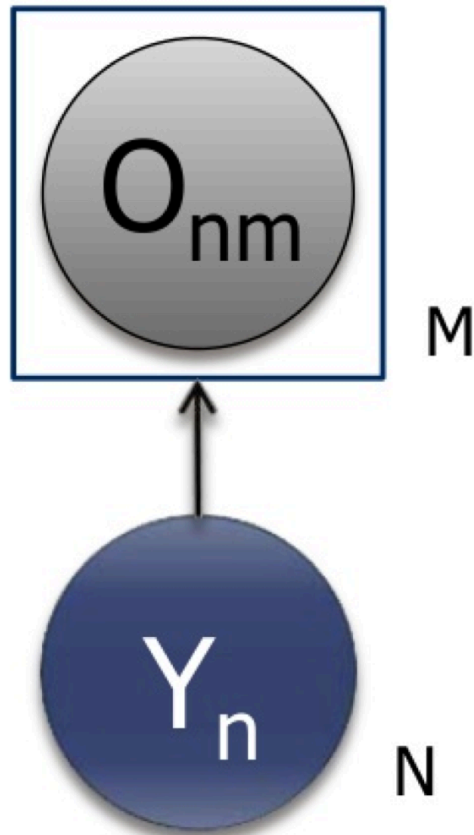
	Tasks					Solution	
Workers	O_{11}	O_{12}	...	O_{1N}	\vdots	Y_1	\vdots
	O_{21}	O_{22}	...	O_{2N}	\vdots	Y_2	\vdots
	\vdots	...	\vdots
	O_{M1}	O_{M2}	...	O_{MN}	\vdots	Y_M	\vdots

- Often, the matrix is incomplete
- Ground truth might never be known – see also <http://crowdtruth.org/framework>

Majority Vote as Latent Model

(materials taken from Alessandro Bozzon / Luca Galli)

Observed Output

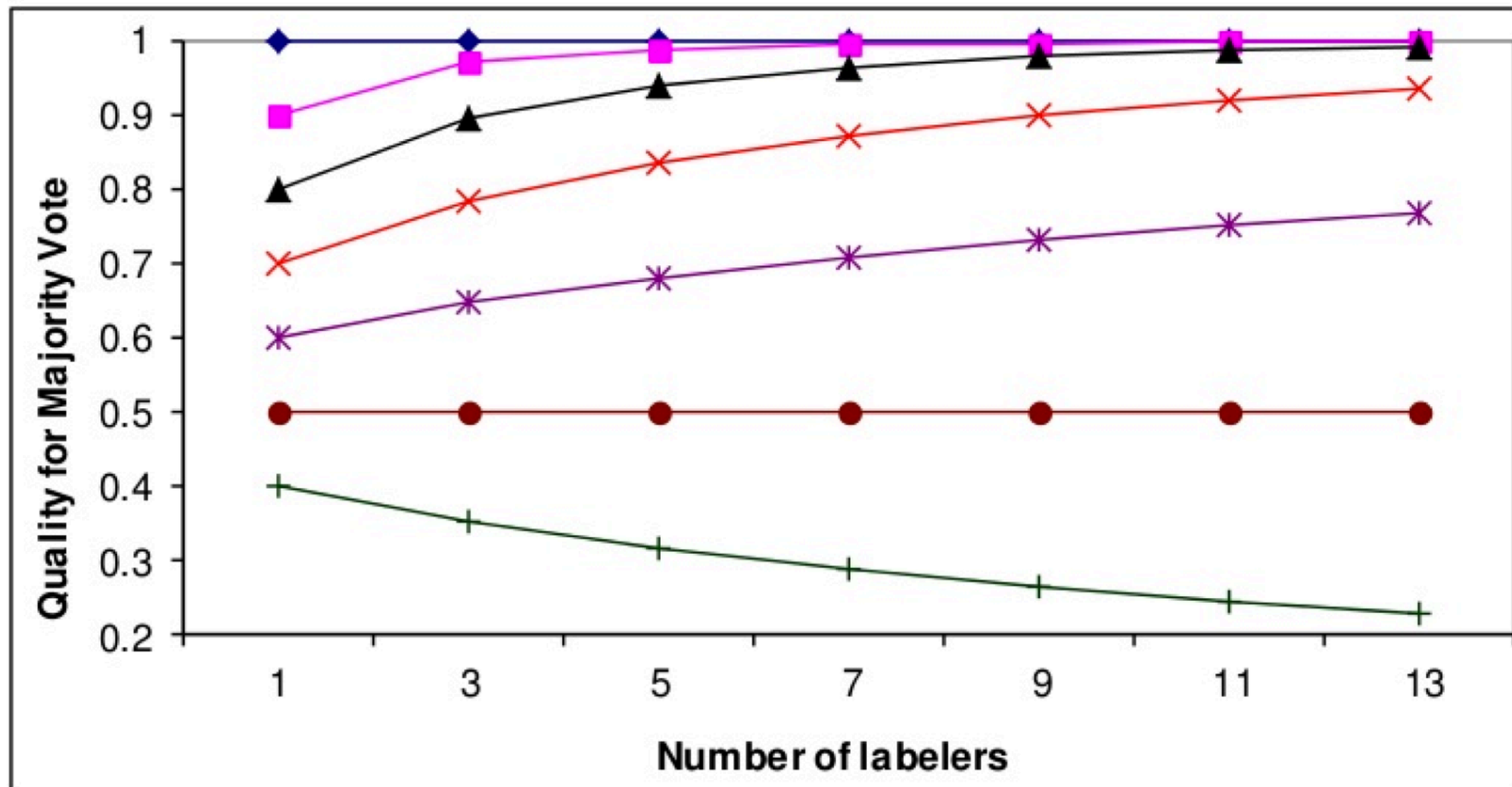


- MV: Ask multiple workers, keep majority as the “true” label
- Assumptions:
 - the output of each worker depends on the true answer
 - outputs are independent of the worker
 - outputs are independent of the (relative) difficulty within the task
- Quality: probability of obtaining the correct label directly depends on probability of correct observation

True Answer

Majority Vote for different levels of single worker quality

(materials taken from Alessandro Bozzon / Luca Galli)



- MV needs a minimum quality of 50%, otherwise it is harmful!

Hidden Factors

(materials taken from Alessandro Bozzon / Luca Galli)

- Majority Vote works best when workers have a similar quality
- Workers could still randomly guess and agree by chance
- Early experiments by Amazon to reject workers that deviated too much from the average (e.g. 5-point relevance scale) were discontinued

- Worker Characteristics

- Expertise (e.g. spelling skills)
- Bias (cultural, age, education)
- Physical Condition (e.g. fatigue)

- Task Characteristics

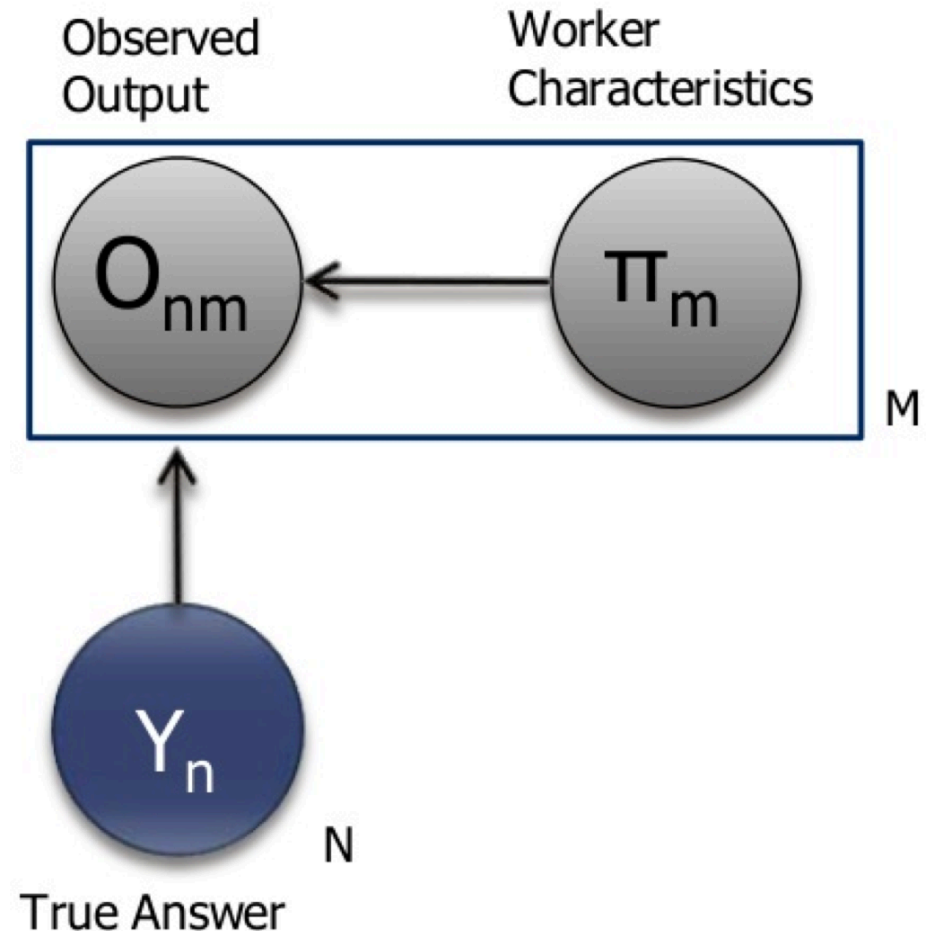
- Quality variance (e.g. longer/shorter text, grammar mistakes)
- Difficulty (e.g. transcription of native vs. non-native speech)



Incorporating Worker Quality

(materials taken from Alessandro Bozzon / Luca Galli)

- Example: Medical diagnosis by doctors
- Model: Doctors have different **rates** and **types** of errors
 - $\pi_{jl}^{(k)}$ defines the probability of doctor k to declare a patient in state l when the true state is j
 - $\eta_{il}^{(k)}$ is the number of times the doctor k gets response l from patient i



Incorporating Worker Quality

(materials taken from Alessandro Bozzon / Luca Galli)

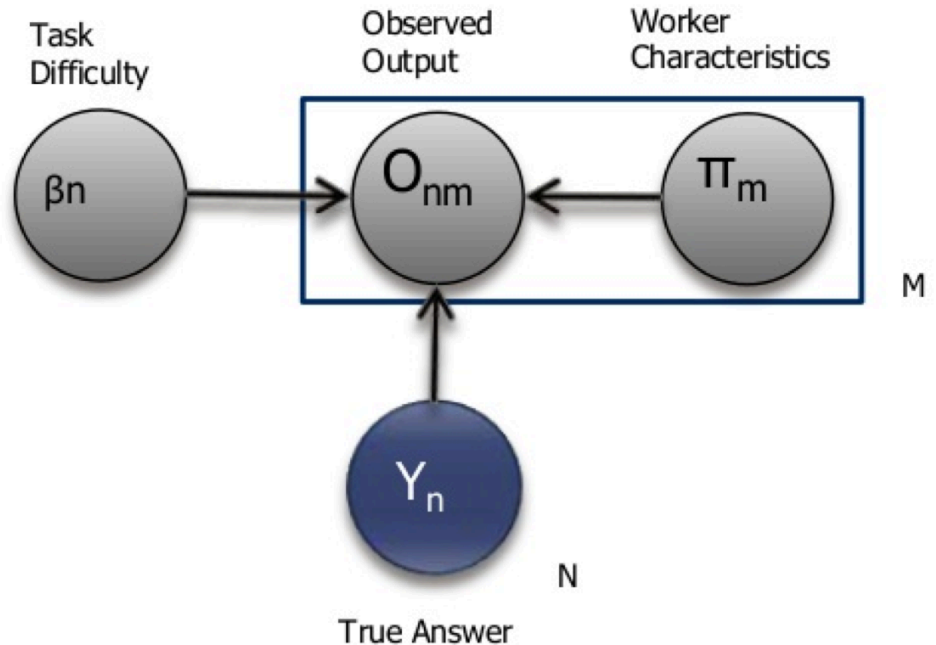
- Solving the equations: Expectation Maximization (EM) algorithm
- Estimate the confusion matrix and the true state simultaneously using EM, which
 - **estimates the true states** of each object by weighting the label votes **according to their current estimates of label quality** (from confusion matrix)
 - **re-estimate the confusion matrices** based on the **current beliefs about the true states** of each object

Dawid, A.P. and Skene, A.M. (1979): Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied Statistics*, 28(1):20–28
Dempster, A. P.; Laird, N.; Rubin, D. B. (1977): Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* 39 (1): 1–38

Incorporating Task Difficulty

(materials taken from Alessandro Bozzon / Luca Galli)

- Example: label images containing at least one “duck”
- Competency varies with image (blurry, b/w, background features)



Comparing Crowd Consensus Methods

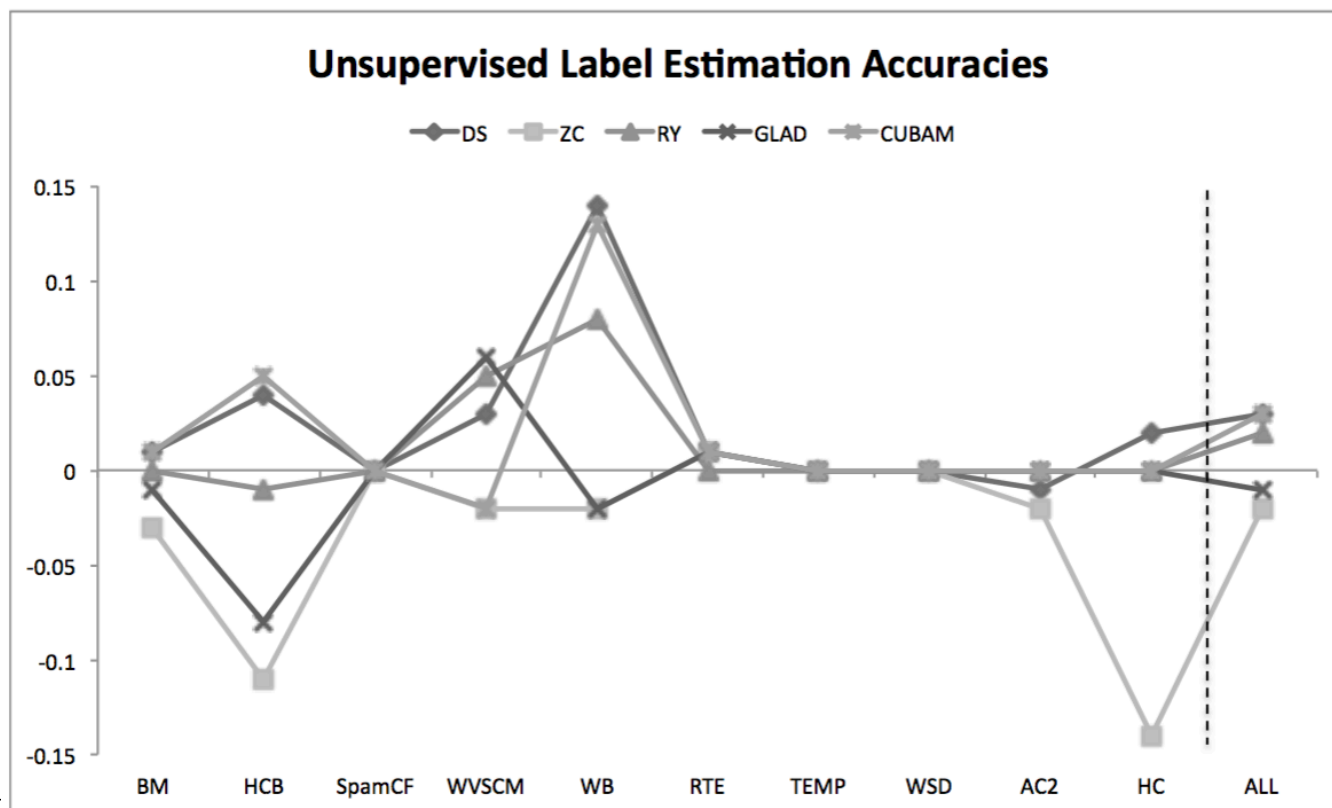
- offline consensus: given multiple noisy labels per example, how do we infer the best consensus label
- comparative evaluation on 5 NLP datasets and 4 non-NLP datasets
- all methods rely on EM in one way or another
- Mean worker accuracy is varied from 0.55 to 0.95

Dataset	Categories	Examples	Workers	Labels	MV Acc.
AC2	4	333	269	3317	88.1
BM	2	1000	83	5000	69.6
HC	3	3275	722	18479	64.9
HCB	2	3275	722	18479	64.8
RTE	2	800	164	8000	91.9
SpamCF	2	100	150	2297	66.0
TEMP	2	462	76	4620	93.9
WB	2	108	39	4212	75.9
WSD	3	177	34	1770	99.6
WVSCM	2	159	17	1221	72.3

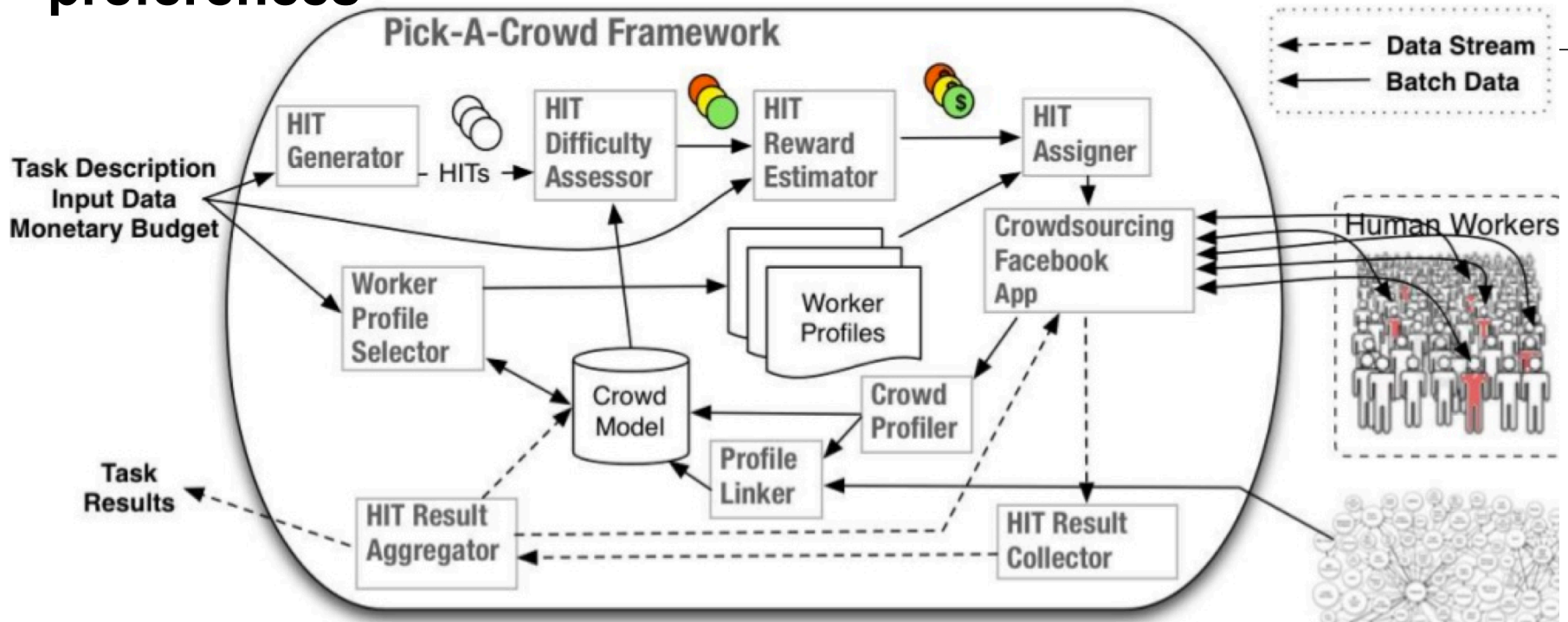
Table 2.1: Public datasets used in the SQUARE benchmark.

Comparing Crowd Consensus Methods II

- MV was often outperformed by some other method.
- classic DS (EM) and version that adds priors fares remarkably well
- each method was seen to outperform every other method in some condition

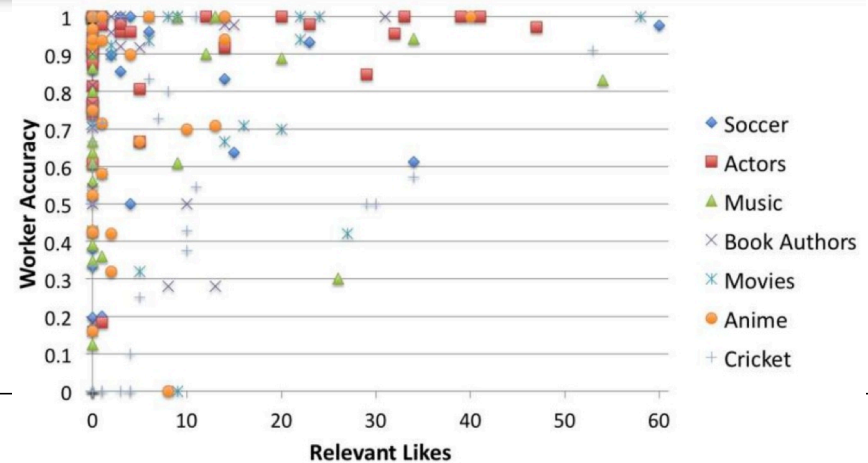


Pick-a Crowd: select set of workers based on their preferences



- Social Media Profiles from workers are correlated to quality

Difallah, D.E., Demartini, G., Cudré-Mauroux, P. (2013): Pick-a-crowd: tell me what you like, and i'll tell you what to do. In Proceedings of the 22nd international conference on World Wide Web (WWW '13). ACM, New York, NY, USA, 367-374



Summary on Modeling Agreement beyond MV

▪ Takeaways:

- Other signals can be employed that tell us about the worker's quality
- Methods based on Expectation Maximization improve label quality
- Main Idea: High-quality workers get a larger vote (implemented in Crowdfower's auto-aggregate)
- Caveat: only pays off for high volume tasks / requesters, as all these parameters have to be estimated reliably

Some more references:

- Demartini, G., Difallah, D. E., Cudré-Mauroux, P. (2012): ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In WWW, pages 469-478, New York, NY, USA
- Hovy, D., Berg-Kirkpatrick, T., Vaswani, A., Hovy, E. (2013): Learning Whom to Trust with MACE. In: Proceedings of NAACL-HLT 2013, pages 1120–1130, Atlanta, GA, USA
- Inel, O., Aroyo, L., Welty, C., Sips, R.-J. (2013): Domain-independent quality measures for Crowd Truth Disagreement. Detection, Representation, and Exploitation of Events in the Semantic Web 2 (DeRiVE 2013), Sydney, Australia
- Wang, J., Ipeirotis, P.G., Provost, F. (2011): Managing Crowdsourcing Workers. Winter Conference on Business Intelligence, 2011
- Jagabathula, S., Subramanian, L., Venkataraman, A. (2014): Reputation-based Worker filtering in crowdsourcing. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14). MIT Press, Cambridge, MA, USA, 2492-2500.
- Zhuang, H., Parameswaran, A., Roth, D., Han, J. (2015): Debiasing Crowdsourced Batches. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15). ACM, New York, NY, USA, 1593-1602.

Let's Crowdfsource!

POS tagging with UPOS

Universal POS tags

These tags mark the core part-of-speech categories. To

Open class words	Closed class words	Other
ADJ	ADP	PUNCT
ADV	AUX	SYM
INTJ	CONJ	X
NOUN	DET	
PROPN	NUM	
VERB	PART	
	PRON	
	SCONJ	

Alphabetical listing

Alice drove down the street in her car

- read instructions on the right, then:
- you have 5 seconds per word

- [ADJ](#): adjective
- [ADP](#): adposition
- [ADV](#): adverb
- [AUX](#): auxiliary verb
- [CONJ](#): coordinating conjunction
- [DET](#): determiner
- [INTJ](#): interjection
- [NOUN](#): noun
- [NUM](#): numeral
- [PART](#): particle
- [PRON](#): pronoun
- [PROPN](#): proper noun
- [PUNCT](#): punctuation
- [SCONJ](#): subordinating conjunction
- [SYM](#): symbol
- [VERB](#): verb
- [X](#): other

Let's Crowdsource! Dependency Parsing

Universal Dependencies

The idea of universal dependencies is to propose a set of universal grammatical relations which can be used with relative fidelity to capture any dependency relation between words in any language. Each dependency relation should be typed with one of the relations in the table below.

The table is adapted from *Universal Stanford Dependencies: A cross-linguistic typology* (de Marneffe et al. 2014). There have been modifications in the relations: we now have 40 universal relations (instead of the 42 ones proposed in the paper).

Note: *nmod*, *neg*, and *punct* appear in two places.

Core dependents of clausal predicates			Non-core dependents of clausal predicates			Special clausal dependents		
Nominal dep	Predicate dep		Nominal dep	Predicate dep	Modifier word	Nominal dep	Auxiliary	Other
nsubj	csubj		nmod	advcl	advmod	vocative	aux	mark
nsubjpass	csubjpass				neg	discourse	auxpass	punct
dobj	ccomp	xcomp				expl	cop	
iobj								
Noun dependents			Compounding and unanalyzed			Coordination		
Nominal dep	Predicate dep	Modifier word				conj	cc	punct
nummod	acl	amod	compound	mwe	goeswith			
appos		det	name	foreign				
nmod		neg						
Case-marking, prepositions, possessive			Loose joining relations			Other		
case			list	parataxis	remnant	Sentence head	Unspecified dependency	
			dislocated		reparandum	root	dep	

- [acl](#): clausal modifier of noun (adjectiva)
- [advcl](#): adverbial clause modifier
- [advmod](#): adverbial modifier
- [amod](#): adjectival modifier
- [appos](#): appositional modifier
- [aux](#): auxiliary
- [auxpass](#): passive auxiliary
- [case](#): case marking
- [cc](#): coordinating conjunction
- [ccomp](#): clausal complement
- [compound](#): compound
- [conj](#): conjunct
- [cop](#): copula
- [csubj](#): clausal subject
- [csubjpass](#): clausal passive subject
- [dep](#): unspecified dependency
- [det](#): determiner
- [discourse](#): discourse element
- [dislocated](#): dislocated elements
- [dobj](#): direct object
- [expl](#): expletive
- [foreign](#): foreign words
- [goeswith](#): goes with
- [iobj](#): indirect object
- [list](#): list
- [mark](#): marker
- [mwe](#): multi-word expression
- [name](#): name
- [neg](#): negation modifier
- [nmod](#): nominal modifier
- [nsubj](#): nominal subject
- [nsubjpass](#): passive nominal subject
- [nummod](#): numeric modifier
- [parataxis](#): parataxis
- [punct](#): punctuation
- [remnant](#): remnant in ellipsis
- [reparandum](#): overridden disfluency
- [root](#): root
- [vocative](#): vocative
- [xcomp](#): open clausal complement

Alice drove down the street in her car
 NOUN VERB ADP DET NOUN ADP PRON NOUN

▪ do this as quickly as possible!

How much should parsing short sentences pay?

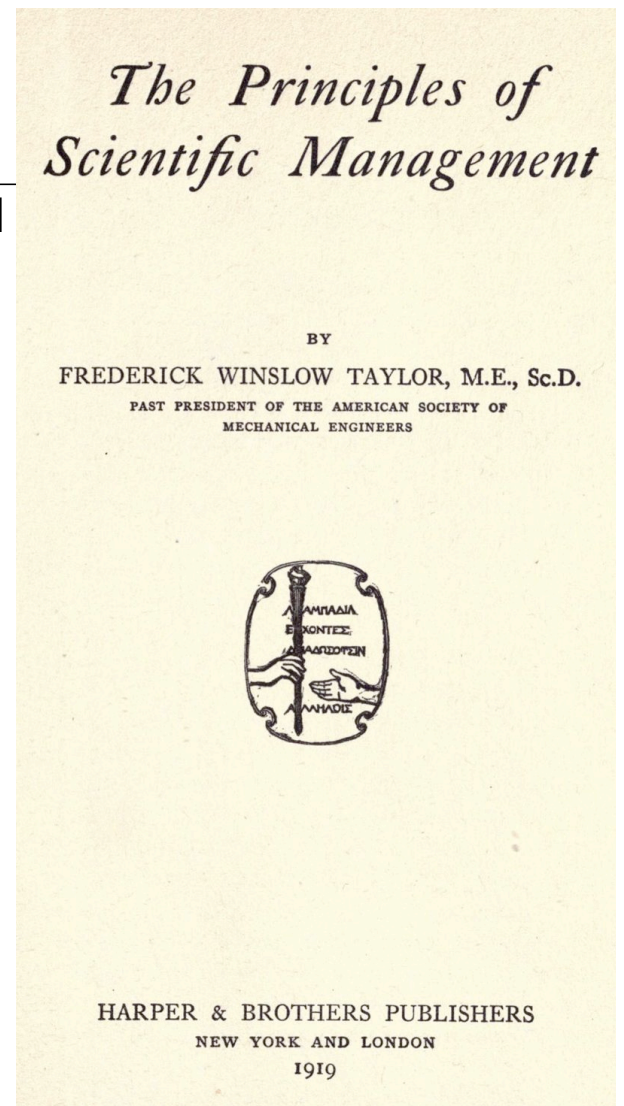
?

Crowdworking on Assembly Lines

In 1911, the American engineer Frederick Taylor delivered a paper in which he announced that **workers' natural laziness and propensity for underworking** was “the greatest evil which the working-people of both England and America are now afflicted.” His solution was a system of “scientific management,” wherein **work would be divided into the smallest repeatable tasks** and assigned a time limit. The aggregate of these tasks would then become the baseline for the workday, and “**those who fail to rise to a certain standard are discharged and a fresh supply of carefully selected men are given work in their place.**”

Almost a century later, **Amazon hit upon a similar approach to worker productivity.** Yet, whereas Taylor's genius was in super-charging the assembly line by reducing all skilled work to tiny micro-tasks, the genius of Mechanical Turk is in creating **virtual assembly lines.**

Moshe Z. Marvit



Home Stories 2014

- Over the last nine years, Milland has completed more than 830,000 tasks on Mechanical Turk, earning an average of 20 cents for each. “I hope our letters make Jeff Bezos realise that there are living, breathing human beings who rely on this service he provides to feed and shelter themselves and their families,” she says. **“I am a human being, not an algorithm, and yet [employers] seem to think I am there just to serve their bidding,”** writes Milland in her letter to Bezos. Amazon does not set minimum rates for work, which can pay less than \$2 a hour, and takes a 10% commission from every transaction. Employers can even refuse to pay for work altogether, with no legal consequences.
- “In the last four years, I have had so many problems getting my cheques that I have considered just quitting,” writes Manish, a Turker from India. **“There have been numerous cases of workers losing their accounts without any fault of their own. Workers are left with little to no support from Amazon.”**

Virtual Online Sweatshops

- Costello describes full-time Turking as “feast or famine,” but years of Turking have diminished her view of the feast. In February 2013, she worked approximately sixty hours a week searching for and performing HITs and made approximately \$150 per week—and that was the feast. The next month, she was unable to find as many “good-paying” HITs and earned only about \$50 per week. She describes how she **often stays up all night with the Mechanical Turk screen open**, because when people post a good batch of HITs, they go quickly.
- As Moshe Marvit pointed out in “The Nation”, cloud labor can spawn a **new breed of sweatshop**. The casualization of labor, in which there’s no longer such a thing as a steady job, much less a career, feeds neoliberal capitalist growth by allowing corporations to skirt regulations that ensure fair pay, guard against discrimination and provide workers with benefits and income supports if the company decides to shift to, say, a cheaper labor market abroad.

<https://www.thenation.com/article/crowdsourcing-bad-workers/>

<https://www.thenation.com/article/how-crowdworkers-became-ghosts-digital-machine/>

Commonly Voiced Issues by Crowdworkers

- Unfair rejection
- slow payment
- low pay
- lack of communication
- threat of suspension
- requester scams
- badly designed tasks
- information asymmetry
- lack/imbalance of power
- lack of search tools/user configuration

Qualitative Study of Turker Nation

- Turking is working → primarily motivated by earning money
- Considerable variation in earnings but it is low wage work
 - Highest earners \$15-16k per year (~ equivalent to 40 hours/per week, US minimum wage - \$7.25per hour) – peak at \$30,000
- Workers generally aspire to earning \$7-10 per hour
 - Newbies do lower paid easy work to increase their reputation and ranking
 - Lower wages off-set against search time, amount of concentration required etc.
- Turkers have preferences and skills
 - E.g. high volume grinding, writing, professional tasks, some multi-skilled
- AMT as a compromise – problems accessing the regular job market or need to supplement income.
 - Some housebound, others are in difficult circumstances

Qualitative Study of Indian Turkers

- Family and community collaboration
 - Word of mouth, Facebook groups etc.
 - Sharing accounts, market in trading accounts, training, CS companies
- Minimum English and some keyboard skills required
 - Lower skilled do simple and intuitive tasks; Higher skilled can earn a good wage by Indian levels
 - Danger of misunderstandings
- Infrastructure challenges, bricolage and back-ups
- Flexibility and turk-life-balance
 - Organise life around turking and are often helped by family
- Accounts/blocking/suspension, getting paid: Many of the participants no longer have accounts
- Cultural questions
 - Some operate on a basis of accepted = allowed

Otey vs. Crowdflower Class-Action Lawsuit

- Crowdflower settled a class action over minimum wage, overtime, and other Fair Labor Standards Act violations, paying out \$585,000 to US workers (Oct 2014) .
- Crux of this dispute: whether the workers are actually employees entitled to minimum wage or independent contractors.
- Line of argumentation: Crowdflower tasks systematically undercut the minimum wage in the US
- Notable:
 - crowdwork relevant for this settlement was all done on Amazon's Mechanical Turk under the requester name "Crowdflower"
 - Pricing is usually set by Crowdflower's customers
 - It is still possible to undercut the minimum wage on both platforms

Crowdworker Forums and Online Resources

- Crowdworkers are people. They talk about their work.
- Some Forums to watch:
 - <http://turkernation.com/> Turker Nation
 - <http://www.turkalert.com/> Turk Alert, also see the Blog
 - <http://www.mturkforum.com/> Mturk Forum
 - <https://turkopticon.ucsd.edu/> Turkopticon
 - <https://www.reddit.com/r/mturk> Reddit Thread
- Advertise your tasks
- Get feedback about your tasks
- notice issues early – don't get yourself into the Hall of Shame!

Turkopticon

“turkers strike back!”

[Help Idealist.org Find New Contact Information for Outdated Nonprofit Organizations](#) [View a HIT in this group](#)

Requester: [Idealist.org](#) HIT Expiration Date: Dec 2, 2009 (14 weeks) Reward: \$0.03
 Time Allotted: 60 minutes HITs Available: 6071

Description: Search the internet for new contact people for organizations on file on Idealist.org

Keywords: [Idealist](#), [Idealist.org](#), [nonprofits](#), [NGO](#), [Action](#), [without Borders](#)

Qualifications Required:
 HIT approval rate (%) is not less than 95

[Find articles on Travel](#) [View a HIT in this group](#)

Requester: [David Smith](#) HIT Expiration Date: Aug 28, 2009 (2 days 17 hours) Reward: \$0.13
 Time Allotted: 30 minutes HITs Available: 2911

Default worker view on MTurk

All HITs

1-10 of 2008 Results

Sort by: HITs Available (most first) GO!

[Show all details](#)

Extract purchased items from a shopping receipt

Requester: [Jon Breliq](#) HIT Expiration Date: Jul 2 h

communicativity:	<div style="width: 55%;"></div>	2.73 / 5
generosity :	<div style="width: 49%;"></div>	2.43 / 5
fairness :	<div style="width: 63%;"></div>	3.13 / 5
promptness :	<div style="width: 56%;"></div>	2.80 / 5

Requester: [What do these scores mean?](#)

Scores based on [241 reviews](#)

Terms of Service violation flags: 2

[Report your experience with this requester »](#)

Requester: [rohzi0d](#) HIT Expiration Date: Aug
 Time Allotted: 48

Turkopticon extensions

Ryan Bowden	FAIR: 1 / 5 <div style="width: 20%;"></div>	60 second time is absolutely ridiculous. I made a real effort to finish it in time but it's just not possible.
A3JUVC4SOAHJTY	FAST: NO DATA	
Averages =>	PAY: 1 / 5 <div style="width: 20%;"></div>	
HIT Group =>	COMM: NO DATA	
Review Requester =>		Avoid this requester until he gets a clue.
		Jul 03 2014 cau...@g... flag

Turkopticon – “turkers strike back”

Turkopticon plugin offers ratings on requesters, by fellow turkers, along these dimensions:

- **communicativity:** how responsive is the requester, how does the requester handle raised issues?
- **generosity:** how well are the HITs of this requester paid?
- **fairness:** how transparent is the rejection behavior of this requester?
- **promptness:** how quickly are the HITs approved/paid after their completion?

Also: comments about requesters.

Default worker view on MTurk

The screenshot shows two HIT listings from the MTurk interface. The top listing is for 'Idealist.org' with a reward of \$0.03 and 6071 available HITs. The bottom listing is for 'David Smith' with a reward of \$0.13 and 2911 available HITs. Both listings include fields for requester, expiration date, time allotted, reward, and number of available HITs, along with a description and keywords.


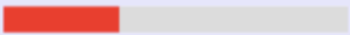
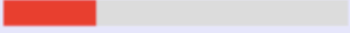

Requester	HIT Expiration Date	Reward	HITs Available
Idealist.org	Dec 2, 2009 (14 weeks)	\$0.03	6071
David Smith	Aug 28, 2009 (2 days 17 hours)	\$0.13	2911

Irani, L.C., Silberman, M.S. (2013): Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. CHI 2013: Changing Perspectives, Paris, France

Choose the best category for this government project (good english important)

Requester: ▼ [The Public Group](#)

HIT Expiration Date: Sep 10, 2013 (6 days)

communicativity:  1.17 / 5
generosity :  1.73 / 5
fairness :  1.39 / 5
promptness :  1.86 / 5

Time Allotted: 60 minutes

[What do these scores mean?](#)

Scores based on [81 reviews](#)

Terms of Service violation flags: 0

[Report your experience with this requester »](#)

[Contact Us](#) | [Careers at Amazon](#) | [Developers](#) |

©2005-2013 Amazon.com, Inc. or its Affiliates



Took a leap of faith on this requester and was rewarded with a %50 reject rate and a broken search feature and no feedback. Would not recommend, even if you have thousands of HITs under your belt to cushion the inevitable rejections.

Aug 29 2013 | [KBH19](#) | [flag](#) | [comment](#)



Arbitrarily rejected over half of the hits I submitted, and then banned me from submitting any more hits for them. I suppose that's a blessing in disguise though, as I had no intention of doing any for them again after the first batch of rejections.

Aug 21 2013 | [bour...@g...](#) | [flag](#) | [comment](#)



Their HIT is very unclear. There is an option to browse for the result, but it does not work.

Aug 20 2013 | [jeff...@g...](#) | [flag](#) | [comment](#)

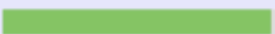

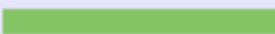

slide stolen from Chris Callison-Burch

Word Alignment

Requester:

▼ [Chris Callison-Burch](#)

HIT Expiration Date: Nov 12, 2013 (9 weeks)

communicativity:  4.05 / 5
generosity :  4.01 / 5
fairness :  4.25 / 5
promptness :  4.01 / 5

Time Allotted: 60 minutes

[What do these scores mean?](#)

Scores based on [95 reviews](#)

Terms of Service violation flags: 1

[Report your experience with this requester »](#)

[Contact Us](#) | [Careers at Amazon](#) | [Developers](#) |

©2005-2013 Amazon.com, Inc. or its Affilia



For those of you who know Arabic, this is a very solid requester with a very fair pay. Highly recommended for those who want to make some real money Payment usually take one week depending on the HITs you are doing

don't waste your time trying to submit machine translated crap, or random answers to multiple choice questions, you will get blocked instantly

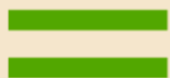
Jul 27 2013 | [hala...@h...](#) | [flag](#) | [comment](#)



Chris is one of the better requesters on MTurk, if you meet his qualifications and actually do the work as he requires. Glad to see that someone out there can finally work on those Arabic translation HITs that we've all seen for months now.

Jul 27 2013 [baudelai...@m...](#)

slide stolen from Chris Callison-Burch



Good requester. Everything approved in a couple of days. I had no problems. This is a safe requester to work for.

Quotes from the Requester Hall of Fame/Shame

A big thumbs up to this hit!! It was fun and actually rewarding. I wish they would contact me with further studies they need help with.

These guys are dishonest, thieving cowards who have no respect what so ever for Turkers, do yourself a favor and AVOID THEM AT ALL COSTS. Your better off doing XXX hits all day, at least you'll get paid.

XXX Inc. == BAD PEOPLE WITH NO RESPECT FOR YOU

I loved doing hits for this requester until I got a bunch of rejections in a row. About 3 of the rejections were legit the rest made no sense to me. The rejection claimed I got a 0 on the test questions, ...

A Final \$1 Mturk Experiment Just for This Tutorial

Tell Future Requesters How To Treat You

Requester: Chris Biemann

Reward: \$0.10 per HIT

HITs available: 0

Duration: 2 Hours

Qualifications Required: Masters has been granted

HIT Preview

Instructions

I am holding a tutorial on crowdsourcing next month. This is **your chance as a worker:** tell **future requesters** how they should treat you -- what is the **most important thing from your perspective?**

All answers will be approved that are 1) written in ENGLISH and have 2) at least 20 words in at least 2 sentences and are 3) on topic!

More is welcome but not required.

Tell your future requesters:

Please fill out this field.

What Crowdworkers would like to tell you

- **Take the task seriously and workers will do the same.** When I come across obvious typos--when it's clear the requestor didn't put in effort to check their own work--I do not take the work seriously either. It's disrespectful to the workers and the task at hand when the survey itself is of low quality.
- Pay a **fair wage** for the work that you are requiring someone to do. At bare minimum, a pay rate to equal to the minimum wage standard. Be available to communicate with the Turk workers. Most of us want to do good work for you and **sometimes we need a little extra guidance so we ask questions. Responding to us promptly** will help ensure we give you what you are looking for which makes all of us happy.
- **Communicate and clarify.** The best way to get usable results is to provide clear instructions and respond when a worker asks for clarification. Value all feedback and remember that **there are no stupid questions**; what might seem clear to you isn't necessarily so for everyone.

What Crowdworkers would like to tell you II

- Honestly, I think the most important thing for requesters to remember is that **crowdworkers are also people** who are just trying to make some money by working like anyone else, except for a myriad of reasons we do it this way instead of going to an office. Those of us who have been doing this long enough understand the problems requesters face: The relatively small number of scamming workers that cause a tremendous amount of problems, etc. I think the good requesters are the ones who understand that workers also deal on a regular basis with requesters who use our work without paying us, who believe we should be grateful to spend our time working for less than half of minimum wage, or who generally insult and belittle us if contacted (often assuming we are all 18-year-old college students with all the bias that entails). **Pay us as fairly as possible within your budget, treat us with respect when we contact you respectfully, and try to make your instructions as clear and coherent as possible to help us understand what you're looking for.** The vast majority of us really do want this to be a mutually beneficial relationship, as do the majority of requesters.

What Crowdworkers would like to tell you III

- The most important thing a requester can do it **to pay a fair amount of money** for each HIT. This should be at least 20 cents per minute
- The most important thing for requesters to know is that **crowdworkers are actual people** and as such should **be treated with respect and fair pay**. Remember that the crowdworkers choose to do your work, and they can just as easily choose not to. Many crowdworkers are highly educated individuals who for one reason or another may not be employed in the traditional workplace. Quite a few are actually employed and do crowdwork for extra income. Few crowdworkers are doing crowdwork just for amusement or to kill time while waiting to meet with the Prime Minister. One final thing, which probably should have been first - **always give clear instructions, and have a system in place for clarifications** when needed, and they will be needed!

What Crowdworkers would like to tell you VI

- You should treat me as **someone who deserves respect**. I think you should also treat me nicely.
- In my experience, workers are looking for a few key things in terms of crowdsourcing work. First, workers **want and need clearly written, precise instructions** that are not verbose. Second, **the pay should be at least ten cents per minute**. Lastly, workers are typically working for extra money, and want a **quick approval process**. In the unlikely event, work is rejected for some reason, there should be a **clear reason why the work is rejected**. The vast majority of the workers do high quality work. Thank you for the opportunity to participate, it is greatly appreciated.

Other Tutorials and Materials (incomplete)

- 2016 UPenn class on “Crowdsourcing and Human Computation”, Chris Callison-Burch, see <http://crowdsourcing-class.org/>
- 2016 ICWSM Tutorial “How to Use Crowdsourcing Effectively for Social Media & Web Science Research” by Ujwal Gadiraju, Gianluca Demartni, Djellel Eddine Difallah, Michele Catasta
- 2015 NAACL-HLT Tutorial “Crowdsourcing for NLP” by Chris Callison-Burch, Lyle Ungar and Ellie Pavlick
- 2013 ICWE Tutorial “An Introduction to Human Computation and Games With A Purpose” by Alessandro Bozzon and Luca Galli
- 2011 WSDM Tutorial “Crowdsourcing 101: Putting the WSDM of Crowds to Work for You” by Omar Alonso and Matthew Lease
- 2010 NAACL-HLT Workshop “Creating Speech and Language Data With Amazon’s Mechanical Turk” by Chris Callison-Burch and Mark Dredze

In a Nutshell: Learned in Lesson 5

Statistical Quality Control

- Methods to improve over majority voting
- approaches based on Expectation Maximization work well
- no “best method”, still ongoing research

Ethics of Crowdsourcing

- Crowdsourcing platforms can be viewed as sweatshops
- After initial euphoria, crowdworkers demand rights
- There are tools that support the choice of genuine/trustworthy requesters
- Build your reputation as a requester
- Don't underpay, don't reject deliberately
- Check the forums!

In a Nutshell: Lessons Learned Overall

- Keep HITs **extremely** simple
- Provide **redundant** instructions
- Crowdworkers are mostly **not linguists**
- **Never trust a single** worker. Trust the crowd.
- Watch out for **scammers**:
 - use qualification requirements
 - use test items
- **Iterate** on small batches, watch closely
- Watch the **forums**
- Always give crowdworkers **an exit** and a room for **comments**
- Be a **crowdworker yourself** for a while – you will notice that they are **human beings**, and you will treat them as such.

THE END



The crowds are waiting for you!

<http://www.greenbookblog.org/wp-content/uploads/2013/11/crowd.jpg>