

Fast Cheap Scalable Good enough

The Promise of Crowdsourcing

Chris Callison-Burch



human language technology
center of excellence



CENTER FOR LANGUAGE
& SPEECH PROCESSING



Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.
 Workers select from thousands of tasks and work whenever it's convenient.

37,649 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



or [learn more about being a Requester](#)

All HITs | HITs Available To You | HITs Assigned To You

Search for containing that pay at least \$ for which you are qualified

Timer: 00:00:00 of 5 minutes Want to work on this HIT? Want to see other HITs?

Accept HIT

Skip HIT

Total Earned: Unavailable
Total HITs Submitted: 0

Enter Postmark & Stamp Information for a Postcard

Requester: Cardcow

Reward: \$0.01 per HIT

HITs Available: 2

Duration: 5 minutes

Qualifications Required: Data Entry for Postcards has been granted

Enter Postmark & Stamp Information for this card



Postmark City:

Postmark State:
(or Country)

Postmark Date:
(Ex: Nov-09)

(month & day)

Postmark Year:
(Ex: 1909)

Stamp: (Ex: 1c, 2c,
half penny)

Requester: Cardcow

Reward: \$0.01 per HIT

Duration: 5 minutes

has been granted



or this card

Postmark City: Barre

Postmark State: MA

Postmark Date: Oct-11

Postmark Year: 1886

Stamp:

1c

Postmark City:

Postmark State:
(or Country)

Postmark 1
(Ex: Nov-09)

(month & day)

: 1c, 2c,

\$ 0.01

NLP Annotation

Snow, O'Connor, Jurafsky and Ng's EMNLP 2008 paper pioneered the use of Mechanical Turk for NLP

- Affect Recognition

fear("Tropical storm forms in Atlantic") > fear("Goal delight for Sheva")

- Word Similarity

sim(lad, boy) > sim(rooster, noon)

- Textual Entailment

if "Microsoft was established in Italy in 1985" then "Microsoft was established in 1985"?

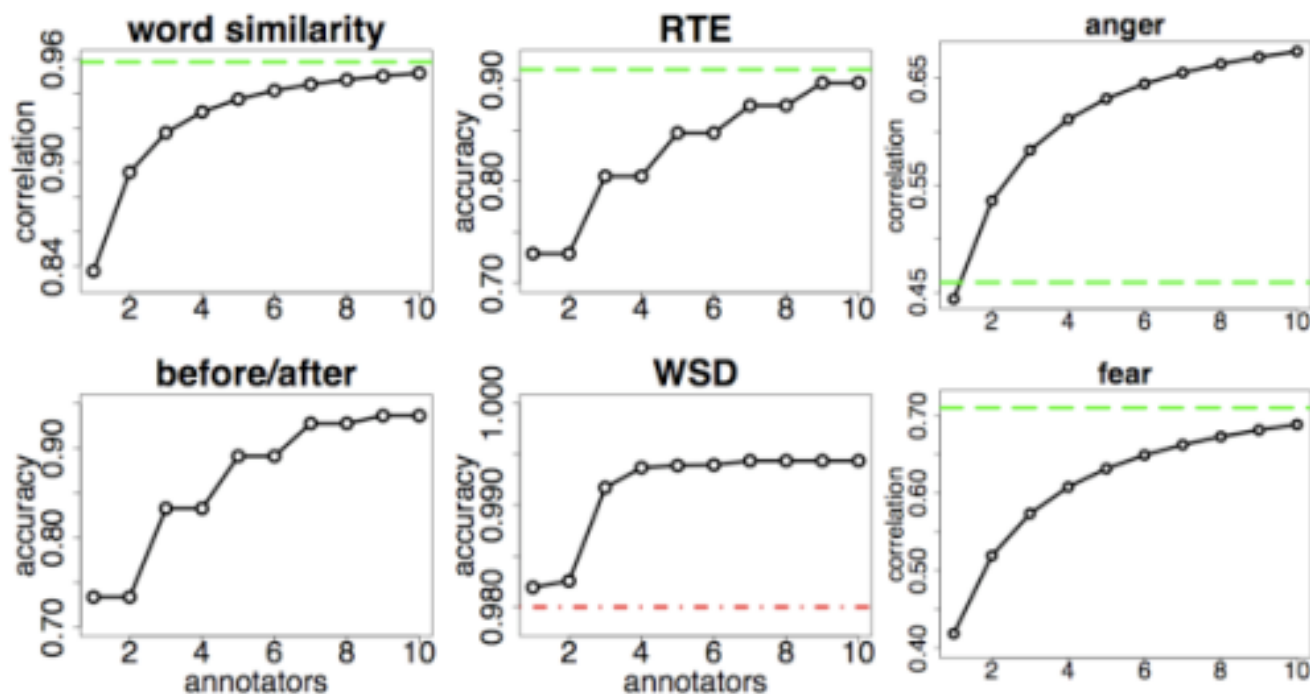
- Word Sense

"a bass on the line" v. "a funky bass line"

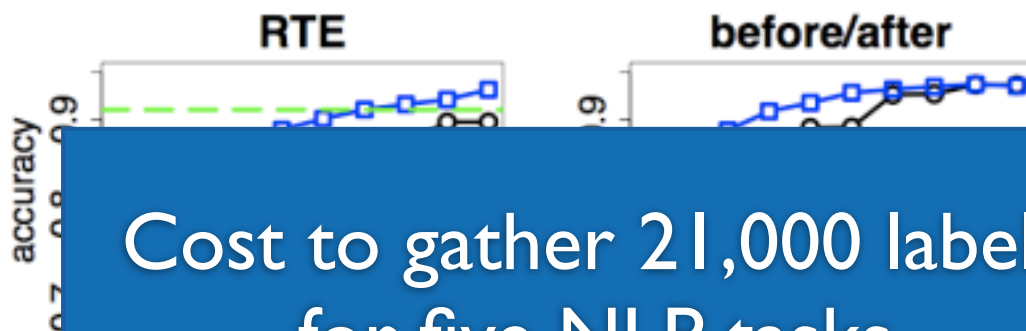
- Temporal Annotation

ran happens before *fell* in:
"The horse raced past the barn fell."

NLP Annotation



Combine non-expert judgments for high correlation with experts



Cost to gather 21,000 labels for five NLP tasks

Weight w/ small amount of expert data

\$ 25.82

Can Turkers Translate?

Urdu

1994 میں اس خطے میں ابتدائی انسانوں کی باقیات جو تقریباً 8 لاکھ سال پرانی مانی جاتی ہے، دریافت کی گئیں جنہیں ہومو اینٹی سیسر یعنی 'بانی انسان' کا نام دیا گیا۔

اس سے قبل 6 لاکھ پرانے انسان جنہیں سائنسی اصطلاح میں ہومو ہیڈلبرجینسس کہا جاتا ہے، اس خطے کے قدیم ترین رہائشی مانے جاتے تھے۔

آثار قدیمہ کے ماہرین کا کہنا ہے کہ انہیں ایسے شواہد ملے ہیں جن سے پتہ چلتا ہے کہ اس خطے کے لوگ ڈھلائی کیے ہوئے اوزار بھی

LDC Translation

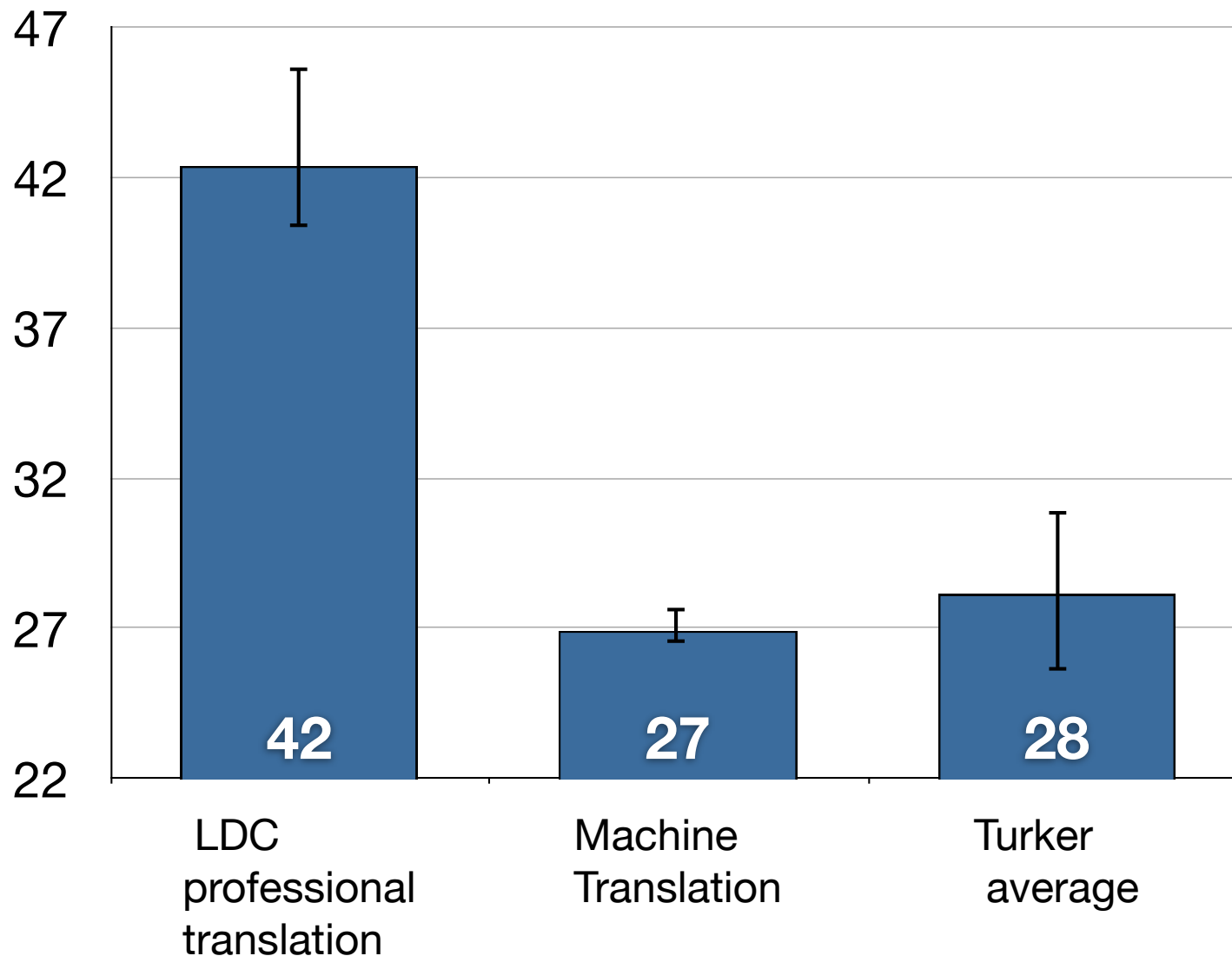
In 1994, the remains of early human beings who were believed to be eight hundred thousand years old were discovered who were given the name homo antecessor meaning the 'founder man'. Prior to this, the six hundred thousand years old man, called homo heidelebergensis in scientific terms, was believed to be the earliest resident of this area.

Turk Translation

In 1994, the remains of pre-historic man, which are believed to be 800,000 years old were discovered and they were named 'Home Antecessor' meaning 'The Founding Man'. Prior to that 6 lac years old humans, named as Homogenisens in scientific terms, were beleived to be the oldest dewellers of this area.

Professionals versus Turkers*

*Without Quality Control



Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mouses.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mouses had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that " It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mice.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mice had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that "It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
This research of American scientists came in front after experimenting on mice.	This research from the American Scientists have come up after the experiments on rats.	This research of American scientists was shown after many experiments on mice.	According to the American Scientist this research has come out after much experimentations on rats.
Experiments proved that mice on a lower calorie diet had comparatively less ability to fight the flu virus.	in has been proven from experiments that rats put on diet with less calories had less ability to resist the Flu virus.	It was proved by experiments the low calories eaters mice had low defending power for flue in ratio.	Experimentaions have proved that those rats on less calories diet have developed a tendency of not overcoming the flu virus.
research has proven this old myth wrong that its better to fast during fever.	Research disproved the old axiom that "It is better to fast during fever"	The research proved this old talk that decrease eating is useful in fever.	This Research has proved the very old saying wrong that it is good to starve while in fever.

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
--------------------------------------	---	---	--

Avoid		dieting		to		prevent	from	flu
Abstention	from	dieting	in order	to		avoid		Flu
Abstain		decrease eating	in order	to		escape		flue
quit		dieting		to	be	safer	from	flu

Avoiding dieting to prevent from flu	abstention from dieting in order to avoid Flu	Abstain from decrease eating in order to escape from flue	In order to be safer from flu quit dieting
--------------------------------------	---	---	--

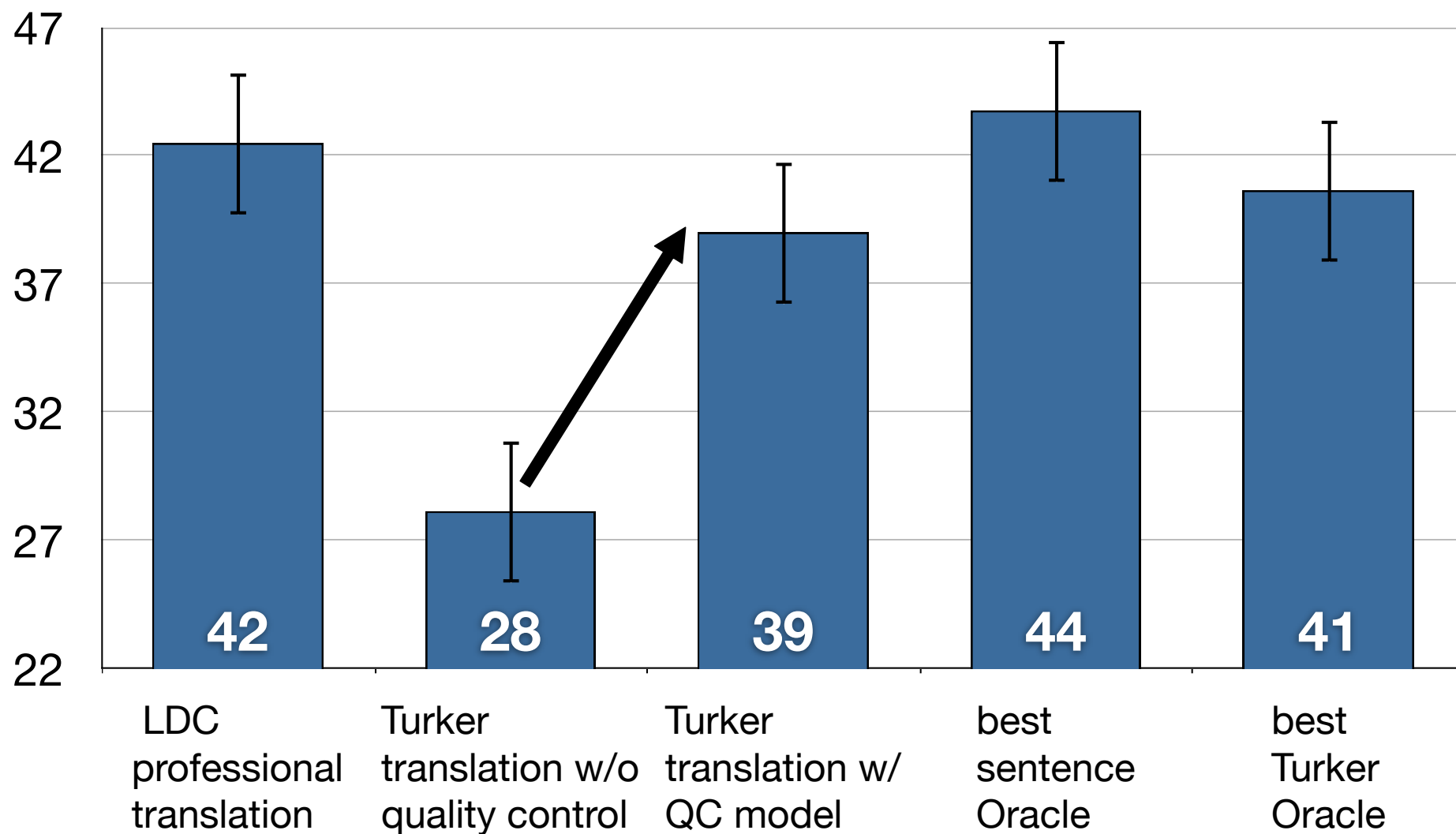
Avoid		dieting		to		prevent	from	flu
Abstention	from	dieting	in order	to		avoid		Flu
Abstain		decrease eating	in order	to		escape	from	flue
quit		dieting		to	be	safer	from	flu

Quality Control Model

- **Sentence features**
 - Language model probability
 - Ratio of source / target sentence lengths
 - Web n-gram match percentage
 - Translation edit rate to other translators
- **Turker features**
 - Aggregate of sentence feature scores
 - Self-reported language abilities
(Is native speaker? How long speaking?)
 - Worker location (Pakistan? India?)
- **Ranking features** (based on second pass vote)
- **Calibration feature** (Bleu against professionals)

Professional Quality from Non-Professionals

Full details in Zaidan and
Callison-Burch (ACL 2011a)

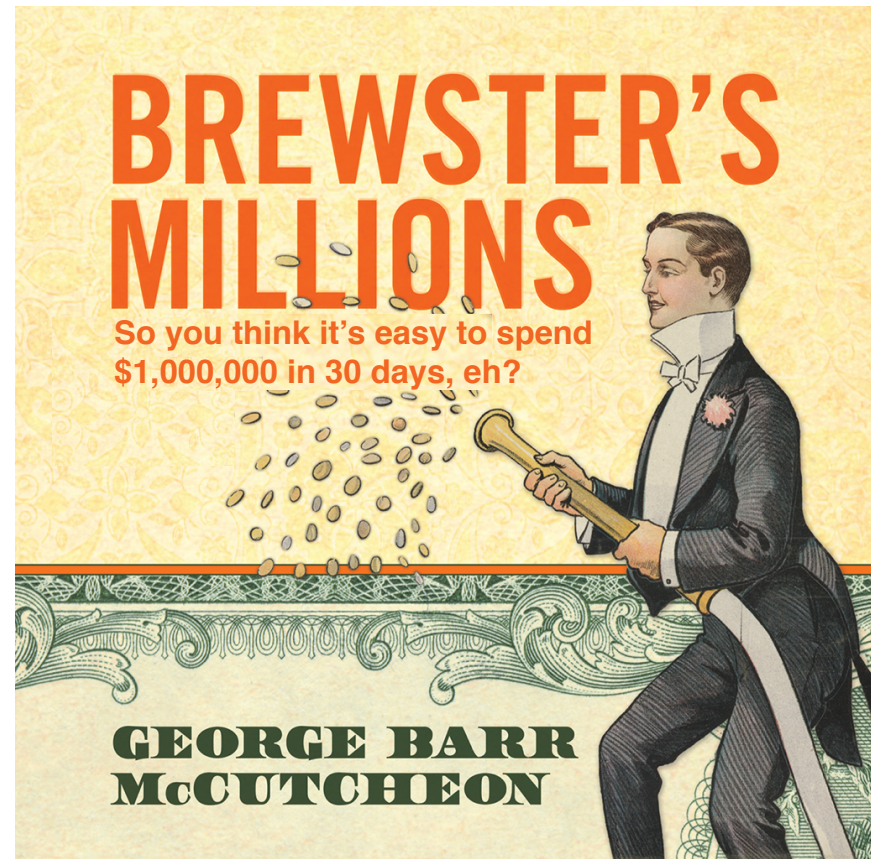


Does it scale? A friendly wager.

Surely it can't scale. I bet you can't spend my money fast enough.

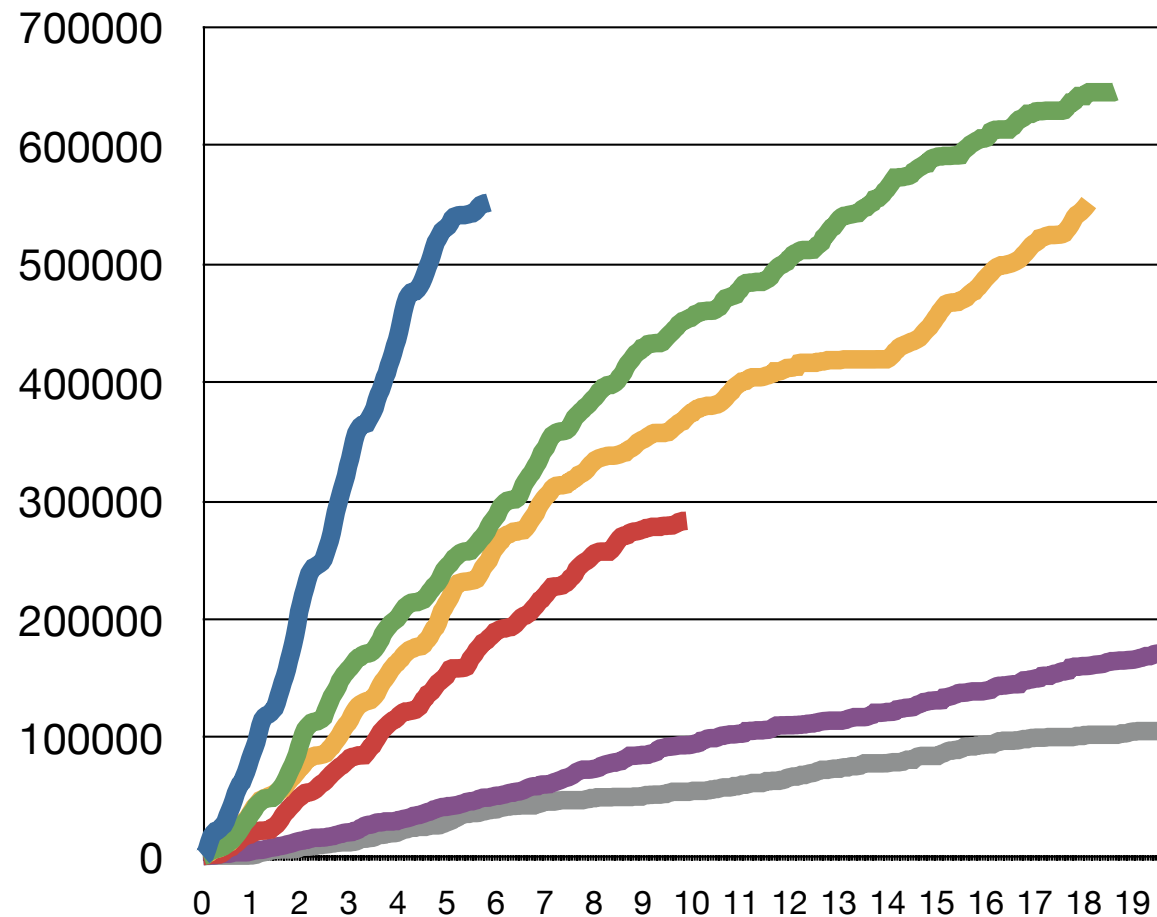
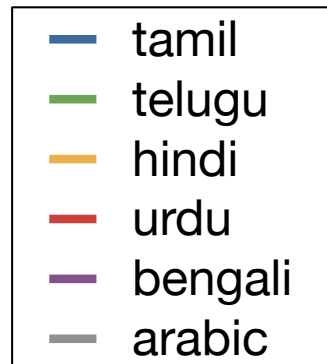


- Ken Church bet me \$50k that I couldn't spend \$50k on MTurk in two months



Fast, Scalable and Cheap

**Translation
Volume: total
English words
after k days**



MTurk translation is 10-60 times less expensive than professional translation

On-Demand Worldwide Workforce

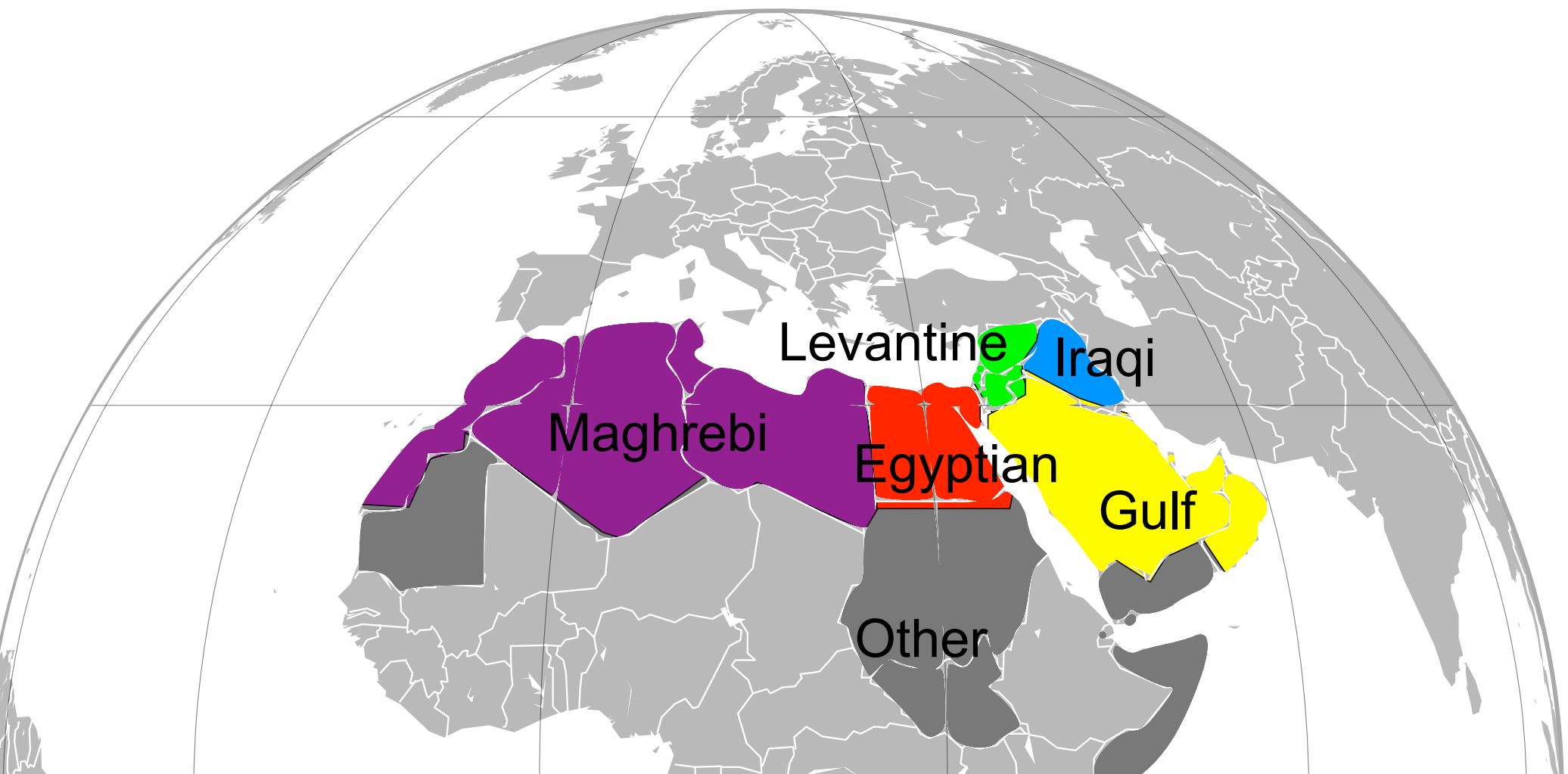
Hindi (352), Tamil (298), Malayalam (282), Spanish (215), Italian (157), Telugu (139), French (136), Malay (128), Kannada (126), Romanian (110), Portuguese (109), German (109), Arabic (106), Marathi (95), Tagalog (92), Bengali (92), Nepali (91), Urdu (89), Polish (89), Russian (86), Cebuano (83), Serbo-Croatian (78), Bulgarian (77), Bosnian (75),

Census of bilinguals on Mechanical Turk by translating 10,000 words in 100 languages into English

Gujarati (73), Croatian (72), Turkish (70), Serbian (69), Punjabi (68), Irish (68), Chinese (64), Macedonian (63), Czech (63), Dutch (60), Swedish (59), Danish (56), Swahili (51), Slovenian (49), Hungarian (48), Catalan (45), Slovak (40), Finnish (39), Indonesian (38), Hebrew (36), Vietnamese (35), Albanian (35), Norwegian (29), Ukrainian (27), Korean (26), Japanese

Machine Translation of Arabic Dialects

Arabic has **different varieties**. MSA is the standardized form but there are many **distinct** regional dialects.



Translating Dialects with MSA MT

MSA:

متى سنرى هذه الفئة من المجرمين تخضع للمحاكمة ؟

*mtY snrY h*h Alvlp mn Almjrmyn txDE llmHAkmp ?*

MT: When will we see this group of offenders subject to a trial ?

Levantine:

ايمتى رح نشوف هالفئة من المجرمين بتتحاكم ؟

AvmtY rH n\$wf hAl\$lp mn Almirmvn bttHAkm ?

MT: Aimity suggested Ncov Halclp Btaathakm of criminals ?

PT: Quando veremos esse grupo de criminosos serem julgados?

ES-EN: **Quando esse** group of criminals see **Serem julgados?**

Arabic Dialect in Online Comments

بلخادم: ما يجمع مصر والجزائر يجعل التقارب أكثر من ضروري

الخميس، 16 يونيو 2011 - 09:27



عبد العزيز بلخادم الممثل الشخصي للرئيس الجزائري

الجزائر (أ.ش.أ.)



أكد عبد العزيز بلخادم الممثل الشخصي للرئيس الجزائري والأمين العام لحزب جبهة التحرير الوطني ذات الأغلبية في البرلمان والحكومة، أن العلاقات بين مصر والجزائر قديمة ومتجذرة علاوة على مصالح مشتركة قوية تربط بين البلدين.

وقال بلخادم إنه بالتالي فإن ما حدث من "سحابات الصيف" عقب مباراة كرة القدم بين فريقى البلدين فى تصفيات كأس العالم عام 2009 لا يؤثر على طبيعة العلاقة والأخوة والروابط بين الشقيقين، مشددا على ضرورة العمل على رفع سقف التعاون بين البلدين على المستوى الحكومى والجهادى والمؤسساتى

Article body (MSA)

1 خليك فى حالك وابعد عنا

بواسطة: bebooo

بتاريخ: الخميس، 16 يونيو 2011 - 09:39

EGY

خليك فى حالك انت وابعدوا عنا وانتم اخر ناس تتكلم عن التقارب والعروبة

2 الجزائر إخوة أعزاء لنا وأهلاً ومرحباً بتقوية العلاقات معهم

بواسطة: سامي شاهين

بتاريخ: الخميس، 16 يونيو 2011 - 09:55

MSA

التقارب بين مصر والجزائر يعتبر حاجة ملحة تستدعيه الظروف الراهنة التي يمر بها الوطن العربي ، أهلاً ومرحباً بأي تقارب عربي عربي بين الأشقاء في الوطن والمصير ، يسقط الاستعمار الجديد المتمثل في أمريكا والنااتو ومن ورائة آل مهيون

3 هزلت

بواسطة: مواطن مصري بسيط

بتاريخ: الخميس، 16 يونيو 2011 - 09:55

EGY

هزلت وعيلت ومشتت سفينة نوح علي اليابس - الدنيا جري فيها ايه علشان نسمع دروس من امثالكم

4 تنبيه

بواسطة: علي

بتاريخ: الخميس، 16 يونيو 2011 - 09:57

MSA

انا الشعب الجزائري يعمل جاهدا من اجل السعي لتوحيد العالم العربي وكذلك السعي ورا طموحات اكبر وهي ، ان تكون كلمت العرب كلما واحدة امام العرب

5 أنت.....

بواسطة: الفيلسوف

بتاريخ: الخميس، 16 يونيو 2011 - 10:14

EGY

أنت تتكلم مين يا بلخادم كل الأنظمة فاسدة و تابعة لأمريكا و إسرائيل، إن أردتم فعلا حل مشاكل العرب في البيت العربي فما على الحكام إلا تكوين جمعية و الذهاب لمنتجع الحكام المخلوعين بالملكة السعودية وسيبوا الشعوب تبحث في قيام الولايات العربية المتحدة و عاصمتها القدس إنشاء الله (معلش حبيبي هو حلم نقطة يمكن يتحقق في يوم من الأيام من منا يعلم الغيب؟)

MSA

EGY

Crowdsourcing Dialect Training Data

Full details in Zaidan and
Callison-Burch (ACL 2011b)

- Labeled 142k comments gathered from online newspapers from Saudi Arabia, Jordan, Egypt
- 59% MSA, 41% dialect
- Trained classifiers with 85-91% accuracy
- Translated dialect-labeled segments
- 1.5M Arabic words / 2.2M English word
- Trained BBN's machine translation system

GLF

LEV

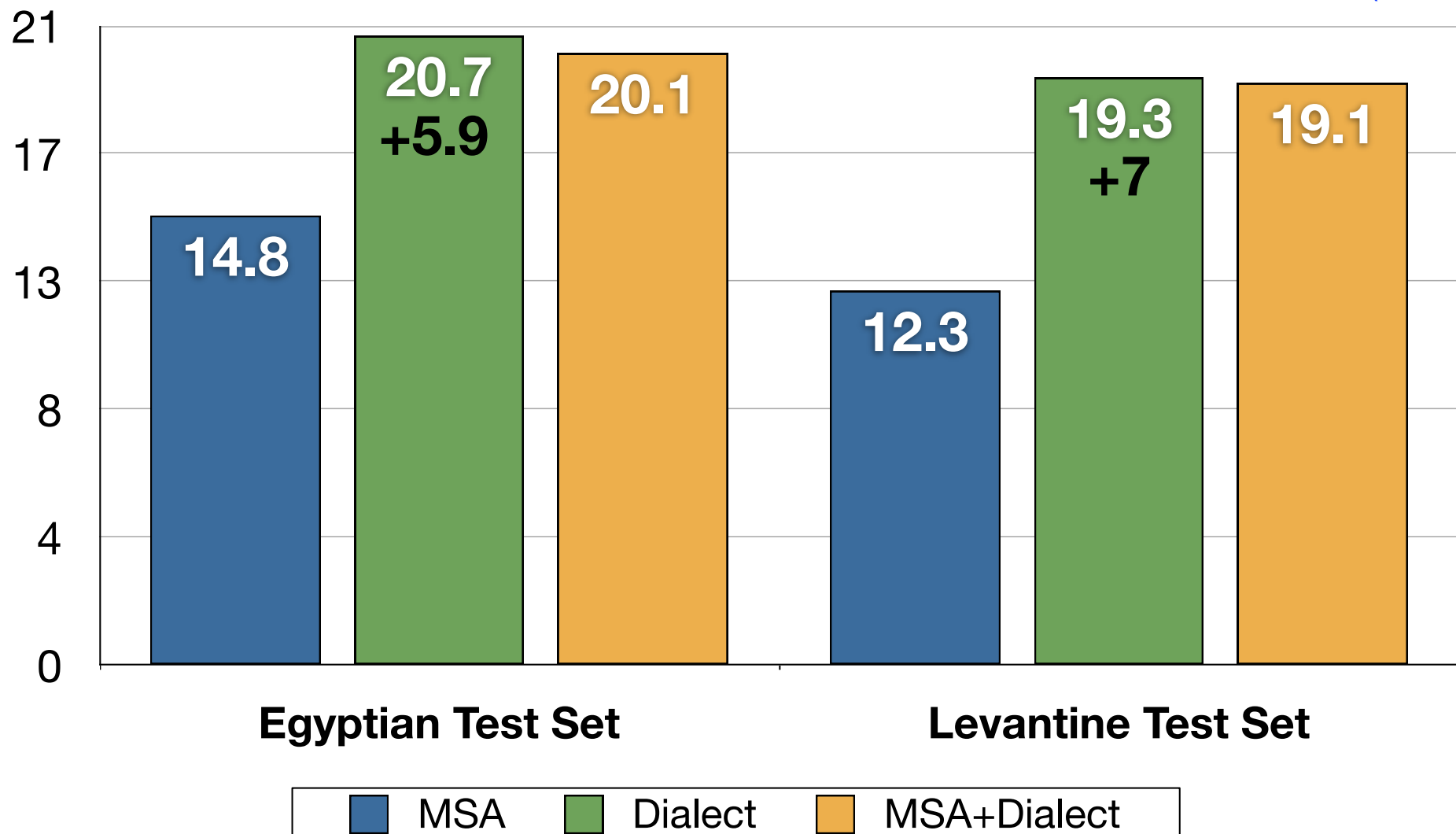
EGY

EGY

LEV

Dialect versus MSA

Full details in Zbib, Malchiodi, Devlin, Stallard, Matsoukas, Schwartz, Makhoul, Zaidan, & Callison-Burch (NAACL 2012)

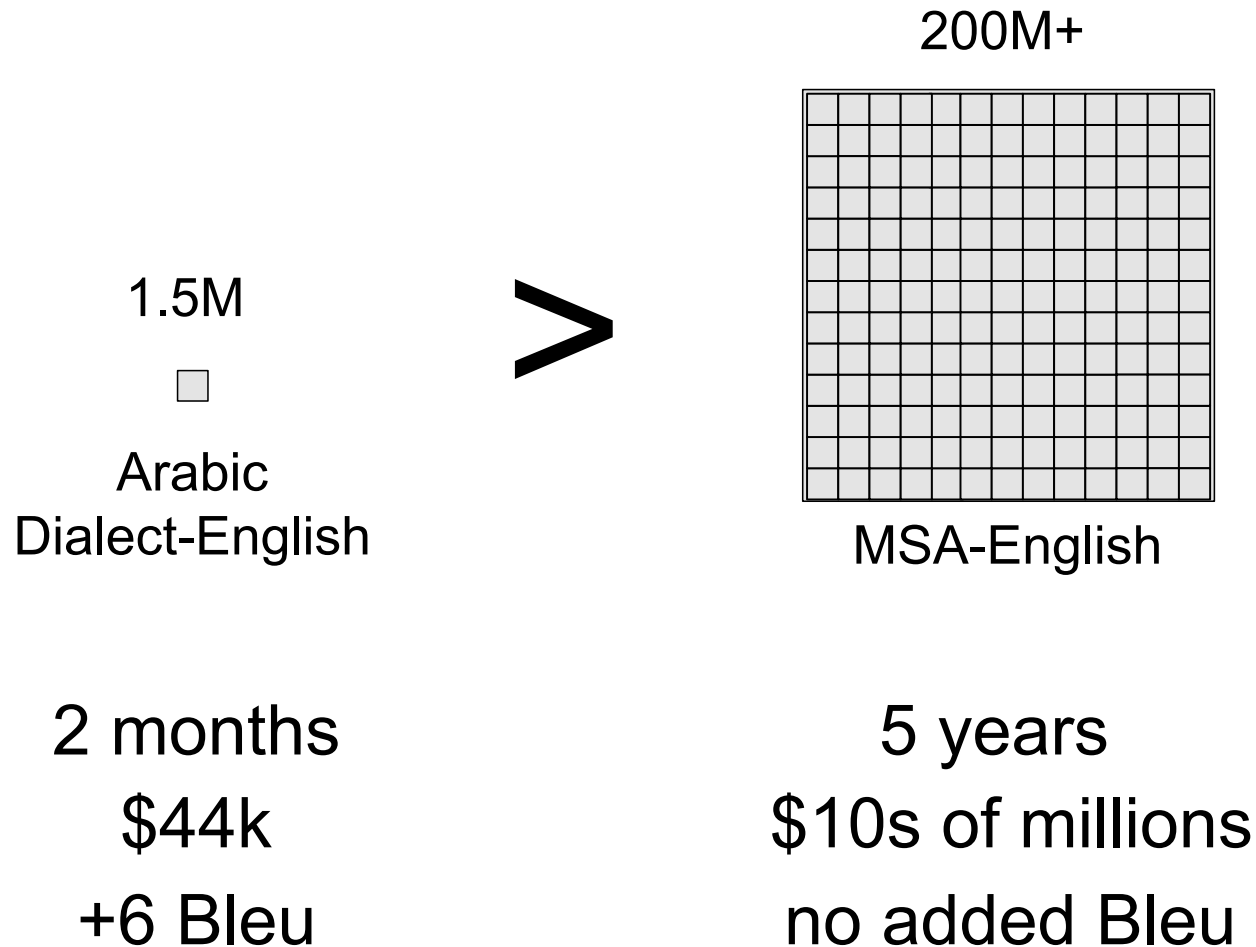


Examples of Dialect Translation

Dialect Input	MSA system	Dialect system	Reference
EGY انت بتعمل له اعلان ولا ايه ؟ !!	You are working for a declaration and not?	You are making the advertisement for him or what?	Are you promoting it or what?!!
EGY نفسي اطمئن عليه بعد ما شاف الصورة دي	Myself feel to see this image.	I wish to check on him after he saw this picture.	I want to be sure that he is fine after he saw the images.
LEV لهيك الجو كتييري كووول	God you the atmosphere.	This is why the weather is so cool	This is why the weather is so cool
LEV طول بالك عم نمزح	Do you think about a joke long.	Calm down we are kidding	Calm down, we are only kidding

Crowdsourcing Takeaways

For dialect translation



Crowdsourcing Takeaways

- Provides a fast, cheap, scalable way of collecting data for a huge range of languages
- Annotators are non-experts, but can do surprisingly complex tasks
 - Give clear, simple instructions & create small subtasks
- Quality control is paramount
 - Embed gold standard data
 - Exploit redundancy

Ominous alternate title (suggested by colleague):

“Why the third decade may
be the last for the LDC”

Prof. Callison-crowdsourced-doom-Burch



Individual researchers now have their own data production companies

My personal budget for MTurk is likely to be \$200k+ over the next 2 years. I expect to be able to create several LDC quality data sets.

- ✓ Parallel corpora for six Indian languages
- ✓ Arabic dialect ID dataset
- ✓ 1.5M word English-Arabic dialect corpus
- ✓ Bilingual dictionaries for ~100 languages
- ✓ English translation of 180 hrs of spoken Spanish
- Corpus annotating and correcting ESL errors
- Speech collection and transcription for new lang.

If individuals like me can do that...

- What role is left for the LDC?
- Will it start losing its contracts?
- How do you compete with agile outfits that can do the job for pennies on the dollar?
- Lots of re-duplication of effort
- Quality control is extremely difficult to get right

How does the LDC stay relevant?

- Become the CrowdFlower for language annotation
 - Create standardized annotation pipelines / quality controls / data formats for common NLP tasks

- NER
- POS tagging
- WSD
- Transcription
- Translation
- Summarization
- Sentiment analysis
- ESL error annotation
- Coreference resolution
- Image annotation
- MT evaluation
- Reading time experiment

How does the LDC stay relevant?

- Establish a trusted workforce for many languages
 - Verify that workers actually understand the languages that they purport to speak
 - Create long-running tasks to ensure that workers always have tasks in their language
 - These tasks can be used to train annotators
- Create a self-service website
 - Groups can upload their data
 - You charge a markup over labor to cover R&D / hosting
 - Give discounts to groups that allow their results to be distributed through LDC

How does the LDC stay relevant?

- Have professional annotators play different roles than they do now
 - Doing tasks that require considerable linguistic training
 - Creating controls for non-trusted / non-professional annotators
 - Decomposing the tasks and annotation guidelines into simple steps that lay people can complete
- Start using the power of crowdsourcing to your advantage

Thanks!

- Here's to another 20 great years!

More info in Omar Zaidan's PhD thesis "Crowdsourcing Annotation for Machine Learning in Natural Language Processing Tasks" and our publications. Thesis defense video at <http://vimeo.com/clsp/omar-zaidan-thesis-defense-talk>