New Approaches for Distributed Non-Expert Annotation and Collection at LDC

Stephanie Strassel (presenter), Ann Bies, Kira Griffitt, Xuansong Li, Justin Mott, Ann Sawyer, Zhiyi Song, Jennifer Tracey, Jonathan Wright
Introduction
Context for LDC Projects

- Linguistic resources for sponsored lang technology programs
- Multiple concurrent projects, tasks and languages
  - Amharic, Arabic, Bengali, English, Farsi, Hausa, Hindi, Hungarian, Indonesian, Mandarin, Russian, Somali, Spanish, Swahili, Tagalog, Tamil, Thai, Turkish, Twi, Uyghur, Uzbek, Vietnamese, Wolof, Yoruba, Zulu
  - Collection of conversational telephone speech, broadcast news and talk shows, amateur video, news text, blogs, Twitter, SMS/chat
  - Annotation ranging from simple
    - Orthographic Transcription, Language ID, Translation Quality Control
  - To complex
    - NP Chunking, Entities and Coref, Semantic Role Labeling, Belief/Sentiment
  - To expert
    - Syntactic Annotation (Treebanking), Lexicon Creation

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
The Workers

- Traditional model: hire staff and pay an hourly wage
  - Local (on site at least occasionally) with some Lx expertise
  - Intensive, lengthy hands-on training
  - Long-term commitment in both directions
- Scope, number, complexity of projects has steadily increased
- Emphasis has shifted toward lower resource languages
  - Short lead time on new languages or tasks
- Need more people than ever, but it’s harder to find them
  - Dozens or even hundreds of short-term workers at a time
- Remote, distributed work by non-experts is essential
Task Engineering
Decision Points

- What questions must be answered to complete the task?
- In what order should they be answered?
- What are the dependences among them?

- Decision Points inform all aspects of task design
  - Training approach including guidelines
  - Workflow
  - User interface design
  - Annotator testing and quality control
BOLT Information Retrieval Query Assessment

- Given a natural language English query, BOLT system automatically processes multi-lingual informal web text corpora and returns answers in English

**Query:** What are the influences of the Euro financial crisis on China?

**System response:** Due to the spread of the European debt crisis has intensified, the US economic recovery is sluggish, further deterioration of the external environment in the development of the Asian economies.

- Poor annotator consistency given a single, simple question “Is the response relevant?”
  - Impact of (often poor) automatic translation
  - Varying degrees of leniency
Unpacking Decision Points

Is the response relevant? → Up to five decision points

Q1: Can you answer relevance questions based on this English response alone, without consulting the original source text?
   YES.
   NO, because the translation is incomprehensible.
   NO, because I need to see the source text to resolve pronouns, get more context and/or clarify the translation.

Q3A: Is the response informative?
   YES, it adds information beyond restating the query.
   NO, it does not add new information.

Q4: Does the response contain any relevant information?
   YES, it addresses the query.
   NO, it does not address the query.

Q5: Were any of your judgments made generously?
   YES, one or more of the judgments was made generously.
   NO, all of the judgments were made without being generous.

Significant increase in annotator consistency using decision points model
No significant decrease in annotator efficiency

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
Decision Trees

Q1: Can I make a relevance judgments from English citation?

- Yes
  - Select YES
- No
  - Select NO because I need to see the source text
- Can't tell
  - Select NO because the translation is incomprehensible
GUI design closely follows decision points

- Dynamic display and skip logic to walk worker through the decision making process
- Prevents illogical judgments, avoids wasting time on irrelevant decisions

Welcome, strasselcmn2test10
Please answer the questions below to set up your call.

Is this a Skype call?
- Yes
- No
GUI Design Factors

- GUI design has a major impact on training time, efficiency, accuracy, even staff retention
- Consider user’s computing background
- Are there assumptions about browser and OS requirements
- What are the network requirements
- Screen real estate
  - May be working on laptop or tablet
- Avoid wasted movements
  - Unnecessary scrolling
  - Mouse vs. keyboard
- Use standard browser shortcuts
- Constrain behavior (decision points)
- Constrain possible answers based on prior decisions
- Provide informative error reports/validation
Workflow Considerations

- Traditional: One annotator makes all task decisions
- Decision points present opportunity to distribute work across team of differently skilled workers

<table>
<thead>
<tr>
<th>Full Entity Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Is there an entity in this passage?</td>
</tr>
<tr>
<td>✓ Is it a specific person, place, organization, location or facility?</td>
</tr>
<tr>
<td>✓ Is it a name, nominal or pronoun?</td>
</tr>
<tr>
<td>✗ What is its extent (full NP)?</td>
</tr>
<tr>
<td>✗ If nom, what is its head?</td>
</tr>
<tr>
<td>✓ What is its type?</td>
</tr>
<tr>
<td>✓ Can it be linked to another entity?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity Annotator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is there an entity in this passage?</td>
</tr>
<tr>
<td>Is it a specific person, place, organization, location or facility?</td>
</tr>
<tr>
<td>Is it a name, nominal or pronoun?</td>
</tr>
<tr>
<td>Intuitive extent</td>
</tr>
<tr>
<td>What is its type?</td>
</tr>
<tr>
<td>Can it be linked to another entity?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NP Chunking Annotator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label the full extent and head of this nominal mention</td>
</tr>
</tbody>
</table>
Many Possible Workflows

Workflow (and deliverable product) depends on many factors

- Cost, quality, timeline, data volume tradeoffs
- How will data be used?
- What is human consistency baseline?
Training
Goal: dramatically advance the state of computational linguistics and human language technology to enable rapid, low-cost development of capabilities for low-resource languages

- Improved situational awareness based on information from any language, in support of emergent missions such as humanitarian assistance/disaster relief, peacekeeping or infectious disease response

- LDC is building data packs for approximately 3 dozen LRLs
  - E.g. Amharic, Hausa, Somali, Uyghur, Zulu

- Remote native speakers collect relevant data (news, informal web text, Twitter etc) and perform linguistic annotation ranging from simple to very challenging
  - Named entities, full entity with coref, NP chunking, POS labeling, semantic annotation, etc.
Applicant Screening

- Applicants directed to online survey to assess...
- Eligibility for position
  - Paperwork, schedule, access to computer & internet
- Language use
  - How often do you …
    - Read, speak, Google, use social media etc
  - Did you grow up speaking …
    - With family? In school? With friends?
- Survey is automatically “scored” and personalized (stock) reply sent
  - Standards vary depending on language, applicant pool
- Promising applicants automatically invited for aptitude test and/or Skype interview
Aptitude Testing

Maghrebi Arabic

Iraqi Arabic

A related language (e.g. Tigrinya)
Self Directed Preliminary Training

- **Primary goals**
  - Introduce core concepts and terminology
  - Assess aptitude for specific linguistic annotation tasks
  - Give annotator a taste of the work they’ll be asked to do
- **Multiple units, multiple short modules per unit**
  - No more than 1-2 hours per unit
- **Logical progression based on task ordering, module complexity**
- **Narrated presentations for each module, followed by a quiz**
  - Must complete the unit within 24 (48, 72) hours after starting
  - Must achieve acceptable score before moving on
  - Two strikes and you’re out (for all tasks that build on this module)
- **Can scale to effectively unlimited number of candidates**
  - Labor-intensive advanced training limited to suitable workers
Chapter 18: Name

The next video explains what the term **name** means for us - what kinds of things are the types of names we care about for annotation.

Please watch the video below (as many times as you want to) then click **Next** to answer questions.

Names can be in different forms

- A name can be a full name, a nickname or alias, or an acronym.
- Philadelphia = City of Brotherly Love
- the United States of America = the United States
- USA = U.S. = America = American
- William Jefferson Clinton = Bill Clinton = B. C.
- UNICEF
- Médecins Sans Frontières = Doctors Without Borders = MSF
## Identify Decision Points

- What questions must be answered to complete the task?
- In what order should they be answered?
- What are the dependences among them?

### Data Scouting
- Is the document suitable for collection?
- Is the document relevant to the domain?
- What is the genre?
- What is the type of incident?
- What is the name of the incident?
- Is there a Wikipedia page about the disaster?
- Does the document contain any eyewitness account(s) of the incident or its aftermath, impact, recovery efforts, etc.?
- How specific/local is the information about the incident?
- What topics are discussed?

### Simple NE
- Is there a name in this (sentence, paragraph, ...)?
- Is it the name of a specific person, place, organization or facility?
- What is its extent?
- What is its type?

### Simple Semantic Annotation
- Does the passage describe an Act or State?
- Does the passage describe a disaster-relevant Act?
- Select the most important word or phrase that expresses the disaster-relevant Act
- Select the word or phrase that corresponds to the Agent(s) of this Act
- Select the word or phrase that corresponds to the Patient(s) of this Act
- etc.
Identify Core Concepts

- What core concepts must be mastered in order to successfully navigate each decision point?

<table>
<thead>
<tr>
<th>Data Scouting</th>
<th>Simple NE</th>
<th>Full Entity</th>
<th>NP Chunking</th>
<th>SSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much of the document needs to be in the target language?</td>
<td>What is a name</td>
<td>What is an entity</td>
<td>Key Concepts</td>
<td>What counts as an Act?</td>
</tr>
<tr>
<td>How to locate dates/times within text</td>
<td>What is an entity</td>
<td>What are specific entities</td>
<td>Tagging and taggability (general)</td>
<td>What counts as a physical Act?</td>
</tr>
<tr>
<td>How to locate names of disasters within text</td>
<td>What is name extent</td>
<td>What is an entity mention</td>
<td>Extents (general)</td>
<td>What counts as a State?</td>
</tr>
<tr>
<td>How to locate names of organizations within text</td>
<td>Entity typing</td>
<td>what is entity mention extent</td>
<td>Grammaticality</td>
<td>What counts as disaster-relevant State?</td>
</tr>
<tr>
<td>How to locate names of people within text</td>
<td>PER entity type</td>
<td>Name extents</td>
<td>Part of Speech (general)</td>
<td>What is a basic word?</td>
</tr>
<tr>
<td>How to locate names of places within text</td>
<td>ORG entity type</td>
<td>Nominal extents</td>
<td>Part of Speech: Noun</td>
<td>What is a phrase?</td>
</tr>
<tr>
<td>How to search for the web using relevant keywords</td>
<td>GPE entity type</td>
<td>Nominal heads</td>
<td>Part of Speech: Pronoun</td>
<td>What is an Agent?</td>
</tr>
<tr>
<td>Etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Identify Shared Concepts

- **Is the same core concept relevant to multiple tasks?**

<table>
<thead>
<tr>
<th>Data Scouting</th>
<th>Simple NE</th>
<th>Full Entity</th>
<th>NP Chunking</th>
<th>SSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much of the document needs to be in the target language?</td>
<td>What is a name</td>
<td>What is an entity</td>
<td>Key Concepts</td>
<td>What counts as an Act?</td>
</tr>
<tr>
<td>How to locate dates/times within text</td>
<td>What is an entity</td>
<td>What are specific entities</td>
<td>Tagging and taggability (general)</td>
<td>What counts as a physical Act?</td>
</tr>
<tr>
<td>How to locate names of disasters within text</td>
<td>What is name extent</td>
<td>What is an entity mention</td>
<td>Extents (general)</td>
<td>What counts as a State?</td>
</tr>
<tr>
<td>How to locate names of organizations within text</td>
<td>Entity typing</td>
<td>What is entity mention extent</td>
<td>Grammaticality</td>
<td>What counts as disaster-relevant State?</td>
</tr>
<tr>
<td>How to locate names of people within text</td>
<td>PER entity type</td>
<td>Name extents</td>
<td>Part of Speech (general)</td>
<td>What is a basic word?</td>
</tr>
<tr>
<td>How to locate names of places within text</td>
<td>ORG entity type</td>
<td>Nominal extents</td>
<td>Part of Speech: Noun</td>
<td>What is a phrase?</td>
</tr>
<tr>
<td>How to search for the web using relevant keywords</td>
<td>GPE entity type</td>
<td>Nominal heads</td>
<td>Part of Speech: Pronoun</td>
<td>What is an Agent?</td>
</tr>
</tbody>
</table>

Etc.
### Identify Shared Concepts

- Is the same core concept relevant to multiple tasks?

<table>
<thead>
<tr>
<th>Data Scouting</th>
<th>Simple NE</th>
<th>Full Entity</th>
<th>NP Chunking</th>
<th>SSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much of the document needs to be</td>
<td>What is a name</td>
<td>What is an entity</td>
<td>Key Concepts</td>
<td>What counts as an Act?</td>
</tr>
<tr>
<td>in the target language?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How to locate dates/times within text</td>
<td>What is an entity</td>
<td>What are specific</td>
<td>Tagging and</td>
<td>What counts as a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>entities</td>
<td>taggability (general)</td>
<td>physical Act?</td>
</tr>
<tr>
<td>How to locate names of disasters within</td>
<td>What is name</td>
<td>What is an entity</td>
<td>Extents (general)</td>
<td>What counts as a</td>
</tr>
<tr>
<td>text</td>
<td>extent</td>
<td>mention</td>
<td></td>
<td>State?</td>
</tr>
<tr>
<td>How to locate names of organizations</td>
<td>Entity typing</td>
<td>What is entity</td>
<td>Grammaticality</td>
<td>What counts as</td>
</tr>
<tr>
<td>within text</td>
<td></td>
<td>mention extent</td>
<td></td>
<td>disaster-relevant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>State?</td>
</tr>
<tr>
<td>How to locate names of people within</td>
<td>PER entity type</td>
<td>Name extents</td>
<td>Part of Speech</td>
<td>What is a basic</td>
</tr>
<tr>
<td>text</td>
<td></td>
<td></td>
<td>(general)</td>
<td>word?</td>
</tr>
<tr>
<td>How to locate names of places within</td>
<td>ORG entity type</td>
<td>Nominal extents</td>
<td>Part of Speech:</td>
<td>What is a phrase?</td>
</tr>
<tr>
<td>text</td>
<td></td>
<td></td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>How to search for the web using</td>
<td>GPE entity type</td>
<td>Nominal heads</td>
<td>Part of Speech:</td>
<td></td>
</tr>
<tr>
<td>relevant keywords</td>
<td></td>
<td></td>
<td>Pronoun</td>
<td></td>
</tr>
<tr>
<td>Etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Identify Shared Concepts

### Is the same core concept relevant to multiple tasks?

<table>
<thead>
<tr>
<th>Data Scouting</th>
<th>Simple NE</th>
<th>Full Entity</th>
<th>NP Chunking</th>
<th>SSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much of the document needs to be in the target language?</td>
<td>What is a name</td>
<td>What is an entity</td>
<td>Key Concepts</td>
<td>What counts as an Act?</td>
</tr>
<tr>
<td>How to locate dates/times within text</td>
<td>What is an entity</td>
<td>What are specific entities</td>
<td>Tagging and taggability (general)</td>
<td>What counts as a physical Act?</td>
</tr>
<tr>
<td>How to locate names of disasters within text</td>
<td>What is name extent</td>
<td>What is an entity mention</td>
<td>Extents (general)</td>
<td>What counts as a State?</td>
</tr>
<tr>
<td>How to locate names of organizations within text</td>
<td>Entity typing</td>
<td>What is entity mention extent</td>
<td>Grammaticality</td>
<td>What counts as disaster-relevant State?</td>
</tr>
<tr>
<td>How to locate names of people within text</td>
<td>PER entity type</td>
<td>Name extents</td>
<td>Part of Speech (general)</td>
<td>What is a basic word?</td>
</tr>
<tr>
<td>How to locate names of places within text</td>
<td>ORG entity type</td>
<td>Nominal extents</td>
<td>Part of Speech: Noun</td>
<td>What is a phrase?</td>
</tr>
<tr>
<td>How to search for the web using relevant keywords</td>
<td>GPE entity type</td>
<td>Nominal heads</td>
<td>Part of Speech: Pronoun</td>
<td>What is an Agent?</td>
</tr>
<tr>
<td>Etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
### One (Set of Related) Core Concept(s) Per Ordered Module

<table>
<thead>
<tr>
<th>Annotation Basics</th>
<th>Ling Basics</th>
<th>Ling Task Branch 1</th>
<th>Ling Task Branch 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Annotation</td>
<td>What Counts as a Language?</td>
<td>Genre</td>
<td>Noun</td>
</tr>
<tr>
<td>2 Units of Annotation</td>
<td>Meaning</td>
<td>Domain</td>
<td>Pronoun</td>
</tr>
<tr>
<td>3 Annotation Task</td>
<td>Ambiguity</td>
<td>Topic</td>
<td>Constituency</td>
</tr>
<tr>
<td>4 Annotation Suitability</td>
<td>Token and Word</td>
<td>Name</td>
<td>Nesting</td>
</tr>
<tr>
<td>5 Tag/Label</td>
<td>Part-of-Speech Basics</td>
<td>Noun Phrases</td>
<td>Coreference</td>
</tr>
<tr>
<td>6 Tagging/Labeling</td>
<td>Structure</td>
<td>Entity: PER, ORG, GPE, LOC</td>
<td></td>
</tr>
<tr>
<td>7 Extents</td>
<td>Grammaticality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Annotation Rules</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Annotation Tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Annotation Pipeline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Quality Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Each Language Is Different</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Task Specific Training

- Formal guidelines
  - Organized around decision points
  - As short as possible; minimize jargon; use diagrams
  - Intuitive basic principles plus a small number of rules of thumb
    - When in doubt, do it this way
  - Many naturally-occurring examples in language, genre, domain
  - Include glossary, FAQ
-Narrated presentations or videos explaining task and working through longer and/or more challenging examples
-Narrated presentations or videos showing how to use GUI
-One or more practice kits to get the hang of the task and GUI
-Self-scoring practice test(s)
-Final test resembles actual task
  - Scoring requirements depend on task, applicant pool

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
Ongoing Training and Testing

- Periodic insertion of “gold standard” work assignments
  - Can be helpful to use English translations of non-English data when language expertise not available
- Periodic insertion of “challenge” assignments to keep annotators attentive, especially for tedious, repetitive tasks
- Double-blind dual annotation of some data to monitor consistency and identify training needs
- “Shepherding” for high-value workers who are struggling
- Group interaction can be useful for consistency, morale
  - Email lists, video hangouts, chat
  - Interactive Q&A
  - Virtual team meetings

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
Incentives
Individual Incentives

- Hourly wage for long term, multi-skilled staff
- Pay by unit of work completed for distributed, short-term workers
- Graduated payments
  - Skills/experience: Training vs. junior vs. senior
  - Difficulty of tasking: topic scouting vs. transcription vs. lexicon development – but also noisy vs. non-noisy calls
- Lotteries
- Bonuses to address particular need
  - Completion of task
  - Data volume
  - Speed
  - Quality
- Combine as needed: e.g. Graduated + completion + speed
  - Get paid $2 per non-noisy call, $3 per noisy call
  - $100 after completing 10 calls
  - $50 if you complete all calls within 2 weeks of enrolling

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
Team Incentives

- **Competition and Collaboration**
- If team reaches goal, every individual on the team earns some bonus
  - *If your language team reaches the goal of completing 50 new 1P kits between midnight tonight and 11:59 PM Thursday, then anyone who works a minimum of 15 hours between midnight tonight and 11:59 PM Thursday will earn 5 extra "free" hours*
- If Team A beats team B, Team A gets a reward
- (Perception of) scarcity of work can also be a strong incentive
  - *Get ‘em while they’re hot! Supplies are limited!*
- Reporting progress, performance (quality, efficiency)
  - Against goal, against team, against self
Different Populations May Require Different Approaches

- BOLT: Large collection of SMS/Chat data in 3 languages
- Online recruiting methods worked well and quickly for English
- Chinese required greater reliance on personal contacts, social networks
- Egyptian challenges: fewer speakers, smaller web presence, more participant reluctance, more participant technical barriers

<table>
<thead>
<tr>
<th>Approach</th>
<th>Egyptian</th>
<th>Chinese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-country and US-based recruiters</td>
<td>9</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Recruiting/participant care hours</td>
<td>&gt;1000</td>
<td>300</td>
<td>250</td>
</tr>
<tr>
<td>Online advertising</td>
<td>92 cities</td>
<td>92 cities</td>
<td>92 cities</td>
</tr>
<tr>
<td>Social networking sites</td>
<td>60</td>
<td>500</td>
<td>200</td>
</tr>
<tr>
<td>Reach out to community orgs</td>
<td>2000</td>
<td>3000</td>
<td>4500</td>
</tr>
<tr>
<td>Fliers, postings</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Word of mouth, friend-of-a-friend/recruiter</td>
<td>300</td>
<td>1000</td>
<td>300</td>
</tr>
<tr>
<td>Contact former collection participants</td>
<td>300</td>
<td>150</td>
<td>1500</td>
</tr>
</tbody>
</table>
Egyptian Collection Challenges

<table>
<thead>
<tr>
<th>What we tried</th>
<th>What we learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meet with numerous cultural, social, industrial and academic organizations in Egypt and US to discuss possible partnership</td>
<td>Potential institutional partners were reluctant to collaborate given perceived risk in current political/social climate</td>
</tr>
<tr>
<td>Emphasize extensive privacy protections, UPenn reputation and oversight of collection, legitimacy of collection for language-related research</td>
<td>Despite careful explanations, most potential participants in Egypt remained suspicious of motives for collection</td>
</tr>
</tbody>
</table>
| Employ young, tech-savvy recruiters with wide social networks in Egypt and among US Egyptians | Technical barriers were higher for Egyptian participants  
  • Required more back-end development to simplify user experience  
  • More direct hand-holding |
| Initially, compensate Egyptians (in Egypt) at same rate as English, Chinese participants (in US); incrementally increase compensation rates and offer bonuses | Even with significantly increased compensation, Egyptians reluctant to sign up due to greater perception of personal risk compared to other participants |
LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
Collection Progress Over Time

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
Collection Results

- Large collection of SMS/Chat data in 3 languages
- Online recruiting methods worked well and quickly for English
- Chinese required greater reliance on personal contacts, social networks
- Egyptian challenges: fewer speakers, smaller web presence, more participant reluctance, more participant technical barriers

<table>
<thead>
<tr>
<th>Lang</th>
<th>Enrolled Users</th>
<th>Collected Messages</th>
<th>Collected Words</th>
<th>Accept Rate</th>
<th>Recruit Hours/Partcipnt</th>
<th>Avg Words/Msg</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>arz</td>
<td>46</td>
<td>161000</td>
<td>690000</td>
<td>69%</td>
<td>21.7</td>
<td>4</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>cmn</td>
<td>118</td>
<td>535000</td>
<td>3700000</td>
<td>70%</td>
<td>2.5</td>
<td>7</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>eng</td>
<td>275</td>
<td>309000</td>
<td>3300000</td>
<td>78%</td>
<td>0.9</td>
<td>11</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

* a few users; ** several users; *** many users
Game-Based Incentives

- Especially effective when the work is monotonous or stressful
- TDT Project required millions of topic relevance: tens of thousands of stories against hundreds of topics
  - Tedious (repetitive news) and stressful (grim news)
  - Pop quizzes and Story of the Day
- HAVIC Project required searching for hundreds of thousands of topic-relevant video clips
  - Need for more variety and atypical videos
    - Big boost in productivity and variety
    - Videos about popcorn, videos about popping balloons, popping bubble wrap, popping bubble gum, videos with pop music, with mom & pop restaurants, with pop goes the weasel, with pop-up books, and more

LDC Workshop on Crowdsourcing and Novel Incentives in Language Resource Creation, October 3-4, 2016
Non-Financial Incentives

- Work is meaningful
  - A good cause
  - Supports my language or culture
- Work is fun/interesting
  - Meet people
  - Free therapy (through sociolx interviews!)
  - Language geek
- Build skills
  - Stepping stone for career
  - Prestigious employer on resume
- Recognition in lieu of financial incentive
  - Job title (Junior \(\rightarrow\) Senior)
  - Certification
  - Public thank you

Researchers at Penn's Linguistic Data Consortium are starting a new project called SIREN to build linguistic resources for dozens of languages, as part of a larger effort to create language technology that can be utilized in disaster relief situations. You will work with project staff to identify appropriate texts in your language and label them for specific linguistic phenomena. The resources you help to develop will stimulate research in natural language processing technology for many new languages. This work could lead to significant advances in machine translation and related technologies that response teams rely on to provide better support during and after a natural disaster.
Conclusion
Response to growing need in sponsored programs to rapidly hire, train and incentivize teams of non-expert native speaker consultants (and human subjects) in dozens of languages.

Decision points inform all aspects of task design and present opportunity to:
- Reconfigure traditional workflow for non-experts
- Simplify training and better align it with worker skills
- Design GUI for efficient, directed annotation

Even with paid (traditional or contract) workers, novel incentives can focus effort on most important problems and make workers more engaged and productive.