Crowdsourcing for MT

João Graça
joao@unbabel.com
75% Does not speak English
Translation market

$34B

Existing Market

Traditional media

Customer Service E-mails

User Generated Content

Newsletters

Product Reviews

UNTAPPED MARKET
Available Solutions

Lack of fast, affordable translation with human quality

Machine Translation
- Affordable
- Fast
- Quality not good enough

Professional Translation
- Expensive
- Slow
This is question, English is faulty. Thank computer to translate to help. SORRY!!!!!

At often, the goat-time install a error is vomit.

How many times like the wind, a pole, and the dragon?

This insult to father’s stones?

Please apologize for your stupidity. There are a many thank you
Use case: customer support

Average Ticket

- 150 Words
- $15
Our vision
Crowd-post-editing

Scratch

Starting Quality

Perfect MT

Post-edited

Cost

Professional

Expertise

Monolingual Speaker

Crowd-sourced

Cost
Post Editing
Reduce Translation Cost
Simple Post-Editing (PE)

The editors are simply asked to correct the initial MT output, with no further interaction

- Nowadays, common practice in most translation agencies

Other early CAT tools:
- **Google Translator Toolkit** (Galvez and Bhansali, 2009)
- **WikiBabel project** (Kumaran et al, 2008).
CAT Tools

- Translation Spell checkers
- Online dictionaries and thesauri
- Translation memory (TM)
- Monolingual and bilingual concordances
- Terminology databases/glossaries
The apartment is free for your required dates.

O apartamento é livre para as datas desejadas.
The apartment is free for your required dates.
Translation Options

- Most machine translation errors are due to poor translation choices.
- Options can be computed using:
  - The raw options generated from MT server while decoding
  - The search graph
  - System Combination
Prediction MT

- Tighter integration of MT + human translation
- Prediction model that interactively suggests translations to the editors (i.e., auto completes), according to their previous translation decisions.
- If the user doesn’t accept a suggestion, new suggestions are generated.
- TransType (Langlais et al., 2000), Caitra (Koehn, 2009), TT2 project (Barrachina et al., 2009), CasMaCat, Green et al. 2014
Main Results

- **Speed:** Reduction in translation time
  - 18% and 34% (Flournoy and Duran, 2009; Guerberof, 2009; Federico et al. 2012) and can be as high as 43% (Plitt and Masselot, 2010)

- **Quality:** Lower Error Rate
  - Both non-professionals (Koehn, 2009) and professional translators (Plitt and Masselot, 2010; Green et al., 2013) are generally faster and produce better translations when post-editing
Crowdsourcing

Google Translate

“In a given day we translate roughly as much text as you’d find in 1 million books. To put it another way: what all the professional human translators in the world produce in a year, our system translates in roughly a single day”

Franz Och,
http://googleblog.blogspot.it/2012/04/breaking-down-language-barriersix-years.html
Crowdsourcing

Professional Translators

Crowd

Translators

Increase the number of existing translators (SCALE)
<table>
<thead>
<tr>
<th>Crowd Type</th>
<th>Represented by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional Translators</td>
<td>![Image]</td>
</tr>
<tr>
<td>Bilinguals</td>
<td>![Image]</td>
</tr>
<tr>
<td>Monolinguals</td>
<td>![Image]</td>
</tr>
</tbody>
</table>
Combine Crowd

• Marketplace model:

• Assign one text to a translator selected from the crowd
Combine Crowd

- Redundant model:
  - Assign same text to multiple translators
  - Combine different translations to maximize quality
Combine Crowd

- Chain model:
  - Assign text to a single translator
  - Have another translator reviewing previous work
Unbabel’s Goal

Affordable Human Quality Translation at Scale
Unbabel Pipeline

Customer submits a text

Machine translation

Translated text is divided into several smaller tasks

Quality check

Not good enough

Unbabel Community

OK

Finished tasks are combined again

Customer gets the translated text
Unbabel Pipeline
1 step - Machine Translation

MT Router
Unbabel Pipeline

2 step - Map step
Unbabel Pipeline
Dynamic Chaining

Text → MT → Editor 1 → Editor N

Confidence
Task Selection

• Route appropriate task for editor
• Check editor profile
• Big impact on quality and speed
Unbabel Pipeline
Mobile Editing experience

OPTIONS

Spellcheck
Unbabel Pipeline

3 step - Reduce step

- Reduce Step
- Spell check
- Inconsistencies (word alignments)
- If necessary add a gold editor
Unbabel Community

- Currently around 8,000 users
- Covering 756 language pairs
- 86 paid language pairs
Unbabel Community

- **Testing Phase**: 7872 Editors
  - Editors are tested when they sign up

- **Free Content**: 4740 Editors
  - Editors get ratings for the tasks

- **Paid Content**: 1268 Paid Editors
  - The best editors have access to paid content

- **Expert**: Specialization layers will grow with time
Testing Phase

- 5 simple tasks
- Explain the system
- Basic Language Knowledge
Free Content

- Leaderboard
- Points for translating
- Points for activity

Free Translations
Blog feeds
Task Stream

- Tasks are chunks of news
- User gets points per task
- News are displayed at news.unbabel.com
Task Feedback

- Crowdsourcing Evaluation
- Community members are asked to give feedback
- Users gain Unbabel points
User Rating

• User Rating based on completed tasks.

• Avg. Rating 4+ to be paid
Paid Editors

- Hourly Rate depending on quality
- Can get demoted if quality decreases
- Can get promoted to gold editors (if consistent high quality)
Feedback Coverage

26% editors are paid

Lack of feedbacks

1268 Paid Editors

The best editors have access to paid content

4740 Editors

Editors get ratings on the tasks

7872 Editors

Editors are tested when they sign up

Paid Content

Free Content

Testing Phase

7872 Editors

4740 Editors

1268 Paid Editors

Feedback Coverage

26% editors are paid
Implicit Feedback

- TER, Time, Words
- Editor 1/2 profile
- Train a regression model
Challenges

Few edits

Good or Bad?
Challenges

Good MT

Few edits

Strong positive signal
Challenges

- Bad MT
- Few edits
- Strong negative signal
How much can we trust the users feedback?
It’s still crowdsourcing
Future Challenges

- MT Evaluation
- User curation
  - Who’s a good editor, who’s a good feedback giver
- Coherence of final text
- User profile and task routing
- Specialized MT engines (Lots of interesting data)
Take away message

• Not enough translators for current translation needs
• MT is not there yet for customer facing content
• New content not valuable enough for current human translation prices
• Tons of interesting challenges to work on
Thank You

João Graça and the Unbabel team