Combining Crowd and AI
to scale professional-quality translation

João Graça
CTO
The Internet, 1997

80% English
The Internet, 2017

30% English

20% Chinese

6% Japanese

5% Portuguese

8% Spanish

4% German

3% Arabic
“All translation firms together are able to translate far less than 1% of relevant content produced everyday”

CSA – MT Is Unavoidable to Keep Up with Content Volumes
Is Machine Translation already here?

Everyone agrees that NMT is here to stay and much better than SMT
Unbabe Experiments with Customer Service Tickets

Moses (generic)  29.6
Neural MT (generic)  35.7
Moses (adapted)  43.5
Neural MT (adapted)  49.9
Is NMT really enough?

A Neural Network for Machine Translation, at Production Scale. Google Blog
Quality per Job

MQM

Job

- Good
- Not sure
- Bad
Will AI solve translation?
Will AI solve translation?

MQM 95

HUMAN EFFORT

TIME

JOBS
Quality per Job

MQM

Job

Good
Not sure
Bad
Unbabel Pipeline
Unbabel Pipeline
Unbabel Pipeline
Unbabel Pipeline

Job

Machine Translation
Unbabel Pipeline

Job ➔ Machine Translation ➔ Q.E. (Quality Estimation)
Unbabel Pipeline

Job → Machine Translation → Quality Estimation → Customer

High Q.E.
Data Generation Engine

Job -> Machine Translation -> Quality Estimation -> Community Translators -> Quality Estimation -> Customer
Data Generation Engine

Before

Initial text

NO DATA POINTS

Submitted text

After

Initial text

All changes in between:
Mouse clicks
Key presses
Timestamps

Submitted text
Keystroke Analysis

Raw data

At 18:03:30: In nugget 3
mouseClick
Cursor at 16
Selected: 0

At 18:03:31: In nugget 3
Pressed Backspace
Cursor at 15
Selected: 0

At 18:03:31: In nugget 3
Pressed Backspace
Cursor at 15
Selected: 0

At 18:03:35: In nugget 3
Pressed Shift
Cursor at 25
Selected: 0

At 18:03:35: In nugget 3
Pressed s
Cursor at 25
Selected: 0

At 18:03:35: In nugget 3
Pressed i
Cursor at 26
Selected: 0

Processed information

Initial text
“Espero que esto es útil”

• Deleted word “es”
• Inserted word “sea”

Submitted text
“Espero que esto sea útil”
Unbabel Pipeline

MACHINE

Job

Machine Translation

Q.E.

Quality Estimation

Q.E.

Low Q.E.

Re-Eval

COMMUNITY

High Q.E.

Crowd

Customer
Machine Translation Models

Phrase-based MT

Neural MT

Er liebte zu essen.
NULL Er liebte zu essen

He loved to eat.
Customer Adaptation

Customer Support Tickets

Bleu
Quality Estimation

**Word-Level QE**
Which words are translated correctly/incorrectly?

**Sentence-Level QE**
How good is the entire translation?
Quality Estimation

Word-level QE example

Hey there, I am sorry about that!
Hey lá, eu sou pesaroso sobre aquele!

Hey lá, eu sou pesaroso sobre aquele!
 QE Training

Unbabel Ticket

source → MT → final

Bad translation

Good translation
QE Word Level
QE Word Level

- Ugent System
- LinearQE
- NeuralQE
- StackedQE
- APE-QE
- FullStackedQE

F1_MULT
QE Word Level

- Ugent System: 41.1
- LinearQE
- NeuralQE
- StackedQE
- APE-QE
- FullStackedQE

F1_MULT
QE Word Level

- **Ugent System**: 41.1
- **LinearQE**: 45.2
- **NeuralQE**
- **StackedQE**
- **APE-QE**
- **FullStackedQE**

F1_MULT
QE Word Level

- Ugent System: 41.1
- LinearQE: 45.2
- NeuralQE: 47.3
- StackedQE
- APE-QE
- FullStackedQE
QE Word Level

- Ugent System: 41.1
- LinearQE: 45.2
- NeuralQE: 47.3
- StackedQE: 50.3
- APE-QE
- FullStackedQE

F1_MULT

WMT 16
WL QE Winner
QE Word Level

- Ugent System: 41.1
- LinearQE: 45.2
- NeuralQE: 47.3
- StackedQE: 50.3
- APE-QE: 55.7
- FullStackedQE: 55.7

WMT 16 WL QE Winner
QE Word Level

Andre F. T. Martins, Marcin Junczys-Dowmunt, Fabio Kepler, Ramon Astudillo, Chris Hokamp, Roman Grundkiewicz. “Pushing the Limits of Translation Quality Estimation.” TACL 2017 (To Appear)
QE Sentence Level
QE Sentence Level

- YANDEX: 52.5
- StackedQE
- APE-QE
- FullStackedQE

Pearson Correlation
QE Sentence Level

- Yandex: 52.5
- StackedQE: 54.9
- APE-QE: 54.9
- FullStackedQE: 54.9

Pearson Correlation
QE Sentence Level

- YANDEX: 52.5
- StackedQE: 54.9
- APE-QE: 61.3
- FullStackedQE: 70.0

Pearson Correlation
QE Sentence Level

<table>
<thead>
<tr>
<th>Model</th>
<th>Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>YANDEX</td>
<td>52.5</td>
</tr>
<tr>
<td>StackedQE</td>
<td>54.9</td>
</tr>
<tr>
<td>APE-QE</td>
<td>61.3</td>
</tr>
<tr>
<td>FullStackedQE</td>
<td>65.6</td>
</tr>
</tbody>
</table>

Andre F. T. Martins, Marcin Junczys-Dowmunt, Fabio Kepler, Ramon Astudillo, Chris Hokamp, Roman Grundkiewicz. “Pushing the Limits of Translation Quality Estimation.” TACL 2017 (To Appear)
QE in the Pipeline
QE in the Pipeline

Job → Machine Translation → Quality Estimation → Customer

High Q.E.
QE in the Pipeline

Document-Level QE
how good is the entire document?
 QE in the Pipeline

Document-Level QE
how good is the entire document?
QE in the Pipeline

- **Job**
- **Machine Translation**
- **Quality Estimation (Q.E.)**
  - Low Q.E.
  - Re-Eval
- **Community Translators**
- **High Q.E.**
- **Customer**

**Document-Level QE**
how good is the entire document?

**Human QE**
Can we evaluate post-edit output?

**Interesting numbers coming soon**
Professional Translation

Goals

★ Quality
★ Speed
★ Cost
<table>
<thead>
<tr>
<th>Goals</th>
<th>Pillars</th>
</tr>
</thead>
<tbody>
<tr>
<td>★ Quality</td>
<td></td>
</tr>
<tr>
<td>★ Speed</td>
<td></td>
</tr>
<tr>
<td>★ Cost</td>
<td></td>
</tr>
</tbody>
</table>
Professional Translation

Goals

- Quality
- Speed
- Cost

Pillars
Professional Translation

Goals
★ Quality
★ Speed
★ Cost

Pillars
• Editors Pool
Professional Translation

Goals

★ Quality
★ Speed
★ Cost

Pillars

• Editors Pool
• Initial Text (MT)
Professional Translation

Goals
- Quality
- Speed
- Cost

Pillars
- Editors Pool
- Initial Text (MT)
- Editor Assignment
Professional Translation

<table>
<thead>
<tr>
<th>Goals</th>
<th>Pillars</th>
</tr>
</thead>
<tbody>
<tr>
<td>★ Quality</td>
<td>• Editors Pool</td>
</tr>
<tr>
<td>★ Speed</td>
<td>• Initial Text (MT)</td>
</tr>
<tr>
<td>★ Cost</td>
<td>• Editor Assignment</td>
</tr>
<tr>
<td></td>
<td>• Interfaces</td>
</tr>
</tbody>
</table>
Professional Translation

Goals

- Quality
- Speed
- Cost

Pillars

- Editors Pool
- Initial Text (MT)
- Editor Assignment
- Interfaces
- Quality Evaluation
Unbabel Community
Unbabel Community

50 000 Users
Distributed Pipeline

Quality Estimation
Editors Pool

- **Testing Phase**: Editors are tested when they sign up.
- **Training Content**: Editors get ratings for the tasks.
- **Paid Content**: The best editors have access to paid content.
- **Expert**: Specialisation layers will grow with time.
Editors Pool

- **Testing Phase**: Editors are tested when they sign up.
- **Training Content**: Editors get ratings for the tasks.
- **Paid Content**: The best editors have access to paid content.
- **Expert**: Specialisation layers will grow over time.

Evaluators
Editors Pool

- **Testing Phase**: Editors are tested when they sign up.
- **Training Content**: Editors get ratings for the tasks.
- **Paid Content**: The best editors have access to paid content.
- **Expert**: Specialisation layers will grow with time.

- **Evaluators**
- **Annotators**
How Editors are Evaluated
Editors Profiling
Editors Profiling
Editor Assignment
Editor Assignment

Tasks/time

- 2 m
- 6 m
- 10 m
- 12 m
- 18 m
- 45 m
Editor Assignment

<table>
<thead>
<tr>
<th>Tasks/time</th>
<th>Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 m</td>
<td><img src="image" alt="Editor" /></td>
</tr>
<tr>
<td>6 m</td>
<td><img src="image" alt="Editor" /></td>
</tr>
<tr>
<td>10 m</td>
<td><img src="image" alt="Editor" /></td>
</tr>
<tr>
<td>12 m</td>
<td><img src="image" alt="Editor" /></td>
</tr>
<tr>
<td>18 m</td>
<td><img src="image" alt="Editor" /></td>
</tr>
<tr>
<td>45 m</td>
<td><img src="image" alt="Editor" /></td>
</tr>
</tbody>
</table>
Editor Assignment

<table>
<thead>
<tr>
<th>Tasks/time</th>
<th>Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 m</td>
<td></td>
</tr>
<tr>
<td>6 m</td>
<td></td>
</tr>
<tr>
<td>10 m</td>
<td></td>
</tr>
<tr>
<td>12 m</td>
<td></td>
</tr>
<tr>
<td>18 m</td>
<td></td>
</tr>
<tr>
<td>45 m</td>
<td></td>
</tr>
</tbody>
</table>

Pull
Editor Assignment

<table>
<thead>
<tr>
<th>SLA</th>
<th>Tasks/time</th>
<th>Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 H</td>
<td>2 m</td>
<td></td>
</tr>
<tr>
<td>30 m</td>
<td>6 m</td>
<td></td>
</tr>
<tr>
<td>2 D</td>
<td>10 m</td>
<td></td>
</tr>
<tr>
<td>6 D</td>
<td>12 m</td>
<td></td>
</tr>
<tr>
<td>20 m</td>
<td>18 m</td>
<td></td>
</tr>
<tr>
<td>40 m</td>
<td>45 m</td>
<td></td>
</tr>
</tbody>
</table>

Pull
Editor Assignment

<table>
<thead>
<tr>
<th>Priority</th>
<th>SLA</th>
<th>Tasks/time</th>
<th>Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>6 H</td>
<td>2 m</td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>30 m</td>
<td>6 m</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>2 D</td>
<td>10 m</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>6 D</td>
<td>12 m</td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>20 m</td>
<td>18 m</td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>40 m</td>
<td>45 m</td>
<td></td>
</tr>
</tbody>
</table>
## Editor Assignment

<table>
<thead>
<tr>
<th>Queue</th>
<th>Priority</th>
<th>SLA</th>
<th>Tasks/time</th>
<th>Editors</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1000</td>
<td>6 H</td>
<td>2 m</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1100</td>
<td>30 m</td>
<td>6 m</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1000</td>
<td>2 D</td>
<td>10 m</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1000</td>
<td>6 D</td>
<td>12 m</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>1100</td>
<td>20 m</td>
<td>18 m</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>1100</td>
<td>40 m</td>
<td>45 m</td>
<td></td>
</tr>
</tbody>
</table>
## Editor Assignment

<table>
<thead>
<tr>
<th>Queue</th>
<th>Priority</th>
<th>SLA</th>
<th>Tasks/time</th>
<th>Editors</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1100</td>
<td>6 m</td>
<td>2 m</td>
<td></td>
<td>4.2</td>
</tr>
<tr>
<td>G</td>
<td>1100</td>
<td>2 D</td>
<td>10 m</td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>G</td>
<td>1100</td>
<td>6 D</td>
<td>12 m</td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>R</td>
<td>1100</td>
<td>20 m</td>
<td>18 m</td>
<td></td>
<td>4.8</td>
</tr>
<tr>
<td>R</td>
<td>1100</td>
<td>40 m</td>
<td>45 m</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Editor Assignment

<table>
<thead>
<tr>
<th>Queue</th>
<th>Topics</th>
<th>Priority</th>
<th>SLA</th>
<th>Tasks/time</th>
<th>Editors</th>
<th>Rating</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>G</td>
<td>1000</td>
<td>6 H</td>
<td>2 m</td>
<td></td>
<td>4.2</td>
<td>🇪🇸</td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>1100</td>
<td>30 m</td>
<td>6 m</td>
<td></td>
<td>3.8</td>
<td>👨‍♂️</td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>1000</td>
<td>2 D</td>
<td>10 m</td>
<td></td>
<td>4.3</td>
<td>🇵🇹</td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>1000</td>
<td>6 D</td>
<td>12 m</td>
<td></td>
<td>4.8</td>
<td>🇬🇧</td>
</tr>
<tr>
<td>R</td>
<td>R</td>
<td>1100</td>
<td>20 m</td>
<td>18 m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>R</td>
<td>1100</td>
<td>40 m</td>
<td>45 m</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pull
## Editor Assignment

<table>
<thead>
<tr>
<th>Queue</th>
<th>Topics</th>
<th>Priority</th>
<th>SLA</th>
<th>Tasks/time</th>
<th>Editors</th>
<th>Rating</th>
<th>Native</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>G</td>
<td>1000</td>
<td>6 H</td>
<td>2 m</td>
<td>4.2</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>1100</td>
<td>30 m</td>
<td>6 m</td>
<td>4.3</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>1000</td>
<td>2 D</td>
<td>10 m</td>
<td>4.8</td>
<td>4.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>1000</td>
<td>6 D</td>
<td>12 m</td>
<td>4.3</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>R</td>
<td>1100</td>
<td>20 m</td>
<td>18 m</td>
<td>4.8</td>
<td>4.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>R</td>
<td>1100</td>
<td>40 m</td>
<td>45 m</td>
<td>4.8</td>
<td>4.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Editor Assignment
Editor Assignment

Regular distribution

old rating 3.8
Editor Assignment

Regular distribution

Smart distribution

old rating 3.8

Improved rating 4.6
Post-Editing Interfaces
Time Spent on Job
Time Spent on Job

Translator 1

MT  WAITING  DELIVERY

TIME
Time Spent on Job

Translator 1

Translator 2

MT
WAITING
WAITING
DELIVERY

TIME
Time Spent on Job: Mobile

 translators

 MT WAITING Waiting DELIVERY

 20% less
Time Spent on Job

Delivery Time (100 words)
Pos-Editing Interfaces

Please fix the remaining issues before submitting

Spelling: ... registrar novamente seus...
- novamente
- novamente, novamente

Inconsistency Issues
Sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam.

Spelling: ... de asdfasdfsdf distante...

3 issues detected

In that way we will be able to register again your tastes and provide you with awesome things.

Dessa forma, seremos capazes de registrar novamente seus gostos e fornecê-lo com sugestões impressionantes.
Smartcheck

External NLP Services
- Spell Check
- Syntax Parser
- Word Aligner
Smartcheck

External NLP Services
- Spell Check
- Syntax Parser
- Word Aligner

Spelling
Tone
Formality
Consistency
Client Rule

Eval
Annotated
Annotation Tool
Smartcheck

External NLP Services
- Spell Check
- Syntax Parser
- Word Aligner

Annotated
Annotation Tool
MACHINE + HUMAN

CORE API

MESSAGING API

CHAT API

CYRANO API

Language OS

Language Engine

MACHINE + HUMAN

Customer Service

Conversational

zendesk

salesforce

salesforce

Messenger

Bots
In Customer Service

Tickets can come in many languages.
Unbabel for Zendesk

Unbabel can offer the **same Customer Satisfaction** as native agents
Customer Replies: Speed & Quality

**SPEED**

20 minutes

**QUALITY**

94
Chat

Unbabel Chat
Understand and be understood in multiple languages.

Start typing here, or paste the text you'd like to be translated.

Translate to Portuguese Translate

22:58

Bom dia, gostava de comprar um chapéu

Bom dia, isso é ótimo! De que cor gosta?

Good morning, that's great! What color do you like?

Translating...

Good morning, I would like to buy a hat

Bom dia, gostava de comprar um chapéu
Editors train the MT engine

Chat Translation Flow

Skips Humans

Community of 50K+ translators
Chat Messages: Speed & Quality

SPEED
2 Minutes

QUALITY
90
What is the future?

HUMAN EFFORT

TIME

MQM 95

JOBS

0%?