Translation as a Collaborative Activity

Benjamin B. Bederson

Computer Science Department
Human-Computer Interaction Lab
University of Maryland
Languages on Internet by Population

2009

English 52%

Japanese 9%
Spanish 8%
Chinese 5%
the rest 29%

2000

English 52%

Japanese 5%
Spanish 5%
Chinese 9%
the rest 29%

Source: Global Reach, Internet World Stats
International Children’s Digital Library

www.childrenslibrary.org
A real-world problem: ICDL

Now:

– Over 4,000 books
– Over 50 languages
– Some translations in a few languages
– ~50,000 adults, ~30,000 children registered
– ~100K unique visitors/month

Goal:

– 10,000 books
– 100 languages
– Every book in every language!
Machine Translation (MT)

Large volume, cheap, fast

Unreliable quality

(餐厅 = restaurant, dining hall)
Professional Translators

High quality, but slow and expensive (even for common language pairs)
Amateur Translators

Contents

1. How can I help make a requested translation? (in short)
3. When and how are changes made to a published translation?
   3.1 Updates to original text
   3.2 Amendments to a translation after publishing
4. Where should I make a translation?
5. I have no idea how to deal with wiki markup. Help!
6. How can I make a translation outside of wiki?
7. Where may I ask a question about translating?

About Us

"中外对话"由国际著名新闻记者伊莎贝尔·希尔顿于2006年创办。该网站提供了高质量和丰富的新闻信息和分析文章，并以中国为重点的环球问题展开讨论。

Founded by international journalist Isabel Hilton in 2006, chinadialogue is the bilingual source of high-quality news, analysis and discussion on all environmental issues, with a special focus on China. Read more...

How I reduced translation costs of 200 articles from $9000 to $46

Friends using this Application

Stephan Vogel
The key idea
Translation with the Crowd
Translation with the **Monolingual** Crowd

Wikipedia: 800 translators  vs. 75,000 contributors
Monolingual translation protocol

- Original source sentence: $F_0$
- Noisy back translation: $F_1$
- Fluent, accurate: $F_2$
- Monolingual post-editing: $E_0$
- Fluent target hypothesis: $E_1$
- Et cetera...: $E_2$

HTER editing
Monolingual translation protocol

• Each participant is performing a monolingual task:
  – Infer partner’s intended meaning as well as possible
  – Express that meaning grammatically in own language

• Source language participant has extra constraint:
  – Expressed meaning must match source sentence
  – Conflict? Original meaning wins
Three Types of Errors
detectable and correctable

Tout le monde doit entendre l'histoire de Cendrillon.

Everybody must to hear story about Cinderella

Everybody must hear the story about Cinderella
Three Types of Errors

detectable but not correctable

Tout le monde doit entendre l'histoire de Cendrillon.

Everybody must *heard the business by* Cinderella

Monolingual post-editing

?
Three Types of Errors

Tout le monde doit entendre l'histoire de Cendrillon.

Everybody *loves* the story about Cinderella
Enrich Translations

Increase *redundancy* and *shared context* ...

... to help make detectable errors correctable
... to help make undetectable errors detectable
Enriching context

- Ppl r rlly gd t gttn g mst f th mnng rght, vn frm dfcnt txt.

{minus, weak}
Monolingual Translation Protocol

Translation Channel:

S₀ → T₁
S₁ ← T₂
S₂ → T₃
⋮

Redundant source text

Enrichment Channel:

Images
Links
Structured Q&A

French speakers

French speaker

English speakers

English speaker
<table>
<thead>
<tr>
<th>Pierre</th>
<th>Mary</th>
</tr>
</thead>
</table>
| Says: *En général, on s'entend bien, tous les deux.*  
(lit. *In general, we get along together, the two of us.*) | Sees: *In general, it means well, both.*    |

MT
<table>
<thead>
<tr>
<th>Pierre</th>
<th>Mary</th>
</tr>
</thead>
</table>
| **Says:** *En général, on s'entend bien, tous les deux.*  
(lit. *In general, we get along together, the two of us.*) | **Sees:** *In general, it means well, both.*  
**Edits into:** *In general, it is about both of us.* |
<p>| <strong>Sees:</strong> <em>En général, Il est à la fois de nous.</em> |</p>
<table>
<thead>
<tr>
<th>Pierre</th>
<th>Mary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Says: <em>En général, on s'entend bien, tous les deux.</em> (lit. <em>In general, we get along together, the two of us.</em>)</td>
<td>Sees: <em>In general, it means well, both.</em></td>
</tr>
<tr>
<td>Sees: <em>En général, Il est à la fois de nous.</em></td>
<td>Edits into: <em>In general, it is about both of us.</em></td>
</tr>
<tr>
<td>Edits into: <em>En général, nous nous entendons bien.</em> (lit. <em>In general, we get along well.</em>)</td>
<td>Sees: <em>In general, we get along fine.</em></td>
</tr>
<tr>
<td></td>
<td>MT</td>
</tr>
<tr>
<td></td>
<td>MT</td>
</tr>
<tr>
<td></td>
<td>MT</td>
</tr>
<tr>
<td></td>
<td>MT</td>
</tr>
<tr>
<td></td>
<td>MT</td>
</tr>
</tbody>
</table>

**enrichment**
Pierre

Says: *En général, on s'entend bien, tous les deux.*
(lit. *In general, we get along together, the two of us.*)

Sees: *En général, Il est à la fois de nous.*

Edits into: *En général, nous nous entendons bien.*
(lit. *In general, we get along well.*)

Sees: *En général, nous sommes de bons amis.*
(lit. *In general, we are good friends.*)

Mary

Sees: *In general, it means well, both.*

MT

Edits into: *In general, it is about both of us.*

Sees: *In general, we get along fine.*

MT

Edits into: *In general, we are good friends.*
<table>
<thead>
<tr>
<th>Pierre</th>
<th>Mary</th>
</tr>
</thead>
</table>
| **Says:** *En général, on s'entend bien, tous les deux.*  
(lit. *In general, we get along together, the two of us.*) | **Sees:** *In general, it means well, both.*  
**Edits into:** *In general, we get along well.*  
(lit. *In general, we get along well.*)  
**Proposes to stop with current translation*** |
| **Sees:** *En général, Il est à la fois de nous.* | **Edits into:** *In general, it is about both of us.*  
**Sees:** *En général, nous sommes de bons amis.*  
(lit. *In general, we are good friends.*) | **Edits into:** *In general, we are good friends.*  
**Agrees to stop with current translation*** |
| **Edits into:** *En général, nous nous entendons bien.*  
(lit. *In general, we get along well.*) | **enrichment*** |
Everyone has heard the story of Cinderella.

It is the story of a girl who lost her mother and her father and was in the care of a person who loved her so much as they wanted their parents.

Does anyone ever thinks about all the children growing up without love and tenderness?

Nobody understands them, reject them all, and they seek care.

I look for when dreams, fairy tales, including toys, among birds, animals and flowers.

All around us, continuously, sad children are growing up and become adults grieve.

This is the story of a girl as well.
It is the story of a girl who lost her mother and her father and everyone was in the same situation of a person who loved her so much as they wanted their parent to live longer.

Does anyone ever thinks about all the children growing up without the love and tenderness?

They are lonely and sad.

Nobody understands them, reject them all, and they seek care.

I look for when dreams, fairy tales, including toys, among birds, animals and flowers.

All around us, continuously, sad children are growing up and become adults grieve.

This is the story of a girl as well.
Es la historia de una niña que perdió a su madre y a su padre y quedó al cuidado de una persona que no la quería tanto como la querían sus padres.

¿Alguien piensa alguna vez en todos los niños que crecen sin amor ni ternura? Están solos y tristes.

Nadie les entiende, todos les rechazan, y ellos buscan cariño.

Lo buscan cuando sueñan, en los cuentos de hadas, entre sus juguetes, entre los pájaros, los animales y las flores.

A nuestro alrededor, continuamente, hay niños tristes que crecen y se convierten en adultos tristes.

Esta es la historia de una niña así.
Preliminary validation of the protocol

- Language pair: Russian to Chinese
  - Hard case: no orthographic cues
  - Easy to find local volunteers
- Two Russian speakers and four Chinese speakers
  - Four Russian-Chinese translation pairs (Russians twice)
  - One hour per pair
- Worked on 44 sentences (6 pages), finished 28
  - = ~8.5 minutes per sentence (~1 word per minute)
  - *N.B. average translators: ~2500 words per 8hr day* (= ~5 words per minute)
Google Translate …

<table>
<thead>
<tr>
<th>Original meaning translated</th>
<th>Google Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>6</td>
</tr>
<tr>
<td>little</td>
<td>6</td>
</tr>
<tr>
<td>much</td>
<td>9</td>
</tr>
<tr>
<td>most</td>
<td>4</td>
</tr>
<tr>
<td>all</td>
<td>3</td>
</tr>
</tbody>
</table>
... Monotrans

![Bar chart showing the comparison between Google Translate and Monotrans for different levels of meaning translated.](chart.png)
One more observation

The original source sentence is not the only way the intended meaning could have been expressed.

Suppose this phrasing is difficult to translate correctly

Perhaps one of these alternatives can be more successful
Polls indicate Brown, a state senator, and Coakley, Massachusetts’ Attorney General, are locked in a virtual tie to fill the late Sen. Ted Kennedy’s Senate seat.

Les sondages indiquent Brown, un sénateur d’ état, et Coakley, Massachusetts’ Procureur général, sont enfermés dans une cravate virtuel à remplir le regrettés énateur Ted Kennedy’s siège au Sénat.

Polls indicate that Brown, a state senator, and Coakley, the Attorney General of Massachusetts, are locked in a virtual tie to fill the Senate seat of the Sen. Ted Kennedy, who died recently.

Les sondages indiquent que Brown, un sénateur d’ état, et Coakley, le procureur général du Massachusetts, sont enfermés dans une cravate virtuel pourvoir le sige au Sénat de Sen. Ted Kennedy, qui est décédé récemment.
Automatically determining where the errors are

The most recent probe to visit Jupiter was the Pluto-bound new horizons spacecraft.

The latest research visit Jupiter was the Pluto-bound new horizons spacecraft.

the most recent probe to visit jupiter was the pluto-bound new horizons spacecraft
Where to from here?

• Larger and more formal validation of the protocol
• Exploring the space of richer annotations
• Reconsidering UI for:
  – Ease of use
  – Throughput
  – Parallel, multi-person contribution
• Exploring the space of automatic and human error detection and paraphrase
Collaborators and Sponsors

Philip Resnik
Linguistics
Professor

Chang Hu
CS Ph.D. student

Olivia Buzek
CS/Linguistics undergrad

Alex Quinn
CS Ph.D. student

Yakov Kronrod
Linguistics Ph.D. student