Towards a Research Infrastructure for Translation Studies

Lars Ahrenberg

Department of Computer and Information Science
Linköping University

1st National SWE-CLARIN Workshop
Uppsala, November 13, 2014
Overview

1. CLARIN and translation studies

2. A micro example

3. Conclusions
The CLARIN mission

"to advance research in humanities and social sciences by giving researchers unified single sign-on access to a platform which integrates language-based resources and advanced tools at a European level."
Towards a Res
Lars Ahrenberg

CLARIN and translation studies
A micro example
Conclusions

Virtual Language Observatory
Translation studies

“the discipline which concerns itself with the theory and practice of translation” (Hatim, 2001: p. 3)
Translation studies

“the discipline which concerns itself with the theory and practice of translation” (Hatim, 2001: p. 3)

Methodology: a combination of close reading and studies of relevant (cultural) context.
“the discipline which concerns itself with the theory and practice of translation” (Hatim, 2001: p. 3)

As texts are concerned, this often means comparisons:

- source text – translation
- translations – originals in target language
- different translations of same source texts
Translation studies and corpus linguistics

- corpora as resources for translation
- corpora for studies of 'translationese'
- corpora for studies of 'translation universals'
- ...

Conclusions
“The underlying assumption is that by exploring the literary record writ large, we will better understand the context in which individual texts exist and thereby better understand those individual texts.” (Jockers, 2013: p. 27)
Towards a Res

Lars Ahrenberg

CLARIN and translation studies

A micro example

Conclusions

1. CLARIN and translation studies

2. A micro example

3. Conclusions
Source text
Joseph Conrad: *Heart of Darkness*, 1899
first published as a serial in *Blackwood’s Magazine*

Note: first 40 pages only, about one third of the full novel.

Translations
Einar Heckscher. Lindelöws, 2006 (rev. 2013)
A comparison by characters
A comparison by words

What, then, could be the cause for the difference?
What is normal?

Word rates from a few LinES corpora.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Word rate</th>
<th>Char rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordimer</td>
<td>1.039</td>
<td>1.178</td>
</tr>
<tr>
<td>Auster</td>
<td>0.987</td>
<td>1.111</td>
</tr>
<tr>
<td>Bellow</td>
<td>0.966</td>
<td>1.091</td>
</tr>
<tr>
<td>TRL 1</td>
<td>0.929</td>
<td>1.067</td>
</tr>
<tr>
<td>TRL 2</td>
<td>0.988</td>
<td>1.154</td>
</tr>
</tbody>
</table>
What is normal?

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Word rate</th>
<th>Char rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordimer</td>
<td>1.039</td>
<td>1.178</td>
</tr>
<tr>
<td>Auster</td>
<td>0.987</td>
<td>1.111</td>
</tr>
<tr>
<td>Bellow</td>
<td>0.966</td>
<td>1.091</td>
</tr>
<tr>
<td>TRL 1</td>
<td><strong>0.929</strong></td>
<td><strong>1.067</strong></td>
</tr>
<tr>
<td>TRL 2</td>
<td>0.988</td>
<td>1.154</td>
</tr>
</tbody>
</table>
How are the rate differences distributed?

In fact, the general tendency is the same over the text as a whole.
Hypothesis

Heckscher’s translation has lower density than Odelbergs and more instances of explicitation.
Vertical cuts based on length differences 1

SRC: He was the only man of us who still followed the sea.
TRL1: Han var den ende av oss som var kvar på sjön.
TRL2: Han var den ende av oss som fortfarande ”vigde sitt liv åt havet”.
Vertical cut 2

SRC: The fascination of the abomination – *you know.*
TRL1: Det gräsligas lockelse, *ni vet.*
TRL2: Det vederstyggligas lockelse – *ni vet vad jag menar.*
Vertical cut 3

SRC: What redeems it *is the idea* only.
TRL1: Det enda försonande *är idén*.
TRL2: Det enda försonande *skulle vara en eventuell idé då*. 
Possible causes

• the translators (?)
• the translation norms current at the time of translation (?)
• the publishing house (?)
• ...

Overview

1. CLARIN and translation studies
2. A micro example
3. Conclusions
What is needed

1. lots of translations
What is needed

1. lots of translations from lots of source texts
What is needed

1. lots of translations from lots of source texts
2. alignment tools at paragraph and sentence level
What is needed

1. lots of translations from lots of source texts
2. alignment tools at paragraph and sentence level
3. support for translation spotting
What is needed

1. lots of translations from lots of source texts
2. alignment tools at paragraph and sentence level
3. support for translation spotting
4. powerful search tools that can display N parallel snippets for large N
What is needed

1. lots of translations from lots of source texts
2. alignment tools at paragraph and sentence level
3. support for translation spotting
4. powerful search tools that can display $N$ parallel snippets for large $N$
5. annotation tools (and taxonomies)
What is needed

1. lots of translations from lots of source texts
2. alignment tools at paragraph and sentence level
3. support for translation spotting
4. powerful search tools that can display $N$ parallel snippets for large $N$
5. annotation tools (and taxonomies)
6. worked out workflows illustrating good practice
The end

Thank you!