Corpora and literary translation research: some methodological issues

Federico Zanettin
Università di Perugia
Thessaloniki, 15 January 2014
Corpora in translation research

- Translation universals
- Translator style
- Translation norms
- Translation and language change
- Contrastive linguistics
- Interpreting studies
- Translation learner corpora
- …
Translator style

• coherent and motivated patterns of choice
  • “recognizable across a range of translations by the same translator”
  • “distinguish that translator’s work from that of other translators”
  • “cannot be explained as directly reproducing the source text’s style or as the inevitable result of linguistic constraints”

(Saldanha, 2011: 240).
• Norms
  • interpersonal and above the individual
  • refer to variant rather than invariant traits of translation

• Translation and language change
  • how evolving translators’ styles and norms relate to evolving language norms
  • how translation affects and is in its turn affected by language change

• Contrastive linguistics
  • provides a basis for assessing translation-specific as opposed to source language-specific constraints
Corpus-based studies of literary translation

- Conventionality vs creativity
- Research aim
- Corpus design
- Data
- Method
Research aim

- Baker (1999) *stylistic variation across translators*
- Kenny (2001) lexical creativity in translation
- Zanettin (2001) *stylistic variation across translators*
- Johansson (2004) *stylistic choices vs source language interference*
- Bosseaux (2007) translators discursive presence and point of view
- Saldanha (2011) *stylistic choices vs source language interference*
- Ji and Oakes (2012) *stylistic variation across translators*
- Rybicki (2012) *stylistic variation across translators*
- Ji (2012) style and language change
Corpus design

- **Monolingual translational corpus**
  - Same target language, different source languages, no source texts (*Baker*)
  - Same target language, different translations of the same source text, no source text (*Ji and Oakes*)

- **Multilingual translational corpus**
  - Different target languages, different source languages, no source texts (*Rybicki*)

- **Monodirectional parallel corpus**
  - One S and one T language, different authors, different translators, not aligned (*Kenny*)
  - One S and one T language, same author, different translators, aligned (*Zanettin*)

- **Multilingual parallel corpus**
  - Same target language, different source languages, different translators, source texts (*Saldanha*)

- **Parallel corpus, “star” model**
  - Multiple translations of the same text(s) in the same language (*Johansson, Bosseaux, Ji*)
Data

- **Kenny: lexis**
  - hapax legomena, writer-specific forms, and unusual collocations
- **Saldanha: typographical markers**
  - Italics, quotation marks
- **Bosseaux: lexis**
  - deixis, modality, transitivity and free indirect discourse
- **Zanettin: lexis**
  - expressions of indeterminacy, reporting verbs + adverbs
- **Johansson: syntax**
  - shifts in translating the subject of the first clause of declarative sentences
- **Ji and Oakes: syntax and lexis**
  - words and sentence length, POS distribution, “emotion/value” words, n-grams
- **Rybicki: lexis**
  - Distribution of most frequent words
- **Ji: lexis**
  - Archaisms, idiomatic expressions
Methods

- **Baker**: wordlists, monolingual concordances, no annotation
- **Kenny**: wordlists, monolingual concordances, no annotation, reference monolingual corpora (S + T)
- **Saldanha**: wordlists, monolingual concordances (limited annotation, reference parallel corpus)
- **Bosseaux**: monolingual and parallel concordances, no annotation
- **Zanettin**: monolingual and parallel concordances, (annotation), reference/control monolingual corpora (S + T)
- **Johansson**: parallel concordances, syntactic annotation
- **Ji and Oakes**: statistics, syntactic and semantic annotation
- **Rybicki**: statistics (Burrow’s Delta, cluster analysis tree diagrams)
- **Ji**: statistics (linear regression analysis of frequency data), diachronic target reference corpora
Parallel corpora construction and analysis

- Literary texts “resistent” to alignment
- Automatic alignment + manual editing of alignment
- Time consuming but important analytical phase
## The Rushdie Parallel Corpus

<table>
<thead>
<tr>
<th>Title</th>
<th>No, of Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midnight's Children</td>
<td>214,982</td>
</tr>
<tr>
<td>I figli della mezzanotte</td>
<td>224,398</td>
</tr>
<tr>
<td>Haroun and the Sea of Stories</td>
<td>43,877</td>
</tr>
<tr>
<td>Harun e il mar delle storie</td>
<td>44,624</td>
</tr>
<tr>
<td>The Moor's Last Sigh</td>
<td>159,752</td>
</tr>
<tr>
<td>L'ultimo sospiro del Moro</td>
<td>167,630</td>
</tr>
<tr>
<td>Shame</td>
<td>105,488</td>
</tr>
<tr>
<td>La vergogna</td>
<td>107,946</td>
</tr>
<tr>
<td>The Satanic Verses</td>
<td>192,634</td>
</tr>
<tr>
<td>I versi satanici</td>
<td>196,646</td>
</tr>
<tr>
<td>Chekov and Zulu</td>
<td>4,687</td>
</tr>
<tr>
<td>Chekov e Zulu</td>
<td>5,025</td>
</tr>
<tr>
<td><strong>Total (English)</strong></td>
<td><strong>721,420</strong></td>
</tr>
<tr>
<td><strong>Total (Italian)</strong></td>
<td><strong>746,269</strong></td>
</tr>
<tr>
<td>English</td>
<td>Italian</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td><em>Midnight’s children</em> by Salman Rushdie</td>
<td><em>I figli della mezzanotte</em> translated by Ettore Capriolo</td>
</tr>
</tbody>
</table>

| '<P>'A jewel,' he said, honking into a handkerchief, 'Sir and Madam, your daughter is a jewel. I am humbled, absolutely. Darn humbled. She has proved to me that a golden voice is preferable even to golden teeth.'</P>' | '<P>'Un gioiello,' disse, strombettando nel suo fazzoletto. "Signore, signora, vostra figlia è un gioiello. Sono mortificato. Assolutamente mortificato. Mi ha dimostrato che una voce d'oro è preferibile persino ai denti d'oro."</P>' |
'A jewel,' he said, honking into a handkerchief, 'Sir and Madam, your daughter is a jewel.

I am humbled, absolutely.

Darn humbled.

She has proved to me that a golden voice is preferable even to golden teeth.'

"Un gioiello," disse, strombettando nel suo fazzoletto.

"Signore, signora, vostra figlia è un gioiello.

Sono mortificato.

Assolutamente mortificato.

Mi ha dimostrato che una voce d'oro è preferibile persino ai denti d'oro."
'A jewel,' he said, honking into a handkerchief, 'Sir and Madam, your daughter is a jewel.

I am humbled, absolutely.

Darn humbled.

She has proved to me that a golden voice is preferable even to golden teeth.'

"Un gioiello," disse, strombettando nel suo fazzoletto.

"Signore, signora, vostra figlia è un gioiello.

Sono mortificato.

Assolutamente mortificato.

Mi ha dimostrato che una voce d'oro è preferibile persino ai denti d'oro."
"A jewel," he said, honking into a handkerchief, 'Sir and Madam, your daughter is a jewel.

I am humbled, absolutely.

Darn humbled.

She has proved to me that a golden voice is preferable even to golden teeth.'


Sono mortificato.

Assolutamente mortificato.

Mi ha dimostrato che una voce d'oro è preferibile persino ai denti d'oro."
'A jewel,' he said, honking into a handkerchief, 'Sir and Madam, your daughter is a jewel. I am humbled, absolutely. Darn humbled. She has proved to me that a golden voice is preferable even to golden teeth.'

Parallel concordances

- *ParaConc* by Michael Barlow
- Reporting direct speech
- Verb of saying post-modified by an adverb
  - “she said dryly”
- Search for “said *ly”
... just a day or two," "Salad baba," she said harshly, "I've got to hand it to you, man. You...

... unnuu, I have a gift for accents," he said haughtily, "Why shouldn't I employ?" "Why I...

... myi is spot on," But I love her," I said helplessly, "I can't stop. I just can't," Myna...

... their is roof, What else to do?" "Hyder said helplessly. CHAPTER TWELVE Stability The great...

... dle money," 'But she is a child,' Raza said hopelessly, Bilquis grew stronger as he weakened...

... d to show Uma around the old place, and said hotly, "You are welcome in our home as often...

... tk to be done," I am ready," Gibeel said humbly, "I was just going, anyway," 'Look,' ...

... ned in puzzlement as Haroun arrived and said hurriedly, "I'll explain later," and gave Iss a g...

... e mouthful of gold?) 'The nation,' he said indelicately through his untoothed gums, 'must not, ...

... is wicked grin, 'A foolish notion,' he said innocently, 'A fancy, a passing whim. Certainly ...

... with skinned apricots. 'Yum, yum,' he said ironically, smiling his crooked smile. He fell as ...

... 'I need to tell me about yourself,' he said jovially. Naturally I know who you are, even i...

... my head, 'I have been in mourning,' I said, lamely, and she, 'But still you could have - m...

... no-one. For public consumption she had said laughingly, "So if this is a racecourse then I wan...

... here is education and breeding," Belle said loudly, 'and we have behaved like dogs. 'Not...

... ace, I was terrified. 'Good luck,' I said loudly, 'Can't hear a word,' she said, and I...

... ay what-what we do for minorities," he said loudly, 'Is it a Hindu who is given this hono...

... k a grip on himself 'The savagery,' he said loudly in his English voice. 'The sheer barba...

... preciate fine contemporary singing," he said loudly to General Kitab and Mudra. 'So I thin...

... t it was going to be all right; then he said meekly, 'Tughlakji, you know, with your previo...

... ze for the wedding fiasco, but Iskander said merrily, 'For God's sake, Raza, Haroun can take ...
Said *ly*, sorted according to target language patterns.
<table>
<thead>
<tr>
<th>%</th>
<th>Pattern</th>
</tr>
</thead>
</table>
| 40%   | **disse (diceva/dissi) + *mente** (e.g.: "mildly = dolcemente,\n                              pacatamente, gentilmente, “quickly” = prontamente, "gummily“ =\n                              gengivamente, "flatly“ = recisamente, categoricamente)\n
| 25%   | **disse + con + abstract noun** (e.g.: "meekly“ = con umiltà, "fiercely“ =\n                              con ferocia, con asprezza, con ardore, "firmly“ = con fermezza)\n
| 15%   | **disse + in/con tono/aria/voce + adj.** (e.g.: "approovingly“ = in tono\n                              d'approvazione, "coquettishly“ = in tono civettuolo, "flatly“ = in tono\n                              reciso, "innocently“ = con aria innocente, "contemptuously“ = con aria\n                              sprezzante, "flatly“ = con voce neutra)\n
| 10%   | **disse + adj.** (e.g.: “she said moodily“ = disse imbronciata, “he said\n                              grimly“ = disse truce, “she said flatly” = disse risoluta\n
| 10%   | **disse + phrase** (e.g.: "I said helplessly“ = dissi, nella mia impotenza,\n                              "…Ayooba Baloch, who spoke coarsely“ = …che Ayooba Baloch, il cui\n                              linguaggio era grossolano)\n
**other** (e.g.: "he said indistinctly through his untoothed gums“ =\n                              farfugliò con le sue sdentate gengive, "he said finally = finì col dire,\n                              disse infine, “she said quickly” = si affrettò a dire)
<table>
<thead>
<tr>
<th></th>
<th>Capriolo</th>
<th>Mantovani</th>
</tr>
</thead>
<tbody>
<tr>
<td>a kind of (82)</td>
<td>una sorta di (39)</td>
<td>una specie di (19)</td>
</tr>
<tr>
<td></td>
<td>una specie di (4)</td>
<td>una sorta di (4)</td>
</tr>
<tr>
<td></td>
<td>come (3)</td>
<td>come (1)</td>
</tr>
<tr>
<td></td>
<td>un modo di (2)</td>
<td>un certo (1)</td>
</tr>
<tr>
<td></td>
<td>per così dire (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>una forma di (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in un certo senso (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>qualcosa di (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>qualche (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[x] omission (4)</td>
<td></td>
</tr>
<tr>
<td>a sort of (51)</td>
<td>una sorta (30)</td>
<td>una specie di (11)</td>
</tr>
<tr>
<td></td>
<td>una specie di (5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in certo qual modo (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in pratica (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>come (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>quasi (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in un certo senso (1)</td>
<td></td>
</tr>
<tr>
<td>something of a (9)</td>
<td>una sorta di (3)</td>
<td>un po' (1)</td>
</tr>
<tr>
<td></td>
<td>una specie di (1)</td>
<td>una certa (1)</td>
</tr>
<tr>
<td></td>
<td>una certa (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a suo modo (1)</td>
<td>[x] omission (1)</td>
</tr>
<tr>
<td>Total: 142</td>
<td>103</td>
<td>39</td>
</tr>
</tbody>
</table>
## Tabella dei dati

<table>
<thead>
<tr>
<th></th>
<th>Capriolo</th>
<th>Mantovani</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>una sorta di</em></td>
<td>69.9 %</td>
<td>10.2 %</td>
</tr>
<tr>
<td><em>una specie di</em></td>
<td>9.7 %</td>
<td>76.9 %</td>
</tr>
<tr>
<td>other</td>
<td>20.3 %</td>
<td>12.8 %</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>
Conclusions

- The analysis of parallel corpora should be carried out in conjunction with the use of other corpus resources in both source and target languages.
- Corpus-based studies of literary translation should combine quantity with quality → annotation.
- Corpus findings should be assessed against knowledge external to the corpus.
Thank you!