Contrastive Analysis, *Tertium comparationis*, and Corpora

Signe Oksefjell Ebeling & Jarle Ebeling
Outline

• Doing contrastive analysis
• Types of corpora: their pros and cons
• Case study
  – Based on comparable data only
  – Based on bidirectional parallel data
Our background/position

- Corpus-based contrastive analysis
- Starting point: perceived (dis)similarities
- Bi-/multidirectional, parallel corpora
- Translations seen as the best *tertium comparationis*
- Interested in finding systemic differences / differences at the system level of languages, i.e. not primarily motivated by applied research


Awareness of research in neighbouring fields, e.g. descriptive translation studies and typology
Starting point of the CA

   
a) Perceived similarity of any kind between phenomenon X in language A and phenomenon Y in language B (cf. Chesterman 1998)
   
b) What is the nature of the similarity (form, meaning, function)?
   
c) Describe the relationship between X and Y in the compared languages; or as is more often the case, the relationship between X in language A and Y₁, Y₂, Y₃, etc. in language B
   
d) Use the description to enrich knowledge of the individual languages and/or the relationship between the languages compared
Starting point of the CA

2. Perceived (quantitative) dissimilarities/difference between original and translated text in the same language

3. Frequent omissions and/or additions in the translation

4. Exploratory by e.g. starting from a unit or item (e.g. word, frame, pattern, construction) in one language and classify the correspondences in one or more other languages

Close, qualitative scrutiny of the differences is necessary.
Other types of corpora used in (other types of) contrastive analysis

- Comparative corpora (regional, dialectal comparisons; also diachronic studies of same language (family)?)
- Translation-only corpora (mono-, bi- or multilingual)
- Multimodal (gestures, images, descriptions/interpretation of situations)
- …
<table>
<thead>
<tr>
<th>Comparables corpora</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Not restricted to translated text types</td>
<td>• Criteria for comparability / TC?</td>
</tr>
<tr>
<td></td>
<td>• More readily available</td>
<td>• No alignment possible</td>
</tr>
<tr>
<td></td>
<td>• Comparison of orig. language</td>
<td>• Cannot reveal sets of cross-linguistic correspondences</td>
</tr>
<tr>
<td>Parallel corpora</td>
<td>Unidirectional translation corpora</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Alignment is possible</td>
<td>• Restricted range of text types</td>
</tr>
<tr>
<td></td>
<td>• Meaning &amp; function constant across the languages <strong>(a relatively sound TC is present)</strong></td>
<td>• Translation effects:</td>
</tr>
<tr>
<td></td>
<td>• Possible to discover (sets of) cross-linguistic correspondences (‘translation paradigms’)</td>
<td>• (i) Traces of source language in translated texts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• (ii) Traces of the translation process, including errors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balanced bi-/multidirectional corpora</td>
<td>Same as unidirectional translation corpora PLUS:</td>
<td>Even more restricted range of text types</td>
</tr>
<tr>
<td></td>
<td>• Possible to check translations in both directions (control for translation effects)</td>
<td>• Achieving balance of text types between directions of translations</td>
</tr>
<tr>
<td></td>
<td>• Comparison of orig. language</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• A sounder TC is present</td>
<td></td>
</tr>
</tbody>
</table>

Short case study

• **Starting point:**
  – Previous contrastive study of two similar-looking patterns in English and Norwegian: *for sake*/*for * skyld* (* = genitive) (Ebeling & Ebeling 2014)

• **Today’s experiment:**
  – Carve up previous study in a different way to demonstrate different TCs:
    • Sameness of form in a comparable corpus;
    • Translation correspondence in a bidirectional parallel corpus.

• **The corpora:**
  – The English and Norwegian original, comparable texts of the English-Norwegian Parallel Corpus+;
  – The English and Norwegian original and translated, comparable texts of the English-Norwegian Parallel Corpus+
Preliminaries
Perceived similarity

for the purpose of (something)
(1) We stayed together for appearances' sake, … (PeRo2E)
(2) … det videre søket etter Leike ville bare være for syns skyld. (JoNe2N)

out of consideration for or in order to help or please someone
(3) Just for a while, then, another chapter or two — for Miriam's sake. (PaAu1E)
(4) … men for Mathias' skyld, herregud, de hadde jo et barn sammen! (JoNe1N)

to express impatience, annoyance, urgency, or desperation (expletive use)
(5) For God's sake, stop! (MiWa1E)
(6) Si noe, for Guds skyld! (LSC2)

(Oxford Dictionaries Online and Altenberg 1982)
Comparable version of study

• Preferred use in English originals:
  – Expletive>Consideration>Purpose

• Preferred use in Norwegian originals:
  – Purpose>Consideration>Expletive
### Summary of the comparable study with ref. to Sinclair’s (1996) Extended-units-of-meaning model

<table>
<thead>
<tr>
<th></th>
<th>English (Core: <em>for sake</em>)</th>
<th>Norwegian (Core: <em>for skyld</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collocation</td>
<td>for [God’s / Christ’s / heaven’s] sake</td>
<td>for [sikkerhets] skyld</td>
</tr>
<tr>
<td>Semantic preference</td>
<td><strong>Annoyance</strong> (shown by words such as God’s (religion) / fuck’s (sex) or “pretend swear words” such as goodness’)</td>
<td><strong>Purpose</strong> (shown by words such as sikkerhets (‘safety’s’) / enkelhets (‘simplicity’s’)</td>
</tr>
<tr>
<td>Semantic prosody</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### How to get at the closest correspondences in the other language?

- E.g.: How is the expletive use of *for sake* typically expressed in Norwegian if not with the formally similar pattern?
Bidirectional version of study

Translations of *for sake* and *for skyld* in the ENPC+ (raw numbers)

<table>
<thead>
<tr>
<th>Congruent / non-congruent</th>
<th>EO &gt; NT</th>
<th>NO &gt; ET</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>for sake = for skyld</em></td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td><em>for sake/skyld = 'other'</em></td>
<td>58</td>
<td>46</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>64</td>
</tr>
</tbody>
</table>

**Congruent** = formally similar correspondences (i.e. *for sake = for skyld*)

**Non-congruent** = formally dissimilar correspondences (e.g. *for God’s sake = Herregud* (‘Lordgod’))

Mutual correspondence (Altenberg 1999) of *for sake* and *for skyld* = 30.2%
Examples of main tendencies

English expletive $\rightarrow$ non-congruent

(7) "They're just traffic cones, for fuck's sake." (PeRo2E)
"Det er jo bare trafikkjegler, for faen." (PeRo2TN)
Lit.: … for the devil

English consideration $\rightarrow$ congruent

(8) For my sake. (JB1)
For min skyld. (JB1T)

English purpose $\rightarrow$ congruent

(9) We stayed together for appearances' sake … (PeRo2E)
Vi holdt sammen for syns skyld … (PeRo2TN)
Norwegian expletive → congruent

(10) - Slipp han nå for guds skyld ned. (PePe1N)
    "Put him down, for God's sake." (PePe1TE)

Norwegian consideration → congruent

(11) ... hun gjorde det for pappas skyld ... (PeRy1N)
    ... she was doing it for Dad's sake ... (PeRy1TE)

Norwegian purpose → non-congruent

(12) Han rygget et skritt for sikkerhets skyld. (KaFo1N)
    Lit.: ... for safety’s sake
    He retreated a step, just to be on the safe side. (KaFo1TE)
Non-congruent English correspondences (translations and sources) of the most frequent Norwegian purpose uses with *skyld*

<table>
<thead>
<tr>
<th>for sikkerhets skyld ('for safety’s sake')</th>
<th>for moro skyld ('for fun’s sake')</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>to be on the safe side</td>
<td></td>
<td></td>
</tr>
<tr>
<td>just in case</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- for N (fun/entertainment/pleasure)
- a matter of N
- for the sake of N
Non-congruent Norwegian correspondences (translations and sources) of the most frequent English expletives with *for sake*

<table>
<thead>
<tr>
<th>for Christ’s sake</th>
<th>for fuck’s sake</th>
<th>for God’s sake</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>for svingende</em> (lit. ‘for swinging’)</td>
<td><em>for svingende</em> (lit. ‘for swinging’)</td>
<td><em>for svingende</em> (lit. ‘for swinging’)</td>
</tr>
<tr>
<td><em>herregud</em> (‘God’, lit. ‘lordgod’)</td>
<td><em>herregud</em> (‘God’, lit. ‘lordgod’)</td>
<td><em>herregud</em> (‘God’, lit. ‘lordgod’)</td>
</tr>
<tr>
<td>helvete (‘hell’)</td>
<td><em>i herrens navn</em> (‘in the lord’s name’)</td>
<td><em>i herrens navn</em> (‘in the lord’s name’)</td>
</tr>
<tr>
<td><em>(for) faen</em> (‘(for) the devil’)</td>
<td><em>(for) faen</em> (‘(for) the devil’)</td>
<td><em>(for) faen</em> (‘(for) the devil’)</td>
</tr>
</tbody>
</table>
Summary of the bidirectional parallel study

• Uncover different preferences of use in English and Norwegian (incl. different extended-units-of-meaning) through both comparable and translation data.

• Translation as a TC offered corresponding, and arguably more equivalent, expressions of *sake / *skyld in the other language.
  - E.g. typical Norwegian expletives corresponding to the expletive *sake pattern emerged.
  - E.g. typical English expressions of purpose corresponding to the purpose *skyld pattern emerged.

• The patterns are cross-linguistically most similar to each other in the consideration use.

• Gained a deeper understanding of the patterns compared.
### Summing up: CA, TC and corpora

<table>
<thead>
<tr>
<th>Comparable only</th>
<th>Parallel (bidirectional) (incl. omissions/additions in translated text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA based on researcher's bilingual knowledge, dictionaries, grammars:</td>
<td>CA based on researcher's bilingual knowledge, dictionaries, grammars, and translators' bilingual knowledge:</td>
</tr>
<tr>
<td>• Perceived (dis)similarity</td>
<td>• Perceived (dis)similarity</td>
</tr>
<tr>
<td>• predefined items/ categories</td>
<td>• predefined items/ categories</td>
</tr>
<tr>
<td>CA based (primarily) on translators' bilingual knowledge/competence:</td>
<td>CA based (primarily) on translators' bilingual knowledge/competence:</td>
</tr>
<tr>
<td>• Exploratory: predefined item with translation paradigm</td>
<td>• Exploratory: predefined item with translation paradigm</td>
</tr>
<tr>
<td>• Exploratory: frequency differences in original vs. translated texts (i.e. undefined starting point)</td>
<td>• Exploratory: frequency differences in original vs. translated texts (i.e. undefined starting point)</td>
</tr>
</tbody>
</table>

**Point for discussion**

Are there other/better ways of doing corpus-based CA (regardless of object of study)?
References


Hasselgård, Hilde. 2017. Corpus-based contrastive analysis: some methodological issues and a talking case study. Guest lecture given on 5 December 2017, Charles University, Prague.


References cont.


*Oxford Dictionaries Online* (https://www.oxforddictionaries.com/)

The corpus “permits contrastive studies […] based on original texts and translations as well as on parallel original texts” (Johansson & Hofland 1994: 27), i.e. the corpus can be seen both as a bidirectional translation corpus and a comparable corpus.
Un-/discover differences and similarities

How to discover and investigate lexical gaps, e.g. Norw. 'jo' (*but, just, only, in fact, of course, after all, Ø*)

Language A: English originals
Language A: English translations of Norwegian originals (*tertium comparationis*)
Language B: Norwegian originals

*Quantitative differences in a balanced corpus: Reason/Cause?*

*mind: 726 occ.*  
*mind: 363 occ.*

*translate* "carry across" ideational and interpersonal functions and stylistic features