The CroCo Project
Cross-linguistic corpora for the investigation of explicitation in translations

Stella Neumann & Silvia Hansen-Schirra
Saarland University, Germany
Overview

- Systematic specificities of translated text
- The property of „explicitation“
- Corpus design issues
- Corpus representation
- Summary and outlook
Motivation
systematic specificities of translated text

Corpus-based comparison of translations with originals
- Simplification
- Normalisation
- Levelling out
- Shining through
- Explicitation
The property „explicitation“

- Stylistic translation technique (Vinay & Darbelnet 1958/1995)
- Blum-Kulka 1986: explicit cohesion markers in translations → increased text length
- Baker 1996: explicitation as possible translation universal
- Olohan & Baker 2000: explicitation of optional *that* in raw corpora
What exactly is explicitation?

An element is explicitated in the target text which must have been somehow implicitly present, i.e. linguistically traceable, in the source text.

English ST [on the Canadian city of Calgary]:
The tower also contains the city’s tourist office.

Modified German TT:
Im Turm ist auch das Tourismusbüro der Stadt Calgary untergebracht.

English gloss:
In the tower is also the tourism office of the city of Calgary accommodated.
What exactly is explicitation?

An element is explicitated in the target text which must have been somehow implicitly present, i.e. linguistically traceable, in the source text.

German ST:
Er erwähnte die Anrufe der beiden Männer...

English gloss:
He mentioned the calls of the two men...

English TT:
He mentioned that both, father and son, had called him...
Factors for the analysis of explicitation

- Explanatory background
  - Language typology
  - Text typology
  - Translation process

- „Explicitation“ as measured by operationalised criteria

- Linguistic level
  - Lexical
  - Grammatical
  - Semantic

- Scope
  - Language pair-specific
    - EN-DE
  - Language pair-specific
    - DE-EN
  - Universal
Corpus design

- **English texts**
  - Reference Corpus ER
  - Register-controlled Corpus EO
  - Translation Corpus GTrans
  - Translation Corpus ETrans
  - 17 registers, 2,000 word samples each
  - 8 registers, at least 10 texts each, 3125 words (av.)
  - 68,000 words

- **German texts**
  - Reference Corpus GR
  - Register-controlled Corpus GO
  - 1 million words
Multilingually comparable corpora
Monolingually comparable corpora

- English texts
  - Reference Corpus ER
  - Register-controlled Corpus EO
  - Translation Corpus ETrans

- German texts
  - Reference Corpus GR
  - Register-controlled Corpus GO
  - Translation Corpus GTrans
Parallel corpora

- English texts
  - Reference Corpus ER
  - Register-controlled Corpus EO

- German texts
  - Reference Corpus GR
  - Register-controlled Corpus GO

- Translation Corpus GTrans
- Translation Corpus ETrans
Selection of texts for parallel corpora

8 registers which
• are relevant for translation
• foreground different register variables
  - Popular scientific texts
  - Tourism leaflets
  - Prepared speeches
  - Political essays on economics
  - Fictional texts
  - Corporate communication
  - Instruction manuals
  - websites
Alignment of raw corpora

English texts

- Register-controlled Corpus EO

German texts

- Register-controlled Corpus GO
- Translation Corpus GTrans
- Translation Corpus ETrans
XML stand-off mark-up

- Corpus
- TSV
- SGML
- TXT
- Layer 1
- Layer 2
- Layer 3
- Indexed corpus

Linked through unique IDs in XML.
Alignment of annotation

Translational Corpus GTrans
Linked through unique IDs

G Layer 1 — E Layer 1
G Layer 2 — E Layer 2
G Layer 3 — E Layer 3

Alignment of translation units
Linked through unique IDs

Register-controlled Corpus EO
 Alignment of annotation units
Example

<nondiscriminatory>trade</nondiscriminatory>

<Gleichbehandlung>Handel</Gleichbehandlung>
Example

cf. the Lincoln road vs. the road to Lincoln (Quirk et al. 1985:1330)
Explicitation indicators

- Relation between content and function words
- Relation between non-finite vs. finite verbs
- Realisation of constituents
- etc.
Summary and outlook

• Creation of a dynamic resource for corpus research in translation studies
• Alignment not only of raw text but also of annotation
• Combination of linguistic approach with rich XML annotation
• Corpus resource available for other researchers
http://fr46.uni-saarland.de/index.php?id=croco