New insights from the study of translations

Stella Neumann*, Silvia Hansen-Schirra°, Kerstin Kunz*, Erich Steiner* & Mihaela Vela*

*Saarland University, Saarbrücken
°Johannes Gutenberg University, Mainz/Germersheim

Project No. STE 840/5-2 sponsored by DFG
Overview

- Motivation
- Empirical investigations with the CroCo Corpus
- Some possible insights
- Conclusion and Outlook
Motivation

- Toury’s (1995) branches of translation studies
  - Type of relationship: possible, probable
  - Criterion: theoretical, conditional
  - “Pure”
  - Theoretical
  - Descriptive
  - Translation Studies
  - Applied
  - Type of relationship: required
  - Criterion: postulated
  - Empirical studies: experimental or quantitative
  - Peculiarities of translations identified on the basis of a quantitative investigation of large amounts of translations
What kind of peculiarities?

- Properties of translations → e.g. explicitation (e.g. Olohan & Baker 2000)
- Detecting contrast-induced translation solutions → e.g. cleft constructions (cf. Hawkins 1986)
- Translation shifts → e.g. transposition (Vinay & Darbelnet 1995, Newmark 1988, van Leuven-Zwart 1989 („syntactic-semantic modification“))
- Register-specific language use in originals and translations → LSP grammar (cf. Halliday & Martin 1993)
- (Outlook: translation unit (cf. Malmkjaer 2006, Kondo 2007))
Why not corpus-driven research?

Olohan & Baker (2000): optional that in TEC and BNC

Operationalisation of explicitation:
Concordances of forms of *say* and *tell* and the string following them

Hypothesis
If the complementizer *that* occurs significantly more frequently after the verbs in TEC than in comparable parts of the BNC → explicitation

Hypothesis is confirmed
→ limitations of string-based approach
→ measures explicitness rather than explicitation

comparable texts display more or less explicitness
parallel texts display more or less explicitation
The CroCo Corpus
Overview 1

- Analysis of corpora, not examples
  → measurable linguistic evidence to establish properties

- Analysis of different translation relevant registers
  → register-specific distinctions available

- Aligned source and target texts
  → to distinguish between comparable and parallel texts

- Annotation of the corpora
  → to analyse lexico-grammatical and cohesive indicators
The CroCo Corpus
Overview 2 (cf. Neumann & Hansen-Schirra 2005)

English texts

Reference Corpus ER
Register-controlled Corpus EO
Translation Corpus GTrans
Register-controlled Corpus GO
Translation Corpus ETrans

German texts

Reference Corpus GR

17 registers, 2,000 word samples each
8 registers, at least 10 texts each, 3,125 words (av.)
68,000 words
1 million words
The CroCo Corpus
Multi-layer annotation and alignment

Translation Corpus

+ Metainformation
+ Tokenization
+ PoS tagging
+ Morphology

Word layer

+ Phrase structure
+ Grammatical functions

Chunk layer

Clause layer

Sentence layer

Register-controlled Corpus

+ Alignment
1. Properties of translations
Cohesive explicitation/implicitation

Operationalization: proportion of explicit to implicit referents in aligned source and target sentences

Der Pkw-Markt in Deutschland ist der einzige wesentliche Markt, in dem ein Rückgang der Auslieferungen an Kunden zu verzeichnen war.
The German passenger car market was the only major market to see a decline in deliveries to customers.

Implicit reference AND no tense, mood, voice

cf. Hansen-Schirra et al. (in print)
2. Contrast-induced translation solutions
cleft constructions

Detect existing solutions for grammatical translation problems on the basis of language typological differences (cf. Hawkins 1986)
Available in both languages but more frequent in English

In EN source text: word=“it” FOLLOWED BY lemma=“be”
(FOLLOWED BY grammatical function=“complement” (INCLUDING part-of-speech =“relative pronoun”))

*It is this ownership that* we truly believe helped our employees to drive toward success, despite the challenges of this year.

Mit dieser Beteiligung am Unternehmen im Rücken haben unsere Mitarbeiter nach unserer Überzeugung maßgeblich zum Erfolg des Unternehmens trotz der großen Herausforderungen dieses Jahres beigetragen.

cf. Vela et al. (2007)
3. Translation shifts
Transpositions in SHARE

Operationalization: aligned lexical words tagged with different parts of speech

EN>GE: 40,090 aligned lexical word pairs, 1,411 shifts (3.52%)
GE>EN: 37,694 aligned lexical word pairs, 1,572 shifts (4.17%)

![Diagram showing transposition percentages and types for EN>GE and GE>EN translations.]

- Verb > noun: 42.95% (EN>GE), 39.69% (GE>EN)
- Noun > verb: 20.03% (EN>GE), 42.95% (GE>EN)
- Adjective > verb: 40% (EN>GE), 40% (GE>EN)
- Adjective > noun: 20% (EN>GE), 20% (GE>EN)
- Adverb > verb: 60% (EN>GE), 60% (GE>EN)
- Adverb > noun: 40% (EN>GE), 40% (GE>EN)
### 3. Translation shifts

**Example verb to noun transposition in E2G_SHARE**

<table>
<thead>
<tr>
<th>word EO</th>
<th>PoS tag</th>
<th>word GTrans</th>
<th>PoS tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>do</td>
<td>vd0</td>
<td>Handeln</td>
<td>nn</td>
</tr>
<tr>
<td>play</td>
<td>vv0</td>
<td>Spielen</td>
<td>nn</td>
</tr>
<tr>
<td>work</td>
<td>vv0</td>
<td>Arbeiten</td>
<td>nn</td>
</tr>
<tr>
<td>programming</td>
<td>vvg</td>
<td>Programme</td>
<td>nn</td>
</tr>
<tr>
<td>communicate</td>
<td>vv0</td>
<td>Kommunikation</td>
<td>nn</td>
</tr>
<tr>
<td>believe</td>
<td>vv0</td>
<td>Auffassung</td>
<td>nn</td>
</tr>
<tr>
<td>computing</td>
<td>vvg</td>
<td>Computers</td>
<td>nn</td>
</tr>
<tr>
<td>compared</td>
<td>vvn</td>
<td>Vergleich</td>
<td>nn</td>
</tr>
<tr>
<td>learn</td>
<td>vv0</td>
<td>Lernen</td>
<td>nn</td>
</tr>
<tr>
<td>enters</td>
<td>vvvz</td>
<td>Schwelle</td>
<td>nn</td>
</tr>
<tr>
<td>integrate</td>
<td>vv0</td>
<td>Integration</td>
<td>nn</td>
</tr>
<tr>
<td>develop</td>
<td>vv0</td>
<td>Entwicklung</td>
<td>nn</td>
</tr>
<tr>
<td>browsing</td>
<td>vvg</td>
<td>Browsen</td>
<td>nn</td>
</tr>
<tr>
<td>manage</td>
<td>vv0</td>
<td>Verwalten</td>
<td>nn</td>
</tr>
<tr>
<td>connect</td>
<td>vv0</td>
<td>Verbindung</td>
<td>nn</td>
</tr>
<tr>
<td>control</td>
<td>vv0</td>
<td>Kontrolle</td>
<td>nn</td>
</tr>
</tbody>
</table>

Whether you want to *communicate, learn, work, or play*, the PC can enrich and improve the experience.

Ganz gleich, ob Sie ein Hilfsmittel zur Kommunikation oder zum Lernen, Arbeiten oder Spielen benötigen, der PC kann diese Erfahrung eindringlicher und besser gestalten.

Non-finite constructions → Nominalizations
4. Register-specific language use
LSP grammar

Operationalization: Proportion of words per sentence/clause/chunk and clauses per sentence

Hypothesis: specialized registers contain higher proportions, SHARE texts more specialized than FICTION texts

→ Hypothesis seems confirmed
5. Translation unit

Overview

Answers available from a corpus-based study:
Which linguistic unit is most frequently aligned?

- Words?
- Phrases?
- Clauses?
- Sentences?
With this important work, AICGS will remain the outstanding institution of German Studies on this side of the Atlantic.

As our first year as Co-Chairmen has ended, we are confident that the work of the Institute will continue to contribute to the German-American dialogue.

Durch diese wichtige Arbeit wird das AICGS eine bedeutende Institution fuer deutsche Studien auf dieser Seite des Atlantiks bleiben.

Nach unserem ersten Jahr als Aufsichtsratsvorsitzende sind wir der Ueberzeugung, dass die Arbeit des Instituts weiterhin zum deutsch-amerikanischen Dialog beitragen wird.
With this important work, AICGS will remain the outstanding institution of German Studies on this side of the Atlantic.

As our first year as Co-Chairmen has ended, we are confident that the work of the Institute will continue to contribute to the German-American dialogue.

Durch diese wichtige Arbeit wird das AICGS eine bedeutende Institution für deutsche Studien auf dieser Seite des Atlantiks bleiben.

Nach unserem ersten Jahr als Aufsichtsratsvorsitzende sind wir der Überzeugung, dass die Arbeit des Instituts weiterhin zum deutsch-amerikanischen Dialog beitragen wird.
5. Translation unit
Aligned clauses = translation units

| Translation Direction | SHARE
|-----------------------|--------
| EO                    | 76.7% |
| Gtrans                | 88.6% |
| GO                    | 91.2% |
| Etrans                | 73.8% |
Why does this matter for translation studies?

- Preparatory steps for an empirical theory of translation with the three steps of
  - Description
  - Explanation
  - Prediction

- Combination of product-oriented and process-oriented research as a requirement (triangulation, cf. Alves 2003)
  - Corpus-based research: high ecological validity
  - Psycholinguistic experiments: insight into cognitive aspects
Conclusion and outlook

- Wealth of information available on the basis of linguistic enrichment of corpora
  - Standard queries for translation problems due to contrastive differences
  - Research into specific topics of translation studies
  - Ultimate goal: empirical theory of translation

- Future work
  - Finish up annotation and alignment
  - Add semantic annotation
  - Create a query interface
  - Conduct psycholinguistic experiments
CroCo Project Web Site

http://fr46.uni-saarland.de/croco/