

## New insights from the study of translations

Stella Neumann\*,  
Silvia Hansen-Schirra<sup>o</sup>, Kerstin Kunz\*,  
Erich Steiner\* & Mihaela Vela\*

\*Saarland University, Saarbrücken

<sup>o</sup>Johannes Gutenberg University, Mainz/Germersheim

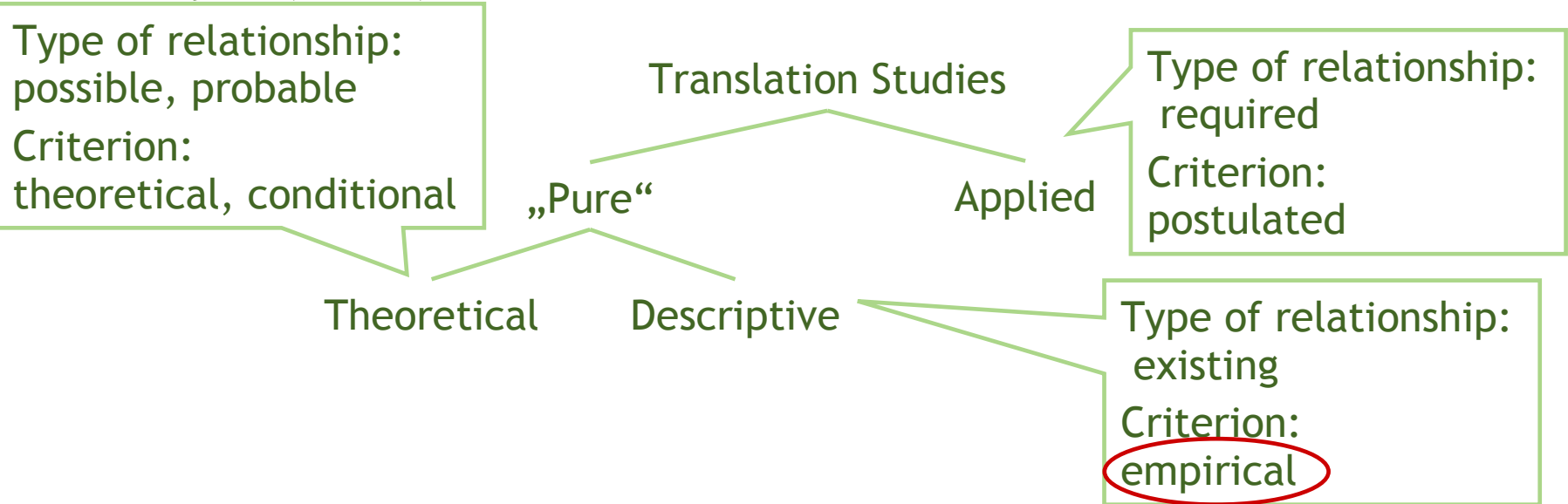
# Overview

---

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

# Motivation

- Toury's (1995) branches of translation studies



- 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

TEC Concordance Browser (v. 0.4)

File Options Plugins Help

Keyword says Search 1 Sort Left 1 Sort Right Extract Metadata

... will be done using a combination of technology and skills. He says that a media database is not just a collection of facts, but is [2657]  
nce, if only they had been available anywhere. However, Nokia says that its production capacity is now up to a level that can meet [2662]  
air is continuing to use MD-11s on its long-haul routes. Potila says that studies conducted during the summer have shown that the dec [2671]  
craft, ATR 72s and Saabs, are in use on domestic routes. Potila says that there will have to be gradual development towards only one [2672]  
hania collected from different parts of the Institute. Antonov says that only ten per cent of the Institute's work is now connected [2681]  
customers or employees unforgettable experiences. K7a7r7inen says that it took him many years to learn the right approach to selli [2685]  
nbourhood feelings of affinity with the Nordic peoples. Rajam7ki says that Nordics and Scots get on remarkably well together. "The Sc [2697]  
its result, but was still in the red. President Jorma Ollila says that the intention is to keep the thrust of Nokia's development [2698]  
ling but heresies. Our contradictor proved obstinate. "The law says that schooling is obligatory for all children." The onlookers sh [2706]  
a study of Virgil Thomson's works for a book. In no. 12, Cage says that in the end, he did not draft this article. He did, however, [2713]  
he other man came to explain to Tanios, in Arabic. 'His Honour says that if you continue to behave insolently, you'll be beaten and [2740]  
ng was leaving, a nurse rushed up to me. "Dr. Li, Zhang Yufeng says that Chairman wants to see you." I rushed to his side. Once a st [2754]  
. "You doctors!" Mao responded. "One of you says this; another says that. I am only going to listen to doctors seventy percent of th [2758]  
across the sea in search of the elixir of immortality. Legend says that the Japanese are their descendants. Shortly after I began w [2759]  
high-ranking patients. ECT, however, was never used. Wang Hebin says that Mao Anqing was sent for treatment to the Soviet Union for s [2776]  
leged wonder drug as procaine (trade name Novocain). The report says that a Dr. Anna Aslan experimented with the drug in Bucharest in [2779]  
partment was abolished: [2782]  
acked her lips in disg 'He says that his teacher hit him. ack him will be beaten abou [2786]  
ue you've actually got m in my next birth. Then a [2804]  
ill by himself; and leave the pots washed clean. His own mother says that! It seems he .can pollish off a half- measure of rice all by [2809]  
... And anyway, that fellow, I forget his name, is the one who says that the mind is capable of reversing the process of entropy, an [2860]  
at someone, while yours truly ... She talks about the book; she says that the book has set her free; that it's all thanks to me ... I [2862]  
l existence? The only drama in my life is having met Elsa. She says that drugs reveal a lot. To me, she reveals a lot. I'm just a de [2863]  
he bigger boy from her hip and sank down against the wall. 'He says that his teacher hit him. The wicked wretch.' Then she covered h [2874]  
ng to tell me what happened?" "It's enough to drive you mad," says the inspector, turning to Giovanni Marden. "There isn't even any [43]  
h fear. A single word is enough to frighten him. "Montani, " says the voice in the corner, "listen to me." How does that brigand [48]

and the skinny woman says the newspapers had to embroider the story,

... is, in a voice that's unexpectedly gentle, if not feminine. He says 'The merchandise is good.' He picks up a brooch, closes his left [122]  
sp. Before dropping me and the case at the bus-stop, the twin says the boss liked my face. And that his twin brother, who knows the [126]  
e glasses. Then the skinny woman backs up the psychologist and says the kid needs to work out her fantasies. My brother-in-law says [148]  
I nodded, wanting to change the subject, and the skinny woman says the newspapers had to embroider the story, because there's no in [156]

comparable texts display more or less explicitness

parallel texts display more or less explicitation

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

# The CroCo Corpus

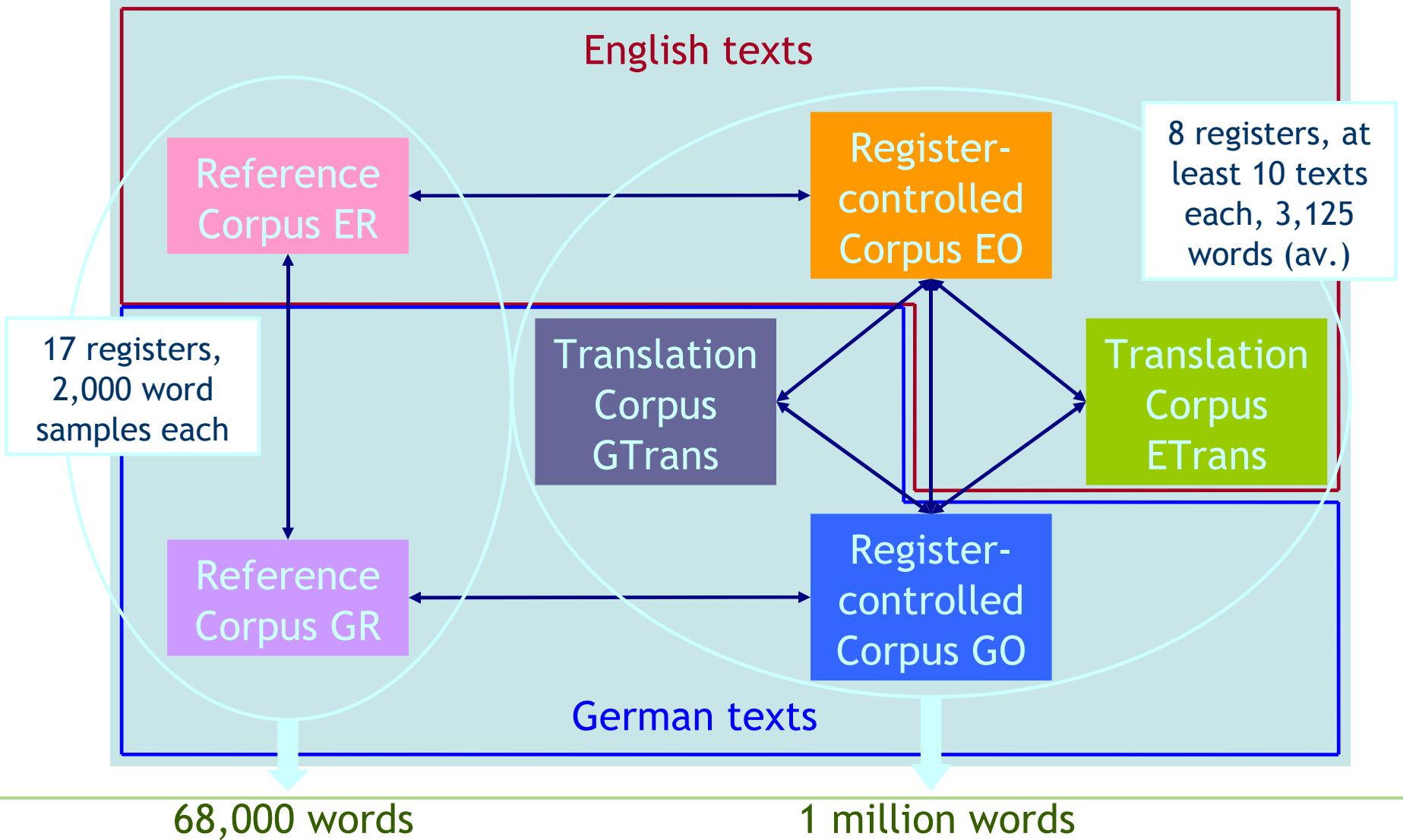
## Overview 1

---

- 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
-

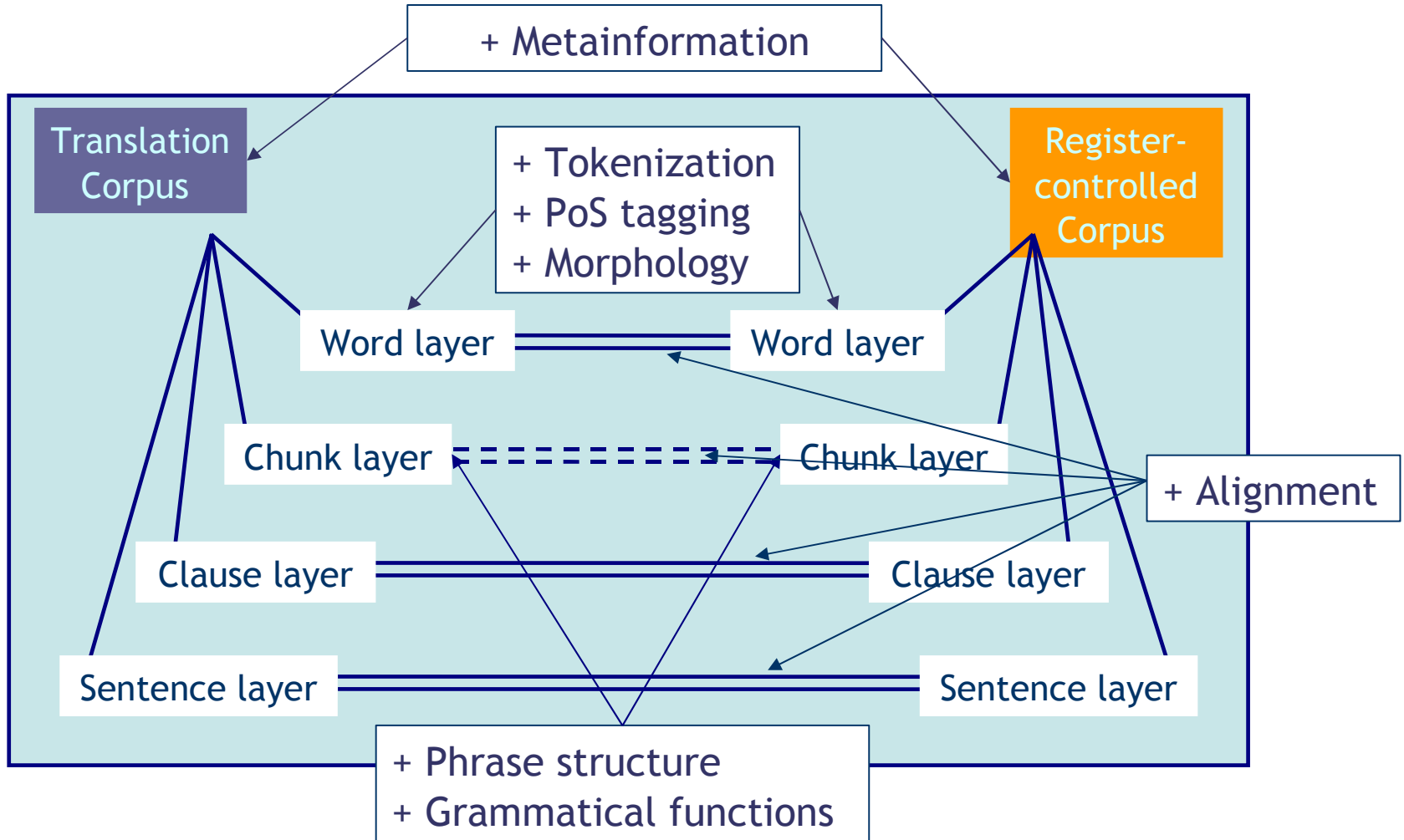
# The CroCo Corpus

Overview 2 (cf. Neumann & Hansen-Schirra 2005)



# The CroCo Corpus

## Multi-layer annotation and alignment

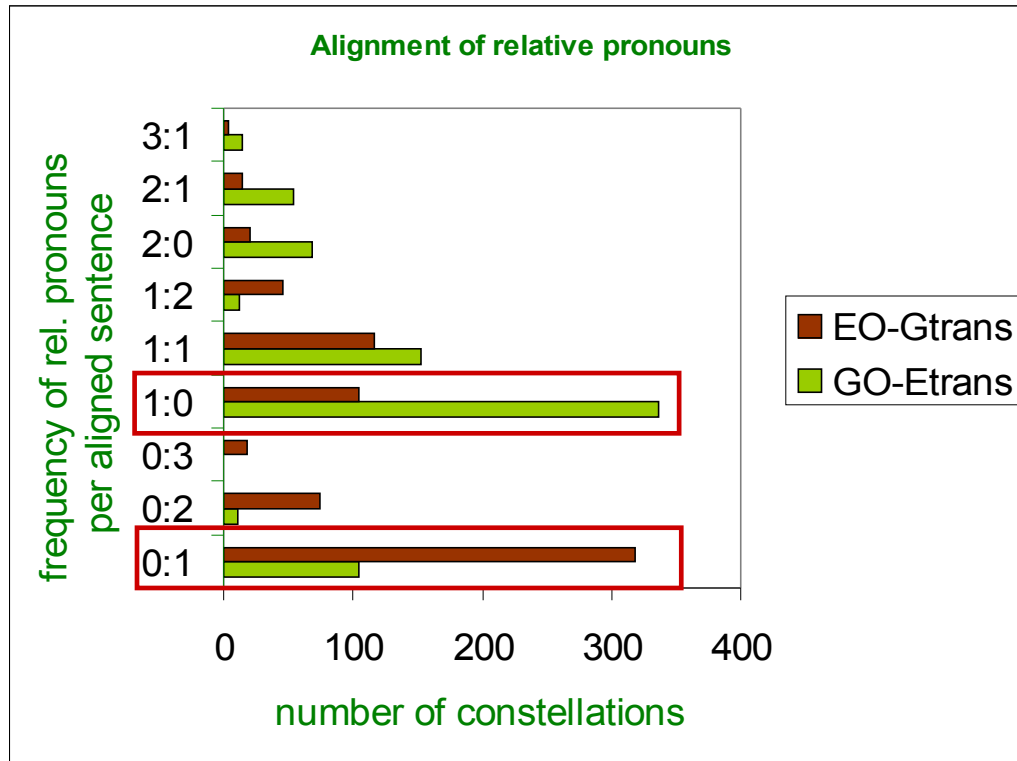




# 1. Properties of translations

## Cohesive explicitation/implicitation

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



Explicit reference

*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

## 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.*

Fronted adverbial in the form of a PP

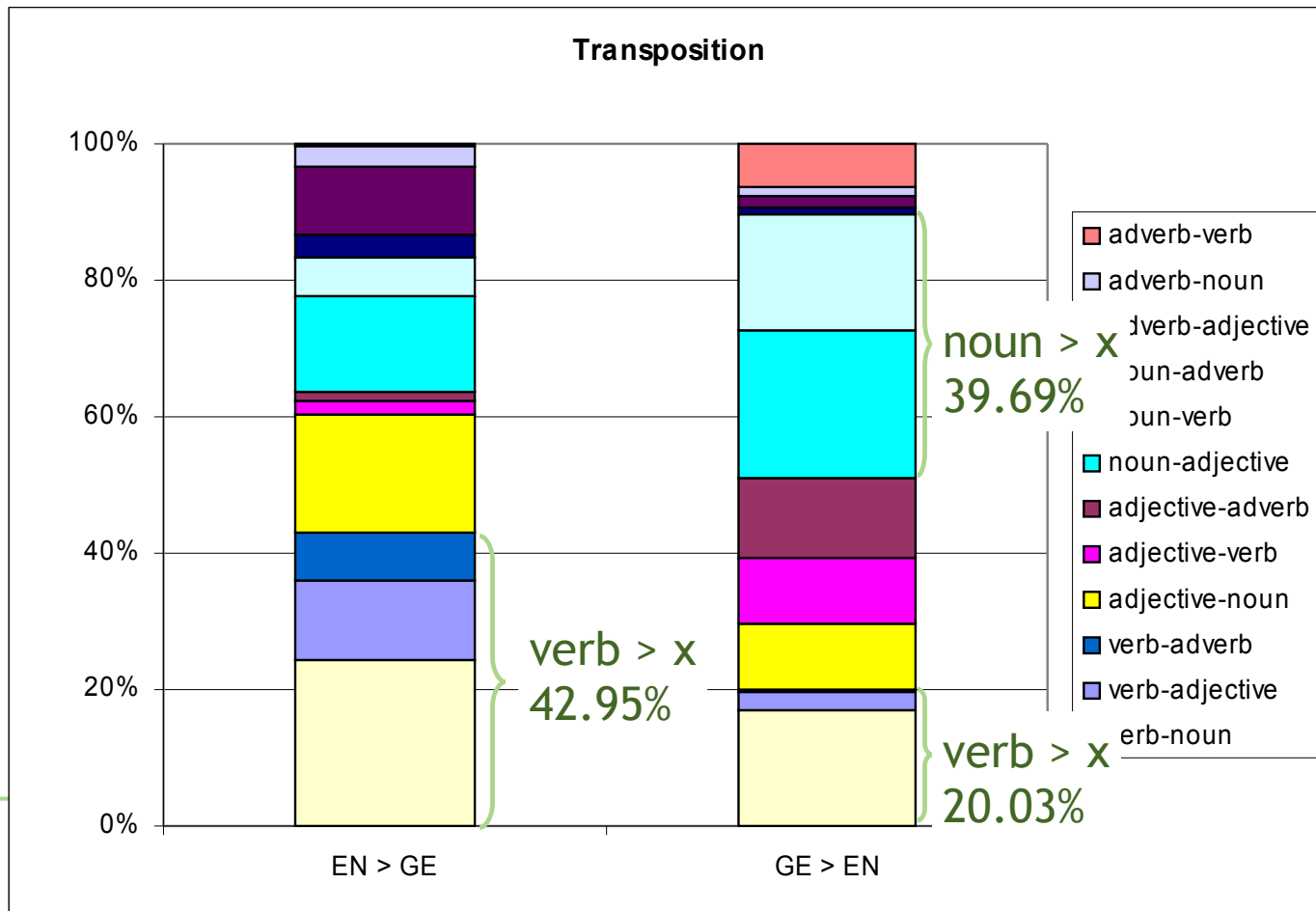
# 3. Translation shifts

## 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%)



# 3. Translation shifts

## Example verb to noun transposition in E2G\_SHARE

word EO	PoS tag	word GTrans	PoS tag
do	vd0	Handeln	nn
play	vv0	Spielen	nn
work	vv0	Arbeiten	nn
programming	vvg	Programme	nn
communicate	vv0	Kommunikation	nn
believe	vv0	Auffassung	nn
computing	vvg	Computers	nn
compared	vvn	Vergleich	nn
learn	vv0	Lernen	nn
enters	vvz	Schwelle	nn
integrate	vv0	Integration	nn
develop	vv0	Entwicklung	nn
browsing	vvg	Browsen	nn
manage	vv0	Verwalten	nn
connect	vv0	Verbindung	nn
control	vv0	Kontrolle	nn

*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.*

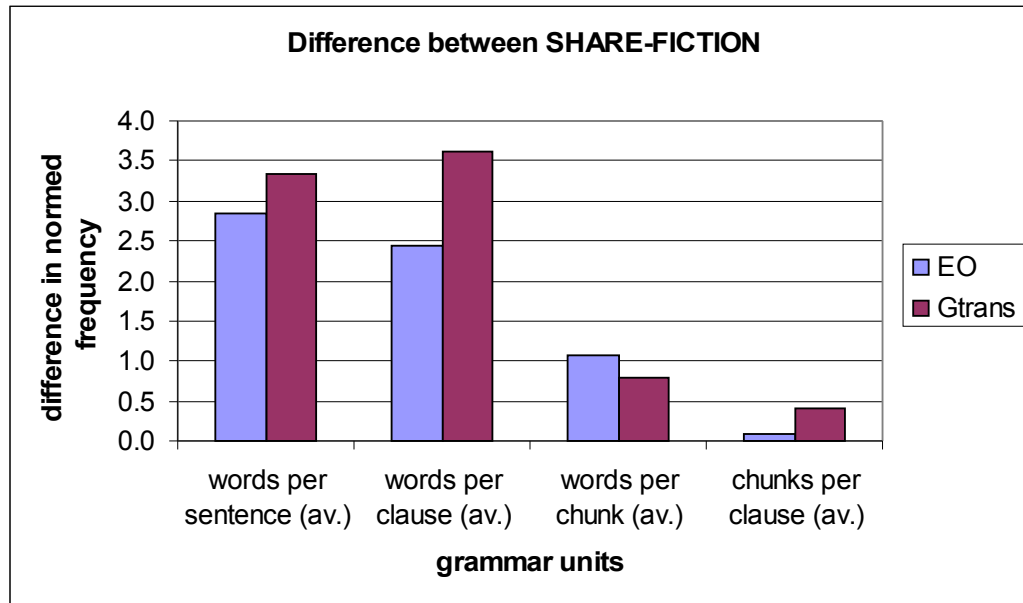
Non-finite constructions →  
Nominalizations

# 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?
-

# 5. Translation unit

## 1. sentence alignment (Trados/WinAlign)

s10

*With this important work, AICGS will remain the outstanding institution of German Studies on this side of the Atlantic.*



s10

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

s11

*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.*

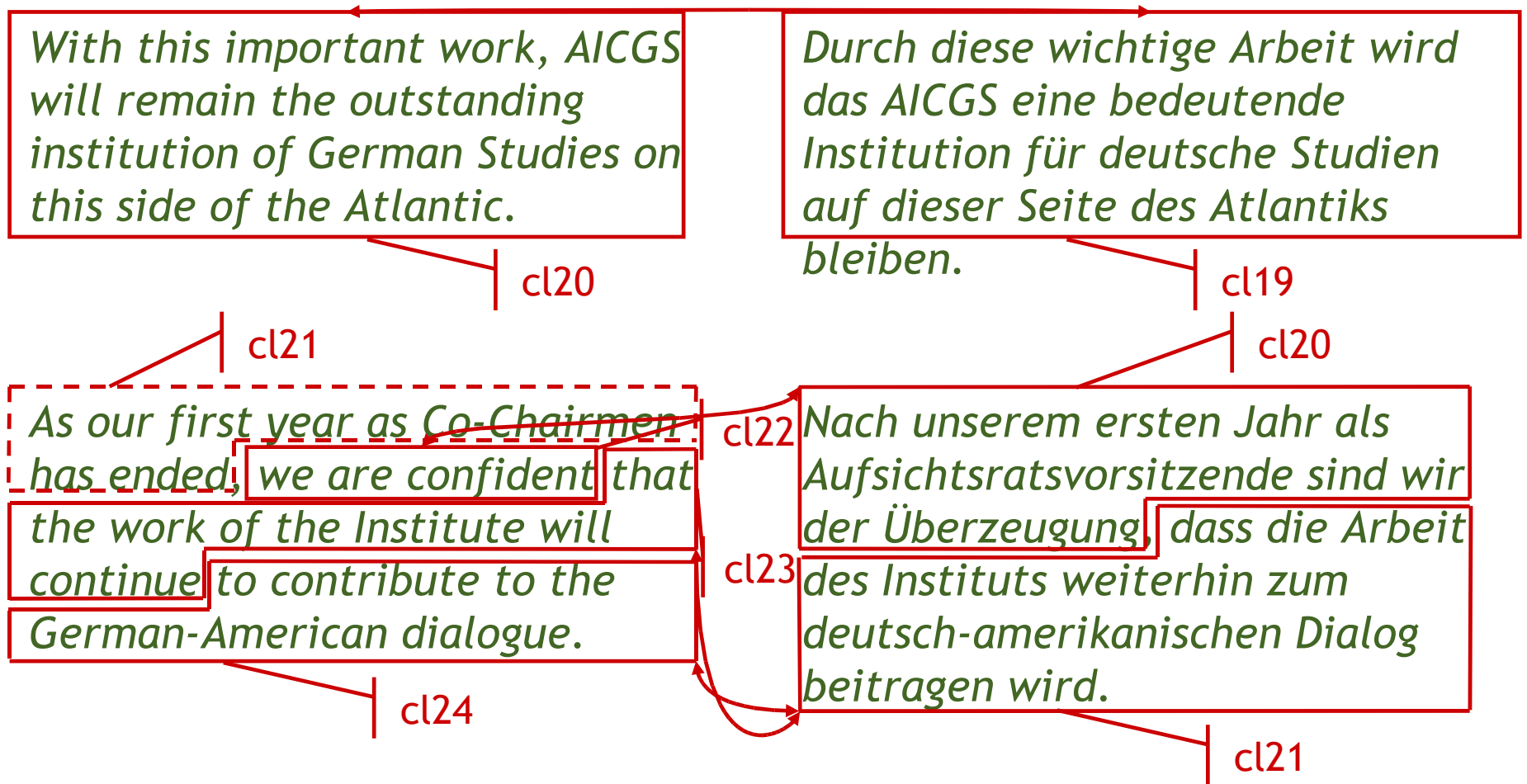


s11

*Nach unserem ersten Jahr als Aufsichtsratsvorsitzende sind wir der Ueberzeugung, dass die Arbeit des Instituts weiterhin zum deutsch-amerikanischen Dialog beitragen wird.*

# 5. Translation unit

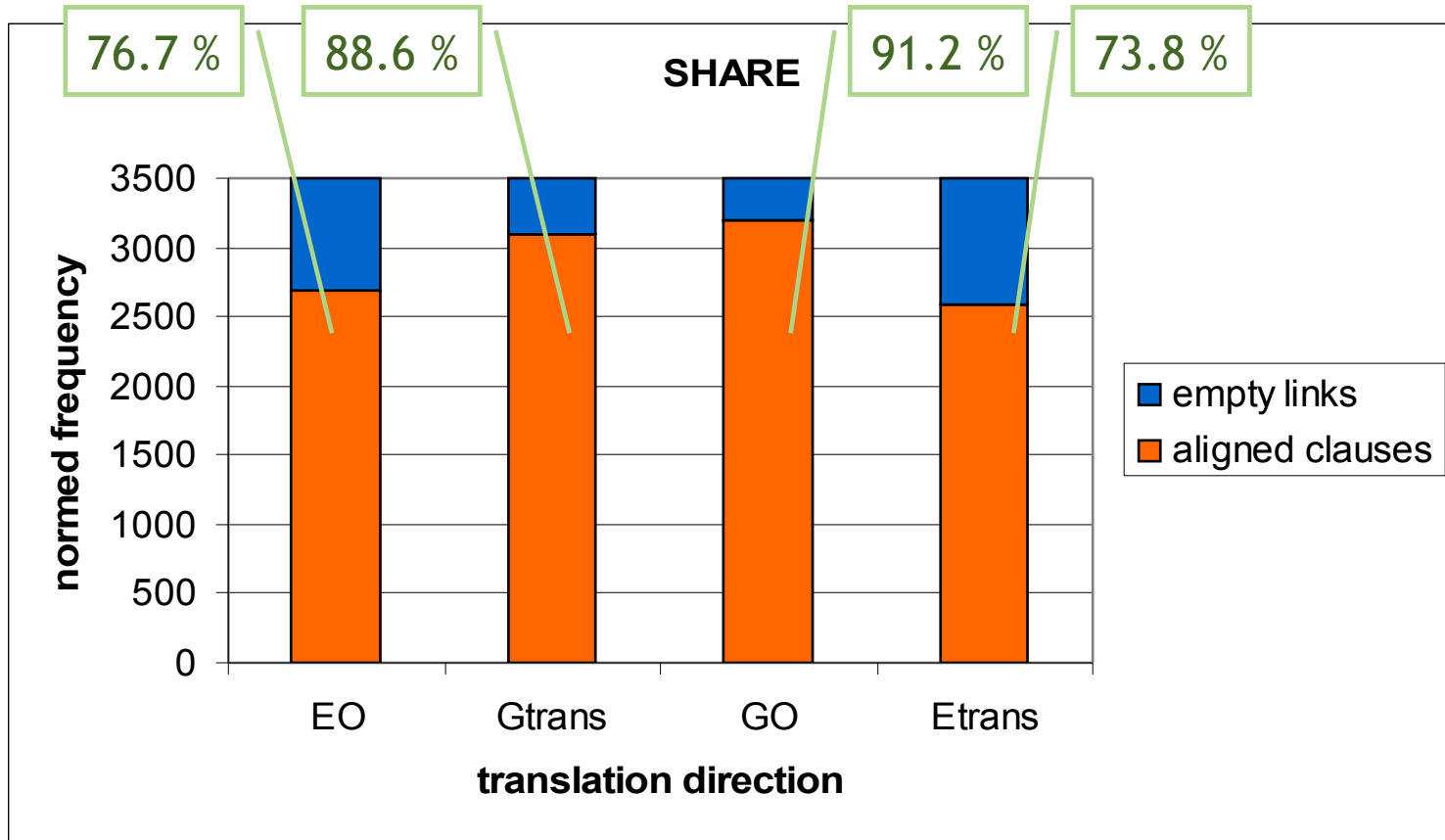
## 2. clause alignment (manual)





# 5. Translation unit

Aligned clauses = translation units



# 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/>

---