

CROCO

LINGUISTIC PROPERTIES OF TRANSLATIONS
A CORPUS-BASED INVESTIGATION FOR THE LANGUAGE PAIR ENGLISH-GERMAN

Cohesion and explicitation in an English-German translation corpus

Silvia Hansen-Schirra
Stella Neumann
& Erich Steiner

Overview

- Background of the study: explicitation and cohesion
 - Brief overview of the project
 - Deriving cohesion indicators for a quantitative study
 - Exemplary findings for the indicators
 - Conclusion and outlook
-

Background of the study

- Explicitness and explicitation:

We assume *explicitation* if in a translation (or language-internally in a pair of register-related texts) meanings (not only ideational, but including interpersonal and textual) are realized in the more explicit variant which are not realized in the less explicit variant, but which are in some theoretically-motivated sense implicit in the latter. The resulting text is more *explicit* than its counterpart.

- Cohesive devices as one gateway for explicitation in translations (e.g. Blum-Kulka 1986)

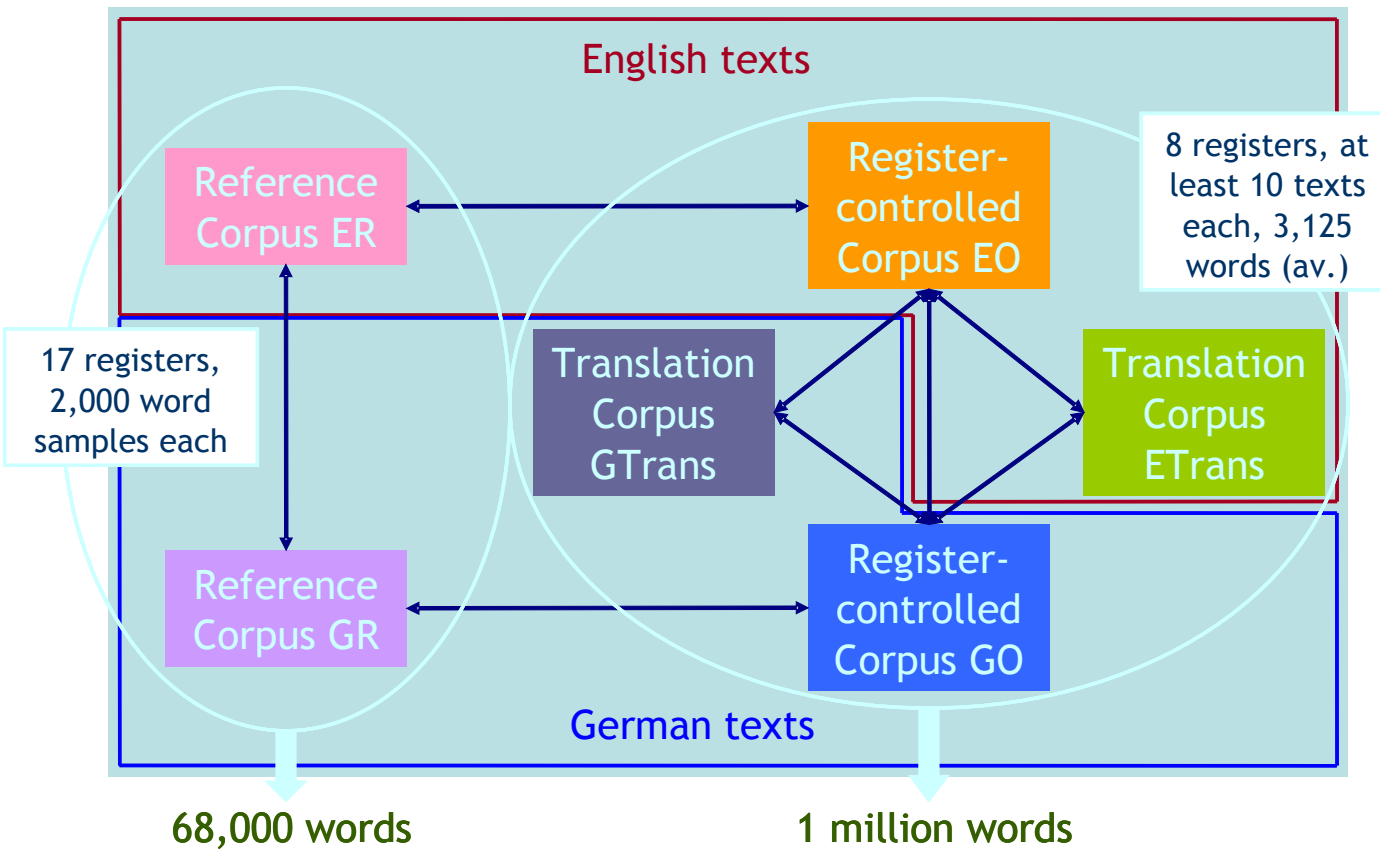
The CroCo research design

- Analysis of **corpora**, not examples
 - measurable linguistic evidence to establish properties
- **Aligned** source and target texts
 - to distinguish between explicitness and explicitation
- **Annotation** of the corpora
 - to analyse lexico-grammatical and cohesive indicators

Basic Principle of the study:

theory-neutral design, but theory-driven deduction of indicators

The CroCo Corpus



Annotation

- Metainformation
- Part of speech
- Morphology
- Phrase structure
- Grammatical functions

Alignment

- Sentence
- Clause
- (Functions)
- Words

Deriving indicators from abstract concepts

- Halliday & Hasan 1976: five cohesive devices for English; operationalization for quantitative studies needed
 - Reference
 - Proportion of explicit to implicit referents
 - Proportion of phoric to fully lexical phrases
 - Substitution and ellipsis
 - Amount of cohesive ellipsis and substitution
 - Conjunction
 - Explicit and implicit encoding of conjunctive relations
 - Lexical cohesion
 - Lexical density and type-token ratio
-

Reference

- Same referential meaning represented by (possibly) different wordings => implicit vs. explicit referents
- Indicator: pronouns

English original

... *a*_{t53} *palmist*_{t54, t55} *inferring*_{t56} *the*_{t57} *future*_{t58} *out*_{t59} *of*_{t60} *his*_{t61} *own*_{t62} *lined*_{t63}
*flesh*_{t64}

German translation

... *ein*_{t64} *Handle*_{t65}*rer*_{t66} *der*_{t67} *seine*_{t68} *Zukunft*_{t69} *aus*_{t70} *den*_{t71} *eigenen*_{t72} *Linien*_{t73}
*ableitete*_{t74}

token index file

<token id="t64" strg="ein"/>

<token id="t65" strg="Handle

<token id="t66" strg=","/>

<token id="t67" strg="der"/>

<token id="t68" strg="seine"/>

<token id="t69" strg="Zukunft"/>

part-of-speech annotation

<token pos="art" xlink:href="#t64"/>

<token pos="nn" xlink:href="#t65"/>

<token pos="yc" xlink:href="#t66"/>

<token pos="prels" xlink:href="#t67"/>

<token pos="pposat" xlink:href="#t68"/>

<token pos="nn" xlink:href="#t69"/>

Reference

- Same referential meaning represented by (possibly) different wordings => implicit vs. explicit referents
- Indicator: pronouns

English original

... *a*_{t53} *palmist*_{t54, t55} *inferring*_{t56} *the*_{t57} *future*_{t58} *out*_{t59} *of*_{t60} *his*_{t61} *own*_{t62} *lined*_{t63}
*flesh*_{t64}

German translation

... *ein*_{t64} *Handle*_{t65, t66} *der*_{t67} *seine*_{t68} *Zukunft*_{t69} *aus*_{t70} *den*_{t71} *eigenen*_{t72} *Linien*_{t73}
*ableitete*_{t74}

word alignment

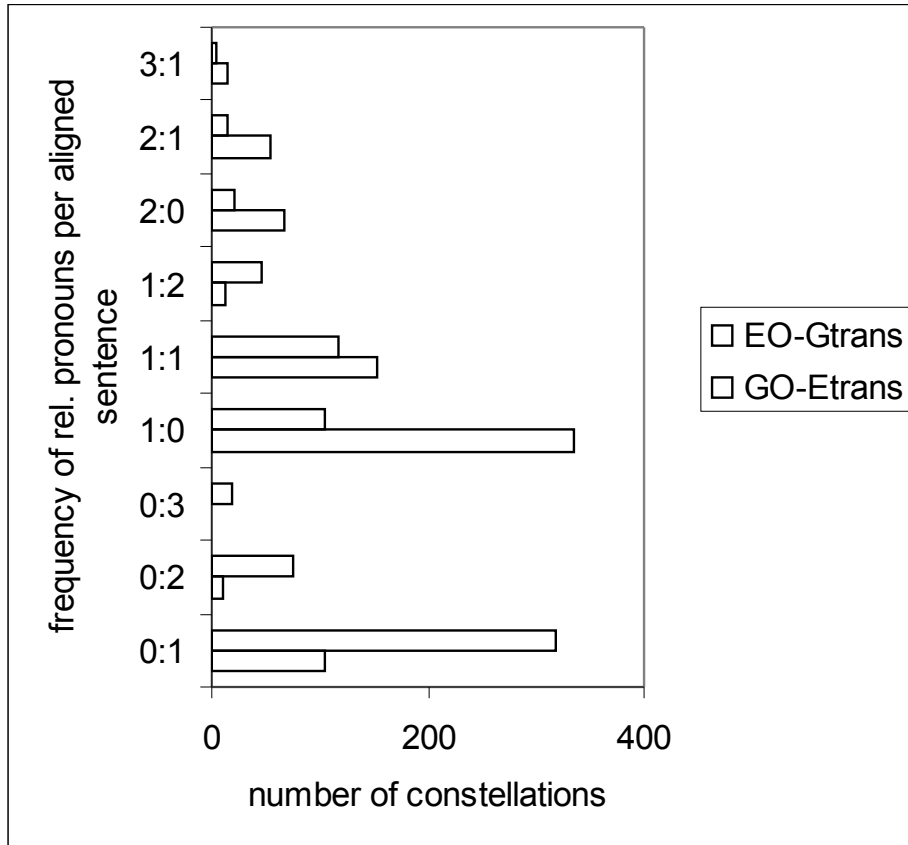
```
<token>  
  <align xlink:href="#t55"/>  
  <align xlink:href="#t66"/> </token>  
<token>  
  <align xlink:href="#t56"/>  
  <align xlink:href="#t74"/> </token>  
<token>  
  <align xlink:href="#tundefined"/>  
  <align xlink:href="#t67"/> </token>
```

ST

TT

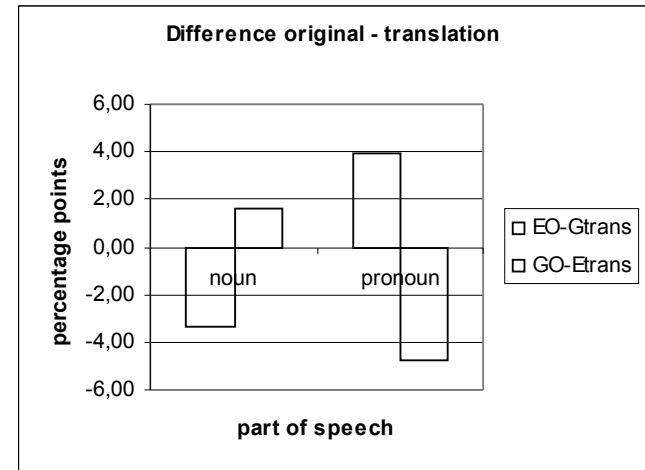
Reference

alignment of relative pronouns



distribution of nouns and pronouns

Subcorpus	Noun	Pronoun
ER	24,60	5,46
EO	27,21	4,73
ETRANS	26,14	4,54
GTRANS	23,84	8,67
GO	24,51	9,32
GR	22,93	8,45



Reference

distribution of nouns and pronouns

EO			Gtrans		
sentence			sentence		
id	no. of nouns	no. of pronouns	id	no. of nouns	no. of pronouns
s12	3	2	s15	6	3
s13	5	2	s16	7	6
s15	6	4	s18	10	7
s18	8	1	s23	5	1
s20	11	1	s27	9	1
s24	10	0	s33	7	1

During challenging market transitions, successful companies usually get surprised by the **downturn**, they determine how long **it** will last and how deep **it** will be, and then they get ready for the upturn.

In einem Marktumfeld, dessen ständige Veränderungen eine große Herausforderung darstellen, werden auch erfolgreiche Unternehmen von einem **Abschwung der Konjunktur** zunächst überrascht, untersuchen dann jedoch umgehend, wie lange **diese schwierige Situation** andauern und wie tief **die konjunkturelle Delle** sein wird, um für den folgenden Aufschwung gerüstet zu sein.

Conjunction

- Conjunctive elements systematically connecting the following to the previous
- Indicators: conjunct(ion)s

<result><ori_en>Baker Hughes Business Support Services has assumed accounting payroll, benefits and IT support duties for many of the company's U.S. operations, **eliminating** duplicate efforts by division personnel. </ori_en>

<trans_ge>Baker Hughes Business Support Services hat die Buchführung, Gehalts- und Sozialleistungen sowie IT-Aufgaben für viele Niederlassungen des Unternehmens in den Vereinigten Staaten übernommen, **wodurch** doppelte Arbeit durch das Personal in den Tochterunternehmen **vermieden werden konnte**.

</trans_ge></result>

<result><ori_en>In this environment, Baker Hughes revenue declined 22% to \$4.5 billion for 1999, **compared to** \$5.8 billion in 1998. </ori_en>

<trans_ge>Vor diesem Hintergrund sanken die Umsatzerlöse von Baker Hughes im Jahre 1999 um 22% auf 4,5 Mrd. Dollar, **während sie** 1998 noch 5,8 Mrd. Dollar **betragen hatten**. </trans_ge></result>

Substitution and ellipsis

- Replacing one item by a semantically weaker or by zero
- Quantification not yet available
- What is it about?

(...) proponents of enlargement, of which I am an enthusiastic **one**, occasionally fall into the rhetorical trap of arguing (...)

(...) die Verfechter der Erweiterung - und ich zähle mich zu den enthusias-
tischen - gehen manchmal in die rhetorische Falle zu argumentieren, (...)

ellipsis

(...) that the US genuinely wants the EU to be a strategic global partner - **one** we can work closely with

(...), dass die Vereinigten Staaten aufrichtig an der EU als strategischen globalen Partner interessiert sind - einen Partner, mit dem wir (...)

repetition

Habt ihr gesehen, ob er sie küßt? Und sie ihn?

Have you seen whether he kisses her? And does she kiss him?

completion

Substitution

Ellipsis

Lexical cohesion

require semantic analysis

- Selection of vocabulary, i.e. replacing a lexical item by ~~a general noun, a (near) synonym, a superordinate, a hyponym,~~ or simply **repeating** the item
- Indirect indicators: type token ratio*, (lexical density)

* Caution: no discrimination between repetition of lexical and function words!

Type token ratio in tourism texts

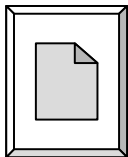
	EO	Etrans	GO	Gtrans
TOU	12.58	13.33	24.17	19.93

	Types	Token
EO_TOU_011	698	5,328
Gtrans_TOU_011	718	4,526

Text no.	EO_TOU	Gtrans
001	17.00	20.40
002	15.96	17.94
003	16.19	18.90
004	18.77	19.71
005	15.35	18.17
006	15.04	16.81
007	15.53	18.03
008	16.24	18.26
009	15.51	16.85
010	14.96	16.09
011	13.10	15.86

Aligned text pair TOU_011

50 most frequent lexical words



	word	absolute freq.	cumulated freq. all words	cumulated freq. content words	word no.	word	absolute freq.	cumulated freq. all words	cumulated freq.
6	San_Francisco	112	1334	112	5	San_Francisco	88	996	88
9	California	60	1572	172	8	los_angeles	64	1230	152
10	Los_Angeles	60	1632	232	10	Kalifornien	52	1340	204
13	mountain	48	1788	280	11	stadt	48	1388	252
14	city	44	1832	324	12	Sierra_Nevada	40	1428	244
19	bay	34	2024	358	19	liegen	30	1670	274
20	valley	34	2058	392	20	bekannt	28	1698	302
22	beach	32	2122	424	22	San_Diego	24	1750	326
24	east	26	2176	450	24	meile	24	1798	350
28	Nevada	24	2276	474	25	Valley	22	1820	372
29	state	24	2300	498	27	bucht	22	1864	394
30	car	22	2322	520	28	strand	22	1886	416
31	coast	22	2344	542	30	Yosemite	20	1926	436
32	coastal	22	2366	564	31	berg	20	1946	456
33	desert	22	2388	586	33	gross	20	1986	476
34	attraction	20	2408	606	34	km	20	2006	496
36	locate	20	2448	626	36	oestlich	20	2046	496
37	mile	20	2468	646	38	staat	20	2086	516
38	Miles	20	2488	666	40	USA	18	2124	534
39	cable	18	2506	684	42	hotel	18	2160	552
40	famous	18	2524	702	46	grenze	16	2226	568
41	golden	18	2542	720	47	kueste	16	2242	584
42	large	18	2560	738	48	tal	16	2258	600
43	near	18	2578	756	49	Big	14	2272	614
45	Sierra	18	2614	774	50	Sur	14	2286	628
46	Yosemite	18	2632	792					
49	bridge	16	2680	808					
50	community	16	2696	824					

Most frequent word
MORE frequent

In total fewer
content words
→ more lexical
variation

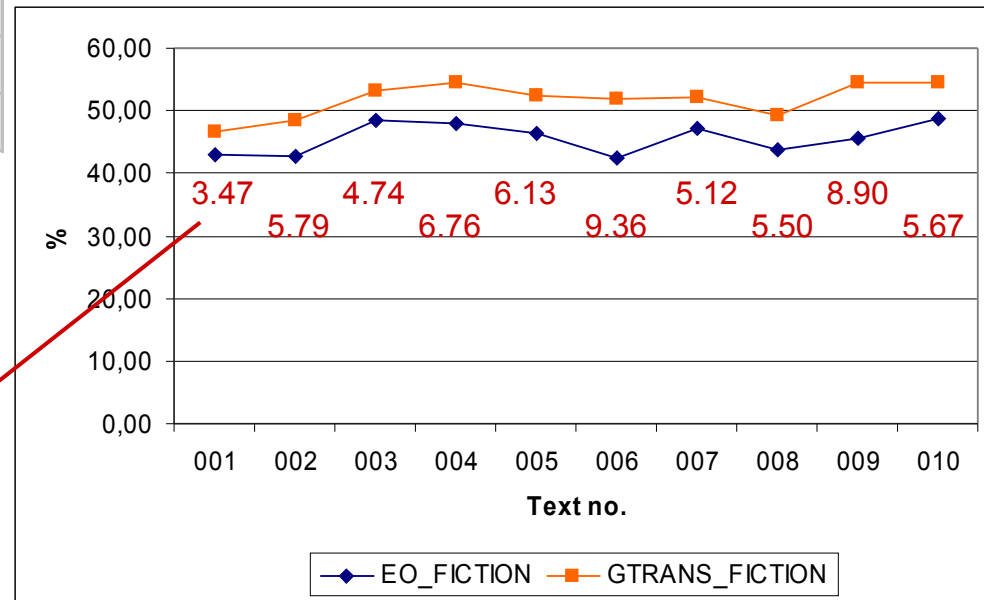
Conclusion

Translation less
repetitive, i.e.
less lexical
cohesion as
reflected by
repetition

Lexical density in fiction texts

Register	EO	Etrans	GO	Gtrans
Essay	54,93	54,52	58,93	57,05
Fiction	45,61	46,49	52,31	51,66
Instr	57,02	57,72	57,66	54,30
Popsci	53,82	49,44	55,38	58,64
Share	56,81	55,03	55,16	55,08
Speech	52,34	49,76	51,74	55,10
Tourism	57,57	56,80	60,29	58,51
Web	56,38	54,85	57,30	55,97
TOTAL	54,31	53,13	56,11	55,73

Difference
between
Gtrans and EO



Conclusion and outlook

- Our analysis gave a first glimpse at quantitative frequencies of cohesion in originals and translations
 - More detailed analyses under way
 - Each cohesive device has a different impact on translations depending on contrastive features of the languages involved
 - Interpretations concerning properties of translation (not only in terms of explicitation) depending on statistical tests
 - Explanations according to language and text typology as well as the translation process
-