Cross-linguistic Analysis of Cohesion

variation across production types and registers

Ekaterina Lapshinova-Koltunski and Kerstin Kunz

Saarland University, Heidelberg University

22 May 2013, Santiago de Compostela

Acknowledgement

Research Project

GECCo: German-English Contrasts in Cohesion

supported by the DFG

Project Team:

- Kerstin Kunz
- Ekaterina Lapshinova-Koltunski
- Marilisa Amoia
- Katrin Menzel
- Erich Steiner

FR 4.6 Applied Linguistics, Interpreting and Translation Studies

www.gecco.uni-saarland.de

Goal of Present Study

Goal of Present Study

cohesive reference:

- types: personal, demonstrative, comparative (cf. Halliday&Hasan, 1976)
- subtypes or functions (cf. Kunz, 2009; Kunz and Steiner, 2012)

across:

1. languages: English vs. German
2. registers: different text types
3. production types: originals vs. translations

Present Study: Linguistic variation

Hypotheses:

- variation is lower between
  GO vs GTRANS than EO vs GTRANS
- we expect more variation in form and function on the fine-grained level (cf. Kunz and Steiner, 2012).

Research Questions:

- Between which subcorpora are the greatest differences: across languages, registers or production types? languages or originals vs translations?
- Which features cause these differences?
- What is the most prominent difference between originals and translations?
- Are differences due to interference or rather to normalisation?

Corpus-based Analysis

Corpus-based Analysis

- Corpus Data
- Data Extraction
- Data Evaluation

Data: GECCo Corpus

subcorpora	registers
	(imported from CroCo)
EO 🇬🇧	FICTION, ESSAY
GO 🇩🇪	INSTR, POPSCI
ETRANS 🇩🇪→ 🇬🇧	TOU, WEB
GTRANS 🇬🇧→ 🇩🇪	SHARE, SPEECH
	(collected)
EO-SPOKEN 🇬🇧	INTERVIEW, ACADEMIC
GO-SPOKEN 🇩🇪	*FORUM, TALKSHOW*

GECCo annotation levels

1) word: ⇒ *word, lemma, pos*
2) chunk:⇒ *sentences, syntactic chunks, clauses, cohesive devices*
3) text: ⇒ *registers*
4) extralinguistic: ⇒ *register analysis, speaker information*

Data: GECCo Corpus

subcorpora	**registers**
	(imported from CroCo)
EO 🇬🇧	FICTION, ESSAY
GO 🇩🇪	INSTR, POPSCI
ETRANS 🇩🇪→ 🇬🇧	TOU, WEB
GTRANS 🇬🇧→ 🇩🇪	SHARE, SPEECH
	(collected)
EO-SPOKEN 🇬🇧	INTERVIEW, ACADEMIC
GO-SPOKEN 🇩🇪	*FORUM, TALKSHOW*

GECCo annotation levels

1) word: ⇒ *word, lemma, pos*

2) chunk:⇒ *sentences, syntactic chunks, clauses, **cohesive devices***

3) text: ⇒ *registers*

4) extralinguistic: ⇒ *register analysis, speaker information*

Corpus Annotation: Reference

- reference_type – types of reference:
  - personal
  - demonstrative
  - comparative
- reference_func – functional subtypes of reference:
  - *it/es* (endophoric and exophoric)
  - head
  - modifier
  - local
  - temporal
  - pronominal adverb
  - general
  - particular

Corpus Extraction: Register Distribution

```
> group Last match reference_type by match text_register;
```

```
FICTION    pers    1376
POPSCI     pers     804
SPEECH     dem      791
POPSCI     dem      706
FICTION    dem      670
```

```
> group Last match reference_func by match text_register;
```

```
FICTION    person-endophoric        1095
           possessive-endophoric     613
           it-endophoric             360
SPEECH     modifier                  294
ESSAY      particular                261
POPSCI     modifier                  259
SHARE      particular                255
POPSCI     particular                238
SHARE      possessive-endophoric     235
TOU        possessive-endophoric     230
```

Data Evaluation

Correspondance Analysis:

- **Input:** frequencies of cohesive devices across registers and production types
- **Output:** a two dimensional graph with:
  - **arrows** for the observed feature frequencies
  - **points** for registers across production types
- **Interpretation:**
  - the length of the **arrows** indicates how pronounced a particular feature is
  - the position of the **points** in relation to the **arrows** indicates the relative importance of a feature for a register.
  - the **arrows** pointing in the direction of an axis indicate a high contribution to the respective dimension

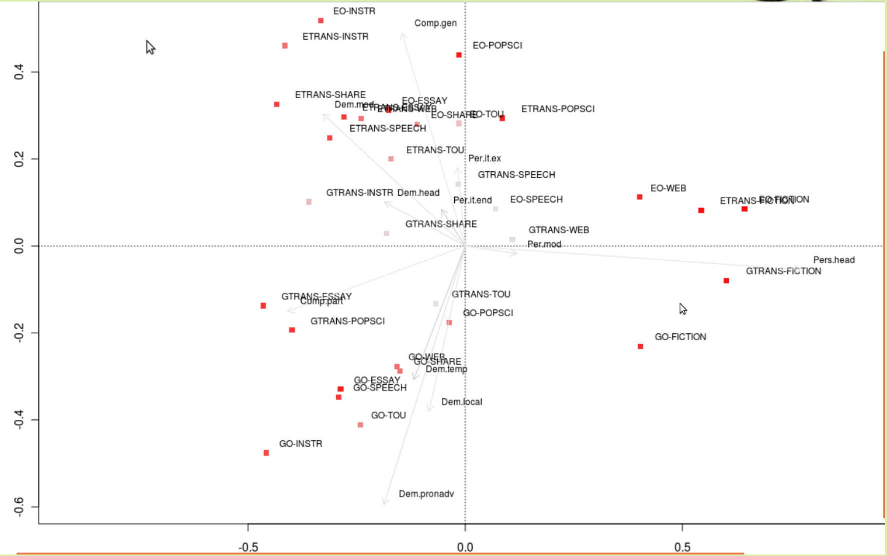
cf. (Glynn, 2012)

Analyses

Correspondence Analysis

EO vs GO vs ETRANS vs GTRANS

Correspondence Analysis



Correspondence Analysis

Observations for *x*-axis separation:

1. EO/GO/ETRANS/GTRANS: FICTION
   EO/GTRANS: WEB
   EO: SPEECH
   ETRANS: POPSCI
   - shared features: pers. head, pers. modifier and *it*-exophoric

   most prominent: pers. head

2. EO/GO/ETRANS/GTRANS: ESSAY, INSTR, SHARE, TOU
   EO/GO/GTRANS: POPSCI
   GO/GTRANS/ETRANS: SPEECH
   GO/ETRANS: WEB
   - shared features: all dem. and comp.

   most prominent: comp. particular

Correspondence Analysis

- **Observations for *y*-axis separation**:
  1. GO/GTRANS: ESSAY, FICTION, POPSCI, TOU
     GO: INSTR, SHARE, SPEECH, WEB
     - shared features: pers. head, pers. modifier, dem. local, dem. pronadv, dem. temporal, comp. particular

       most prominent: dem. pronadv and dem. local

  2. EO/ETRANS/GTRANS: INSTR, SHARE, SPEECH, WEB
     EO/ETRANS: ESSAY, FICTION, POPSCI, TOU
     - shared features: pers. *it*-endo/exophoric, dem. head, dem. modifier, comp. general

       most prominent: comp. general

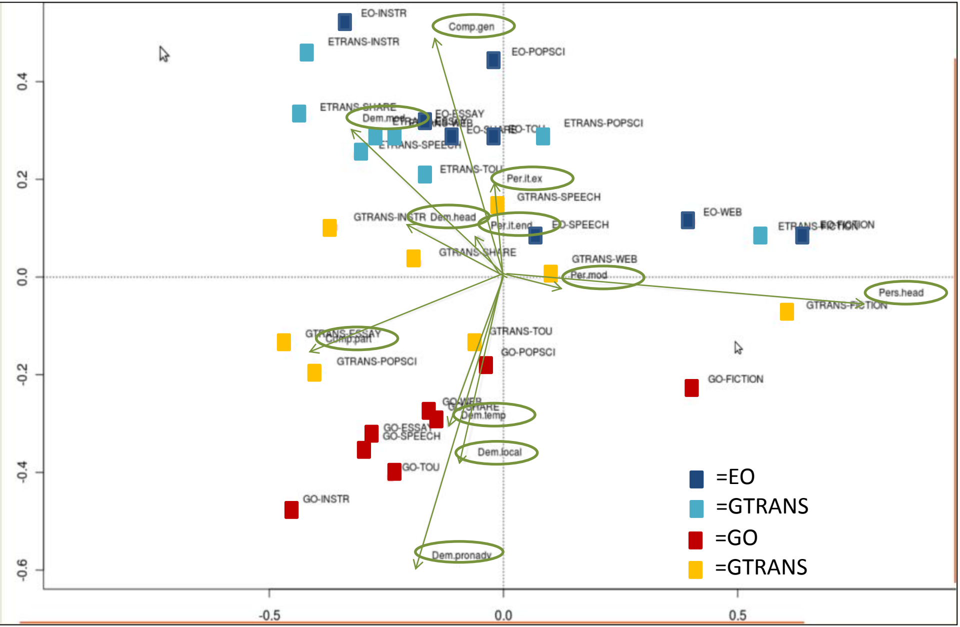
- both *y* and *x*-axis: dem. modifier

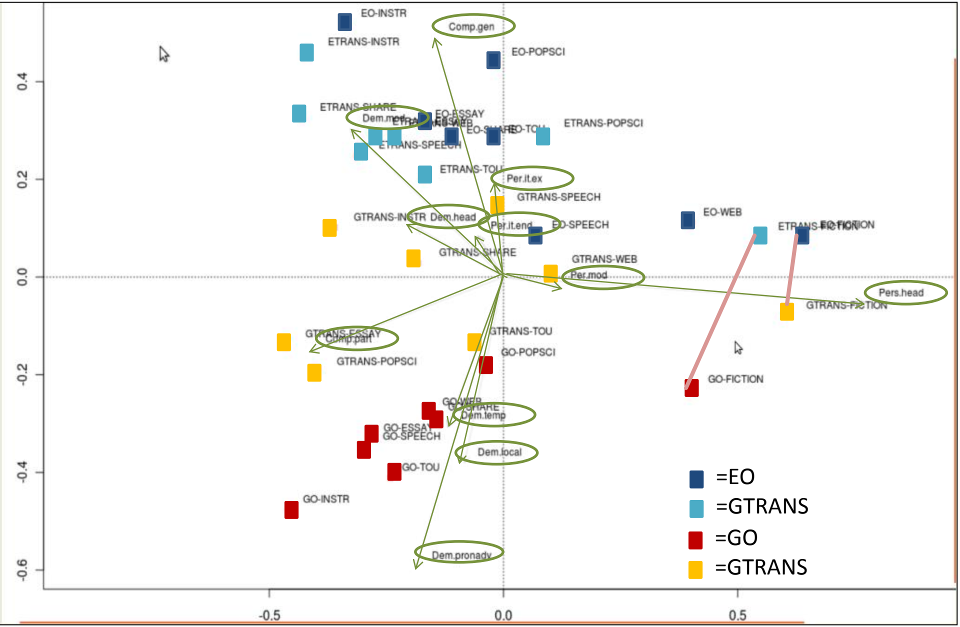
Correspondence Analysis

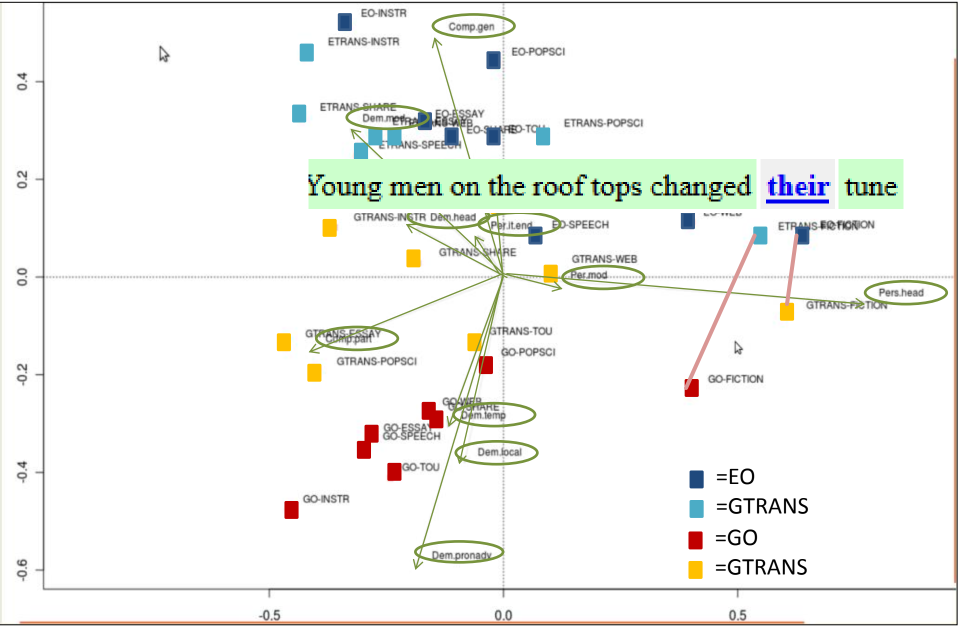
Interpretating Results

- *x*-axis:
  - separation between different registers
  - translations show differences and similarities from/with originals in both languages
  - most prominent features: pers. head and comp. particular
- *y*-axis:
  - clear separation between English and German originals
  - English translations are similar to English originals ⇒ **normalisation**?
  - German translations show more variation:
    - some registers similar to English originals ⇒ **interference**?
    - some registers similar to German originals ⇒ **normalisation**?
  - most prominent features: dem. pronadv, dem. local and comp. general

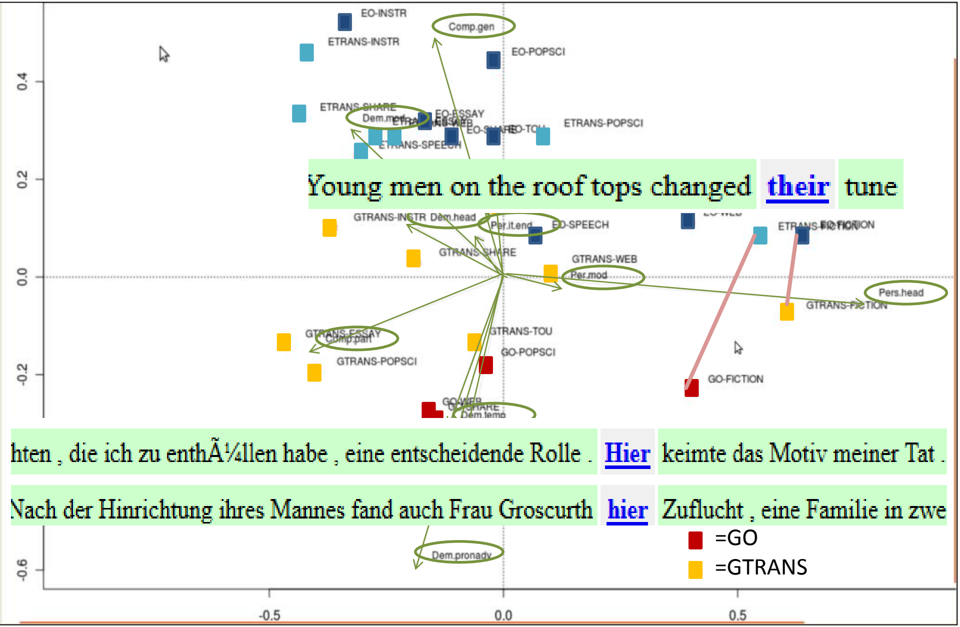
Analyses


EO-INSTR
Comp.gen
ETRANS-INSTR
EO-POPSCI
ETRANS-SHARE
Dem.mod
EO-ESSAY
ETRANS-WEB
EO-SHARE
EO-TOU
ETRANS-POPSCI
ETRANS-SPEECH
ETRANS-TOU
Per.it.ex
GTRANS-SPEECH
GTRANS-INSTR
Dem.head
Per.it.end
EO-SPEECH
EO-WEB
ETRANS-FICTION
EO-FICTION
GTRANS-SHARE
GTRANS-WEB
Per.mod
Pers.head
GTRANS-FICTION
GTRANS-ESSAY
Comp.part
GTRANS-TOU
GO-POPSCI
GTRANS-POPSCI
GO-FICTION
GO-WEB
GO-SHARE
Dem.temp
GO-ESSAY
GO-SPEECH
Dem.local
GO-TOU
GO-INSTR
Dem.pronadv
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.5
0.0
0.5
=EO
=GTRANS
=GO
=GTRANS


EO-INSTR
Comp.gen
ETRANS-INSTR
EO-POPSCI
0.4
ETRANS-SHARE
Dem.mod
EO-ESSAY
ETRANS-POPSCI
EO-TOU
ETRANS-SPEECH
ETRANS-TOU
0.2
Per.it.ex
GTRANS-SPEECH
GTRANS-INSTR
Dem.head
Per.it.end
EO-SPEECH
EO-WEB
ETRANS-FICTION
EO-FICTION
GTRANS-SHARE
GTRANS-WEB
0.0
Per.mod
Pers.head
GTRANS-FICTION
GTRANS-ESSAY
Comp.part
GTRANS-TOU
GO-POPSCI
-0.2
GTRANS-POPSCI
GO-FICTION
GO-WEB
GO-SHARE
Dem.temp
GO-ESSAY
GO-SPEECH
Dem.local
-0.4
GO-TOU
GO-INSTR
-0.6
Dem.pronadv
-0.5
0.0
0.5
🟦 =EO
🟦 =GTRANS
🟥 =GO
🟨 =GTRANS



Young men on the roof tops changed **their** tune



Young men on the roof tops changed **their** tune

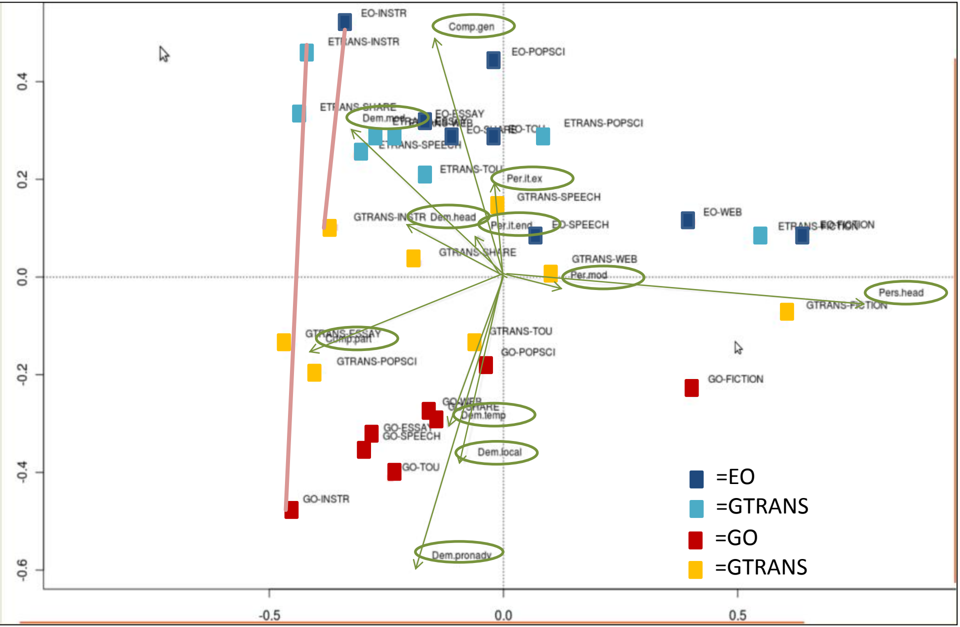
hten , die ich zu enthÃ¼llen habe , eine entscheidende Rolle . **Hier** keimte das Motiv meiner Tat .

Nach der Hinrichtung ihres Mannes fand auch Frau Groscurth **hier** Zuflucht , eine Familie in zwe

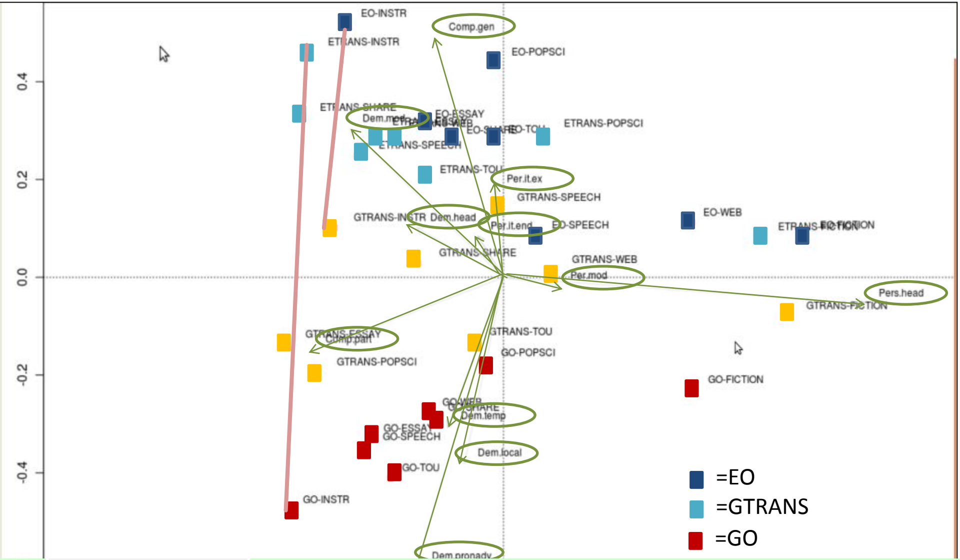
🟥 =GO

🟨 =GTRANS

Analyses

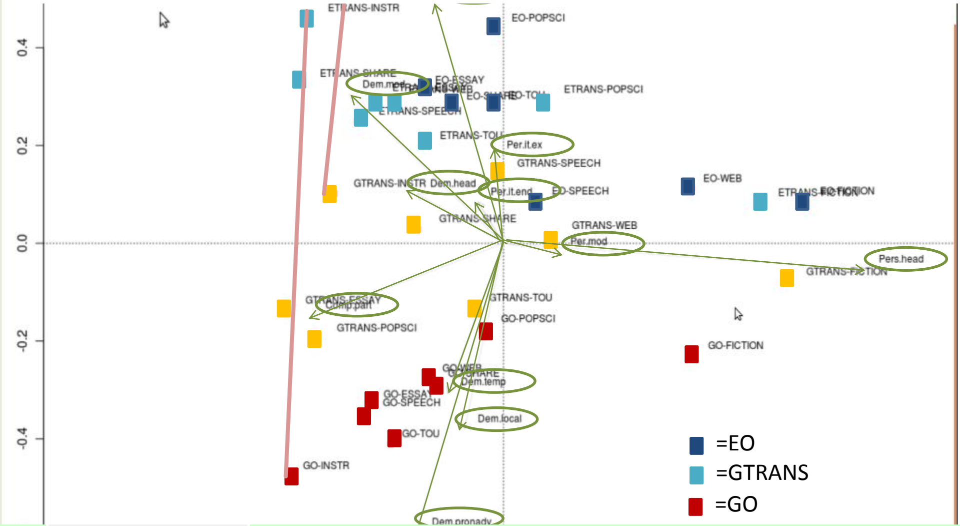


EO-INSTR
ETRANS-INSTR
Comp.gen
EO-POPSCI
ETRANS-SHARE
Dem.mod
EO-ESSAY
ETRANS-ESSAY
ETRANS-WEB
EO-SHARE
EO-TOU
ETRANS-POPSCI
ETRANS-SPEECH
ETRANS-TOU
Per.it.ex
GTRANS-SPEECH
GTRANS-INSTR
Dem.head
Per.it.end
EO-SPEECH
EO-WEB
ETRANS-FICTION
EO-FICTION
GTRANS-SHARE
GTRANS-WEB
Per.mod
Pers.head
GTRANS-FICTION
GTRANS-ESSAY
Comp.part
GTRANS-TOU
GO-POPSCI
GTRANS-POPSCI
GO-FICTION
GO-WEB
GO-SHARE
Dem.temp
GO-ESSAY
GO-SPEECH
Dem.local
GO-TOU
GO-INSTR
Dem.pronadv
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.5
0.0
0.5
=EO
=GTRANS
=GO
=GTRANS

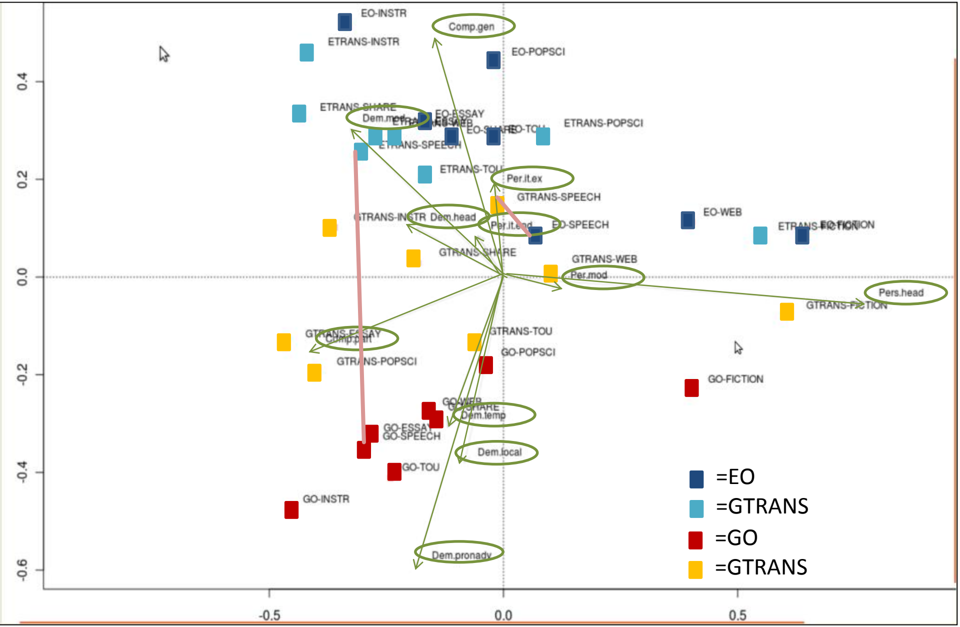


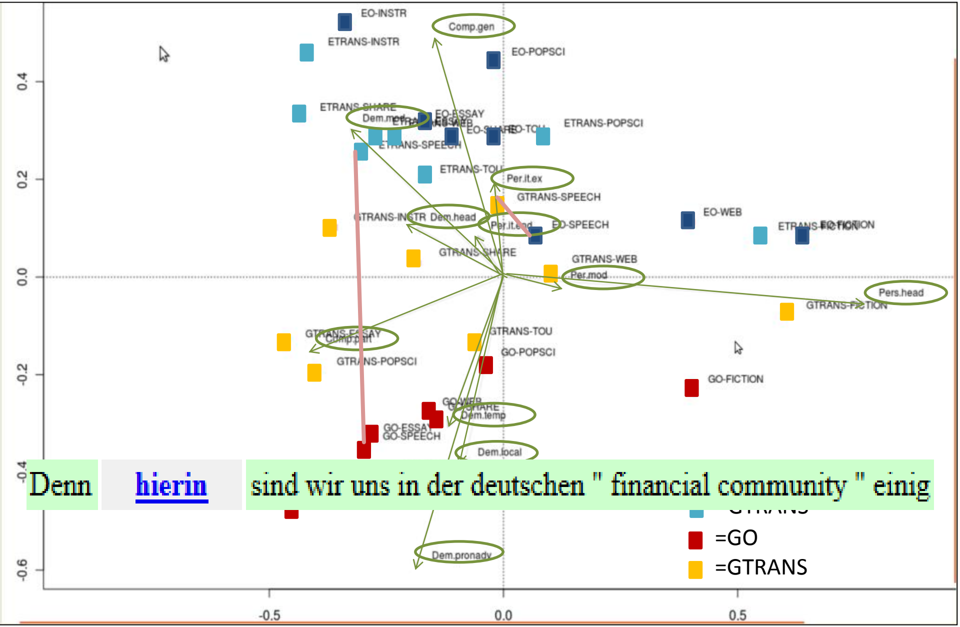
Noch gravierendere Probleme treten auf , wenn mehrere Betriebssysteme auf einem Rechner installiert wurden

If you see such interference , move the monitors apart until the interference disappears .



Noch gravierendere Probleme treten auf , wenn mehrere Betriebssysteme auf einem Rechner installiert wurden


EO-INSTR
Comp.gen
ETRANS-INSTR
EO-POPSCI
ETRANS-SHARE
Dem.mod
EO-ESSAY
ETRANS-ESSAY
ETRANS-WEB
EO-SHARE
EO-TOU
ETRANS-POPSCI
ETRANS-SPEECH
ETRANS-TOU
Per.it.ex
GTRANS-SPEECH
GTRANS-INSTR
Dem.head
Per.it.end
EO-SPEECH
EO-WEB
ETRANS-FICTION
EO-FICTION
GTRANS-SHARE
GTRANS-WEB
Per.mod
Pers.head
GTRANS-FICTION
GTRANS-ESSAY
Comp.part
GTRANS-TOU
GO-POPSCI
GTRANS-POPSCI
GO-FICTION
GO-WEB
GO-SHARE
Dem.temp
GO-ESSAY
GO-SPEECH
Dem.local
GO-TOU
GO-INSTR
Dem.pronadv
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.5
0.0
0.5
=EO
=GTRANS
=GO
=GTRANS

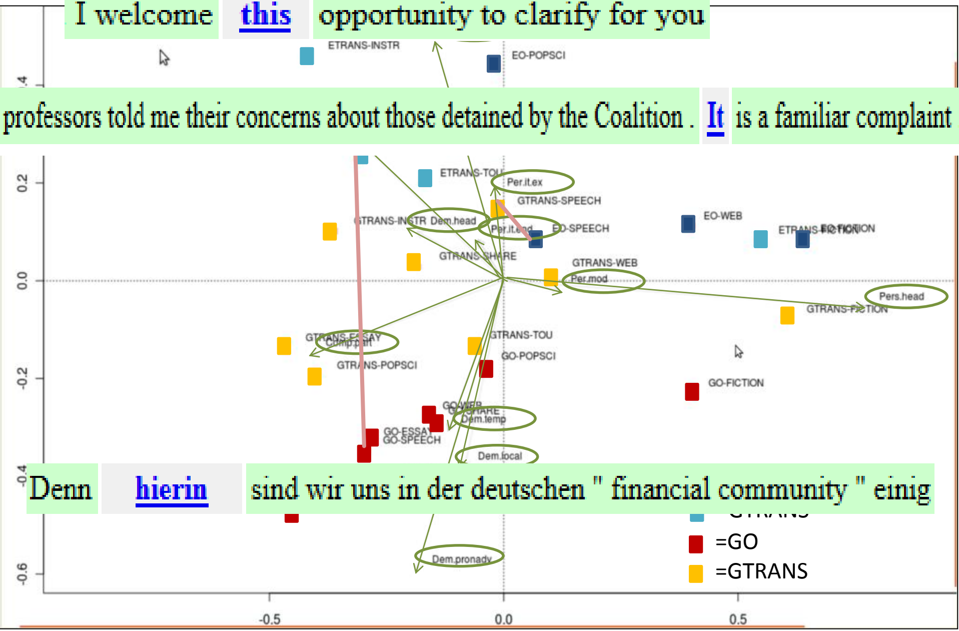


Denn **hierin** sind wir uns in der deutschen " financial community " einig

I welcome **this** opportunity to clarify for you

professors told me their concerns about those detained by the Coalition . **It** is a familiar complaint .

Denn **hierin** sind wir uns in der deutschen " financial community " einig



Discussion

Discussion

Research Questions:

1. Between which subcorpora are the greatest differences ?
2. Which features cause these differences ?
3. What is the most prominent difference between originals and translations ?
4. Are differences due to interference or rather to normalisation ?

Discussion

Research Questions:

1. Between which subcorpora are the greatest differences: across languages, registers or production types?

   ⇒ greatest differences between original subcorpora! translations are in between but ETRANS is closer to EO

2. Which features cause these differences?

   ⇒ ENGLISH:
   preference for pers. reference and comp. general
   and dem. modifier

   ⇒ GERMAN:
   preference for dem. pron. adverbs + dem. adverbials
   and comp. particular

Discussion

Research Questions:

3 What is the most prominent difference between originals vs. translations (of the same language)?

register-dependent:

- GTRANS-FICTION:
  more pers. heads and modifiers, less pron. adverbials and loc. dem. than GO
- GTRANS-SPEECH:
  more pers. modifiers, dem. modifiers, and es-exophoric than GO
- GTRANS INSTR:
  less temp. and loc. adverials and less comp. particular

Discussion

Research Questions:

4 Are differences due to interference or rather to normalisation?

language-/translation direction-dependent:

- EO ⇒ GTRANS:
  1. strong interference
  2. normalisation (=exaggeration of TL Conventions) for particular registers on the other hand
  3. lower distributions than both original subcorpora
     ⇒ strongly depends on register and devices of reference

  ⇒ more heterogeneity!
- GO ⇒ ETRANS:
  1. interference but not too such a strong degree
  2. ETRANS generally shows more commonalities to EO

  ⇒ less distinct properties of translation,
  less dependence on register

Thank you!

Questions? Comments? Suggestions?

Ekaterina Lapshinova
e.lapshinova@mx.uni-saarland.de
Kerstin Kunz
kerstin.kunz@iued.uni-heidelberg.de