Empirical insights into Cohesion
Contrasting English and German across Modes

Kerstin Kunz (Heidelberg),
Ekaterina Lapshinova-Koltunski, José Martínez-Martínez,
Katrin Menzel, Erich Steiner (Saarbrücken)

IÜD, Heidelberg
28/10/16
Research project supported by the DFG (GZ STE 840/6-2 and KU 3129/1-2)

Universität des Saarlandes & Universität Heidelberg

Current team

- Kerstin Kunz
- Ekaterina Lapshinova-Koltunski
- Erich Steiner
- José M. Martínez
- Katrin Menzel

Former members
Stefania Degaetano-Ortlieb, Marilisa Amoia

www.gecco.uni-saarland.de
Overview

1. Defining Concepts
2. Research Agenda and Methodology
3. Analyses and Results
   - Strength of Cohesive Relation
   - Type of Meaning Relation
4. Conclusion
Defining Concepts
**Cohesion** is an important component of effectively organised and meaningful *discourse*, as the *message* being communicated in discourse is not just a set of clauses, but forms a **unified, coherent whole**.

<table>
<thead>
<tr>
<th>Types of Cohesion (cf. Halliday &amp; Hasan 1976)</th>
<th>Meaning relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coreference&lt;br&gt; <em>An option ... it/ this option</em></td>
<td>identity</td>
</tr>
<tr>
<td>Substitution&lt;br&gt; <em>Many options ... a good one</em></td>
<td>Type reference/ comparison</td>
</tr>
<tr>
<td>Ellipsis&lt;br&gt; <em>You will feel disappointment. []</em>&lt;br&gt; <em>Maybe.</em>&lt;br&gt; <em>Many options ... a good [].</em></td>
<td></td>
</tr>
<tr>
<td>Comparative Reference&lt;br&gt; <em>One option ... another/ better option</em></td>
<td></td>
</tr>
<tr>
<td>Cohesive conjunction&lt;br&gt; <em>X. But/ And/ However Y</em></td>
<td>Logico-semantic relations (addition, contrast, cause, …)</td>
</tr>
<tr>
<td>Lexical cohesion&lt;br&gt; <em>factors ... factors</em></td>
<td>Similarity (repetition, general nouns, meronymy, Synonymy …)</td>
</tr>
</tbody>
</table>
Case studies


cf. also [http://www.gecco.uni-saarland.de/GECCo/deliverables.html](http://www.gecco.uni-saarland.de/GECCo/deliverables.html)
Several studies have shown that two of [the factors affecting regret] are how much one feels personal responsibility for the result and how easy it is to imagine a better alternative. The availability of choice obviously exacerbates both these factors. When you have no options, what can you do? You will feel disappointment, maybe, regret, no. With no options, you just do the best you can. But with many options, the chances increase that [a really good one] is out there, and you may well feel that you ought to have been able to find it.
Several studies have shown that two of [the factors affecting regret] are how much one feels personal responsibility for the result and how easy it is to imagine a better alternative. The availability of choice obviously exacerbates both these factors. When you have no options, what can you do? You will feel disappointment, maybe even regret, no. With no options, you just do the best you can. But with many options, the chances increase that [a really good one] is out there, and you may well feel that you ought to have been able to find it.

Several studies have shown that **two of the factors affecting regret** are how much one feels personal responsibility for the result and how easy it is to imagine a better alternative. The availability of choice obviously exacerbates both **these factors**.

When you have no options, what can you do? You will feel disappointment, maybe; regret, no. With no options, you just do the best you can. But with many options, the chances increase that **a really good one** is out there, and you may well feel that you ought to have been able to find it.

Mehreren Studien zufolge wird das Gefühl der Reue zum einen stärker, je mehr man sich für das Resultat persönlich verantwortlich fühlt, und zum anderen, je leichter man sich eine bessere Alternative vorstellen kann. Ein Auswahlangebot verschlimmert offensichtlich beide Faktoren. Was kann man schon groß anstellen, wenn man keine Wahl hat? Vielleicht ist man enttäuscht, aber Reue empfindet man nicht. Wenn es hingegen viele Optionen gibt, wächst das Risiko, dass man meint, **eine besonders gute übersehen** zu haben, und dies nun bereut.
Several studies have shown that two of the factors affecting regret are how much one feels personal responsibility for the result and how easy it is to imagine a better alternative. The availability of choice obviously exacerbates both these factors. When you have no options, what can you do? You will feel disappointment, maybe; regret, no. With no options, you just do the best you can. But with many options the chances increase that a really good one is out there, and you may well feel that you ought to have been able to find it.
Defining Concepts

Differences across Modes and Registers

EO-INTERVIEW-08
A: Well, the reality is they’re going to have a nuclear program anyway.
B: But, I mean, in other words, it’s OK?
A: No, I think, that’s – well, I think ... It’s a major thing here.
B: Well, I think one of the things the Bush administration has always emphasized, ..., it’s the nature of the regime, stupid.
A: And so they’ve also emphasized: You judge countries by how they behave. If they’re a democratic country, transparent, they behave constructively in the world, the standards are different. That’s why the standards are different on this trip between India and Pakistan.

EO-POPSCI-004
Several studies have shown that two of the factors affecting regret are how ... and how easy it is to imagine a better alternative. The availability of choice obviously exacerbates both these factors. When you have no options, what can you do? You will feel disappointment, maybe; regret, no. With no options ... But with many options, the chances increase that a really good one is out there, and you may well feel that you ought to have been able to find it.
Research Agenda and Methodology
Research Design - Overall

Mode 1-2

Register 1-14

Language 1-2
Research Agenda and Methodology

Research Design - Case Study

Strength

Type

Lexical Cohesion
Conjunction
Ellipsis
Substitution
Reference

Language 1-2
Register 1-14
Model 1-2
Theoretical Background

- Language: EO vs. GO
  (Hawkins, 1986; König&Gast, 2012; Königs, 2011, etc.)

- Mode of production: spoken vs. written
  (Biber, 1988; Mair, 2006; Leech et al., 2009)

- Register
  (Biber, 1995; House 2002; Hansen-Schirra et al., 2012; Neumann, 2013)
Abstract Dimensions

1. Overall Degree of Cohesion (Kunz et al. in press.a)
2. Strength of Cohesive Relations
3. Types of Meaning Relations
4. Breadth of Variation (Kunz et al. in press.b)
Empirical Methodology

- define operationalisations: relate cohesive features to abstract dimensions

- for each independent variable
  - extract instances/frequencies for dependent variables from a corpus
  - evaluate frequencies statistically

- interpret results in terms of abstract dimensions
Operationalisations

1. Degree of Cohesion:
   What is the average proportion of cohesive devices per text?

2. Strength of Cohesive Relations:
   - How explicit are cohesive devices?
   - How close are elements in cohesive chains?
   - How long are cohesive chains?

3. Types of Meaning Relations:
   - Which meaning relations are more important in languages/modes/registers?
   - Which meaning relations are most important for the distinction of languages/modes/registers

4. Breadth of variation: How much cohesive variation is there for
   - Language: English <=> German,
   - Mode: spoken <=> written,
   - Register?
Corpus Resources

Register 1-14

English

German

written

spoken

written

spoken
Corpus Versions

<table>
<thead>
<tr>
<th>GECCOCOH</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>register</td>
<td>EO</td>
<td>GO</td>
</tr>
<tr>
<td>ACADEMIC</td>
<td>40.559</td>
<td>43.703</td>
</tr>
<tr>
<td>ESSAY</td>
<td>34.998</td>
<td>35.668</td>
</tr>
<tr>
<td>FICTION</td>
<td>36.996</td>
<td>36.778</td>
</tr>
<tr>
<td>INSTR</td>
<td>36.167</td>
<td>36.880</td>
</tr>
<tr>
<td>INTERVIEW</td>
<td>37.898</td>
<td>40.198</td>
</tr>
<tr>
<td>POPSCI</td>
<td>35.148</td>
<td>36.177</td>
</tr>
<tr>
<td>SHARE</td>
<td>35.824</td>
<td>35.235</td>
</tr>
<tr>
<td>SPEECH</td>
<td>35.062</td>
<td>35.399</td>
</tr>
<tr>
<td>TOU</td>
<td>35.907</td>
<td>36.574</td>
</tr>
<tr>
<td>WEB</td>
<td>36.119</td>
<td>35.779</td>
</tr>
<tr>
<td>TOTAL</td>
<td>364.678</td>
<td>372.391</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GECCOCHAIN</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>register</td>
<td>EO</td>
<td>GO</td>
</tr>
<tr>
<td>ESSAY</td>
<td>27.171</td>
<td>31.407</td>
</tr>
<tr>
<td>FICTION</td>
<td>36.996</td>
<td>36.778</td>
</tr>
<tr>
<td>INTERVIEW</td>
<td>30.057</td>
<td>35.036</td>
</tr>
<tr>
<td>POPSCI</td>
<td>27.055</td>
<td>32.639</td>
</tr>
<tr>
<td>TOTAL</td>
<td>121.279</td>
<td>135.860</td>
</tr>
</tbody>
</table>

- both subsets of GECCo (Lapshinova et al., 2012 and Hansen-Schirra et al., 2012)

GECCo annotation levels

- word: ⇒ word, lemma, pos, chunk:⇒ sentences, syntactic chunks, clauses, text: ⇒ registers, extralinguistic: ⇒ register analysis, speaker information, cohesion (Lapshinova & Kunz, 2014; Martinez Martinez et al., 2016)
Analyses and Results
Strength of Cohesive Relation

see Kunz et al (2016)

1. How close are elements in cohesive chains (distance between members of cohesive chains)?

2. How long are cohesive chains?
1. Chain Distance

Analyses and Results  
Strength of Cohesive Relation

- lexcoh.chaindist.mean
- coref.chaindist.mean

ESSA Y FICTION INTERVIEW
POPSCI ESSA Y FICTION INTERVIEW
POPSCI
EO
GO

Empirical insights into Cohesion
2. Chain Length

Analyzes and Results

Strength of Cohesive Relation

28/10/16

Empirical insights into Cohesion
### Meaning Relations

**through cohesive devices**

1. Which meaning relations are more important in languages/modes/registers?
2. Which meaning relations are most important for the distinction of languages/modes/registers?

<table>
<thead>
<tr>
<th>cohesive devices</th>
<th>meaning relations</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>coreference</td>
<td>identity</td>
<td>1,2</td>
</tr>
<tr>
<td>substitution</td>
<td>type reference/comparison</td>
<td>2</td>
</tr>
<tr>
<td>ellipsis</td>
<td>logico-semantic relations: addition, contrast, cause, etc.</td>
<td>2</td>
</tr>
<tr>
<td>comparative reference</td>
<td>similarity: repetition, general nouns, meronymy, synonymy, etc.</td>
<td>1</td>
</tr>
<tr>
<td>cohesive conjunction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lexical cohesion</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 1. Important Relations

#### Identity (Coreference) and Similarity (Lexical Cohesion)

<table>
<thead>
<tr>
<th></th>
<th>cohyp</th>
<th>coinst</th>
<th>comer</th>
<th>hol</th>
<th>hyper</th>
<th>hypo</th>
<th>inst</th>
<th>mer</th>
<th>repet</th>
<th>syn</th>
<th>type</th>
<th>ident</th>
</tr>
</thead>
</table>

**association between lang/modes/registers and sem. relations (likelihood ratio)**

- If ratio < 1 => log(ratio) < 0 (negative values) => red color
- If ratio > 1 => log(ratio) > 0 (positive values) => blue color
2. Distinction: Languages

<table>
<thead>
<tr>
<th>Type of Meaning Relation</th>
<th>GO</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO</td>
<td>99.3</td>
<td>98.6</td>
<td>99.0</td>
</tr>
<tr>
<td>EO</td>
<td>98.5</td>
<td>99.2</td>
<td>98.9</td>
</tr>
<tr>
<td>Weight. Av.</td>
<td>98.9</td>
<td>98.9</td>
<td>98.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of Meaning Relation</th>
<th>GO</th>
<th>EO</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>F-Measure</td>
</tr>
<tr>
<td>coref:dem-pronadv</td>
<td>1.0173</td>
<td>coref:comp-general</td>
<td>1.0173</td>
</tr>
<tr>
<td>conj:adversat-adverb</td>
<td>1.9675</td>
<td>coref:dem-mod</td>
<td>1.9675</td>
</tr>
<tr>
<td>conj:causal-adverb</td>
<td>1.6618</td>
<td>subst:nom</td>
<td>1.6618</td>
</tr>
<tr>
<td>subst:claus</td>
<td>1.6057</td>
<td>coref:pers-it</td>
<td>1.6057</td>
</tr>
<tr>
<td>coref:dem-local</td>
<td>1.4476</td>
<td>conj:causal-subjun</td>
<td>1.4476</td>
</tr>
<tr>
<td>coref:dem-artic</td>
<td>1.3708</td>
<td>conj:temp-subjun</td>
<td>1.3708</td>
</tr>
<tr>
<td>conj:addit-adverb</td>
<td>1.1606</td>
<td>subst:verb</td>
<td>1.1606</td>
</tr>
<tr>
<td>conj:temp-adverb</td>
<td>1.0173</td>
<td>coref:comp-particular</td>
<td>1.0173</td>
</tr>
<tr>
<td>conj:modal-adverb</td>
<td>0.9146</td>
<td>conj:adversat-conn</td>
<td>0.9146</td>
</tr>
<tr>
<td>conj:adversat-subjun</td>
<td>0.8751</td>
<td>coref:pers-mod</td>
<td>0.8751</td>
</tr>
</tbody>
</table>
## 2. Distinction: Modes

<table>
<thead>
<tr>
<th>Type of Meaning Relation</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>spoken</td>
<td>97.7</td>
<td>91.3</td>
<td>94.4</td>
</tr>
<tr>
<td>written</td>
<td>98.3</td>
<td>99.6</td>
<td>98.9</td>
</tr>
<tr>
<td>Weight.Av.</td>
<td>98.2</td>
<td>98.2</td>
<td>98.2</td>
</tr>
</tbody>
</table>

| written                  | 0.8866 conj:temp-subjun | 1.3347 conj:modal-adverb |
| 0.8543 coref:pers-mod    | 1.1550 coref:pers-it    |
| 0.7204 coref:obj         | 1.1275 subst:verb       |
| 0.7177 elli:verb         | 1.0998 coref:dem-head   |
| 0.6995 conj:causal-adverb| 0.9904 conj:adversat-conn|
| 0.5736 conj:causal-conn  | 0.8789 conj:addit-conn   |
| 0.5107 antec:subj        | 0.6856 conj:addit-subjun|
| 0.4760 conj:adversat-adverb| 0.6239 subst:nom        |
| 0.4485 conj:adversat-subjun| 0.4989 antec:other      |
| 0.4120 antec:obj         | 0.4709 antec:event-vp    |
CONCLUSIONS
Conclusions

Cohesive devices:
GO vs. EO: wider scope, more explicit and focused
spoken vs. written: wider scope, less explicit, more focused

Cohesive chains:
GO = EO
Register specific!
Conclusions

Type

Lexical Cohesion
Conjunction
Ellipsis
Substitution
Reference

Distribution:
Language and Mode:
similarity > logico-sem > identity > comparison

Distinction:
German: logic-semantic
English: identity, compar.
Spoken: logico-sem (modal)
Written: identity, temp, cause
Conclusions

Degree

Lexical Cohesion
Conjunction
Ellipsis
Substitution
Reference

German
(<)=
English

spoken >
written
Conclusions

Variation

Lexical Cohesion
Conjunction
Ellipsis
Substitution
Reference

German > English
References


