



# Beyond Identity Coreference

## Contrasting Indicators of Textual Coherence in English and German

*Kerstin Kunz (Heidelberg),  
Ekaterina Lapshinova-Koltunski, José Martínez-Martínez  
(Saarbrücken)*

CORBON, San Diego  
16/06/16



Research Project

# GECCo: German-English Contrasts in Cohesion

supported by the DFG

## Project Team:

- Kerstin Kunz
- Ekaterina Lapshinova-Koltunski
- Erich Steiner
- Jose Manuel Martinez
- Katrin Menzel

Acknowledgement: Stefania Degaetano-Ortlieb, Marilisa Amoia

[www.gecco.uni-saarland.de](http://www.gecco.uni-saarland.de)

# Overview



- 1 Defining Concepts
- 2 Goals and Features
- 3 Methodology
- 4 Results
  - Topic Development
  - Semantic Variation
- 5 Summary



# Defining Concepts

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 personally responsible 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.

## Coreference

Mehreren Studien zufolge wird das Gefühl der Reue zum einen stärker, je mehr man sich für das Ergebnis persönlich verantwortlich fühlt, und zum anderen, je leichter man sich eine bessere Alternative vorstellen kann. Ein reiches 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 personally responsible 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 Ergebnis 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.

## Lexical Cohesion

# Cohesion



Types of Cohesion (cf. Halliday & Hasan 1976)	Meaning relations
Coreference <i>An option ... <u>it/ this</u> option</i>	identity
Substitution <i>Many options ... a good <u>one</u></i>	Type reference/ comparison
Ellipsis <i>You will feel disappointment. []</i> <i>Maybe.</i> <i>Many options ... a good [].</i>	
Comparative Reference <i>One option ... <u>another/ better</u> option</i>	
Cohesive conjunction <i>X. <u>But/ And/ However</u> Y</i>	Logico-semantic relations (addition, contrast, cause, ...)
Lexical cohesion <i>factors ... factors</i>	Similarity (repetition, general nouns, meronymy, Synonymy ...)





# Goals and Features

# Research Goals



- Variation in coreference and lexical chains as an indicator of
  - **topic development**
    - topic continuity of individual referents
    - topic continuity within one domain
    - topic variation
    - topic interaction
  - **semantic variation**
- **cross-linguistically**: contrasts between English and German, see House (1997, 2015), Hansen-Schirra et al. (2012) and Neumann (2013)
- **across registers**: between spoken and written registers, see Hundt & Mair (1999), Mair (2006) or Leech et al. (2009)

# Goals and Features



topic continuity of individual referents	coreference chain length and coref. chain distance
topic continuity within one domain	lexical chain length and lexical chain distance and lexical chain number
topic variation	chain length and distance and chain number
topic interaction	chain length and chain distance
semantic variation	repetition and identity
semantic variation	synonymy, meronymy, (co-)hyponymy, hyperonymy, (co-)meronymy, holonymy, (co-instance), type, antonymy, synonymy

# Example



## Differences across registers

### EO-INTERVIEW-010

I live in a **town** called (**Reigate**). **London** and the countryside which is quite nice. It takes us about 25 minutes to get to **it's** a **town**, **it's** more of a **village**. **It's** quite small. **It's** very nice actually, **it's** a nice **place** to live. And I grew up in a **place** called **Banstead** which is fairly close to **Reigate**.

### 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**.



# Methodology

# Corpus Data



in tokens

register	EO	GO
ESSAY	27.171	31.407
FICTION	36.996	36.778
INTERVIEW	30.057	35.036
POPSCI	27.055	32.639
TOTAL	121.279	135.860

- subset of GECCo  
(Lapshinova et al., 2012 and Hansen-Schirra et al., 2012)
- **ESSAY** and **POPSCI**: written discourse
- **INTERVIEW**: spoken discourse
- **FICTION**: spoken elements in form of dialogues





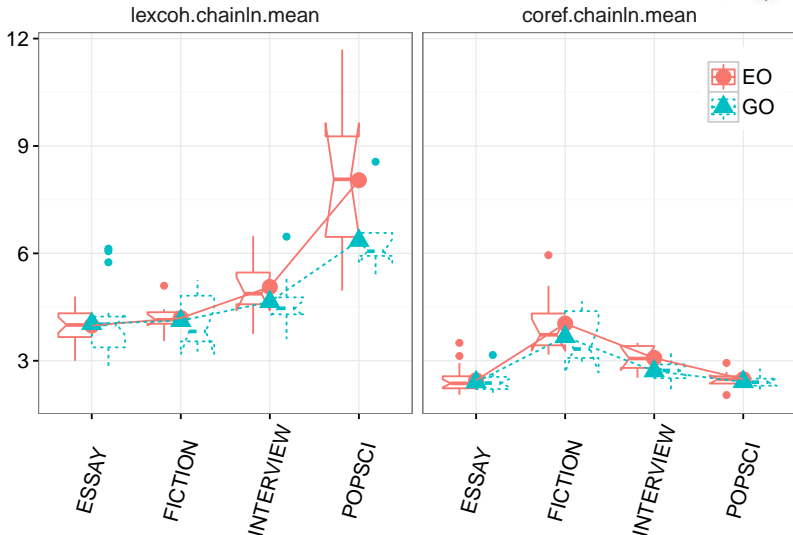
# Results



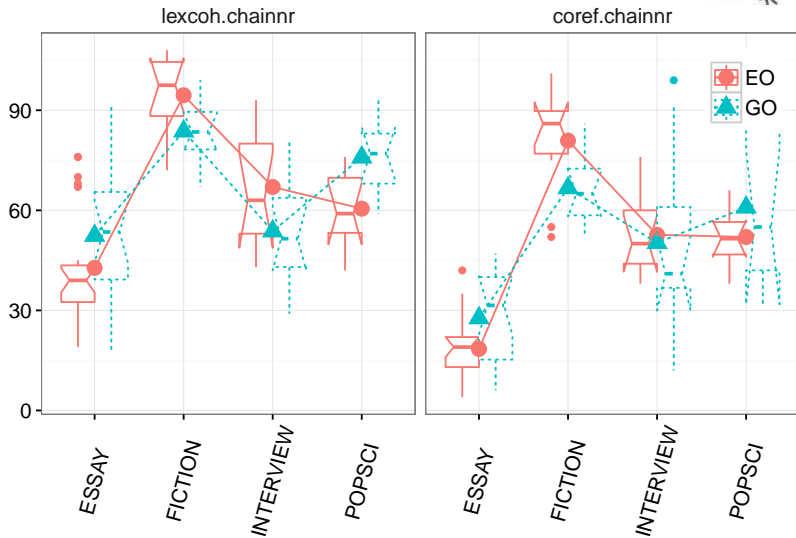


# Topic Development

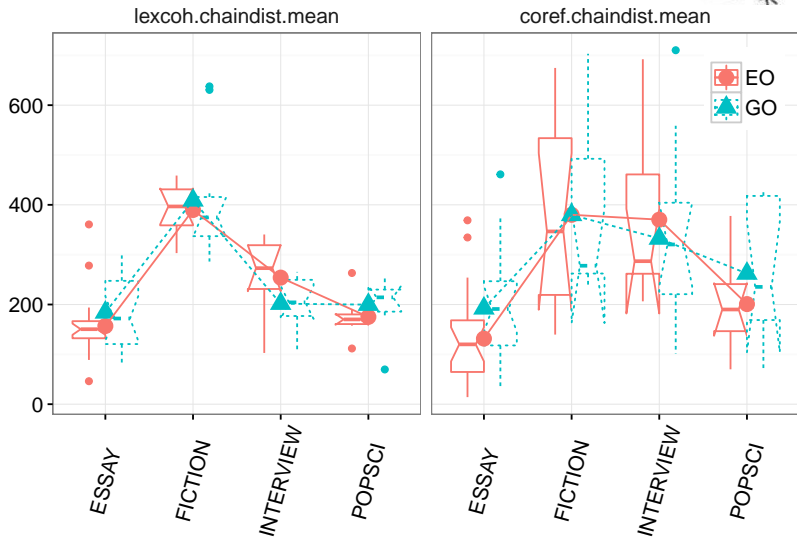
# Chain Length



# Chain Number



# Chain Distance



# Summary Topic Development



- language NOT significant
- language:register partially significant
- register significant
- chain types: partially different behavior

REGISTER	TOPIC DEVELOPMENT	CHAIN FEATURES
POPSCI	topic continuity within one domain	lexical chain length and distance in lexical chains and lexical chain number
ESSAY	topic variation	chain length and number and distance
FICTION	topic continuity of referents and topic interaction	coref. chain length and lexical chain length and chain number and chain distance
INTERVIEW	topic continuity of referents and topic variation	coref chain length and lex. chain distance



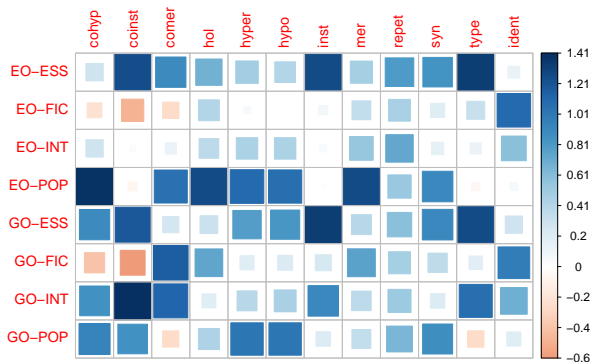
# Semantic Variation

# Semantic Variation and Registers



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

- If ratio  $< 1 \Rightarrow \log(\text{ratio}) < 0$  (negative values)  $\Rightarrow$  red color
- If ratio  $> 1 \Rightarrow \log(\text{ratio}) > 0$  (positive values)  $\Rightarrow$  blue color



# Summary Semantic Variation



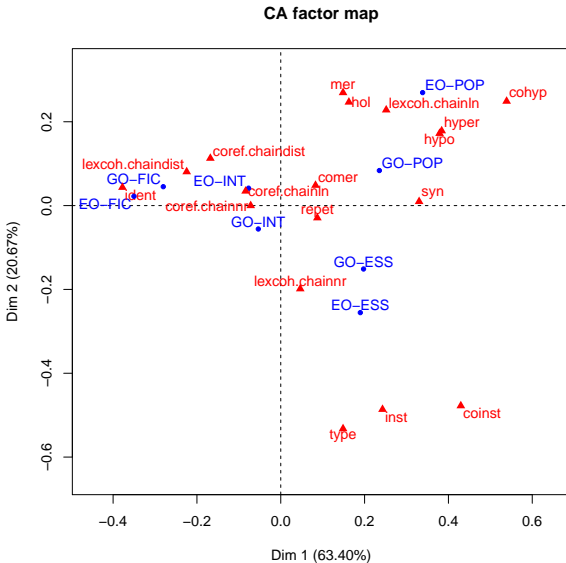
REGISTER	SEM. VARIATION	SEM. FEATURES
POPSCI	yes! E>G	identity, (co-)instance, (co-)hyponymy, hyperonymy, holonymy, (co-)meronymy
ESSAY	no and specific	identity, repetition, instance, type, coinstance
FICTION	yes and specific	identity, coinstance, hyperonymy, hyponymy, GE:comeronymy
INTERVIEW	no and specific, G>E	identity, EN:repetition, GE:type, coinstance, comeronymy





# Summary

# CA with All Features



# Summary



- more differences between registers than between languages
- German registers are more alike
- English registers are more marked
- most contrasts  $E \Leftrightarrow G$  in POPSCI

# Summary



REGISTER	TOPIC DEVELOPMENT	SEM. VARIATION
POPSCI	topic continuity within one domain	yes! E>G
ESSAY	topic variation	no and specific
FICTION	topic continuity of referents and topic interaction	yes and specific
INTERVIEW	topic continuity of referents and topic variation	no and specific, G>E



Thank you!  
Questions?

Information:  
[www.gecco.uni-saarland.de](http://www.gecco.uni-saarland.de)

# Analysed Features



- chain features (coreference, lexical)
  - chain length: number of elements in a chain
  - chain number: number of different chains
  - chain distance: between elements in the same chain
- semantic relations between adjacent chain elements
  - identity
  - repetition
  - antonymy
  - synonymy
  - hyperonymy
  - hyponymy
  - co-hyponymy
  - holonymy
  - meronymy
  - co-meronymy
  - type-instance
  - instance-type
  - co-instance

# Annotation



## Coreference: semi-automatic

- mark coreferential elements
- (distinguish antecedents and anaphors)
- (define antecedent type)
- (define anaphor type)
- assign coreferential elements to a coreference chain

## Lexical chains: mostly manual

- mark lexical elements = nominal expressions
- define the type of sense relation between two adjacent chain elements
- assign lexical elements to a lexical chain

# Sense Relations



- **REPETITION**: orthographical repetition of nominal expressions such as *London* and *London*, or *place* and *place*.
  - *I live in a town called Reigate. It's between **London** and the countryside which is quite nice. It takes us about 25 minutes to get to **London** on the train. It's very nice actually, it's a nice **place** to live. And I grew up in a **place** called Banstead.*
- \* In case of compounding, the second element is the determining factor: *stem cell* but not *stem cell behaviour* or *stem cell research*.
  - *For instance, just identifying a true **stem cell** can be tricky. For scientists to be able to share results and gauge the success of techniques for controlling stem cell behavior... Of course, the goal of stem cell research is to replace...*



# Sense Relations



- **ANTONYMY**: relation of contrast *halving* and *doubling*
  - *Dazu gehören zum Beispiel **die Halbierung** der Energie- und Rohstoffintensität bis 2020 gegenüber 1990 (bzw. 1994) und **die Verdoppelung** des Anteils erneuerbarer Energien am Energieverbrauch bis 2010.*
  
- **SYNONYMY**: total synonymy (*Lebewesen – Organismen*) but also near synonymy, such as between technical and common-language terms, e.g. *belly* and *abdomen*
  - *Wie ist die Übereinstimmung unter diesen Umständen anders zu erklären als durch eine Beziehung, die alle die **Lebewesen** miteinander verbindet...? Und wie anders wäre diese Beziehung zu verstehen als die einer durch Vererbung, durch “genetische Überlieferung” entstandenen Gemeinschaft von **Organismen***

# Sense Relations



- **HYPERONYMY**: in case the superordinate term follows the more specific term: *village* and *place*
  - **HYPONYMY**: in case the specific term follows the superordinate one: *place* and *town*
  - **COHYPONYMY**: between two elements on the same level of specification: *town* and *village*
- *I live in a town called Reigate. It's between London and the countryside which is quite nice. It takes us about 25 minutes to get to London on the train. I say it's a **town**, it's more of a **village**. It's quite small. It's very nice actually, it's a nice **place** to live. And I grew up in a **town** called Banstead.*

# Sense Relations



- **HOLONYMY**: relation, where the whole follows the part: *the EU* following *Britain*
  - **MERONYMY**: part-whole relation, where the part follows the whole: *Bulgaria* following *the EU*
  - **CO-MERONYMY**: succession of two parts that belong to a whole: *Romania, Croatia, Britain* following *Bulgaria*
- *This means that this will not be the last enlargement of **the EU**. **Bulgaria** and **Romania** are waiting in the wings and likely to join in the next 2-3 years. The candidatures of Turkey and **Croatia** are well advanced. We in **Britain** welcome both. In particular we believe that Turkey, if she meets all the criteria, would add a new dimension to **the EU**, and represent a vital reaching out to the Islamic world at a time when such links are more than ever needed.*

# Sense Relations



- **TYPE**: relation between a common noun and a named entity: *town* and *Reigate*
  - **INSTANCE**: relation where the named entity follows the common noun: *Reigate* and *town*, *Banstead* and *town*
  - **CO-INSTANCE**: relation between two named entities (*Reigate* and *London*)
- *I live in a **town** called **Reigate**. It's between London and the countryside which is quite nice. It takes us about 25 minutes to get to London on the train. I say it's a **town**, it's more of a village. It's quite small. It's very nice actually, it's a nice place to live. And I grew up in a **town** called **Banstead** which is fairly close to Reigate.*