



Empirical insights into Cohesion

Contrasting English and German across Modes

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

IÜD, Heidelberg

28/10/16



GECCo

GERMAN ENGLISH CONTRASTS IN COHESION

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

Cohesive Phenomena



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**

Types of Cohesion (cf. Halliday & Hasan 1976)	Meaning relations
Coreference <i>An option ... it/ this option</i>	identity
Substitution <i>Many options ... a good one</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 ... another/ better option</i>	
Cohesive conjunction <i>X. But/ And/ However Y</i>	Logico-semantic relations (addition, contrast, cause, ...)
Lexical cohesion <i>factors ... factors</i>	Similarity (repetition, general nouns, meronymy, Synonymy ...)

Case studies



- **Coreference:** Kunz, K. & E. Steiner (2012). Towards a comparison of cohesive reference in English and German: System and text. In: M. Taboada, et al. (eds.), Contrastive Discourse Analysis. Functional and Corpus Perspectives. London: Equinox.
- **Substitution:** Kunz, K. & E. Steiner (2013). Cohesive substitution in English and German: a contrastive and corpus-based perspective, in: Aijmer, K. & Altenberg, B. eds. 2013. Advances in corpus-based contrastive linguistics. Studies in honour of S. Johansson. Amsterdam: John Benjamins. pp. 201–231.
- **Conjunctions:** Kunz, K. & E. Lapshinova-Koltunski (2014). Cohesive conjunctions in English and German: Systemic contrasts and textual differences. In: Vandelanotte, L. et al. (eds). Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora. Amsterdam/New York: Rodopi, pp. 229-262.
- **Ellipses:** Menzel, K. (2016). Understanding English-German contrasts - a corpus-based comparative analysis of ellipses as cohesive devices, PhD dissertation, Saarland University.
- **Lexical Cohesion:** Kunz, K., E. Lapshinova-Koltunski, J. M. Martínez-Martínez, K. Menzel & E. Steiner (in press). Shallow features as indicators of English-German contrasts in lexical cohesion. In: Languages in Contrast.

cf. also <http://www.gecco.uni-saarland.de/GECCo/deliverables.html>

Cohesive devices



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.

Coreference

Lexical cohesion

Ellipsis

Conjunction

Substitution

Cohesive devices



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.

Cohesive chains - Coreference



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.

Cohesive chains - Lex. Cohesion



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.

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?

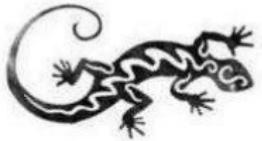
A: No, I think, that's – well, I think ... It's a major thing here.

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.

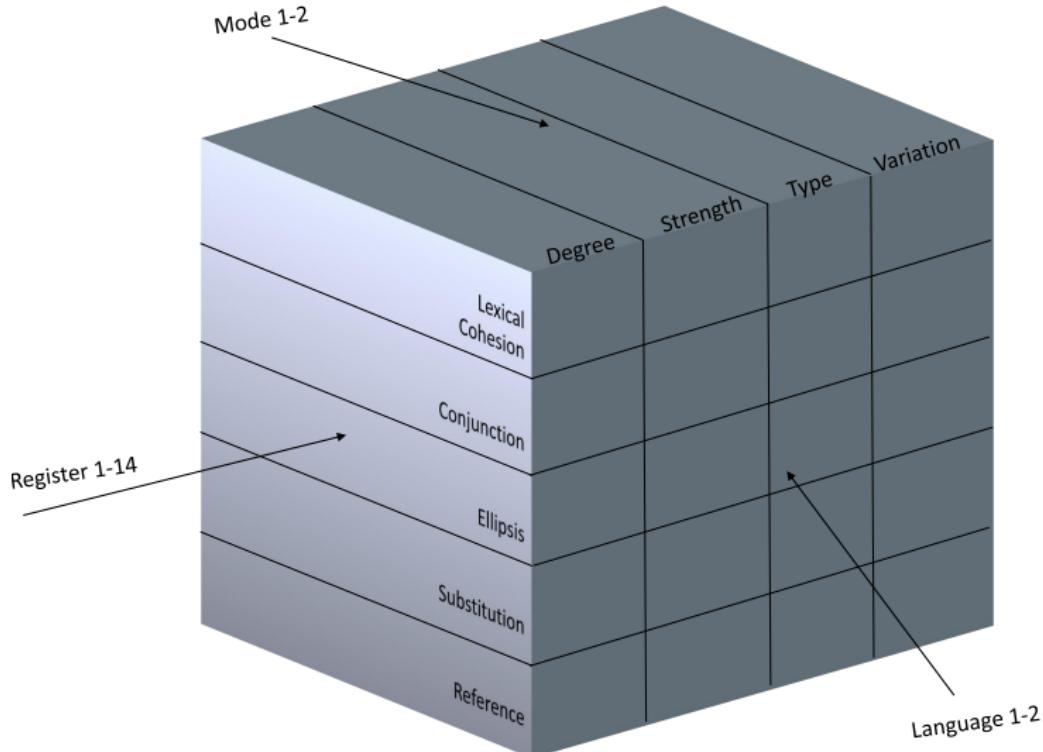
EO-POPSCI-004

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.

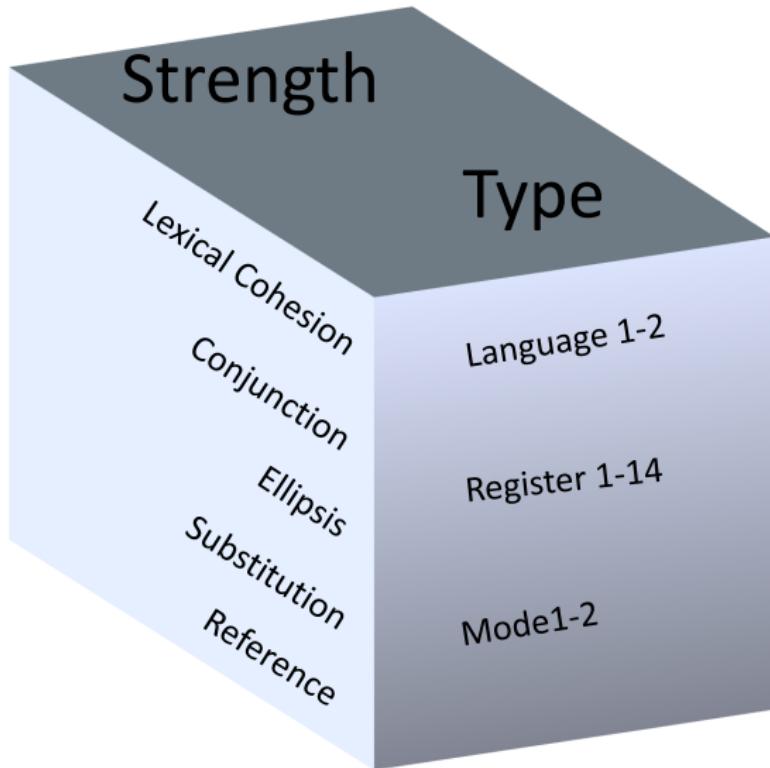


Research Agenda and Methodology

Research Design - Overall



Research Design - Case Study



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**

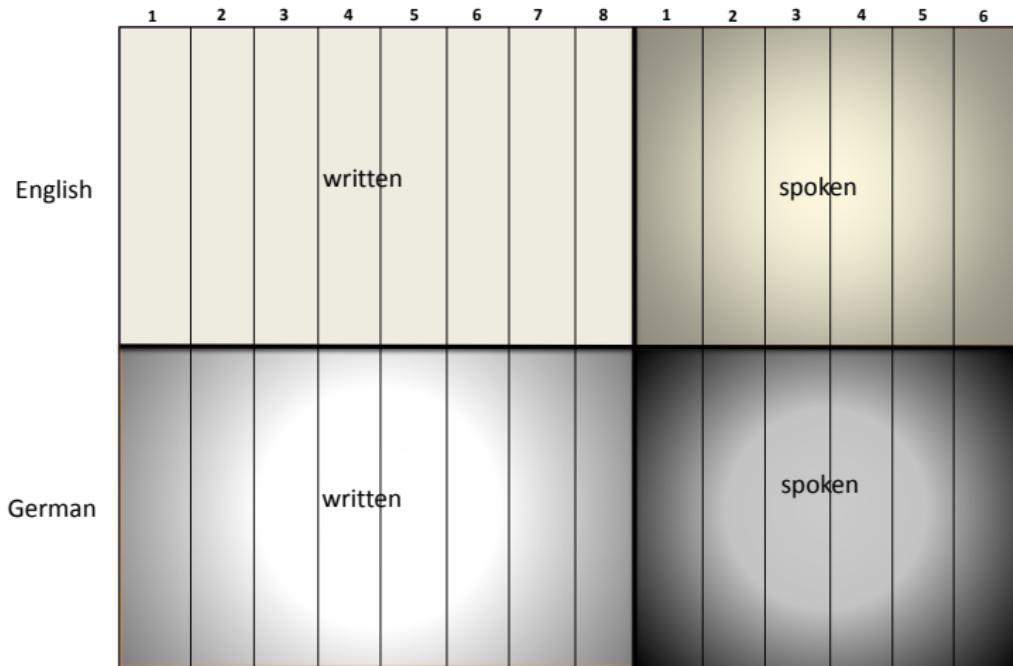
④ Breadth of variation: How much cohesive variation is there for

- Language: English <=> German,
- Mode: spoken <=> written,
- Register?



Corpus Resources

Register 1-14



Corpus Versions



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

GECCOCHAIN		
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

- 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



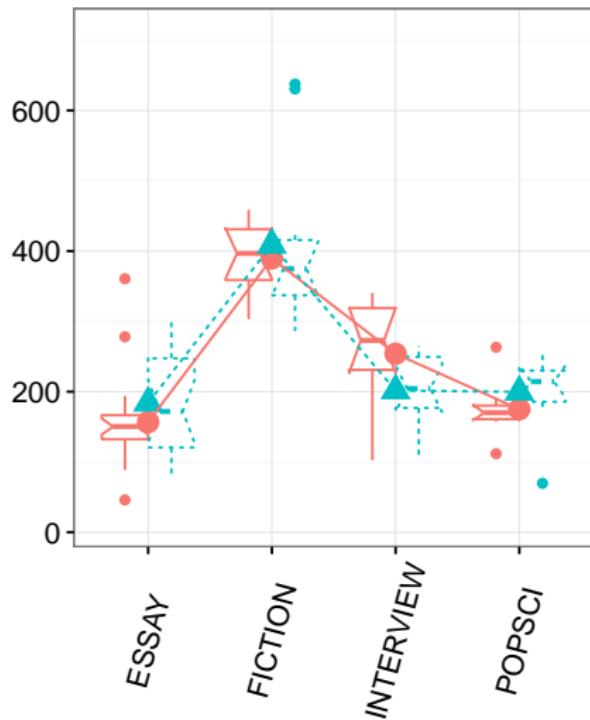
see Kunz et al (2016)

- ① How close are elements in cohesive chains (distance between members of cohesive chains)?
- ② How long are cohesive chains?

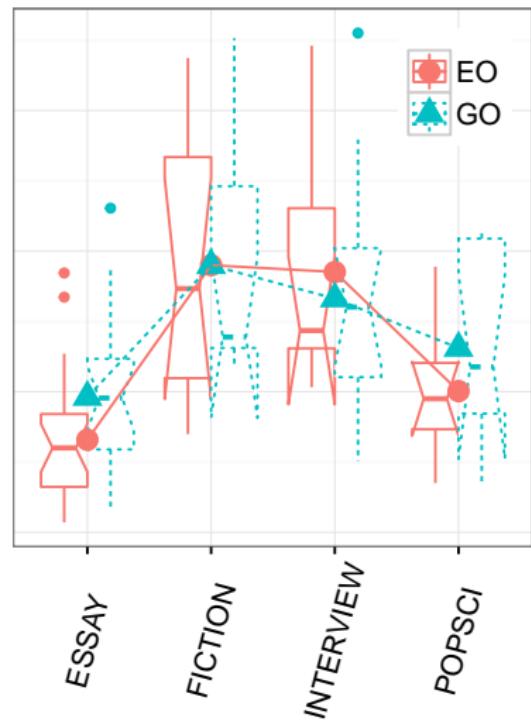
1. Chain Distance



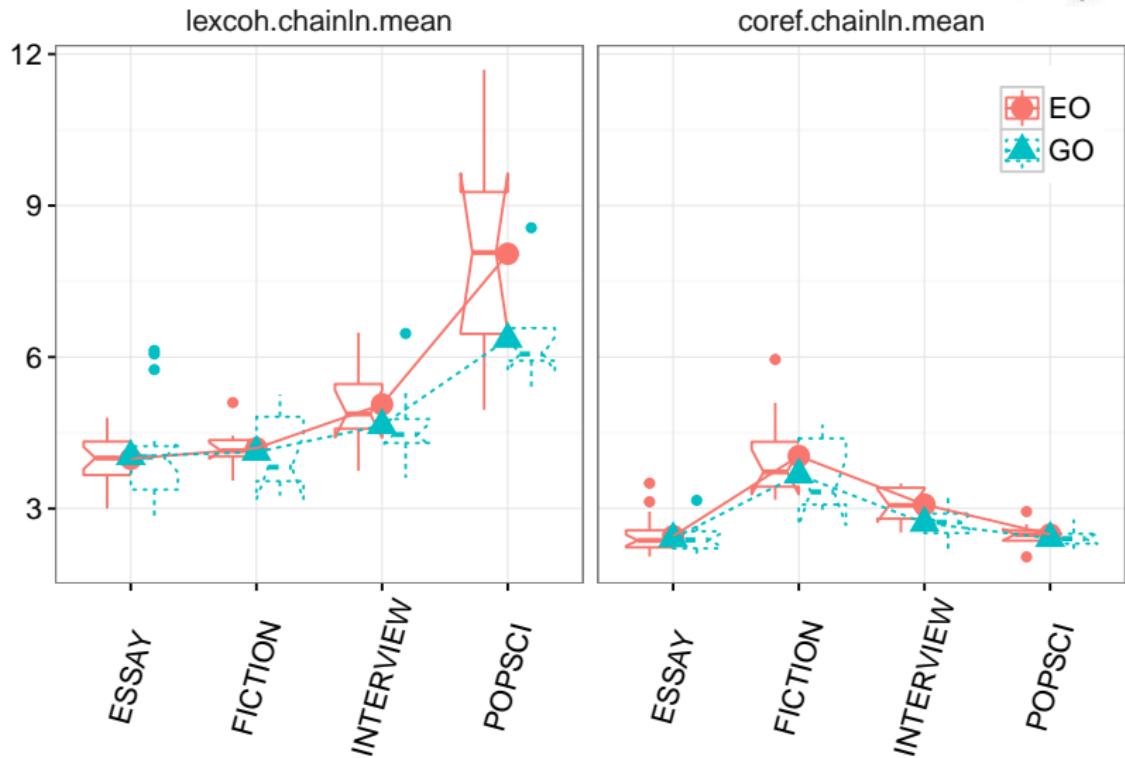
lexcoh.chaindist.mean



coref.chaindist.mean



2. Chain Length



Meaning Relations



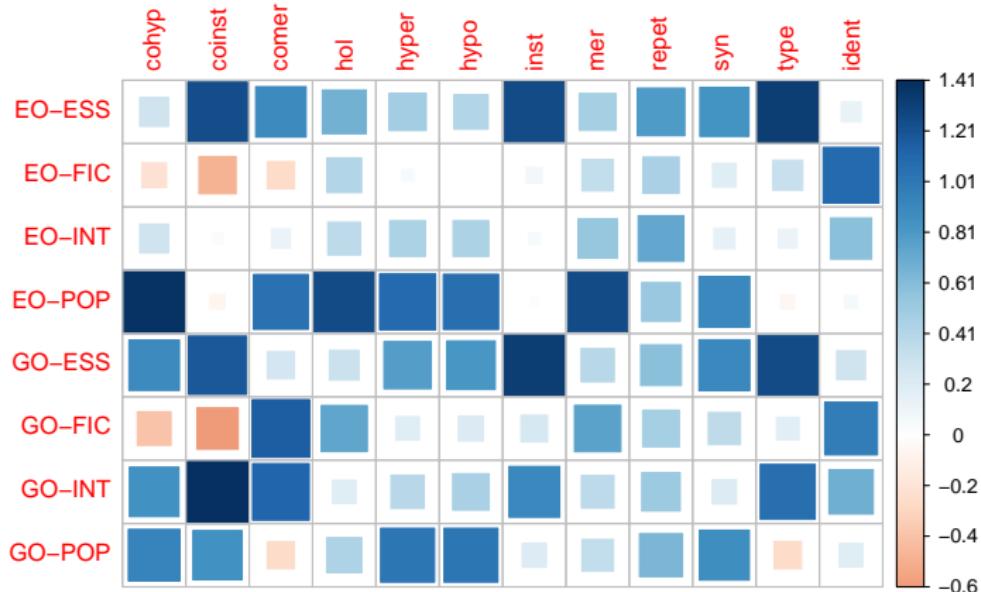
through cohesive devices

- 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

cohesive devices	meaning relations	
coreference	identity	1,2
substitution		
ellipsis	type reference/comparison	2
comparative reference		
cohesive conjunction	logico-semantic relations: addition, contrast, cause, etc.	2
lexical cohesion	similarity: repetition, general nouns, meronymy, synonymy, etc.	1

1. Important Relations

Identity (Coreference) and Similarity (Lexical Cohesion)



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



	Precision	Recall	F-Measure
GO	99.3	98.6	99.0
EO	98.5	99.2	98.9
Weight.Avg.	98.9	98.9	98.9

GO		EO	
4.1984	coref:dem-pronadv	1.0173	coref:comp-general
2.0487	conj:adversat-adverb	1.9675	coref:dem-mod
1.4926	conj:causal-adverb	1.6618	subst:nom
1.4716	subst:claus	1.6057	coref:pers-it
1.1850	coref:dem-local	1.4476	conj:causal-subjun
1.1568	coref:dem-artic	1.3708	conj:temp-subjun
1.0611	conj:addit-adverb	1.1606	subst-verb
1.0585	conj:temp-adverb	1.0173	coref:comp-particular
0.9209	conj:modal-adverb	0.9146	conj:adversat-conn
0.9135	conj:adversat-subjun	0.8751	coref:pers-mod

2. Distinction: Modes



	Precision	Recall	F-Measure
spoken	97.7	91.3	94.4
written	98.3	99.6	98.9
Weight.Av.	98.2	98.2	98.2

	written	spoken
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



Strength

Lexical Cohesion

Conjunction

Ellipsis

Substitution

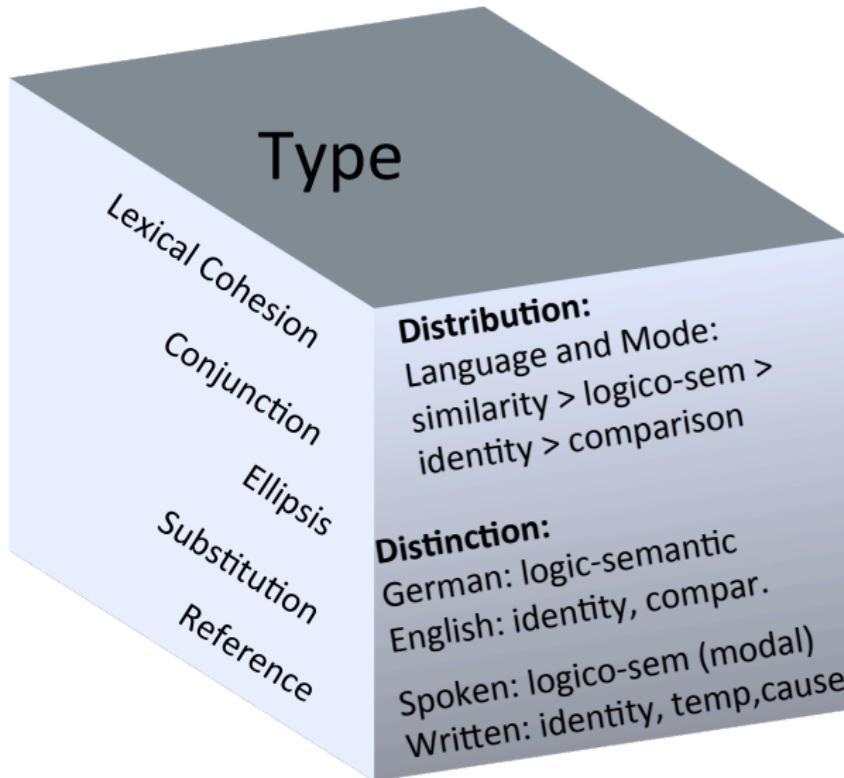
Reference

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

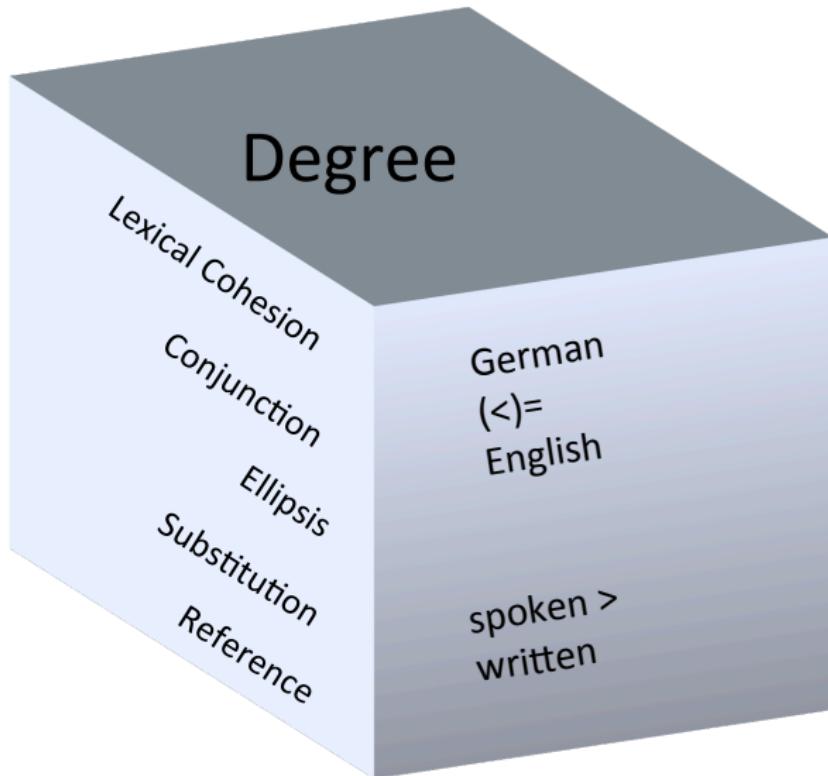
spoken vs. written: wider
scope, less explicit, more
focused

Cohesive chains:
GO = EO
Register specific!

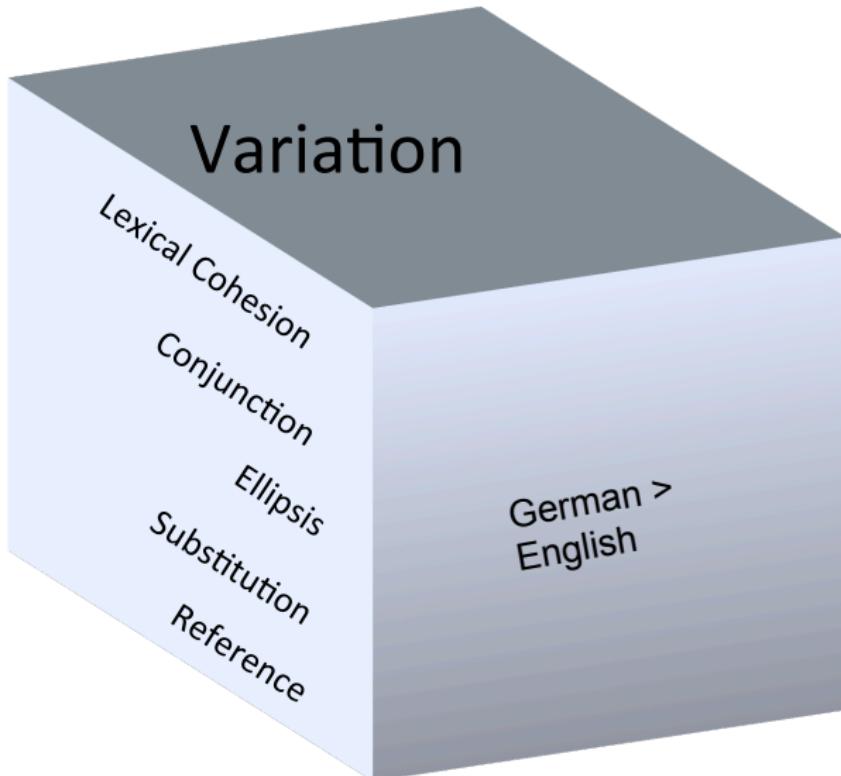
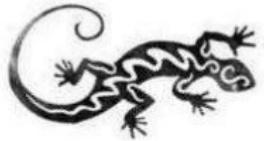
Conclusions



Conclusions



Conclusions



References



- Biber, D. (1995). Dimensions of register variation: A cross-linguistic comparison. Cambridge: CUP.
- Biber, D. (1988). Variation across speech and writing. Cambridge: CUP.
- Halliday, M.A.K. & R. Hasan (1976). Cohesion in English. Longman, London.
- Hansen-Schirra, S., S. Neumann & E. Steiner (2012). Cross-linguistic Corpora for the Study of Translations. Insights from the language pair English - German. Series Text, Translation, Computational Processing. Berlin, New York: Mouton de Gruyter.
- Hawkins, J. A. (1986). A comparative typology of English and German. Unifying the contrasts. London etc. Croom Helm.
- House, J. (2002). Maintenance and convergence in covert translation English - German. In: Hasselgard, H., S. Johansson, B. Behrens and C. Fabricius-Hansen. (eds.). Information Structure in a Cross-Linguistic Perspective. Amsterdam - New York: Rodopi
- König, E. & V. Gast (2012). Understanding English-German Contrasts. Grundlagen der Anglistik und Amerikanistik. Berlin: Erich Schmidt Verlag. [3rd, extended edition].
- Königs, K. (2011). Übersetzen Englisch-Deutsch, Lernen mit System. Oldenbourg Verlag.
- Kunz, K., E. Lapshinova-Koltunski & J. M. Martínez-Martínez (2016). Beyond Identity Coreference: Contrasting Indicators of Textual Coherence in English and German. In: Proceedings of CORBON at NAACL-HLT2016, San Diego.
- Kunz, K., S. Degaetano-Ortlib, E. Lapshinova-Koltunski, K. Menzel & E. Steiner. (in press.a). GECCo - an empirically-based comparison of English-German cohesion. In: De Sutter, G., I. Delaere & M.-A. Lefer (eds.). New Ways of Analysing Translational Behaviour in Corpus-Based Translation Studies. TILSM series. Mouton de Gruyter.
- Kunz, K., E. Lapshinova-Koltunski, J. M. Martínez-Martínez, K. Menzel & E. Steiner (in press.b). Shallow features as indicators of English-German contrasts in lexical cohesion. In: Languages in Contrast.
- Lapshinova-Koltunski, E., K. Kunz & M. Amoia (2012). Compiling a Multilingual Corpus. In Mello, H., and M. Pettorino (eds). Proceedings of the VIIthGSCP-2012. Speech and Corpora. Firenze: Firenze University Press.
- Lapshinova-Koltunski, E. & K. Kunz (2014). Detecting Cohesion: semi-automatic annotation procedures. In: Proceedings of Corpus Linguistics. Lancaster, UK.

References



- Leech, G., M. Hundt, C. Mair, & N. Smith (2009). Change in Contemporary English. A Grammatical Study. Cambridge: CUP.
- Mair, C. (2006). Twentieth-Century English. History, variation and standardization. Cambridge: CUP.
- Martínez Martínez, J. M., E. Lapshinova-Koltunski & K. Kunz (2016). Annotation of Lexical Cohesion in English and German: Automatic and Manual Procedures. In: Proceedings of the Conference on Natural Language Processing (Konferenz zur Verarbeitung natürlicher Sprache) - KONVENS-2016, September, Bochum, Germany.
- Menzel, Katrin (2016). Understanding English-German contrasts - a corpus-based comparative analysis of ellipses as cohesive devices, PhD Dissertation. Universität des Saarlandes.
- Neumann, Stella (2013). Cross-linguistic Register variation. Mouton De Gruyter.