Beyond Identity Coreference: Contrasting Indicators of Textual Coherence in English and German

Kerstin Kunz
Universität Heidelberg
Heidelberg, Germany

Ekaterina Lapshinova-Koltunski
Universität des Saarlandes
Saarbrücken, Germany

José Manuel Martínez
Universität des Saarlandes
Saarbrücken, Germany

kerstin.kunz@iued.uni-heidelberg.de
{e.lapshinova,j.martinez}@mx.uni-saarland.de

Abstract

This paper focuses on the interaction of chains of coreference identity with other types of relations, comparing English and German data sets in terms of language, mode (written vs. spoken) and register. We first describe the types of coreference and the chain features analysed as indicators of textual coherence and topic continuity. After sketching the feature categories under analysis and the methods used for statistical evaluation, we present the findings from our analysis and interpret them in terms of the contrasts mentioned above. We will also show that for some registers, coreference types other than identity are of great importance.

1 Introduction

This paper presents the findings from an empirical analysis of different types of coreference chains in a corpus of English and German written and spoken registers. The motivation lying behind our study is twofold.

First, our objective is to analyse the interaction between typical relations of coreference identity and other relations in chains, e.g. type-instance, part-whole, etc. Example (1) illustrates a coreference chain consisting of the antecedent *town* and anaphors expressed via It/it, and a lexical chain with the elements *town*, *Reigate*, *London*, *village*, *place*, *Banstead*, establishing relations of type/(co-)instance (e.g. *town – Reigate/London*), hyper-/hyponymy (*place – village/town*).

(1) *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 place called Banstead which is fairly close to Reigate.*

We focus on the features of such chains, including the type of the semantic relations, the distance between chain elements, the number of chains as well as chain length. Hence, the focus is not on the formal types of anaphors and antecedents but on the relations themselves. These features serve as indicators of how coherence and topic continuity are overtly expressed in texts.

Second, another aim of this study is to analyse specific properties of textual coherence. The intention here is to see which factor plays a more important role for variation in coherence: language (English vs. German), mode of production (written vs. spoken) or register (political essays, interviews, popular scientific texts and fictional texts containing dialogues).

Knowledge about these coherence types may advance automatic multilingual coreference resolution, which is, however, beyond the scope of this study.

2 Related Work

Most existing studies concentrate on the properties of anaphors and antecedents only, describing their structural and functional properties. Some of them,

mostly theoretical, e.g. (Ariel, 2001; Kibrik, 2011), are related to the model of referential choice based on the degree of referent salience. Algorithms describing the degree of salience were presented by Hajičová et al. (2006); Lambrecht (1994), Strube & Hahn (1999), etc.

There also exists a considerable amount of large-scale annotated data for coreference, anaphoric relations, event anaphora, bridging relations and so on, compiled mostly for automatic anaphora resolution (MUC-6 (1995), MUC-7 (1997) or ACE NIST (Doddington et al., 2004) and more recently CoNNL 2011 (Pradhan et al., 2011). However, most of these corpora are monolingual and cannot be applied for a multilingual analysis, as they do not contain comparable registers across languages. For instance, the largest coreference corpus for English is OntoNotes (Technologies, 2006) containing several genres. For German, the TüBa-D/Z corpus (Telljohann et al., 2012) was annotated with semantic and coreference information, but contains newspaper texts only.

Besides, the number of studies that base their analysis on corpora annotated with chains, e.g. as described in (Zikánová et al., 2015; Lapshinova-Koltunski and Kunz, 2014a) , is rather small in contrast to those just using the annotation of relations. Yet, an extensive comparison from a multilingual perspective is missing.

3 Phenomena under analysis

Our focus therefore is on the textual relations set up in what we call **chains** in our study. A chain minimally consists of a tie between an anaphor and an antecedent, yet many chains are larger and contain several anaphors. We mainly distinguish two types of chains: **coreference chains** and **lexical chains**.

The coreference chains in our framework not only include relations of identity between **entities**, as between *town* and *it* in example (1) above, but also **abstract** and **situation anaphora** as in (2), where *That* refers to the underlined preceding clause.

(2) *They may cry, and we find it very hard to find out why ... That's difficult.*

The formal types of anaphors annotated in coreference chains are mainly function words (i.e. grammatical types of cohesion) and include personal and possessive pronouns, demonstrative determiners, demonstrative pronouns, the definite article, and local and temporal adverbs (*here, there, now, then*). The annotated antecedents may include NPs, clauses, clause complexes (see example (2)) or even larger textual chunks.

The lexical chains analysed in this study contain lexical relations between nominal antecedents and anaphors (nouns or nominal compounds), which vary in terms of the semantic relation between the chain elements. Relations include repetitions as between *London - London*, *place - place*, hyperonymy as between *place - town* and *place - village* in example (1) in Section 1 above and others. They are comparable to what is called **bridging** or **indirect anaphora** in the state-of-the-art literature.

Note that the two types of chains may interact as in example (3), where coreference relations are set up between the two grammatical anaphors (the demonstrative pronoun *this* and the definite article *the*) to the antecedent *reward system*. In addition, a lexical relation of hyper-/hyponymy holds between *this system* (hyperonym) and a relation of repetition between *the system* and this antecedent.

(3) *Neurobiologists have long known that the euphoria induced by drugs of abuse arises because all these chemicals ultimately boost the activity of the brain's reward system: a complex circuit of nerve cells, or neurons, that evolved to make us feel flush after eating or sex... At least initially, goosing this system makes us feel good... But new research indicates that chronic drug use induces changes in the structure and function of the system's neurons...*

The chain features of coreference and lexical cohesion analysed in this study include: 1) **chain length** concerns the number of elements in a coreference or a lexical chain; 2) **chain number** concerns the number of different coreference and lexical chains per text; 3) **chain distance** is the distance in tokens between different elements in the same chain. 4) **semantic relation** is the type of semantic relation between adjacent chain elements. The types analysed in this study are: 1) identity, for all elements in a coreference chain; and all kinds of other re-

lations, namely: 2) repetition, 3) antonym, 4) synonym, 5) hyperonym, hyponym and cohyponym, 6) holonym, meronym, comeronym, 7) type, instance, coinstance. Apart from analysing each of these features separately we also study their interaction.

4 Methods and resources

4.1 Research questions

The chain features described in Section 3 indicate how coherence is created and how and to which degree topics are distributed throughout the texts. For instance, we may find long coreference chains in combination with a small distance between the respective chain elements. This points to high **topic continuity in terms of certain referents**. Furthermore, **topic continuity within one domain** is expressed by long lexical chains with small distance between elements. Small chains and low distance in combination with a high number of different chains hints at a high degree of **topic variation**, i.e. that text producers often jump from one topic to another. By contrast, long chains and high distances between elements indicates **topic interaction** i.e. that there are several important topics which are interwoven. Moreover, repetition in combination with coreference points to low **semantic variation** whereas relations of lexical cohesion such as type-instance and meronymy point to high variation.

We are additionally interested in which contrasts are more pronounced, those concerning language, mode or register (see Section 1) and which of the analysed features mainly contribute to these contrasts.

4.2 Data

For the research aims within this study, we use a data set containing texts of both written and spoken discourse. The written part was extracted from the corpus described in (Hansen-Schirra et al., 2012), whereas the spoken subcorpus was extracted from the corpus described in (Lapshinova-Koltunski et al., 2012). The whole corpus is annotated on various linguistic levels, including parts-of-speech (POS), chunks, clauses, sentences, and various devices of cohesion, i.e. coreference, discourse relations, elliptical constructions and substitution annotated as described by Lapshinova & Kunz (2014b)[1]. Relations of coreference other than identity (synonymy, antonymy, hyponymy, etc.) were annotated for the subset of the data analysed in this study. The registers included are political essays (ESSAY), popular-scientific articles (POPSCI), fictional excerpts (FICTION) and transcribed interviews (INTERVIEW). ESSAY and POPSCI represent written discourse, INTERVIEW represents spoken discourse, whereas FICTION is on the borderline, as it contains both written and spoken elements in the form of dialogues. INTERVIEW and FICTION additionally share narrative elements. The details on the analysed subset are provided in Table 1.

	EO		GO	
register	**texts**	**tokens**	**texts**	**tokens**
ESSAY	23	27171	20	31407
FICTION	10	36996	10	36778
INTERVIEW	9	30057	12	35036
POPSCI	8	27055	9	32639

Table 1: Corpus description.

4.3 Visualisation techniques

We use various techniques to investigate the distributional characteristics of subcorpora in terms of occurrences of the features described in Section 3 and to answer the questions in Section 4.1.

Box plots are used to visualise a summary of the distribution underlying a particular sample and to compare central measure values and spread of the data across groups. Special attention is given to the median (second quartile) and the IQR (range between the first and third quartile). Box plots have lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles. We use notched box plots to see if the differences between the variables are significant: if the notches of two box plots overlap, there is no evidence that their medians differ (Chambers et al., 1983). The means are also plotted for the sake of completeness. The evidence from box plots is confirmed with two-way factorial **ANOVA** tests for the significance of the differences between languages

[1] More information about the corpus and how to gain access to it can be found at `http://hdl.handle.net/11858/00-246C-0000-0023-8CF7-A`.

(English vs. German), registers (ESSAY vs. FICTION, etc.) and the interaction between these two variables. η^2 is calculated to show the variance explained by the variables and their interaction.

Mosaic plots are used to visualise a table and to examine the association between the variables. For each cell, the height of bars is proportional to the observed relative frequency. The colours indicate the standard deviation of the expected count in chi-square testing (or standardised residuals). If row and column variables are completely independent (no association), the mosaic bars for the observed values are aligned to the mosaic bars for the expected values. In case of an association, the bars are coloured according to the standardised residuals. Standardised residual is a measure of the strength of the difference between observed and expected values, and thus, a measure of how significant your cells are to the chi-square value. This helps to see which cells are contributing the most to the value.

Correlation plots are used to visualise correlations between various variables under analysis. For this, we calculate row and column profiles. The profile of a given row/column is calculated by taking each row/column point and dividing by the sum of all row/column points. Then, the squared distance is computed between each row/column profile and the other rows/columns in the table, resulting in a distance matrix (a kind of correlation matrix), which can be visualised with a correlation plot.

Correspondence analysis (CA) is a multivariate technique to observe similarities and differences between the variables under analysis using an entire set of features in interaction. It enables us to see how certain features are grouped together and where the biggest differences and similarities lie, see (Venables and Smith, 2010; Baayen, 2008; Greenacre, 2007) for details. Moreover, we are able to trace the interplay of categories of the cohesive devices under analysis. The output is plotted into a two dimensional graph. The position of the points indicates the relative importance of a feature for a subcorpus.

5 Results

5.1 Chain length

Figure 1 visually summarises the average chain length distributions across languages and registers

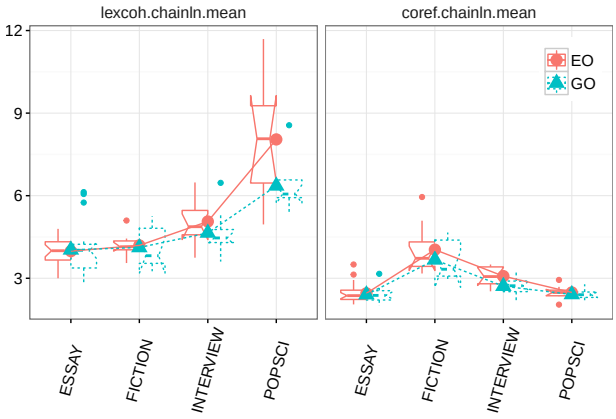

Figure 1: Average length chain: lexical cohesion vs. coreference.

for lexical and coreference chains. Regarding **lexical chains**, the data shows a significant difference between registers ($\nmid .05$) and for the interaction *language:register*, whereas the difference between languages is not significant (see p values in Table 2). The register featuring the longest lexical chains and a more distinctive behaviour is POPSCI, while the other registers tend to show ($\approx 50\%$) shorter chains and differences are not so marked. As for **coreference chains**, only the difference between registers is statistically significant (see Table 3). FICTION is the register showing the longest coreference chains and the clearest difference when compared to the other registers. η^2 in Tables 2 and 3 confirms that the independent variable *register* is the factor explaining a higher proportion ($\approx 60\%$) of the variation observed for both types of chains, whereas the effect of *language* and the interaction *language:register* is negligible.

	p	η^2
language	0.0512459	0.02
register	0.0000000	0.59
language:register	0.0135995	0.04

Table 2: Two-way factorial ANOVA significance tests and effect sizes for lexical chains.

	p	η^2
language	0.0488657	0.02
register	0.0000000	0.60
language:register	0.4213339	0.01

Table 3: Two-way factorial ANOVA significance tests and effect sizes for coreference chains.

5.2 Chain number

Figure 2 shows a significant difference between registers for **lexical chains** with respect to the number of chains, and for the interaction *language:register* ($\nmid$ .05). The register featuring the highest number of lexical chains is FICTION, at the other end of the spectrum we find ESSAY ($\approx 33\%$ of FICTION), both registers show a clearcut difference when compared with INTERVIEW and POPSCI, which are located somewhere in the middle. Although the difference between languages is not significant (see p values in Table 4), there is an interesting difference in terms of register ranking. As for **coreference chains**, the picture is almost the same. We observe a slightly lower number of coreference chains than lexical chains. Nevertheless, the difference is not so marked as with chain length. η^2 in Tables 4 and 5 confirms that the independent variable *register* is the factor explaining a higher proportion ($\approx 50\%$ for lexical chains, $\approx 60\%$ for coreference) of the variation observed for both types of chains, whereas the effect of *language* and the interaction *language:register* is negligible.

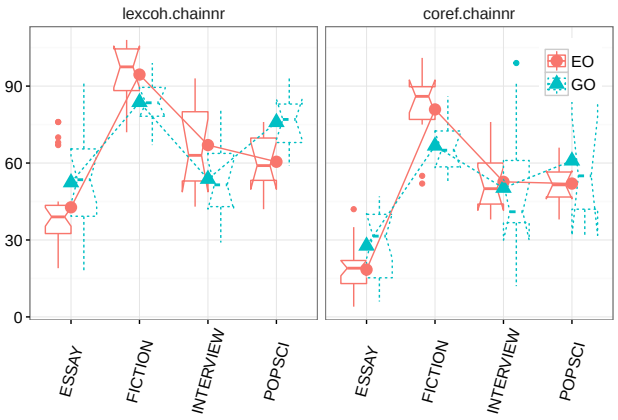

Figure 2: Number of chains: lexical cohesion vs. coreference.

	p	η^2
language	0.5652659	0.00
register	0.0000000	0.47
language:register	0.0060930	0.06

Table 4: Two-way factorial ANOVA significance tests and effect sizes for lexical chains.

	p	η^2
language	0.4941296	0.00
register	0.0000000	0.62
language:register	0.0338786	0.03

Table 5: Two-way factorial ANOVA significance tests and effect sizes for coreference chains.

5.3 Chain distance

Regarding **lexical chains**, Figure 3 shows a significant difference between registers, taking into account the average distance between elements of the chains. By contrast, the difference between languages is not significant (see p values in Table 6). The register showing the greatest average distance between elements in lexical chains is FICTION, at the other end of the spectrum we find ESSAY ($\approx 50\%$ of FICTION), both registers disassociate from INTERVIEW and POPSCI, which are located somewhere in the middle. **Coreference chains** show a completely different picture this time: Differences between registers are again significant. However, FICTION and INTERVIEW stand out as the registers with the highest distance between elements of chains, ESSAY is again the register showing the lowest distance, and POPSCI is situated in the middle. Quite remarkably, there clearly is a higher spread of the distributions for **coreference** than for **lexical chains** denoted by the IQR and the standard deviation reaching proportions up to 1 to 3 in some cases. The magnitude and range of the values is very similar for both lexical and coreference chains. Finally, η^2 in Tables 6 and 7 confirms that the independent variable *register* is the factor explaining a higher proportion of the variation observed for both lexical chains ($\approx 60\%$) and to a less extent for coreference ($\approx 30\%$), whereas the effect of *language* and the interaction *language:register* is negligible.

	p	η^2
language	0.5291255	0.00
register	0.0000000	0.61
language:register	0.1872871	0.02

Table 6: Two-way factorial ANOVA significance tests and effect sizes for lexical chains.

5.4 Combination of chain features

First, we prove if there is an association between the variables under analysis using a mosaic plot illus-

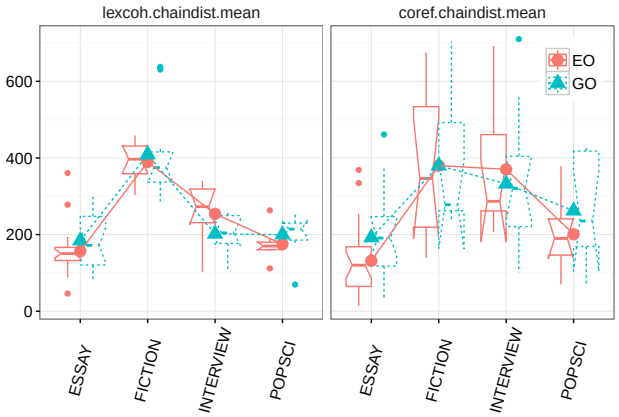

Figure 3: Average chain distance: lexical cohesion vs. coreference.

	p	η^2
language	0.3216957	0.01
register	0.0000001	0.32
language:register	0.5594535	0.02

Table 7: Two-way factorial ANOVA significance tests and effect sizes for coreference chains.

trated in Figure 4.

Blue colour indicates that the observed value is higher than the expected value if the data were random, whereas red colour specifies that the observed value is lower. The number of lexical chains is very important in both English and German for ESSAY, distance between elements in coreference chains plays a great role in INTERVIEW (however, more in English than in German). The distance between elements in lexical chains is strong in FICTION (however, more in German than in English). Overall, this confirms our observations in Sections 5.1, 5.2 and 5.3 above.

We then produce a correlation plot on the basis of squared distances as explained in 4.3 above. The size and the colour of the circle in the plot is proportional to the magnitude of the distance between register profiles, see Figure 5.

We see that cross-lingual differences between registers (e.g. EO-ESSAY vs. GO-ESSAY) are smaller than intralingual distances between registers of one language (e.g. GO-FICTION vs. GO-ESSAY). This, again, confirms our observations in the previous Sections, where we saw a prevalence of the variable *register* in the variation in our data.

Next, we analyse the association between chain

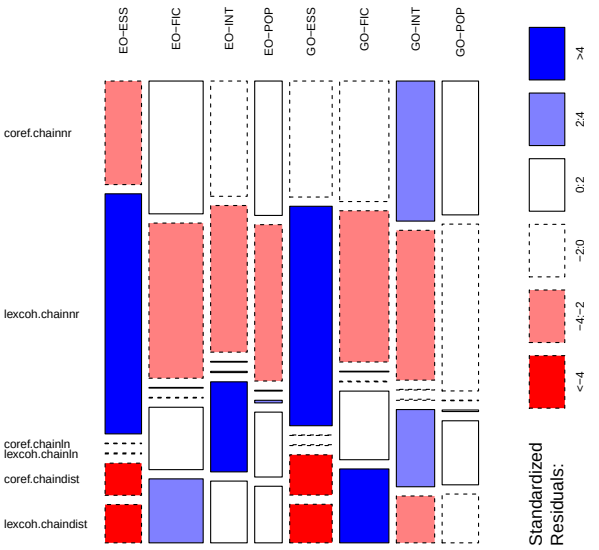

Figure 4: Standardized residuals for chain features.

properties and the registers, which is possible with a correlation graph on the basis of log-likelihood ratio (calculated on the basis of observed and expected values of chi-squared test), see Figure 6. Blue colour of the cell means a positive value and a log(ratio) that is higher than 0, whereas red colour would mean a negative value with a log(ratio) below 0. Cell size and colour intensity indicate the strength of the association.

As seen in Figure 6, all chain features are positively associated with all registers of both languages though there are certain preferences. For instance, length of lexical chains is of special importance in POPSCI, especially in English, whereas their number is more specific for English and German ESSAY. Distance between elements in chains play a greater role for INTERVIEW and FICTION, as already seen above (see Figure 4).

5.5 Semantic relations

For the semantic relations under analysis (see Section 4.1 above), we start with the association between the variables proved with a mosaic plot illustrated in Figure 7.

This plot clearly shows that identity relations are more important in both fictional registers, repetitions in English essays and interviews, hyponymy and meronymy relations are more typical for both popular-scientific registers. Fictional texts in both languages show strong preferences for using coref-

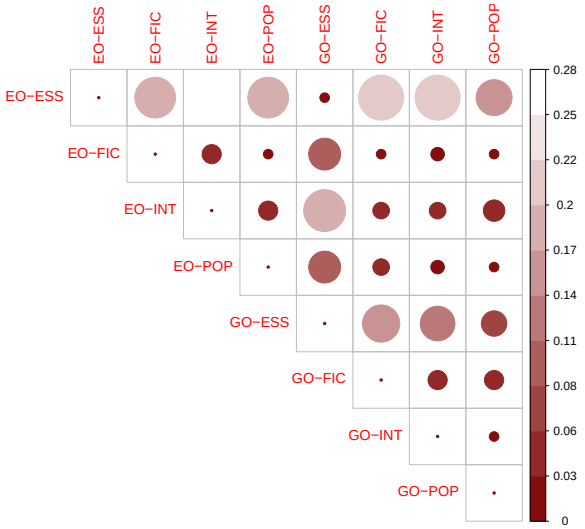

Figure 5: Correlations between language-register profiles.

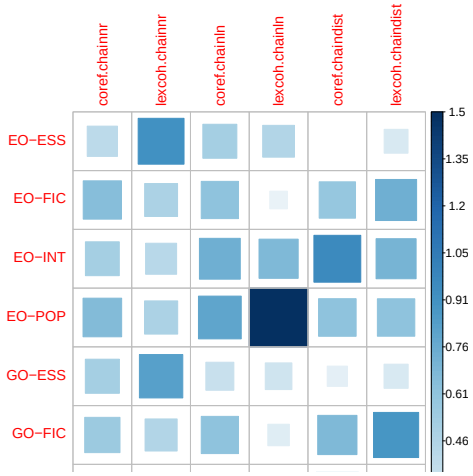

Figure 6: Associations between chain features and registers.

erence, whereas this semantic type is much less important for political essays and popular-scientific articles. In general, the coloured patterns for registers of both languages seem to be similar, which confirms the observation on the strength of registerial contrast in our data.

Next, we analyse the correlations between register profiles in our data (based on distance matrix), visualised in Figure 8.

Again, cross-lingual differences between registers are smaller here. At the same time, the intralingual differences between registers seem to be greater in English than in German, since the circles are bigger on the left upper part of the plot. This confirms the observations from our previous analyses on lexical

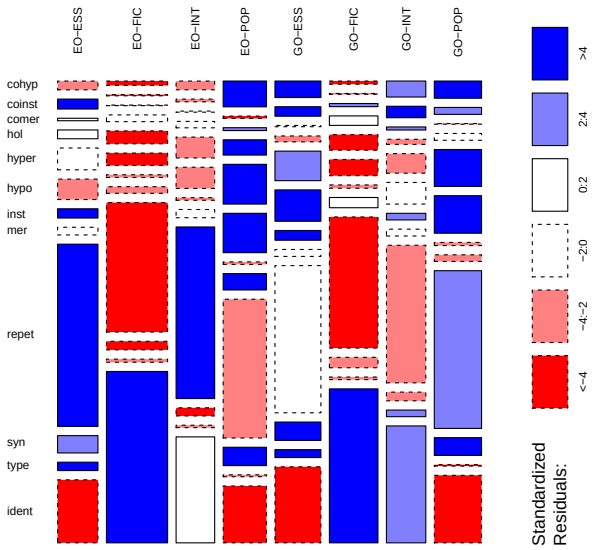

Figure 7: Standardized residuals for semantic relations.

cohesion, in which we used a set of shallow lexical features (TTR, LD, most frequent words, etc.). As for coreference, an opposite effect was observed in (Kunz and Lapshinova-Koltunski, 2015) for the same language pair.

Analysing association between semantic relations and the registers in a correlation plot (Figure 9) produced on the basis of log-likelihood ratio, we see that our previous observations are confirmed here too: relations of identity are strongly associated with FICTION, hyponymy and meronymy with POPSCI. Instance-type relations are typical for ESSAY.

In general, the registers with week identity associations (ESSAY and POPSCI) tend to show a strong association to other relations, i.e. hyper-/hyponymy, type-instance, etc., whereas semantic relations tend to show a lower association (FICTION and partly INTERVIEW) when the identity association is strong. This means that for certain registers (e.g. narrative ones), chain relations other than identity play a minor role.

5.6 Feature combination

In the last analysis step, we combine all the features under analysis , to map the correlations between them, as well as between registers applying CA, see Figure 10.

The plot provides us with two multilingual subcorpora groupings: FICTION and INTERVIEW on the left side, and ESSAY and POPSCI on the right

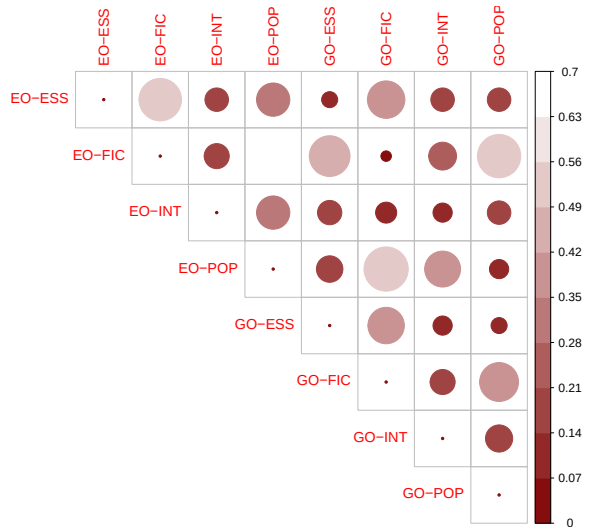

Figure 8: Correlations between language-register profiles.

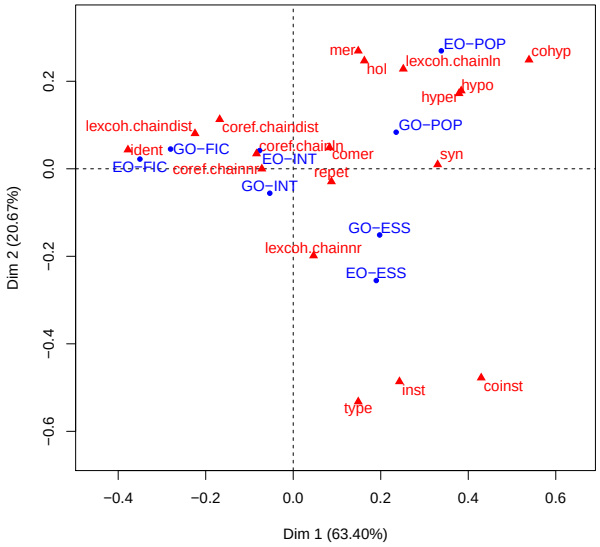

Figure 10: Correspondence analysis for all features.

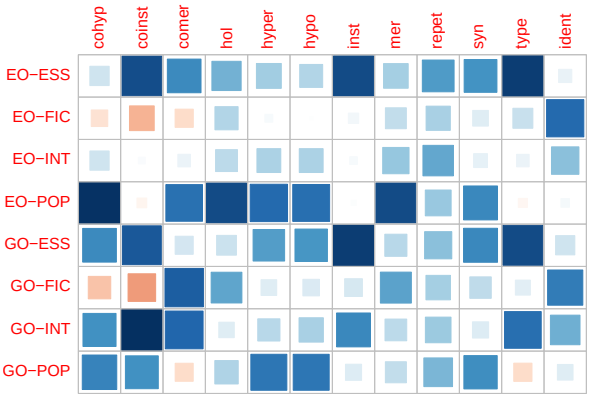

Figure 9: Associations between semantic relations and registers.

side. This suggests a subdivision according to mode: written vs. spoken. In terms of features, those related to lexical cohesion are on the right side of the x-axis, and coreference-related ones on the left side. However, the distance between elements of lexical chains seems to have a correlation with the relation of identity and its chain properties, especially with the distance between elements of coreference chains (as their points are situated close to each other on the plot). Length and number of coreference chains also have a correlation and are especially important for interviews. We also observe groupings of the subtypes of semantic relations on the plot, e.g. meronym and holonym; hyperonym, hyponym and cohyponym; type, instance and coinstance.

6 Discussion

Altogether, registerial differences are more pronounced than language differences, at least for the language pair English-German. The differences and similarities observed between the registers seem to reflect typical situational configurations, some of which pointing to differences between written and spoken discourse.

In POPSCI, we find a relatively low number of long lexical chains in which the distance between elements is relatively low, in combination with a medium number of short coreference chains with low distance. This goes along with relatively high semantic variation and few repetitions, as compared to the other registers. The chain features express continuity within one topic domain and a detailed development of these topics, reflecting the intention of information distribution. In ESSAY, we observe the lowest frequencies for chain number, chain length and distance in both coreference and lexical chains, pointing to a generally lower textual coherence and much topic variation. The frequent use of repetitions serve the communicative goal of persuasion. FICTION is characterized by a high number of short lexical chains and long coreference chains, with a high distance between elements of the two chain types, and with much semantic variation in lexical chains. Thus there is a focus on specific referents reflecting a narrative style together with

extensive use of lexical resources available in the two languages. INTERVIEW features longer chains than ESSAY and POPSCI but shorter ones than FICTION, a medium number of chains which is below FICTION and a chain distance as high as in FICTION for coreference. Apart from that, INTERVIEW equals ESSAY in low distance in lexical chains and frequent use of repetitions. This however may rather be attributed to constraints of short term memory capacity in a spoken context rather than the intention to manipulate opinions as in ESSAY.

Last but not least, our findings show that identity is not the only and most important coreference relation to build textual coherence, at least for some registers. This all the more calls for an extensive exploration of such relations in future analyses.

References

M. Ariel. 2001. Accessibility theory: An overview. In T. Sanders, J. Schliperoord, and W. Spooren, editors, *Text representation: Linguistic and psycholinguistic aspects*, pages 29–87. John Benjamins Publishing, Amsterdam, Philadelphia.

R. Harald Baayen. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge University Press.

John Chambers, William Cleveland, Beat Kleiner, and Paul Tukey. 1983. *Graphical Methods for Data Analysis*. Wadsworth.

G. Doddington, A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel. 2004. Automatic content extraction (ace) program – task definitions and performance measures. In *Proceedings of LREC-2004: 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, May 24-30.

Michael J. Greenacre. 2007. *Correspondence analysis in practice*. Chapman & Hall/CRC, Boca Raton.

E. Hajičová, B. Hladká, and L. Kučová. 2006. An annotated corpus as a test bed for discourse structure analysis. In *Proceedings of the Workshop on Constraints in Discourse*, pages 82–89, Maynooth, Ireland. National University of Ireland.

Silvia Hansen-Schirra, Stella Neumann, and Erich Steiner. 2012. *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. de Gruyter, Berlin, New York.

A. A. Kibrik. 2011. *Reference in discourse*. Oxford University Press, Oxford.

Kerstin Kunz and Ekaterina Lapshinova-Koltunski. 2015. Cross-linguistic analysis of discourse variation across registers. *Special Issue of Nordic Journal of English Studies*, 14(1):258–288.

K. Lambrecht. 1994. *Information structure and sentence form. Topic, focus and the mental representation of discourse referents*. Cambridge University Press, Cambridge.

Ekaterina Lapshinova-Koltunski and Kerstin Kunz. 2014a. Annotating cohesion for multillingual analysis. In *Proceedings of the 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Reykjavik, Iceland, May. LREC.

Ekaterina Lapshinova-Koltunski and Kerstin Kunz. 2014b. Annotating cohesion for multillingual analysis. In *Proceedings of the 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Reykjavik, Iceland, May. LREC.

Ekaterina Lapshinova-Koltunski, Kerstin Kunz, and Marilisa Amoia. 2012. Compiling a multilingual spoken corpus. In Tommaso Raso Heliana Mello, Massimo Pettorino, editor, *Proceedings of the VIIth GSCP International Conference: Speech and corpora*, pages 79–84, Firenze. Firenze University Press.

S. Pradhan, L. Ramshaw, M Mitchell, M. Palmer, R. Weischedel, and N. Xue. 2011. Conll-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*.

M. Strube and U. Hahn. 1999. Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

BBN Technologies. 2006. *Coreference Guidelines for English OntoNotes – Version 6.0*. Linguistic Data Consortium. BBN Pronoun Coreference and Entity Type Corpus.

H. Telljohann, E. Hinrichs, S. Kbler, H Zinsmeister, and K. Beck. 2012. Stylebook for the tübingen treebank of written german (tüba-d/z). Technical report, Tübingen, Germany. Revised Version.

William N. Venables and David M. Smith. 2010. *An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics*.

Š. Zikánová, E. Hajičová, B. Hladká, Jínová P., J. Mírovský, A. Nedoluzhko, L. Poláková, K. Rysová, M. Rysová, and J. Václ. 2015. *Discourse and Coherence. From the Sentence Structure to Relations in Text*. ÚFAL, Prague, Czech Republic.