

English vs. German from a textual perspective:

Looking inside chain intersection

Kerstin Kunz, Ekaterina Lapshinova-Koltunski

University of Heidelberg, University of Saarland (Germany)

Abstract: This paper presents a cross-lingual corpus-based study on the intersection of chains of coreference and lexical cohesion. The two types of cohesion are often combined and thus play an important role for the development of discourse topics. We analyse chain intersection as cases where chain elements of lexical cohesion occur inside of coreference chains. We use a corpus of English and German original texts from four written and spoken registers which is annotated for both types of cohesion. Our analyses point to contrasts between the two languages and across the four registers under analysis in the types and the number of intersections in coreference chains. This variation has an effect on the way important topics develop in a text.

Keywords: coreference, lexical cohesion, discourse topics, register analysis, English, German

1. Motivation and state of the art

This paper presents a corpus-based analysis of particular types of interaction between chains of coreference and lexical cohesion that we call **chain intersection**. Our main focus is on the comparison of English and German and variation in written and spoken registers in these two languages. We argue that different types of chain intersection and the number of chain intersections reflect continuity and development of important discourse topics. They impact on how topics are perceived by text recipients as such.

There is general agreement in the literature that cohesion is an important linguistic device to explicitly establish coherence and continuity in texts. The interaction of coreference and lexical chains is regarded as essential to a text's cohesive harmony. The notion of chain interaction discussed in existing studies (Hasan, 1984; Hoey, 1991; Martin, 2015; Song *et al.*, 2015), however, differs considerably from our concept of chain intersection, as will be seen in Section 2 below.

For instance, chain interaction in Hasan's (1984, 1985) model of cohesive harmony applies when elements of different chains are realized as different constituents of the same clause – in the theme or the rheme – and as different participants involved in the same process (e.g. actor and beneficiary). Clauses or sentences in the same text containing such

reoccurring chain interactions of the same coreference and/or of lexical chains are considered to form key sentences. They are used in different approaches to measure local coherence, e.g. Grosz *et al.* (1995) in Centering Theory, or Strube and Hahn (1999) and Hoey (1991). Our concept does not consider intra-clausal interaction between different chain elements. It accounts for elements in coreference chains in which grammatical and lexical devices of cohesion are combined and where the lexical device is at the same time integrated in an element of a lexical chain. In most existing corpus-based studies, the two types of cohesion are either studied separately, are not distinguished or do not deal with discourse topics as an aspect of language contrast and register variation.

Most computational studies, such as Doddington *et al.* (2004) and more recently CoNNL 2011 (Pradhan *et al.*, 2011), are monolingual. They focus on automatic anaphora resolution and draw data from large corpus resources such as OntoNotes (Technologies, 2006). They contain information on coreference relations and bridging but do not analyse chain interaction. Computational models developed by Morris and Hirst (1991), Barzilay and Elhadad (1999) and other works building on them apply chain interaction to extract key sentences for automatic text summarization on monolingual corpora containing individual registers. Their notion of chain interaction is based on the study by Hoey (1991) and, again, differs from the concept proposed in the study presented here (see Section 2). Besides, these studies are restricted to lexical cohesion. There are also computational models which identify key sentences for automatic assessment of local coherence and are concerned with the interaction of coreference chains. They are largely based on Centering Theory (see Grosz *et al.* 1995). Several computational works analyse the interaction of coreference and lexical chains or bridging, such as Mesgar and Strube (2015). These, however, do not focus on the fine-grained analysis of language contrast and register variation as they develop algorithms for automatic text analysis.

There are few multilingual studies that base their analysis on corpora that are manually annotated with coreference and bridging (e.g. Zikánová *et al.*, 2015; Lapshinova-Koltunski and Kunz, 2014). They do not consider interaction between the two types of chains. The multilingual corpus-based studies by Kerremans (2014) involve fine-grained manual annotations and use lexical patterns in coreference chains to analyse terminological variation and equivalence in originals and translations to build terminological databases. This model differs from ours in that they do not analyse whether the lexical patterns in coreference chains are also part of lexical chains. So, a corpus-based account applying fine-grained annotations in order to see how languages and registers differ in terms of chain intersection in the sense explained below does not exist so far. Moreover, we argue that our approach offers new insights into the interplay of coreference chains and lexical cohesion and how this intersection contributes to the linguistic reflection of discourse topics.

The remainder of this paper is organised as follows: we start with a clarification of the main concepts used: Discourse topics, coreference and lexical cohesion as indicators of discourse topics, and chain intersection. We discuss the linguistic indicators of different aspects of chain intersection, allowing us to interpret the latter as an indicator of topic continuity and development. In Section 3, we describe our methods and resources. We use a corpus of English and German comparable texts from four registers (political essays, fictional texts, popular scientific texts and spoken interviews). The corpus is annotated for both lexical cohesion (Martínez Martínez *et al.*, 2016) and coreference chains (Lapshinova-Koltunski and Kunz, 2014) and allows filtering out the chains that represent cases of intersection. We report on the results of the above research questions in detail in Section 4 and interpret them in terms of contrasts in thematic progression between English and German, also encompassing register variation in Section 4.10.

2. Chain intersection

As mentioned above, the focus of this paper is on the interplay between two types of cohesive chains: coreference and lexical chains. We therefore begin with a brief definition of the two types before we discuss our concept of chain intersection.

2.1 Discourse topics

It has previously been established that cohesive chains often interact in texts and that this interplay is an important factor influencing how textual coherence and the development of discourse topics are perceived by text recipients, see e.g. Tanskanen (2006) and Todd (2016). The term *discourse topic* still remains somewhat undefined and fuzzy as it has been used in the literature from a variety of different perspectives, sociological and pragmatic or cognitive and semantically oriented ones. It is understood here in the sense of Chafe (1976) and Brown and Yule (1983) from a textual rather than a grammatical perspective. Intra-clausal relations between ‘theme’ and ‘rheme’, ‘topic’ or ‘comment’ or ‘topic’ and ‘background’ are therefore not considered in this paper, although they interact with cohesive chains. Discourse topics are topics that unfold throughout the text.

For our purpose, Todd’s approach seems most fitting, in which topics are defined as “clustering of concepts which are associated or related from the perspective of the interlocutors in such a way as to create connectedness and relevance” (Todd, 2003: 2009). This ideational clustering of extralinguistic concepts may stay rather implicit and may depend heavily on the text recipient’s inference of knowledge about the world and the context of situation. However, conceptual associations are indicated, at least to some extent, by linguistic patterns in the text. One essential mechanism to explicitly express connectedness in semantic space (see also Hoey, 1991) are cohesive chains.

A discourse topic may be global and concern the whole text or it may be rather local and be important to a smaller part of the text. But even on a more local level it often extends beyond clause boundaries. Coreference and lexical chains are employed as textual means to indicate both local and more global relations, depending on the number of elements and the distance between elements in a chain (see Kunz *et al.* 2016).

This paper is concerned with how these two types of cohesion interact. From a textual semantic point of view, we are interested in how and when coreference chains are integrated into chains of lexical cohesion. From a more conceptual point of view we investigate, how and when concepts about central individual referents (explicitly indicated by coreference) are integrated into clusterings of associated concepts, which are explicitly indicated by lexical cohesion. Let us therefore first take a look at both types of cohesive chains in turn and see how they contribute to the creation of discourse topics and then discuss how one type can intersect with the other.

2.2 Coreference and lexical cohesion

Most existing models refer to coreference and lexical cohesion as two separate relations. This distinction is grounded in formal criteria – grammatical vs. lexical devices of cohesion – and also in conceptual differences in the meaning relations established, e.g. identity vs. similarity (Hasan, 1985), or coreference vs. bridging (Clark, 1975; Poesio *et al.*, 1997). Our notions of coreference and lexical chains combine these two aspects, as illustrated in example (1).¹

¹ Note that in all our examples, the extension of elements in coreference chains is marked by brackets; elements of lexical chains are underlined.

- (1) This past spring, the U.S. Department of Education issued < a report, The Condition of Education 2000>. Some of the trends < it> pinpointed offer evidence that < The report> found that the benefits of attending college are greater today than ever before. With significant increases in the number of students who may not speak English at home, < this report> suggests that ...

In **coreference chains**, **grammatical** devices signal a textual relation to other coreferring expressions in the same text. The conceptual association evoked in this way is **identity** between conceptual referents. The first element in a coreference chain, the **antecedent** introduces a new extralinguistic referent into the textual world. Linguistic forms of antecedents can be manifold: they quite often contain an indefinite article and, most essential to this paper, a lexical nominal head, as in example (1) above. The subsequent elements of the coreference chain, the **anaphors**, contain grammatical devices, signalling that the same extralinguistic referent is mentioned again (and again). These devices may either serve as a modifier of a coreferring noun phrase, like *the* and *this* in example (1), or they may function as a pronominal head, such as *it* in example (1). It is widely accepted in the literature that different anaphoric forms indicate different degrees of accessibility, or givenness (see e.g. Ariel 2001, Prince 1981, Gundel *et al.*, 2003) but variation in anaphoric forms may also reflect pragmatic choice subject to register (see e.g. Kunz 2010). These aspects are however not the focus of the present paper.

What is important here is, first, that coreference chains reflect linguistically that individual conceptual referents play a central role in the textual world. Most often they are not the only participants of the textual world but they contribute to the development of discourse topics. Second, these central referents are related to other concepts in the textual world. All elements in coreference chains that contain a lexical head, be it in the antecedent or in the anaphor, have a potential to intersect with lexical chains and thus to reflect a relation to other concepts in the discourse (see below).

In chains of **lexical cohesion**, the relation between lexical devices of the chain elements is relevant. Our study includes relations between nominal expressions, which may consist of multiple words. Adjacent elements in lexical chains are connected by repetition, as in example (1), or sense relations such as hyperonymy, synonymy and meronymy, antonymy, and relations between named entities (see Martínez Martínez *et al.* 2016 for more details about the sense relations analysed). While grammatical devices in coreference chains are employed to signal identity between individual instantiated referents, lexical devices signal conceptual **similarity** between types of referents. As will be explained in more detail below, the two types of chains may intersect under certain conditions. In any case, lexical chains are an explicit means to create semantic space (see also Hoey 1991) in a text. They indicate linguistically how concepts in the textual world are clustered, evoking associations between types of referents. They are an essential linguistic mechanism to reflect discourse topics. So for an operationalisation of our approach, we regard lexical chains as **explicit discourse topics**. Our aim is to see if, when and to which degree central individual referents evoked by coreference chains contribute to these explicit discourse topics.

A cohesive chain minimally consists of a tie between an antecedent (first element in a chain, see above) and an anaphor (subsequent element(s)). As can be seen in example (1), many chains consist of more than two elements and contain several anaphors. Typically texts contain both types of chains, although to varying degrees. Kunz *et al.* (2016) analysed the variation in cohesive chains with respect to three chain features, the number of elements in chains, the distance between members and the number of different chains, as well as the

interaction of these chain features. They further discussed how this variation impacts on the way discourse topics are structured linguistically in texts: whether the organization of cohesive chains reflect more topic continuity or more variation, whether there is an abrupt change, a continuous modification or a constant interaction of topics. However, chains of coreference and lexical cohesion were analysed separately.

2.3 Chain intersection

In this paper, we focus on the combination of coreference and lexical cohesion in cohesive elements which belong to both chain types. As already stated in Section 1, our concept of chain intersection differs from other approaches in that we do not investigate different chains linked by elements of chains that are realized as different syntactic constituents. Our concept does not consider intra-clausal interaction between different chain elements.

Generally speaking, chain intersection takes place whenever a lexical item that is part of an element in a lexical chain also occurs inside an element of a coreference chain. The two chains ‘meet’ at the point of the intersection. From the perspective of the coreference chain, our approach accounts for elements in coreference chains in which grammatical and lexical devices are combined and where the lexical device is at the same time integrated in an element of a lexical chain. This is illustrated in example (2).

- (2) 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...

In example (2), we have a coreference chain (marked with brackets) and a lexical chain (marked with underlining) that intersect. Both chains consist of three chain elements. In this case, the intersection starts in the antecedent of the coreference chain, with the lexical head, a compound noun. The nominal expression forms the antecedent of a lexical chain. The conceptual relation of identity in the coreference chain is indicated in the two anaphoric chain elements by two grammatical devices (the demonstrative pronoun *this* and the definite article *the*) to the antecedent *reward system*. These devices do not serve as heads but function as modifiers of the whole chain element. So in this coreference chain not only the antecedent but also the anaphors contain a lexical nominal head.

These lexical heads establish a relation of lexical cohesion. In example (2), the noun *system* is a hyperonym of the antecedent, and the noun in the nominal phrase *the system* is a repetition of the preceding noun. Other possible sense relations are synonymy and hyponymy. So a lexical chain and a coreference chain meet or intersect because the lexical items in their chain members overlap. In the case of example (2), the intersection takes place already in the coreferential antecedent and the lexical antecedent and goes on in the coreferential and lexical anaphors. In this way strong bonds are established inside an explicit discourse topic. There is an explicit linguistic signal indicating not only that an individual referent plays an important role in the textual world but also that this referent is central to an explicit discourse topic. We will see below that there are different types of chain intersection depending on where in the coreference chain and where in the lexical chain the intersection takes place. Moreover, variation in chain intersection concerns the number of intersections. In this study

we are interested in how the two languages English and German as well as the four registers differ with respect to a number of variations discussed in more detail in the next section.

2.4 Features and types of chain intersection

In this section, we define the nine features of chain intersection we analyse in our study on a more refined level. The explanations provided here serve as background information for the overview of the operationalisations given in 3.1. Note that we cannot discuss all the features that may be important in the frame of this paper. We will address them shortly in our outlook and hope to deal with them in the future. We start here by a definition of ‘shallow’ features which serve as general indicators of the degree of explicit marking of discourse topics by chain intersection and explain how they impact on the continuity of explicit discourse topics in general. We then define features on the basis of which different types of chain intersection are distinguished.

2.4.1 General features of chain intersection

Generally, the higher number of intersections between coreference and lexical chains is measured per text, the more central referents contribute to the explicit discourse topics indicated by lexical cohesion in this text. So, the first feature we are interested in is whether there are differences between English and German and between the four registers in our corpus in terms of **the overall number of chain intersections** (feature 1 in 3.1). This feature is obtained by counting the total number of tokens (i.e. nouns and nominal phrases) that are included in intersections of coreference and lexical chains. It is important to note here that an element in a coreference or in a lexical chain may contain several lexical nominal items, e.g. in the case of compounding. They are counted as separate overlaps with this measure. We therefore add the two other features below, in order to obtain the number of intersections per chain elements.

Most of the coreferential antecedents in our corpus contain lexical nouns and therefore have a potential for overlapping chains. This does not always apply to the subsequent elements in coreference chains, the coreferential anaphors, which may consist of pronouns, as can be seen in example (3) below. **The number of coreference anaphors with a lexical head** informs about this general potential of intersection for coreferential anaphors (feature 2 in 3.1).

The next two features are an elaboration of the first one defined above, distinguishing the perspective taken: The length of a chain element may differ in both chain types. For instance, the coreferential antecedent *the brain's reward system* in example (2), which is one single element in the coreference chain, contains three different nouns. Each of these nouns may, however, overlap with one single element of three different lexical chains, or the nouns taken together may overlap with one element of a lexical chain. A coreferential antecedent may even be more complex. We therefore include two additional features, which account for the perspective taken, the extension of the coreferential element or of the lexical element (see below).

Feature 3, **the number of intersections per coreferring element**, counts how many coreferring elements contain intersections with elements in lexical chains (and not whether the same coreference chain is affected by chain intersections again and again).

Furthermore, feature 4, **the number of intersections per lexical chain element**, informs about how many lexical chain elements intersect with coreferring elements.

Feature 5 serves to see how often an intersection takes place in one coreference chain, counting **the number of coreference chains with only one intersection**. One intersection means that only one element (i.e. an antecedent or also one, two or more anaphors) in the

whole coreference chain is responsible for the intersection. The first element in a coreference chain is always involved whenever there is a chain intersection. This implies that the coreferential antecedent is solely responsible if there is one intersection only.

In order to account for differences in the number of intersections per chain and thus between types of intersection, an additional feature is relevant: **the position of the intersection in a coreference chain** (feature 6). We here explore whether the intersection takes place only in the antecedent (first) or additionally also in another element – an anaphor of the coreference chain (non-first).

2.4.2 Features related to specific types of chain intersection

Example (2) above and examples (3) and (4) below illustrate three different types of chain intersection, which can be distinguished on the basis of several chain features. Examples (2) and (3) share one characteristic feature distinguishing them from example (4): Chain intersection is observed for the antecedents – the first chain elements – in both chains. This implies that the newly introduced individual referent in the textual world is at the same time used to introduce an explicit discourse topic. Hence, an important feature is **the number of antecedents of coreference chains that intersect with antecedents of a lexical chain**, operationalised with feature 7.

- (3) Well, in Edinburgh most of the water comes from <reservoirs> which are more towards the a lot of <them> are more towards the Borders, and then <they>'re actually quite old, I think. They first tried to sort out water in Edinburgh, ... But now, a lot of the water comes from the hills on the outskirts of Edinburgh. And then it's brought into holding reservoirs, and then it's brought into the treatment works... And it's the company looks at different ways of helping water companies manage those assets better... if you've got a set of pipes and you've got a set of, say, service reservoirs, which is where you store the clean water, ...

Example (4) shows a type of chain intersection which exhibits more apparent differences from (2) and (3). It demonstrates the importance of one more feature: **the number of antecedent(s) in coreference chain(s) that are anaphors in a lexical chain** (feature 8 in Section 3.1). The example is taken from our English subcorpus of fictional texts.

- (4) This evening I find myself settled here in this comfortable guest house in a street not far from the centre of Salisbury.... <The landlady, a woman> of around forty or so, appeals to regard me as a rather grand visitor ... <She> informed me that <a double room> at the front was available, though I was welcome to <it> for the price of a single. I was then brought up to <this room>, in which ... On inquiring where the bathroom was, <the woman> told me ... I asked <her> to bring me up a pot of tea, and when <she> had gone, inspected <the room> further...

In example (4), we have one lexical chain starting with *guest house* and two coreference chains starting with *The landlady ...* and *a double room*. This type of intersection differs from the ones described above in that the antecedents of the two coreference chains are anaphors in the lexical chain. The two referents pointed at by the two coreference chains play a central role to the explicit discourse topic but they are not used to introduce it. It is introduced by the general concept *the guest house*, which is not itself mentioned again with coreferring expressions later in the text. Instead, other referents, which are mentioned for the first time in the text afterwards and which are related to *guest house* by meronymy are taken up again. In

this way the topic is broadly introduced and configurations of specific referents that are involved in it are specified afterwards.

There is a difference between examples (2) and (3) concerning **the number of anaphors in a coreference chain that are also anaphors in a lexical chain**, which is operationalised with feature 9. In example (2), an individual referent introduces an explicit discourse topic – the first intersection of the two chain types takes place in the antecedents. Moreover the anaphors of the coreference chain, apart from the grammatical devices signalling identity, contain a lexical head and therefore overlap with anaphors of the same lexical chain again. This type of intersection is comparable to Halliday and Hasan (1976, 277ff)’s notion of **reiteration**. It accounts for chains of lexical cohesion where the anaphors, not the nominal lexical head, are combined with (i.e. preceded by) a grammatical item – the definite article or a demonstrative determiner – that indicates coreference. Thus, the individual referent is conceptually enriched and contributes to a very great extent to the explicit discourse topic because there is an intersection between a coreference and a lexical chain not only with respect to the antecedents but also with respect to the anaphors.

In example (3) above, we have a coreference chain with the elements *reservoirs* – *them* – *they* and a lexical chain with the elements *reservoirs* – *holding reservoirs* – *service reservoirs*. As in example (2), the antecedents of both chains overlap: The referring expression *reservoirs* in example (3) serves as an antecedent for a coreference chain and a lexical chain. What is different is that the intersection holds for the antecedents only (i.e. the first mentions) whereas the rest of the coreference and the lexical chains do not overlap: The coreferential anaphors are made up of coreferential pronouns functioning as nominal heads. The lexical anaphors, which follow the coreferential anaphors are nominal expressions in the plural. So the coreferential anaphors do not contain a lexical element and the lexical anaphors do not contain a grammatical coreferential element. The semantic relation between *reservoirs* – *holding reservoirs* – *service reservoirs* is not that of identity, it is a relation of hyperonymy/hyponymy between different conceptual referents. This type of chain intersection is a typical mechanism to establish a smooth transition from one central referent to other concepts within one explicit discourse topic.

To sum up, we aim to compare the two languages and the four registers with respect to the overall number of chain intersections as well as to the types of chain intersection which depend on where, when and how often a lexical chain ‘meets’ a coreference chain. As illustrated by the above examples, these variations have an effect on the development of explicit discourse topics. They reflect variation as to the general degree of importance central referents have for explicit discourse topics, at which point and how often they contribute to the discourse topic. With a final comparison of all features we want to explore whether contrasts are greater between languages or between registers, and also compare the registers language-internally to find out if the breadth of variation is greater in English or in German.

3. Analysis design

In the following sections, we present the set of operationalisations formulated on the basis of the features presented in Sections 2.4.1 and 2.4.2 that we use in our analysis. Apart from that, we describe the corpus resources at hand.

3.1 Operationalisations of chain intersection

For the sake of convenience, we here provide a concise summary of the nine features, which were already introduced in 2.4.1 and 2.4.2, along with the operationalisations used for the

corpus linguistic analysis. This structure serves as a basis for the analyses presented in Section 4 below.

1. Overall number of chain intersections (`nr.inters`): obtained by computing the number of tokens involved in the intersection between coreference chain elements and lexical chain elements, i.e. the total number of overlapping tokens in coreference and lexical chains.
2. Number of coreference anaphors that have a lexical head calculated as the proportion of all anaphors of coreference chains whose head is a lexical item – full nominal phrases (`nr.corefana.lexhead`).
3. Number of intersections per coreferring element (`nr.intersec.percor`): obtained by computing the number of coreferring elements that also include elements of lexical cohesion chains.
4. Number of intersections per lexical chain element (`nr.inters.perlexcoh`): obtained by computing the number of lexical chain elements that intersect with coreferring elements.
5. Number of coreference chains with only one intersection (`nr.corefchain.one.inters.percor`).
6. Average position of intersection (`nr.intersec1st` and `nr.intersec.non1st`): In our study we distinguish between the number of intersections which take place in the first element/ position of a coreference chain (coreferential antecedents) and the number of intersections in a position different from the first position of a coreference chain (coreferential anaphors).
7. Overlapping antecedents (`nr.intersec.ante.ante`): measuring the number of antecedents in coreference chains that are also antecedents (first elements) of lexical chains.
8. Number of antecedents in coreference chains that are anaphors (not the first member) in lexical chains (`nr.intersec.ante.ana`).
9. Overlapping anaphors (`nr.intersec.ana.ana`): number of anaphors in coreference chain that are anaphors (not the first element) in a lexical chain.

In a final step, we analyse the overall variation in the two languages and the four registers comparing them with respect to all the nine features. This is done with the help of correspondence analysis, which was applied, for instance, by Kunz *et al.* (2017) in their study of cohesive features in English and German. The findings will be presented in Section 4.10.

3.2 Corpus design and annotation

The dataset we use for our analysis contains 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 registers included in our sub-subcorpus 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.

Table 1. Corpus description.

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

The whole corpus is annotated on various levels of lexicogrammar, e.g. parts-of-speech (POS), chunks, clauses, sentences. As mentioned above, the corpus contains manual annotation of various cohesive devices, including coreference (Lapshinova-Koltunski and Kunz 2014) and lexical cohesion (Martínez Martínez *et al.* 2016).² The annotation of coreferential devices includes possessive determiners and pronouns, personal pronouns, demonstrative determiners and pronouns as well as coreferential adverbs such as *here* and *there*, *now* and *then*, *hereby* and *therewith* (pronominal adverbs). Moreover, annotation of situational coreference (or complex anaphors), where the antecedent consists of a longer textual chunk than just a noun phrase (e.g. a clause, sentence or text paragraph), is also included.

The subset of the corpus presented here additionally provides relational information about lexical chains. Adjacent elements in lexical chains were annotated manually for the type of semantic relation holding between them (e.g. synonymy, antonymy, hyponymy, etc.). All manual corrections and annotations were performed with the tool MMAX2 (Müller and Strube, 2006).

3.3 Analysis techniques

We apply a descriptive data analysis with bar plots as visualisation techniques in Sections 4.1 to 4.9 to observe frequencies of the selected features and to derive general tendencies in their distribution across English and German texts in our data. We use them to relate their frequencies to the total number of chains per language and register to obtain an insight into their distributions. The results are tested for significance using the Pearson's chi-squared test with Yate's continuity correction,³ with the help of which we can prove if the observed differences between languages (English vs. German) and registers (ESSAY vs. FICTION, etc.) are significant. The Chi-square test measures how well the observed distribution of data fits with the distribution that is expected if the variables are independent.

In Section 4.10, we describe the results of correspondence analysis (CA, Nenadić and Greenacre, 2007) performed for all the features taken together. This technique is explorative and allows us to discover structures in the data in terms of groupings of observations Baayen (2008), for instance, groupings of subcorpora according to their similarities. Besides that, this technique helps to see possible correlation of dependent and independent variables. The

² 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>

³ A correction for the Chi-square test to use with small data sets.

correlation of our features with the corresponding subcorpora indicates the contribution of these features to the similarities between languages and registers. In CA, distances between dependent and independent variables are calculated and represented in a two-dimensional map, and the larger the differences between subcorpora or texts, the further apart they are on the map. Likewise, dissimilar categories of features are further apart. The correlations between dependent and independent variables are transformed into a set of uncorrelated variables, called principal axes or dimensions. The first two principal axes account for as much variation as possible in two dimensions. In the present paper, this technique will provide a better overview of the interdependence of the features as well as over the breadth of variation between registers and languages.

4. Analyses, results and discussion

In the present Section, we describe the results of our analysis. As already mentioned in Section 3 above, the section is structured according to the features under analysis that we operationalised in Section 3.1. We will restrict ourselves to a mere description of the findings for each separate feature.

4.1 Overall number of intersections

In Figure 1, we provide a bar plot representing the proportion of the overlapping tokens against the total number of tokens that are elements in coreference and lexical chains.

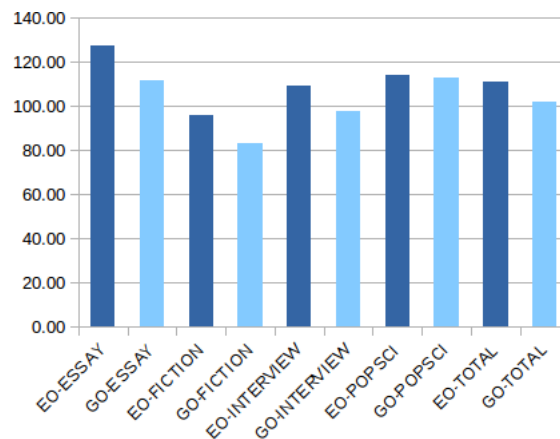


Figure 1. Intersecting tokens in coreference and lexical cohesion chains.

As seen from the plot, English texts reveal a higher number of total intersections when separate tokens (nouns) are counted. This applies to all registers except popular-scientific texts. The latter show similar tendencies in both English and German. The highest number is observed in English essays. Overall, the difference between English and German is significant, as we achieve a very low p-value ($p < 0.00001$, $\chi^2 = 57.369$, $df = 3$) with Pearson's Chi-squared test.

As noted in Section 2.4.1, an element in a chain may contain several lexical tokens that are nouns. These may intersect with elements of different chains. The higher number of intersections measured in English as compared to German may in part be explained by the fact that all nouns separated by a white space (which is more often the case in English) are

counted as separated tokens whereas compounds without a white space (more common in German) count as one word. More frequent compounding would explain why popular scientific texts and political essays contain a higher number of intersections than the fictional texts and the interviews.⁴ The findings for political essays seem to point to the frequent repetition of compounds. This serves as a precision of information (as in scientific texts) and reflects the ideational function of persuasion.

4.2 Number of coreference anaphors with a lexical head

The proportion of all anaphors of coreference chains whose head is a lexical item measured against the total number of coreferring expressions is presented in Figure 2.

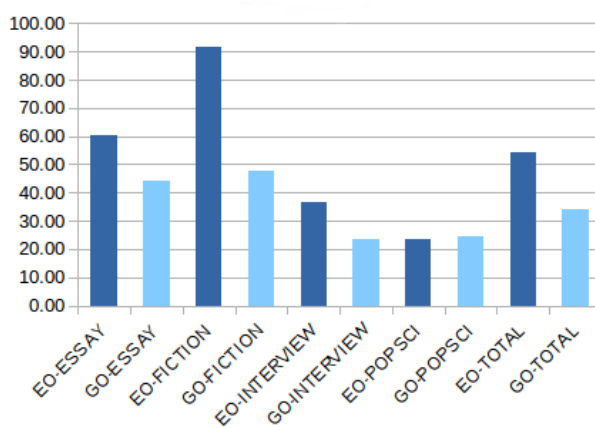


Figure 2. Coreference anaphors with lexical heads.

This number is much higher for all English texts if compared to the German ones, implying that the potential of a coreferential anaphor to take part in an intersection with a lexical chain is generally higher in English than German. In other studies, we could observe two language-specific factors, in addition to the ones mentioned above, which seem to be at play here: a higher number of coreferential anaphors are made up of pronouns and there is a higher amount of extended co-reference with non-nominal expressions (clauses, sentences or textual paragraphs) in German than in English. As for register variation, the popular-scientific articles show very similar, yet low proportions. These texts primarily have an informative communicative purpose, and high information density is expressed by high lexical density. The texts of this register contain many nouns (mostly terms) building chains of lexical cohesion. However, there are few coreference chains and many elements of the lexical chains do not intersect with coreference at all: They are often related by meronymy and repetition and indicate generic relations between types of referents (not instantiated ones) within a discourse topic. The English fictional texts show the highest number of coreferential anaphors with nouns. This is surprising, as fiction contains many coreferential pronouns (a feature of narrative style and spoken language). Looking into the texts reveals that the narrative parts of the English texts frequently contain descriptions of the settings in which the protagonists act, similar to example (4). The different components of the settings are mentioned again but alternate throughout stretches of text. They thus have to be resumed by a

⁴ The proportion of compound nouns in our corpus comprises 25% in political essays, 19% in popular-science, 16% in fictional texts and 11% in the transcribed interviews.

fully lexical phrase. In German fictional texts, there is less alternation and the focus is more on the main protagonists.

The results of Pearson's Chi-squared test confirms that the differences across registers between the two languages are significant ($p < 0.00001$, $\chi^2 = 88.771$, $df = 3$).

4.3 Number of intersections per coreferring element

The proportion of intersecting coreferring elements measured against the total number of (both lexical and coreference) chains is given in Figure 3.

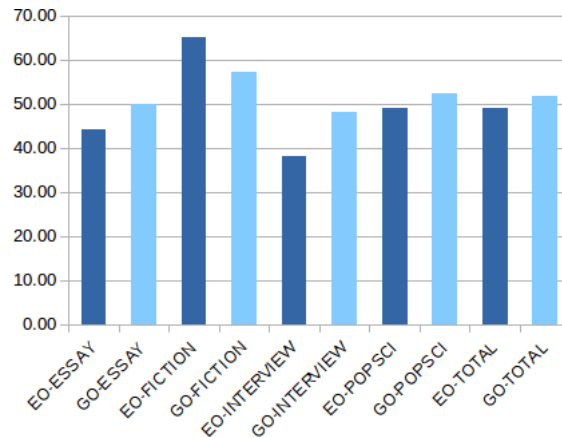


Figure 3. Intersections measured per coreferring element.

We observe an opposite tendency for this feature: German texts show a slightly higher number of intersections. However, an exception is provided by English fictional texts – here we have the highest number of intersections amongst all the texts analysed in both languages. The number in this register is higher than for all others across languages. This seemingly has to do with the exceptional length of the coreference chains as well as the high number of different coreference chains. One possible reason for the general differences to the above findings in terms of general language contrast and register variation has already been suggested above: Chain elements rather than tokens serve as a basis for the feature here. Hence, another reason for the high value for fiction seems to be the lower number of multiple nouns contained per coreferring element. So in German, more coreferring elements overlap with lexical chains elements than in English although the lexical potential in anaphors is lower. This may even strengthen the explicit effect, from a contrastive perspective. Individual referents thus seem to be connected more strongly to an explicit discourse topic in German than English, except for the fictional texts. The difference between the two languages is also significant in this case ($p < 0.00001$, $\chi^2 = 48.843$, $df = 3$).

4.4 Number of intersections per lexical chain element

Figure 4 illustrates the proportion of intersecting lexical chain elements calculated against the total number of all chains in the corpus. Since this feature is measured on the basis of lexical cohesion elements rather than coreference chain elements, it conveys a different perspective on chain intersection than that in 4.3 above.

In this case, we observe a similar tendency as in Section 4.3, if all German and English texts are considered: The German texts use more intersections than the English ones, except

for the fictional texts. This difference is also significant ($p < 0.00001$, $\chi^2 = 81.064$, $df = 3$). Thus, more elements of lexical chains are connected to a central referent in German than in English. However, the register-specific figures show that political essays and popular-scientific texts contribute the most to the language contrast. In the latter register, language contrast is more pronounced for this feature than for all other features. This time, the numbers for the fictional texts lie below those of the other registers in both languages.

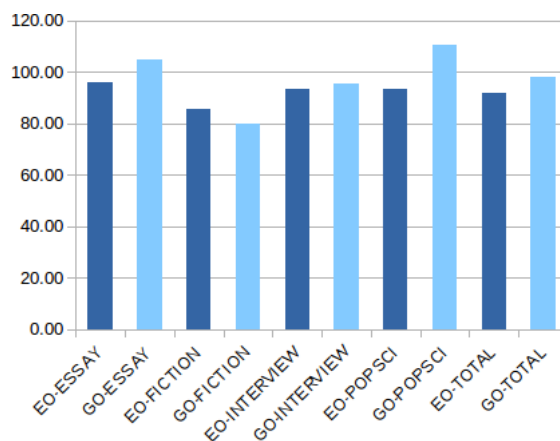


Figure 4. Intersections measured per lexical chain element.

4.5 Number of coreference chains with only one intersection

In Figure 5, we provide the proportion of coreference chains with one intersection measured only against the total number of intersections.

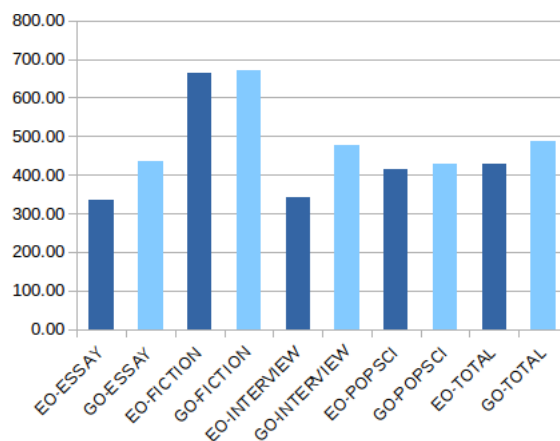


Figure 5. Chains with one intersection only

The findings show a higher number of chains with only one intersection in German compared to English. This means that there are more cases in German than English where only the coreferential antecedent but not the rest of the coreference chain intersects with a lexical chain, as shown in example (3). We observe similar tendencies for both languages in popular-scientific and fictional texts. The fictional texts outperform all other registers again. This may generally be connected to two facts: First, a low number of multiple nouns in noun phrases and second, a very high number of different chains of both chain types in both languages

(based on the findings by Kunz *et al.* 2016). In interviews and political essays, the number of coreference chains with only one intersection is higher in German than in English. Overall, languages turn out to differ significantly across registers ($p < 0.00001$, $\chi^2 = 46.478$, $df = 3$).

4.6 Average position

The average position of the intersections in a coreference chain is defined as a binary category: first and non-first. In Figure 6, we present the proportions of these two types of intersections calculated against the total number of intersections.

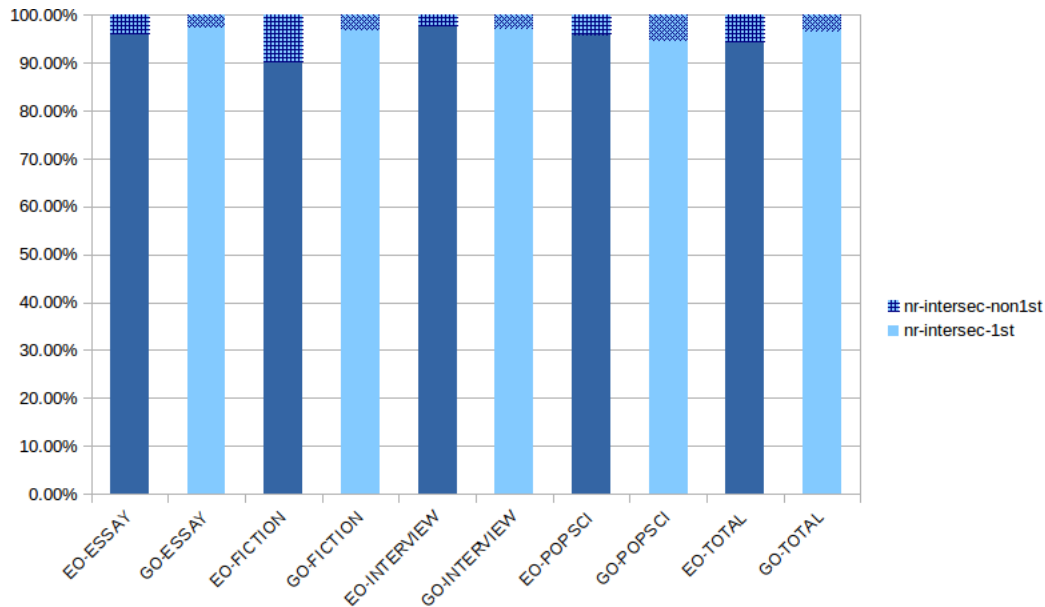


Figure 6. Average intersection position.

We generally note a much higher number of intersections in the first position (the first chain elements) than in all other positions of coreference chains. This is of course due to the fact that the coreferential antecedent is always involved whenever chain intersection takes place, no matter which type of chain intersection. More interestingly, the proportion of non-first chain element intersections in relation to other chain elements is higher in English than in German, i.e. more coreferential anaphors are involved in English than German, leading to an intersection type such as in example (2). The English fictional texts contain the highest numbers of all texts, whereas the English interviews show the lowest frequencies. The significance test shows that the difference between English and German texts is significant (with a p-value of ca. 0.002, which is considerably higher than the results for the other features, but still below 0.05).

4.7 Overlapping antecedents

Figure 7 presents the proportion of overlapping antecedents calculated against the total number of intersections in English and German registers. In German fictional texts, interviews and political essays, antecedents tend to overlap more often than in the corresponding registers in English, which means that explicit discourse topics are introduced more often with central referents here. In popular-scientific articles, whose discourse structure is more standardised than in the other registers, we observe a similar number of

overlapping antecedents in both languages. However, the overall difference between different registers in both languages is significant ($p < 0.0001$). Within each language, fictional texts reveal most frequent cases of an overlap, with the highest number again shown for the English fictional texts and the lowest for English Interviews.

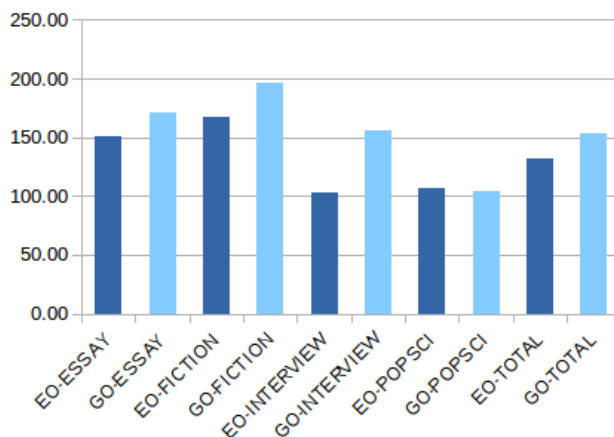


Figure 7. Overlapping antecedents in coreference and lexical chains.

4.8 Number of antecedents of coreference chains that are anaphors of lexical chains

Figure 8 displays this proportion which is also measured against the total number of intersecting elements. What has to be noted first is that the numbers for this feature are generally higher than those for overlapping antecedents for all registers in both languages. The degree of the difference is register-specific. For instance, it is less pronounced in the political essays.

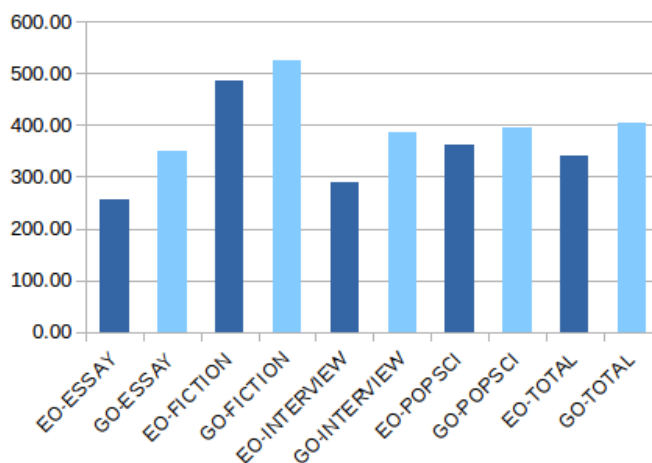


Figure 8. Intersections between coreference antecedents and lexical chain anaphors.

Second, in all German texts, there are more coreferential antecedents that intersect with anaphors (rather than with antecedents in lexical chains) than in English. Therefore, explicit discourse topics are introduced more often with a lexical antecedent in an earlier stretch of text, preceding the whole coreference chain, as exemplified in (4). This lexical chain element

is related to the coreferential antecedent by similarity of sense. It implies that central referents less often serve to introduce explicit discourse topics and more often play a role as topics unfold. Overall, we observe a significant difference between the languages ($p < 0.00001$, $\chi^2 = 31.74$, $df = 3$).

4.9 Overlapping anaphors

The proportion of all overlapping anaphors in coreference and lexical chains presented in Figure 9 is measured against the total number of chain intersections.

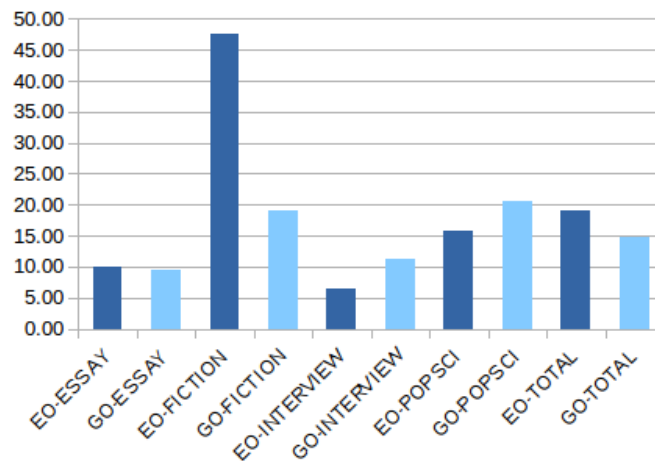


Figure 9. Overlapping anaphors in coreference and lexical chains.

Generally speaking, the numbers of overlapping anaphors is very low in both languages, when compared to the numbers for overlapping antecedents and coreferential antecedents that intersect with anaphors in lexical chains. Hence the intersection type as shown in example is not very frequent. The feature also indicates significant differences between the two languages ($p < 0.001$, $\chi^2 = 16.72$, $df = 3$) if analysed across all registers. But the tendency of German numbers lying below those of English holds only for two registers, ESSAY and FICTION. In addition, we observe different rankings of registers within the languages: while in English, fictional texts show the highest amount of overlapping anaphors, popular-scientific articles occupy the first position in German. These texts have more central referents with an important role for the explicit discourse topic that are also conceptually enriched throughout the text.

4.10 Overall variation

Figure 10 illustrates the output of the correspondence analysis. As seen from the two-dimensional plot (which explains 89.1% of variation in our data), we observe heterogeneous tendencies across languages and registers. English and German popular-scientific texts seem to be very similar, as they are situated very close to each other on the x-axis and even overlap on the y-axis. This coincides with the tendencies we observed for these texts analysing individual features. Interestingly, the x-axis separates fictional texts in both languages from the other registers, which again concurs with the results observed for individual features – fictional texts in both languages often behave differently from the other texts. However, they do reveal some language-specific features found along the y-axis. Correspondence analysis does not show a clear distance between languages, i.e. a consistent language contrast: It is

extension in coreference chains is longer than those of lexical chain elements. Both measures taken together show the same tendency, namely that central referents seem to be more relevant to explicit discourse topics in German than in English texts. The next two features reveal that the higher number of intersections in German mainly stems from intersections in which coreferential antecedents are involved: The higher number of coreference chains with only one intersection in German along with the higher number of intersections in the first position as compared to English means that more topics in German are introduced by a central referent directly. This observation finds further support by the three remaining features, which are more closely linked to specific types of intersection and thus to variation in the development of explicit topics by central referents.

5.2 Types of chain intersection

In both languages, the number of coreferential antecedents that intersect with anaphors in lexical chains is higher than the number of overlapping antecedents. In addition, overlapping antecedents play a more important role for intersection than overlapping anaphors. This entails a general ranking with respect to the types of intersection: More often a smooth transition is preferred, in which the explicit discourse topic is not introduced by a central referent directly. The antecedent of a lexical chain introduces a configuration of concepts. The central referent established by coreference, which constitutes one important concept within this configuration, is mentioned later. Explicit discourse topics that are introduced by a central referent directly are less frequent. In any case, further continuity in explicit discourse topics is not upheld to a great extent by central referents, as there are few anaphors of coreference chains overlapping with anaphors in lexical chains. Hence, most explicit discourse topics are reflected by lexical relations without conceptual identity being involved.

As for language contrast, we observe more overlapping antecedents in German than in English, hence more discourse topics are introduced by important individual referents directly. The same tendency however applies to the number of antecedents of coreference chains that are anaphors in lexical chains. Again we find more intersections in German than English, which possibly results from the fact that German has more intersections than English in general. In these constellations, the discourse topic is introduced by a nominal expression that is a holonym or a meronym of the following central referent in most cases. Contrary to the other two features, the number of overlapping anaphors is higher in English than in German.

Furthermore, our findings show that there is at least one register which is in contrast to the overall tendency observed in terms of language contrast, for most of the features. General features and particular types of intersection heavily depend on the register. They may thus relate to specific configurations of field, tenor and mode. Quite interestingly, the tendencies for registers in the two languages sometimes coincide but sometimes they do not. The fictional texts quite often stand out and seem to bear least resemblance to the other registers. Hence chain intersection and its impact on discourse topics seem to be a reflection of the distinction between fiction and non-fiction. Within the fictional texts, we note a dramatic difference with respect to the number of overlapping antecedents, the numbers for the English texts being much higher than those for German. This difference contributes the most to the general language contrast observed for the feature. A reason for this could be that settings and interaction between objects play a greater role in English, and main protagonists are favoured in German. The German fictional texts also contain more dialogic parts than the English ones. However EO FICTION is in even sharper contrast to EO INTERVIEW, a register within the same language, which contains a very low number of overlapping anaphors. EO INTERVIEW is also the register with the lowest number of overlapping antecedents and a

relatively low number of intersections between coreference antecedents and lexical chain anaphors, so there is not much chain intersection in general. This may be caused by the mode of spoken language in that there is a more frequent use of grammatical anaphors of coreference and a frequent occurrence of extended reference. It may also stem from colloquial style. The differences across languages with this spoken register are greater than for the written registers ESSAY and POPSCI, the latter being more standardised than all other registers.

6. Outlook

In this study we could not integrate all features of coreference and lexical chains that are relevant to the development of discourse topics. These deserve further exploration in the future. For instance, we only differentiate intersections that are contained in coreferential antecedents vs. coreferential anaphors, but we do not specify further which position the anaphor has in the coreference chain and the lexical chain (the second, the third or another element). This would, however, be interesting for longer chains and inform about ‘interrupted’ intersection, which may have an interpersonal function in argumentative and persuasive texts (e.g. introduction and synopsis). Moreover, a more precise interpretation of the role of intersections to explicit discourse topics can be obtained if the features introduced in this study are related to the other chain features (as it was discussed in Kunz *et al.* 2016): chain length, distance in chains and number of different chains. The features of this study have to be brought together with the features of other models dealing with chain interaction and cohesive harmony, as mentioned in Section 1.

References

- Ariel, M. 2001. Accessibility theory: An overview. In *Text Representation: Linguistic and Psycholinguistic Aspects*, T. Sanders, J. Schliperoord and W. Spooren (eds), 29–87. Amsterdam/Philadelphia: John Benjamins.
- Baayen, R. H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Barzilay, R. and Elhadad, M. 1999. Text summarizations with lexical chains. In *Advances in Automatic Text Summarization*, I. Mani and M. Maybury (eds), 111–121. Cambridge, MA: MIT Press.
- Brown, G. and Yule, G. 1983. *Discourse Analysis*. [Cambridge Textbooks in Linguistics.] Cambridge: Cambridge University Press.
- Chafe, W.L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In *Subject and Topic*, C. N. Li (ed.), 25–55. New York: Academic Press.
- Clark, H. H. 1975. Bridging. In R. C. Schank & B. L. Nash-Webber (Eds.) *Proceedings of the 1975 Theoretical Issues in Natural Language Processing*, 169–174. New York: Association for Computing Machinery.
- Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S. and Weischedel, R. 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.
- Grosz, B., Weinstein, S. and Joshi, A. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2): 203–225.
- Gundel, J., Hegarty, M. and Borthen, K. 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information*, 12(3): 281–299.

- Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London/New York: Longman.
- Hansen-Schirra, S., Neumann, S. and Steiner, E. 2012. *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. Berlin/New York: de Gruyter.
- Hasan, R. 1984. Coherence and cohesive harmony. In *Understanding Reading Comprehension*, J. Flood (ed.), 181–219. Delaware: International Reading Association.
- Hasan, R. 1985. The texture of a text. In *Language, Context and Text: Aspects of Language in a Social-semiotic Perspective*, M.A.K. Halliday and R. Hasan (eds), 70–96. Oxford: Oxford University Press. Reprinted 1989.
- Hoey, M. 1991. Another perspective on coherence and cohesive harmony. In *Functional and Systemic Linguistics: Approaches and Uses*, E. Ventola (ed.), 385–414. Berlin and New York: Mouton de Gruyter.
- Kerremans, K. 2014. *Terminological Variation in Multilingual Europe. The Case of English Environmental Terminology Translated into Dutch and French*. PhD thesis, Vrije Universiteit Brussel.
- Kunz, K. 2010. *Variation in English and German Nominal Coreference: A Study of Political Essays*. PhD thesis, Saarland University, Frankfurt am Main.
- Kunz, K., Degaetano-Ortlieb, S., Lapshinova-Koltunski, E., Menzel, K. and Steiner, E. 2017. Gecco – an empirically-based comparison of English-German cohesion. In *Empirical Translation Studies: New Methodological and Theoretical Traditions*, volume 300 of *TILSM series*, G. De Sutter, M.-A. Lefer and I. Delaere (eds), 265–312. Berlin and New York: Mouton de Gruyter.
- Kunz, K., Lapshinova-Koltunski, E. and Martínez Martínez, J. M. 2016. Beyond identity coreference: Contrasting indicators of textual coherence in English and German. In *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes co-located with NAACL 2016*, San Diego, California. ACL.
- Lapshinova-Koltunski, E. and Kunz, K. 2014. Annotating cohesion for multilingual analysis. In *Proceedings of the 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Reykjavik, Iceland. LREC.
- Lapshinova-Koltunski, E., Kunz, K. and Amoia, M. 2012. Compiling a multilingual spoken corpus. In *Proceedings of the VIIth GSCP International Conference: Speech and Corpora*, H. Mello, M. Pettorino and T. Raso. (eds), 79–84, Firenze: Firenze University Press.
- Martin, J. 2015. Cohesion and texture. In Tannen, D., Hamilton, H. and Schiffrin, D. (eds), *The Handbook of Discourse Analysis*, Blackwell Handbooks in Linguistics, 61–81. Malden, Mass: Wiley.
- Martínez Martínez, J. M., Lapshinova-Koltunski, E. and Kunz, K. A. 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*, Bochum, Germany.
- Mesgar, M. and Strube, M. 2015. Graph-based coherence modeling for assessing readability. In *Proceedings of STARSEM 2015: The Fourth Joint Conference on Lexical and Computational Semantics*, 309–318, Denver.
- Morris, J. and Hirst, G. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1): 21–48.
- Müller, C. and Strube, M. 2006. Multi-level annotation of linguistic data with MMAX2. In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, S. Braun, K. Kohn and J. Mukherjee (eds), 197–214. Frankfurt a.M.: Peter Lang.
- Nenadić, O. and Greenacre, M. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: The *ca* package. *Journal of Statistical Software*, 20(3): 1–13.
- Poesio, M., Vieira, R. and Teufel, S. 1997. Resolving bridging references in unrestricted text. In *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, ANARESOLUTION '97, 1–6, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Pradhan, S., Ramshaw, L., Mitchell, M., Palmer, M., Weischedel, R. and Xue, N. 2011. CoNLL-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*.

- Prince, E. 1981. Toward a taxonomy of given-new information. In *Radical Pragmatics*, P. Cole (ed.), 223–255. New York: Academic Press.
- Song, W., Fu, R., Liu, L. and Liu, T. 2015. Discourse element identification in student essays based on global and local cohesion. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2255–2261, Lisbon, Portugal. Association for Computational Linguistics.
- Strube, M. and Hahn, U. 1999. Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3): 309–344.
- Tanskanen, S.-K. 2006. *Collaborating towards Coherence: Lexical Cohesion in English Discourse*. Amsterdam/Philadelphia: John Benjamins.
- Technologies, B. 2006. *Coreference Guidelines for English OntoNotes – Version 6.0*. Linguistic Data Consortium. BBN Pronoun Coreference and Entity Type Corpus.
- Todd, R. 2003. *Topics in Classroom Discourse*. PhD thesis, University of Liverpool.
- Todd, R. 2016. *Discourse Topics*. Pragmatics & Beyond new series. Amsterdam/Philadelphia: John Benjamins.
- Zikánová, Š., Hajičová, E., Hladká, B., P., J., Mírovský, J., Nedoluzhko, A., Poláková, L., Rysová, K., Rysová, M. and Václ, J. 2015. *Discourse and Coherence. From the Sentence Structure to Relations in Text*. ÚFAL, Prague, Czech Republic.

Authors' addresses

Kerstin Kunz
Institut für Übersetzen und Dolmetschen
Universität Heidelberg
Plöck 57a
69117 Heidelberg

kerstin.kunz@iued.uni-heidelberg.de

Ekaterina Lapshinova-Koltunski
Department of Language Science and Technology
Universität des Saarlandes
Building A2.2, Room 1.02
Campus, 66123 Saarbrücken

e.lapshinova@mx.uni-saarland.de