Cross-lingual Incongruences in the Annotation of Coreference

Ekaterina Lapshinova-Koltunski[1]   Sharid Loáiciga[2]
Christian Hardmeier[3]   Pauline Krielke[1]
[1]Department of Language Science and Technology, Saarland University, Germany
[2]CLASP, University of Gothenburg, Sweden
[3]Department of Linguistics and Philology, Uppsala University, Sweden
e.lapshinova@mx.uni-saarland.de   sharid.loaiciga@gu.se
christian.hardmeier@lingfil.uu.se
pauline.krielke@uni-saarland.de

Abstract

In the present paper, we deal with incongruences in English-German multilingual coreference annotation and present automated methods to discover them. More specifically, we automatically detect full coreference chains in parallel texts and analyse discrepancies in their annotations. In doing so, we wish to find out whether the discrepancies rather derive from language typological constraints, from the translation or the actual annotation process. The results of our study contribute to the referential analysis of similarities and differences across languages and support evaluation of cross-lingual coreference annotation. They are also useful for cross-lingual coreference resolution systems and contrastive linguistic studies.

1 Introduction

Linguistically annotated parallel corpora in multiple languages are valuable resources for language technology, linguistic research and translation studies. To be maximally useful for such applications, the annotation should accurately reflect linguistically relevant contrasts across the languages. Ideally, parallel structures should be annotated identically in all languages, and any differences in the annotated structures should indicate either language contrasts or non-trivial effects of the translation process. Unfortunately, experience shows that this is very difficult to achieve in practice. Annotated resources with texts in multiple languages invariably exhibit cross-linguistic variation that arises spuriously as a result of the annotation process and does not reflect any linguistically relevant phenomena.

We refer to the differences in annotated parallel texts as annotation *incongruences*. Manual detection of incongruences is not only time- and effort-consuming, but also inefficient. Despite the ubi-quity of this problem in all kind of parallel linguistic annotation, it has received little attention in the existing works.

In this paper, we address the problem of cross-lingual incongruences in the manual annotation of coreference. To our knowledge, none of the existing studies on parallel coreference annotation (Dipper and Zinsmeister, 2012; Zikánová et al., 2015; Grishina and Stede, 2017) addresses this issue. We analyse the incongruences in coreference chains in a subset of the corpus ParCorFull (Lapshinova-Koltunski et al., 2018), an English-German parallel corpus containing manual annotations of coreference.

We automatically extract annotated chains from the corpus, create an alignment between chains in the source and target language and identify those that do not have parallel equivalents in the source or the target language. Among the parallel chains, we detect those with differences in English and German. Our use of word alignments to map coreference structures between language is similar to the existing studies on annotation projection (e.g. Yarowsky et al., 2001) or specifically on multilingual coreference projection (Postolache et al., 2006; Ogrodniczuk, 2013; Grishina and Stede, 2015; Novák, 2018). However, in contrast to annotation projection, we do not aim to produce any automatic annotations. Instead, we use automated methods to discover incongruences in the existing annotations produced manually on parallel texts.

The cross-lingual variation in the chains is then analysed both quantitatively and qualitatively. We develop a typology of the incongruences encountered in ParCorFull, illustrated with corpus examples, and present empirical results on the prevalence of different types of variations using a corpus sample.

The results of this study facilitate the analysis of similarities and idiosyncracies in coreference

across languages and support the evaluation of cross-lingual coreference annotation. In this way, they contribute to contrastive linguistic and translation studies as well as to cross-lingual coreference resolution. Moreover, the method applied in this work can be used for automatic evaluation of other manually annotated structures in parallel data.

2 Annotation Incongruences

2.1 Definition of Incongruences

Annotation of multilingual data requires definition of universal categories that exist in all the languages involved. For instance, the English chain in example (1) is represented by a nominal phrase, a relative and a personal pronoun (*a close friend – who – she*). The corresponding German translation contains the three-member chain that also consists of a nominal phrase, a relative and a personal pronoun (*eine enge Freundin – die – ihr*).

(1) a. *I had [a close friend] from college [who]'d gone through a divorce and wanted to have children. And so [she] and I have a daughter, and mother and daughter live in Texas.*
   b. *Ich hatte [eine enge Freundin] aus Uni-Tagen, [die] eine Scheidung hinter sich hatte und Kinder wollte. Mit [ihr] habe ich also eine Tochter, und Mutter und Tochter leben in Texas .*

This is an ideal case of a parallel coreference chain, the only difference being the case of the personal pronoun *she/ihr*. However, even typologically close languages, like English and German, have systemic differences in the range of linguistic means triggering coreference (Kunz and Lapshinova-Koltunski, 2015; Novák and Nedoluzhko, 2015; Kunz and Steiner, 2012; Kunz, 2010). Moreover, translated texts differ from non-translated ones in terms of language use, as shown by corpus-based studies on translationese (Baroni and Bernardini, 2006; Ilisei et al., 2010, among others). As a result, parallel texts do not always contain identical chains. Equivalent chains may differ in the types of referring expressions[1] or they may differ in the number of referring expressions. Besides that, translations may contain chains that are not present in the source texts, and source

---

[1]Note that the differences in the types of referring expressions within the parallel chains are beyond the scope of this study.

texts may contain chains that do not appear in the target. We refer to these language-typology and translation-process-driven differences in the parallel chains as *annotation incongruences*. Furthermore, we realise that the differences in the annotated structures may also have their origin in the process of manual annotation, i.e. human annotators may have interpreted the source and the target text in a different way (especially if the annotation of the source and the target texts was performed independently), or may simply have made errors.

2.2 Typology of incongruences

We suggest that we can classify incongruences into four groups according to their sources: 1. explicitation; 2. implicitation; 3. annotation interpretation and 4. annotation error. The first two groups (1 and 2) are related to the hypothesis of explicitation, while the latter two groups (3 and 4) are related to the annotation process rather than the translation process.

1. Explicitation The hypothesis of explicitation was formulated by Blum-Kulka (1986). We adopt the definition of explicitation (and implicitation) introduced by Klaudy and Karoly (2005, p. 15), according to which explicitation takes place when a translation contains more specific linguistic units (instead of more general units in the source), or new linguistic units (not present in the source), a phrase is extended to clause level, a sentence is split into two sentences, etc. For the explanation of explicitation-induced incongruences, we use Klaudy's notion of operational asymmetry and her classification of explicitation into obligatory, optional, pragmatic and translation-inherent (see Klaudy, 2008, 106–107). Obligatory explicitation is dictated by differences in the syntactic and semantic structure of languages. Optional explicitation is related to the differences in text-building strategies and stylistic preferences. In the case of pragmatic explicitation, there is implicit cultural information and translators often need to include explanations. In example (2), the English source does not contain any coreference chains, whereas its German translation does. The German coreference chain includes two members: a noun phrase and a relative pronoun introducing a relative clause. The information contained in this relative clause was packed into a participle construction (non-finite *ing*-construction) in the English source. This clause type has a direct equival-

ent in German, the present participle *schreibend* ("writing"). However, the English *ing*-form is used much more widely than the German present. In particular, participial clauses are restricted to formal written registers in German and can sound stilted and they are used much less frequently than clauses with *ing*-forms in English (Durrell, 2011, p.281–285). This is a case of an obligatory explicitation – a translator has to add a relative clause and thus, a pronoun, to express the information, as the language system requires one. Further cases include an addition of a reflexive pronoun required by the verb valency.

(2) a. *It turned out that tens of thousands of autonomous individuals writing an encyclopedia...*

b. *Es stellte sich heraus, dass [Zehntausende von autonomen Einzelpersonen], [die] ein Lexikon schreiben...*

The decision to use more explicit constructions can have stylistic reasons, as in example (3). Here, the English source has a coordinated verb phrase which does not require a subject for the second verb, and the source chain has two members only. Instead, we find two subordinate clauses in the German target (corresponding to the two verbal clauses in English) which require another mention of the subject, and thus, a chain of three members.

(3) a. *...business strategy has always been premised on [assumptions about technology], that [those assumptions] are changing, and, in fact, changing quite dramatically...*

b. *...dass Geschäftsstrategie immer schon auf [Annahmen über Technologie] basiert, dass sich [diese Annahmen] ändern, und dass [sie] sich sogar ziemlich dramatisch ändern....*

Other explicitation cases are translation-inherent as they depend on a translator's decision, as in example (4). The German chain has the element *ihre* ("their"). This is a possessive pronoun modifying the noun phrase *ihre Kinder*, which is, however, a part of another coreference chain. In the source, this information is expressed via the pronoun *them*, where no modifying element is necessary.

(4) a. *And I suddenly thought, most deaf chil-*

dren are born to [hearing parents]. [Those hearing parents] tend to try to cure them.

b. *Und da dachte ich plötzlich, dass die meisten tauben Kinder [hörende Eltern] haben, und [diese] in der Regel versuchen, [ihre] Kinder zu heilen.*

While obligatory (language-typology-driven) explicitation is easy to identify, it is difficult to differentiate between optional, pragmatic and translation-inherent cases that are translation-process-driven. For this reason, we classify the analysed cases exposing explicitation into two groups only: obligatory and non-obligatory.

2. Implication Implication is an opposite process to explicitation and means that translations can be shorter, more compressed; e.g., the subject (which was a member of a chain in the source) was omitted and a coordinated verb phrase was used instead. In other cases, the information is packed into a different, more compact construction without a mention. As well as explicitation, implication can be obligatory and non-obligatory (see Klaudy and Károly, 2005, p. 16–17). In example (5), the English source contains a chain of three members and the third mention (*that*) is not present in the corresponding German sentence. This German sentence contains the discourse element *hier* which links this to the previous context. However, this element is not a member of the coreference chain, as the relation and its scope is different. The element *hier* refers to the whole situation and not to *logic*.

(5) a. *...and it maps exactly on to [the kind of Porter-Henderson logic] [that] we've been talking about. And [that] is, about data.*

b. *und es ordnet sich genau in [die Art der Porter-Henderson-Logik] ein, über [die] wir gesprochen haben. Es geht [hier] um Daten .*

Implication cases can also be related to the specific genre of our data – TED talks are subtitled and not translated, and compressing information is a core strategy in subtitling. This is a frequent cause of optional, non-obligatory implication. The kind of compression we observe in our translations results from the guidelines of reducing information to tackle reading-speed issues[2]. In the

German sentence in (6), we observe a compression of the information contained in the English source: *weekend I spent with them* vs. *gemeinsames Wochenende* ("joint weekend/weekend together").

(6) a. *And the first weekend I spent with [them] – the first of many – I recorded more than 20 hours of conversation.*
   b. *Am ersten gemeinsamen Wochenende – einem von vielen – zeichnete ich mehr als 20 Stunden an Gesprächsstoff auf.*

3. Different interpretations Annotators sometimes interpret German texts differently from the English sources. This is especially frequent with ambiguous cases, when it is difficult to understand exactly which components participate in the coreference relation. In example (7), the English texts contains two chains (*when we collect... – This – it – this* and *a revolution in medicine – this*). In the German translation, there is only one non-entity chain *wenn wir... – dies – es – darüber*. A different interpretation in the translation results from the fact that the English sentence contains the full verb *drive* with *a revolution in medicine* as a direct object. This creates a different identity and thus, a different coreference chain. In the German translation, the nominal phrase *eine Revolution in der Medizin* is linked to *es* with a copula verb and is, therefore, a part of the same identity expressed via pronoun.

(7) a. *Think what happens [when we collect all of that data and we can put it together in order to find patterns we wouldn't see before]$_1$. [This]$_1$, I would suggest, perhaps [it]$_1$ will take a while, but [this]$_1$ will drive [a revolution in medicine]$_2$. Fabulous, lots of people talk about [this]$_2$.*
   b. *Was passiert, [wenn wir all diese Daten sammeln und wir sie zusammenfügen können, um Muster zu erkennen, die wir nicht vorher sehen konnten]$_1$. Vielleicht dauert [dies]$_1$ ja noch eine Weile, aber [es]$_1$ wird eine Revolution in der Medizin. Fabelhaft – sehr viele Leute sprechen [darüber]$_1$.*

4. Annotation error This type of incongruences emerges due to errors in the manual annotation, such as if mentions were not included into the chains they should have been included to or there

were two shorter chains annotated instead of one longer one.

3 Data

For our analysis, we use a subset of the parallel corpus ParCorFull (Lapshinova-Koltunski et al., 2018). This corpus contains English texts and their German translations that were annotated with coreference chains. The underlying coreference scheme was designed for uniform coreference annotations of a multilingual corpus (see Lapshinova-Koltunski and Hardmeier, 2017, for details).

The annotated elements (markables) in this corpus include pronouns, nouns, nominal phrases or elliptical constructions that are parts of a coreference pair (antecedent-anaphora), as well as verb phrases or clauses being antecedents of event anaphora. The annotated antecedents are of two different types: entities and events. For the analysis in this paper, we restrict ourselves to chains with nominal antecedents, excluding event reference with verbal and clausal antecedents.

Entities can be represented by a pronoun or a noun phrase. Antecedents can be split, i.e. two pronouns or two nouns (disjoint in a text) constitute one antecedent – all components of the antecedent are linked to a referring expression. The annotated referring expressions (anaphors) are represented by pronouns (personal, demonstrative, relative and reflexive) and nominal phrases. Demonstrative pronouns may also refer to locations (*there*, *here*) and time (*then*, *now*). There are also pronominal adverbs formed by replacing a preposition and a pronoun, like *für+das* → *dafür* ("for this"). These are very common in German, but sound rather archaic and are generally avoided in English. Coreferring nominal phrases include proper names, nominal premodifiers, full nominal phrases and nominal phrases with quantifiers (see more details in Lapshinova-Koltunski et al., 2018). Linguistic chains may also include substitution and ellipsis in addition to referring expressions[3] – they often occur in similar contexts as coreference if considered cross-lingually.

The whole version of ParCorFull contains ca. 161,000 words. Our subset includes 77,216 word of TED talks (39,764 of the English TED talks and

---

[3]Although substitution and ellipsis do not express identity, they are included into the annotation scheme of ParCorFull, as they express the relations of near-identity and may occur in the same context as coreference.

37,452 of their German translations) and 21,237 words of news (10,644 English and 10,593 German texts).

4 Extraction Method

As mentioned in Section 3, we concentrate on the extraction of entity coreference chains only. We start by computing word level alignments between the source and target sides of the corpus in both directions using Giza++ (Och and Ney, 2003) with grow-diag-final symmetrization (Koehn et al., 2005). To align chains, we compute a matching score between each pair of source and target chains in the document, based on the alignment points they share. Alignment points are words in the source language aligned to words in the target language. Since word alignments are not necessarily one to one, each word may have no, one or multiple alignment points (Koehn, 2010).

For each potential chain pair, we take all the words in the source chain (all mentions) and count their alignment points with the words in the target chain, and repeat the process in the other direction. We then compute the average between the alignment points source-target and target-source and take the pair with the highest score as a pair match:

$$C_1 = |\{s \in S | \exists t \in T : (s,t) \in A\}|$$
$$C_2 = |\{t \in T | \exists s \in S : (s,t) \in A\}|$$
$$\text{score} = (C_1 + C_2)/2$$

where S and T are the sets of word indices belonging to the English and German chain, respectively, and A is the set of alignment points (pairs of source and target indices).

Potentially, two pairs of chains could have the same score. This never occurred with the relatively short nominal chains we considered for this paper. However, we expect it to happen with longer chains, for instance those corresponding to events.

5 Results

5.1 Automatic extraction

Our extraction procedure yields different categories of automatically identified incongruences. We group them according to the categories in Table 1.

I **Matching chains.** We approximate the concept of matching chains by considering pairs of chains in which both the source and the target chains contain the same number of mentions – as in example (1), where the English source chain *a close friend – who – she* corresponds to the German chain *eine enge Freundin – die – ihr*. While this simple operationalisation misses certain interesting transformation (such as alternations between pronouns and named entities, or changes in the order of the mentions), it allows us to concentrate on cases II and III, where changes are happening with certainty.

II **Overlapping chains.** Here we have matching chain pairs with a different number of mentions in either side of the corpus. We subdivide them according to the chain length with longer chain either in the English source, i.e., English has more mentions, or in the German translations, i.e., more German mentions.

III **Unpaired chains.** These are chains in either side of the corpus for which no chain correspondence is found. In the sample data under analysis, all cases of this type are German chains without a correspondence in English.

The results show that the analysed subcorpus contains approximately 32% chain matches, i.e., equivalent chains that have the same number of referring expression. These chains may still differ in the type of referring expressions contained in these chains. However, this variation is beyond the scope of this study. We restrict our analysis to the equivalent chains with a different number of mentions (overlapping chains) that constitute approximately 36% of the extracted chains. In this case, it is difficult to extract mention correspondences automatically. However, it can be seen that among the overlapping chains, most frequently, the German chains are longer than their English counterparts (334 vs. 210). Last, we also observe a considerable number of chains annotated in the German translation only (31%), whereas there are just a few cases of the unpaired English chains, i.e., those annotated in the source text only.

We assume that explicitation in our data is represented by cases where we observe more German mentions and unpaired German chains, whereas implicitation is related to the categories of more English mentions and unpaired English chains. Along these lines, explicitation would comprise

Chain category		TED		News		Total		Analysis
		#	%	#	%	#	%	
Matching	same number of mentions	392	32	84	33	476	32	
Overlapping	more English mentions	174	14	36	14	210	14	Implicitation
	more German mentions	293	24	41	16	334	22	Explicitation
Unpaired	English chains	0	0	9	4	9	1	Implicitation
	German chains	376	30	84	33	460	31	Explicitation
Total number of chains		1235	100	254	100	1489	100	

Table 1: Incongruences found automatically in the annotation of coreferential chains in the ParCorFull corpus.

Types of incongruences	TED		News	
	#	%	#	%
Explicitation	40	37.4	27	41.5
Implicitation	7	6.5	8	12.3
Dif. interpetation	16	15.0	0	00.0
Annontation error	44	41.1	30	46.2
Total chains	107	100.0	65	100.0

Table 2: Result of manual analysis of incongruences.

around 80% of the extracted incongruences, and implicitation around 20%. However, the extracted incongruences might contain phenomena not related to the annotation or the translations itself. In the following section we present a manual analysis to investigate this further.

5.2 Manual analysis

We select a set of overlapping and unpaired chains from a TED text and from several news texts[4] for our manual analysis. The cases are classified according to the four categories defined in Section 2 above. We summarise their distributions in Table 2. We also try to identify the reasons for the specific incongruences.

Explicitation The results of the manual analysis, however, show that 37.4% of the analysed TED talk chains and 41.5% chains in news are cases of explicitation – German translations are longer than the corresponding English sources, and thus, they contain additional elements of coreference chains. Most of these cases are represented by relative clauses as illustrated in example (2) in Section 2 above. In this example, we find a non-finite construction in English that has to be transferred into German and has no equivalent con-

struction. Non-finite constructions contain participles and also infinitives like the one in example (8). These are cases of obligatory explicitation.

(8) a. *[the first one in his family] to go to college.*
    b. *[der Erste in seiner Familie], [der] an einer Universität studierte.*

Our data also contains examples where a relative pronoun is omitted in English, as in example (9). The mismatch between the chains is not caused by the difference of the constructions used – there are relative clauses in both the source and the target. The difference is in the degree of explicitation of this clause – English does not require a relative pronoun, whereas German does. Therefore, a translator has to make the German target sentence more explicit.

(9) a. *Those are [things] you have in common with your parents and with your children.*
    b. *[Dinge], [die] Sie mit Ihren Eltern und Kindern gemein haben.*

Example (10) illustrates another mismatch between relative clauses in English and German. In this case, the relative clause is not obligatory, and we describe it as a case of non-obligatory explicitation: There is a temporal clause introduced with *those moments... when* in the source, which is transferred with a relative clause (with a locative function) into German: *jene Momente... in denen.* The temporal clause does not belong to a coreference chain and is, therefore, not annotated in our corpus.

(10) a. *...like [those moments] in grand opera [when] the hero realizes he loves the heroine.*

---

[4]News texts are shorter than TED talks.

b. ...an [jene Momente] in der Oper erinnert, in [denen] der Held erkennt, dass er die Heldin liebt.

Explicitation through relative pronouns makes up 70% of the observed cases. Further examples of explicitation include adding possessives, like in example (4) in Section 2, where the pronoun *them* is transferred into *ihre Kinder*. This is a case of non-obligatory explicitation. Example (11) contains a similar transformation pattern (*parenting – ihre Mutterrolle*) with a different source of explicitation. In English, *parenting my brother and me* does not require any modifier, whereas the German noun phrase *Mutterrolle für ich und meinen Bruder* requires either the definite article *the* or the possessive pronoun *ihre*. In this way, German is more explicit.

(11) a. ...[my mother] used to say... I took it as the greatest compliment in the world that [she] would say that about parenting my brother and me.
b. sagte [meine Mutter] immer... Als Kind nahm ich das als das größte Kompliment, dass [sie] so [ihre] Mutterrolle für mich und meinen Bruder beschreiben würde.

Implicitation Implicitation comprises 6.5% in the analysed TED talks and 12.3% in the analysed news. In most cases, we observe omission of the subject pronoun in the German sentence, and a verb phrase is used instead of a clause, see example (12). In the English sentence, the first relative clause introduced with *who* contains a verb in passive voice, whereas the second has an active verb. Therefore, the second subject expressed through the relative pronoun *who* is necessary here. In the German translations, both clauses are active.

(12) a. [mice] [who] have been given that substance and [who] have the achondroplasia gene, grow to full size.
b. [Mäuse], [die] diesen Wirkstoff erhalten haben, und das Achondroplasie-Gen aufweisen.

In example (13), both English and German sentences contain clauses that have the same verb tense and voice. However, the translator decided to omit the subject in the target.

(13) a. And [Sue] looked at the floor, and [she] thought for a minute.
b. Und [Sue] schaute auf den Boden und dachte eine Minute nach.

Simplifying syntax by merging sentences is recommended as a strategy for subtitle compression[5]. Thus, the analysed cases of implicitation in our data could be genre-specific.

Different interpretations Differences in how the source and the target text were interpreted only affect the incongruences in the analysed TED talk (15%) and do not occur in the news sample. These are mostly cases of overlapping chains containing more German mentions or German unpaired chains. In example (14), the annotator identified different antecedents in the source and the target sentence. The English chain contains two elements – a split antecedent that consists of three noun phrases (*self-acceptance, family acceptance* and *social acceptance*) and the anaphor *they* referring to them. The German chain starts in the previous sentence and has the antecedent *drei Stufen der Akzeptanz* that corresponds to *three levels of acceptance* which is not marked in the English sentence, the anaphor *die* that corresponds to the English relative *that* and the anaphor *die drei* corresponding to the pronoun *they* in the English source. Both chain variants can be considered as correct chains depending on how the text is interpreted.

(14) a. ...that there were three levels of acceptance that needed to take place. There's [self-acceptance]$_1$, there's [family acceptance]$_1$, and there's [social acceptance]$_1$. And [they]$_1$ don't always coincide.
b. dass es [drei Stufen der Akzeptanz]$_2$ gibt, [die]$_2$ alle zum Tragen kommen mussten. Da war die Eigenakzeptanz, die Akzeptanz der Familie und die gesellschaftliche Akzeptanz. Und [die drei]$_2$ überschneiden sich nicht immer.

The scope of a relation can be interpreted in a different way, if an anaphor is ambiguous. In example (15), the pronoun *it* refers to an event which might be expressed by either *putting ... away* or *is put away*. The annotator marks the first one in the

English source and the second in the German target. The latter is expressed via the deverbal noun *Weggeben* ("putting away"), which is annotated as a nominal antecedent.

(15) a. *"There is no reason to feel guilty about [putting a Down syndrome child away]₁ whether it is put away in the sense of hidden in a sanitarium... [It]₁ is sad, yes – dreadful. But [it]₁ carries no guilt.*

b. *"Es gibt keinen Grund, sich schuldig zu fühlen, wenn man ein Kind mit Down-Syndrom weggibt, egal ob es sich dabei um [ein " Weggeben"]₂ im Sinne von 'in einem Heim verstecken" handelt... [Es]₂ ist traurig, ja – und schrecklich. Aber [es]₂ entbehrt jeder Schuld.*

Annotation error   The analysed incongruences contain 43% of annotation errors (44% in the TED talk and 30% in the news). These errors can be classified into the following categories: (a) different chain membership as illustrated in example (7) in Section 2 above; (b) non-marked mentions or chains – annotation is missing in either English or German; and (c) incorrectly marked mentions. The first case is especially frequent in overlapping chains (containing both more English and German mentions). The second error type is represented mostly by the unpaired German chains. The last error category is scattered across different types of incongruences.

6   Conclusion and Discussion

In this paper, we analysed incongruences in parallel coreference annotation and suggested a typology based on their sources. The results showed that many incongruences in our data are due to explicitation, i.e., German translations contain more explicit linguistic means that trigger coreference. We also showed that explicitation has its origin either in language typology – idiosyncracies between the two languages under analysis in terms of coreference, or in the translation process. Besides that, we detected differences in the interpretation of the source and target texts along with annotation errors. They both result from the way the annotation was performed – although the annotation scheme includes universal categories for both languages, the annotation process itself was not parallel, and the source and the target texts were annotated independently. This raises the question of annotation strategies, when working cross-lingually. Could those incongruences be avoided if the work were performed in parallel? This would require the annotators to have a very good command of both languages. However, parallel annotation could cause different problems, e.g., by biasing the annotation of the target text excessively towards the source. Another option is to annotate texts independently and then cross-check them in parallel, which might help to detect chains and mentions that were "overseen" in the independent annotation procedure.

In the future, we aim at automating the classification of the extracted incongruences according to the suggested typology. Automatic extraction of annotation error candidates can also help in the improvement of the existing annotation and saves time, as the annotators do not have to read all the texts from scratch, which reduces manual correction effort.

Moreover, we plan to analyse more texts manually to find out if the incongruence categories are systemic across the whole corpus at hand. As there is a subcorpus of news in our data, we can also investigate genre-related effects. Furthermore, we will perform analysis of further types of coreference chains, i.e. non-entity coreference. Although annotated in our data, they were excluded from analysis for practical reasons.

Another extension of the study includes analysis of the differences in the type of referring expressions in parallel chains, as mentioned in Section 5.1 above. Non-equivalence of the referring expressions in parallel chains represents coreference transformations in English-German translations (for instance, a nominal phrase in English is translated as a pronoun in German, etc.). This kind of information is valuable for contrastive linguistics and translation studies as it delivers information on different strategies in information status presentation in English and German.

The problem of making annotations of parallel texts consistent across languages was here studied in the context of coreference annotation, but it clearly poses a challenge for all types of multilingual linguistic annotation. More systematic and automatic methods to improve cross-lingual annotation congruence have the potential to benefit applications and research in language technology, contrastive linguistics and translation studies alike.

References

Marco Baroni and Silvia Bernardini. 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21(3):259–274.

Shoshana Blum-Kulka. 1986. Shifts of cohesion and coherence in translation. In J. House and S. Blum-Kulka, editors, *Interlingual and intercultural communication*, pages 17–35. Gunter Narr, Tübingen.

Stefanie Dipper and Heike Zinsmeister. 2012. Annotating abstract anaphora. *Language Resources and Evaluation*, 46(1):37–52.

Martin Durrell. 2011. *Hammer's German Grammar and Usage*, 5 edition. Routledge, London and New York.

Yulia Grishina and Manfred Stede. 2015. Knowledge-lean projection of coreference chains across languages. In *Proceedings of the 8th Workshop on Building and Using Comparable Corpora*, page 14, Beijing, China.

Yulia Grishina and Manfred Stede. 2017. Multi-source annotation projection of coreference chains: assessing strategies and testing opportunities. In *Proceedings of the 2nd Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2017)*, pages 41–50. Association for Computational Linguistics.

Iustina Ilisei, Diana Inkpen, Gloria Corpas Pastor, and Ruslan Mitkov. 2010. Identification of translationese: A machine learning approach. In *Computational linguistics and intelligent text processing*, pages 503–511. Springer Berlin Heidelberg.

Kinga Klaudy. 2008. Explicitation. In M. Baker and G. Saldanha, editors, *Routledge Encyclopedia of Translation Studies*, 2 edition, pages 104–108. Routledge, London & New York.

Kinga Klaudy and Krisztina Károly. 2005. Implicitation in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures*, 6:13–28.

Philipp Koehn. 2010. *Machine Translation*. Cambridge University Press, Cambridge.

Philipp Koehn, Amittai Axelrod, Alexandra Birch Mayne, Chris Callison-Burch, Miles Osborne, and David Talbot. 2005. Edinburgh system description for the 2005 IWSLT speech translation evaluation. In *International Workshop on Spoken Language Translation*, Pittsburgh, Pennsylvania.

Kerstin Kunz. 2010. *Variation in English and German Nominal Coreference: A Study of Political Essays*. Saarbrücker Beiträge zur Sprach- und Translationswissenschaft. Peter Lang.

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.

Kerstin Kunz and Erich Steiner. 2012. Towards a comparison of cohesive reference in English and German: System and text. In M. Taboada, S. Doval Suárez, and E. González Álvarez, editors, *Contrastive Discourse Analysis. Functional and Corpus Perspectives*. Equinox, London.

Ekaterina Lapshinova-Koltunski and Christian Hardmeier. 2017. *Coreference Corpus Annotation Guidelines*.

Ekaterina Lapshinova-Koltunski, Christian Hardmeier, and Pauline Krielke. 2018. ParCorFull: a parallel corpus annotated with full coreference. In *Proceedings of 11th Language Resources and Evaluation Conference*, pages 423–428, Miyazaki, Japan. European Language Resources Association (ELRA).

Michael Novák and Anna Nedoluzhko. 2015. Correspondences between Czech and English coreferential expressions. *Discours*, 16.

Michal Novák. 2018. A fine-grained large-scale analysis of coreference projection. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 77–86, New Orleans, Louisiana. Association for Computational Linguistics.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29:19–51.

Maciej Ogrodniczuk. 2013. Translation- and projection-based unsupervised coreference resolution for Polish. *Language Processing and Intelligent Information Systems, IIS 2013*, 7912.

Oana Postolache, Dan Cristea, and Constantin Orasan. 2006. Tranferring coreference chains through word alignment. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the First International Conference on Human Language Technology Research*.

Šárka Zikánová, Eva Hajičová, Barbora Hladká, Pavlína Jínová, Jiří Mírovský, Anna Nedoluzhko, Lucie Poláková, Kateřina Rysová, Magdaléna Rysová, and Jan Václ. 2015. *Discourse and Coherence. From the Sentence Structure to Relations in Text*. ÚFAL, Praha, Czechia.