A Pilot Experiment on Exploiting Translations for Literary Studies on Kafka's "Verwandlung"

Fabienne Cap, Ina Rösiger and Jonas Kuhn
Institute for Natural Language Processing
University of Stuttgart
Germany
[cap|roesigia|kuhn]@ims.uni-stuttgart.de

Abstract

We present a manually annotated word alignment of Franz Kafka's *"Verwandlung"* and use this as a controlled test case to assess the principled usefulness of word alignment as an additional information source for the (monolingually motivated) identification of literary characters, focusing on the technically well-explored task of co-reference resolution. This pilot set-up allows us to illustrate a number of methodological components interacting in a modular architecture. In general, co-reference resolution is a relatively hard task, but the availability of word-aligned translations can provide additional indications, as there is a tendency for translations to *explicate* underspecified or vague passages.

1 Introduction

We present a pilot study for a methodological approach starting out with combinations of fairly canonical computational linguistics models but aiming to bootstrap a modular architecture of text-analytical tools that is more and more adequate from the point of view of literary studies. The core modeling component around which our pilot study is arranged is word-by-word alignment of a text and its translation, in our case Franz Kafka's *"Verwandlung"* (= *Metamorphosis*) and its translation to English. As research in the field of statistical machine translation has shown, word alignments of surprisingly good quality can be induced automatically exploiting co-occurrence statistics, given a sufficient amount of parallel text (from a reasonably homogeneous corpus of translated texts). In this pilot study, we present a manually annotated reference word alignment and use this to assess the principled usefulness of word alignment as an additional information source for the (monolingually motivated) task of identifying mentions of the same literary character in texts ((Bamman et al., 2014)). This is a very important analytical sub-step for further analysis (e.g., network analysis, event recognition for narratological analysis, stylistic analysis of character speech etc.). Literary character identification is related to, but not identical to named entity recognition, as is pointed out in Jannidis et al. (2015). In addition, co-reference resolution is required to map pronominal and other anaphoric references to full mentions of the character, using his or her name or other characteristic descriptions. In our present study we focus on the technically well-explored task of co-reference resolution.[1] This pilot set-up allows us to illustrate a number of methodological components interacting in a modular architecture.

2 Background

Whenever computational linguists exchange thoughts with scholars from literary studies who are open-minded towards "the Digital Humanities", the feeling arises that the machinery that computational linguistics (CL)/Natural Language Processing (NLP) has in their toolbox should *in principle* open up a variety of analytical approaches to literary studies – if only the tools and models were

[1] In their large-scale analysis of >15,000 English novels, Bamman et al. (2014) adopt a simpler co-reference resolution strategy for character clustering. Our work explores the potential for a more fine-grained analysis.

Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature, pages 48–57,
Denver, Colorado, June 4, 2015.

appropriately adjusted to the underlying research questions and characteristics of the targeted texts. At a technical level, important analytical steps for such studies are often closely related to "classical" tasks in CL and NLP. Possible applications include the identification of characters in narrative texts, extraction of interactions among them as they are depicted in the text, and possibly more advanced analytical categories such as the focalization of characters in the narrative perspective. Based on preprocessing steps that extract important feature information from formal properties of the text, algorithmic models that are both theoretically informed and empirically "trained" using reference data, should lead to automatic predictions that will at least support exploration of larger collections of literary texts and ideally also the testing of quantitative hypotheses.

Of course, the qualifying remark that the tools and models would first need to be adjusted to the higher-level questions and the characteristics of the texts under consideration weighs quite heavily: despite the structural similarity with established NLP analysis chains, the *actual* analytical questions are different, and NLP tools optimized for the standard domains (mostly newspaper text and parliamentary debates) may require considerable adjustment to yield satisfactory performance on literary texts.

2.1 Methodological Considerations

The large-scale solution to these challenges would be to overhaul the entire chain of standard NLP processing steps, adjusting it to characteristics of literary texts and then add new components that address analytical questions beyond the canonical NLP offerings. This would however presuppose a master plan, which presumably requires insights that can only come about during a process of stepwise adjustments. So, a more realistic approach is to bootstrap a more adaquate (modular) system architecture, starting from some initial set-up building on existing machinery combined in a pragmatic fashion. Everyone working with such an approach should be aware of the limited adequacy of many components used – but this may be a "healthy" methodological training: no analytical model chain should be relied on without critical reflection; so an imperfect initial architecture may increase this awareness and help adopting to meta-level analysis tools for visualization and stepwise evaluation. Ideally, it should also make it clear that the literary scholars' insights into the texts and the higher-level questions are instrumental in improving the system at hand, moving to more appropriate complex models (taking a view on modeling in the sense of McCarty (2005) as an iterative cycle of constructing, testing, analyzing, and reconstructing intermediate-stage models, viewed as tools to advance our understanding of the modeled object).[2]

2.2 Pilot Character of Word-aligned Text

This paper tries to make a contribution in this spirit, addressing a methodological modular "building block" which (i) has received enormous attention in technically oriented NLP work, and (ii) intuitively seems to bear considerable potential as an analytical tool for literary studies – both for in-depth analysis of a small selection of texts and for scalable analysis tasks on large corpora – and for which (iii) a realistic assessment leads one to expect the need for some systematic adjustments in the methods and machinery to make the established NLP approach fruitful in literary studies. We are talking about the use of translations of literary works into other languages and techniques from NLP research on statistical machine translation and related fields. Literature translations exist in abundance (relatively speaking), often in multiple variants for a given target language, and multiple target languages can be put side by side. Such translation corpora ("parallel corpora") are a natural resource for translational studies, but research in various subfields of NLP has shown that beyond this, the interpretive work that a translator performs as a by-product of translating a text, can often be exploited for analytical questions that are not *per se* taking a cross-linguistic perspective: Tentative word-by-word correspondences in a parallel corpus can be induced automatically, taking advantage of co-occurrence statistics from a large collection, with surprising reliability. These "word alignment" links can then be used to map concepts that are sufficiently invariant across languages. The so-called paradigm of "annotation projection" (pio-

[2] For a more extensive discussion of our methodological considerations, see also Kuhn and Reiter (2015 to appear).

neered by Yarowsky et al. (2001)) has been enormously successful, even for concepts that one would not consider particularly invariant (such as grammatical categories of words): here, strong statistical tendencies can be exploited.

Since the statistical induction of word alignments requires no knowledge-based preprocessing (the required sentence alignment can also be calculated automatically), it can in principle be applied to any collection of parallel texts. Hence, it is possible to test quite easily for which analytical categories that are of interest to literary scholars the translator's interpretive guidance could be exploited.

As pointed out in the introduction, we pick literary character identification as an analytical target category that is very central to literary studies of narrative text, focusing on the task of building chains of co-referent mentions of the same character. Co-reference resolution is a relatively hard task, but the availability of word-aligned translations can provide additional indications: surprisingly often, translators tend to use a different type of referential phrase in a particular position: pronouns are translated as more descriptive phrases, and vice versa. A hypothesis that is broadly adapted in translational studies states there is a tendency for translations to *explicate* underspecified or vague passages (Blum-Kulka, 1986). An example of "explication" that affects character mentions is found in the second sentence of the English translation of Kafka's *"Der Prozess"*:[3] the apposition *seiner Zimmervermieterin* ("his landlady"), whose attachment is structurally ambiguous, is translated with a parenthetical that makes the attachment to *Mrs. Grubach* explicit:

(DE) Die Köchin der Frau Grubach, seiner Zimmervermieterin, die ihm jeden Tag gegen acht Uhr früh das Frühstück brachte, kam diesmal nicht.

(EN) Every day at eight in the morning he was brought his breakfast by Mrs. Grubach's cook – Mrs. Grubach was his landlady – but today she didn't come.

As the example also illustrates, potential referential ambiguity is however only one aspect translators deal with. Here, the translation avoids the long relative clause (*die ... brachte*) from the original after the initial subject, at the cost of using a completely different sentence structure. As a side effect, an additional referential chain (*Mrs. Grubach's cook – she*) is introduced in the English translation. So, it is an open question how effective it is in practice to use translational information in co-reference resolution of the original.[4]

For the purposes of this paper, which are predominantly methodological, aiming to exemplify the modular bootstrapping scenario we addressed above, the combination of word-alignment and co-reference is a convenient playing ground: we can readily make use of existing NLP tool chains to reach analytical goals that are structurally not far from categories of serious interest. At the same time, the off-the-shelf machinery is clearly in a stage of limited adequacy, so we can point out the need for careful critical reflection of the system results.

2.3 Reference Data and Task-based Evaluation

To go beyond an impressionistic subjective assessment of some analytical tool's performance (which can be highly misleading since there are just too many factors that will escape attention), it is crucial to rely on a controlled scenario for assessing some modular component. It is indeed not too hard to arrive at a manually annotated reference dataset, and this can be very helpful to verify some of the working assumptions. In our case, working with translated literary texts, we made various assumptions: automatic word alignment will be helpful for accessing larger amounts of text; standard techniques from machine translation are applicable to this; word alignment can be used to "project" invariant analytical categories for text segments etc.

A manual word alignment gives us a reference dataset that can give us a clearer picture about many of these assumptions: we can compare an automatic alignment obtained by various techniques with the reference alignment; we can check how "projection" works in the ideal case that the alignment is correct

[3] www.farkastranslations.com/books/Kafka_Franz-Prozess-en-de.html

[4] An additional advantage of literary texts that we are not going into here is that often multiple translations (to the same or different languages) are available, which a robust automatic approach could exploit, hence multiplying the chance to find a disambiguating translation (see also (Zhekova et al., 2014)).

etc. With Kafka's *"Verwandlung"* we chose a literary text that has some convenient properties at a superficial level. At the same time, Kafka's extremely internalized narrative monoperspective makes this text highly marked. So, in a future study that goes deeper into narrative perspectives, this reference text may be complemented with other examples.

3 Manual Word Alignment

The basis for a good word alignment is a reliable sentence alignment. However, the latter is a challenging task on its own – especially when it comes to literature translations – and is thus beyond the scope of this paper. We start from a freely available version of Franz Kafka's *"Verwandlung"* which has been carefully sentence-aligned and provided for download[5]. We selected *"Verwandlung"* for two reasons: the first has to do with the original language of the work, here: German. Usually, human translators tend to *explicate* ambiguities in their translations. We thus assume that the word alignment will be useful for German co-reference resolution. The other has to do with the limitation of this work, both in terms of quantity (the parallel text to be manually aligned consisted of only 675 lines), and in terms of the low number of characters involved. It is fairly simple to resolve ambiguities occurring in this limited set of characters for a human annotator, but that does not neccessarily apply to an automatic co-reference resolver.For this pilot study, manual word alignments were established by a German native speaker with a computational linguistics background. The annotator was asked to mark translational correspondences. In lack of suitable correspondences, words are to be left unaligned.

3.1 Results

In the following, we quantitatively summarize the alignments we established for *"Verwandlung"*, focussing on personal pronouns and their respective translations. Let us first consider the translations for the pronouns of the original, German, language in Table 1 (a): it can be seen that indeed there are a few cases where a German pronoun is translated into a more specified person in English (see highlighted markup of the respective table cells). An ex-

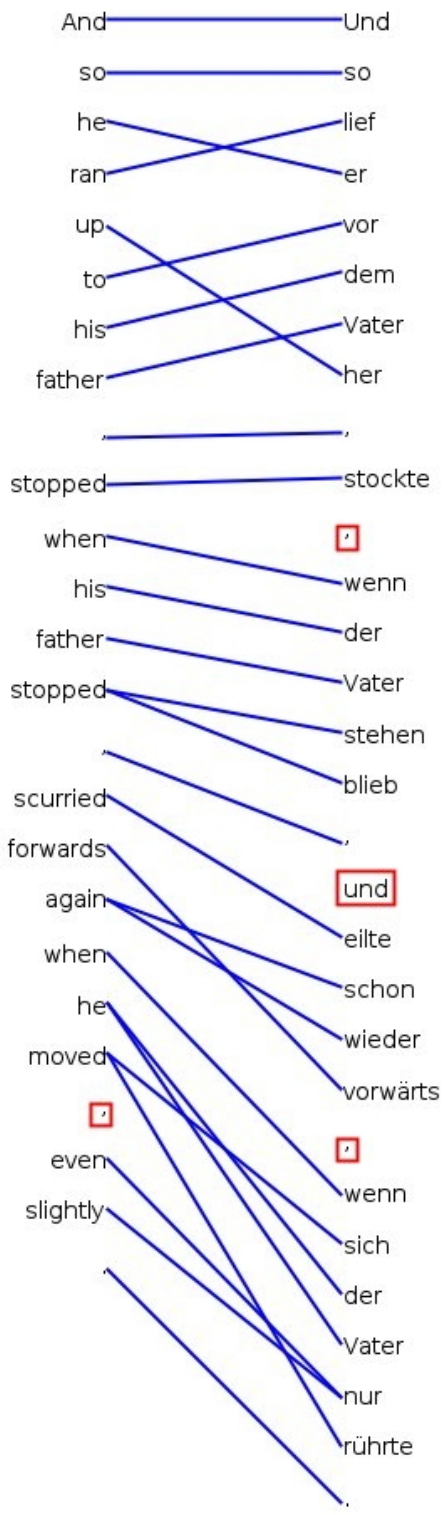

Figure 1: Screenshot of manual word alignment tool. It can be seen how the alignment from the second **he** to **der Vater** helps disambiguate the co-reference.

[5] http://opus.lingfil.uu.se/Books.php

(a) German pronouns together with their English translations

German	English translation					all
er	he	his	himself	him	Gregor's father	
	315	12	9	2	1	339
ihn	him	it	he	his		
	50	17	9	2		78
ihm	him	he	his	it		
	42	11	4	2		59
sein/seine[nmrs]	his	he	the	him	Gregor's	
	142	5	5	4	1	157
sie *(sg)*	she	her	his sister			
	107	24	1			132
ihr/ihre[nmrs] *(sg)*	her	she	the	his sister's		
	57	10	4	1		72
sie *(pl)*	they	their/them	the food	the pain	Gregor's father and mother	
	52	8	1	1	1	63
ihr/ihre[nmrs] *(pl)*	their	they	them			
	25	1	1			15

(b) English pronouns together with their alignment to the original German text

English	German original text							all
he	er	ihm	ihn	Gregor	man	sein/sein[er]	der Vater	
	349	12	9	10	5	5	2	390
him	ihn	ihm	Gregor	er	sein/sein[emrs]	sich	Vater	
	51	43	15	11	7	6	1	134
his	die	sein/sein[emrs]	de[nmrs]	ihm	Gregors	eine[nr]	des Vaters	
	140	140	111	4	4	4	2	380
she	sie	ihr/ihre[rm]	die	die Schwester	die Bedienerin	die Mutter	Grete	
	115	11	2	2	2	1	1	134
her	ihr/ihre[nmr]	de[mnr]	sie	die	(die) Mutter	die Schwester	das	
	51	24	22	17	2	2	2	120

Table 1: Overview of how German (a) and English (b) pronouns have been translated. The translations are obtained through manual alignment of the parallel German and English editions of the work. Highlighting indicates pronouns where the translations might actually help co-reference resolution.

ample is *"er"* (= "he"), which in one case is translated as "Gregor's father". In case of the plural pronoun *"sie"* (= "they") we can see that it is translated as "Gregor's father and mother" in one case and into the more abstract entities "the food" and "the pain". Even though not denoting personal entities, the latter can still help resolving the other pronouns that might occur in the close context. In Table 1 (b), we give results for the opposite direction, namely we show the German original words from which the English pronouns have been translated. Comparing these two tables, we can see that in general, more English pronouns are used than German ones (cf. last column "all" indicating the frequency with which the pronoun has occurred overall in the text). Be it a consequence thereof or not, we also find more resolved co-referential ambiguities in this translation direction. While "he" has been translated 10 times from *"Gregor"* and two times from *"der Vater"* (= "the father"), we find a more diverse distribution when looking at the female counterpart *"she"*, which has amongst others been translated from *"die Schwester"* (= "the sister"), *"die Bedienerin"* (= "the charwoman"), *"die Mutter"* (= "the mother") or *"Grete"*. In the next section, we will run a state-of-the-art co-reference resolver on both the German original and the English translation of the novel. For a subset of pronouns, we will then manually compare the outcome of the resolver with the translation to see in which of the cases highlighted in Table 1 (where access to a translation might help co-reference resolution), the access to the translation actually can improve co-reference resolution.

4 Automatic Coreference Resolution

Noun phrase coreference resolution is the task of determining which noun phrases (NPs) in a text or dialogue refer to the same real-world entities (Ng, 2010). Coreference resolution has been extensively addressed in NLP research, e.g. in the CoNLL shared task 2011 and 2012 (Pradhan et al., 2011; Pradhan et al., 2012)[6] or in the SemEval shared task 2010 (Recasens et al., 2010)[7]. State-of-the-art tools that take into account a variety of linguistic rules and features, most of the time in a machine learning setting, achieve promising results. However, when applying co-reference resolution tools on out-of-domain texts, i.e. texts that the system has not been trained on, performance typically decreases. Thus, co-reference resolution on literature text is even more challenging as most state-of-the-art co-reference resolver are trained on newspaper text. For a system that has been trained on newspaper articles, it is difficult to resolve the longer literary texts that typically revolve around a fixed set of characters. In our experiments, we for example observe a tendency of the resolver to create several co-reference chains for one character. Domain-adaptation is time-consuming, as it often requires manually designed gold data to increase performance. Moreover, recall that Kafka's *"Verwandlung"* has been written 100 years ago and that German language has been changing in this time period. This might lead to additional sparsities.

Apart from that, there are even some more general difficulties an automatic co-reference resolver has to deal with: First, it is difficult for a system to resolve an NP that has more than one potential antecedent candidates that match morpho-syntactically, i.e. that agree, for example, in number and gender. Second, often background or world knowledge is required to find the right antecedent. Consider the following examples taken from *"Verwandlung"* showing gold co-reference annotations through different colour markup: **Gregor** and **the father**.

(1) And so **he** ran up to **his father**, stopped when **his father** stopped, scurried forwards again when **he** moved, even slightly.

(2) **He** had thought that nothing at all remained from **his father's** business, at least **he** had never told **him** anything different, and **Gregor** had never asked **him** about it anyway.

For an automatic system, it is easy to confuse the two male persons present in the sentence, as they are both singular and masculine. Interestingly, humans can easily resolve these cases.

Due to the above mentioned reasons, it is particularly important to exploit given resources in a certain domain. In the literature area, translations into many languages are typically available. In the following, we will explain the benefits of using such translations by showing examples of how manual word alignment can help co-reference resolution, here in the case of *"Verwandlung"*. We will also talk about the prospects of automatic word alignment.

4.1 Experimental Setup

For English, we perform our experiments using the IMS HotCoref co-reference resolver (Björkelund and Kuhn, 2014) as a state-of-the-art co-reference resolution system that obtains the best results published to date on the English CoNLL 2012 Shared Task data[8]. It models co-reference as a directed rooted tree, making it possible to use non-local features that go beyond pair-based information. We use the English non-local model[9] that comprises the training and development datasets from the CoNLL 2012 shared task. These datasets, as common in most NLP tasks, mainly consists of news text and dialogue.

For German, our experiments are also based on the IMS HotCoref system, but as German is not a language that is featured in the standard resolver, we first had to adapt it to it. These adaptations include gender and number agreement, lemma-based (sub)string match and a feature that addresses German compounds, to name only a few. Again, the training data consists of newspaper texts as we base our experiments on the Tueba-D/Z SemEval data[10] (version 8).

[6] http://conll.cemantix.org/2012

[7] http://stel.ub.edu/semeval2010-coref/

[8] www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/HOTCoref.en.html

[9] www2.ims.uni-stuttgart.de/cgi-bin/anders/dl?hotcorefEngNonLocalModel

[10] http://www.sfs.uni-tuebingen.de/de/ascl/ressourcen/corpora/tueba-dz.html

4.2 Using Alignments

In order to assess the usefulness of word alignments in co-reference resolution, we ran IMS-HotCoref on both the German original text and its English translation. Then, we had a closer look at the sentences in which having access to the translation of a pronoun presumably helps its resolution. In Table 2, we show how the co-reference resolver performs for each of the German translations being highlighted in Table 1(a). It can be seen that in 3 out of 7 cases, the word alignment would indeed improve the resolution.

German	English	IMS-HotCoref	correct?
er	Gregor's father	der	NO
seiner	Gregor's	der Zug	NO
sie *(sg)*	his sister	Schwester	YES
ihre	his sister's	Schwester	YES
sie *(pl)*	the food	Herren	NO
	the pain	Schmerzen	YES
	Gregor's mother and father	Eltern	YES

Table 2: Results of the German co-reference resolver.

For the opposite direction, where the German original text is assumed to help resolve coreferential ambiguities, Table 3 contains the results of the co-reference resolver for all crucial occurrences of the pronoun "he", as highlighted in Table 1(b).

English	German	IMS-HotCoref	correct?
he	Gregor	one of the trainees	NO
	Gregor	Gregor	YES
	Gregor	Gregor	YES
	Gregor	(one after) another	NO
	Gregor	Gregor	YES
	Gregor	Gregor	YES
	Gregor	father	NO
	Gregor	Gregor	YES
	Gregor	Gregor	YES
	Gregor	Gregor	YES
	der Vater	Gregor	NO
	der Vater	Gregor	NO

Table 3: Results of the English co-reference resolver.

And even here, we find evidence in 5 of 12 cases that the alignment to the original language would help co-reference resolution. Thereof, the latter two cases are particularly challenging. In fact, we have already introduced them as Examples (1) and (2) above. For Example (1), Figure 2 (a) shows the proposed co-reference annotations by the DE and EN co-reference resolver on the right hand side while gold annotations are shown on the left hand side.

The German sentence is much easier to process for an automatic system, as it contains fewer pronouns and many identical definite descriptions, and so unsurprisingly, the output of the German tool is correct. The English system, however, wrongly links the second pronoun *he* (marked with a box) to Gregor (=the first *he*), as there are two potential antecedents in the sentence that both agree in number and gender with the anaphor.
If we have word alignments, as shown in Figure 1, we can see that the second *he* is aligned with *der Vater* (again, marked with a box), and therefore we now know that the tool's assignment was wrong.

For Example (2), the word alignment again helps predict the right co-reference links. The output of the tool and the right gold annotation is shown in Figure 2 (b). The English co-reference resolver wrongly puts the second *he* (marked by a box) in the co-reference chain describing Gregor, but the word alignment (not shown for Example (2)) tells us that this is not the case: it actually refers to the father.

We also experimented with a second EN co-reference resolver, the Stanford co-reference system as part of the Stanford Core NLP tools[11], but the results were similar to the IMS HotCoref system. When comparing two system outputs or gold annotations with the output predicted by the system, the ICARUS Coreference Explorer[12] (Gärtner et al., 2014) is a useful tool to browse and search co-reference-annotated data. It can display co-reference annotations as a tree, as an entity grid, or in a standard text-based display mode. Particularly useful in our case is the fact that the tool can compare two different annotations on the same document. In the differential view, the tool analyses discrepancies between the predicted and the gold annotation (or two predicted annotations, respectively) and marks different types of errors with different colors. Figure 3 shows an exemplary differential view between the Stanford and the HotCoref system for Kafka's *"Verwandlung"*.

[11] nlp.stanford.edu/software/corenlp.shtml

[12] www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/icarus.html

	gold co-reference annotations	output of IMS-HotCoref
EN	And so **he** ran up to **his father**, stopped when **his father** stopped, scurried forwards again when **he** moved, even slightly.	And so **he** ran up to **his father**, stopped when **his father** stopped, scurried forwards again when **he** moved, even slightly.
DE	Und so lief **er** vor **dem Vater** her, stockte, wenn **der Vater** stehen blieb, und eilte schon wieder vorwärts, wenn sich **der Vater** nur rührte.	Und so lief **er** vor **dem Vater** her, stockte, wenn **der Vater** stehen blieb, und eilte schon wieder vorwärts, wenn sich **der Vater** nur rührte.

(a) Example (1)

	gold co-reference annotations	output of IMS-HotCoref
EN	**He** had thought that nothing at all remained from **his father's** business, at least **he** had never told **him** anything different, and **Gregor** had never asked **him** about it anyway.	**He** had thought that nothing at all remained from **his father's** business, at least **he** had never told **him** anything different, and **Gregor** had never asked **him** about it anyway.
DE	**Er** war der Meinung gewesen , daß **dem Vater** von jenem Geschäft her nicht das Geringste übriggeblieben war , zumindest hatte **ihm der Vater** nichts Gegenteiliges gesagt , und **Gregor** allerdings hatte **ihn** auch nicht darum gefragt .	**Er** war der Meinung gewesen , daß **dem Vater** von jenem Geschäft her nicht das Geringste übriggeblieben war , zumindest hatte **ihm der Vater** nichts Gegenteiliges gesagt , und **Gregor** allerdings hatte **ihn** auch nicht darum gefragt .

(b) Example (2)

Figure 2: Illustration of gold co-reference annotations and tool outputs for Examples (1)+(2).

5 Related Work

As mentioned earlier, the "annotation projection" paradigm was first described by Yarowsky et al. (2001) in order to improve POS-tagging. However, it has proven useful for a number of other applications, e.g. for multilingual co-reference resolution. Most approaches aim at projecting co-references which are available for one (usually well-resourced) language to another (less-resourced) language for which no tools or not even annotated training data are available. The degree of automatic processing ranges from using manually annotated co-references and hand-crafted translations (e.g. Harabagiu and Maiorano (2000)) to automatically obtained word alignments and combined with a manual post-editing of the obtained co-references (e.g. (Postolache et al., 2006)). Finally, some approaches make use of automatic word alignment, but instead of manual post-editing, they access the quality of the projection through training an own co-reference resolver for the under-resourced language based on the projected data (e.g. de Souza and Orăsan (2011) and Rahman and Ng (2012)). Zhekova et al. (2014) were to our knowledge the first ones who projected co-reference annotations in the literature domain, in contrast to general language texts. In lack of a reasonable amount of parallel training data to train an automatic word alignment of the language pair Russian to German, they developed an alignment tool which facilitates manual alignment. In contrast to previous works, they used not only one translation but different German translations of a Russian novel. Mikhaylova (2014) applied automatic word alignment to the same Russian novel and its German translations, trained on the novel itself. In order to improve word alignment performance on such a comparably small training corpus, she made use of simple generalisation techniques (e.g. lemmatisation). Most previous works on multilingual co-reference resolution focus on using co-referential annotations in one language to obtain the same kind of annotations in another language. In our work, we want to improve co-reference resolution in one (sufficiently well resourced) language by using translations from or into another language. This underlying concept has already been described by Harabagiu and Maiorano (2000). However, they rely on manual projections instead of automatic word alignment and moreover, they apply their approach to general language text and not to literature.

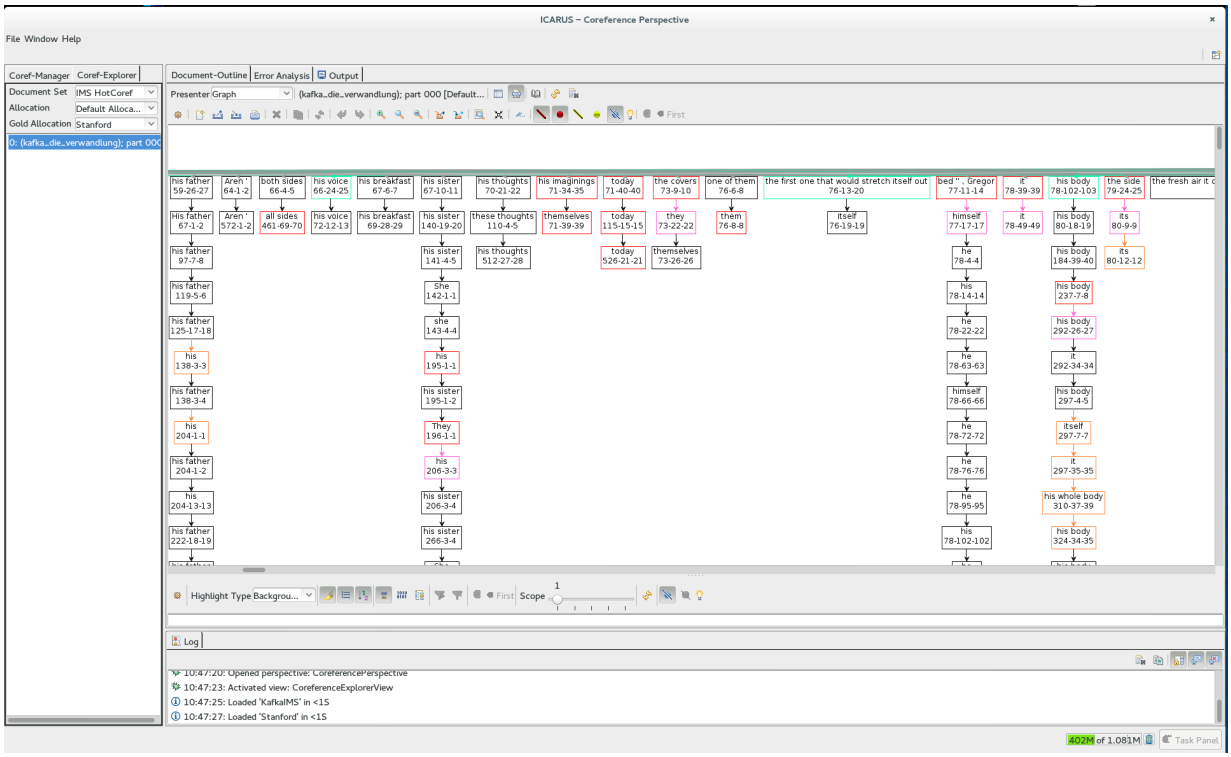

Figure 3: The ICARUS differential error analysis explorer illustrates the differences between the Stanford and the IMS HotCoref system using colour markup.

6 Conclusion and Future Work

We presented a pilot study in which we show how computational linguistic tools and resources can help to improve the identification of character/persona references in literary texts. Our test case is fairly controlled, which enables us to assess the modular components on their own. Based on a manual gold standard word alignment of Franz Kafka's *"Verwandlung"* we show numerous examples for which the translation of a pronominal referent can help to resolve coreferential ambiguites which otherwise may not be resolved accordingly. Having shown that, we will in the near future focus on substituting the manual alignment with an automatic word alignment approach. Due to the limited training data, we will examine different possibilities to improve automatic word alignment of literature text. As German is a morphologically rich language, lemmatisation should definitely be considered. Moreover, the training text for automatic word alignment might be enriched with general language data. In order to enhance the positive matching of personal pronouns to names used in the underlying novel, we will use a POS-tagger to identify proper nouns and then either replace them with names used in the literature or leave them underspecified. The manual gold standard alignment we presented for *"Verwandlung"* is useful in at least two respects for future works: on the one hand it serves us as an upper bound for annotation projection beyond standard co-reference resolution (e.g. distinguishing canonical stative present tense usage and historical/scenic present tense usage), on the other hand we can use it to approximate the quality of different automatic word alignment approaches on literary texts. In the future, we will adopt our automatic co-reference resolver to use the word alignment of pronominal referents, which should lead to an improved performance. While our pilot study is on literary texts, word alignments can be used for co-reference resolution even for texts from other genres, as long as parallel translations are available.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grants for the projects D2 and A6 of the SFB 732.

References

David Bamman, Ted Underwood, and Noah A. Smith. 2014. A Bayesian Mixed Effects Model of Literary Character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 370–379, Baltimore, Maryland.

Anders Björkelund and Jonas Kuhn. 2014. Learning Structured Perceptrons for Coreference Resolution with Latent Antecedents and Non-local Features. In *ACL'14: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland.

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 Verlag, Tübingen, Germany.

José Guilherme Camargo de Souza and Constantin Orăsan. 2011. Can Projected Chains in Parallel Corpora Help Coreference Resolution? In *Anaphora Processing and Applications*, pages 59–69. Springer.

Markus Gärtner, Gregor Thiele, Wolfgang Seeker, Anders Björkelund, and Jonas Kuhn. 2014. ICARUS – An Extensible Graphical Search Tool for Dependency Treebanks. In *ACL'13: Proceedings of the 51nd Annual Meeting of the Association for Computational Linguistics: Systems Demonstrations.*, Sofia, Bulgaria.

Sanda M Harabagiu and Steven J Maiorano. 2000. Multilingual Coreference Resolution. In *Proceedings of the 6th Conference on Applied Natural Language Processing*, pages 142–149.

Fotis Jannidis, Markus Krug, Isabella Reger, Martin Toepfer, Lukas Weimer, and Frank Puppe. 2015. Automatische Erkennung von Figuren in deutschsprachigen Romanen. Conference Presentation at "Digital Humanities im deutschsprachigen Raum".

Jonas Kuhn and Nils Reiter. 2015, to appear. A Plea for a Method-Driven Agenda in the Digital Humanities. In *Proceedings of Digital Humanities 2015: Global Digital Humanities*, Sydney.

Willard McCarty. 2005. *Humanities Computing*. Palgrave Macmillan.

Alena Mikhaylova. 2014. Koreferenzresolution in mehreren Sprachen. Master's thesis, Ludwig Maximilians Universität München.

Vincent Ng. 2010. Supervised Noun Phrase Coreference Research: The First Fifteen Years. In *ACL'10: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1396–1411.

Oana Postolache, Dan Cristea, and Constantin Orasan. 2006. Transferring Coreference Chains Through Word Alignment. In *LREC'06: Proceedings of the 5th International Conference on Language Ressources and Evaluation*.

Sameer Pradhan, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, and Nianwen Xue. 2011. CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes. In *CoNLL'11: Proceedings of the 15th Conference on Computational Natural Language Learning: Shared Task*, pages 1–27.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing (EMNLP) and the Conference on Computational Natural Language Learning (CoNLL): Shared Task*, pages 1–40.

Altaf Rahman and Vincent Ng. 2012. Translation-based Projection for Multilingual Coreference Resolution. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 720–730.

Marta Recasens, Lluís Màrquez, Emili Sapena, M Antònia Martí, Mariona Taulé, Véronique Hoste, Massimo Poesio, and Yannick Versley. 2010. Semeval-2010 Task 1: Coreference Resolution in Multiple Languages. In *Semeval'10: Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 1–8. Association for Computational Linguistics.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing Multilingual Text Analysis Tools via Robust Projection across Aligned Corpora. In *HLT'01: Proceedings of the 1st International Conference on Human Language Technology research*, pages 1–8. Association for Computational Linguistics.

Desislava Zhekova, Robert Zangenfeind, Alena Mikhaylova, and Tetiana Nikolaienko. 2014. Alignment of Multiple Translations for Linguistic Analysis. In *Proceedings of the The 3rd Annual International Conference on Language, Literature and Linguistics (L3)*, pages 9–10.