Data-Driven Detection of General Chiasmi
Using Lexical and Semantic Features

Felix Schneider[1] and **Phillip Brandes**[2] and **Björn Barz**[1] and **Sophie Marshall**[2] and **Joachim Denzler**[1]

[1] Computer Vision Group
[2] Institut für Germanistische Literaturwissenschaft
Friedrich Schiller University Jena
Jena, Germany
`firstname.lastname@uni-jena.de`

Abstract

Automatic detection of stylistic devices is an important tool for literary studies, e.g., for stylometric analysis or argument mining. A particularly striking device is the rhetorical figure called chiasmus, which involves the inversion of semantically or syntactically related words. Existing works focus on a special case of chiasmi that involve identical words in an A B B A pattern, so-called antimetaboles. In contrast, we propose an approach targeting the more general and challenging case A B B' A', where the words A, A' and B, B' constituting the chiasmus do not need to be identical but just related in meaning. To this end, we generalize the established candidate phrase mining strategy from antimetaboles to general chiasmi and propose novel features based on word embeddings and lemmata for capturing both semantic and syntactic information. These features serve as input for a logistic regression classifier, which learns to distinguish between rhetorical chiasmi and coincidental chiastic word orders without special meaning. We evaluate our approach on two datasets consisting of classical German dramas, four texts with annotated chiasmi and 500 unannotated texts. Compared to previous methods for chiasmus detection, our novel features improve the average precision from 17% to 28% and the precision among the top 100 results from 13% to 35%.

1 Introduction

Knowledge about the use and distribution of stylistic devices, such as the chiasmus, can be used for stylometric purposes such as attribution to a certain author or genre (Pasanek and Sculley, 2008). Current computational methods for stylometry focus on analyzing word distributions and do not take more complex stylistic devices into account (Burrows, 2002; Jannidis, 2014; Dimpel et al., 2016). However, since styles may also differ in their use of stylistic devices, this is an important factor to consider. Different stylistic devices can also be an indicator of the structure of arguments (Mitrović et al., 2017; Lawrence et al., 2017). Automatic detection of stylistic devices is an important prerequisite for enabling such analyses at scale.

The chiasmus is a stylistic device that is found in texts since the antiquity. Chiasmus is defined as an inversion of semantically or syntactically related words, phrases, or sentences in an *A B B' A'* pattern (Fauser, 1994). It can be used, for example, to emphasize contrasts. One example for it is:

> **Eng** ist die **Welt**,
> und das **Gehirn** ist **weit**
> (**Narrow** is the **world**,
> and the **brain** is **wide**)
> *Wallensteins Tod* (Schiller, 1799)

The semantically related words are narrow and wide, as well as world and brain.

Existing works for automatic chiasmus detection focus on a special case called antimetabole, which consists of an inversion of identical lemmata (i.e. *A=A', B=B'*). An example for this is:

> **Un** pour **tous**, **tous** pour **un**
> (**One** for **all**, **all** for **one**)
> *Les Trois Mousquetaires* (Dumas, 1844)

Here, the inverted words are "all" and "one".

In contrast to this, we are interested in finding general chiasmi, where the lemmata can be different. In this work we present the first machine learning based method for finding general chiasmi. Existing methods find candidates by searching for inversions of lemmata. However, such an approach cannot find general chiasmi like the one by Schiller mentioned above, which does not consist of an inversion of lemmata but of semantic concepts.

We propose searching for chiasmus candidates by finding inversions of part-of-speech *(PoS)* tags in an *A B B' A'* pattern instead of inversions of lemmata. This requires more sophisticated filtering of the candidates, which include much more

false positives without rhetorical intention. To this end, we use a linear classifier and propose a novel set of lexical and semantic features. Our semantic features are based on word embeddings (Mikolov et al., 2013; Bojanowski et al., 2017) to include information about the semantic relationship between the words constituting the chiastic pattern.

Compared with existing methods, our novel features improve the average precision from 17% to 28% in an experiment on four fully annotated dramas by Friedrich Schiller. Further experiments on 493 unannotated texts confirm the usefulness of our automatic general chiasmus detector, which improves the precision among the top 100 rated chiasmus candidates from 13% to 35%.

2 Related Work

Only few works address the detection of chiasmi, and the majority of those focus on antimetaboles and not the general case of chiasmus. While it is easy to capture all occurences of antimetabole in a text by just gathering all cross-wise repeating lemmata, this approach also yields a disproportionately high number of false positives (Gawryjolek, 2009). Thus the main challenge is to remove the false positives from the set of chiasmus candidates.

The approach of Dubremetz and Nivre (2015) introduces a filtering step for the chiasmus candidates. Here, the authors use hand-crafted features based on punctuation and word repetitions together with a maually tuned classifier. They expanded on their work with more features and machine learning (Dubremetz and Nivre, 2017, 2018), and trained a Kernel SVM and a linear classifier to detect antimetaboles, which constitutes the current state of the art. In contrast to their method, we search for candidates not based on lemmata but on PoS tags. We furthermore include novel features incorporating lemma information and semantic information about the conceptual similarity of the words constituting the chiasmus.

A first approach for general chiasmus detection was presented by Java (2015) by searching for inversions in syntax trees. However, this misses many true positives and they do not filter their results, leading to numerous false positives as well. In our work, we use inversions of PoS tags with an additional learning-based filtering step to find more candidates and discard more false positives.

3 Method

Our method consists of two steps; first the detection of suitable chiasmus candidates, second filtering these candidates using machine learning.

Candidate Extraction First we extract candidates from the text. In contrast to Dubremetz and Nivre (2018), we do not search for inversions of lemmata but of PoS tags. Like them, we limit the candidate phrase length to 30 tokens. With this stragety we find candidates for general chiasmi as well as antimetaboles. However, the candidates do not include lemma information a priori, in contrast to lemma-based candidate mining. We also find more false positive candidates, which raises the performance requirements for the subsequent filtering step. Thus, in addition to the Dubremetz features we include two new types of features in our work.

Dubremetz Features The features proposed by Dubremetz and Nivre (2018) comprise a combination of various elements like the number of identical words between the supporting words, the number of hard and soft punctiation marks at different positions, grammatical dependencies of the words, and n-gram-repetitions. For a full list of these features please refer to their work.

Lexical Features Since our generalized search procedure based on PoS tags does not consider lemma information anymore, we need to provide this information explicitly for the filter. Therefore, we include a novel set of *lexical features* to still capture this information. For all six pairs *AB, AB', AA', BB', BA', B'A'* of supporting tokens, we add a binary feature equaling 1 if the lemmata of the two tokens are identical and 0 otherwise.

Embedding Features To distinguish true chiasmi from random PoS tag inversions, the semantic relationship between the supporting tokens needs to be considered. We use word embeddings to integrate this semantic information, since they represent words as vectors in a space, where the distance between two vectors signifies their semantic relation (Bojanowski et al., 2017). For each pair of supporting tokens, we add an *embedding feature* equaling the cosine similarity (Salton et al., 1975) of the word embeddings of the two tokens.

Candidate Ranking We train a logistic regression classifier using the aforementioned features to distinguish true chiasmi from random PoS tag in-

	Schiller dramas			Dubremetz data
	Antimetaboles	Chiasmi	Combined	Antimetaboles
(baseline) D	0.21 ± 0.18	0.15 ± 0.07	0.17 ± 0.04	$\mathbf{0.73 \pm 0.24}$
L	0.07 ± 0.08	0.01 ± 0.00	0.02 ± 0.00	0.06 ± 0.01
E	0.10 ± 0.07	0.06 ± 0.03	0.09 ± 0.07	0.25 ± 0.13
LE	0.11 ± 0.06	0.05 ± 0.02	0.08 ± 0.04	0.25 ± 0.13
DL	$\mathbf{0.49 \pm 0.32}$	0.14 ± 0.09	0.22 ± 0.07	0.72 ± 0.24
DE	0.48 ± 0.30	$\mathbf{0.23 \pm 0.13}$	$\mathbf{0.28 \pm 0.01}$	$\mathbf{0.73 \pm 0.24}$
DLE	0.48 ± 0.30	0.19 ± 0.09	$\mathbf{0.28 \pm 0.08}$	$\mathbf{0.73 \pm 0.24}$

Table 1: Average precision for different feature combinations. D=Dubremetz features, L=lexical features, E=embedding features.

versions. For ranking candidates, we use the score obtained from the classifier's decision function.

4 Experiments

We perform two types of experiments[1]. First we evaluate the average precision (AP) of different feature combinations using 5-fold cross-validation on an annotated dataset to conduct an ablation study regarding the different feature types. AP describes the area under the precision-recall curve and is common in information retrieval. Additionally, we conduct this experiment on antimetabole candidates used by Dubremetz and Nivre (2018) for a comparison with their approach on their task.

In the second experiment, we evaluate how well our model generalizes to texts from different authors not included in the training data. To this end, we extract PoS tag inversions from the GerDraCor corpus (Fischer et al., 2019) and retrieve the top-scoring 100 candidates using our pre-trained classifier. These results are evaluated manually to assess the precision among the top 100 results.

For PoS tagging, dependencies, and word embeddings we use the *de_core_news_lg* model from spaCy (Honnibal et al., 2020). For the English data from Dubremetz and Nivre (2018), we use the spaCy *en_core_web_lg* model. To create syntax trees, we used CoreNLP (Manning et al., 2014).

Feature Combinations Our dataset[2] consists of four annotated texts by Friedrich Schiller, "Die Piccolomini", "Wallensteins Lager", "Wallensteins Tod", and "Wilhelm Tell". We annotated the whole texts, finding 45 general chiasmi and 9 antimetaboles. To obtain negative examples, we randomly chose 4000 PoS tag inversions from the 1,006,487 inversions in these texts.

Table 1 shows the AP of 5-fold cross-validation conducted with both the antimetaboles, chiasmi, and a combination of both as positive examples. It can be seen that all our additional features improve over the Dubremetz features alone, which are designed for finding antimetaboles in lemma inversions. Adding our novel features improves the detection of general chiasmi in PoS tag inversions. The combination of Dubremetz and lexical features improves the detection especially for antimetaboles. This shows that the lexical features include lemma information that is not present a priori with PoS tag inversions, including the information in the classifier, not in the candidate choice. When using only the novel features, the combination of both improves the results.

When we only search for chiasmi and exclude the antimetabole examples from the data, the combination of Dubremetz and embedding features yields the best results. Both the combinations of either embedding or all new features with the Dubremetz features improve over the baseline. However, in contrast to the antimetaboles-only experiment, the embedding features bring a stronger improvement than the rest. The combination with lexical features even slightly decreases the AP. This can also be seen when using only the novel features.

For the experiment with antimetaboles and chiasmi both considered as positive, the combination of the baseline and both novel feature sets or of the baseline and embedding features improves AP the most. The baseline features with the embedding features yield a better performance than the combination with the lexical features, both improving over the baseline. Lexical features are more important for Antimetabole detection, while embedding features help with the general chiasmus detection.

In addition, we tested our approach on English

---

[1]Code: https://git.io/DetectChiasmus
[2]Data: https://git.io/ChiasmusData

	Schiller dramas			Rest of the GerDraCor corpus		
	D method	D features	DLE	D method	D features	DLE
antimetaboles	6	2	**8**	7	2	**25**
chiasmi	5	5	**10**	6	9	**10**
combined	11	7	**18**	13	11	**35**

Table 2: Number of correct examples among the top 100 ranked ones in unseen texts for the Dubremetz method baseline, the PoS inversions with Dubremetz features and the Dubremetz+lexical+embedding (DLE) features.

data used by Dubremetz and Nivre (2018). This data consist of annotated candidates based on lemmata instead of PoS tags, so only the antimetabole detection can be tested. Since the novel features do not introduce additional information for candidates based on lemma inversions, the new features do not improve the detection.

We also tested the approach by Java (2015) on the *Combined* candidates, yielding a precision of 0.2% and a recall of 5.6%, compared to 10% and 69% with our *DLE* approach. This shows that in their approach most positive examples are discarded, while due to the missing filtering step many false positives occur.

Chiasmi in Unseen Texts   Since we have shown that both the novel lexical and embedding features and their combination can help with the task of antimetabole and general chiasmus detection, we evaluate the generalization performance of our chiasmus classifier trained on the four annotated Schiller dramas to other texts. The first set of texts comprises seven other dramas by Friedrich Schiller, which contain a total of 2,822,313 PoS tag inversions and 80,606 lemma inversions. To see how well our method generalizes to different authors, we tested it on the remaining 493 documents from GerDraCor. For evaluation, we annotated the top-scoring 100 candidates manually.

We compare the baseline Dubremetz method based on lemma inversions with the full combination of the Dubremetz, lexical, and embedding features based on PoS tag inversions. Table 2 shows the results of this experiment. We can see that the precision among the top 100 rated candidates improved from 11% to 18% for the Schiller dramas and from 13% to 35% for the rest of the dataset. Interestingly, even the lemma-based baseline method finds some chiasmi. This is mainly due to some words with a common lemma, for example, *sich* can take the forms of *mich* and *dich* in the sentence and thus carry a chiasmus. We furthermore observe that the Dubremetz features together with the PoS

tag inversions yield less positive examples than with lemma inversions. Since the number of PoS tag inversions is vastly higher than the number of lemma inversions, this is a much harder problem when the same feature set is used. The results show that our additional novel features are not only able to bridge this gap but also find substantially more chiasmi than the baseline.

The number of chiasmi compared to the number of PoS tag inversions in a text is extremely small; in our annotated dataset from the previous experiment, only 0.005% of the inversions were chiasmi or antimetaboles. Thus, our approach is suited to find these stylistic devices.

Since our candidate mining approach also allows all four PoS tags to be identical (i.e., A A A A instead of A B B A), we found that many of the false positives among the top 100 displayed another stylistic device: the parallelism.

5   Conclusion

We present a novel machine learning approach for a problem seldom addressed: the detection of general chiasmi. First, we extend the candidate mining scheme used by methods for the special case of antimetabole detection to general chiasmi by searching for inversions in PoS tags instead of lemmata. We then propose two new sets of *lexical* and *embedding* features, which encode information about the lemmata and word semantics.

Our experiments show that our approach outperforms the state of the art. Especially the semantic embedding features proved useful for detecting chiasmi and antimetaboles. Our approach still falls short of human performance, which is expected, as the number of chiasmi in a text is extremely low compared to the number of potential chiasmus candidates.

An interesting next research step is to test the model on different languages without retraining. Also, applying the method to search for parallelisms using an *A B A' B'* pattern seems promising.

References

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

John Burrows. 2002. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, 17(3):267–287.

Friedrich Michael Dimpel, Victoria Gutsche, and René Wundke. 2016. Problematische Autorschaft. Zu ersten Anwendungsversuchen stilometrischer Methoden im Rahmen der Edition der Deutschen Politischen Schriften Zincgrefs. *Magazin für digitale Editionswissenschaften*, 2:23–32.

Marie Dubremetz and Joakim Nivre. 2015. Rhetorical figure detection: the case of chiasmus. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 23–31, Denver, Colorado, USA. Association for Computational Linguistics.

Marie Dubremetz and Joakim Nivre. 2017. Machine learning for rhetorical figure detection: More chiasmus with less annotation. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 37–45, Gothenburg, Sweden. Association for Computational Linguistics.

Marie Dubremetz and Joakim Nivre. 2018. Rhetorical figure detection: Chiasmus, epanaphora, epiphora. *Frontiers in Digital Humanities*, 5:10.

Alexandre Dumas. 1844. Les trois mousquetaires. *Le Siècle*.

Markus Fauser. 1994. *[Art.] Chiasmus, in Historisches Wörterbuch der Rhetorik*, volume 2, pages 171–173.

Frank Fischer, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke. 2019. Programmable corpora: Introducing dracor, an infrastructure for the research on european drama.

Jakub Jan Gawryjolek. 2009. Automated annotation and visualization of rhetorical figures.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Fotis Jannidis. 2014. *Der Autor ganz nah – Autorstil in Stilistik und Stilometrie, in Theorien und Praktiken der Autorschaft*, pages 169–195. de Gruyter, Berlin.

James Java. 2015. *Characterization of Prose by Rhetorical Structure for Machine Learning Classification*. Ph.D. thesis, Nova Southeastern University.

John Lawrence, Jacky Visser, and Chris Reed. 2017. Harnessing rhetorical figures for argument mining.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60. The Association for Computer Linguistics.

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Jelena Mitrović, Cliff O'Reilly, Miljana Mladenović, and Siegfried Handschuh. 2017. Ontological representations of rhetorical figures for argument mining.

Brad Pasanek and D. Sculley. 2008. Mining millions of metaphors. *Literary and Linguistic Computing*, 23(3):345–360.

G. Salton, A. Wong, and C. S. Yang. 1975. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620.

Friedrich Schiller. 1799. *Wallensteins Tod*.