Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)

MONAPipe: Modes of Narration and Attribution Pipeline for German Computational Literary Studies and Language Analysis in spaCy

Tillmann Dönicke[1], **Florian Barth**[1], **Hanna Varachkina**[2], **Caroline Sporleder**[1]
[1]Göttingen Centre for Digital Humanities, [2]Department of German Philology
University of Göttingen
{tillmann.doenicke@,florian.barth@,hanna.varachkina@stud.,caroline.sporleder@cs.}
uni-goettingen.de

Abstract

MONAPipe is a collection of pipeline components for the open-source Python library spaCy. The components perform a broad range of morphological, syntactic, semantic and pragmatic analyses for German texts and are mostly developed specifically for the literary domain. MONAPipe[1] combines implementations from various separate resources with new ones in one place, constituting a convenient tool for computational linguistics and literary studies.

1 Introduction

When working with text using computational methods, one has to follow a series of standard processing steps that are often combined into a pipeline for efficiency. Although the choice of the existing pipelines is large, there are only a view which focus on the literary domain (e.g. BookNLP[2]), from which to our knowledge none is usable for German. It is well known that literary texts have properties which pose challenges for natural language processing (NLP), such as non-standard orthography, long and complex sentences, long-distance coherence and possibly multi-layered narrative levels to name but a few. MONAPipe presents an extension of the spaCy pipeline which provides basic NLP components based on high-performance German models. Our custom pipeline consists of numerous components that can be divided into six categories: preprocessing, morphosyntactic analysis, semantic analysis, speech and coreference resolution, feature extraction and discourse units, narration and attribution. Some components are domain-independent (e.g. tense tagging), while others are specifically created to analyze fiction and literary concepts (e.g. literary comment).

2 SpaCy

MONAPipe is developed for spaCy (v2.3[3]), which is an open-source software library for crosslinguistic natural language processing in Python. An input text is converted to a document object and then consecutively piped through a series of (built-in or custom) pipeline components which can be arranged by the user. The components enrich the document with information that can be attributed to the document, its tokens or spans (of tokens).

3 Pipeline Components

The main contribution of MONAPipe are new pipeline components for spaCy. Some of the components were developed from scratch whereas others are reimplementations or wrappers of existing tools. Table 1 provides an overview of the currently usable MONAPipe components, which we will discuss in the following.

3.1 Preprocessing

If one wants to process a text which is not already tokenized, one can use spaCy's built-in **Tokenizer**. Built-in follow-up components are a part-of-speech (POS) **Tagger** which assigns both German (Smith, 2003b, p. 12 f.) and universal (de Marneffe et al., 2021, p. 261) POS tags, a dictionary-based **Lemmatizer**, and a named entity recognizer (**NER**) that recognizes persons, locations, organizations and miscellaneous entities (Nothman et al., 2013).

Older texts commonly exhibit non-standard orthography, which can cause problems in follow-up language processing. We therefore provide a **Normalizer** that replaces every out-of-vocabulary word by its most frequent normalized form in the German Text Archive[4] (DTA), a collection of 4,160

---

[1] https://gitlab.gwdg.de/mona/pipy-public
[2] https://github.com/booknlp/booknlp
[3] https://v2.spacy.io/usage
[4] https://www.deutschestextarchiv.de/download

Component	Type	Main Reference(s)
Preprocessing		
Tokenizer	B	spaCy
Tagger	B	spaCy
Lemmatizer	B	spaCy
NER	B	spaCy
Normalizer	I	this paper
Morphosyntactic Analysis		
Sentencizer	B/W	spaCy, NLTK
DependencyParser	B/I	spaCy, Dönicke (2020)
Clausizer	I	Dönicke (2020)
Analyzer	I	Altinok (2018), Dönicke (2020)
TenseTagger	I	Dönicke (2020)
Semantic Analysis		
TemponymTagger	R	Strötgen and Gertz (2010, 2015)
GermanetTagger	I	Hamp and Feldweg (1997), this paper
EmotionsTagger	I	Mohammad and Turney (2010), this paper
Speech and Coreference Resolution		
SpeechTagger	W/I	Brunner et al. (2020)
SpeakerExtractor	I	this paper
Coref	R	Krug et al. (2015), this paper
Feature Extraction and Discourse Units		
FeatureExtractor	I	Dönicke (2021), this paper
DiscourseSegmenter	I	Dönicke (2021)
Modes of Narration and Attribution		
EventTagger	W	Vauth et al. (2021)
AnnotationReader	I	this paper
CommentTagger	I	Weimer et al. (to appear)
GenTagger	I	Gödeke et al. (to appear)
EntityLinker	I	Barth et al. (2022)
AttributionTagger	I	Dönicke et al. (2022)

Table 1: Overview of MONAPipe components with origin (B: built-in in spaCy, R/I: re-/implemented by MONAPipe authors, W: wrapper for external tool). See text for more information.

texts (480M tokens) from 1600–1900. This approach correctly normalizes over 99.9% of tokens and types in the DTA. Original forms and character positions of tokens are preserved as attributes.

3.2 Morphosyntactic Analysis

The **Sentencizer** (i.e. sentence splitter) adds sentence spans to the document. Currently, one can use either a sentencizer from spaCy or NLTK[5].

The **DependencyParser** adds a dependency tree to each sentence. Which dependency scheme is used depends on the spaCy model, where the German model provided by spaCy produces trees in the TIGER scheme (Smith, 2003b). An alternative to TIGER is the Universal Dependencies (UD) scheme (de Marneffe et al., 2021). While some of our components function in either scheme, most do either require UD parses or function significantly better with them. We therefore recommend using MONAPipe with a UD-based spaCy model and use the model provided by Dönicke (2020).

Dönicke (2020) also provides a **Clausizer** that splits UD trees into clauses and adds clause spans to the document and its sentences, a morphological **Analyzer** based on DEMorphy (Altinok, 2018), and a **TenseTagger** that extracts grammatical features (finiteness, tense, mood, voice) and modal verbs like *müssen* 'must' from a clause's (potentially composite) verb. Dönicke (2020) reports accuracies of 93% for tense, 79% for mood, 94% for voice and 80% for modal verbs in the literary domain. We integrate these components into MONAPipe and make a small change in the handling of modal verbs, so that semi-modal verbs like *pflegen (zu)* 'use (to)' are properly recognized as modal verbs in according contexts (and not always treated as full verbs).[6]

3.3 Semantic Analysis

The **TemponymTagger** extracts and normalizes temporal expressions from a document. The component is a reimplementation of the HeidelTime[7] system (Strötgen and Gertz, 2010, 2015) and uses its resource files for German.

The **GermanetTagger** assigns Levin (1995)'s semantic categories to verbs and clauses (in case the verb is the root) and Hundsnurscher and Splett (1982)'s categories to adjectives, which are extracted from GermaNet (Hamp and Feldweg, 1997). Using the lemmas of verbs and adjectives, possible word senses (synsets) are identified and disambiguated using the synsets from the token's context.

The **EmotionsTagger** adds scores for sentiment (positive, negative) and basic emotions as defined by Ekman (1992) (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) from the NRC Word-Emotion Association Lexicon[8] (Mohammad and Turney, 2010, 2013) to tokens.

3.4 Speech and Coreference Resolution

The **SpeechTagger** assigns scores for speech[9] types to tokens and clauses. We provide two im-

---

[5] https://www.nltk.org/

[6] For example, the semi-modal verb *use* is a full verb in *John used a lighter* and a modal verb in *John used to smoke*. We distinguish the two cases as follows: A semi-modal verb is a modal verb if it is accompanied by a subordinate verb and it is a full verb otherwise.

[7] https://github.com/HeidelTime/heideltime

[8] https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[9] We use the term "speech" for any speech, thought or writing representation in texts (cf. Brunner et al., 2020).

plementations of this component. The first one uses Brunner et al. (2020)'s Redewiedergabe tagger to predict token-wise scores for direct, indirect, free indirect and reported speech. It achieves 85% F1 for direct, 76% F1 for indirect, 60% F1 for reported and 59% F1 for free indirect speech for texts from the 19th to the 20th century (both fiction and non-fiction). The second, faster implementation simply labels tokens within quotation marks as direct speech (ignoring other speech types) and achieves 70% F1 on the same test set (since direct speech is not always marked by quotation marks in older texts). The clause-wise scores are calculated from the product of the token-wise scores.

The **SpeakerExtractor** then adds direct speech spans to the document and tries to identify speaker and addressee for each span. We use a small set of rules to identify a preceding/succeeding verbum dicendi first and then select its subject as speaker and object as addressee.

The development of our **Coref** (coreference) component was driven by the aim to resolve anaphoric pronouns and coreferent nominal phrases (NPs) in a text. We therefore consider all NPs as mentions (including pronouns[10], common NPs and named entities), which contrasts other works. For example, in DROC – a corpus of German novels – (Krug et al., 2018) only mentions of literary characters are annotated, and in ParCorFull – a parallel corpus of news and other domains – (Lapshinova-Koltunski et al., 2018) mentions can be non-nominal and the annotation of a generic NP depends on whether it is a common NP or a pronoun. The corpus with the most similar concept of mentions to ours is GerDraCor-Coref – a corpus of German dramatic texts – (Pagel and Reiter, 2020), although non-nominal mentions are also annotated in part of the corpus.

The Coref component is a UD-based reimplementation of Krug et al. (2015)'s rule-based system which consecutively executes 11 passes to find the antecedent of a mention. Since Krug et al. (2015)'s system was developed for DROC, we made some adjustments to handle a wider variety of NPs (passes 3, 5–7). We use the Extended Open Multilingual Wordnet[11] (Bond and Foster, 2013) to find synonyms in the semantic pass (pass 8) and the results from the SpeechTagger and SpeakerExtractor to resolve pronouns in direct speech (passes 10–11). We store coreference clusters in the same format as NeuralCoref[12], so that one can replace our Coref component by a (currently non-existent) German NeuralCoref model in the future without producing errors in follow-up components.

	Mentions	MUC	B^3	CEAFe	CoNLL
GerDraCor					
HotCoref	–	56.55	14.98	14.84	28.79
DramaCoref	60.00	42.54	19.87	18.97	27.12
full mentions	56.24	43.21	19.78	12.56	25.18
mention heads	70.25	58.20	29.18	15.04	34.14
NP heads	74.36	57.10	31.91	18.18	35.73
gold NP heads	97.03	68.22	39.91	33.97	47.37
DROC					
Schröder et al. (2021)	–	–	–	–	64.72
Krug (2020)	–	87.50	40.40	31.60	53.17
full mentions	38.25	30.67	11.92	3.99	15.53
mention heads	57.04	45.55	24.06	10.88	26.83
NP heads	61.97	50.78	29.60	12.28	30.89
gold NP heads	97.85	68.14	39.42	28.85	45.47
ParCorFull					
Pražák et al. (2021)[13]	–	–	–	–	55.40
full mentions	36.98	24.19	18.76	16.15	19.70
mention heads	41.04	26.68	21.63	18.12	22.14
NP heads	43.21	28.23	23.73	20.63	24.20
gold NP heads	96.99	62.67	68.04	57.58	62.76

Table 2: Coref evaluation on three corpora. The first numeric column shows the F1 for mention identification. MUC, B^3 and CEAFe are F1-based metrics for coreference resolution (cf. Moosavi and Strube, 2016). The CoNLL score is the average of the three.

Despite contrasts to other works, we score our system on GerDraCor, DROC and ParCorFull (see Table 2) using the scorer from Moosavi et al. (2019) to get a rough impression on its performance and to compare it to previous works. We accede to Nedoluzhko et al. (2021) and consider an evaluation on mention heads in a cross-resource scenario as more meaningful than using full mentions, but show scores for full mentions for comparison. For example, mention identification scores 14% higher for mention heads than for full mentions on GerDraCor.[14] Since our system only links NPs, we also show the scores when (heads of) non-nominal mentions are excluded.[15] Our system achieves sim-

---

[10] We exclude indefinite, interrogative and expletive pronouns since they do not have antecedents. Possessive pronouns are de facto exclduded since they usually appear within a larger mention but we do not consider nested mentions.

[11] http://compling.hss.ntu.edu.sg/omw/summx.html

[12] https://github.com/huggingface/neuralcoref

[13] The performance of Pražák et al. (2021)'s system on ParCorFull is listed at https://github.com/ondfa/coref-multiling.

[14] One reason is that mentions in GerDraCor may include succeeding punctuation which is not the case for our mentions.

[15] According to the UD guidelines, we define a mention as nominal if its head has one of these relations: nsubj, obj, iobj, obl, vocative, expl, dislocated, nmod, appos, nummod.

ilar results to those of the recently tested systems HotCoref (Roesiger and Kuhn, 2016) and DramaCoref (Pagel and Reiter, 2021).[16] For DROC and ParCorFull, the F1 for mention identification suffers from a low precision, since we consider much more NPs to be mentions than those in the corpora, and our system performs much lower than the neural systems presented in Krug (2020, p. 173) and Schröder et al. (2021) for DROC[17] and Pražák et al. (2021) for ParCorFull. We therefore also provide the scores for evaluating on gold NPs only: the gold NPs in DROC are linked with a similar performance as those in GerDraCor, and even better in ParCorFull.

3.5 Feature Extraction and Discourse Units

The **FeatureExtractor** combines the information from previous components and some additional information in a (mostly) delexicalized functional grammar (DFG) structure. DFG structures combine rudiments of lexical functional grammar (LFG) and UD grammar and are created for each clause. We take over the basic set-up of Dönicke (2021), who includes grammatical features from the clause, the complex verb, NPs and discourse markers, and add separate levels for adjectives, articles and quantifiers. We further integrate all available semantic information, including GermaNet category and emotion (see Section 3.3), sentiment from SentiWS[18] (Remus et al., 2010), speech type (see Section 3.4) as well as overt quantifier type (using Dönicke et al. (2021)'s categories), and link pronominal anaphora to their antecedents. An example is shown in the appendix.

Dönicke (2021) uses the feature structures for discourse unit segmentation and we also integrate his German model as **DiscourseSegmenter**. The model achieved 92% F1 for German in the DISRPT 2021 Shared Task on Elementary Discourse Unit Segmentation (Zeldes et al., 2021) (4% lower than the best-performing, neural system).

3.6 Narration and Attribution

The **EventTagger** is a wrapper for the event-classification model from Vauth et al. (2021)[19], which classifies clauses into four event types: non-event, stative event, process event and change of state. The model was trained on works of literature and achieves accuracies of 84% for non-event, 75% for stative event, 79% for process event and 56% for change of state. Note that Vauth et al. (2021)'s event types are based on narrative theory (e.g. Schmid, 2014; Prince, 2012) but there are parallels to discourse/situation entity types (also known as clause-level aspect) from linguistic theory (e.g. Vendler, 1957; Smith, 2003a; Friedrich and Palmer, 2014), most importantly the distinction between dynamic and stative events, which is why we consider the EventTagger a useful component for both narratological and linguistic analyses.

MONAPipe further includes components for the automatic identification of narrative modes, which are especially useful for the analysis of fictional literature. The components were developed on the Modes of Narration and Attribution Corpus (MONACO) (Barth et al., 2021), a corpus of fictional texts from 1600 to 1950 which are annotated with narratological information. The annotations in MONACO are saved in a CoNLL-based format and the XML-based output format of the annotation tool CATMA[20]. We provide an **AnnotationReader** that can read CATMA files for the piped document and assigns the annotations to its tokens and clauses. In this way, predictions and annotations (e.g. gold annotations) can be directly accessed at an element of interest.

The term 'narrative mode' itself is a cover term for various stylistic devices that shape the narration of a story. Bonheim (1975) distinguishes four narrative modes: description (depiction of things in motion), report (depiction of things in motion), speech (utterances, thoughts etc. of characters), and comment. In comment, the narration pauses and additional information is provided, e.g. when the narrator interprets what just happened. A text example with all narrative modes is shown in Figure 1. Since report and description usually constitute the most part of a narrative text and speech can be identified by the SpeechTagger, we consider comment to be the most interesting narrative mode to automatically identify in a text.

The annotation guidelines in MONACO follow Chatman (1980) and distinguish three subtypes of comment: interpretation (of story elements), judgment/attitude (towards story elements), and metafictional comment (about the story or the narra-

---

[16] Like Pagel and Reiter (2021), we also randomly selected 80% of the texts in GerDraCor-Coref (1.2.1) as test set but chances are high that our test sets are not identical.

[17] We use the same 18 texts from DROC as test set.

[18] https://github.com/Liebeck/spacy-sentiws

[19] https://github.com/uhh-lt/event-classification

[20] https://catma.de/

[Dr. Johnson was well along in years]$_{\text{DESCRIPTION}}$ [when Boswell explained to him the solipsism of Bishop Berkeley, yet Johnson was still nimble enough to kick a pebble down the path and exclaim,]$_{\text{REPORT}}$ ['thus do I refute him, Sir!']$_{\text{SPEECH}}$ [His was the voice of common sense kicking logic out of the way.]$_{\text{COMMENT}}$

Figure 1: Example text with annotated narrative modes (Bonheim, 1975). Brackets mark annotation spans.

tion itself). The fourth subtype included by Chatman (1980), generalization (i.e. general truths that "reach beyond the world of the fictional work into the real universe", p. 243), is not treated as a subtype of comment in MONACO. Instead, generalization and non-fictionality are treated as separate modes with own subtypes.

Special difficulty when developing text-classification systems for narrative modes is posed by the fact that they can span arbitrarily long text passages and overlap with each other. Since 'passages' in MONACO are defined as sequences of clauses, one can approach the task as multi-class multi-label classification of clauses and address the reconnection of subsequent clauses with the same labels to passages in a postprocessing step.

The statistical **CommentTagger** of MONAPipe (described in Weimer et al. (to appear)) uses the features from Section 3.5. When tested on two held-out texts, the binary model achieves 59% F1, which we consider to be a good state of the art given the difficulties of the task and the literary domain. The multi-class model achieves 36% F1 for interpretation, 28% F1 for attitude and 48% for meta-fictional comment. Taggers for generalizing and non-fictional passages are still in development but MONAPipe also includes the current versions of a rule-based and a statistical **GenTagger** to recognize generalizations (described in Gödeke et al. (to appear)) as well as an **EntityLinker** (described in Barth et al. (2022)), which links named entities to Wikidata[21] entries and determines whether they are fictional or real entities.

MONACO also contains annotations for speaker attribution, i.e. whether the content of a clause is conveyed by a character, the narrator and/or the author of the text. In Dönicke et al. (2022), we trained a neural classifier on MONACO, which we also wrap in a spaCy component. The **AttributionTagger** and the SpeechTagger are indeed somewhat similar, e.g. free indirect speech is typically attributed to a character and the narrator. However, while the task of the SpeechTagger is to identify certain constructions, the AttributionTagger labels the supposed source of information (independently from preselected constructions). In Dönicke et al. (2022), the model achieves 84% accuracy on a held-out test set.

4 Other Features

Automatic saving/loading of intermediate results can be enabled to avoid unnecessary recomputation, which is especially useful for long texts.

We also include functions to 1) calculate inter-annotator agreement in terms of Fleiss's κ, Krippendorff's α and Mathet et al.'s γ after adding annotations to documents, and 2) compare annotations to automatically assigned labels in terms of accuracy, precision, recall and F1 or with a confusion matrix. Agreement and evaluation measures can be executed for tokens and clauses.

In addition, we developed a **CorpusReader** that reads metadata from the source files (TEI-XML) of our literary corpus and provides structured metadata, e.g. GND-identifiers[22] for a work's author, that can be accessed within the pipeline. Furthermore, we enrich existing metadata, e.g. we detect Wikidata entries for a literary work. These metadata is used in MONAPipe components such as the EntityLinker.

5 Conclusion and Future Work

MONAPipe is a custom spaCy pipeline that provides a set of tools for the linguistic and literary analysis of German texts. Many of its components do not have equivalents and present state of the art in the field of computational literary studies or show competitive results compared to the existing tools.

We plan to add further components for natural and narratological language processing as well as new versions of existing components, e.g. taggers for generalization and non-fictionality. The current coreference system is meant to be a make-shift implementation and we want to develop wrappers for other tools in the future. We also plan to upgrade MONAPipe from spaCy v2.x to v3.x.

---

[21] https://www.wikidata.org/wiki/Wikidata:Main_Page

[22] GND: Integrated Authority File, German for "Gemeinsame Normdatei", https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html.

References

Duygu Altinok. 2018. DEMorphy, German language morphological analyzer. arXiv:1803.00902.

Florian Barth, Tillmann Dönicke, Benjamin Gittel, Luisa Gödeke, Anna Mareike Weimer, Anke Holler, Caroline Sporleder, and Hanna Varachkina. 2021. MONACO: Modes of Narration and Attribution Corpus. https://gitlab.gwdg.de/mona/korpus-public.

Florian Barth, Hanna Varachkina, Tillmann Dönicke, and Luisa Gödeke. 2022. Levels of non-fictionality in fictional texts. In *Proceedings of the 18th Joint ACL - ISO Workshop on Interoperable Semantic Annotation within LREC2022*, pages 27–32, Marseille, France. European Language Resources Association.

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362, Sofia, Bulgaria. Association for Computational Linguistics.

Helmut Bonheim. 1975. Theory of narrative modes. *Semiotica*, 14(4):329–344.

Annelen Brunner, Ngoc Duyen Tanja Tu, Lukas Weimer, and Fotis Jannidis. 2020. To BERT or not to BERT – comparing contextual embeddings in a deep learning architecture for the automatic recognition of four types of speech, thought and writing representation. In *SwissText/KONVENS*.

Seymour Benjamin Chatman. 1980. *Story and Discourse: Narrative Structure in Fiction and Film*. Cornell paperbacks. Cornell University Press.

Tillmann Dönicke. 2020. Clause-level tense, mood, voice and modality tagging for German. In *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, pages 1–17, Düsseldorf, Germany. Association for Computational Linguistics.

Tillmann Dönicke. 2021. Delexicalised multilingual discourse segmentation for DISRPT 2021 and tense, mood, voice and modality tagging for 11 languages. In *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*, pages 33–45, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Tillmann Dönicke, Luisa Gödeke, and Hanna Varachkina. 2021. Annotating quantified phenomena in complex sentence structures using the example of generalising statements in literary texts. In *Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 20–32, Groningen, The Netherlands (online). Association for Computational Linguistics.

Tillmann Dönicke, Hanna Varachkina, Anna Mareike Weimer, Luisa Gödeke, Florian Barth, Benjamin Gittel, Anke Holler, and Caroline Sporleder. 2022. Modelling speaker attribution in narrative texts with biased and bias-adjustable neural networks. *Frontiers in Artificial Intelligence*, 4.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Annemarie Friedrich and Alexis Palmer. 2014. Situation entity annotation. In *Proceedings of LAW VIII - The 8th Linguistic Annotation Workshop*, pages 149–158, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

Luisa Gödeke, Florian Barth, Tillmann Dönicke, Anna Mareike Weimer, Hanna Varachkina, Benjamin Gittel, Anke Holler, and Caroline Sporleder. to appear. Generalisierungen als literarisches Phänomen. Charakterisierung, Annotation und automatische Erkennung. *Zeitschrift für digitale Geisteswissenschaften*.

Birgit Hamp and Helmut Feldweg. 1997. Germanet-a lexical-semantic net for german. *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.

Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

Markus Krug. 2020. *Techniques for the Automatic Extraction of Character Networks in German Historic Novels*. doctoral thesis, Universität Würzburg.

Markus Krug, Frank Puppe, Fotis Jannidis, Luisa Macharowsky, Isabella Reger, and Lukas Weimar. 2015. Rule-based coreference resolution in German historic novels. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 98–104, Denver, Colorado, USA. Association for Computational Linguistics.

Markus Krug, Frank Puppe, Isabella Reger, Lukas Weimer, Luisa Macharowsky, Stephan Feldhaus, and Fotis Jannidis. 2018. Description of a Corpus of Character References in German Novels - DROC [Deutsches ROman Corpus]. DARIAH-DE Working Papers 27, Göttingen: DARIAH-DE. URN: urn:nbn:de:gbv:7-dariah-2018-2-9.

Ekaterina Lapshinova-Koltunski, Christian Hardmeier, and Pauline Krielke. 2018. ParCorFull: a parallel corpus annotated with full coreference. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Beth Levin. 1995. English verb classes and alternations. *A preliminary Investigation*, 1.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Yann Mathet, Antoine Widlöcher, and Jean-Philippe Métivier. 2015. The unified and holistic method gamma (γ) for inter-annotator agreement measure and alignment. *Computational Linguistics*, 41(3):437–479.

Franz Hundsnurscher and Jochen Splett. 1982. *Semantik der Adjektive des Deutschen. Analyse der semantischen Relationen*.

Saif Mohammad and Peter Turney. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34, Los Angeles, CA. Association for Computational Linguistics.

Saif Mohammad and Peter Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465.

Nafise Sadat Moosavi, Leo Born, Massimo Poesio, and Michael Strube. 2019. Using automatically extracted minimum spans to disentangle coreference evaluation from boundary detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4168–4178, Florence, Italy. Association for Computational Linguistics.

Nafise Sadat Moosavi and Michael Strube. 2016. Which coreference evaluation metric do you trust? a proposal for a link-based entity aware metric. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 632–642, Berlin, Germany. Association for Computational Linguistics.

Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, and Daniel Zeman. 2021. Is one head enough? mention heads in coreference annotations compared with UD-style heads. In *Proceedings of the Sixth International Conference on Dependency Linguistics (Depling, SyntaxFest 2021)*, pages 101–114, Sofia, Bulgaria. Association for Computational Linguistics.

Joel Nothman, Nicky Ringland, Will Radford, Tara Murphy, and James R. Curran. 2013. Learning multilingual named entity recognition from wikipedia. *Artificial Intelligence*, 194:151–175. Artificial Intelligence, Wikipedia and Semi-Structured Resources.

Janis Pagel and Nils Reiter. 2020. GerDraCor-coref: A coreference corpus for dramatic texts in German. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 55–64, Marseille, France. European Language Resources Association.

Janis Pagel and Nils Reiter. 2021. DramaCoref: A hybrid coreference resolution system for German theater plays. In *Proceedings of the Fourth Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 36–46, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ondřej Pražák, Miloslav Konopík, and Jakub Sido. 2021. Multilingual coreference resolution with harmonized annotations. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1119–1123, Held Online. INCOMA Ltd.

Gerald Prince. 2012. A grammar of stories. In *A Grammar of Stories*. De Gruyter Mouton.

Robert Remus, Uwe Quasthoff, and Gerhard Heyer. 2010. SentiWS - a publicly available German-language resource for sentiment analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Ina Roesiger and Jonas Kuhn. 2016. IMS HotCoref DE: A data-driven co-reference resolver for German. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 155–160, Portorož, Slovenia. European Language Resources Association (ELRA).

Wolf Schmid. 2014. *Elemente der Narratologie*. de Gruyter.

Fynn Schröder, Hans Ole Hatzel, and Chris Biemann. 2021. Neural end-to-end coreference resolution for German in different domains. In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, pages 170–181, Düsseldorf, Germany. KONVENS 2021 Organizers.

Carlota S. Smith. 2003a. *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press.

George Smith. 2003b. A brief introduction to the TIGER treebank. Technical report, Universität Potsdam.

Jannik Strötgen and Michael Gertz. 2010. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 321–324, Uppsala, Sweden. Association for Computational Linguistics.

Jannik Strötgen and Michael Gertz. 2015. A baseline temporal tagger for all languages. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 541–547, Lisbon, Portugal. Association for Computational Linguistics.

Michael Vauth, Hans Ole Hatzel, Evelyn Gius, and Chris Biemann. 2021. Automated event annotation in literary texts. In *Proceedings of the Conference on Computational Humanities Research 2021 (CHR 2021)*, pages 333–345, Amsterdam, the Netherlands.

Zeno Vendler. 1957. Verbs and times. *The philosophical review*, 66(2):143–160.

Anna Mareike Weimer, Florian Barth, Tillmann Dönicke, Luisa Gödeke, Hanna Varachkina, Anke Holler, Caroline Sporleder, and Benjamin Gittel. to appear. The (in-)consistency of literary concepts – formalising, annotating and detecting literary comment. *Journal of Computational Literary Studies*.

Amir Zeldes, Yang Janet Liu, Mikel Iruskieta, Philippe Muller, Chloé Braud, and Sonia Badene. 2021. The DISRPT 2021 shared task on elementary discourse unit segmentation, connective detection, and relation classification. In *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*, pages 1–12, Punta Cana, Dominican Republic. Association for Computational Linguistics.

A Appendices

Aber Peter kauft sich jeden Morgen einen schlechten Kaffee.
‘But Peter buys himself a bad coffee every morning.’

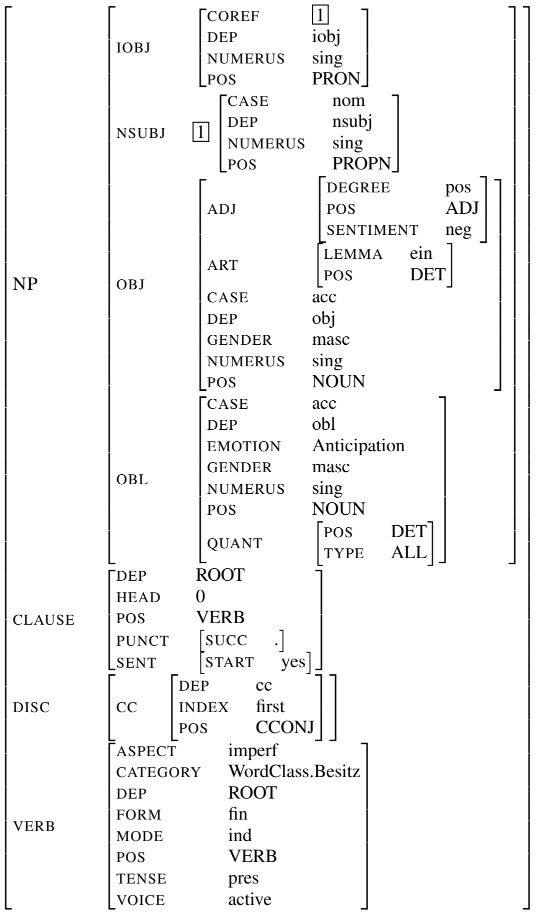

Figure 2: Sample DFG structure.

Acknowledgements

We thank all the researchers and developers who made their code and models publicly and freely accessible so that we could integrate them into MONAPipe. When using MONAPipe, you should reference all contributors of the components that you use (as listed in this paper or MONAPipe’s project documentation).

We further thank the anonymous reviewers for their valuable comments. This work is funded by Volkswagen Foundation (Dönicke, Sporleder), and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 424264086 (Barth, Varachkina, Sporleder).