

DramaCoref: A Hybrid Coreference Resolution System for German Theater Plays

Janis Pagel and Nils Reiter

Department of Digital Humanities, University of Cologne

Nov 11, 2021 @CRAC2021



University of Cologne



Volkswagen**Stiftung**

Drama

- ▶ Literary genre
- ▶ Usually meant to be performed on stage
- ▶ Distinct structural elements: dramatis personæ, segmentation tags (acts and scenes), stage directions, speaker tags, speech

ACT I

Scene I.

The Prince's Cabinet

The Prince, seated at a desk, which is covered with papers.

PRINCE. Complaints; nothing but complaints! [...] To be sure, if we could relieve every one, we might indeed be envied. Emilia? *opening a petition, and looking at the signature.* An Emilia? Yes—but an Emilia Bruneschi—not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? *Reads* She asks much—too much. But her name is Emilia. It is granted *signs the paper, and rings.*

Figure: Opening scene of Gotthold Ephraim Lessing's *Emilia Galotti*.
Translation from German by the authors.

Drama

- ▶ Literary genre
- ▶ Usually meant to be performed on stage
- ▶ Distinct structural elements: dramatis personæ, segmentation tags (acts and scenes), stage directions, speaker tags, speech

ACT I

Scene I.

The Prince's Cabinet

The Prince, seated at a desk, which is covered with papers.

PRINCE. Complaints; nothing but complaints! [...] To be sure, if we could relieve every one, we might indeed be envied. Emilia? *opening a petition, and looking at the signature.* An Emilia? Yes—but an Emilia Bruneschi—not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? *Reads* She asks much—too much. But her name is Emilia. It is granted *signs the paper, and rings.*

Figure: Opening scene of Gotthold Ephraim Lessing's *Emilia Galotti*.
Translation from German by the authors.

Drama

- ▶ Literary genre
- ▶ Usually meant to be performed on stage
- ▶ Distinct structural elements: dramatis personæ, segmentation tags (acts and scenes), stage directions, speaker tags, speech

ACT I

Scene I.

The Prince's Cabinet

The Prince, seated at a desk, which is covered with papers.

PRINCE. Complaints; nothing but complaints! [...] To be sure, if we could relieve every one, we might indeed be envied. Emilia? *opening a petition, and looking at the signature.* An Emilia? Yes—but an Emilia Bruneschi—not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? *Reads* She asks much—too much. But her name is Emilia. It is granted *signs the paper, and rings.*

Figure: Opening scene of Gotthold Ephraim Lessing's *Emilia Galotti*.
Translation from German by the authors.

Drama

- ▶ Literary genre
- ▶ Usually meant to be performed on stage
- ▶ Distinct structural elements: dramatis personæ, segmentation tags (acts and scenes), stage directions, speaker tags, speech

ACT I

Scene I.

The Prince's Cabinet

The Prince, seated at a desk, which is covered with papers.

PRINCE. Complaints; nothing but complaints! [...] To be sure, if we could relieve every one, we might indeed be envied. Emilia? *opening a petition, and looking at the signature.* An Emilia? Yes—but an Emilia Bruneschi—not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? *Reads* She asks much—too much. But her name is Emilia. It is granted *signs the paper, and rings.*

Figure: Opening scene of Gotthold Ephraim Lessing's *Emilia Galotti*.
Translation from German by the authors.

Drama

- ▶ Literary genre
- ▶ Usually meant to be performed on stage
- ▶ Distinct structural elements: dramatis personæ, segmentation tags (acts and scenes), stage directions, speaker tags, speech

ACT I

Scene I.

The Prince's Cabinet

The Prince, seated at a desk, which is covered with papers.

PRINCE. Complaints; nothing but complaints! [...] To be sure, if we could relieve every one, we might indeed be envied. Emilia? *opening a petition, and looking at the signature.* An Emilia? Yes—but an Emilia Bruneschi—not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? *Reads* She asks much—too much. But her name is Emilia. It is granted *signs the paper, and rings.*

Figure: Opening scene of Gotthold Ephraim Lessing's *Emilia Galotti*.
Translation from German by the authors.

Drama

- ▶ Literary genre
- ▶ Usually meant to be performed on stage
- ▶ Distinct structural elements: dramatis personæ, segmentation tags (acts and scenes), stage directions, speaker tags, speech

ACT I

Scene I.

The Prince's Cabinet

The Prince, seated at a desk, which is covered with papers.

PRINCE. Complaints; nothing but complaints! [...] To be sure, if we could relieve every one, we might indeed be envied. Emilia? *opening a petition, and looking at the signature.* An Emilia? Yes—but an Emilia Bruneschi—not Galotti. Not Emilia Galotti. What does she want, this Emilia Bruneschi? *Reads* She asks much—too much. But her name is Emilia. It is granted *signs the paper, and rings.*

Figure: Opening scene of Gotthold Ephraim Lessing's *Emilia Galotti*.
Translation from German by the authors.

Data

- ▶ Data described in [Pagel and Reiter, 2020, @LREC2020]
- ▶ 31 texts annotated for coreference
- ▶ noun phrases, predicates, generics, abstract anaphora
- ▶ around 300 000 tokens, 60 000 mentions and 5400 entities

Neural Mention Detection

- ▶ Comparison of parsing baseline, German BERT and BERT fine-tuned on literary data

Model	MacroPrec	MacroRec	MacroF1	Acc
Parsing Baseline	0.5	0.57	0.53	0.7
Base BERT	0.54	0.51	0.52	0.84
Literary BERT	0.24	0.33	0.28	0.73

Rule-based system

- ▶ Based on existing system by [Lee et al., 2013]
- ▶ Several sieves, ordered by precision
 - ▶ Exact match
 - ▶ Head match
 - ▶ Pronouns
 - ▶ Hyponymy
 - ▶ ...
- ▶ Extention:
 - ▶ GermaNet instead of WordNet as resource
 - ▶ Lemma-based instead of token-based
 - ▶ use of speaker information for pronoun resolution and proper nouns

Main results

	Mentions	MUC	B ³	CEAF _m	CEAF _e	CoNLL
act	auto	0.45	0.23	0.32	0.19	0.29
	gold	0.79	0.42	0.43	0.41	0.54
castmembers-only	auto	0.44	0.26	0.35	0.15	0.28
dramatispersonae	auto	0.46	0.23	0.33	0.19	0.29
	gold	0.79	0.44	0.45	0.42	0.55
scene	auto	0.43	0.27	0.37	0.24	0.32
	gold	0.72	0.47	0.49	0.42	0.54

Main results

	Mentions	MUC	B ³	CEAF _m	CEAF _e	CoNLL
act	auto	0.45	0.23	0.32	0.19	0.29
	gold	0.79	0.42	0.43	0.41	0.54
castmembers-only	auto	0.44	0.26	0.35	0.15	0.28
dramatispersonae	auto	0.46	0.23	0.33	0.19	0.29
	gold	0.79	0.44	0.45	0.42	0.55
scene	auto	0.43	0.27	0.37	0.24	0.32
	gold	0.72	0.47	0.49	0.42	0.54

Main results

	Mentions	MUC	B ³	CEAF _m	CEAF _e	CoNLL
act	auto	0.45	0.23	0.32	0.19	0.29
	gold	0.79	0.42	0.43	0.41	0.54
castmembers-only	auto	0.44	0.26	0.35	0.15	0.28
dramatispersonae	auto	0.46	0.23	0.33	0.19	0.29
	gold	0.79	0.44	0.45	0.42	0.55
scene	auto	0.43	0.27	0.37	0.24	0.32
	gold	0.72	0.47	0.49	0.42	0.54

Results on several corpora

M.	Corpus	M. F1	MUC	B ³	CEAF _m	CEAF _e	CoNLL
auto	Fairytales	0.70	0.51	0.30	0.34	0.26	0.36
	Radio news	0.43	0.36	0.30	0.38	0.37	0.38
	Newspaper	0.46	0.26	0.25	0.33	0.31	0.28
	Drama	0.60	0.45	0.23	0.32	0.19	0.29
gold	Fairytales	0.96	0.72	0.46	0.46	0.47	0.55
	Radio news	0.70	0.61	0.58	0.63	0.61	0.61
	Newspaper	0.86	0.66	0.62	0.65	0.64	0.64
	Drama	0.96	0.79	0.42	0.43	0.40	0.54

Results on several corpora

M.	Corpus	M. F1	MUC	B ³	CEAF _m	CEAF _e	CoNLL
auto	Fairytales	0.70	0.51	0.30	0.34	0.26	0.36
	Radio news	0.43	0.36	0.30	0.38	0.37	0.38
	Newspaper	0.46	0.26	0.25	0.33	0.31	0.28
	Drama	0.60	0.45	0.23	0.32	0.19	0.29
gold	Fairytales	0.96	0.72	0.46	0.46	0.47	0.55
	Radio news	0.70	0.61	0.58	0.63	0.61	0.61
	Newspaper	0.86	0.66	0.62	0.65	0.64	0.64
	Drama	0.96	0.79	0.42	0.43	0.40	0.54

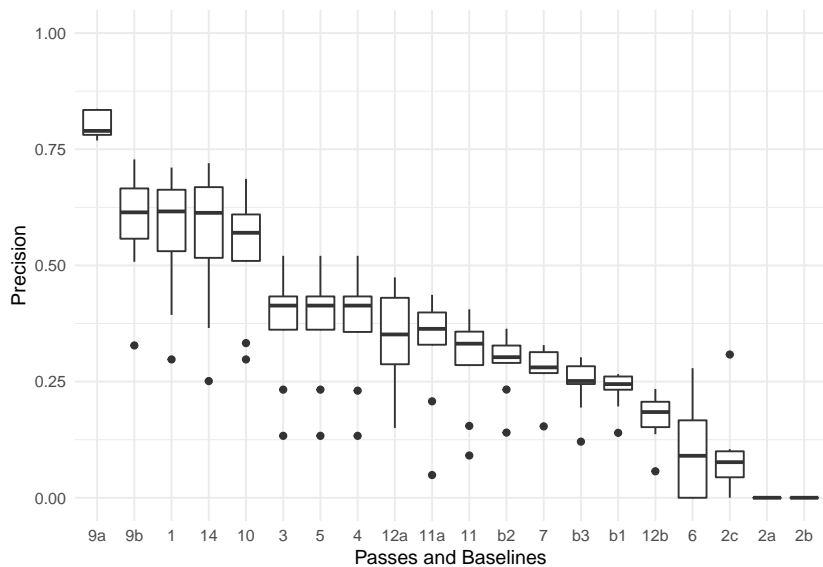
Results on several corpora

M.	Corpus	M. F1	MUC	B ³	CEAF _m	CEAF _e	CoNLL
auto	Fairytales	0.70	0.51	0.30	0.34	0.26	0.36
	Radio news	0.43	0.36	0.30	0.38	0.37	0.38
	Newspaper	0.46	0.26	0.25	0.33	0.31	0.28
	Drama	0.60	0.45	0.23	0.32	0.19	0.29
gold	Fairytales	0.96	0.72	0.46	0.46	0.47	0.55
	Radio news	0.70	0.61	0.58	0.63	0.61	0.61
	Newspaper	0.86	0.66	0.62	0.65	0.64	0.64
	Drama	0.96	0.79	0.42	0.43	0.40	0.54

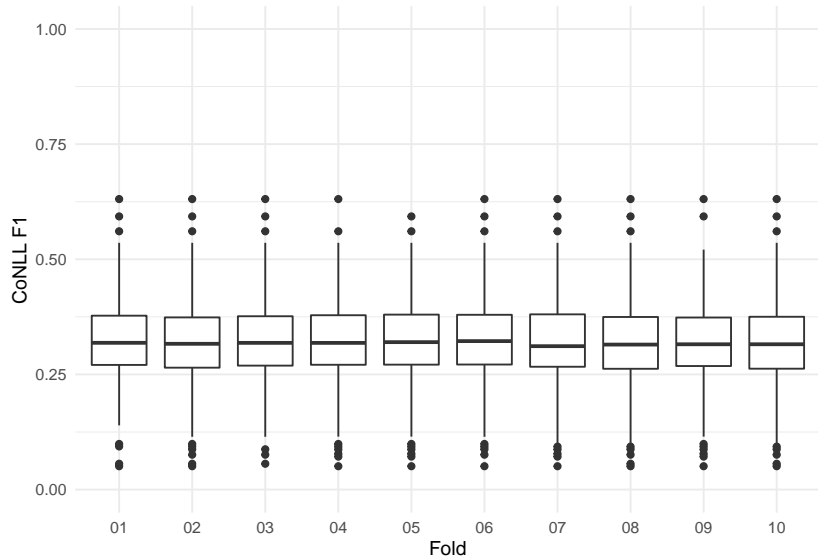
Comparison to state-of-the-art

System	MUC	B ³	CEAF _e	CoNLL
HotCoref	56.55	14.98	14.84	28.79
DramaCoref	42.54	19.87	18.97	27.12

Error analysis: Pass precision



Error analysis: Cross-validation



References



Lee, H., Chang, A., Peirsman, Y., Chambers, N., Surdeanu, M., and Jurafsky, D. (2013).

Deterministic coreference resolution based on entity-centric, precision-ranked rules.

Computational Linguistics, 39(4):885–916.



Pagel, J. and Reiter, N. (2020).

GerDraCor-Coref: A coreference corpus for dramatic texts in German.

In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, pages 55–64.