

# Romantic Code

Eine Computerlinguistische Modellierung der Literarischen Romantik

Vortrag am Kolloquium »Phänomenologie der Digital Humanities«  
an der Freien Universität Berlin (Prof. Dr. Frank Fischer)

online, 23. Juni 2022

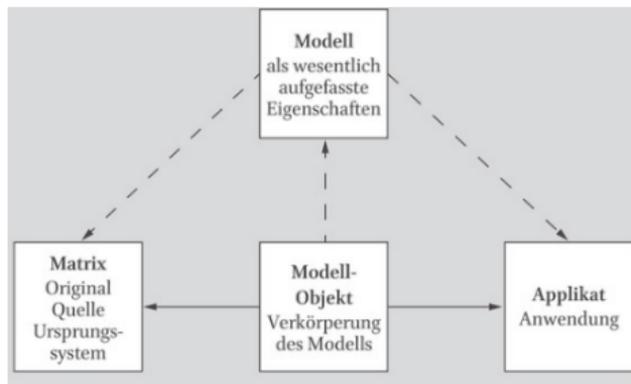
Tinghui Duan (段庭辉), M.A.  
DFG-Graduiertenkolleg „Modell Romantik“  
Friedrich-Schiller-Universität Jena

# Was ist ein Modell?

„ $X$  ist ein Modell des Originals  $Y$  für den Verwender  $v$  in der Zeitspanne  $t$  bezüglich der Intention  $Z$ .“ (Stachowiak 1980)

# Was ist ein Modell?

„ $X$  ist ein Modell des Originals  $Y$  für den Verwender  $v$  in der Zeitspanne  $t$  bezüglich der Intention  $Z$ .“ (Stachowiak 1980)



(Vgl. Kerschbaumer 2018)

# Zwecke eines (computerlinguistischen) Modells Romantik

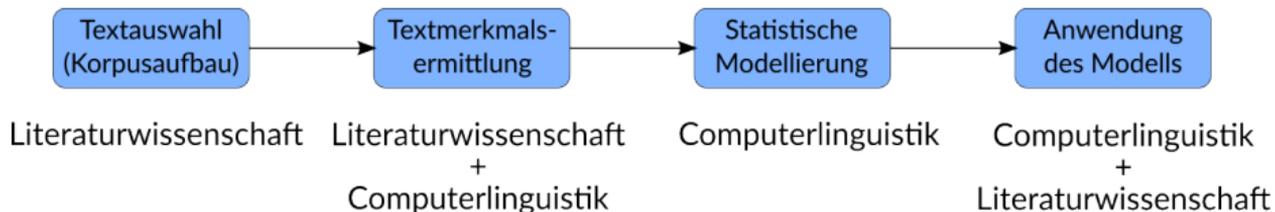
1. (Hermeneutische) Hypothesen zu romantischen Texten zu überprüfen
2. Erkenntnisse über stilistische Merkmale der romantischen Texte zu gewinnen
3. Alte und neue romantische Texte zu entdecken

# Was ist ein „Modell Romantik“?

Ein „Modell Romantik“ ( $X$ ) ist ein Modell der literarischen Texte aus der Romantikepoche ( $Y$ ) für die Literaturwissenschaft ( $v$ ) im aktuellen Forschungsdiskurs ( $t$ ) bezüglich der Intention, sprachliche Merkmale der romantischen Texte zu untersuchen und Texte in Bezug auf ihre Romantizität auszuwerten ( $Z$ ).

# Was ist ein „Modell Romantik“?

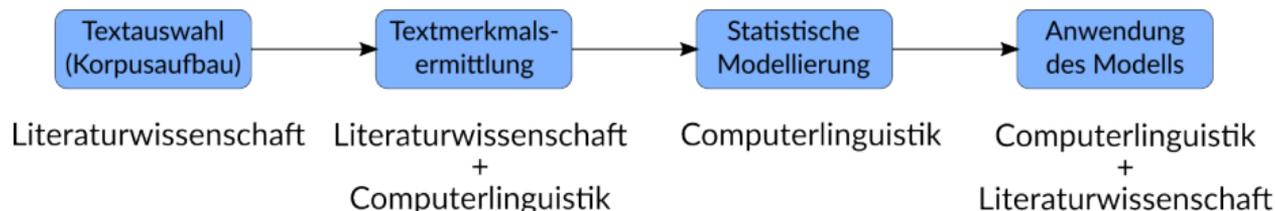
Ein „Modell Romantik“ ( $X$ ) ist ein Modell der literarischen Texte aus der Romantikepoche ( $Y$ ) für die Literaturwissenschaft ( $v$ ) im aktuellen Forschungsdiskurs ( $t$ ) bezüglich der Intention, sprachliche Merkmale der romantischen Texte zu untersuchen und Texte in Bezug auf ihre Romantizität auszuwerten ( $Z$ ).



# Herausforderungen eines (computerlinguistischen) Modells Romantik

## Herausforderungen

- keine klare Definition von Romantik
- historische Sprachentwicklung
- Unbefriedigende Leistung von NLP-Tools
- Interpretierbarkeit der Modellaussagen



## Workflow

1. Textannotation (Kategorisierung)
2. Merkmalermittlung (NLP)
3. Statistische Modellierung + Evaluation (Machine Learning)
4. Anwendung (Auswertung unbekannter Texte)

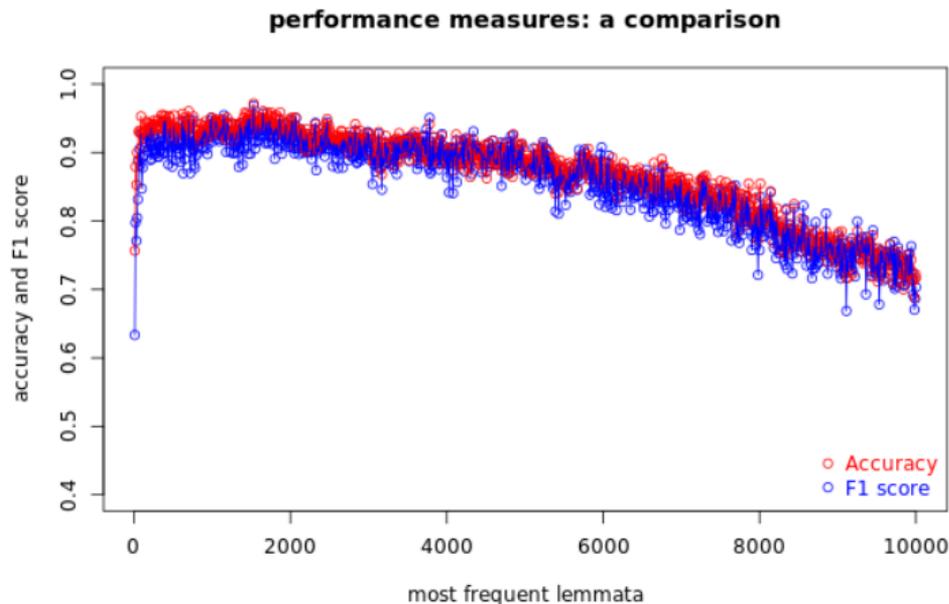
## 1. Textannotation (Kategorisierung)

- Datenquelle: *TextGrid Repository* (<https://textgridrep.org/>)
- Kategorisierung basiert auf Einführungs-/Lehrbüchern der germanistischen Literaturwissenschaft
- Drei benachbarte Epochen

	Autor:innen	Texte	Token
Weimarer Klassik	5	21	1,708,643
Romantik	25	235	6,531,396
Realismus	14	212	8,297,365
$\Sigma$	44	468	16,537,404

## 2. Merkmalermittlung

die  $n$  häufigsten Wörter (Lemmata) — ein Experiment mit Stylo (Burrows's Delta)



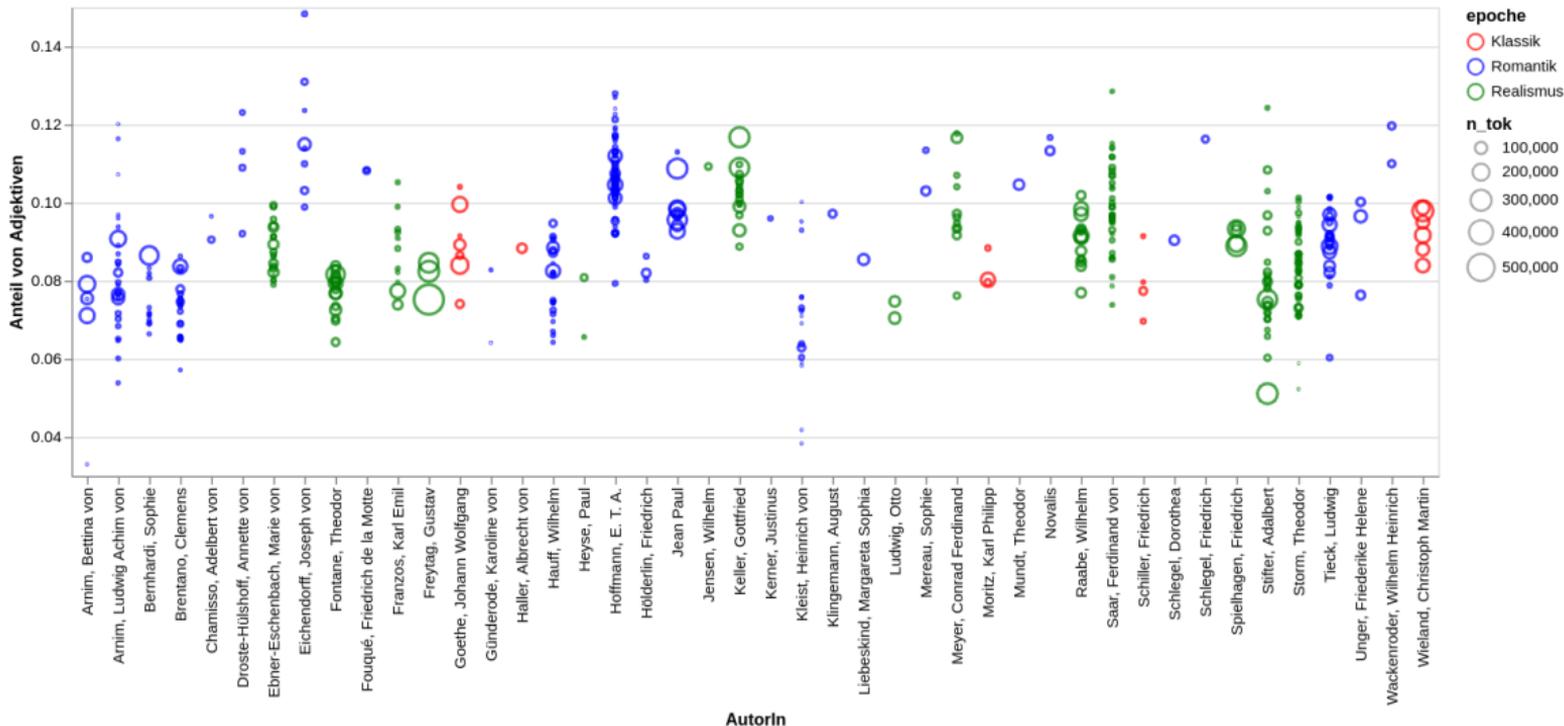
(Vgl. Eder, Rybicki und Kestemont 2018; Jannidis und Lauer 2014)

„Der französische Dichter Alfred de Musset hat das [romantische Stilphänomen, TD] in die spöttische Bemerkung gefasst, die Romantik bestehe darin, lauter Adjektive zu gebrauchen. Wer nicht einfach von Herz, Liebe und Frömmigkeit schreibe, sondern vom kindlichen Herzen, von unschuldiger Liebe und süßer Frömmigkeit, werde zum Romantiker.“

(zitiert nach Matuschek 2021)

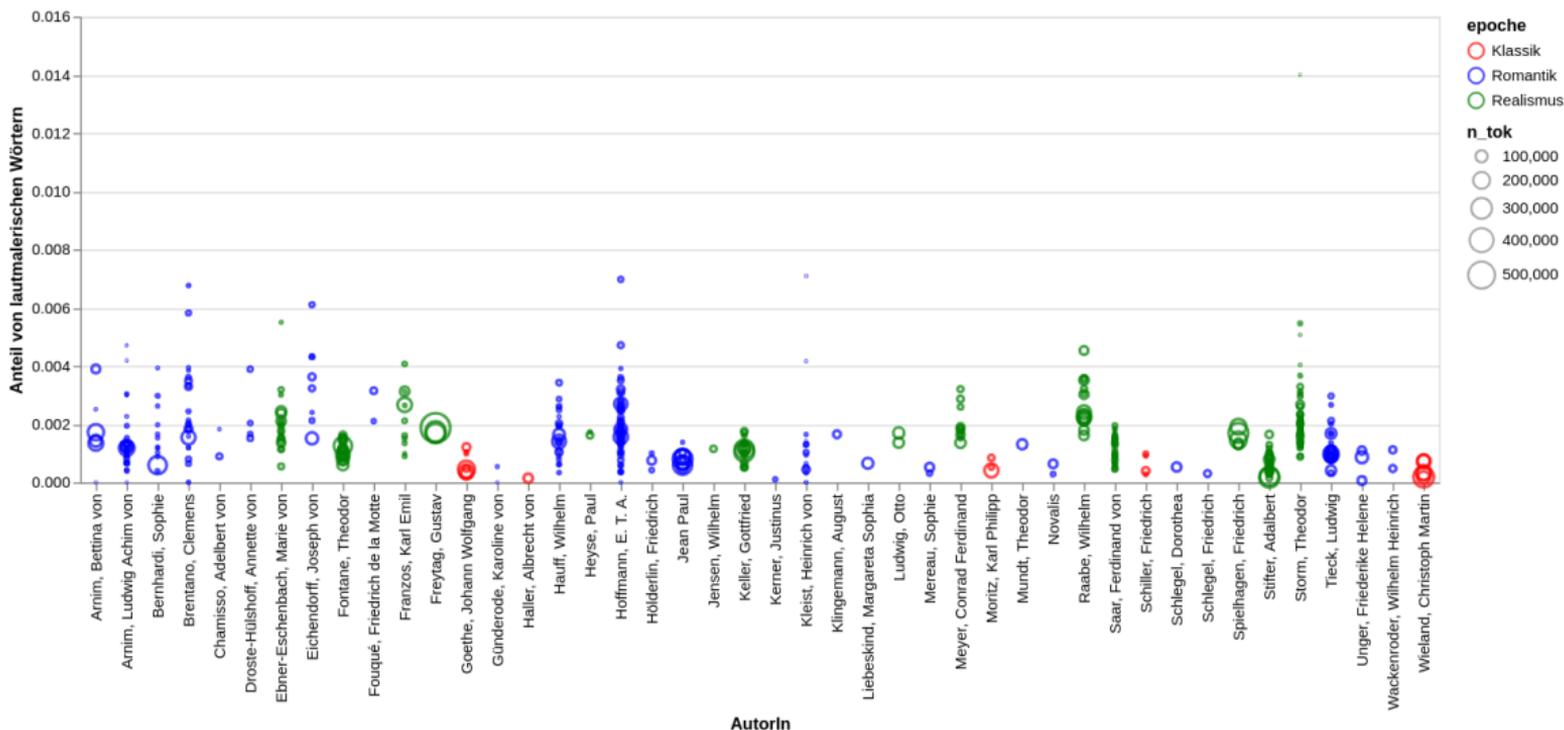
# Hypothese zu romantischen Texten 1: hoher Anteil von Adjektiven

## Anteil von Adjektiven (POS="ADJA")



# Hypothese zu romantischen Texten 2: Volkstümlichkeit/Lautmalerei

## Anteil von lautmalersichen Wörtern



# Hypothese zu romantischen Texten 3: Phantasie

So ist mir noch nie zu Muthe gewesen: es ist, als hätt' ich vorhin geträumt, oder ich wäre in eine andere Welt hinübergeschlummert; denn in der Welt, in der ich sonst lebte, wer hätte da sich um Blumen bekümmert, und gar von einer so seltsamen Leidenschaft für eine Blume hab' ich damals nie gehört.

(Novalis: *Heinrich von Ofterdingen*)

# Hypothese zu romantischen Texten 3: Phantasie

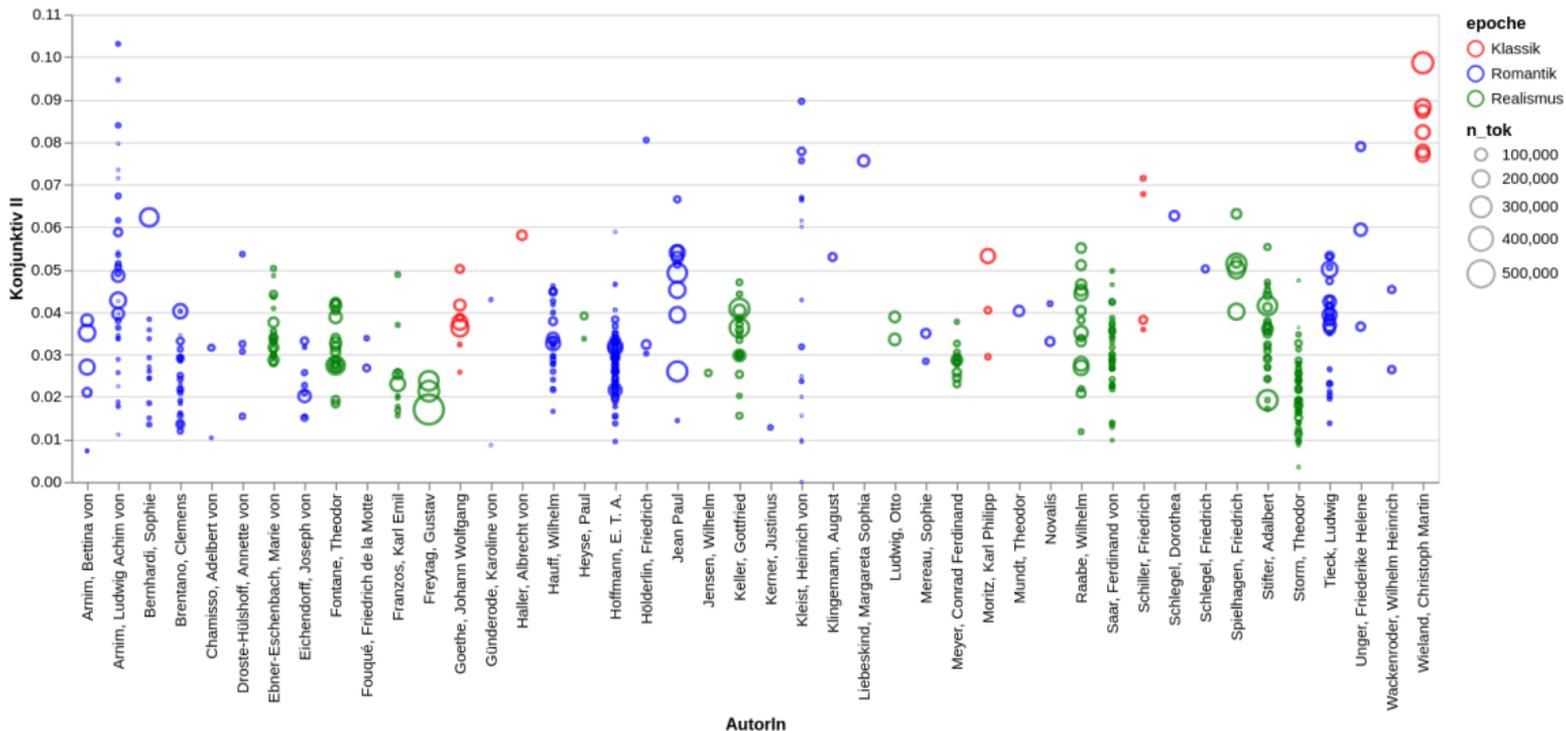
So ist mir noch nie zu Muthe gewesen: es ist, als hätt' ich vorhin geträumt, oder ich wäre in eine andere Welt hinübergeschlummert; denn in der Welt, in der ich sonst lebte, wer hätte da sich um Blumen bekümmert, und gar von einer so seltsamen Leidenschaft für eine Blume hab' ich damals nie gehört.

(Novalis: *Heinrich von Ofterdingen*)

ID	Token	Lemma	POS	Feats
1	es	es	PPER	3 Sg Neut Nom
2	ist	sein	VAFIN	3 Sg Pres Ind
3	,	,	\$,	
4	als	als	KOKOM	
5	<b>hätte</b>	haben	VAFIN	1 Sg  <b>Past Subj</b>
6	ich	ich	PPER	1 Sg _ Nom
7	vorhin	vorhin	ADV	
8	geträumt	träumen	VVPP	
9	,	,	\$,	
10	oder	oder	KON	
11	ich	ich	PPER	1 Sg _ Nom
12	<b>wäre</b>	sein	VAFIN	1 Sg  <b>Past Subj</b>
13	in	in	APPR	Acc
14	eine	eine	ART	Indef Fem Acc Sg
15	andere	andere	ADJA	_ Fem Acc Sg _
16	Welt	Welt	NN	Fem Acc Sg
17	hinübergeschlummert	hinüberschlummern	VVPP	
18	;	;	\$.	

# Hypothese zu romantischen Texten 3: Phantasie

## Anteil von Konjunktiv II



## Hypothese zu romantischen Texten 4: Dynamik

Sie lächelte still und sah mich recht vergnügt und freundlich an, und von fern schallte immerfort die Musik **herüber**, und Leuchtkugeln flogen vom Schloß **durch die** stille Nacht **über die** Gärten, und die Donau rauschte **dazwischen herauf** — und es war alles, alles gut!

(Eichendorff: *Aus dem Leben eines Taugenichts*)

# Hypothese zu romantischen Texten 4: Dynamik

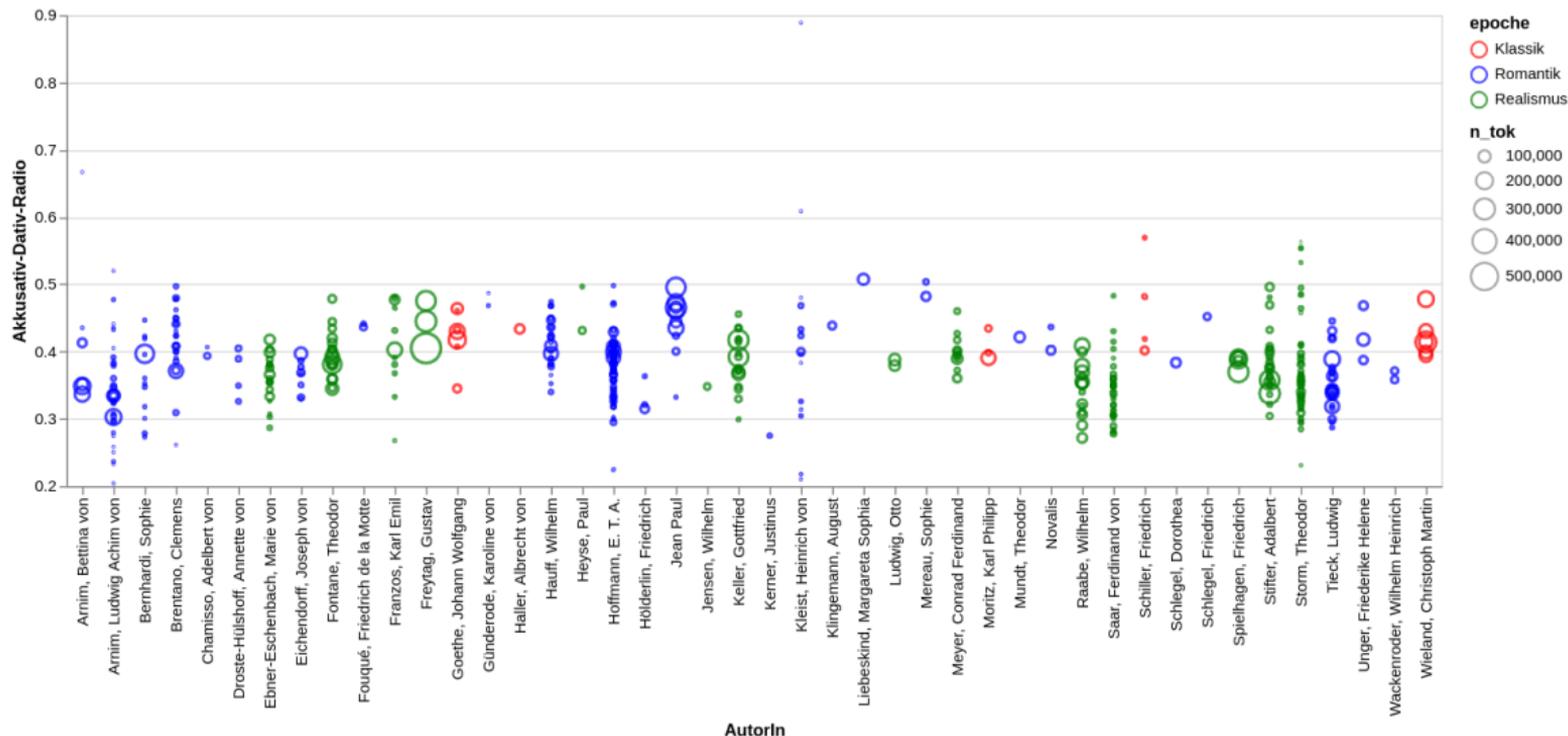
Sie lächelte still und sah mich recht vergnügt und freundlich an, und von fern schallte immerfort die Musik **herüber**, und Leuchtkugeln flogen vom Schloß **durch die** stille Nacht **über die** Gärten, und die Donau rauschte **dazwischen herauf** — und es war alles, alles gut!

(Eichendorff: *Aus dem Leben eines Taugenichts*)

ID	Token	Lemma	POS	Feats
54	und	und	KON	_
55	Leuchtkugeln	Leuchtkugel	NN	Fem Nom Pl
56	flogen	fliegen	VVFIN	3 Pl Past Ind
57	vom	von	APPRART	Dat
58	Schloß	Schloß	NN	Neut Dat Sg
59	durch	durch	APPR	Acc
60	die	die	ART	Def Fem Acc Sg
61	stille	still	ADJA	Pos Fem Acc Sg _
62	Nacht	Nacht	NN	Fem Acc Sg
63	über	über	APPR	Acc
64	die	die	ART	Def Masc Acc Pl
65	Gärten	Garten	NN	Masc Acc Pl

# Hypothese zu romantischen Texten 4: Dynamik

## Akkusativ/Dativ-Ratio



## Neue Liebe

Durch den Wald, im Mondenscheine  
Sah ich jüngst die Elfen reuten;  
Ihre Hörner hört ich klingen,  
Ihre Glöckchen hört' ich läuten.

Ihre weißen Rößlein trugen  
Güldnes Hirschgeweih' und flogen  
Rasch dahin, wie wilde Schwäne  
Kam es durch die Luft gezogen.

Lächelnd nickte mir die Kön'gin,  
Lächelnd, im Vorüberreiten.  
Galt das meiner neuen Liebe,  
(0.004) (0.099)  
Oder soll es Tod bedeuten?  
(-0.07) (0.004)

(Heinrich Heine: XXXII)

# Hypothese zu romantischen Texten 5: Kippfigur

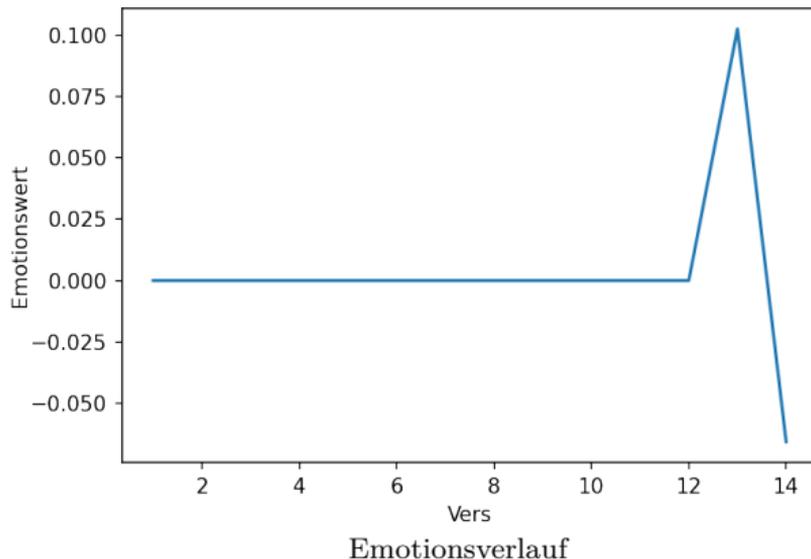
## Neue Liebe

Durch den Wald, im Mondenscheine  
Sah ich jüngst die Elfen reuten;  
Ihre Hörner hört ich klingen,  
Ihre Glöckchen hört' ich läuten.

Ihre weißen Rößlein trugen  
Güldnes Hirschgeweih' und flogen  
Rasch dahin, wie wilde Schwäne  
Kam es durch die Luft gezogen.

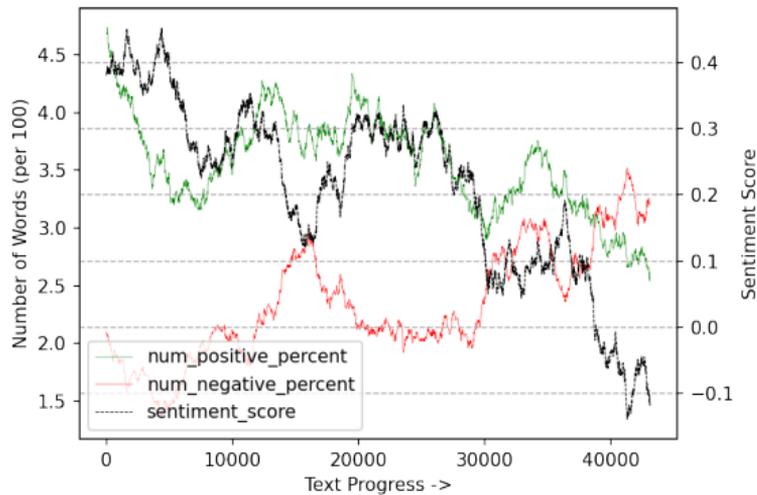
Lächelnd nickte mir die Kön'gin,  
Lächelnd, im Vorüberreiten.  
Galt das meiner neuen Liebe,  
(0.004) (0.099)  
Oder soll es Tod bedeuten?  
(-0.07) (0.004)

(Heinrich Heine: XXXII)

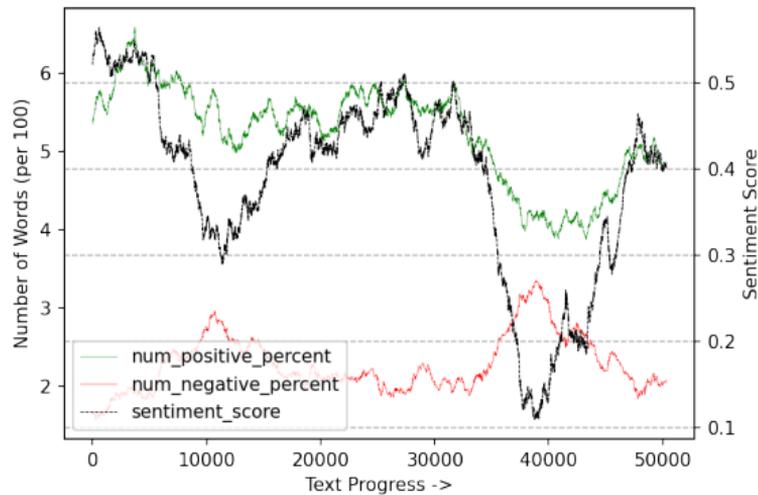


(Emotionslexikon: SentiWS, vgl. Remus, Quasthoff und Heyer 2010)

# Emotionsanalyse



*Die Leiden des jungen Werthers*



*Heinrich von Ofterdingen*

(Emotionslexikon: SentiWS, vgl. Remus, Quasthoff und Heyer 2010)

## 2. Merkmalsermittlung

literaturwissenschaftlich motivierte Merkmale

- Adjektive
- Lautmalerei (Vorkommen bestimmter Wörter)
- Phantasie (Konjunktiv II)
- Dynamik (Akkusativ)
- Kippfigur (starke emotionale Schwankung)
- ...

## 2. Merkmalsermittlung

### literaturwissenschaftlich motivierte Merkmale

- Adjektive
- Lautmalerei (Vorkommen bestimmter Wörter)
- Phantasie (Konjunktiv II)
- Dynamik (Akkusativ)
- Kippfigur (starke emotionale Schwankung)
- ...

### (computer)linguistisch motivierte Merkmale

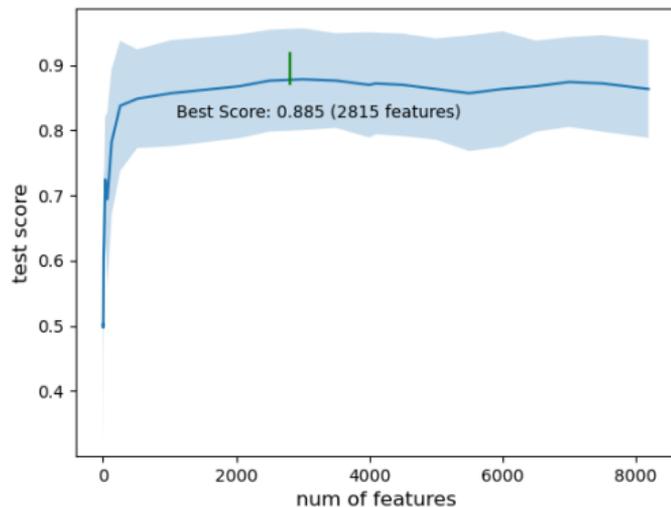
- Text-, Satz- und Wortlänge
- Lexikalische Vielfältigkeit (z.B. TTR)
- Verteilung von Wortarten (syntaktisch, lex)
- Vokal/Konsonant Relation (phonologisch, sublex)
- Emotionalität (semantisch, lex)
- ...

(zu Interpretation von TTR vgl. Long, Detwyler und Zhu 2018)

## Workflow

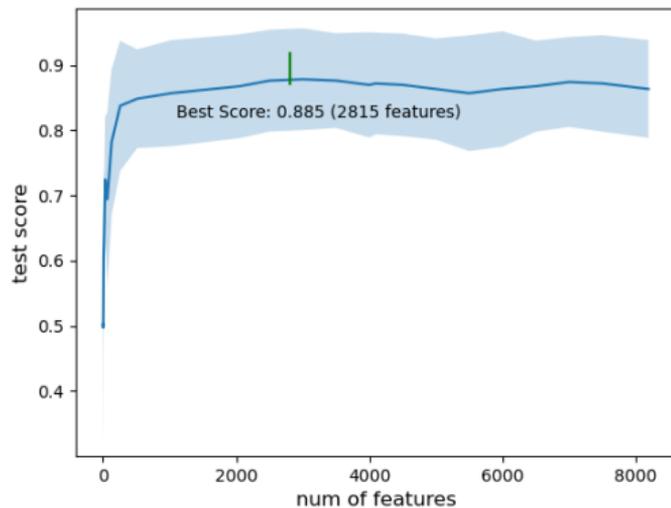
1. Textannotation (Kategorisierung)
2. Merkmalermittlung (NLP)
3. Statistische Modellierung + Evaluation (Machine Learning)
4. Anwendung (Auswertung unbekannter Texte)

## TF-IDF (Unigram, Lemmata) mit Logistic Regression

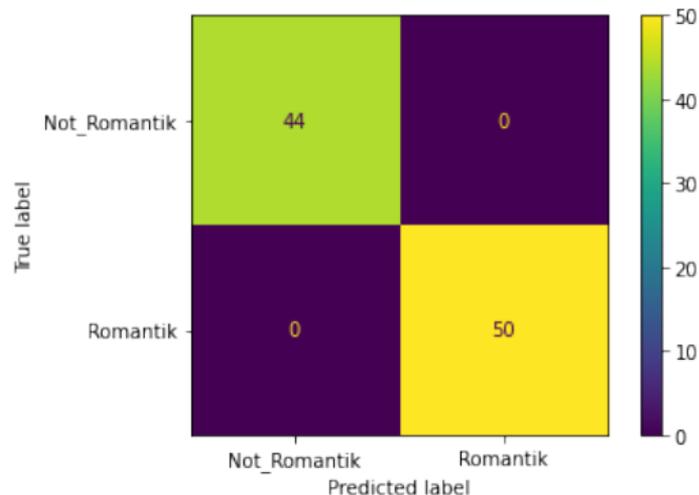


Gridsearch for best score parameter

## TF-IDF (Unigram, Lemmata) mit Logistic Regression



Gridsearch for best score parameter



Confusionmatrix der Testdatenevaluation (features=2815)

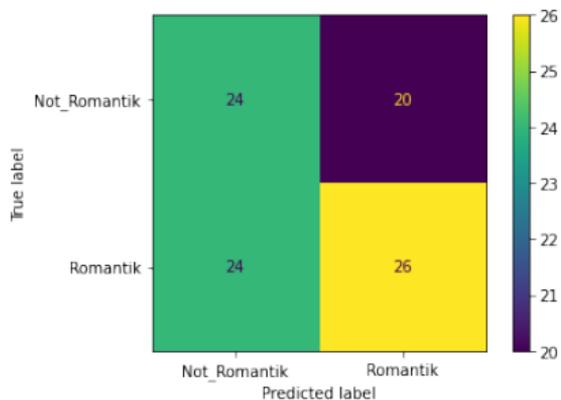
# Evaluation des Modells

<b>Autor:in</b>	<b>Werk</b>	<b>Prediction</b>	<b>predict_proba</b>
Thea Dorn	<i>Die Unglückseligen</i>	Not Romantik	0.991, 0.009
Wilhelm Genazino	<i>Ein Regenschirm für diesen Tag</i>	Not Romantik	1.000, 0.000
Wolfgang Herrndorf	<i>Diesseits des Van-Allen-Gürtels</i>	Not Romantik	1.000, 0.000
Wolfgang Herrndorf	<i>Tschick</i>	Not Romantik	1.000, 0.000
Wolfgang Hilbig	<i>Alte Abdeckerei</i>	Not Romantik	0.991, 0.009
Wolfgang Hilbig	<i>Die Kunde von den Bäumen</i>	Not Romantik	0.996, 0.004
Felicitas Hoppe	<i>Paradiese, Übersee</i>	Not Romantik	0.997, 0.003
Helmut Krausser	<i>UC</i>	Not Romantik	0.999, 0.001
Klaus Modick	<i>Fahrtwind</i>	Not Romantik	0.999, 0.001
Matthias Zschokke	<i>Der Mann mit den zwei Augen</i>	Not Romantik	0.999, 0.001
Simon Strauss	<i>Römische Tage</i>	Not Romantik	0.997, 0.003

(Vgl. bspw. Bartsch, Annika. 2019. *Romantik um 2000: zur Reaktualisierung eines Modells in deutschsprachigen Romanen der Gegenwart*. Heidelberg: Universitätsverlag Winter.)

# Evaluation des Modells

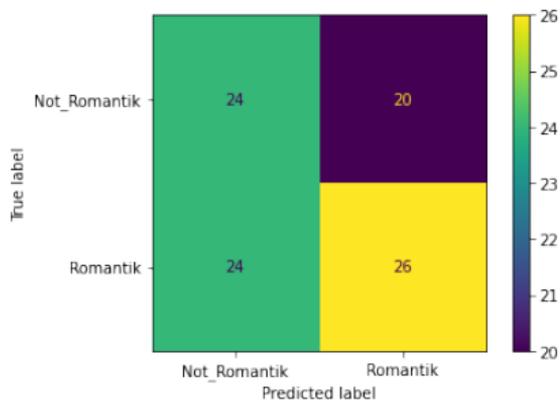
num of features: 5  
evaluation score: 0.532



Confusionmatrix der Testdatenevaluation  
(features=5)

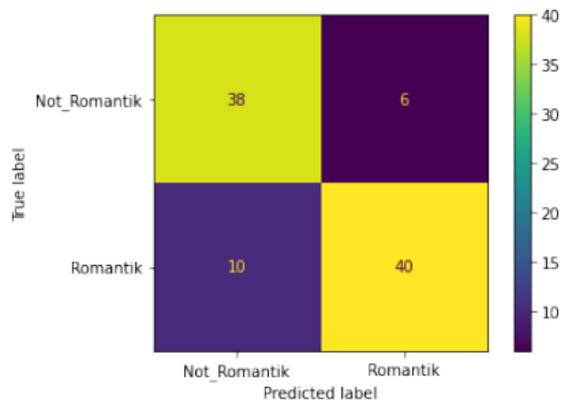
# Evaluation des Modells

num of features: 5  
evaluation score: 0.532



Confusionmatrix der Testdatenevaluation  
(features=5)

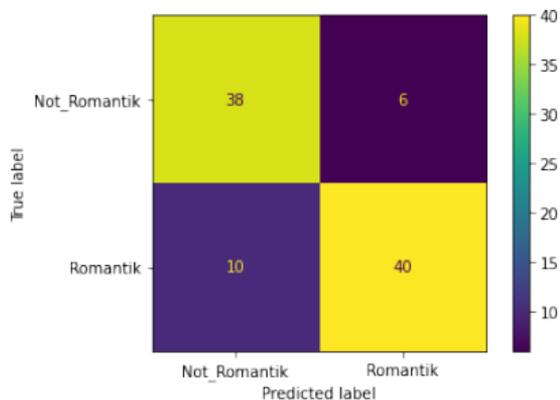
num of features: 20  
evaluation score: 0.829



Confusionmatrix der Testdatenevaluation  
(features=20)

# Interpretation des Modells

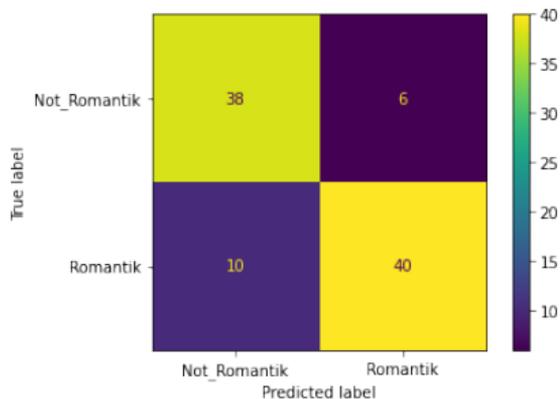
num of features: 20  
evaluation score: 0.829



Confusionmatrix der Testdatenevaluation  
(features=20)

# Interpretation des Modells

num of features: 20  
evaluation score: 0.829



Confusionmatrix der Testdatenevaluation  
(features=20)

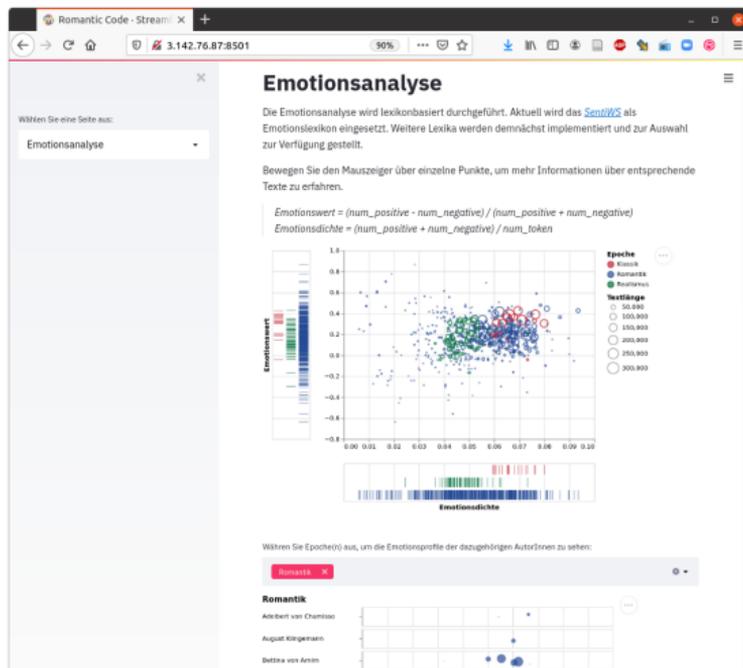
Feature	Koeffizient
sonscore	3.122
wlc	1.845
n_N	1.338
sl	1.318
n_V	1.141
n_ADJ	0.832
rhyme	0.618
vcr	0.538
adr	0.466
konjunktiv	0.439
iforger_score	0.239
sentiment_score	0.149
lautmalerei	-0.011
n_tok	-0.086
alliteration	-0.138
assonance	-0.284
homoioteleuton	-0.285
n_ADV	-0.436
n_sent	-0.448
wls	-1.073

# Anwendung des Modells

Vorhersage des Modells mit 20 Features

<b>Autor:in</b>	<b>Werk</b>	<b>Prediction</b>	<b>predict_proba</b>
Thea Dorn	Die Unglückseligen	Not Romantik	0.984, 0.016
Wilhelm Genazino	Ein Regenschirm für diesen Tag	Not Romantik	0.524, 0.476
Wolfgang Herrndorf	Diesseits des Van-Allen-Gürtels	Not Romantik	0.854, 0.146
Wolfgang Herrndorf	Tschick	Not Romantik	0.951, 0.049
Wolfgang Hilbig	Alte Abdeckerei	Romantik	0.128, 0.872
Wolfgang Hilbig	Die Kunde von den Bäumen	Not Romantik	0.734, 0.266
Felicitas Hoppe	Paradiese, Übersee	Romantik	0.376, 0.624
Helmut Krausser	UC	Not Romantik	0.558, 0.442
Klaus Modick	Fahrtwind	Not Romantik	0.882, 0.118
Matthias Zschokke	Der Mann mit den zwei Augen	Not Romantik	0.807, 0.193
Simon Strauss	Römische Tage	Romantik	0.412, 0.588

## In Entwicklung: ein exploratives Tool für Romantikforschung



- Archer, Jodie und Matthew Lee Jockers (2016). *The bestseller code: anatomy of the blockbuster novel*. New York: St. Martin's Press. 242 S.
- Eder, Maciej, Jan Rybicki und Mike Kestemont (2018). *Stylometry with R: A Package for Computational Text Analysis*. URL: [https://github.com/computationalstylistics/stylo\\_howto/blob/master/stylo\\_howto.pdf](https://github.com/computationalstylistics/stylo_howto/blob/master/stylo_howto.pdf).
- Jacobs, Arthur M. (2015). „Neurocognitive poetics: methods and models for investigating the neuronal and cognitive-affective bases of literature reception“. In: *Frontiers in Human Neuroscience* 9. DOI: 10.3389/fnhum.2015.00186. URL: <https://www.frontiersin.org/articles/10.3389/fnhum.2015.00186/full> (besucht am 10.04.2021).
- Jacobs, Arthur M. und Annette Kinder (2020). *Quasi Error-free Text Classification and Authorship Recognition in a large Corpus of English Literature based on a Novel Feature Set*. arXiv: 2010.10801 [cs.CL].
- Jakobson, Roman (1960). „Linguistics and Poetics“. In: *Style in Language*. Hrsg. von Thomas Albert Sebeok. New York & London: MIT Press, S. 350–377.
- Jannidis, Fotis und Gerhard Lauer (2014). „Burrows's Delta and Its Use in German Literary History“. In: Erlin, Matt und Lynne Tatlock. *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. Rochester: Camden House, S. 29–54.
- Kerschbaumer, Sandra (2018). *Immer wieder Romantik: modelltheoretische Beschreibungen ihrer Wirkungsgeschichte*. Jenaer Germanistische Forschungen. Heidelberg: Universitätsverlag Winter.
- Long, Hoyt, Anatoly Detwyler und Yuancheng Zhu (2018). „Self-Repetition and East Asian Literary Modernity, 1900-1930“. In: *Journal of Cultural Analytics* 2.2. DOI: 10.22148/16.022. URL: <https://culturalanalytics.org/article/11040-self-repetition-and-east-asian-literary-modernity-1900-1930>.
- Matuschek, Stefan (2021). *Der gedichtete Himmel: Eine Geschichte der Romantik*. C.H.Beck.

- Remus, Robert, Uwe Quasthoff und 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)*. LREC 2010. Valletta, Malta: European Language Resources Association (ELRA). URL: [http://www.lrec-conf.org/proceedings/lrec2010/pdf/490\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/490_Paper.pdf) (besucht am 20.02.2021).
- Stachowiak, Herbert, Hrsg. (1980). *Modelle und Modelldenken im Unterricht: Anwendungen der allgemeinen Modelltheorie auf die Unterrichtspraxis*. Forschen und Lernen. Bad Heilbrunn/Obb.: Klinkhardt.