

# Zum Verhältnis von “close” und “distant reading” in der digitalen Textanalyse

*Jun.-Prof. Ulrike Henny-Krahmer*  
*[ulrike.henny-krahmer@uni-rostock.de](mailto:ulrike.henny-krahmer@uni-rostock.de)*



## Überblick

- Begriffe
  - Digitale Textanalyse
  - Close & distant reading
  - Zwischenformen
  - Qualitative & quantitative Analysen
- Anwendungsfälle
  - Klassifikation literarischer Gattungen
  - Mikroanalyse von Stilvariation
  - Topic Modeling für Digitale Editionen
- Fazit

## Begriffe: Was ist digitale Textanalyse?

“Literatur rechnen, oder digitale Textanalyse, ist eine wissenschaftliche Tätigkeit, die digitalisierte (literarische) Texte mithilfe von rechnerischen Programmen analysiert und beschreibt.”

(Herrmann 2012, <http://litre.uni-goettingen.de>)

## Begriffe: Was ist digitale Textanalyse?

Jannidis 2010:

- Computergestützte Textanalyse: Teil der Computerphilologie
- Computerphilologie: “Die → Computerphilologie, also die Verwendung des Computers in der Literaturwissenschaft, vor allem zur Erstellung und Verwendung digitaler Editionen (Jannidis 2007), hat dabei eine Brückenfunktion und verbindet die Literaturwissenschaft mit den stärker formalisierten Wissenschaften der Computer- bzw. Korpuslinguistik sowie der Informatik. Ihre Verfahren sind daher in den meisten Fällen keineswegs einzigartig für die Literaturwissenschaft, sondern verwenden bereits bestehende Konzepte und Verfahren, um literaturwissenschaftliche Probleme zu lösen.”

## Begriffe: Was sind „close“ und „distant reading“?

**Close reading:** “Close Reading bezeichnet die sorgfältige Lektüre und Interpretation eines einzelnen oder weniger Texte. Close Reading ist in der digitalen Literaturwissenschaft außerdem mit der manuellen Annotation textueller Phänomene verbunden (vgl. auch Distant Reading als Gegenbegriff).”

**Distant reading:** “Distant Reading ist ein Ansatz aus den digitalen Literaturwissenschaften, bei dem computationelle Verfahren auf häufig große Mengen an Textdaten angewandt werden, ohne dass die Texte selber gelesen werden. Meist stehen hier quantitative Analysen im Vordergrund, es lassen sich jedoch auch qualitative Metadaten quantitativ vergleichen. Als Gegenbegriff zu Close Reading wurde der Begriff insbesondere von Franco Moretti (2000) geprägt.”

(Quelle: forText. Literatur digital erforschen. 2016ff. <https://fortext.net/ressourcen/glossar>)

## Begriffe: Was sind „close“ und „distant reading“?

Moretti 2000:

„we know how to read texts, now let's learn how *not* to read them. Distant reading: where distance, let me repeat it, is a *condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, Less is more. If we want to understand the system in its entirety, we must accept losing something.“



## Begriffe: Was sind „close“ und „distant reading“?

“Computer können lesen, nur ist ihr Lesen genau genommen ein Rechnen. Distant reading meint eine distanzierte, weil auf große Korpora statt auf Einzelwerke ausgerichtete Erschließung der Literatur, die statistische Methoden statt Interpretation, informationstechnische Suchen statt Einzellektüre und mathematische Modellierungen statt Kulturtheorien nicht scheut.”

(Lauer 2009)

## Begriffe: Zwischenformen

Distant reading

Middle-distance reading

Scalable reading

„mixed methods“

Close reading

Wechsel zwischen quantitativen Metriken  
und Detailstudien von Mustern in  
Einzeltexten;  
Auch: mittelgroße Korpora  
(u.a. Craig 2013, Herrmann 2017)



## Begriffe: Quantitative vs. qualitative Analysen

- Qualitativ
  - textbasiert
  - Einzellektüre (Werke, Textstellen)
  - Interpretation
- Quantitativ
  - numerisch
  - algorithmisch
  - mathematisch modelliert
  - statistisch fundiert

Analog vs.  
digital?

Close vs. distant  
reading?

## Distant reading: Möglichkeiten

- Für die Literaturgeschichte
  - große Zeiträume analysieren
  - große Textmengen: Texte über den Kanon hinaus
  - vergleichende / kontrastive Analysen
- Für Stilanalysen:
  - systematische, formale Analysen ganzer/vieler Texte
  - auch hier: vergleichende/kontrastive Analysen

## Anwendung: Klassifikation literarischer Gattungen

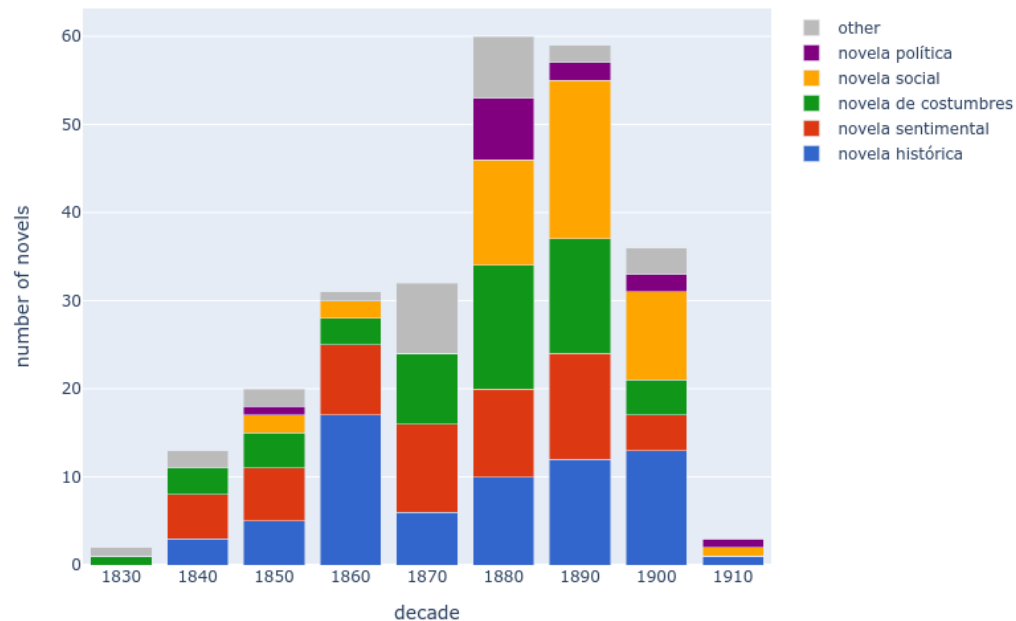
- Dissertation „Genre Analysis and Corpus Design: 19th Century Spanish American Novels (1830-1910)“
- Hintergrund: Projekt CLiGS („Computergestützte Literarische Gattungsstilistik“), Universität Würzburg, 2015-2020 (BMBF Nachwuchsgruppe)
  - Aufbau eines digitalen Textkorpus: 256 Romane in XML-TEI aus Argentinien, Kuba, Mexiko, <https://github.com/cligs/conha19>
  - Digitale Bibliographie: 829 Werke, „Bib-ACMé: Bibliografía digital de novelas argentinas, cubanas y mexicanas (1830-1910)“, <https://github.com/cligs/bibacme> & <http://bibacme.cligs.digital-humanities.de/>
  - Klassifikation & Clustering der Texte nach literarischen Gattungen und literarischen Strömungen



# Anwendung: Klassifikation literarischer Gattungen

## Metadaten: Zuordnung von Romanuntergattungen

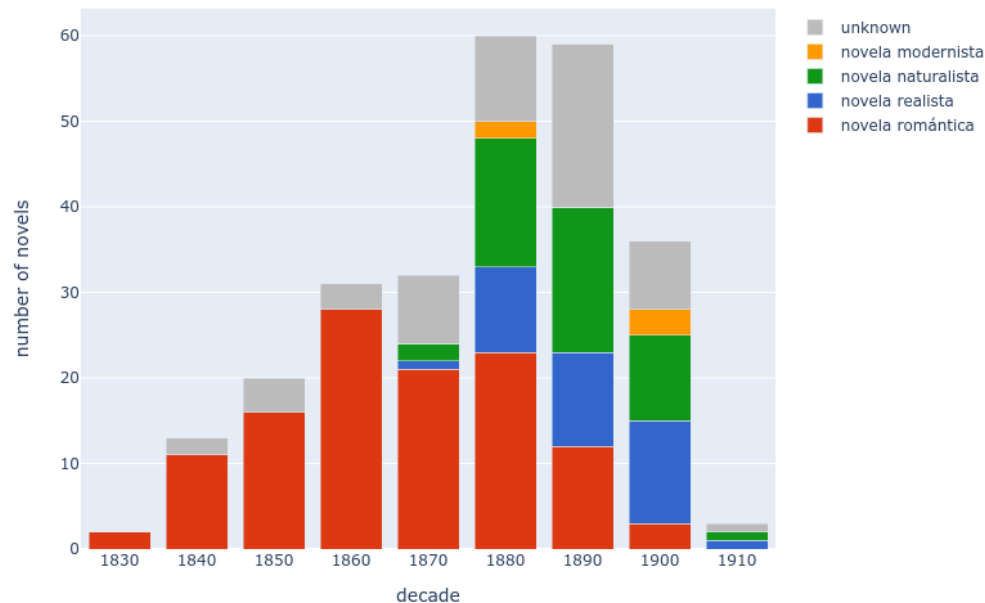
Novels by decade and thematic subgenre



# Anwendung: Klassifikation literarischer Gattungen

## Metadaten: Zuordnung von literarischen Strömungen

Novels by decade and literary current



# Anwendung: Klassifikation literarischer Gattungen

## Metadaten: Modellierung der Gattungsinformationen

```
<bibl xml:id="W563">
  <author key="A227">Ortega, Enrique</author>
  <title>Los casamientos del diablo</title>
  <term type="subgenre.title.explicit">novela histórica americana</term>
  <term type="subgenre.title.explicit.norm" resp="#uhk">novela histórica</term>
  <term type="subgenre.title.explicit.norm" resp="#uhk">novela americana</term>
  <term type="subgenre.title.explicit.norm" resp="#uhk">novela</term>
  <term type="subgenre.title.implicit" resp="#uhk">novela sentimental</term>
  <term type="subgenre.title.implicit" resp="#uhk">novela romántica</term>
  <term type="subgenre.summary.signal.explicit" resp="#uhk">novela histórica</term>
  <term type="subgenre.summary.signal.explicit" resp="#uhk">novela americana</term>
  <term type="subgenre.summary.signal.explicit" resp="#uhk">novela</term>
  <term type="subgenre.summary.signal.implicit" resp="#uhk">novela sentimental</term>
  <term type="subgenre.summary.signal.implicit" resp="#uhk">novela romántica</term>
  <term type="subgenre.summary.theme.explicit" resp="#uhk" cligs:importance="2">novela
    histórica</term>
  <term type="subgenre.summary.theme.implicit" resp="#uhk">novela sentimental</term>
  <term type="subgenre.summary.current.implicit" resp="#uhk">novela romántica</term>
  <term type="subgenre.summary.identity.explicit" resp="#uhk">novela americana</term>
  <term type="subgenre.summary.mode.reality.explicit" resp="#uhk">novela histórica</term>
  <term type="subgenre.summary.mode.representation.explicit" resp="#uhk">novela</term>
  <country>Argentina</country>
  <note type="source">
    <ptr target="#Lichtblau_1997"/>
  </note>
</bibl>
```

## Anwendung: Klassifikation literarischer Gattungen

### Textanalyse: Klassifikationsergebnisse

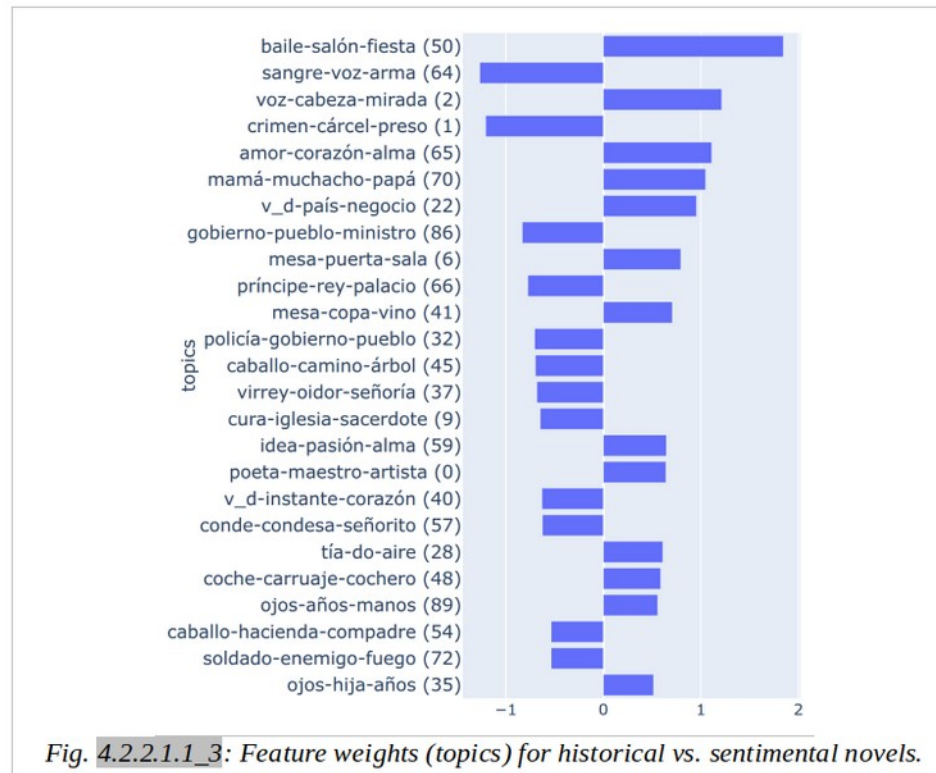
Subgenre 1	Subgenre 2	Top accuracy	Mean accuracy	SD accuracy	Top F1	Mean F1	SD F1
novela histórica	other	1	0.82	0.10	1	0.82	0.09
novela sentimental	other	1	0.81	0.12	1	0.80	0.13
novela de costumbres	other	1	0.71	0.14	1	0.72	0.16
novela histórica	novela sentimental	1	0.89	0.11	1	0.89	0.10
novela histórica	novela de costumbres	1	0.84	0.10	1	0.84	0.10
novela sentimental	novela de costumbres	1	0.77	0.13	1	0.75	0.16

Tab. 4.2.2.1.1\_2: Classification results for primary thematic subgenres (SVM, 90 topics, optimization interval of 250).



# Anwendung: Klassifikation literarischer Gattungen

## Textanalyse: Klassifikationsergebnisse



## Anwendung: Klassifikation literarischer Gattungen

### Textanalyse: Klassifikationsergebnisse

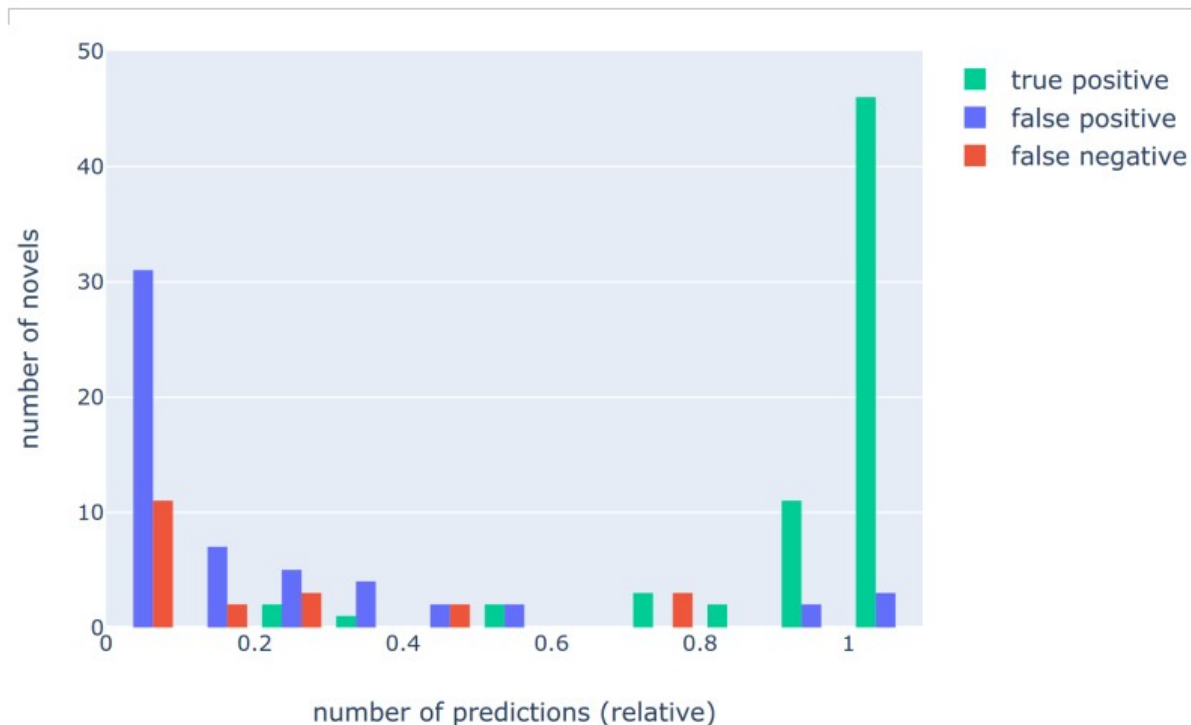


Fig. 4.2.2.1.1\_9: Predictions for "novela histórica" vs. "other" (topics).

# Anwendung: Klassifikation literarischer Gattungen

## Textanalyse: Klassifikationsergebnisse

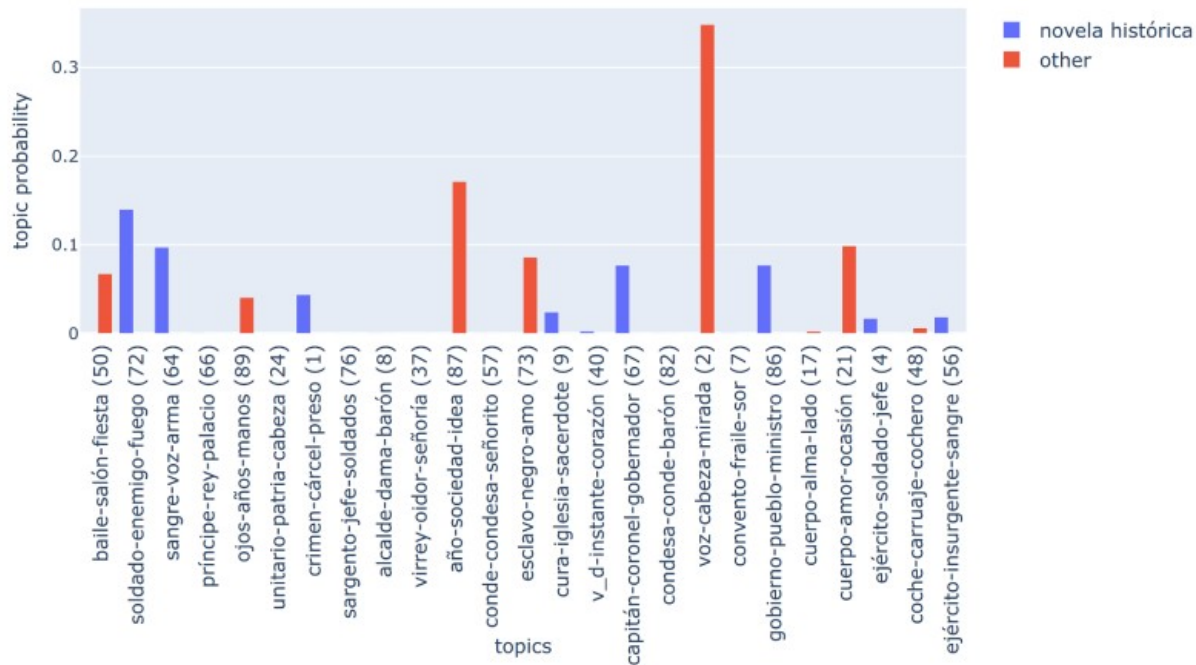


Fig. 4.2.2.1.1\_14: Top topics for "novela histórica" vs. "other" in the novel "Vía Crucis".

## Anwendung: Mikroanalyse von Stilvariation

### Stilometrische Analyse literarischer Ästhetik in Roberto Bolaño's 2666

- Studie gemeinsam mit Robert Hesselbach, Friedrich-Alexander-Universität Erlangen-Nürnberg
- Bolaño: chilenischer Autor (1953-2003)
- wichtigstes Werk: Roman 2666 (posthum 2004 erschienen)
- 5 Teile: komplexer Inhalt, stilistische Variation, insb. Teil 4: „La parte de los crímenes“
- Ansatz: “microanalysis of style variation” (Hoover 2017, Krautter 2018)
  - wahrgenommene Stilunterschiede durch Vergleich der Teile prüfen
  - Methoden: MFW, N-Gramme (Stylo), Satzlängen, Anteil direkter Rede, Komplexität des Vokabulars



## Anwendung: Topic Modeling für Digitale Editionen

- in Zusammenarbeit mit Frederike Neuber (BBAW)
- Topic Modeling: Verfahren der digitalen Textanalyse zur Ermittlung von Themen, Topoi, Motiven
- Fragen:
  - Wo wird das schon in digitalen Editionen eingesetzt?
  - Warum könnte es da sinnvoll eingesetzt werden?
  - Was sind Herausforderungen und Bedarfe?
- Zwei Korpora: Jean Paul (1763-1825), Briefe Uwe Johnson (1934-1984), bisher je ca. 1.000 Briefe

## Anwendung: Topic Modeling für Digitale Editionen

- Beispieltopics Johnson-Korpus, verschiedene Topic-Modelle mit je 20 Topics



## Anwendung: Topic Modeling für Digitale Editionen

- Brief aus dem Johnson-Korpus im Text-Viewer des Tools „SerendipSlim“



Text: uj0097

Beginning of document

Gross Friedenau 24 . Dezember , 1968 Liebe Johanna , wie gewünscht wankt meine Fassung unter dem Eindruck Ihrer Namensgeschichte , um so mehr , als deren Verzweigungen ja nicht nur durch gewöhnliches Wachstum verursacht scheinen , sondern eigentlich durch die Zunahme der Dimensionen , in denen Sie anderen im Schlaf und auf der Treppe erscheinen können . Zum Aussuchen nähme ich mir die unausweichlichen , zum Beispiel die , in denen sich die Gegenwehr der Schwester oder die Hilflosigkeit der Eltern ausdrückt : Montsch , das Kind . Für die meisten von den anderen fehlt mir sogar der referierende Anlass , und es ist lediglich Prinzipientreue von der geschmäckerlichen Art , die mich jeglicher Absicht bei der Schaffung neuer Volkskultur misstrauen lässt . Vorziehen wenn ich dürfte täte ich noch jene Benennungen , die einen Respekt ausdrücken , etwa jene entfernte Verwandte Judith , weil sie den Holofernes schliesslich zur Strecke brachte . Der neueste Vorschlag , Schnörkel , will mir bei aller Wissenschaftlichkeit , selbst bei dem gefühlvollen Zugeständnis ans Sächsische , geringschätzig vorkommen . Er legt Entbehrlichkeit nahe , wengleich der Sinn für nette Dekoration nicht zu kurz kommt . Die Nichtswürdigkeit dieser Bemerkungen ergibt sich aus dem Namen ihres Urhebers , der genannt ist nach Einem , den es nicht gab , und dessen Hervorbringungen gefälschte sind , untergeschobene : Ossian , als der Sie mit den vorsichtigsten Wünschen zum Neuen Jahr begrüsst , liebe Johanna , Ihr Durch Boten .



## Fazit

- Qualitative und quantitative Textanalysen folgen verschiedenen Paradigmen: trotzdem – Abstufungen und Verbindungen
  - Bei „distant reading“ ist Nähelektüre, also „close reading“, sehr oft Voraussetzung, z.B. für Metadatenerfassung und Hypothesenbildung
  - Aber auch immer wieder im Analyseprozess von Bedeutung (Arbeitsschritte der digitalen Textanalyse)
  - Nicht zuletzt: wichtig für die Interpretation und Einordnung von Ergebnissen
- In der Praxis ist beides in der digitalen Textanalyse ohnehin miteinander verbunden
  - Verbindung kann Erkenntnispotential und Qualität von Analysen steigern
  - Verbindung kann gezielt gesucht werden

## Literaturhinweise

- Craig, H. (2013, September). Middle-distance reading with information-theory metrics. (Talk abstract). Retrieved from <http://www.gcdh.de/en/events/calendar-view/prof.-hugh-craig-middle-distance-reading-with-information-theory-metrics1/>
- forText (2016ff.): forText. Literatur digital erforschen. <https://fortext.net>
- Herrmann, J. Berenike (Hg.) (2012): Literatur rechnen. Ein Wiki zur digitalen Textanalyse. [Digitale Ressource]. <http://litre.uni-goettingen.de/index.php>
- Herrmann, J. Berenike (2017): "In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics", in: Digital Humanities Quarterly 11 (4). <http://www.digitalhumanities.org/dhq/vol/11/4/000341/000341.html>
- Jannidis, Fotis (2010): "Methoden der computergestützten Textanalyse", in: Methoden der literatur- und kulturwissenschaftlichen Textanalyse. Hrsg. von Ansgar Nünning und Vera Nünning. Stuttgart und Weimar: Metzler: 109-132.
- Lauer, Gerhard (2009): "Literatur rechnen. Lektüre im Computerzeitalter", in: FAZ.NET. <https://www.faz.net/aktuell/feuilleton/geisteswissenschaften/literatur-rechnen-lektuere-im-computerzeitalter-1840973.html>
- Moretti, Franco (2000): "Conjectures on World Literature", in: New Left Review 1: 54-68. <https://newleftreview.org/issues/ii1/articles/franco-moretti-conjectures-on-world-literature>