

Schulunterricht: Textvisualisierung mit Voyant unterrichten

Marie Flüh ¹

1. Universität Hamburg

forTEXT

Thema:	Textvisualisierung	DOI:	10.48694/fortext.3774
Jahrgang:	1	Ausgabe:	5
Erscheinungsdatum:	2024-08-07	Erstveröffentlichung:	2019-10-07 auf fortext.net
Lizenz:			open  access

Allgemeiner Hinweis: Rot dargestellte **Begriffe** werden im Glossar am Ende des Beitrags erläutert. Alle externen Links sind auch am Ende des Beitrags aufgeführt.

Eckdaten der Unterrichtseinheit

- Thema der Sitzung: Themen und Motive in Gedichten von der Klassik bis zur Gegenwart – ein digitaler Einstieg in die Analyse von Gedichten
- Klassenstufe: Sekundarstufe II, Literaturunterricht im Fach Deutsch
- Phasen: Einführung in die Methodik der quantitativen Textanalyse anhand einer **Wordcloud** zu Friedrich Schillers Ballade *Die Bürgschaft* (1798) und kritische Diskussion der Visualisierung; Einführung in die Arbeit mit dem Textanalysetool Voyant (Lehrvortrag und Präsentation eines Tutorials über Funktionen von Voyant); Organisation der Partnerarbeit; Präsentation und Diskussion der Ergebnisse im Plenum
- Einbettung der Stunde in die Unterrichtseinheit: Interpretation lyrischer Texte, Einstieg in Unterrichtseinheit zur Lyrikanalyse, Distant Reading (vgl. **Distant Reading**) mit repräsentativen Gedichten von der Klassik bis zur Gegenwart
- Sozialformen: Vortrag, Diskussion im Plenum, Partnerarbeit
- Medien/Materialien: Lernende sollten max. zu zweit an einem PC arbeiten, Lehrende benötigen einen Laptop und Beamer. Alle von uns bereitgestellten Materialien erhalten sie unter: <https://web.archive.org/save/https://zenodo.org/records/10519367>
- Dauer der Unterrichtseinheit: 90 Minuten
- Schwierigkeitsgrad des Tools: einfach

Fachliche Lernziele:

- Sprachliche Handlungsfähigkeit: Ergebnispräsentation in Form strukturierter Vorträge, Reflexion über Sprache, indem sprachliche Einheiten und inhaltliche Schwerpunkte lyrischer Texte zunächst mit Voyant herausgefiltert (vgl. **Text Mining**) und isoliert betrachtet und anschließend in Nahtsicht (vgl. **Close Reading**) detailliert sowie kontextsensitiv untersucht werden
- Auseinandersetzung mit Texten und Medien: Auseinandersetzung mit literarischen Texten in digitaler Form, Texterschließung hier als analytische quantitative Korpusanalyse (vgl. **Korpus**) und „Blick von oben“ auf repräsentative, poetische Texte
- Aufbau von Methodenkompetenz: Kennen und Anwenden des Textanalystools Voyant, Auswertung von Grafiken und statistischen Daten, kritische Reflexion der Ergebnisse, Erstellen von Interpretationshypothesen
- Aufbau von Orientierungswissen: Die Schüler*innen lernen erste Besonderheiten der Gattung Lyrik aus der Zeit der Klassik bis zum beginnenden 20. Jhd. kennen

KMK-Lernziele:

- Suchen und Verarbeiten (Auswerten und Bewerten): Daten analysieren, auswerten, interpretieren und kritisch bewerten
- Produzieren und Präsentieren: Nutzung von Voyant als technisches Bearbeitungswerkzeug zur Textvisualisierung; Erkennen der Bedeutung von Urheberrecht und geistigem Eigentum literarischer Texte
- Problemlösen und Handeln: Bedarfsgerechter Einsatz von Voyant als Werkzeug zur Textanalyse; Erkennen und Formulieren der algorithmischen Struktur von Voyant
- Analysieren und Reflektieren: Nutzung und kritische Bewertung von Voyant als Gestaltungsmittel, Verstehen und Reflektieren der Medien in der digitalen Welt; Erkennen und Bewerten der Chancen und Risiken des Gebrauchs von Distant-Reading-Verfahren

Bausteine

- Verlaufsraaster der Unterrichtseinheit: Aus welchen Phasen setzt sich die Unterrichtsstunde zusammen? Dem Verlaufsplan entnehmen Sie Inhalte und Schwerpunkte.
- Anwendungsbeispiel: Anhand welcher Texte führen Sie in die digitale Analyse von Gedichten ein? Leiten Sie die Lerngruppe dazu an, mit einem Textanalysetool eine literarische Großform digital zu untersuchen.
- Verlauf der Unterrichtseinheit(en): Wie sieht die konkrete Ausgestaltung der Phasen aus und welche Arbeitsschritte werden vorgenommen? Erfahren Sie, wie die Unterrichtseinheit strukturiert ist und welche Beispielaufgaben Sie Ihren Schüler*innen stellen können.
- Lösungen zu den Beispielaufgaben: Hat die Lerngruppe die Beispielaufgaben richtig gelöst? Hier finden Sie Antworten.

Verlaufsraaster der Unterrichtseinheit

Phase	Impulse des/der Lehrenden	Erwartete Aktivität der Lernenden	Sozialform	Medien / Materialien
Vorab und Einstieg (ca. 10 Min.)	Präsentation der Wordcloud und Leitung der Diskussion	Interpretation der Wordcloud, Abgleich mit eigenen Lektüreerfahrungen	Diskussion im Plenum	Beamer/Smartboard, Laptop, PowerPoint-Präsentation Voyant
Problematisierung (ca. 15 Min.)	Impuls: Wie werden Wordclouds erstellt? Hinleitung zum Paradigma Distant Reading, Vortrag und Videopräsentation	Mutmaßungen/Erklärungen über die Zusammensetzung	Diskussion im Plenum, Lehrvortrag	Beamer/Smartboard, Laptop, PowerPoint-Präsentation Voyant
Erarbeitung (ca. 40 Min.)	Gruppeneinteilung, technische Fragen beantworten, Koordination der Toolauswahl, um Überschneidungen zu vermeiden	Lösen der Aufgaben auf dem Arbeitsblatt „Literatur digital erforschen“, Vorbereitung eines Kurzvortrags, in dem ein Tool aus der Voyant-Toolsuite und die Ergebnisse der Korpusanalyse vorgestellt werden	Partnerarbeit	Arbeitsbogen „forTEXT_Textvisualisierung_Voyant_AB.pdf“ (forTEXT 2019), PC-Ausstattung 1:2, Beamer, Laptop
Sicherung (ca. 20 Min.)	Leitung Ergebnisvergleich, Sammlung an der Tafel: Welche Aussagen werden über das Korpus gemacht?	Vorstellung der Ergebnisse	Vortrag, Diskussion im Plenum	Smartboard, Tafel oder Whiteboard
Reflexion & Transfer				

Phase	Impulse des/der Lehrenden	Erwartete Aktivität der Lernenden	Sozialform	Medien / Materialien
<i>(Hausaufgabe oder Folgestunde)</i>	Überleitung zum Close Reading	Lesen unterschiedlicher Gedichte	Einzelarbeit	Gedicht aus Korpus in ausgedruckter Form

Das Verlaufsraaster stellen wir Ihnen als PDF-Datei („forTEXT_Textvisualisierung_Voyant_LM_Verlaufsraaster.pdf“) auf Zenodo zur Verfügung (forTEXT 2019).

1. Anwendungsbeispiel

Die hier skizzierte Doppelstunde ist die erste Stunde einer Unterrichtseinheit in der Sekundarstufe II zum Thema Lyrikanalyse, die einen festen Bestandteil des schulischen Kerncurriculums für das Fach Deutsch darstellt. Im Literaturunterricht der gymnasialen Oberstufe für das Fach Deutsch ist die Auseinandersetzung mit Werken von Autor*innen, die epochemachend waren, vorgesehen. Neben dem Aufbau von Orientierungswissen über Werk und Wirken kanonisierter Autor*innen sollen in der Oberstufe anhand unterschiedlicher Textgattungen fachspezifische Kompetenzen aufgebaut werden. Als verbindliche Referenztexte, anhand derer die Interpretation literarischer Texte der Gattung Lyrik unterrichtet werden können, werden repräsentative Gedichte von der Klassik bis zur Gegenwart angeführt. An diese Rahmenbedingungen knüpft die Konzeption der Unterrichtsstunde „Themen und Motive in Gedichten von der Klassik bis zur Gegenwart – Ein digitaler Einstieg in die Analyse von Gedichten“ an.

Im Hintergrund der Voyant-Tools stehen zwei Ansätze: Distant Reading und Textvisualisierung (Horstmann und Stange 2024). Distant-Reading-Verfahren eignen sich sowohl für die Exploration großer Textmengen (z. B. das gesamte Œuvre von Autor*innen) als auch für die Analyse vergleichsweise kleiner Textmengen (wie einzelner Werke). Im Kern geht es um die statistische Auswertung – d. h. die quantitative Analyse – von Textdaten.

Textvisualisierung ist ein Teilbereich der Informationsvisualisierung. Diese befasst sich mit der visuellen Repräsentation komplexer Texte und der Anpassung dieser Darstellungsformen durch interaktive Softwares wie Voyant (Flüh 2024a). Beide Methoden werden in diesem Fall eingesetzt, um ein aus 40 Gedichten bestehendes Lyrikkorpus zu explorieren und Hypothesen über Besonderheiten und häufig vorkommende Themen in lyrischen Texten zu bilden. Um das Textkorpus zu erstellen, wurden u. a. die folgenden Anthologien konsultiert: Bode (2018); Detering (2007); Gnüg und Möhrmann (1999); Hahn (2008). Die einzelnen und individuell auswählbaren Voyant-Tools visualisieren die Ergebnisse dieser quantitativen Textanalyse auf ganz unterschiedliche Art und Weise. Essentiell ist also nicht nur ein sicherer Umgang mit dem Tool unter Einbezug unterschiedlicher Funktionalitäten, sondern v. a. auch die kritische Auswertung der Grafiken und die Ableitung von Interpretationshypothesen.

Bei der Arbeit mit Voyant im Schulunterricht sollten immer Fragestellungen ausgewählt werden, die selbstständig bearbeitet und beantwortet werden können. Hierzu zählen z. B. Fragen nach übergeordneten Textinhalten (wie Themen und Motiven), der Entwicklung einer Figur und Personenbeziehungen oder der sprachlichen Gestaltung. Ein Mehrwert für Ihren Literaturunterricht entsteht dann, wenn auf Grundlage der statistischen Auswertung Muster erkennbar werden und Interpretationshypothesen gebildet werden können. Hierbei gilt: Die Ergebnisse einer rein quantitativen Abfrage reichen nicht aus, um stichhaltige Aussagen über die Konzeption lyrischer Texte zu machen. Sie eignen sich aber, um die Aufmerksamkeit auf bestimmte Textphänomene zu lenken, das Interesse für Besonderheiten lyrischer Texte zu wecken, den kritischen Umgang mit der Auswertung statistischer Erhebungen zu üben, eine Methode des Distant Readings kennenzulernen und v. a. um ausgehend von der Entdeckung eines auffälligen und interessanten Textphänomens in die tiefergehende Interpretation und Analyse von Gedichten einzusteigen.

Der Zeitraum, aus dem die Werke in dieser Unterrichtseinheit stammen, umfasst – abgeleitet aus dem Hamburger Lehrplan für Gymnasien – die Jahre von 1786 bis zur Gegenwart. Eine Aufschlüsselung der Zusammensetzung des Lyrikkorpus und das gesamte für die Unterrichtseinheit benötigte Korpus als TXT-Datei (vgl. [Reintext-Version](#)) finden Sie hier und am Ende der Unterrichtseinheit (Material: „forTEXT_Textvisualisierung_Voyant_Zusammensetzung_Lyrikkorpus.pdf“ und „forTEXT_Textvisualisierung_Voyant_Lyrikkorpus.txt“ (forTEXT 2019)). Die Stunde zielt darauf ab, mittels einer quantitativen Textanalyse und Textvisualisierung einen spielerischen und epochenübergreifenden Überblick über Themen, Motive und Schreibweisen der Lyrik der letzten Jahrhunderte zu geben.

Für die erste Auseinandersetzung mit der digitalen Methode der quantitativen Textvisualisierung untersuchen Sie in der Einstiegsstunde zunächst ein mittelgroßes Textkorpus aus 40 Gedichten mithilfe von Voyant, die aus unterschiedlichen Anthologien ausgewählt wurden. Bei der Zusammensetzung des Textkorpus, das Sie auf Zenodo (forTEXT 2019) herunterladen können, haben wir darauf geachtet, dass Autor*innen der enthaltenen Texte mehr als 70 Jahre verstorben sind. Dadurch gelten ihre Texte als gemeinfrei. Ist dies der Fall, können die Texte problemlos unter Wahrung des Urheberrechts für Ihren digital unterstützten Schulunterricht verwendet

werden. Im Rahmen einer kurzen Einführung, in der Sie das Korpus und die Methode vorstellen (siehe Material Präsentation: „forTEXT_Textvisualisierung_Voyant_Praesentation.pdf“ (forTEXT 2019)), weisen Sie auf diesen Aspekt hin, damit sich auch die Schüler*innen über die geltenden rechtlichen Rahmenbedingungen bewusst sind, die es bei jeglicher Form der digitalen Arbeit mit literarischen Texten zu bedenken gilt. Wenn Sie in Ihrem Unterricht ein eigens zusammengestelltes Korpus verwendet möchten, z. B. zu einer anderen Gattung, sollten Sie die folgenden Aspekte beachten:

- Um ein möglichst repräsentatives und für die quantitative Textanalyse geeignetes Textkorpus zu schaffen – also eine gute Stichprobe auszuwählen –, gelten dieselben Rahmenbedingungen wie in der Statistik: Die Ausgangsbedingungen entscheiden maßgeblich über die Aussagekraft des Ergebnisses. Wird schon bei der Auswahl eine Verzerrung eingeführt, schlägt sich das im Ergebnis nieder. Hierfür sollten Sie Ihre Schüler*innen sensibilisieren, wofür sich z. B. eine fachübergreifende Zusammenarbeit mit den Kolleg*innen aus der Mathematik umsetzen ließe.
- Einer Verzerrung der Daten (in der Statistik als systematische Fehler oder auch Bias bekannt) lässt sich u. a. durch eine zufällig gewählte Stichprobe entgegenwirken. Dennoch repräsentiert die Stichprobe nicht die Gesamtheit der Lyrik von der Klassik bis zur Moderne. Hierfür wäre ein deutlich größeres Textkorpus notwendig, das sich z. B. im Rahmen einer größer angelegten Projektarbeit erstellen und untersuchen ließe.
- Um weitere Verzerrungen zu vermeiden, sollten Sie bei der Korpuserstellung stets auf eine ausgewogene Verteilung zwischen Gedichten, die aus weiblicher Feder stammen, und solchen, die von Männern verfasst wurden, achten. Um eine ausgewogene Datengrundlage zu schaffen, sollten Sie darüber hinaus darauf achten, dass der jeweilige Einzeltextumfang nicht zu stark vom durchschnittlichen Umfang aller Texte abweicht.

2. Verlauf der Unterrichtseinheiten

2.1 Vorarbeiten

Bevor Sie Voyant in Ihrem Fachunterricht einsetzen, sollten Sie die grundlegenden Funktionalitäten kennen und beherrschen. Hierfür empfiehlt es sich, den Toolbeitrag zu Voyant (Flüh 2024a) zu lesen und im Anschluss die Lerneinheit zu Voyant (Flüh 2024b) Schritt für Schritt durchzuführen. Um vertiefendes Fachwissen über die Möglichkeiten der Textvisualisierung (Horstmann und Stange 2024) aufzubauen, nutzen Sie den entsprechenden Methodeneintrag.

Das beschriebene Lernszenario ist für Schüler*innen der Sekundarstufe II bestimmt, da die Benutzeroberfläche (vgl. GUI) und v. a. das *Voyant-Handbuch*, welches Informationen über die Funktionen der unterschiedlichen Tools beinhaltet, nur englischsprachig zur Verfügung steht. Die Unterrichtseinheit richtet sich also an eine Lerngruppe, die auf Englischkenntnisse zurückgreifen kann.

Bevor Sie die Stunde eröffnen, ist es sinnvoll, das Textkorpus „forTEXT_Textvisualisierung_Voyant_Lyrikkorpus.txt“ (forTEXT 2019) zentral auf dem Schulserver (vgl. Server) zu hinterlegen, sodass die Schüler*innen von ihrem Computer aus darauf zugreifen und es in Voyant hochladen können. Die Unterrichtsstunde sollte im Computerlabor durchgeführt werden. Sofern Sie über eine entsprechende Ausstattung an Laptops und auch in Ihrem Klassenraum über eine stabile WLAN-Verbindung verfügen, können Sie die Stunde natürlich auch hier durchführen. An einem Computer sollten nicht mehr als zwei Personen arbeiten, damit möglichst jedes Gruppenmitglied Arbeitsschritte mit Voyant umsetzen kann.

Ihre Schüler*innen sollten vorbereitend den Kommentar „Franco Moretti „Distant Reading“ – Ein Lektüre-Kommentar“ von Mareike Schumacher gelesen und als Hausaufgabe eine Definition von Distant Reading ausgearbeitet haben. Hieran knüpfen Sie in der Einstiegsphase der Unterrichtsstunde an. Um den Artikel herunterladen zu können und ihn an Ihre Schüler*innen verteilen zu können, bedarf es einer kostenfreien Anmeldung bei ZEITonline. Sofern Sie den Artikel an die gesamte Lerngruppe verteilen möchten, ohne sich anzumelden, senden wir Ihnen den Artikel zusammen mit den Musterlösungen zu.

2.2 Einstieg und Problematisierung

Ein visueller Impuls eignet sich, um die Motivation der Schüler*innen für ein Thema zu wecken und um abstrakte Unterrichtsgegenstände zu veranschaulichen. Das gilt nicht nur für die Naturwissenschaften, wo Grafiken und Schaubilder beinahe allgegenwärtig sind, sondern auch für den Literaturunterricht. Gerade die sprachliche Besonderheit literarischer Texte macht es für Schüler*innen nicht selten schwierig, in den Texten „verborgene“ abstrakte Inhalte – wie Informationen über den Inhalt oder nur implizit vermittelte Bedeutung – zu erkennen. Eine Visualisierung der Texte kann dabei helfen, in einem ersten Schritt das Augenmerk auf die Besonderheiten eines Textes zu richten, die sich in einem zweiten Schritt näher untersuchen lassen.

Für den geplanten Einstieg eignet sich ein Text, der den Schüler*innen – z. B. aus einer bereits abgeschlossenen Unterrichtseinheit – schon bekannt ist. In dieser Unterrichtseinheit verwenden wir eine Wordcloud von Friedrich

Schillers Ballade *Die Bürgschaft* (siehe Abb. 1 und in der Präsentation „forTEXT_Textvisualisierung_Voyant_Praesentation.pdf“ (forTEXT 2019) enthalten).



Wordcloud zu Friedrich Schillers *Die Bürgschaft* (1799)

Nutzen Sie das bestehende Wissen der Lerngruppe (Wissen über Inhalt und Konzeption der Ballade) und beziehen es auf den neuen Unterrichtsgegenstand (Textvisualisierung). Im Kern geht es darum, deutlich zu machen, dass Distant Reading ein Paradigma von Analyseverfahren bezeichnet, mit denen große Mengen an Textdaten ausgewertet werden, ohne dass der Text vorher gelesen wurde. Im Rahmen der Einführung überprüft die Lerngruppe die Zuverlässigkeit dieses Ansatzes. Leiten Sie nun eine Diskussion über die Visualisierung an, indem Sie die Schüler*innen Hypothesen über die Bedeutung der Visualisierung anstellen lassen und darüber diskutieren, ob die eigenen Lektüreerfahrungen sich mit der Wordcloud in Verbindung bringen lassen oder nicht. Hierbei sollte außerdem das aus dem Artikel erschlossene Wissen über den Distant Reading-Ansatz in die Diskussion einbezogen werden. Außerdem sollte deutlich werden, dass die Wordcloud nur ein mögliches Ergebnis darstellt, hinter dem eine quantitative Auswertung des Textes steht. Leiten Sie die Schüler*innen außerdem dazu an, die Visualisierung auch kritisch zu bewerten. Hier ist schließlich ein lyrischer Text, der zuvor in minutiöser Kleinarbeit analysiert und interpretiert wurde, in einzelne Teile zerlegt worden. Die ästhetische Wirkung, die sich beim Lesen der Ballade entfalten kann, geht hier verloren.

Mögliche **Leitfragen und Impulse** (Folie 3) – neben der Aufforderung zur Beschreibung der Wordcloud – sind:

- Wie könnte die Wordcloud entstanden sein?
- Beschreiben Sie, inwieweit Wordcloud und eigene Lektüreerfahrungen zueinander passen.
- Gibt es ein nicht-digitales Verfahren, um eine Wordcloud zu entwerfen?

Von der Vorstellung und Diskussion über eine hier exemplarisch vorgestellte Form der Textvisualisierung leiten Sie nun durch einen kurzen Impulsvortrag über zu der konkreten Umsetzung dieser Methode. In einem kurzen Lehrvortrag bereiten Sie die Schüler*innen auf die nächste Unterrichtsphase vor. Der Vortrag beinhaltet eine kurze Vorstellung des Lyrikkorpus (Folie Nr. 4) und der Vorstellung von Voyant. Hierfür spielen Sie das Voyant-Video vor (Folie Nr. 5), um die Lerngruppe mit den grundlegenden Funktionen von Voyant vertraut zu machen.

2.3 Erarbeitung

Anschließend teilen Sie die gesamte Schulklasse in Zweiergruppen ein und verteilen das Arbeitsblatt (Material: „forTEXT_Textvisualisierung_Voyant_AB.pdf“ (forTEXT 2019)), dessen Bearbeitung mit zentralen Funktionen von Voyant vertraut macht. Das Arbeitsblatt ist so konzipiert, dass zunächst die grundlegenden Funktionalitäten des Textanalysetools vermittelt werden und die Schüler*innen mit den Standardeinstellungen arbeiten. Weisen Sie Ihre Schüler*innen darauf hin, die Lösungen schriftlich in Stichworten festzuhalten, damit sich die Ergebnisse in einem späteren Teil der Stunde noch vergleichen und diskutieren lassen. Während die Schüler*innen die Aufgaben bearbeiten, nehmen Sie eine zurückhaltende Position ein. Interpretationshypothesen sollten eigenständig entwickelt werden, während bei technischen Problemen oder Fragestellungen Ihre Expertise gefragt ist.

2.4 Sicherung

In der Sicherungsphase stellen einige Zweiergruppen ihre Textanalyselandschaft und ihre Interpretationshypothesen vor (Aufgabe 6). Hierbei werden die zentralen Interpretationshypothesen von der nicht vortragenden Person an der Tafel, dem Smart- oder Whiteboard festgehalten. Der Vortrag sollte nicht länger als fünf Minuten dauern. Bei einer geschätzten Klassengröße von 20 Schüler*innen werden nicht alle Gruppen ihre Ergebnisse präsentieren können. Lassen Sie also drei bis vier Gruppen ausführlich vortragen und leiten dann die Frage nach weiteren, noch nicht genannten Interpretationshypothesen an das Plenum weiter und ergänzen das Tafelbild um neue Vorschläge.

2.5 Transfer & Reflexion

In der letzten Phase geht es darum, die erstellten Visualisierungen und die daraus abgeleiteten Interpretationen mit den tatsächlichen Lektüreerfahrungen in Verbindung zu setzen. Die Interpretationshypothesen sollen nun durch die tatsächliche Lektüre der Gedichte überprüft werden. Teilen Sie hierzu an jede Gruppe ein Gedicht aus (Material: „forTEXT_Textvisualisierung_Voyant_Material_Close_Reading.zip“ (forTEXT 2019)) und lassen dieses in Ruhe lesen. Sollte die Zeit nicht mehr ausreichen, können Sie die folgenden Aufgaben als Hausaufgabe formulieren:

- In welchen Gedichten kommen die Most Frequent Words „o“ (33), „leben“ (24), „gott“ (23) und „ach“ vor?
- Auf welche Art und Weise lassen sich die von dir vorgestellte Visualisierung und das gelesene Gedicht in Beziehung zueinander setzen?
- Wie verändert sich die Wahrnehmung der Texte, wenn du sie liest?

Einige der Gedichte sind kürzer als andere. Achten Sie bei der Verteilung der Gedichte also darauf, dass einige Gruppen ggf. zwei Gedichte erhalten. Besprechen Sie in der nächsten Stunde gemeinsam die Beispielaufgaben, ob die Interpretationshypothesen standhalten oder nicht. Im Unterrichtsgespräch sollten außerdem die Grenzen der quantitativen Textanalyse thematisiert werden. Für die Interpretation der Textvisualisierungen gilt: Es stehen weder der Lesegenuss noch die durch die Gedichte vermittelte ästhetische Wirkung im Vordergrund. Wenn Sie die Lektüreerfahrungen der Schüler*innen besprechen, sollte deutlich werden, dass es sich bei Distant- und Close-Reading-Verfahren um zwei völlig unterschiedliche Paradigmen handelt. Ersteres nähert sich der Interpretation lyrischer Texte durch eine vornehmlich quantitative Auswertung. Aspekte der ästhetischen Wirkung und zahlreiche gattungstypische Textmerkmale wie z. B. die Versstruktur – welche im Korpus ja entfernt wurde – spielt bei diesem Ansatz keine Rolle. Bei der genauen Lektüre der Gedichte sollte aber deutlich werden, dass die Art und Weise der Rezeption nicht zuletzt durch das Versmaß bestimmt wird. Der Blick „von oben“ auf die Textmenge vermag Motive und Themen zu verdeutlichen, die in lyrischen Texten epochenübergreifend vorkommen. Daraus lassen sich durchaus erste Interpretationshypothesen ableiten. Die Frage danach, welche Textmerkmale und -strukturen dazu führen, dass z. B. *Die Bürgschaft* bei Leser*innen ein Gefühl der Rührung entstehen lässt oder auf welche Art und Weise das Versmaß die Rezeption lyrischer Texte beeinflusst, sollte im Anschluss thematisiert werden.

3. Lösungen zu den Beispielaufgaben

Die Lösungsvorschläge erhalten Sie auf [Zenodo](https://zenodo.org/records/12530163) (forTEXT 2019).

Externe und weiterführende Links

- Arbeitsmaterialien zur Unterrichtseinheit: <https://web.archive.org/save/https://zenodo.org/records/12530163> (Letzter Zugriff: 18.06.2024)

- Kommentar „Franco Moretti“Distant Reading“ - Ein Lektüre-Kommentar“: <https://web.archive.org/save/https://lebelieberliterarisch.de/franco-moretti-distant-reading/> (Letzter Zugriff: 18.06.2024)
- Voyant-Handbuch: <https://web.archive.org/save/https://voyant-tools.org/docs/#!/guide/tools> (Letzter Zugriff: 18.06.2024)

Bibliographie

- Bode, Dietrich. 2018. *Deutsche Gedichte. Eine Anthologie*. Stuttgart: Reclam.
- Detering, Heinrich. 2007. *Reclams grosses Buch der deutschen Gedichte. Vom Mittelalter bis ins 21. Jahrhundert*. Stuttgart: Reclam.
- Flüh, Marie. 2024b. Lerneinheit: Textvisualisierung mit Voyant. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 5. Textvisualisierung (7. August). doi: 10.48694/fortext.3773, <https://fortext.net/routinen/lerneinheiten/textvisualisierung-mit-voyant>.
- . 2024a. Toolbeitrag: Voyant. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 5. Textvisualisierung (7. August). doi: 10.48694/fortext.3775, <https://fortext.net/tools/tools/voyant>.
- forTEXT. 2019. Textvisualisierung mit Voyant unterrichten. 7. Oktober. doi: 10.5281/zenodo.10519366, <https://zenodo.org/records/12530163>.
- Gnüg, Hiltrud und Renate Möhrmann. 1999. *Frauen Literatur Geschichte. Schreibende Frauen vom Mittelalter bis zur Gegenwart*. Stuttgart: Metzler.
- Hahn, Ulla. 2008. *Stechäpfel. Gedichte von Frauen aus drei Jahrtausenden*. Stuttgart: Reclam.
- Horstmann, Jan und Jan-Erik Stange. 2024. Methodenbeitrag: Textvisualisierung. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 5. Textvisualisierung (7. August). doi: 10.48694/fortext.3772, <https://fortext.net/routinen/methoden/textvisualisierung>.

Glossar

- Annotation** Annotation beschreibt die manuelle oder automatische Hinzufügung von Zusatzinformationen zu einem Text. Die manuelle Annotation wird händisch durchgeführt, während die (teil-)automatisierte Annotation durch **Machine-Learning-Verfahren** durchgeführt wird. Ein klassisches Beispiel ist das automatisierte **PoS-Tagging** (Part-of-Speech-Tagging), welches oftmals als Grundlage (**Preprocessing**) für weitere Analysen wie Named Entity Recognition (NER) nötig ist. Annotationen können zudem deskriptiv oder analytisch sein.
- Browser** Mit Browser ist in der Regel ein Webbrowser gemeint, also ein Computerprogramm, mit dem das Anschauen, Navigieren auf, und Interagieren mit Webseiten möglich wird. Am häufigsten genutzt werden dafür Chrome, Firefox, Safari oder der Internet Explorer.
- 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).
- Commandline** Die Commandline (engl. *command line interface* (CLI)), auch Kommandozeile, Konsole, Terminal oder Eingabeaufforderung genannt, ist die direkteste Methode zur Interaktion eines Menschen mit einem Computer. Programme ohne eine grafische Benutzeroberfläche (**GUI**) werden i. d. R. durch Texteingabe in die Commandline gesteuert. Um die Commandline zu öffnen, klicken Sie auf Ihrem Mac „cmd“ + „space“, geben „Terminal“ ein und doppelklicken auf das Suchergebnis. Bei Windows klicken Sie die Windowstaste + „R“, geben „cmd.exe“ ein und klicken Enter.
- CSV** CSV ist die englische Abkürzung für *Comma Separated Values*. Es handelt sich um ein Dateiformat zur einheitlichen Darstellung und Speicherung von einfach strukturierten Daten mit dem Kürzel `.csv`, sodass diese problemlos zwischen IT-Systemen ausgetauscht werden können. Dabei sind alle Daten zeilenweise angeordnet. Alle Zeilen wiederum sind in einzelne Datenfelder aufgeteilt, welche durch Trennzeichen wie Semikola oder Kommata getrennt werden können. In Programmen wie Excel können solche Textdateien als Tabelle angezeigt werden.
- Data Mining** Data Mining gehört zum Fachbereich **Information Retrieval** und bezieht sich auf die systematische Anwendung computergestützter Methoden, die darauf abzielt, in vorhandenen Datenbeständen Muster, Trends oder Zusammenhänge zu erkennen. Textbasierte Formen des Data Minings sind u. a. **Text Mining**, **Web Mining** und **Opinion Mining**.
- Distant Reading** Distant Reading ist ein Ansatz aus den digitalen Literaturwissenschaften, bei dem computationale 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.

- GUI** GUI steht für *Graphical User Interface* und bezeichnet eine grafische Benutzeroberfläche. Ein GUI ermöglicht es, Tools mithilfe von grafischen Schaltflächen zu bedienen, um somit beispielsweise den Umgang mit der **Commandline** zu umgehen.
- HTML** HTML steht für *Hypertext Markup Language* und ist eine textbasierte Auszeichnungssprache zur Strukturierung elektronischer Dokumente. HTML-Dokumente werden von **Webbrowsern** dargestellt und geben die Struktur und Online-Darstellung eines Textes vor. HTML-Dateien können außerdem zusätzliche **Metainformationen** enthalten, die auf einer Webseite selbst nicht ersichtlich sind.
- Information Retrieval** Die Teildisziplin der Informatik, das Information Retrieval, beschäftigt sich mit der computergestützten Suche und Erschließung komplexer Informationen in meist unstrukturierten Datensammlungen.
- Korpus** Ein Textkorpus ist eine Sammlung von Texten. Korpora (Plural für „das Korpus“) sind typischerweise nach Textsorte, Epoche, Sprache oder Autor*in zusammengestellt.
- Lemmatisieren** Die Lemmatisierung von Textdaten gehört zu den wichtigen **Preprocessing**-Schritten in der Textverarbeitung. Dabei werden alle Wörter (**Token**) eines Textes auf ihre Grundform zurückgeführt. So werden beispielsweise Flexionsformen wie „schneller“ und „schnelle“ dem Lemma „schnell“ zugeordnet.
- Machine Learning** Machine Learning, bzw. maschinelles Lernen im Deutschen, ist ein Teilbereich der künstlichen Intelligenz. Auf Grundlage möglichst vieler (Text-)Daten erkennt und erlernt ein Computer die häufig sehr komplexen Muster und Gesetzmäßigkeiten bestimmter Phänomene. Daraufhin können die aus den Daten gewonnen Erkenntnisse verallgemeinert werden und für neue Problemlösungen oder für die Analyse von bisher unbekannten Daten verwendet werden.
- Markup Language** Markup Language bezeichnet eine maschinenlesbare Auszeichnungssprache, wie z.B. **HTML**, zur Formatierung und Gliederung von Texten und anderen Daten. So werden beispielsweise auch **Annotationen** durch ihre Digitalisierung oder ihre digitale Erstellung zu Markup, indem sie den Inhalt eines Dokumentes strukturieren.
- Metadaten** Metadaten oder Metainformationen sind strukturierte Daten, die andere Daten beschreiben. Dabei kann zwischen administrativen (z. B. Zugriffsrechte, Lizenzierung), deskriptiven (z. B. Textsorte), strukturellen (z. B. Absätze oder Kapitel eines Textes) und technischen (z. B. digitale Auflösung, Material) Metadaten unterschieden werden. Auch **Annotationen** bzw. **Markup** sind Metadaten, da sie Daten/Informationen sind, die den eigentlichen Textdaten hinzugefügt werden und Informationen über die Merkmale der beschriebenen Daten liefern.
- Named Entities** Eine Named Entity (NE) ist eine Entität, oft ein Eigenname, die meist in Form einer Nominalphrase zu identifizieren ist. Named Entities können beispielsweise Personen wie „Nils Holgerson“, Organisationen wie „WHO“ oder Orte wie „New York“ sein. Named Entities können durch das Verfahren der Named Entity Recognition (NER) automatisiert ermittelt werden.
- Opinion Mining** Unter Opinion Mining, oder Sentiment Analysis, versteht man die Analyse von Stimmungen oder Haltungen gegenüber einem Thema, durch die Analyse natürlicher Sprache. Das Opinion Mining gehört zu den Verfahren des **Text Minings**.
- POS** PoS steht für *Part of Speech*, oder „Wortart“ auf Deutsch. Das PoS- **Tagging** beschreibt die (automatische) Erfassung und Kennzeichnung von Wortarten in einem Text und ist ein wichtiger **Preprocessing**-Schritt, beispielsweise für die Analyse von **Named Entities**.
- Preprocessing** Für viele digitale Methoden müssen die zu analysierenden Texte vorab „bereinigt“ oder „vorbereitet“ werden. Für statistische Zwecke werden Texte bspw. häufig in gleich große Segmente unterteilt (*chunking*), Großbuchstaben werden in Kleinbuchstaben verwandelt oder Wörter werden **lemmatisiert**.
- Reintext-Version** Die Reintext-Version ist die Version eines digitalen Textes oder einer Tabelle, in der keinerlei Formatierungen (Kursivierung, Metadatenauszeichnung etc.) enthalten sind. Reintext-Formate sind beispielsweise TXT, RTF und **CSV**.
- Server** Ein Server kann sowohl hard- als auch softwarebasiert sein. Ein hardwarebasierter Server ist ein Computer, der in ein Rechnernetz eingebunden ist und der so Ressourcen über ein Netzwerk zur Verfügung stellt. Ein softwarebasierter Server hingegen ist ein Programm, das einen spezifischen Service bietet, welcher von anderen Programmen (Clients) lokal oder über ein Netzwerk in Anspruch genommen wird.
- Text Mining** Das Text Mining ist eine textbasierte Form des **Data Minings**. Prozesse & Methoden, computergestützt und automatisch Informationen bzw. Wissen aus unstrukturierten Textdaten zu extrahieren, werden als Text Mining zusammengefasst.
- Type/Token** Das Begriffspaar „Type/Token“ wird grundsätzlich zur Unterscheidung von einzelnen Vorkommnissen (Token) und Typen (Types) von Wörtern oder Äußerungen in Texten genutzt. Ein Token ist also ein konkretes Exemplar eines bestimmten Typs, während ein Typ eine im Prinzip unbegrenzte Menge von Exemplaren (Token) umfasst.

Es gibt allerdings etwas divergierende Definitionen zur Type-Token-Unterscheidung. Eine präzise Definition ist daher immer erstrebenswert. Der Satz „Ein Bär ist ein Bär.“ beinhaltet beispielsweise fünf Worttoken („Ein“, „Bär“, „ist“, „ein“, „Bär“) und drei Types, nämlich: „ein“, „Bär“, „ist“. Allerdings könnten auch vier Types, „Ein“, „ein“, „Bär“ und „ist“, als solche identifiziert werden, wenn Großbuchstaben beachtet werden.

Web Mining Unter Web Mining versteht man die Anwendung von Techniken des **Data Mining** zur Extraktion von Informationen aus dem World Wide Web. Das Web Mining ist ein Teilbereich des Data Minings und zählt zu einem der wichtigsten Anwendungsgebiete für das **Text Mining**.

Wordcloud Eine *Wordcloud*, oder auch Schlagwortwolke, ist eine Form der Informationsvisualisierung, beispielsweise von Worthäufigkeiten in einem Text oder einer Textsammlung. Dabei werden unterschiedlich gewichtete Wörter, wie die häufigsten Wörter, i.d.R. größer oder auf andere Weise hervorgehoben dargestellt. Die horizontale/vertikale Ausrichtung und die Farbe der dargestellten Wörter hat meistens allerdings keinen semantischen Mehrwert.