

Lerneinheit: Sentimentanalyse mit SentText

Marie Flüh  ¹

1. Universität Hamburg

forTEXT

Thema:	Sentimentanalyse	DOI:	10.48694/fortext.3798
Jahrgang:	1	Ausgabe:	7
Erscheinungsdatum:	2024-07-10	Erstveröffentlichung:	2020-06-29 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 Lerneinheit

- Anwendungsbezug: Lexikonbasierte Sentimentanalyse bedeutender griechischer Dramen (*Der gefesselte Prometheus* von Aischylos, *Die Wolken* von Aristophanes, *Medea* von Euripides), Epen (*Odyssee* und *Ilias* von Homer, *Theogonia* von Hesiodos) und Lyrik (*Der goldene Esel* von Apuleius, *Metamorphosen* von Ovid, *Auf die Liebe* von Anakreon)
- Methodik: Lexikonbasierte Sentimentanalyse
- Angewendetes Tool: SentText
- Lernziele: Zentrale Schritte der Korpuskonstituierung umsetzen, Hochladen der Texte und Kennenlernen der Analyseparameter, Analyse einzelner Texte und Interpretation der Analyseergebnisse, vergleichende Analyse der Homerischen Epen und Interpretation der Analyseergebnisse, manuelle Korrektur der Sentimentwerte (vgl. *Sentimentwert*), gattungstypologische Korpuskonstituierung, vergleichende Sentimentanalyse der drei Teilkorpora
- Dauer der Lerneinheit: 90 Minuten
- Schwierigkeitsgrad des Tools: einfach

Bausteine

- Anwendungsbeispiel: In welcher textuellen Gestalt erscheinen Empfindungen in Homers *Ilias* und welche Stimmung bestimmt die *Odyssee*? Welche Stimmungen bestimmen unterschiedliche Werke der griechischen Antike – und lassen sich gattungstypologische Differenzen ausmachen? Führen Sie eine lexikonbasierte Sentimentanalyse (Flüh 2024a) durch, bestimmen Sie den Sentimentgehalt (vgl. *Sentimentwert*) eines eigens erstellten Antike-Korpus und entwickeln Sie aus Ihren Analyseergebnissen bzw. -visualisierungen erste Interpretationshypothesen.
- Vorarbeiten: Wie gelangen Sie an Ihre Textbeispiele? Erfahren Sie, welche digitalen Repositorien Primärtexte für digitale Textanalysen zur Verfügung stellen und wie Sie die Texte herunterladen können.
- Funktionen: Welche Funktionen bietet Ihnen SentText (Flüh 2024b)? Lernen Sie die unterschiedlichen Module sowie deren Funktionen kennen und führen Sie eine Sentimentanalyse durch.
- Lösungen zu den Beispielaufgaben: Haben Sie die Beispielaufgaben richtig gelöst? Hier finden Sie Antworten.

1. Anwendungsbeispiel

In dieser Lerneinheit werden Sie anhand unterschiedlicher Schriftzeugnisse der griechischen Antike die wichtigsten Elemente einer lexikonbasierten Sentimentanalyse (Flüh 2024a) kennenlernen. Die Analyse zielt darauf ab, herauszufinden, welche Terminologie für die Beschreibung von Stimmungen (vgl. *Opinion Mininig*) in literarischen Texten unterschiedlicher Gattungen verwendet wird und welche Gefühlslagen die Erzählwelten bestimmen. Um unterschiedliche Analyseperspektiven auf sentimentbezogene Textmerkmale in dramatischen, epischen und lyrischen Texten der griechischen Antike werfen zu können, nutzen Sie das Textanalysetool SentText.

Als Textgrundlage dienen mehrere der wohl frühesten und bedeutendsten Epen, Dramen und Dichtungen der Weltliteratur, deren Wirkung bis in die Gegenwart reicht. Vor allem die Homerischen Epen gelten als Urquell der griechischen, römischen, byzantinischen und europäischen Kultur, deren Einfluss sich noch heute in der Denkweise der westlichen Welt widerspiegelt; Homers Werke bestimmen – mal mehr, mal weniger subtil – europäische Dichtung, Malerei, bildende Kunst und Literatur sowie Theater und Film (Latacz 2014). Gleichmaßen zählen die dramatischen Werke Aischylos', Aristophanes' und Euripides' sowie die Gedichte Apuleius', Ovids und

Anakreons bis heute zu prägenden Werken der Literaturgeschichte. Aufgrund dieser Tiefenwirkung finden sich deutschsprachige Übersetzungen aller genannten sowie diverser weiterer Autoren der griechischen Antike im ansonsten auf die Langzeitarchivierung originär deutschsprachiger Literatur ausgerichteten **Textgrid Repository**. Diese Lerneinheit befähigt Sie zum adäquaten Einsatz von **SentText**. Sie behandelt grundlegende Fragestellungen der literaturwissenschaftlich ausgerichteten Sentimentanalyse. Darauf aufbauend können Sie eigenen Forschungsfragen nachgehen und die im Rahmen der Analyse herausgefilterten emotionsbezogenen Signalwörter (vgl. **SBW**) bspw. als Grundlage für eine Emotionsanalyse heranziehen.

2. Vorarbeiten

Um die Methode kennenzulernen, der sich diese Lerneinheit widmet, lesen Sie zunächst den Methodeneintrag über Sentimentanalyse (Flüh 2024a). Lesen Sie hier (Flüh 2024b) im Anschluss die Besprechung des Textanalyse-tools **SentText** und machen sich auf diese Weise mit den wichtigsten Funktionen des Analysewerkzeugs vertraut. Die Korpusbildung (Bläß 2024) markiert neben der Herausbildung einer Forschungsfrage die Anfangsphase eines literaturwissenschaftlichen Projekts. In einem ersten Schritt erstellen wir also ein Antike-Korpus, welches Texte aller genannten Autoren beinhaltet.

Laden Sie sich die digitale Version der *Ilias* und der *Odyssee* herunter, indem Sie das **Textgrid Repository** aufrufen und dort eine Suchabfrage nach sämtlichen im Langzeitarchiv enthaltenen Texten Homers durchführen. Wählen Sie hierfür neben dem Eingabefeld für Suchabfragen am oberen rechten Bildschirmrand die Filterfunktion „Inhalte“ und dann „...nach Autor“ aus. Nun wird eine Liste aller im Repository enthaltener Autor*innen generiert. Scrollen Sie bis zu „Homer“, sehen Sie, dass das Archiv sechs seiner Werke beinhaltet. Um zu sehen, um welche Werke es sich handelt, klicken Sie auf „Homer“ (vgl. Abb. 1).

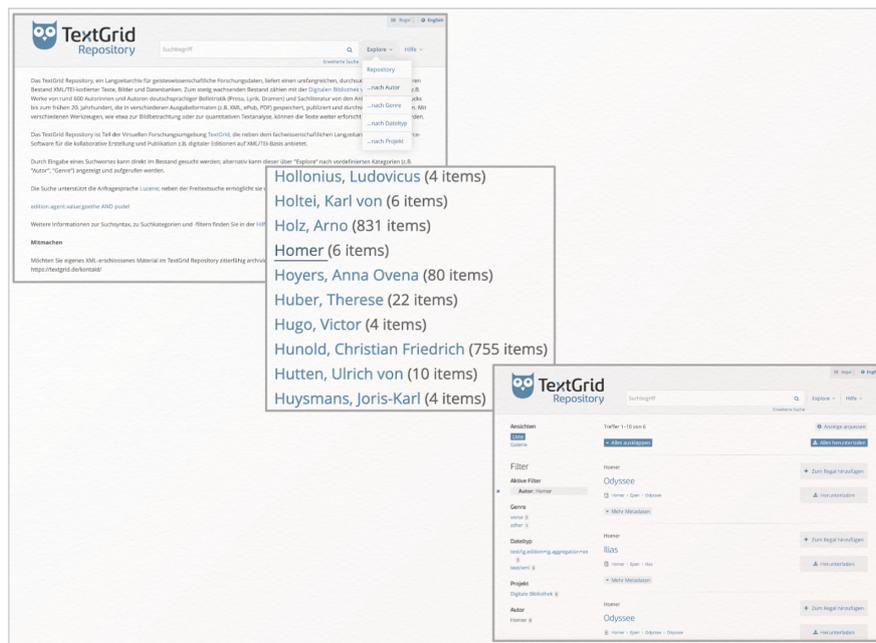


Abb. 1: Von der Suchabfrage bis zum Suchergebnis

SentText arbeitet ausschließlich mit Dateien im TXT- und XML-Format (vgl. **XML**; **Reintext-Version**). Am linken Bildschirmrand können Sie unter der Kategorie „Dateityp“ Ihre Suchabfrage weiter ausdifferenzieren. Klicken Sie in diesem Bereich auf „text/xml“ (vgl. Abb. 2).

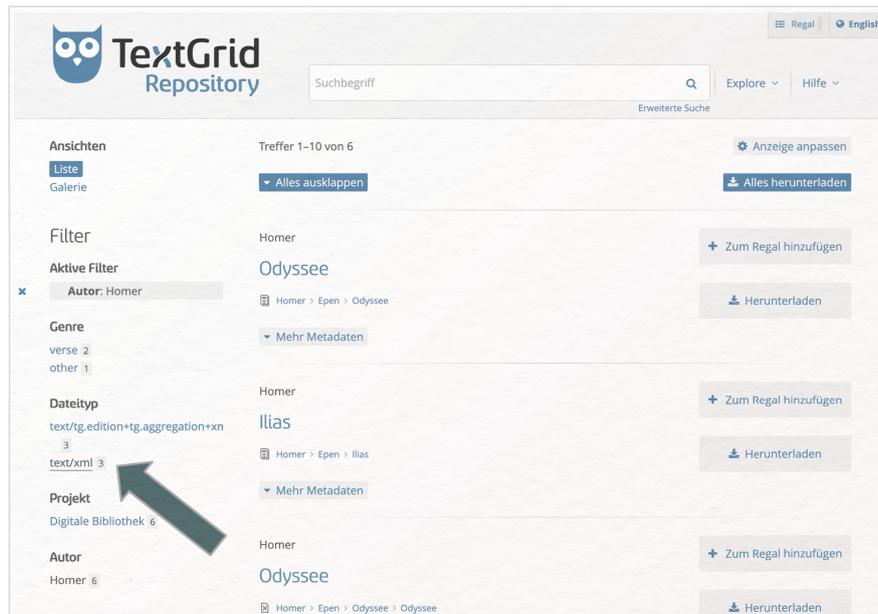


Abb. 2: Auswahl der Filterfunktion „Dateityp“ im TextGrid Repository

Jetzt erscheinen in der Listenansicht ausschließlich Dateien, deren Formate dem für unsere Analyse benötigten Typus entsprechen. An oberer Position der Liste wird bereits die *Odyssee* angeführt. Klicken Sie in dieser Kachel als nächstes auf den „Herunterladen“-Button. Nun öffnet sich ein neues Browserfenster, in dem die *Odyssee* im XML-Format angezeigt wird (vgl. Abb. 3). Wenn Sie etwas nach unten scrollen, wird deutlich, dass sich hinter allen – hier lilafarbenen – Programmierbefehlen unterschiedliche Textelemente verbergen.

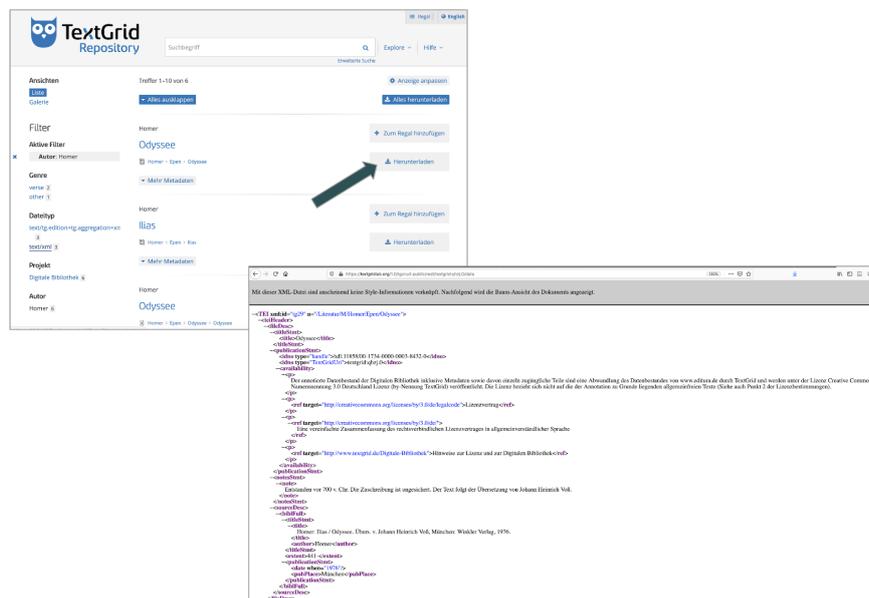


Abb. 3: Download der *Odyssee* im XML-Format

Um die Datei auf Ihrem PC zu speichern, klicken Sie am obersten Bildschirmrand in die Menüleiste Ihres Computers und wählen unter dem Reiter „Datei“ die Option „Seite speichern unter...“ aus (vgl. Abb. 4). Legen Sie in dem „Sichern unter“-Textfeld einen aussagekräftigen Namen für die Datei fest (Autor_Titel) und speichern diese auf Ihrem Desktop ab, indem Sie Ihre Auswahl per Klick auf „Sichern“ bestätigen (vgl. Abb. 4).

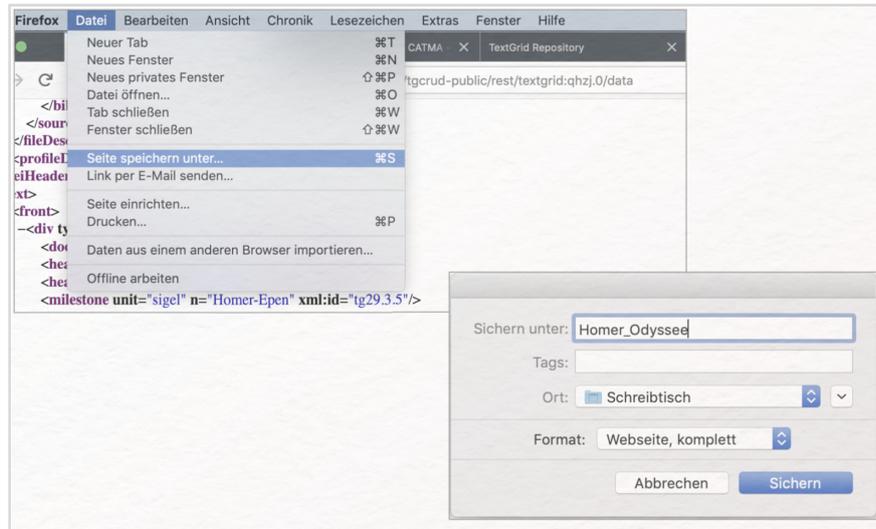


Abb. 4: Speichern der XML-Datei

Dieserart verfahren Sie auch mit der *Ilias*: Per Mausklick auf das Pfeilsymbol im oberen linken Bildschirmbereich kehren Sie von der XML-Ansicht der *Odyssee* in den Auswahlbereich des Textarchivs zurück. Direkt unter der soeben heruntergeladenen *Odyssee* finden Sie eine digitale Variante der *Ilias* im XML-Format. Wählen Sie nun in der *Ilias*-Kachel den „Herunterladen“-Button aus und speichern das XML-Dokument über die Menüleiste Ihres Computers. Sollten Sie einen anderen Browser verwenden, kann der Downloadprozess leicht variieren. Bei Microsoft Edge klicken Sie zuerst auf „Odyssee“, wählen in der linken Spalte unter „Herunterladen“ aus den aufgelisteten Dateitypen „plain text (txt)“ oder „Metadateien (xml)“ aus und laden die Datei mit einem Rechtsklick unter „Ziel speichern unter...“ herunter. Falls Sie mit Google Chrome arbeiten, öffnet sich das Fenster mit der txt- oder xml-Datei automatisch, anschließend können Sie den Text per Rechtsklick auf „Speichern unter...“ als xml-Datei speichern. Kehren Sie nun auf die Startseite des [TextGrid Repository](#) zurück und laden auf die gleiche Weise (1. Alphabetisch geordnete Autorenliste aufrufen; 2. Autorennamen auswählen; 3. Auswahl der Filterfunktion text/xml; 4. Herunterladen der XML-Datei; 5. Speichern auf dem Desktop) auch die übrigen der folgenden Texte herunter (vgl. Abb. 5):

Epiker	Dramatiker	Dichter
Homer: <i>Die Ilias</i>	Aischylos: <i>Der gefesselte des Prometheus</i>	Apuleius: <i>Der goldene Esel</i>
Homer: <i>Die Odyssee</i>	Aristophanes: <i>Die Wolken</i>	Ovid: <i>Metamorphosen</i>
Hesiodos: <i>Theogonie</i>	Euripides: <i>Medea</i>	Anakreon: <i>Auf die Liebe</i>

Abb. 5: Bestandteile des Antike-Korpus

Nun haben Sie die Bestandteile des Antike-Korpus im XML-Format auf Ihrem Desktop gespeichert und sind bereit für die Sentimentanalyse mit SentText. Das Tool arbeitet webbasiert und ohne Account-Pflicht. Folgen Sie [diesem Link](#), um auf die Startseite zu gelangen. Nun befinden Sie sich im „ABOUT“-Bereich, wo Tool und Funktionsweisen kurz vorgestellt werden. Lesen Sie die Einführung und kopieren im Anschluss probeweise die ersten Sätze des ersten Gesanges der *Odyssee* in das Demo-Feld:

„Sage mir, Muse, die Taten des vielgewanderten Mannes,
Welcher so weit geirrt nach der heiligen Troja Zerstörung,
Vieler Menschen Städte gesehn und Sitte gelernt hat [...].“

Aufgabe 1: Welche Wörter werden rot markiert, welche Wörter werden blau markiert und welche Stimmung assoziieren Sie mit den farblich hervorgehobenen Begriffen? Welches Wort scheint Ihnen eine besonders positive Gefühlslage zu repräsentieren und welches ist negativ konnotiert? Bringen Sie die Wörter – absteigend von positiv zu negativ konnotierten Wörtern – in eine Reihenfolge.

Aufgabe 2: Die Demoversion greift auf das Sentimentwörterbuch *SentimentWortschatz (SentiWS)* zurück. „heiligen“ hat hier einen Sentimentwert von 0.0727, der Sentimentwert von „gelernt“ beträgt 0.2492. „Zerstörung“ hat einen Sentimentwert von -0.3578. Passen die errechneten Sentimentwerte zu Ihrer persönlichen Reihenfolge?

In den Aufgaben 1 und 2 haben Sie die Grundlagen der lexikonbasierten Sentimentanalyse in sehr reduzierter Form kennengelernt: Der Computer vergleicht Textdaten – den ersten Gesang der *Odyssee* – und **Sentimentwörterbuch** – *SentiWS* – miteinander und legt für die Schnittmenge (sämtliche im Text und im Wörterbuch enthaltenen Wörter) die Polarität (Verortung auf einer Skala von positiv bis negativ) und den **Sentimentwert** fest. Zurück zu SentText! Die Startseite setzt sich aus drei Sparten zusammen. Unter „DOCUMENTATION“ finden Sie wichtige Informationen über verschiedene Parameter, die bei der Analyse berücksichtigt werden. Nehmen Sie sich einen Augenblick Zeit, um die unterschiedlichen Einflussgrößen zu begutachten und behalten Sie diese im Hinterkopf; sie bestimmen die Analyseergebnisse und -visualisierungen und beeinflussen dadurch schlussendlich auch Ihre Interpretation.

Klicken Sie nun auf den mittleren Reiter „SENTIMENT ANALYSIS“. Jetzt befinden Sie sich in dem Bereich, in dem Sie Ihre Texte hochladen und den Analyseprozess in Gang setzen (vgl. Abb. 6).

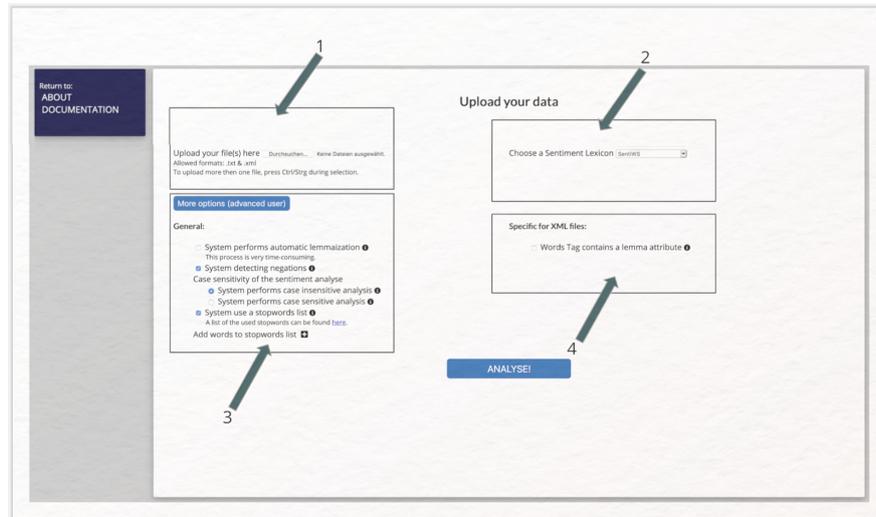


Abb. 6: Uploadbereich bei SentText: 1) Dateien auswählen und hochladen (obligatorisch); 2) Analysegrundlage festlegen (obligatorisch); 3) Analyseparameter anpassen (fakultativ); 4) Verarbeitung von XML-Dateien spezifizieren (fakultativ): Dateien im XML-Format sind deutlich umfangreicher aufgebaut als Dateien im TXT-Format. Sie wurden oft bereits bearbeitet und enthalten entsprechende Zusätze wie die Speicherung des Lemmas als Attribut. Ob diese Eigenschaften beim Import beibehalten werden sollen (höherer Rechenaufwand) oder nicht, steht Ihnen frei

Klicken Sie zunächst im Upload-Bereich auf „Durchsuchen...“ (vgl. Abb. 6, Punkt 1). Navigieren Sie zu Ihrem Desktop, wo Sie die neun XML-Dateien abgelegt haben. Aktivieren Sie die Strg- oder Ctrl-Taste und halten Sie diese gedrückt, während Sie per Mausclick alle neun Textdateien für die Analyse auswählen. Bestätigen Sie Ihre Auswahl via Mausclick auf „Öffnen“ (vgl. Abb. 7).

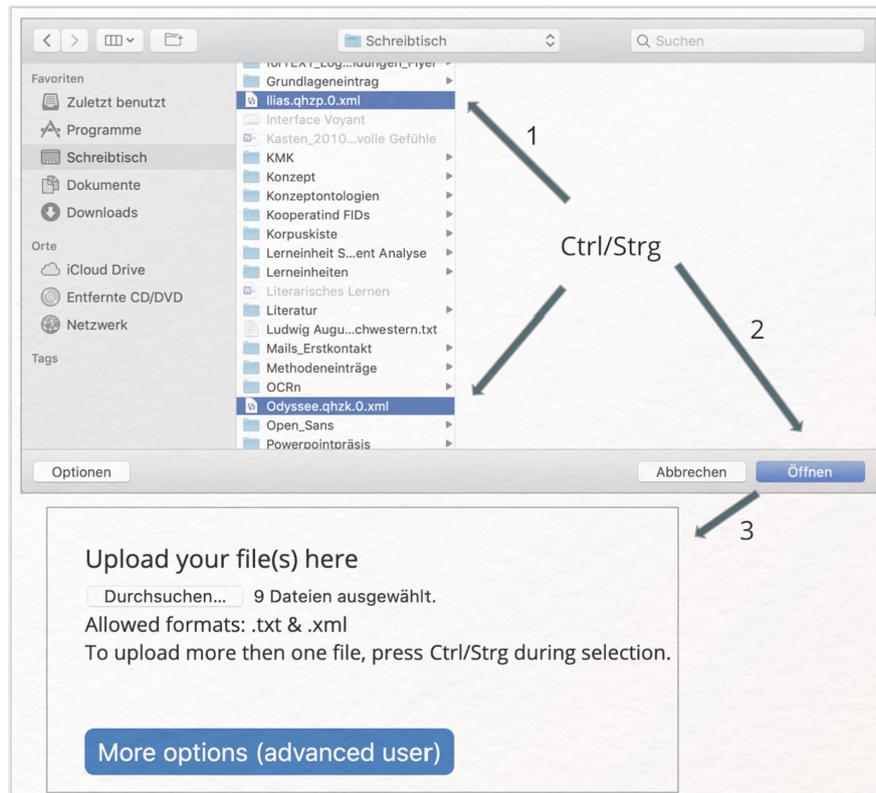


Abb. 7: Upload der Texte in drei Schritten: Auswahl beider Dateien bei gedrückter Ctrl-/Strg-Taste, bestätigen der Auswahl, Anzeige der Dateien im Sentimentanalysetool SentText

Wie Sie sehen, zeigt SentText nun die Anzahl der für die Analyse ausgewählten Dateien an. Rechts neben dem Upload-Bereich können Sie festlegen, auf welches Sentimentwörterbuch das Tool bei der Analyse zurückgreifen soll (vgl. Abb. 6, Punkt 2). Zur Auswahl stehen *SentiWS* der Universität Leipzig und die *Berlin Affective Wordlist – Reloaded* der Freien Universität Berlin. Darüber hinaus können Sie in diesem Bereich ein eigenes Wörterbuch einspeisen. Hierbei handelt es sich um ein wichtiges **Feature**, da sich auf diese Weise bspw. historisch gewachsene orthographische Besonderheiten in die Analyse einbeziehen lassen. In dieser Lerneinheit greifen wir zunächst auf *SentiWS* zurück, welches sich als verlässliche Grundlage für lexikonbasierte Sentimentanalysen literarischer Texte erwiesen hat (Schmidt und Burghardt 2018, 144 f.). Sie behalten die Voreinstellungen (vgl. **Default**) also bei. Klicken Sie nun auf den „More options (advanced user)“-Button (vgl. Abb. 7, Punkt 3). Hier sehen Sie unterschiedliche Einstellungen, mit denen Sie die Analyse in Zukunft anpassen können. In diesem Bereich können Sie u. a. die **Stoppwortliste** erweitern und auf diese Weise für die Analyse irrelevante Wörter ausschließen (bspw. Regieanweisungen, Auftritte, Abgänge oder andere Elemente des Nebentextes in Dramen). Für den ersten Durchlauf der Analyse greifen Sie auf die bereits aktivierten Voreinstellungen zurück. Starten Sie nun per Klick auf „ANALYZE!“ die Sentimentanalyse. Die Analyse aller Dateien dauert nur einen kurzen Moment (vgl. Abb. 8). Sollten Sie Dateien größeren Umfangs hochladen oder andere Analyseparameter auswählen, kann dieser Arbeitsprozess mehr Zeit in Anspruch nehmen.

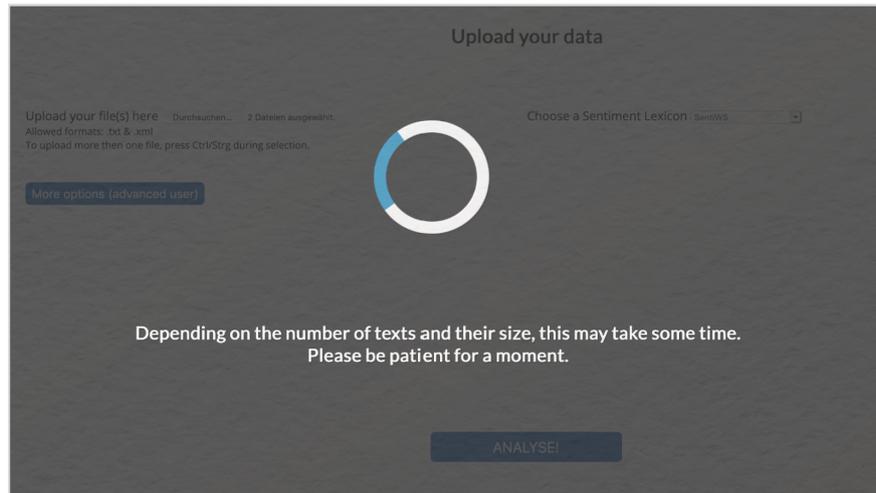


Abb. 8: Analyse des Antike-Korpus binnen weniger Sekunden

3. Funktionen

Im folgenden Teil lernen Sie die unterschiedlichen Module und deren Funktionen exemplarisch kennen. Wie Sie nun auf Ihrem Bildschirm sehen, besteht das Interface von SentText aus vier Panels oder Modulen (vgl. Abb. 9).

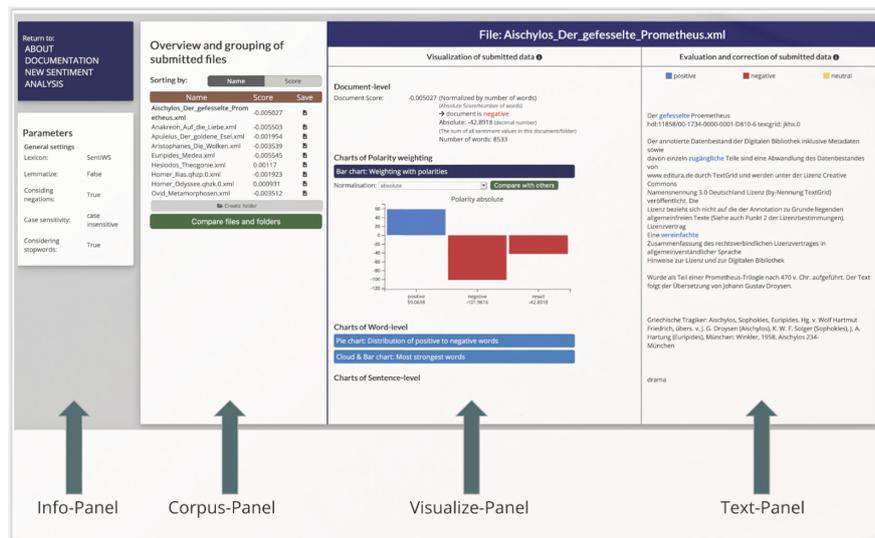


Abb. 9: Vier Panels in SentText; hier nicht sichtbar: das „Compare“-Panel

Im linksseitig angeordneten Info-Panel werden die Analyseparameter angezeigt. Im Gegensatz zu den drei übrigen Modulen erfüllt das Info-Panel ausschließlich eine Orientierung stiftende Funktion und ist nicht klickbar. Die Parameter der Analyse – also die variablen Rahmenbedingungen, unter denen sie durchgeführt wird – sollten Sie bei der Interpretation der Ergebnisse stets berücksichtigen bzw. transparent machen, um die Nachvollziehbarkeit Ihrer Ergebnisse zu gewährleisten. Im oberen Bereich des Panels können Sie per Mausklick auf „NEW SENTIMENT ANALYSIS“ eine neue Analyse mit anderen Grundeinstellungen starten. Achtung: Die Daten der aktuellen Analyse gehen dadurch verloren. Das Corpus-Panel – der Bereich unter „Overview and grouping of submitted files“ – dient der (Re)Organisation einzelner Textdateien zu Textkorpora und ermöglicht eine vergleichende Analyse mehrerer Textkorpora, die in separaten Ordnern angelegt werden. Im Visualize-Panel („Visualization of submitted data“) werden die Analyseergebnisse des im Corpus-Panel ausgewählten Dokuments in unterschiedliche Diagramme übersetzt (in diesem Fall Aischylos' *Der gefesselte Prometheus*, der der alphabetischen Anordnung aller hochgeladenen Dateien folgend am Anfang steht). Wählen Sie im Corpus-Panel einen anderen Text aus, erscheinen die entsprechenden Analyseergebnisse im Visualize-Panel. Hinter den ovalen blauen Schaltflächen im Visualize-Panel verbergen sich interaktive Diagramme, die verschiedene Analyseperspektiven auf den ausgewählten Text ermöglichen. Per Mausklick auf den gewünschten Diagrammtyp öffnet sich die Visualisierung. Sobald Sie im Corpus-Panel einen anderen Text auswählen, reagiert auch das

Visualize-Panel mit der Anzeige der entsprechenden Visualisierungen und das Text-Panel mit der Anzeige des jeweiligen Textes. Im Text-Panel („Evaluation and correction of submitted data“) sehen Sie die hochgeladenen XML-Dateien als Reintext, Sentiment-tragende Wörter (vgl. **SBW**) sind wie in der Demoversion ihrer Valenz entsprechend blau, rot und manchmal auch gelb eingefärbt. Darüber hinaus können Sie die Ergebnisse der Analyse in diesem Modul manuell korrigieren, indem Sie individuelle Sentimentwerte vergeben.

Vergleich einzelner Texte: Im Corpus-Panel werden Ihnen die hochgeladenen Texte in alphabetischer Reihenfolge angezeigt („Sorting by: Name“). Lassen Sie sich per Klick auf „Sorting by: Score“ eine nach Sentimentwerten (vgl. **Sentimentwert**) geordnete Übersicht erstellen (vgl. Abb. 10).



Abb. 10: Organisation der hochgeladenen Dateien in alphabetischer Reihenfolge (dunkelgrau hinterlegt) oder mit dem Sentimentwert des Dokuments entsprechender Anordnung (hellgrau hinterlegt)

Aufgabe 3: In welchen Texten herrscht eine positive, in welchen Texten eine negative Stimmung vor? Erstellen Sie eine Rangliste und nutzen Sie die empirischen Daten, um eine erste Interpretationshypothese über das Stimmungsbild abzuleiten.

Zur Erinnerung: Die Polarität wird bei *SentiWS* jeweils mithilfe einer Skala von -1 (maximal negativ) bis $+1$ (maximal positiv) ausgedrückt und auf vier Nachkommastellen gerundet. „Die Ausprägung der einzelnen Wörter wurden mittels der „Pointwise Mutual Information“-Methode (PMI) berechnet. Bei diesem Ansatz wird die semantische Ausprägung eines Wortes (SO: semantic orientation) aus semantischen Assoziationen des Wortes (SA: semantic association) zu ausgewählten Wörtern (sog. seed words: ausgewählte Wörter, die bei der Berechnung quasi als exemplarische Schlüsselwörter verwendet werden), die stark positiv oder negativ konnotiert sind, abgeleitet. Die semantische Ausprägung eines Wortes bezeichnet folglich den Grad der Verbindung eben dieses Wortes zu einem manuell erstellten Set an Wörtern, die positiv oder negativ konnotiert sind. Für jedes Wort wird berechnet, ob und wie stark es mit positiven und negativen Wörtern assoziiert wird.“ (s. Artikel 30 im Methodeneintrag Sentimentanalyse (Flüh 2024a)). Die Dezimalwerte in SentText sind teilweise deutlich höher. In diesen Fällen handelt es sich um den absoluten Sentimentwert, bei dem – bspw. für die Darstellung der *Strongest Words* – sämtliche Sentimentwerte addiert werden.

Aufgabe 4: Unterhalb der blauen Titelzeile im Visualize-Panel, in der der Dateiname der ausgewählten Datei angezeigt wird, können Sie einer kleineren Kopfzeile die Wortanzahl des Dokuments und die Gesamtanzahl der Sentiment-tragenden Wörter entnehmen. Wie lauten die absoluten Sentimentwerte (Addition der Werte aller Sentiment-tragenden Wörter, angegeben als Dezimalzahl) beider Homerischen Epen?

Um die Funktionen des Text-Panels auszuprobieren, wählen Sie im Corpus-Panel nun Aischylos' *Der gefesselte Prometheus* aus. Im Visualize-Panel öffnen sich dadurch unterschiedliche Visualisierungen, während im Text-Panel Aischylos' Drama erscheint. Scrollen Sie im Text-Panel durch den Text und hovern über die farblich markierten Wörter.

Aufgabe 5: Nutzen Sie die hinter dem Informationssymbol verborgenen Tooltips im Text-Panel. Welche Wörter werden gelb markiert – und beinhaltet der Text gelb hervorgehobene Wörter?

Aufgabe 6: Welchen Sentimentwert und welche Polarität haben „gefesselt“, „Fesseln“, „Einsamkeit“, „Kraft“ und „Gewalt“? Passen die generierten Polaritäten zu Ihren eigenen Assoziationen? Begründen Sie Ihre Antwort.

Wie Sie feststellen, ist es im Rahmen einer lexikonbasierten Sentimentanalyse nicht ohne Weiteres möglich, Doppeldeutigkeiten kontextsensitiv zu behandeln. Die Analyse bewegt sich ausschließlich auf der Textoberfläche des ausgewählten Textes und filtert sentimentbezogene Textmerkmale heraus. Durch den Abgleich mit *SentiWS* erhält „fesseln“ eine positive Polarität, obwohl der/die aufmerksame Leser*in weiß, dass hier der qualvolle Zustand Prometheus beschrieben wird. Anhand dieser Polaritätszuschreibung wird eine grundlegende Problematik sichtbar: Lexikonbasierte Ansätze werden der Spezifik literarischer Texte – sei es in Form einer kontextgebundenen Semantik, sei es in Form implizit im Text verankerter Gefühlsausdrücke oder einer ironischen Ausdrucksweise – nicht vollständig gerecht. Um falsch erkannten Wörtern die richtige Polarität zuordnen zu können, beinhaltet SentText eine manuelle Korrekturfunktion. Wie Sie im Zuge der Bearbeitung von Aufgabe 6 erkannt haben,

besitzt das Lexem „Fesseln“ einen negativen Sentimentwert von -0.0048 . Dieser Wert erscheint in Anbetracht der desaströsen Lage Prometheus auch für das titelgebende „gefesselte“ passend, welches einen positiven Sentimentwert erhalten hat. Um einen passenden Sentimentwert festzulegen, markieren Sie „gefesselte“. Nun öffnet sich das Dialogfeld, in dem Sie den Sentimentwert manuell eingeben können (vgl. Abb. 11). Jetzt haben Sie die Wahl: Entweder Sie entfernen die Markierung des ausgewählten Wortes als SBW via Klick auf „- Remove“ oder Sie legen einen neuen Sentimentwert fest.

Evaluation and correction of submitted data ⓘ

■ positive ■ negative ■ neutral

Edit sentiment of word: **gefesselte**

Current sentiment: 0.004

Add sentiment of word:

Add negation:

Der **gefesselte** Prometheus

Abb. 11: Manuelle Korrektur durch Auswahl des Wortes und Eingabe eines passenden Sentimentwerts

Tippen Sie in das Textfeld hinter „Add sentiment of word“ den negativen Sentimentwert -0.0048 ein und bestätigen Ihre Eingabe durch einen Klick auf den „Save sentiment“-Button. Wie Sie sehen, wird „gefesselt“ fortan rot markiert, weist den zugewiesenen negativen Sentimentwert auf (vgl. Abb. 12) und entspricht damit einer kontextsensitiv korrekten Semantik.

Evaluation and correction of submitted data ⓘ

■ positive ■ negative ■ neutral

Der **gefesselt** Prometheus

Abb. 12: Ergebnis der manuellen Korrektur

Ausgewählte Visualisierungen: Einen Großteil der Visualisierungen werden wir uns zu einem späteren Zeitpunkt im Rahmen einer vergleichenden Analyse anschauen. Zwei ausgewählte Formen lassen sich jedoch nur für einzelne Texte abrufen: Die „Table“- und die „Timeline“-Visualisierung. Beide Visualisierungen arbeiten auf der Satzebene und sind im „Visualize-Panel“ dementsprechend unter der Überschrift „Charts of Sentence Level“ zu finden. Klicken Sie hier auf „Timeline: Development of the Sentiment“ oder „Table: The strongest ten sentences“, öffnen sich die Darstellungen. Beide Visualisierungen interagieren mit dem Text-Panel und lassen sich modifizieren.

Aufgabe 7: Wählen Sie die „Table“-Visualisierung aus. Wie lautet der Satz mit dem negativsten Sentimentwert? Welche Figur äußert ihn und an wen richtet er sich?

Aufgabe 8: Wählen Sie die „Timeline“-Visualisierung aus. In welchem Textsegment finden sich besonders positive bzw. negative Sätze und wie lauten diese?

Vergleich zweier Texte: Nun möchten wir die Analyseergebnisse beider „Extremfälle“ etwas genauer unter die Lupe nehmen und die unterschiedlichen Formen der Visualisierung erkunden. Wählen Sie im Corpus-Panel zunächst Homers *Odysee* aus und richten Ihre Aufmerksamkeit nun wieder auf das Visualize-Modul. Über den grünen „Compare with others“-Button im Visualize-Panel können Sie mehrere Textdateien miteinander vergleichen. Wählen Sie die *Ilias* und die *Odysee* zum Vergleich aus, indem Sie zuerst auf die grüne „Compare with others“-Schaltfläche klicken. Nun werden Ihnen die Analyseergebnisse aller Einzeltexte angezeigt. Um aus dem Antike-Korpus einzelne Texte auszuwählen, klicken Sie hinter „Select files“ – wo zum jetzigen Zeitpunkt noch „9 selected“ angezeigt wird – auf das kleine Pfeilsymbol. Deaktivieren Sie die übrigen Texte, indem Sie das Häkchen entfernen oder die „unselect all“-Funktion nutzen und ausschließlich *Ilias* und *Odysee* mit einem Häkchen versehen (vgl. Abb. 13).

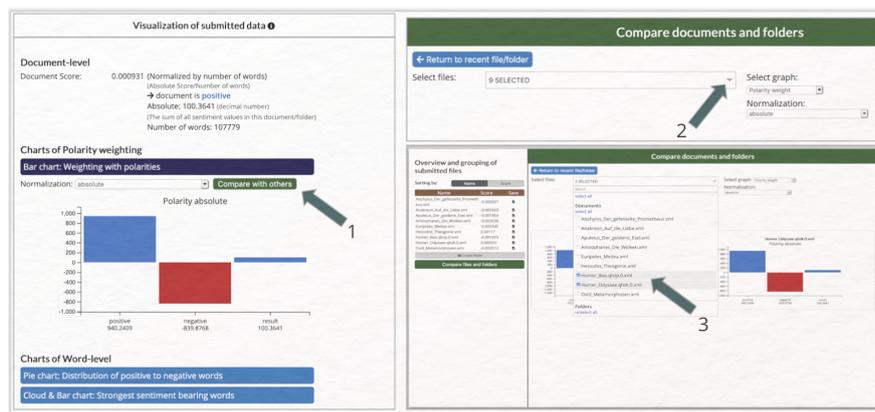


Abb. 13: Auswahl zweier Dateien für die vergleichende Analyse in drei Schritten

Wie auch bei der Analyse von Einzeltexten können Sie nun hinter „Select graph“ für die ausgewählten Texte unterschiedliche Visualisierungen abrufen. Während alle unter „Scores“, „Polarities“ und „Distribution“ abrufbaren Visualisierungen quantitative Analyseergebnisse in den Vordergrund stellen (Verteilung der SBWs und der Sentiment-tragenden Sätze im Text oder Korpus), ermöglichen die „Strongest Words“-Visualisierungen eine Exploration semantischer Strukturen.

Aufgabe 9: Werten Sie die Tortendiagramme aus, die die Verteilung der Sentiment-tragenden Sätze und Wörter anzeigen (Select graph, Distribution, Words/Sentences).

Wählen Sie unter „Select graph“ nun die Wordcloud aus. Jetzt werden Ihnen für beide Texte Wordclouds (vgl. **Wordcloud**) angezeigt, die die besonders häufig vorkommenden positiven Sentimentwörter beinhalten. Die Größe der Buchstaben repräsentiert die Vorkommenshäufigkeit. Die Farbe steht für die Polarität (blau: positiv; rot: negativ). Sobald Sie mit der Maus über einzelne Wörter hovern, werden Ihnen das Wort, der Sentimentwert, die Vorkommenshäufigkeit und der absolute Sentimentwert (also die pro Vorkommen addierten Sentimentwerte) angezeigt (vgl. Abb. 14 und 15). Unter „Polarity“ können Sie auswählen, zu welcher Polarität Sie sich eine Wordcloud anzeigen lassen möchten.



Abb. 14: Wordclouds der positiven Sentiment bearing words im Vergleich

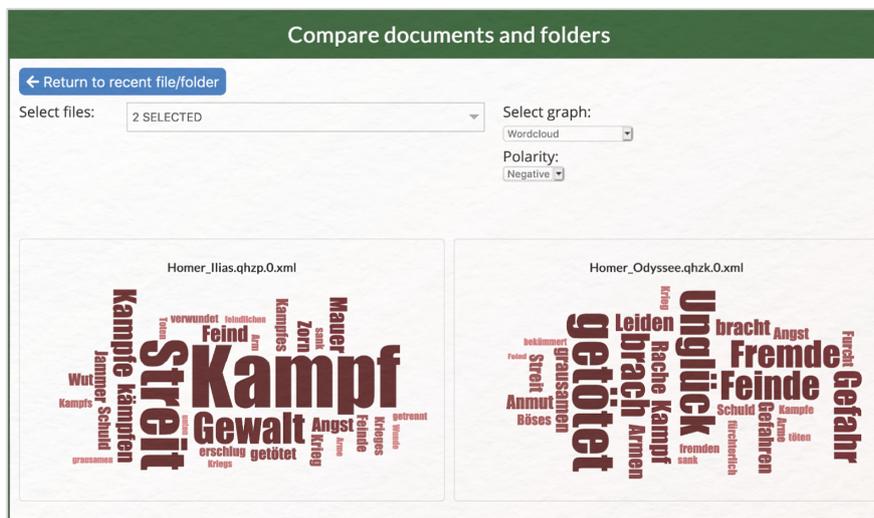


Abb. 15: Wordclouds der negativen Sentiment bearing words im Vergleich

Aufgabe 10: Welche positiven und welche negativen Sentimentwörter bestimmen die *Ilias* und die *Odyssee*? Werten Sie die Wordclouds aus.

Welche Wörter mit besonders hohem positiven und negativen Sentimentwert kommen in der *Ilias* und der *Odyssee* vor? Beziehen Sie in Ihre Auswertung auch die Balkendiagramme mit ein, welche Sie unter „Select Graph → (‘Strongest words’) → „Bar“ auswählen können.

Nun haben Sie einige der Analyseergebnisse innerhalb der Toolumgebung miteinander verglichen. Wenn Sie alle Analyseergebnisse herunterladen möchten, um diese bspw. auch offline und zu einem anderen Zeitpunkt analysieren zu können, nutzen Sie die Download-Funktion im Corpus-Panel (vgl. Abb. 16). Hierfür hovern Sie über dem Downloadsymbol rechts neben dem Dateititel und klicken auf das gewünschte Dateiformat.

Overview and grouping of submitted files

Sorting by: Name Score

Name	Score	Save
Aischylos_Der_gefesselte_Pro metheus.xml	-0.005027	📄
Anakreon_Auf_die_Liebe.xml	-0.005503	📄
Apuleius_Der_goldene_Esel.x ml	-0.001954	📄
Aristophanes_Die_Wolken.xml	-0.003539	📄
Euripides_Medea.xml	-0.005545	📄
Hesiodos_Theogonie.xml	0.00117	📄
Homer_Ilias.qhzp.0.xml	-0.001923	📄
Homer_Odyssee.qhzp.0.xml	-0.001923	📄
Ovid_Metamorphosen.xml	-0.003512	📄

CSV PNG XML

Create folder

Compare files and folders

Abb. 16: Download der Analyseergebnisse in drei Formaten

Aufgabe 11: Laden Sie die Analyseergebnisse der *Odyssee* herunter und speichern Sie diese im xml-Format auf Ihrem Desktop ab.

Vergleich mehrerer Korpora (Corpus-, Visualize-Panel): Im folgenden Teil der Lerneinheit werden Sie das noch ungeordnete Antike-Korpus so anordnen, dass ein genretypologischer Vergleich möglich wird. Hierfür richten Sie sich wieder dem „Overview and grouping of submitted files“-Modul zu. Legen Sie zunächst drei unterschiedliche Ordner an. Klicken Sie hierzu auf „create folder“, geben als Ordnername „Dramatiker“ in das Textfeld ein und bestätigen Ihre Eingabe per Klick auf „Submit“ (vgl. Abb. 17).

The screenshot displays the SentText interface. On the left, a table lists files with their names and scores. A modal dialog is open in the center, prompting for a folder name, with 'Dramatiker' entered. On the right, the 'Document-level' sentiment analysis for a document is shown, indicating it is negative with an absolute score of -42.8918. Below this, a bar chart titled 'Polarity absolute' shows the distribution of sentiment scores: positive (59.0698), negative (-101.9616), and result (-42.8918).

Name	Score
Aischylos_Der_ge	
fesselte_Prometh	-0.005027
eus.xml	
Anakreon_Auf_die	
_Liebe.xml	-0.005503
Apuleius_Der_gol	
dene_Esel.xml	-0.00
Aristophanes_Die	
_Wolken.xml	-0.00
Euripides_Medea.	
xml	-0.00
Hesiodos_Theogoni	
e.xml	0.00117
Homer_Ilias.qhzp.	
0.xml	-0.001923
Homer_Odyssee.q	
hzk.0.xml	0.000931
Ovid_Metamorph	
osen.xml	-0.003512

Document-level
 Document -0.005027 (Normalized by number of words)
 Score:
 (Absolute Score/Number of words)
 → document is **negative**
 Absolute: -42.8918 (decimal number)
 (The sum of all sentiment values in this document/folder)
 Number of words: 8533

Please enter folder name:
 Dramatiker
 Submit Cancel

Polarity absolute

Category	Score
positive	59.0698
negative	-101.9616
result	-42.8918

Abb. 17: Einen Ordner erstellen: „Create folder“, Texteingabe, „Submit“

Wiederholen Sie diesen Vorgang und legen Sie einen Epiker- und einen Dichter-Ordner an. In Ihrem Corpus-Panel befinden sich nun drei Ordner und neun Textdateien (vgl. Abb. 18).

Overview and grouping of submitted files

Sorting by: Name Score

	Name	Score	Save
 	Dramatiker	0	
 	Epiker	0	
 	Dichter	0	
	Aischylos_Der_gefesselte_Prometheus.xml	-0.005027	
	Anakreon_Auf_die_Liebe.xml	-0.005503	
	Apuleius_Der_goldene_Esel.xml	-0.001954	
	Aristophanes_Die_Wolken.xml	-0.003539	
	Euripides_Medea.xml	-0.005545	
	Hesiodos_Theogonie.xml	0.00117	
	Homer_Ilias.qhzp.0.xml	-0.001923	
	Homer_Odyssee.qhzk.0.xml	0.000931	
	Ovid_Metamorphosen.xml	-0.003512	

 Create folder

Compare files and folders

Abb. 18: Das unsortierte Corpus-Panel

Bevor die vergleichende Analyse durchgeführt werden kann, müssen die verschiedenen Dateien einem entsprechenden Ordner zugeordnet werden. Ziehen Sie hierfür die einzelnen Dateien per Drag & Drop in den passenden Ordner (vgl. Abb. 19).

Overview and grouping of submitted files

Sorting by: Name Score

	Name	Score	Save
📁 🗑️	Dramatiker	-0.004573	📁
	Aischylos_Der_gefesselte_Pro metheus.xml	-0.005027	📁
	Aristophanes_Die_Wolken.x ml	-0.003539	📁
	Euripides_Medea.xml	-0.005545	📁
📁 🗑️	Epiker	-0.000624	📁
	Homer_Ilias.qhzp.0.xml	-0.001923	📁
	Homer_Odyssee.qhzk.0.xml	0.000931	📁
	Hesiodos_Theogonie.xml	0.00117	📁
📁 🗑️	Dichter	-0.002962	📁
	Anakreon_Auf_die_Liebe.xml	-0.005503	📁
	Apuleius_Der_goldene_Esel.x ml	-0.001954	📁
	Ovid_Metamorphosen.xml	-0.003512	📁

📁 Create folder

Compare files and folders

Abb. 19: Das nach Gattungen sortierte Corpus-Panel

Klicken Sie nun auf den grünen „Compare files and folders“-Button und starten die vergleichende Sentimentanalyse. Nun öffnet sich automatisch ein zuvor nicht sichtbares Panel: das „Compare documents and folders“-Modul. Die generalisierten Analyseergebnisse werden Ihnen zunächst für alle Einzeltexte und für die drei Textkorpora angezeigt (vgl. Abb. 12). Sollte dies nicht der Fall sein, wählen Sie im Eingabefeld hinter „Selected files“ zunächst alle 12 Dateien aus. Für jedes Dokument bzw. jedes Korpus werden die folgenden Metadaten angezeigt: Titel, normalisierter Sentimentwert (= absoluter Sentimentwert, geteilt durch die Anzahl aller Wörter eines Textes, um die Vergleichbarkeit unterschiedlich langer Texte zu gewährleisten), absoluter Sentimentwert und Wortanzahl. Unter „Select graph“ können Sie aus unterschiedlichen Visualisierungsformen (Verteilungsgraf, Tortendiagramm, Wordcloud) auswählen, die die verschiedenen Analyseergebnisse (Polarität, Verteilung positiver/negativer Wörter und Sätze, stärkste Sentimentwörter/Sentiment-tragende Sätze) für die einzelnen Textdateien sowie die drei Korpora darstellen. Zunächst bleiben Sie aber bei der Startansicht (vgl. Abb. 20).

Document Name	Normalized	Absolute	Number of words
Aischylos_Der_gefoessete_Prometheus.xml	-0.005027	-42.8918	8533
Anakreon_Auf_die_Liebe.xml	-0.005503	-1.0181	185
Apuleius_Der_goldene_Esel.xml	-0.001954	-122.193	62520
Aristophanes_Die_Wolken.xml	-0.003539	-49.2032	13903
Earlphaen_Medea.xml	-0.005545	-60.0062	10822
Hesioden_Theogonie.xml	0.00117	11.0367	9430
Homer_Ilias@hp.0.xml	-0.001923	-373.3258	142153
Homer_Odyssee@hp.0.xml	0.000931	100.3641	107779
Quint_Metamorphosen.xml	-0.003512	-399.193	113650
Dramatiker	-0.002425	-425.427	175411
Epiker	0.00095	111.4008	117209
Dichter	-0.002962	-522.4041	176355

Abb. 20: Ergebnis der vergleichenden Analyse

Aufgabe 12: In welchen Texten herrscht eine positive, in welchen Texten eine negative Stimmung vor? Erstellen Sie eine Rangliste.

Zur Erinnerung: Die Polarität wird bei *SentiWS* jeweils in einer Skala von -1 (maximal negativ) bis $+1$ (maximal positiv) ausgedrückt und auf vier Nachkommastellen gerundet. Um die Analyseergebnisse in eine Reihenfolge zu bringen, ist der Blick auf den normalisierten Sentimentgehalt des Dokuments hilfreich. Nach einem ersten Überblick über den Sentimentgehalt der einzelnen Texte widmen wir uns nun dem detaillierten Vergleich der drei Teilkorpora. Sie haben in dem Compare-Modul die Möglichkeit, zuvor angelegte Ordner miteinander vergleichen zu lassen und die Ergebnisse genauer unter die Lupe zu nehmen. Per Default (vgl. **Default**) werden Ihnen im Moment die Analyseergebnisse aller 9 Dateien und die der drei Textkorpora angezeigt („12 selected“). Klicken Sie im Auswahlfeld neben „Selected files“ zunächst auf „unselect all“ und setzen dann ein Häkchen vor dem Dramatiker-, Epiker- und Dichter-Ordner (vgl. Abb. 21). Auf diese Weise schließen Sie die Analyse der Einzeltexte aus und können sich auf den Vergleich der Teilkorpora konzentrieren.

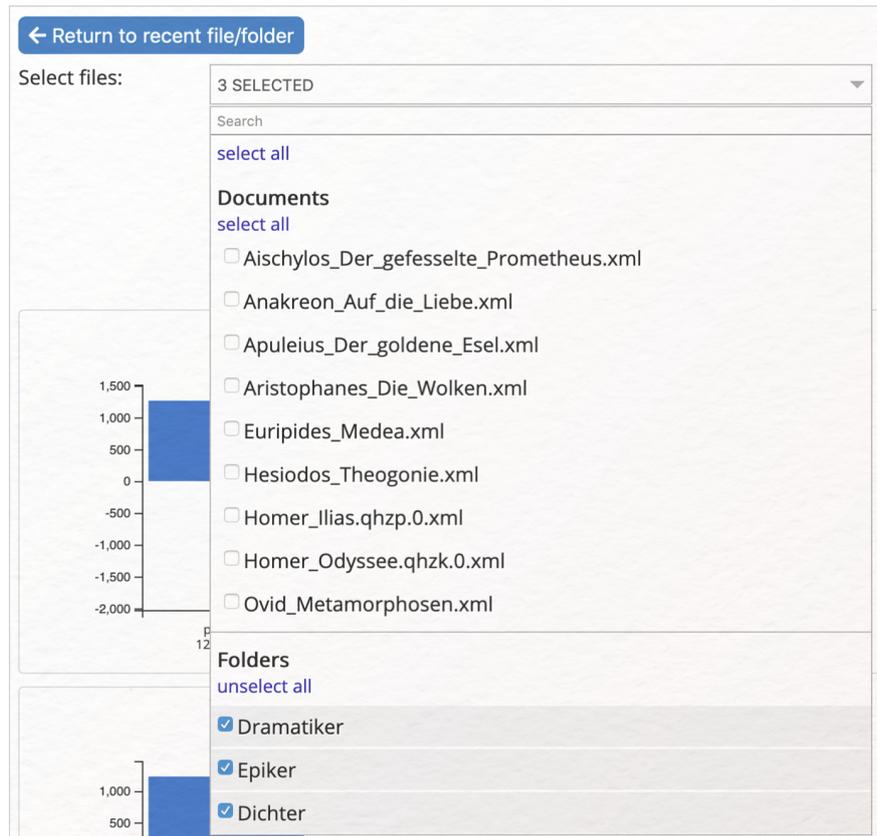


Abb. 21: Auswahl der drei Korpora für eine vergleichende Sentimentanalyse

Sollten Sie einzelne Texte miteinander vergleichen wollen, können Sie diese übrigens ebenfalls in diesem Bereich auswählen. Durch die Auswahl der drei Ordner gelangen Sie nun zu der bereits bekannten und jetzt etwas größer abgebildeten Gesamtübersicht, in der die absolute und die normalisierte Polarität aller Ordner angezeigt wird. Neben dem „Select graph“-Schriftzug haben Sie wieder die Möglichkeit, sich einzelne Analyseergebnisse in Form unterschiedlicher Visualisierungen anzeigen zu lassen. Unter „Polarity“ wird die Polarität der einzelnen Korpora angezeigt. Unter „Distribution“ wird die Verteilung der positiven und negativen sentiment bearing words bzw. der Sätze mit positiver und negativer Polarität in den einzelnen Teilkorpora angezeigt. Unter „strongest words“ können Sie sich die besonders positiven und besonders negativen Wörter in Wordclouds oder Grafen anzeigen lassen. Bearbeiten Sie nun die folgenden Aufgaben, indem Sie die Einstellungen hinter „select graph“ den Angaben in den Klammern entsprechend anpassen.

Aufgabe 13 (Polarity Weight; Normalization: Sentiment Bearing Words): Schauen Sie sich die Grafen an. Welches Korpus enthält die meisten *sentiment bearing words*? Welches Korpus ist am „stimmungsvollsten“?

Aufgabe 14 (Distribution; Words): Beschreiben und interpretieren Sie die Verteilung der positiven und negativen Sentiment-tragenden Wörter innerhalb der drei Teilkorpora. Wie lassen sich die Diagramme interpretieren, wenn die Verteilung der positiven und negativen Sätze (Distribution; Sentences) angezeigt wird?

Aufgabe 15 (Strongest Words; Wordcloud, Polarity: negative/positive): Welche negativen und welche positiven Sentiment-tragenden Wörter bestimmen die drei Textkorpora? Lassen sich aus den Wordclouds erste Interpretationshypothese über inhaltliche Schwerpunkte ableiten?

Damit sind Sie am Ende der Lerneinheit zur lexikonbasierten Sentimentanalyse mit SentText gelangt. Sie haben ein Antike-Korpus erstellt und dieses bei SentText hochgeladen. Hier haben Sie das Corpus-, das Visualize- sowie das Text-Modul und deren Funktionen kennengelernt.

4. Lösungen zu den Beispielaufgaben

Aufgabe 1: Welche Wörter werden rot markiert, welche Wörter werden blau markiert und welche Stimmung assoziieren Sie mit den farblich hervorgehobenen Begriffen? Welches Wort scheint Ihnen eine besonders positive Gefühlslage zu repräsentieren und welches ist negativ konnotiert? Bringen Sie die Wörter – absteigend von positiv zu negativ konnotierten Wörtern – in eine Reihenfolge.

„heiligen“ und „gelernt“ werden blau markiert. Sie sind positiv konnotiert und könnten mit Reinheit und einem

bereichernden Wissenszuwachs assoziiert werden. „Zerstörung“ wird rot eingefärbt und ist negativ konnotiert. Der Begriff entwirft eine düstere, destruktive Stimmung. Die folgenden Reihenfolgen erscheinen plausibel: 1) heiligen (sehr positiv); 2) lernen (positiv); 3) Zerstörung (negativ) oder 1) lernen (sehr positiv); 2) heiligen (positiv); 3) Zerstörung (negativ).

Aufgabe 2: Die Demoversion greift auf das Sentimentwörterbuch *SentimentWortschatz (SentiWS)* zurück. „heiligen“ hat hier einen Sentimentwert von 0.0727, der Sentimentwert von „gelernt“ beträgt 0.2492. „Zerstörung“ hat einen Sentimentwert von -0.3578. Passen die errechneten Sentimentwerte zu Ihrer persönlichen Reihenfolge?

Die errechneten Sentimentwerte entsprechen der Reihenfolge. Dass „Zerstörung“ einen äußerst negativen Sentimentwert erhält, erscheint plausibel. „gelernt“ ist *SentiWS* zufolge positiver konnotiert als „heiligen“. Die positive Bewertung beider Verben ist nachvollziehbar, könnte je nach wertendem Individuum und zeitlichem Kontext allerdings auch in umgekehrter Reihenfolge ausfallen. Ein religiöser Mensch würde evtl. „heiligen“ positiver bewerten als „lernen“. Mit dem Beginn der Aufklärung ab 1700 dürfte „lernen“ deutlich positiver bewertet werden als „heiligen“. An dieser Stelle wird erneut der Einfluss des Sentimentwörterbuchs auf Analyseergebnisse und Interpretation deutlich. Die Konnotation der Lexeme wird durch den zeitlichen Kontext und die menschlichen Bewertungen anhaftende Subjektivität mitbestimmt.

Aufgabe 3: In welchen Texten herrscht eine positive, in welchen Texten eine negative Stimmung vor? Erstellen Sie eine Rangliste und nutzen Sie die empirischen Daten, um eine erste Interpretationshypothese über das Stimmungsbild abzuleiten.

Texte mit negativer Polarität in aufsteigender Reihenfolge (negativ zu sehr negativ)

Homer: *Ilias* -0.001923

Apuleius: *Der goldene Esel* -0.001954

Ovid: *Metamorphosen* -0.003512

Aristophanes: *Die Wolken* -0.003539

Aischylos: *Der gefesselte Prometheus* -0.005027

Anakreon: *Auf die Liebe* -0.005503

Euripides: *Medea* -0.005545

Texte mit positiver Polarität in aufsteigender Reihenfolge (positiv zu sehr positiv)

Homer: *Odyssee* 0.000931

Hesiodos: *Theogonia* 0.00117

Für einen Großteil der Texte ist ein negatives Stimmungsbild konstitutiv; nur zwei Texte weisen einen positiven durchschnittlichen Sentimentwert auf. Homers Epen führen beide Listen an, hier lassen sich sowohl besonders positive als auch negative gefühlsbezogene Extremsituationen in Form Sentiment-tragender Signalwörter auf der Textoberfläche nachweisen.

Aufgabe 4: Unterhalb der blauen Titelzeile, in der der Dateiname der ausgewählten Datei angezeigt wird, können Sie einer kleineren Kopfzeile die Wortanzahl des Dokuments und die Gesamtanzahl der Sentiment-tragenden Wörter entnehmen. Wie lauten die absoluten Sentimentwerte (Addition der Werte aller Sentiment-tragender Wörter, angegeben als Dezimalzahl) beider Homerischer Epen?

SentText addiert sämtliche Sentimentwerte, die für die *Odyssee* vergeben wurden und errechnet einen positiven Gesamtwert von 100.3641. SentText addiert sämtliche Sentimentwerte, die für die *Ilias* vergeben wurden und errechnet einen negativen Gesamtwert von -273.3258.

Aufgabe 5: Nutzen Sie die hinter dem Informationssymbol verborgenen Tooltips im Text-Panel. Welche Wörter werden gelb markiert – und beinhaltet der Text gelb hervorgehobene Wörter?

Gelb markiert sind diejenigen Wörter, die im ausgewählten Lexikon enthalten sind und deren Sentimentwert 0 beträgt. Es handelt sich um neutrale Wörter. Der *Gefesselte Prometheus* enthält keine neutralen Sentimentwörter.

Aufgabe 6: Welchen Sentimentwert und welche Polarität haben „gefesselt“, „Fesseln“, „Einsamkeit“, „Kraft“ und „Gewalt“? Passen die Polaritäten und Ihre eigene Vorstellung zusammen? Falls nicht: Woran könnte der Unterschied liegen?

„gefesselt“: Polarität: positiv; Sentimentwert: 0.004

„Fesseln“: Polarität: negativ; Sentimentwert: -0.0048

„Einsamkeit“: Polarität: negativ; Sentimentwert: -0.339

„Kraft“: Polarität: positiv; Sentimentwert: 0.004

„Gewalt“: Polarität: negativ; Sentimentwert -0.5363

In den meisten Fällen stimmen wahrscheinlich Sentimentwert und eigene Konnotation überein. Die negative Konnotation der Lexeme „Gewalt“ und „Fesseln“ erscheint genauso plausible wie die positive Polarität von „Kraft“. Dass „fesseln“ als verwandte Verbform von „Fesseln“ eine positive Polarität aufweist, erscheint fragwürdig. Hier müsste eine manuelle Änderung vorgenommen werden. Eine Erklärung für die Vergabe eines positiven Sentimentwerts liegt in der Bedeutungsvielfalt der Verbform. „fesseln“ im Sinne einer durch das Anlegen von

Fesseln oder Festbinden beraubten Bewegungsfreiheit ist negativ konnotiert. Versteht man „fesseln“ allerdings als faszinierendes Beanspruchen von Aufmerksamkeit, ist der positive Sentimentwert nachvollziehbar.

Aufgabe 7: Wählen Sie die „Table“-Visualisierung aus. Wie lautet der Satz mit dem negativsten Sentimentwert? Welche Figur äußert ihn und an wen richtet er sich?

Der Satz lautet: „Ja, krank, wenn Krankheit seine Feinde hassen heißt!“. Er wird von Prometheus ausgesprochen und richtet sich an Hermes.

Aufgabe 8: Wählen Sie die „Timeline“-Visualisierung aus. In welchem Textsegment finden sich besonders positive bzw. negative Sätze und wie lauten diese?

In dem Textsegment zwischen 600 und 650 finden sich Sätze mit besonders positiver Polarität. Einen besonders positiven Sentimentwert von 0.201 weist „Behaglich“ auf, wobei es sich hierbei nicht um einen vollständigen Satz handelt. Im Segment zwischen 620 und 675 finden Sätze mit besonders negativer Polarität, hierzu gehört z. B. „So wirst du mir auch Schuld an deinem Leide vor?“.

Aufgabe 9: Werten Sie die Tortendiagramme aus, die die Verteilung der Sentiment-tragenden Sätze und Wörter anzeigen (Select graph, Distribution, Words/Sentences).

Der Vergleich der vier Diagramme zeigt eine sehr ähnliche Verteilung der Sentiment-tragenden Wörter und Sätze. Beide Epen weisen ein relativ ausgeglichenes Verhältnis zwischen positiven und negativen Sentiment-tragenden Wörtern und Sätzen auf. Die Anzahl an positiven SBWs überwiegt, wobei diese Tendenz in der *Ilias* ausgeprägter ist als in der *Odyssee*. Die Verteilung positiver und negativer Sätze ist ebenfalls ähnlich, positive Sätze überwiegen in beiden Texten. Es zeigt sich ein ausgeglichenes Stimmungsbild.

Aufgabe 10: Welche positiven und welche negativen Sentimentwörter bestimmen die *Ilias* und die *Odyssee*? Werten Sie die Wordcloud aus.

Welche Wörter mit besonders hohem positiven und negativen Sentimentwert kommen in der *Ilias* und der *Odyssee* vor? Beziehen Sie in Ihre Auswertung auch die Balkendiagramme mit ein, welche Sie unter „Select Graph → (Strongest words)“ → „Bar“ auswählen können.

Die acht negativsten Wörter in Homers *Odyssee* lauten „Gewalt“, „Jammer“, „getötet“, „Unglück“, „Feinde“, „brach“, „Fremde“ und „Gefahr“.

Die acht negativsten Wörter der *Ilias* sind „Kampf“, „Streit“, „Gewalt“, „Kämpfe“, „kämpfte“, „Mauer“, „Feind“ und „Zorn“. Aus der Distanz betrachtet, spiegelt sich das zentrale Thema des Epos – der erbitterte Kampf zwischen Achaern und Trojanern um die Stadt Troja – deutlich in den besonders negativen Wörtern wider. Der Bau und die Verteidigung der Mauer um das Lager der Griechen lässt sich ebenfalls ablesen, darüber hinaus findet sich auch der berühmte iliadische Zorn im Grafen wieder. Die *Odyssee* erscheint auf den ersten Blick weniger Kampfdarstellungen zu enthalten. Zentrale Themen wie die Reise in die Fremde oder der Verlust von Freunden lassen sich aus dem Grafen ableiten.



Abb. 22: Die acht besonders negativen Wörter in der *Ilias* und der *Odyssee*

Die acht besonders positiven Wörter der *Ilias* lauten „Genossen“, „Mut“, „mutige“, „Mutes“, „mutigen“, „Muts“, „Geschenk“ und „Schnell“. Hier wird die wichtigste Charaktereigenschaft der Helden deutlich. Gleichzeitig offenbart sich der Mehrwert einer Lemmatisierung. Bei einem weiteren Analysedurchlauf sollte eine Lemmatisierung durchgeführt werden, damit Wörter desselben Wortstammes nicht als unterschiedliche Lexeme gezählt werden. Bei „großen“, „herrliche“, „edlen“, „Geschenke“, „Genossen“, „kluge“, „prächtigen“ und „Ruhe“ handelt es sich um

die acht besonders positiven SBWs der *Odyssee*. Die Zusammensetzung ist deutlich heterogener als in der *Ilias*. Die Adjektive groß, herrlich, edel, klug und prächtig rücken Charaktereigenschaften oder körperliche Eigenschaften in den Vordergrund.

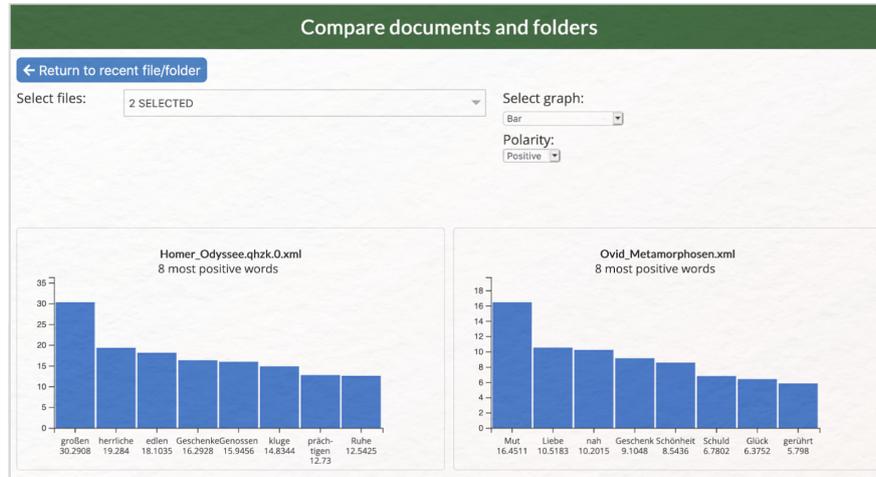


Abb. 23: Die acht besonders negativen Wörter in der *Ilias* und der *Odyssee*

Aufgabe 11: Laden Sie die Analyseergebnisse der *Odyssee* herunter und speichern Sie diese auf Ihrem Desktop ab.

Aufgabe 12: In welchen Texten herrscht eine positive, in welchen Texten eine negative Stimmung vor? Erstellen Sie eine Rangliste.

Positiver durchschnittlicher Sentimentwert (absteigend, sehr positiv bis weniger positiv)

1) Hesiodos: *Tehogonie* (0.0017)

2) Homer: *Odyssee* (0.000931)

Negativer durchschnittlicher Sentimentwert (absteigend, sehr negativ bis weniger negativ)

1) Homer: *Ilias* (-0.001923)

2) Apuleius: *Der goldene Esel* (-0.001954)

3) Ovid: *Metamorphosen* (-0.003512)

4) Aristophanes: *Die Wolken* (-0.003539)

5) Aischylos: *Der gefesselte Prometheus* (-0.00491)

6) Anakreon: *Auf die Liebe* (-0.005503)

7) Euripides: *Medea* (-0.005545)

Aufgabe 13 (Polarity Weight; Normalization: Sentiment Bearing Words): Schauen Sie sich die Grafen an. Welches Korpus enthält die meisten „sentiment bearing words“? Welches Korpus ist am „stimmungsvollsten“?

Das Korpus der Epiker enthält die meisten sentiment bearing words und könnte demnach als besonders stimmungsvoll bezeichnet werden.

Epiker: 23.030 SBWs (-0.0006)

Dichter: 5.770 SBWs (-0.0030)

Dramatiker: 2.990 SBWs (-0.0046)

Bei der Bezeichnung als „stimmungsvollstes Korpus“ ist Vorsicht geboten, da die Textlänge natürlich die Anzahl der SBWs beeinflusst. Die Epen sind bspw. deutlich umfangreicher als die lyrischen Werke. Bei einer solchen quantitativen Analyse ist es also ratsam, Texte ähnlichen Umfangs auszuwählen. Um eine fehlerhafte Interpretation zu vermeiden, sollten außerdem die normalisierten Werte einbezogen werden.

Aufgabe 14 (Distribution; Words): Beschreiben und interpretieren Sie die Verteilung der positiven und negativen Sentiment-tragenden Wörter innerhalb der drei Teilkorpora. Wie lassen sich die Diagramme interpretieren, wenn die Verteilung der positiven und negativen Sätze (Distribution; Sentences) angezeigt wird?

Die Verteilung der positiven und negativen SBWs erscheint auf den ersten Blick sehr ausgeglichen. Das gilt v. a. für das Korpus der Dramatiker. Hier befindet sich das Verhältnis fast in der Wage (48,7 % negative SBWs, 51,3 positive SBWs). Mit einem Verhältnis von +60 zu -40 weist das Korpus der Epiker die größte Diskrepanz zwischen positiven und negativen SBWs auf. Das Korpus der Dichter nimmt eine Mittelposition ein (+54,49 zu -45,51). Die vergleichende Analyse der drei Antike-Korpora zeigt zunächst, dass im Durchschnitt alle drei Gattungen einen normalisierten negativen Sentimentwert aufweisen. Das bestätigt der Blick auf die einzelnen Texte, von denen insgesamt nur zwei einen positiven Sentimentwert aufweisen. Bei den epischen Texten (*Ilias*, *Odyssee* und *Theogonia*) schlägt der Sentimentwert besonders stark in den Minusbereich aus. Das ist überraschend, da

zwei der im Epiker-Korpus enthaltenen Texte zu den Texten gehören, für die ein durchschnittlich positiver Sentimentwert errechnet wurde. Dieser Befund geht konform mit den übrigen Analyseergebnissen, da die *Ilias* insgesamt den zweitnegativsten Sentimentwert aufweist und den Sentimentwert des Epiker-Korpus maßgeblich in den negativen Bereich führt.

Aufgabe 15 (*Strongest Words; Wordcloud, Polarity: negative/positive*): Welche negativen und welche positiven Sentiment-tragenden Wörter bestimmen die drei Textkorpora? Lassen sich aus den Wordclouds erste Interpretationshypothesen über inhaltliche Schwerpunkte ableiten?

Die vergleichende Analyse der drei Antike-Korpora zeigt zunächst, dass im Durchschnitt alle drei Gattungen einen negativen Sentimentwert aufweisen. Das bestätigt der Blick auf die einzelnen Texte, von denen nur zwei einen positiven Sentimentwert aufweisen.

Epiker: -161.925

Dramatiker: -152.1012

Dichter: -123.2111

Aus den Wordclouds treten die folgenden SBWs besonders hervor:

Negative zentrale Begriffe:

Dramatiker: Leid, schlechten, Feinden, Schuld

Epiker: Streit, Gewalt, Mauer, Kampfe, kämpfen

Dichter: Arme, Gewalt, Feind, Feinde, Gefahr, Furcht, Krieg, Streit

Positive zentrale Begriffe:

Dramatiker: Heil, mag, gut, Freude, treu

Epiker: Mut, großen, herrliche, edlen

Dichter: Mut, Liebe, Schönheit, Freuden, Glück, Geschenk

Dramatiker: Hier ließe sich eine Auseinandersetzung mit Schuld und der Konflikt zwischen Feind und Freund ableiten. Die Wordcloud lässt vermuten, dass Fragen der Moral verhandelt werden.

Epiker: In der Wordcloud rücken Gewalt und Kampf in den Vordergrund, was ebenfalls auf das Austragen eines Konflikts deutet. Mut tritt als wichtige Eigenschaft des Helden hervor.

Dichter: Das zentrale Thema der kriegerischen Auseinandersetzung ist in der Wordcloud deutlich zu erkennen und steht mit konventionell negativ konnotierten Wörtern wie Gewalt, Furcht, Feind und Streit in Verbindung. Dem gegenüber stehen konventionell positiv konnotierte Lexeme wie Liebe, Glück, Freuden, Geschenk oder Schönheit. In dieser Wordcloud steht ein gefühlsbetontes Vokabular im Vordergrund, was gattungstypologisch plausibel erscheint.

Externe und weiterführende Links

- Textgrid Repository: <https://web.archive.org/save/https://textgridrep.org/> (Letzter Zugriff: 04.09.2024)
- SentText: <https://web.archive.org/save/http://thomasschmidtur.pythonanywhere.com/> (Letzter Zugriff: 04.09.2024)

Bibliographie

- Bläß, Sandra. 2024. Methodenbeitrag: Korpusbildung. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 2. Korpusbildung (12. Juni). doi: 10.48694/fortext.3708, <https://fortext.net/routinen/methoden/korpusbildung>.
- Flüh, Marie. 2024a. Methodenbeitrag: Sentimentanalyse. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 7. Sentimentanalyse (7. Oktober). doi: 10.48694/fortext.3797, <https://fortext.net/routinen/methoden/sentimentanalyse>.
- . 2024b. Toolbeitrag: SentText. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 7. Sentimentanalyse (7. Oktober). doi: 10.48694/fortext.3799, <https://fortext.net/tools/tools/senttext>.
- Latacz, Joachim. 2014. *Homers Ilias. Studien zu Dichter, Werk und Rezeption*. Hg. von Thierry Greub, Krystyna Greub-Fracz, und Arbogast Schmitt. Berlin, Boston: de Gruyter. <https://doi.org/10.1515/9783110306361> (zugegriffen: 18. Mai 2020).
- Remus, R., U. Quasthoff und G. 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)*, 1168–1171. Valetta, Malta: European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2010/pdf/490_Paper.pdf.
- Schmidt, Thomas und Manuel Burghardt. 2018. An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing. In: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 139–149. Santa Fe, New Mexico: Association for Computational Linguistics.

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.
- 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**.
- Default** Das/der Default (engl. für Voreinstellung oder Standardwert) bezeichnet den Wert einer Softwareeinstellung oder einer Eingabevariable, der verwendet wird, falls Nutzer*innen selbst keinen Wert oder keine Einstellungen vornehmen. Es handelt sich also um die standardmäßig festgelegten Einstellungen eines Tools oder Programms. Per Default festgelegte Parameter lassen sich i. d. R. manuell umstellen.
- Feature** Unter Features können Einzelfunktionen eines Tools verstanden werden, die beispielsweise komplexe Funktionen wie die Visualisierung eines Textes als **Wordcloud** ermöglichen, oder auch kleinere Funktionseinheiten wie den Abgleich einzelner Spracheigenschaften (**Properties**) mit **annotierten** Beispieltexten darstellen.
- 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.
- 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 unbekanntem Daten verwendet werden.
- Markup (Textauszeichnung)** Die Textauszeichnung (eng. *Markup*) fällt in den Bereich der Daten- bzw. Textverarbeitung, genauer in das Gebiet der Textformatierung, welche durch **Auszeichnungssprachen** wie **XML** implementiert wird. Dabei geht es um die Beschreibung, wie einzelne Elemente eines Textes beispielsweise auf Webseiten grafisch dargestellt werden sollen.
- 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 of 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**.
- Property** Property steht für „Eigenschaft“, „Komponente“ oder „Attribut“. In der automatischen **Annotation** dienen konkrete Wortheigenschaften wie Groß- und Kleinschreibung zur Klassifizierung von Wörtern oder Phrasen. Durch die Berücksichtigung solcher Eigenschaften in den **Features** eines Tools kann **maschinelles Lernen** bestimmter Phänomene umgesetzt werden. In der manuellen Annotation können als Properties auch Eigenschaften von **Annotationen** benannt werden.
- 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**.
- SBW** SBW steht für *Sentiment Bearing Word* (Sentimentwort) und bezeichnet ein Lexem, das eine Stimmung, eine Bewertung, ein Gefühl, eine Einstellung oder auch eine Emotion zum Ausdruck bringt. Für SBWs werden „semantische Orientierung“ und „Stärke“ in Form des **Sentimentwerts** festgelegt. SBWs werden in **Sentimentlexika** gesammelt und als Grundlage für lexikonbasierte Sentimentanalysen verwendet.
- Sentimentwert** Der Sentimentwert oder Sentimentgehalt eines Wortes beschreibt, meistens auf einer Skala von -1 (maximal negativ; bspw. „schädlich“: -0,9269) bis +1 (maximal positiv, bspw. „Freude“: 0,6502) die Polarität von Sentimentwörtern (siehe auch **SBWs**). Der Wert wird bei der Generierung von **Sentimentlexika** für jedes enthaltene Wort errechnet.
- Sentimentwörterbuch** Ein Wörterbuch, das ausschließlich Lexeme enthält, die **Träger von Sentiments** sind, wird als Sentimentwörterbuch definiert. Sentimentlexika fungieren als wichtige Ressource für lexikonbasierte Sentimentanalysen, bei denen die Wörter des Wörterbuchs und die Wörter eines Primärtextes miteinander abgeglichen werden.
- Stoppwortliste** Stoppwörter sind hochfrequente Wörter, meist Funktionswörter, die, aufgrund ihrer grammatisch bedingten Häufigkeit, beispielsweise die Ergebnisse von inhaltlichen oder thematischen Analysen verzerren können. Deshalb werden diese Wörter, gesammelt in einer Stoppwortliste, bei digitalen Textanalysen meist nicht berücksichtigt.
- TEI** Die *Text Encoding Initiative* (TEI) ist ein Konsortium, das gemeinsam einen Standard für die Darstellung von Texten in digitaler Form entwickelt. Die TEI bietet beispielsweise Standards zur Kodierung von gedruckten Werken und zur Auszeichnung von sprachlichen Informationen in maschinenlesbaren Texten (siehe auch **XML** und **Markup**).
- 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.

XML XML steht für *Extensible Markup Language* und ist eine Form von **Markup Language**, die sowohl computer- als auch menschenlesbar und hochgradig anpassbar ist. Dabei werden Textdateien hierarchisch strukturiert dargestellt und Zusatzinformationen i. d. R. in einer anderen Farbe als der eigentliche (schwarz gedruckte) Text dargestellt. Eine standardisierte Form von XML ist das **TEI-XML**.