

ドイツにおけるデジタル・ヒューマニティーズ 日本学を中心に

Michael Kinski, Frankfurt



デジタル・ヒューマニティーズ(DH)とは何か

- 人文科学は、人間の社会、文化、言語、歴史、思考、コミュニケーションのあらゆる側面を扱う個々の分野で構成されています。
- デジタル人文科学は、これらの関心領域を共有し、デジタル化されつつある働動環境およびメディアの世界の条件下で、新しい知識を獲得して伝えるプロセスをさらに発展させようとしています。
- デジタル人文科学は知識と文化遺産のデジタル化、ツールの使用と開発を目的しながら、デジタル世界における人文科学の方法論的および理論的基盤の考察を行っています。

デジタル・ヒューマニティーズ(DH)とは何か

- デジタル人文科学は、従来の学問分野に関連してさまざまなよう認識できます。一部の人々は、それらを人文科学への補助科学でありながらそれらの問題関心に従うものと見なしています。
- 従来の学問分野向けのコンピュータサイエンスによる研究法を開発する応用コンピュータサイエンスと見なす人もいます。
- さらに別の人は、それを方法やデジタル化されたレパートリーのレベルですべての問題提携を網羅する人間科学と見なしています。
- その認識と応用が深まれば、DHはDH自身の主題とDH自身の方法を持ち、最終的に独立した研究分野になるでしょう。

フランクフルト大学の日本学科におけるデジタル・ヒューマニティーズ ワークショップ・シリーズ 第4号

TALKS + WORKSHOP
SOCIAL NETWORK ANALYSIS

KINSKI
HAERPFER
BINGENHEIMER
VOIGTMANN
BEERENS
BORN
ITO

DIGITAL
APPROACHES
4000

20./21.08.
NEUE MENSA K111

DIGIJAPPOFFM
FRANKFURT

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

日本学

REGISTRATION REQUIRED @ DIGIJAPOFFM@EM.UNI-FRANKFURT.DE

...その知らせを受けてのコメント

- Sehr geehrte Listenmitglieder, auch in der **Japanologie** gibt es **recht viele Initiativen für digitale Anwendungen in Forschung und Lehre**. Die Ankündigung von Herrn Voigtmann (Japanologie Univ. Frankfurt a.M.) von heute ist nur eines von vielen Beispielen. **Aber kommt dabei das "deep reading" eines Textes und die darauf aufbauende Interpretation nicht zu kurz?** Oder habe ich das missverstanden, denn **es scheint bei dieser Veranstaltung gar nicht ums Lesen und Interpretieren zu gehen, sondern um Social Network Analysis mittels digitaler Techniken**. Zu diesem Thema passt jedoch irgendwie der gestern in der FAZ erschienene Text von Rüdiger Görner. Siehe hier: <https://www.faz.net/aktuell/feuilleton/hoch-schule/wenn-die-saat-des-digitalen-aufgeht-digital-humanities-16255522.html> Darin heißt es: "Das „deep reading“ sei verschwunden (Nietzsche sprach vom „langsamen Lesen“ als einer Kulturtechnik, ohne die es weder kritisches Bewusstsein noch ästhetisches Vergnügen geben könne) und damit das eigenständige Sicherarbeiten von Textverständnis. Der kanadische Romancier und Beobachter dieses Kongresses, Stephen Marche, befand dazu neulich im „Times Literary Supplement“, dass selbst Gebildete zunehmend unwillig seien, zwischen fälschlicher und geprüfter Information zu unterscheiden, und zu einer Gemeinschaft von Menschen werden, die sich selbst nur noch als eine Ansammlung von undifferenzierten und uninterpretierten Ziffern verstehen. Die bloße Existenz von Instrumentarien wie Facebook belege, was geschehe, wenn wir unsere Kommunikationsformen reinen Technokonstrukteuren überließe." Vielleicht macht dieser Artikel sensibel für die andere, auf der Frankfurter Veranstaltung nicht im Zentrum stehende Frage, ob die instrumentelle / digitale Nutzung von Texten japanischer Autoren nicht auch in der Japanologie **das eigene "Sich-Erarbeiten von Textverständnis" eher verhindert als dass es erleichtert würde. "Netzwerke von Gelehrten und Autoren digital darstellen", oder aber "Textverständnis erarbeiten" - Ist beides gleich wichtig für die Japanologie? Ist das eine wichtiger und dringlicher als das andere?** Mit besten Grüßen, Wolfgang Seifert

コメントの要点

- 日本学の分野でも研究や授業において「デジタルなアプローチ」を使う試みあり
- フ大日本学科のワークショップはその一つの表れ
- 目的は本来日本学の使命であるべき文献の解読や解釈ではなく、デジタルな技法によるsocial network analysis
- 危険の1 „deep reading“による文献との取り組みが蔑ろにされるという恐れ
- 危険の2 若い研究者が自分の努力をもって文献を解明する能力が伸びないという恐れ
- 結論 将来の日本学にどれが大事だろう？ コンピューターによるsocial network analysis？それとも「文献解読の努力」？

理解度の問題

- 概していえばドイツの学术界ではデジタル・ヒューマニティーズについて誤解や偏見が根付いている
- その1 デジタル・ヒューマニティーズはつまりデータバンクを作ることを意味している
- その2 デジタル・ヒューマニティーズはズルい
つまり本来文献学的な日本学の本業を怠けた
偽ったりする作業
- Franco Morettiのいう„distant reading vs. close reading“という役割分担を認識せずに、ほとんどの知識なくデジタル・ヒューマニティーズを軽んじる傾向が受け止められる

...その反面

- 連邦政府による支援(科研費)
- 州政府のレベルでの支援(科研費)
- DFG (Deutsche Forschungsgemeinschaft = 科学研究費を提供する公の機関)による支援

Bundesministerium für Bildung und Forschung (BMBF)

ドイツ連邦の文部科学省

„Digital humanities: the rise of a new academic discipline“ (2018)

Research in
Germany



Land of Ideas

BMBF 2018:

Digital humanities: the rise of a new academic discipline

- **Computerised access to sources:** Manuscripts, images and archaeological artefacts are scanned and prepared for online access. The Cuneiform Digital Library for example contains tens of thousands of cuneiform scripts from collections worldwide, complete with transcriptions of the inscriptions and information about where they were found and their dating. This is a project that was initiated back in the 1990s by the Max Planck Institute for the History of Science in Berlin and the University of California in Los Angeles.
- **Data mining and visualisation:** data analysis and data processing are other fields in the digital humanities. One example is the Leibniz ScienceCampus "[Empirical Linguistics & Computational Language Modeling](#)", where new methods of recording and analysing digital language documents are being developed.
- **Digital research environments for humanities and social science scholars:** The [TextGrid](#) platform was the first virtual research environment to be set up for the humanities in Germany. Funded by Germany's Federal Ministry of Education and Research (BMBF), the platform has been assisting researchers, particularly with the creation of digital editions, since the summer of 2011.

支援プログラムの3つの側面

- **Computerised access to sources:** Manuscripts, images and archaeological artefacts are scanned and prepared for online access. The Cuneiform Digital Library for example contains tens of thousands of cuneiform scripts from collections worldwide, complete with transcriptions of the inscriptions and information about where they were found and their dating. This is a project that was initiated back in the 1990s by the Max Planck Institute for the History of Science in Berlin and the University of California in Los Angeles.
- 資料・研究対象へのコンピュータ化されたアクセス:
原稿、画像、考古学的遺物がスキャンされ、オンラインアクセス用に準備
- 例えば楔文字のデジタルライブラリ作成

BMBF 2018:

Digital humanities: the rise of a new academic discipline

- **Data mining and visualisation:** data analysis and data processing are other fields in the digital humanities.
- **データマイニングと視覚化:** データ分析とデータ処理
- **例** Leibniz Science Campus
Empirical Linguistics & Computational Language Modeling: 自然言語処理モデルの導入を通じてドイツ語の大規模コーパスリソースの高品質な**自動注釈**をサポートする革新的な研究を実施し、言語学における高度な実証的研究を可能にすることを目的とする

BMBF 2018:

Digital humanities: the rise of a new academic discipline

- **Digital research environments for humanities and social science scholars:** The [TextGrid](#) platform was the first virtual research environment to be set up for the humanities in Germany. Funded by Germany's Federal Ministry of Education and Research (BMBF), the platform has been assisting researchers, particularly with the creation of digital editions, since the summer of 2011.
- 人文科学および社会科学の学者のためのデジタル研究環境: デジタル版の作成で研究者を支援するためのバーチャル研究環境を設置 例 TextGrid platform (2011より支援)

BMBF 2018:

Digital humanities: the rise of a new academic discipline

**DARIAH-DE Digital Forschungsstruktur für die
Geistes- und Kulturwissenschaften**

人文科学および文化研究のためのデジタル研究構造

- Text Grid → DARIAH-DE
- <https://de.dariah.eu/en/web/guest/dienste-und-werkzeuge>

TextGrid>>DARIAH-DEのツール

- **CosmoTool**は、さまざまな資料からの伝記情報を歴史的人物の国内および外国における移動プロフィールに統合するデジタルツールです。

CosmoToolは「DARIAH-DEフェデレーションアーキテクチャ」に基づいており、非構造化テキストからデータを抽出できます。現在、CosmoToolは開発段階にあり、まだ機能が制限されています。

- <https://cosmotool.de.dariah.eu/cosmotool/personsearch/>

Johann Wolfgang von Goethe

* 28. August 1749

† 22. März 1832

cosmotool / Biographische Daten / Johann Wolfgang von Goethe

Biographische Daten

from structured data

Zeitleiste

1749-8-28 (1) Geburt
Frankfurt am Main1749-8-28 (2) aus Textanalyse
Goethe-Haus1749-8-28 (3) aus Textanalyse
Frankfurter Großen Hirschgraben1765-1-1 (4) aus Textanalyse
Leipzig1768-3-1 (5) aus Textanalyse
Dresden1770-4-1 (6) aus Textanalyse
Universität Straßburg

(7) Studium

Ereignis-Details

Ereignis 5

aus Textanalyse: Dresden

Textstelle: "Auf Oesers Empfehlung besuchte er im März 1768 Dresden und die Gemäldegalerie."

Quelle: <http://de.wikipedia.org>*Extracted from full-text*

Kartendarstellung



Data Modeling Environment

- DARIAH-DEによるデータモデリング環境は、データのモデリングと関連付けのためのツールです。DMEというデータモデリング環境の特徴は、その研究指向の焦点と、ドメイン知識の解明のための基礎となる概念です。
- コードはGitHub:
<https://github.com/tgradl/dme>



Mapping-Editor

Data Modeling Environment (DME) / Datenmodelle und Mappings / Mapping-Editor

Quellmodell: **goerep_ksw** Zielmodell: **MMM**



Beispieltransformation

Input Ergebnisse 3 Ausführen ⬆ ☰

Quellmodell Zielmodell 3 / 3 ⬅ ➡

- M. (Absendeort)
 - **Spatial** [Neuhof](Empfangsort)
 - **TechnicalInformations**
 - **Format** text/XML
 - **Type** Handschrift
 - **Subtype** Brief
 - **Identifizier**
 - **Mww** ksw_goerep_10468
 - **External**
 - ksw_goerep_10468
 - **LicenceInformations**
 - **Rights** Bitte beim

Elementmodell

2017-08-30 17:01:10.695 SUCCESS Beispielinput transformiert (Dauer 127ms): 3 Ressourcen erkannt
2017-08-30 17:01:10.526 SUCCESS Beispielinput verarbeitet (Dauer 48ms): 3 Ressourcen erkannt
2017-08-30 17:01:10.323 SUCCESS Beispielinput transformiert (Dauer 145ms): 3 Ressourcen erkannt

州レベル: LOEWE

- 例: フランクフルト大学のCEDIFOR (人文科学、社会科学、教育学のデジタル研究センター)
- **目的** 人文系の研究とコンピュータサイエンスの方法論の間のギャップを埋めること
- 人文科学、社会科学、教育学者の研究におけるコンピュータサイエンスの方法の使用について助言およびサポートする方法論の専門知識を提供すること
- 研究データの持続可能な確保だけでなく、研究課題の計画と実施にも関係する

連邦政府の科研費機構 DFG

- Beispiel: Transkribus



Transkribus = 文字起こしのツール

- 歴史的文献の文字起こし・翻刻を行いますか？中世からの文献でも、20世紀の文献でも高度に標準化された、柔軟で信頼性の高い方法で翻刻を実行したいのですか？また、手書きのテキスト認識やレイアウト分析などの自動ツールのサポートを利用したいのですか？
- それならTranskribusはぴったり

Transkribus の特徴

- プラットフォームに登録する
- ニーズに合わせて特別に設計されたエキスパートツールをダウンロード
- 自分のドキュメント/画像をアップロードする
- 自分のプライベートコレクションを管理する(他の誰もあなたのドキュメントにアクセスできない！)
- レイアウト分析ツールのサポートにより、画像をブロック、ライン/ベースライン、単語に分割します
- 文字起こしの価値を大幅に高める画像とテキストをリンクします(実際には、リンクなしでテキストを文字起こしすることはできません！)

- 任意の言語および任意の文字セットでテキストを転記(独自の仮想キーボードをロード)
- TEI、RTF、PDF、XMLなどのいくつかの形式で常にドキュメントをエクスポートします

Transkribusのインターフェイス

The screenshot displays the Transkribus software interface. At the top, there is a toolbar with various icons for navigation and editing. Below the toolbar, the interface is divided into several sections:

- Left Sidebar:** Contains a 'Tags of current Transcript' table and a 'Tags Properties' table.
- Center:** A large image of a handwritten document in German, with some text highlighted in yellow.
- Bottom:** A numbered list of transcribed text segments, corresponding to the lines in the document image.

Tags of current Transcript Table:

Tag	Value	Text
1	date	17.1.6. [242] 17.1.6. Rilke Lieber
2	person	Rainer M. Rilke Lieber Herr Rainer M
3	abbrev	R. Herr Rainer Maria R. Ich h
4	place	Ham- gleich an die beste Ham-
5	place	burger die beste Ham- burger St
6	organization	Firma die Antwort der Firma. Es
7	place	Lokal ein ähnliches Lokal wie da
8	organization	Keller & Lokal wie das von Keller &
9	place	Berlin Keller & Reiner in Berlin, <
10	organization	Kunstge Ihren Brief an das Kunstge
11	organization	haus Hu Kunstgewerbe- haus Hult
12	textStyle	zu Hänc Kunstgewerbe- haus Hult
13	textStyle	des Her Hulbe zu Händen des Her
14	person	Herrn A zu Händen des Herrn Agt
15	textStyle	Matinée den Vortrag als Matinée u
16	textStyle	Sonntag und auf einen Sonntag ar

Tags Properties Table:

Tag specification	Color	Shortcut
abbrev	Red	Ⓜ
add	Cyan	Ⓜ
Address	Purple	Ⓜ
blackening	Black	Ⓜ
date	Blue	Ⓜ
div	Teal	Ⓜ
gap	Pink	Ⓜ
organization	Magenta	Ⓜ
person	Green	Ⓜ
place	Purple	Ⓜ
etc	Yellow	Ⓜ

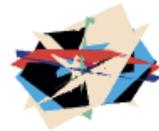
Transcribed Text Segments:

- 12 Doch möchte ich Ihnen in keiner Weise vorgreifen, teile
- 13 also der Firma Ihre Adresse nicht mit, sondern über-
- 14 lasse Ihnen selbst die Anknüpfung der Verhandlungen;
- 15 Sie würden dann gut tun, Ihren Brief an das Kunstgewerbe-
- 16 haus Hulbe zu Händen des Herrn Agte zu adressieren.
- 17 Ich empfehle Ihnen noch, den Vortrag als Matinée und
- 18 auf einen Sonntag anzusetzen; das entspricht am besten

ドイツ語圏における デジタル・ヒューマニティーズ

Programm Tagung Tickets Standorte Call for Papers

DHd 2019



multimedial &
multimodal

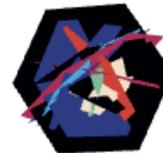


6. Jahrestagung
des Verbands
Digital Humanities im
deutschsprachigen Raum

Frankfurt &
Mainz



25.03. —
29.03.2019



DhD 学会、2019年

- 会議はメディアの概念とデジタル人文科学におけるその内訳に取り組むことに焦点を当てることを意図しています。
- 目的は、デジタル研究対象へのマルチモーダルアクセスと分析の可能性と制限について議論することです。
- 焦点は、モデリング（エンコーディング、インターフェイス、注釈）、分析（分類、セグメンテーション、ネットワーキング）の理論的および方法論的基盤（インター/クロスモダリティ、メディア理論）、合成（インタラクティブ3D再構築、設計主導型アプローチへの統合、インフラストラクチャの概念）に関する問題に置く
- もう一つの主題は、調査方法のアルゴリズムの基礎を議論・検討することです。

- 会議は86のワークショップ、パネル、講演
- 例 `Texterkennung mit Ocropy – Vom Bild zum Text`
`Ocropy`によるテキスト認識-画像からテキストへ
- 歴史的文献の光学式文字認識(OCR)は、認識率が低い。巧みな前処理とニューロネットワークに基づくモジュラーコマンドラインプログラムである`ocropy`(`OCRopus`)を使用すると、大幅に優れた結果を得ることができます。`Ocropy`はPythonで書かれており、ラスタグラフィックの生成、セグメンテーション(行に分割されたドキュメント)、誤った認識テキストの修正、新しい文字のトレーニング、そしてドキュメントの認識のためのモジュールを包含。
- ワークショップは、「ゼロから認識されたテキストファイルまで」すべての段階を紹介し、参加者も実習することが目的

DARIAH-DEがまとめたデータによると14ヶ所の大学が
DHを専攻にする学士課程・修士課程を提供する

- Bamberg
- Bielefeld
- Darmstadt
- Erlangen
- Frankfurt
- Gießen
- Göttingen
- Graz
- Groningen
- Hamburg
- Köln
- Lüneburg
- Saarbrücken
- Würzburg

フランクフルト大学におけるDH

- Cediforというサポート機構
- **研究** 例えば日本語・日本学に背景を持っているArmin Hoenen氏による文献系譜学 (Stemmatologie) 関係の研究

フランクフルト大学におけるDH

- Armin Hoenen: **An open problem in computational stemmatology - a model for contamination (2020)**

汚染は、テキストをコピーする時に複数の原稿または自分の記憶を基盤にした筆記者に起因するテキストバリエーションの混合のすべての現象を指す包括的な用語

「テキスト汚染」は、伝統のコピー履歴の再構築、ひいては元のテキストの再構築へのまったく異なるアプローチを意味するため、ステムマトロジーでは最大の課題の一つ

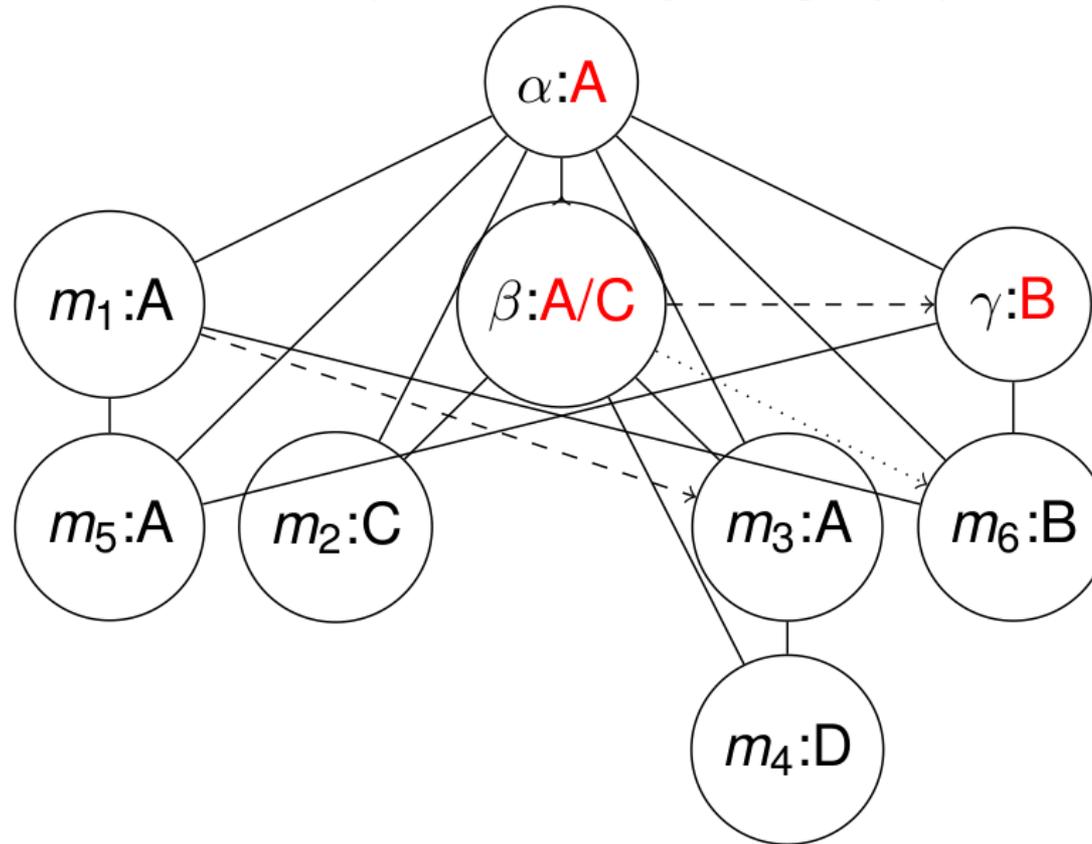
時間的な順序を維持しながら、汚染の程度のある伝統に対応するグラフ理論モデルを提示し、組み合わせの数と可能なシナリオの数を配慮に入れる

Model コンタミネーションと写本の系統樹



Horizontal transmission **コンタミネーションと****写本の系統樹**

A DAG (directed acyclic graph)



日本学におけるDH

- Heidelberg大学 > Judit Árokay教授
- 近世日本の旅行記を使用して、文学的空間と場所の地理的視覚化

このプロジェクトでは、文化研究の地形的変化 (topographical turn) を取り上げ、テキスト内の空間と場所を視覚化するための表現方法を開発します

日本学におけるDH

- Tübingen大学 > Robert Horres教授
- 日本の伝統的色彩

辞書式の視点と歴史的な色の問題の紹介から出発して、日本の伝統的な色の分野を考察する。伝統的な色彩システムの象徴性、使用状況、社会的エンコーディングは参照システムとして理解されており、今日までの日本社会のさまざまな状況で効果的

フランクフルト大学 日本学科

- Bastian Voigtmann M.A.
- Koray Birenheide B.A.
- Michael Kinski

Bastian Voigtmann 教育に関する勅語と注釈本

- 教育勅語関係資料 (1974-91) 164冊の注釈本を挙げる
- 教育勅語釈義全書 (1936) 140冊の注釈本を挙げる
- 多様な著者
- 何人かの有名な著者(内藤鶴鶴、井上哲次郎、井上円了)に加えて、あまり知られていない/ほとんど知られていない著者

- 問題提起

文体分析による実質的な近接性を実証可能？

- RstudioでStyloパッケージを使用した試験体の検査

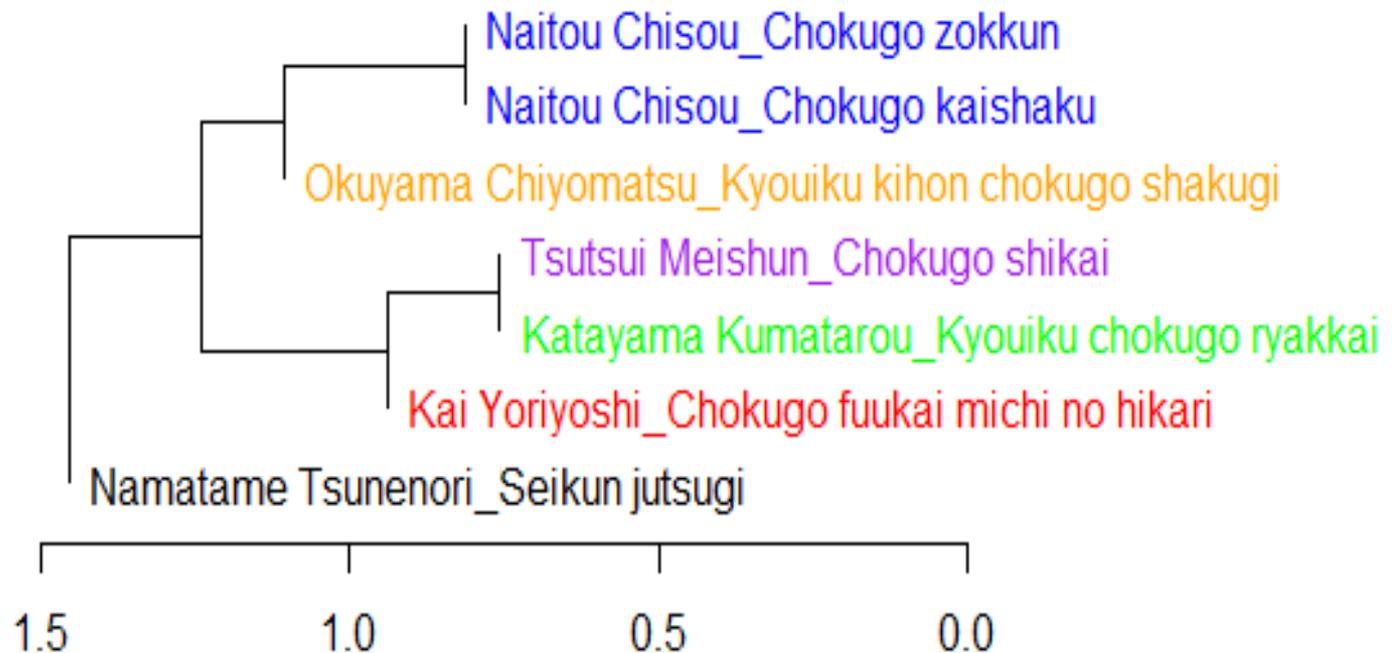
```
>>library(stylo)
```

```
>>stylo()
```

Eder, M., Rybicki, J. and Kestemont, M. (2016). Stylometry with R: a package for computational text analysis. R Journal, 8(1): 107-121, url: <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>

結果

Vortrag DA 2 Cluster Analysis

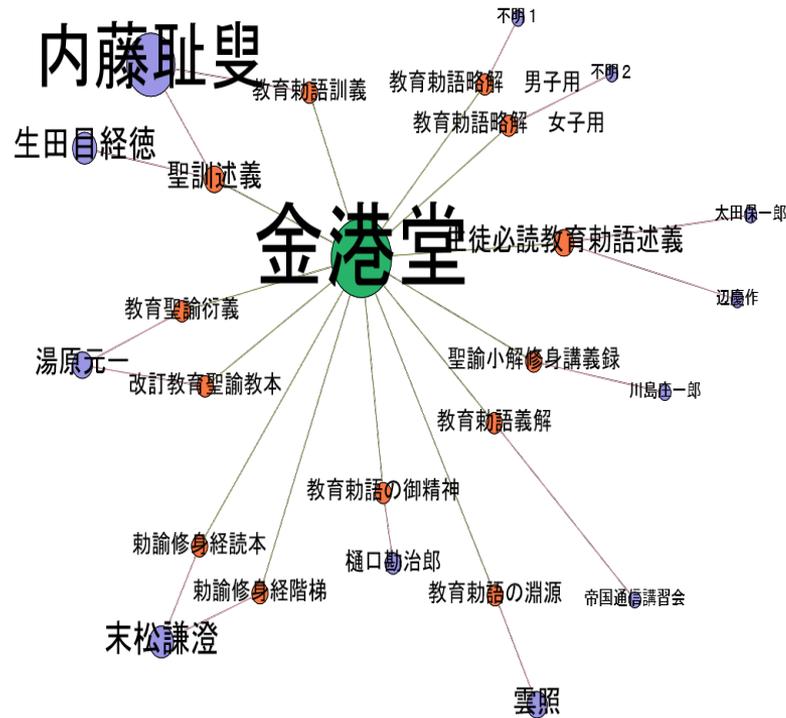


1000 MFC 3-grams Culled @ 0%
Classic Delta distance

- Digital Approaches 4000, 20./21.08.19
- Bastian Voigtmann (Frankfurt)

教育勅語に関する明治時代の解説者の 著者と出版社のネットワークの例

著者と出版社の関連



Koray Birenheide
DemiScriptの開発(プログラミング)

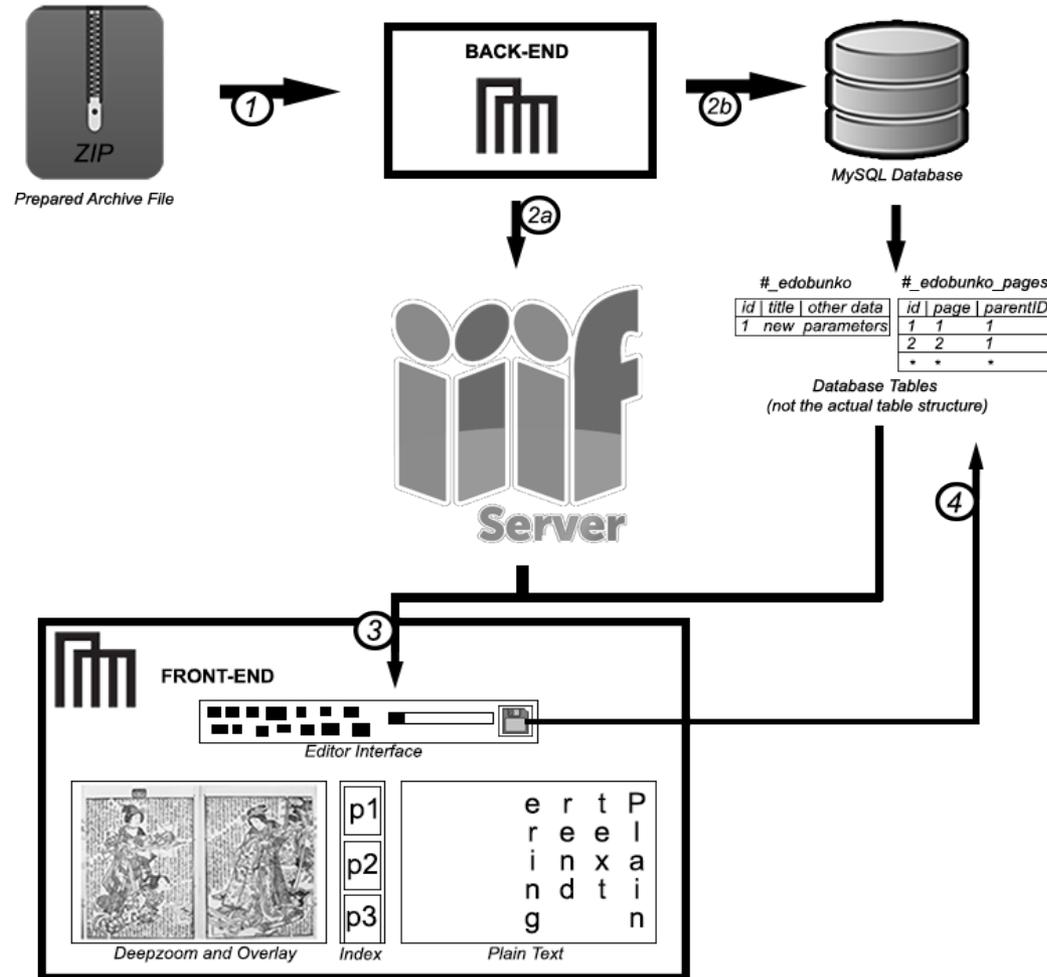
- **Planning and Implementation of a Digital Transcription and Annotation Software using Joomla! CMS and current Academic and Web Standards**

The Digital Corpus System: DemiScript

Utilized Web Technologies

- Involved languages are PHP, HTML5, CSS3, XML and JavaScript
- Joomla!: Advantages of a Content Management System
- IIF Server
- Utilized Libraries: JQuery, JQueryUI, OpenSeadragon und Fabric.js

The Digital Corpus System: DemiScript Compiling and Management of Digitized Documents



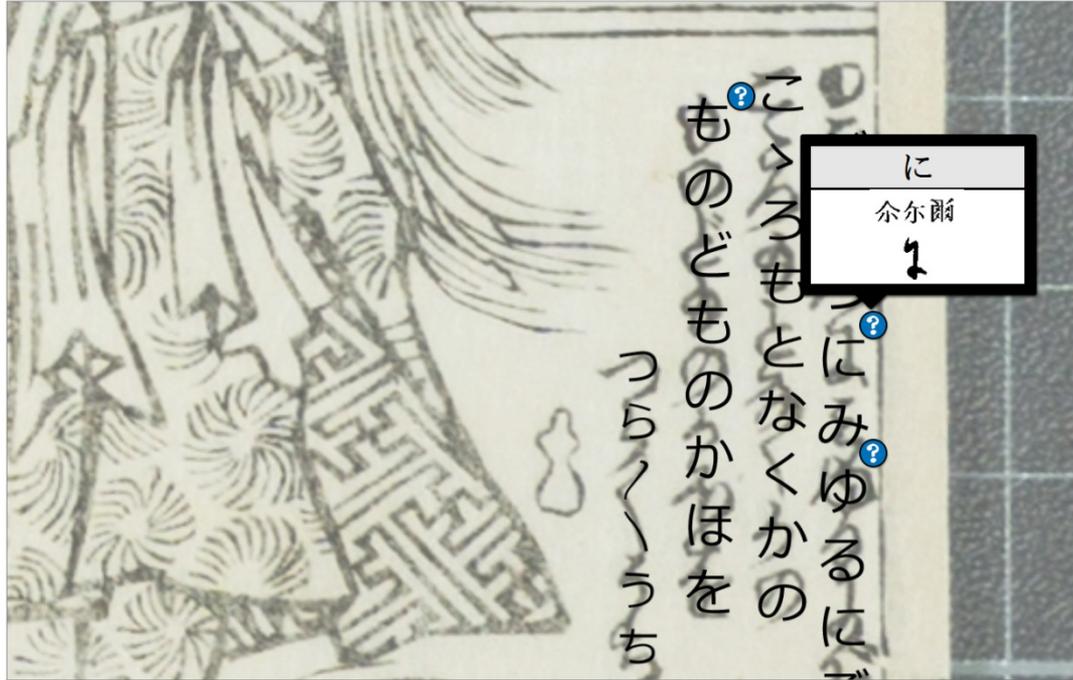
文字起こしとクラウドソーシング

Add Line or Edit Line Directly.

Export Image

セリフ

セリフ

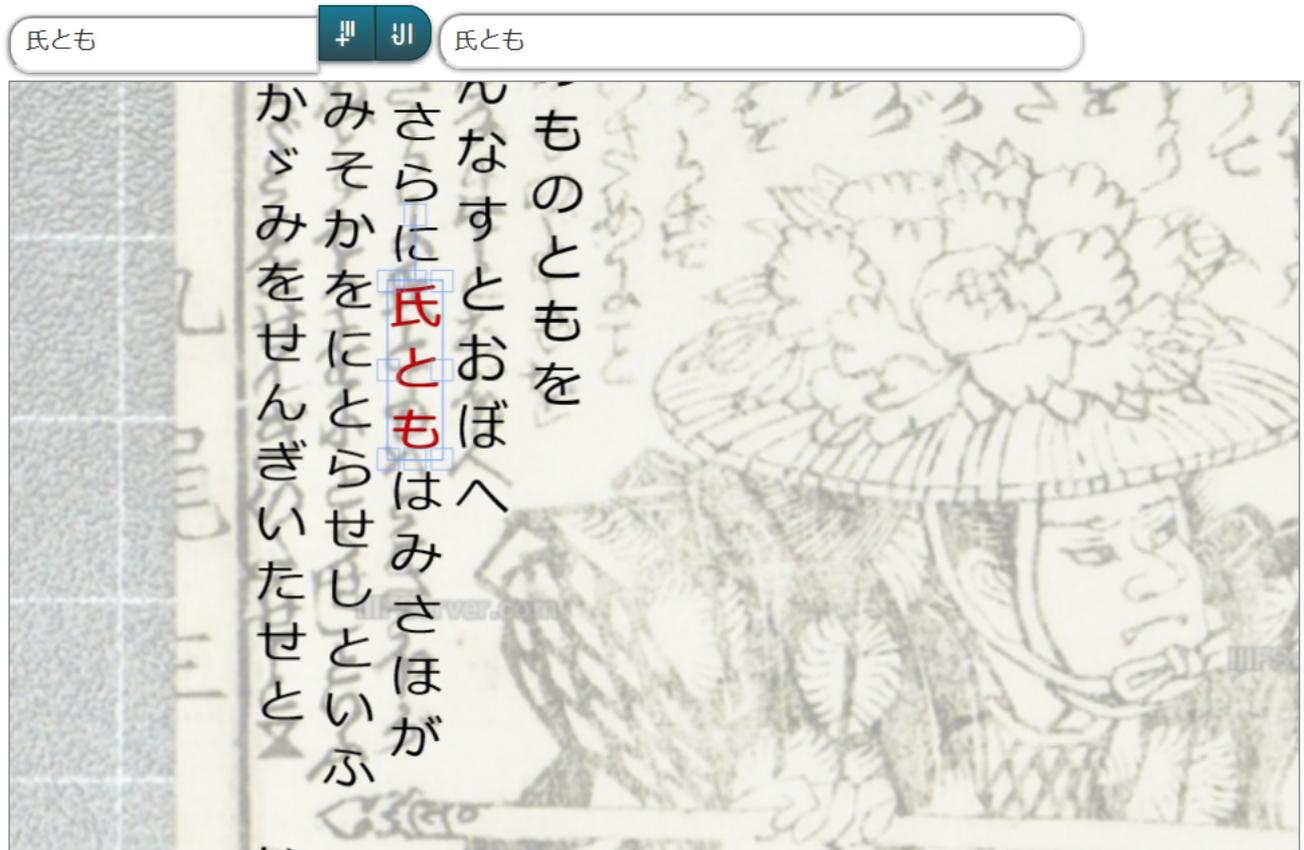


Sortable Text Segmented Text TEI Formatted



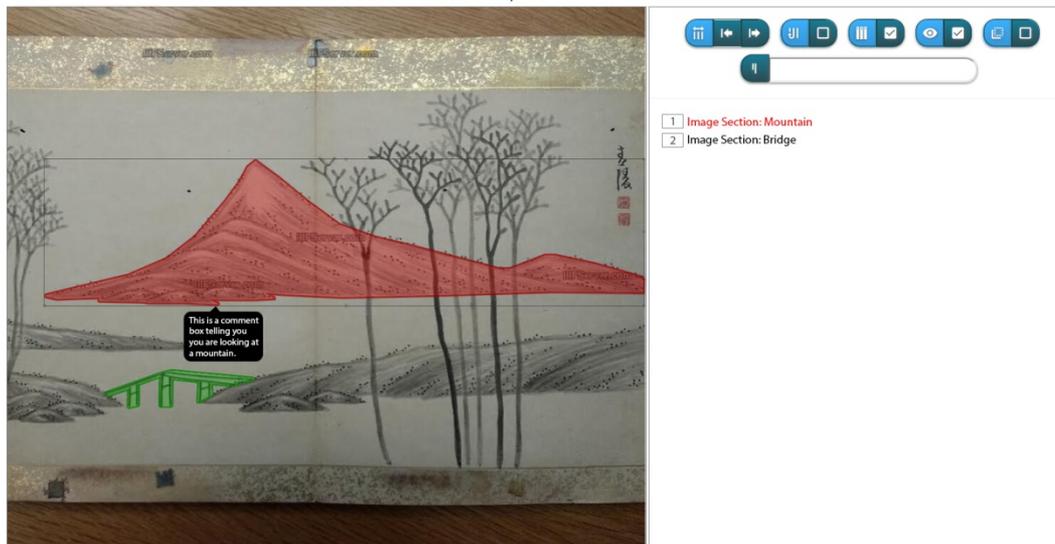
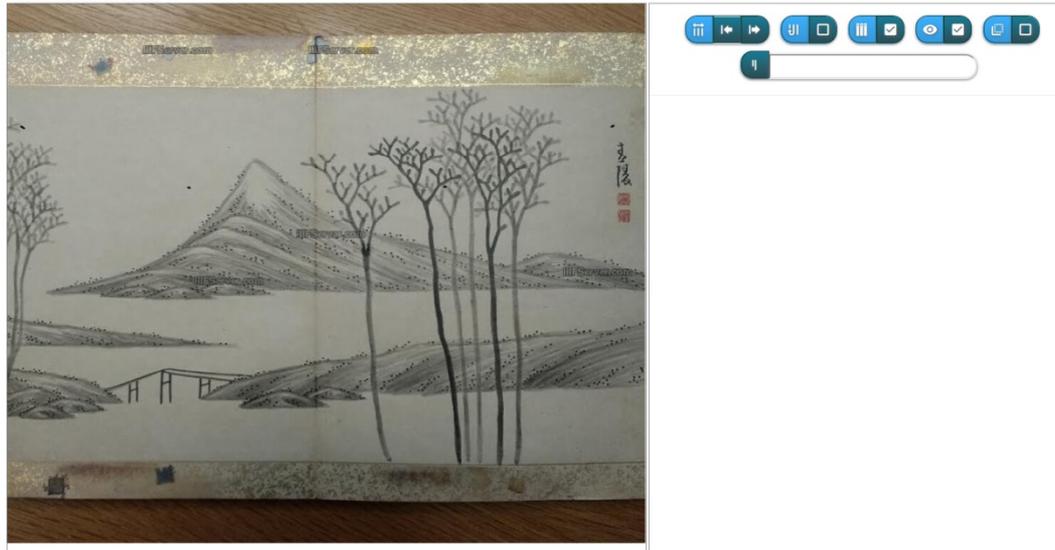
58 たげ
57 さま
56 いかにも
55 ながら
54 こと
53 一の巻より
52 二のまきへ
51 いったる
48 ちたげ
47 氏ともかうべを
46 ぬかづきたるゝなかに
45 さみおんまへちかく
44 大小とつてこゝにたは
43 うちにかくしちつたる
42 とりのけつゝいしやつの
41 両にんはかぶりしものを
40 そといわれてくだんの
39 めにものみすべき
38 たてばよしさなくは
37 そこそ
36 きりく
35 ためか
34 なさんが
33 さまたけ
32 しゆゑんの
31 さんせしは
30 めどほりへすい
29 にうちまぎれわが
28 ながらでんがくどち
27 さざるゝ身であり
26 たてしゆつしさへいた
25 してひやうきといひ
24 たてしゆつしさへいた
19 ふたりはさきたたつて
18 たねなんぢら
17 なり氏とちさだ
16 たてゝ「りよぐわい
15 たちまちこゑをふり
14 おどろきたるち氏が
13 両にんなるにぞうち
12 あらでちほとゆうきの
11 まゝとのやくしやに
10 ながむるにこほこれ
9 つらくうち
8 ものどものかほを
7 こゝろもたなくかの
6 ぶこつにみゆるにぞ
5 獅子しし
3 つゞききよに

文字起こしとクラウドソーシング



The Future of the Project

Photoshop Mock-ups



例: MK

- DHの「実験台」としての海保儀兵（海保青陵、鶴）

海保青陵（1755–1817）

- 中小規模の大名家に使えた家老クラスの家
族に長男として生まれた
- 荻生徂徠の弟子宇佐美瀧水に学んだ
- 儒学者でありながら、儒学を厳しく批判した
- 現代日本の先駆者として高く評価されること
もある

評判のきかけ

- 大名や将軍が持っている領土は財産で、そこで生産されているものは全部領主が利益をとるべきもの
- 宇宙全体の原理は「売り買い」に従う
- 大名が家来の労働力を買って、代金としてお米を上げる。侍が自分の力を主君に売り、代わりに知行か扶持米などをもらう

有名な主張

- 人間関係は全て市場の原理に基づいている

評判の裏腹

- 宇宙全体の原理を「売り買い」にみるのことは江戸時代の文脈でみて過激的
- 青陵の主張に過激的な見方は少なくない
- 非差別民(穢多、非人間)などについてにそれが端的に現れる

過激的な意見

- 非差別民(穢多、非人間)を例に
 - 外の遊民
 - 穢多の特殊性

- Luke Roberts: *Mercantilism in a Japanese Domain*, 1998: 200.

The **eccentric** Kaiho (Kaibo) Seiryô (1755–1817) was an independent scholar who gave economic advice to a number of domains in the early nineteenth century. When Kaiho wrote his treatise *Keiko dan* sometime after 1811, he performed **impressive and humorous verbal acrobatics** by substituting the character “profit” (*ri*) for “principle” (*ri*) as a guide to good government. He **stood the Confucian verities on their heads** and turned the proper management of profit into a public activity of the highest order. The world had changed such that his **festive inversion** could not be taken purely as **joke**, and yet because the rhetoric was so **subversive**, he **could not be taken in full seriousness** by contemporaries or even some modern.

風変わりな 印象的でユーモラスな口頭アクロバット 儒学の価値観を覆した 滑稽な逆転 真剣に受け止められがたい

青陵の著作

- 海保青陵が残した著作の多くには「稽古談」「升小談」「富貴談」といった具合に「談」という名がつけられている
- どれもまるで目の前の誰かに語りかけているかのように、ざっくばらんで軽快な青陵の語り口を伝えている
- おしゃべりで座を盛り上げ、時には酒も交えたにぎやかな場面が彼の行く先々で見られたのだろう
- 当時の評判 「談話おもしろき人」

青陵の著作 二つの側面

- 青陵の特異的な言葉遣いまたは話法に注目しながら、経済思想の提唱者としての青陵と、ふざけた印象を与える文章のスタイルや飛んだ発想との持ち主としての青陵
- 矛盾？

青陵の著作

- 矛盾ではない。むしろ ...
- 自分の言葉使いを楽しみながら、聞き手を混乱させたり面白がらせたりしているのではないのでしょうか
- 思想的內容より青陵が文章の雰囲気と話法を大事にしている印象は強い
- 聞き手・読者の先入観を弄びながら、テキストに対する反応を操作しようとしている
- 読書の世界より、演説や談話の世界で聞き手との対話と共通認識を可能にするために、話し手との繋がりを強化して聞き手の反応を誘導することが大事
- そういう意味で青陵の作品が彼の話法を垣間見させる

青陵の著作

- 聞き手との共通の思想空間、言説空間を前提にし、それとの一致を守りながら、そこに隠されている誰も意識していないおかしい側面を浮き彫りさせたり、あるいは意外な方向へ向かわせたり、共通認識を誇張したりしている
- 思想家や哲学者がやることよりはむしろ政治風刺家の目指すことを思わせているのではないのでしょうか
- パロディー Simon Dentith identifies two characteristics: it works as a conservative force „in the way that it is used to mock literary and social innovation, policing the boundaries of the sayable,“ but it can also be subversive in that it „attacks the official word, mocks the pretensions of authoritative discourse, and undermines the seriousness“ through which domination is achieved. (Parody, 2000: 20)

青陵の著作

- Simon Dentithによる
パロディーの定義

- 二つの特徴:

- 保守的な要因 „文学と社会の革新を嘲笑する
ために使用されます“
- 反体制的な要因 „社会の支配が達成される深
刻さを損なう
- . (Parody, 2000: 20)



青陵の著作

- 青陵が共通認識を基礎に聞手・読者の理解を操作したり支配的・慣例的な価値体制をパロディー化することを証明できるのか
- それを証明するために青陵や聞手・読者が共通した言論の空間を復元し、彼の言葉遣いをその中に位置付ける必要がある
- text mining / topic modellingをしようとするのはそのためだ

現在の段階：初期段階

- text miningやtopic modellingを以て得られる結果
- 手がかかり：
 - Franco Moretti „distant reading vs. Close reading“
 - Sara KerrのJane Austen研究

F. Moretti



Figure 1: Length of Titles

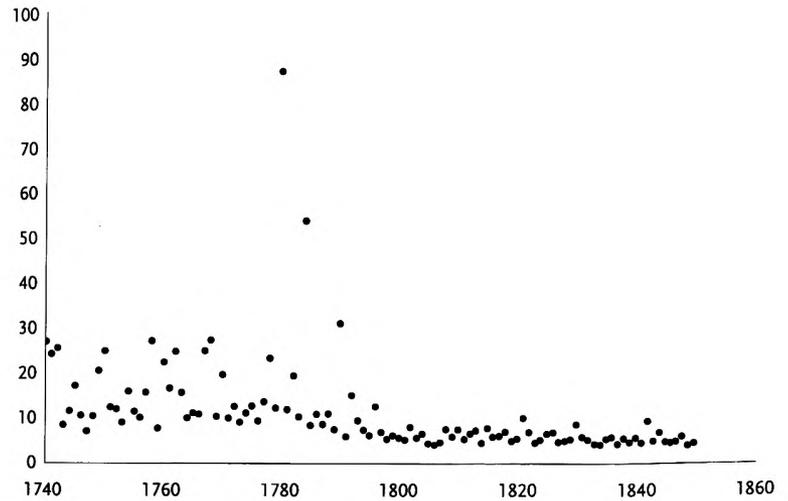
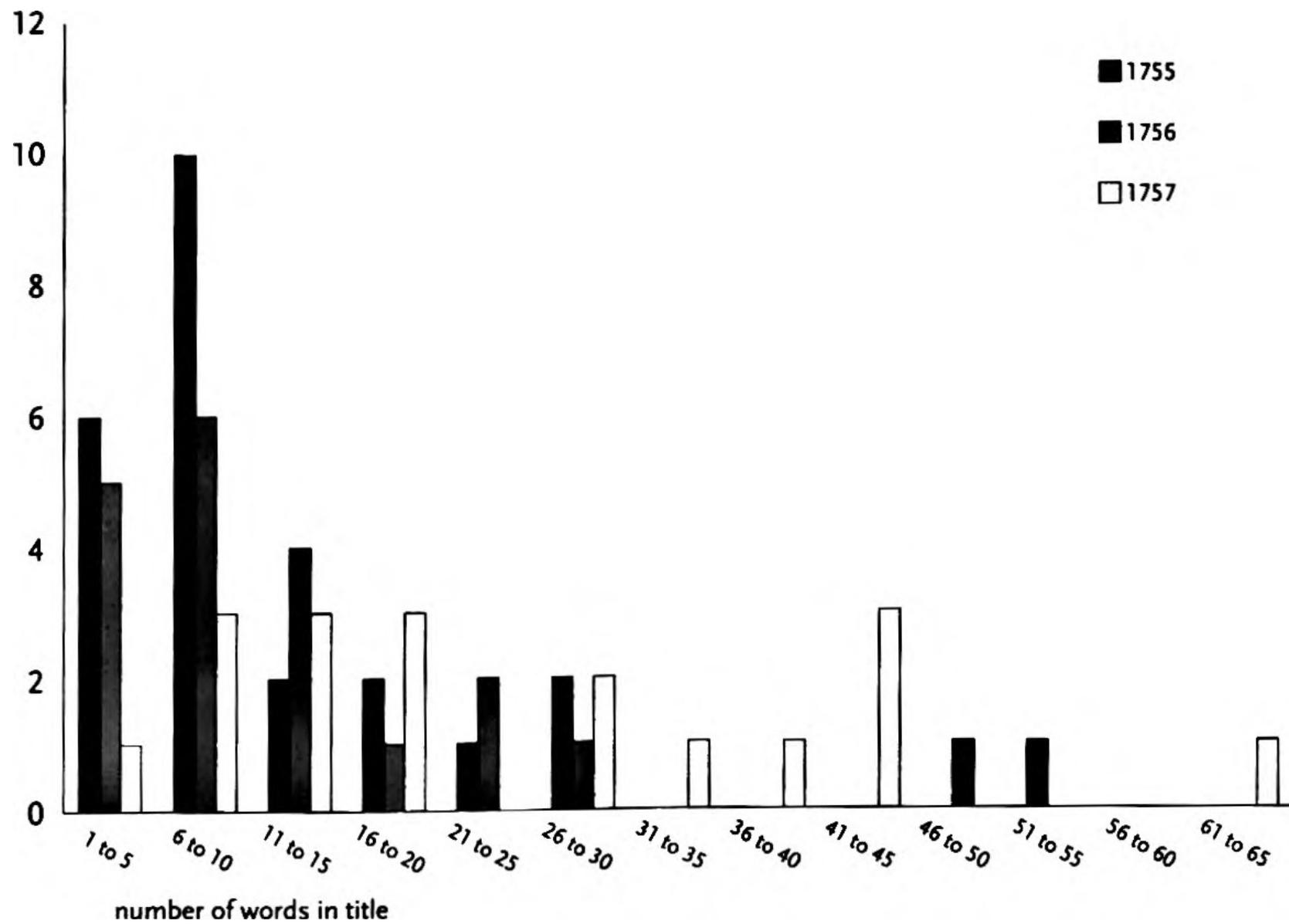
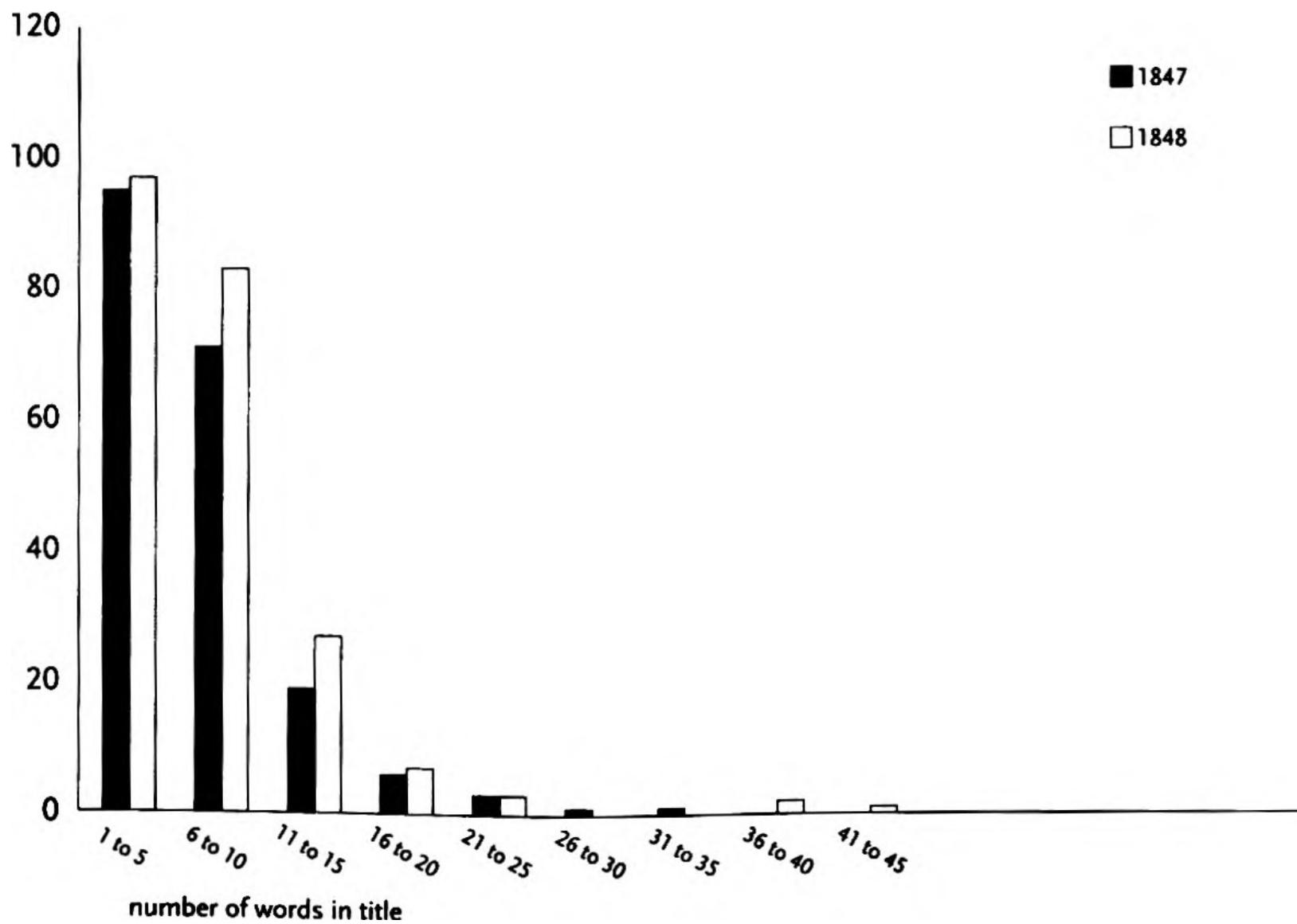


Figure 2: British novels 1740-1850, standard deviation from the mean





Sara Kerr

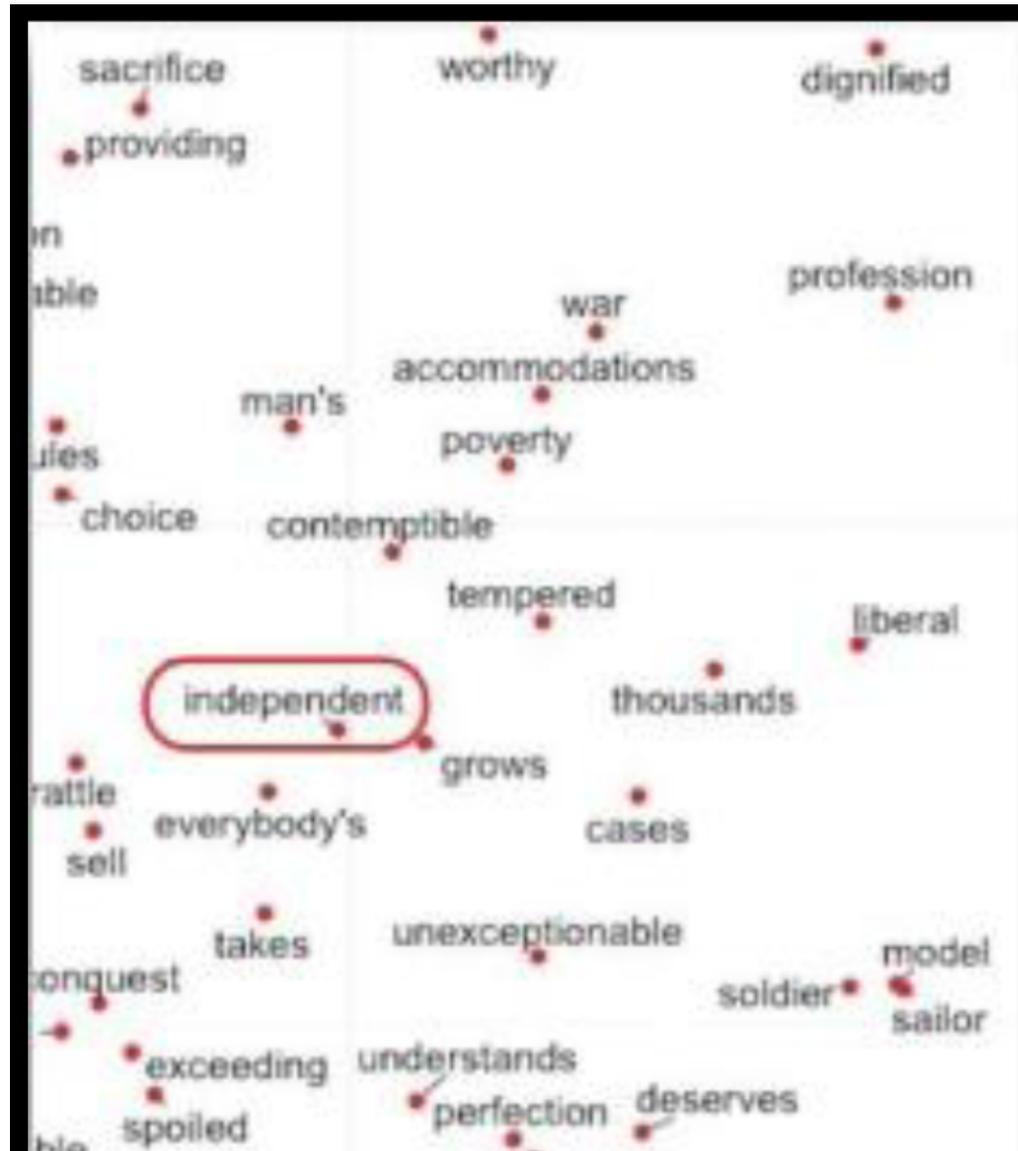


Figure 1: Independence in Austen

....問題点が多い

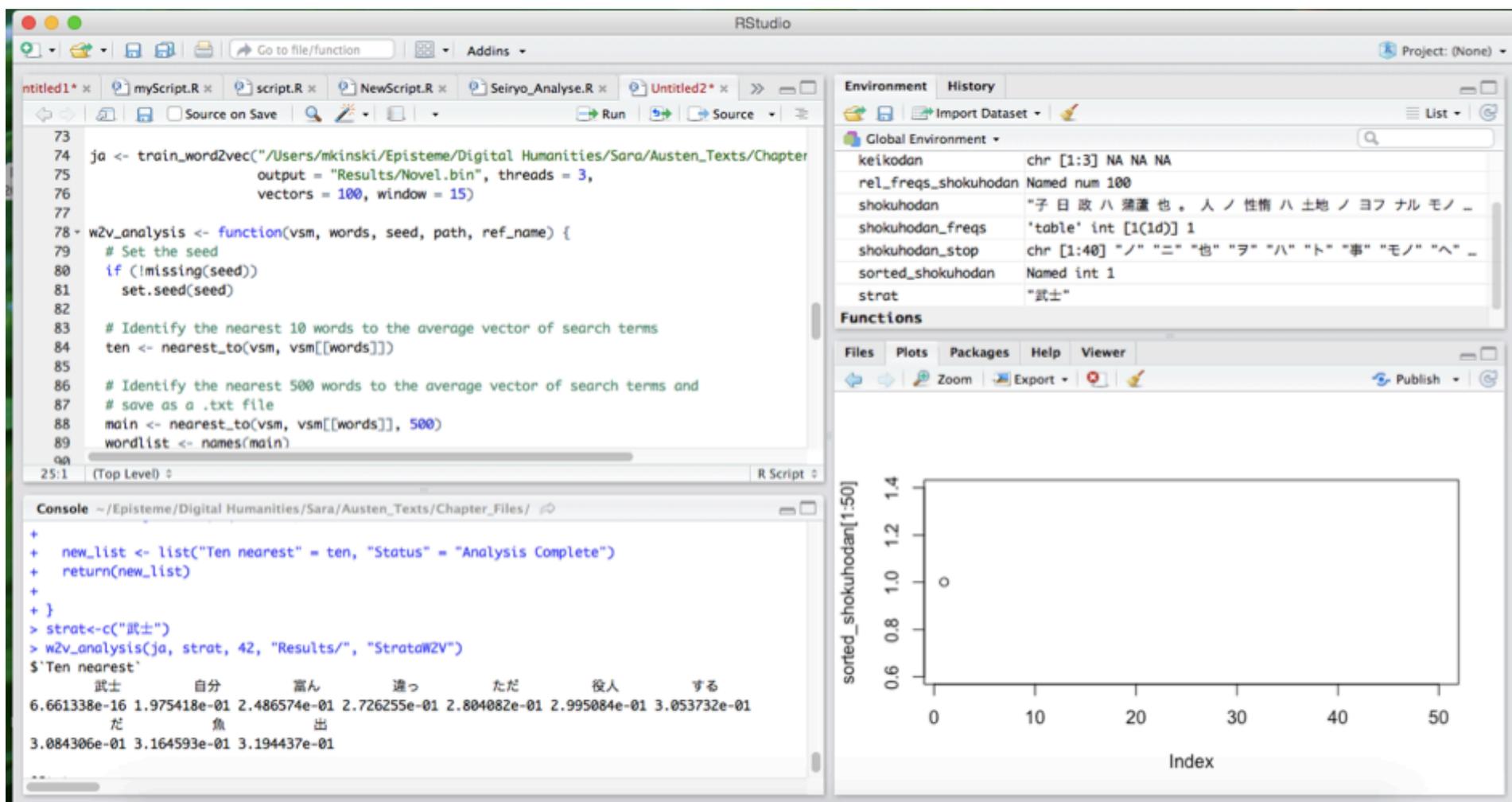
- コーパスの有無
- 日本語環境に対応できるか
- DHツールの活かし方 & 分析方法

Rstudioでの分析手順

コーパスの問題

- 狙い: 文献をOCR化して、分析可能にすることだ
- 例の紹介
 - 1. word2vec
 - 2. vector space

Rstudioによる分析 1. word2vec 言葉と言葉の意味上の距離を測り、関連性を 視覚化する



```

15 vectors = 100, window = 15, force=TRUE)
16 words<-c("悪党")
17 main <- nearest_to(vsm, vsm[[words]], 500)
18 wordlist <- names(main)
19 new_model <- vsm[[wordlist, average = F]]
20 reduction <- Rtsne(as.matrix(new_model), dims = 2, initial_dims = 50,
21                   perplexity = 30, theta = 0.5, check_duplicates = F,
22                   pca = F, max_iter = 1000, verbose = F,
23                   is_distance = F, Y_init = NULL)
24
25 df <- as.data.frame(reduction$Y)
26 rows <- rownames(new_model)
27 rownames(df) <- rows #wenn man das überspringt werden Zahlen geplottet
28 Encoding(rows)<-"UTF-8"
29 filepath <- paste0("Results/", "TestStrata")
30 # Save dataframe as .csv file
31 write.csv(df, paste0(filepath, ".csv"))
32
33 #jetzt den plot; ggf. müssten Sie noch die quartz HiraMaru zeile voransetzen und das theme
34 ref_name<-"TestStrata"
35 path<-"Results/"
36 theme_set(theme_classic(base_size = 18, base_family = "HiraKakuProN-W3"))
37 ggplot(df, aes(x = V1, y = V2, label = rows)) +
38   geom_point(size=2, color = "red") +
39   geom_text_repel(size = 8, family = "HiraKakuProN-W3") +
40   xlab("Dimension 1") +
41   ylab("Dimension 2 ") +
42   # geom_text(fontface = 2, alpha = .8) +
43   theme_bw(base_size = 12) +
44   theme(legend.position = "none") +
45   ggtitle(paste0("2D reduction of VSM ", ref_name, " using t_SNE"))
46
47 ggsave(paste0(ref_name, ".jpeg"), path = path, width = 24,
48        height = 18, dpi = 100)
49
50
48:39 (Top Level)

```

```

Console /Frankfurt/Episteme/DH Projekt/Sara/Austen_Texts/Chapter_Files/
> wordlist <- names(main)
> new_model <- vsm[[wordlist, average = F]]
> reduction <- Rtsne(as.matrix(new_model), dims = 2, initial_dims = 50,
+                   perplexity = 30, theta = 0.5, check_duplicates = F,
+                   pca = F, max_iter = 1000, verbose = F,
+                   is_distance = F, Y_init = NULL)
> df <- as.data.frame(reduction$Y)
> rows <- rownames(new_model)
> rownames(df) <- rows #wenn man das überspringt werden Zahlen geplottet
> Encoding(rows)<-"UTF-8"
> filepath <- paste0("Results/", "TestStrata")
> write.csv(df, paste0(filepath, ".csv"))
> ref_name<-"TestStrata"
> path<-"Results/"
> theme_set(theme_classic(base_size = 18, base_family = "HiraKakuProN-W3"))
> ggplot(df, aes(x = V1, y = V2, label = rows)) +
+   geom_point(size=2, color = "red") +
+   geom_text_repel(size = 8, family = "HiraKakuProN-W3") +
+   xlab("Dimension 1") +
+   ylab("Dimension 2 ") +
+   # geom_text(fontface = 2, alpha = .8) +
+   theme_bw(base_size = 12) +
+   theme(legend.position = "none") +
+   ggtitle(paste0("2D reduction of VSM ", ref_name, " using t_SNE"))
> ggsave(paste0(ref_name, ".jpeg"), path = path, width = 24,
+        height = 18, dpi = 100)
>

```

Environment History

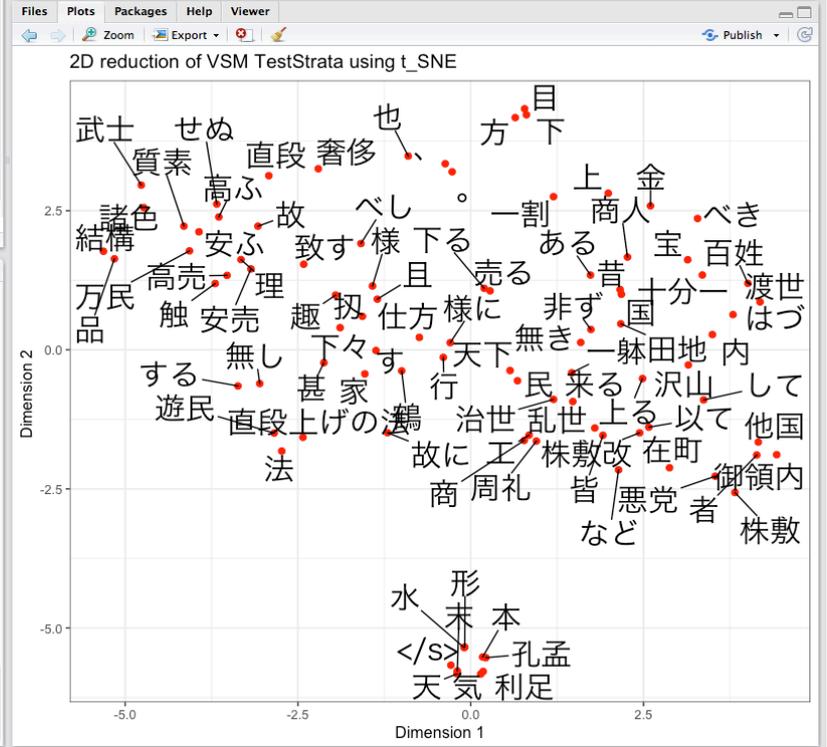
Global Environment

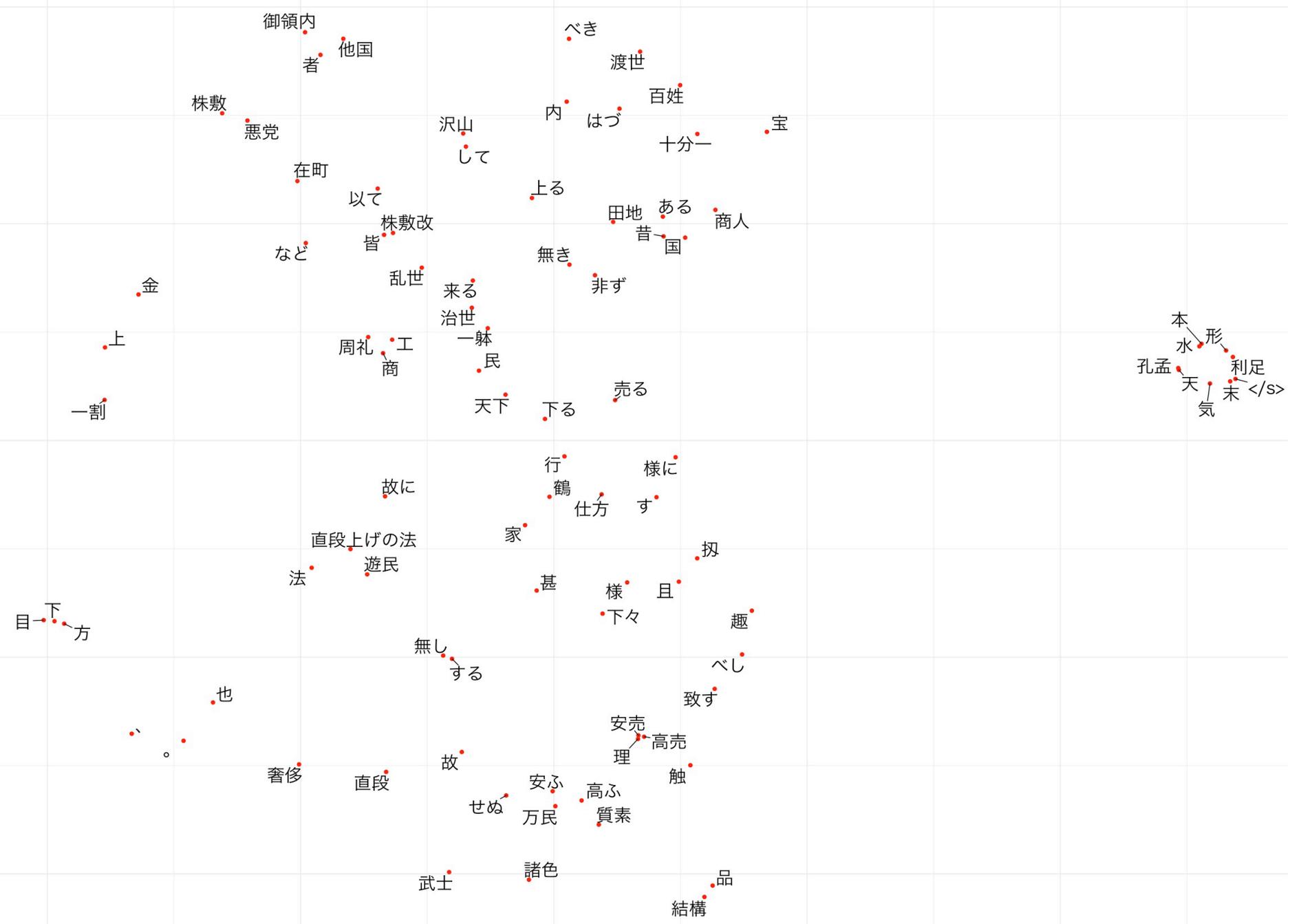
Data

df 92 obs. of 2 variables

Values

filepath	"Results/TestStrata"
main	Named num [1:92] 1.33e-15 9.14e-03 3.01e-02 3.90e-02 4.68e-02 ...
new_model	Formal class VectorSpaceModel
path	"Results/"
reduction	List of 13
ref_name	"TestStrata"
rows	chr [1:92] "</s>" "、" "、" "也" "上" "奢侈" "べし" "する" "下" "べき" "故に" "様...
vsm	Formal class VectorSpaceModel
wordlist	chr [1:92] "悪党" "在町" "株敷改" "株敷" "など" "他国" "乱世" "以て" "苗" "御領内" ...
words	"悪党"





本水形
孔孟天氣
利足
未

安売理
高売
触

武士
諸色

品
結構

御領内
他国
者

株敷
悪党

沢山
して

べき
渡世

内
上る

はづ
百姓
十分一

宝

在町

以て
株敷改
皆

田地
ある
昔
国
商人

金

など

乱世

無き

非ず

上

周礼
工商

来る
治世
一寐
民

下る
売る

一割

天下

下る

本水形
孔孟天氣
利足
未

行
鶴

仕方

様に

故に

家

甚

様

且

扱

趣

目
下
方

無し
する

べし

致す

也

直段
遊民
法

甚

様

且

扱

趣

奢侈

直段

故

安ふ
万民

高ふ
質素

安売
理

高売

触

、

。

せぬ

万民

高ふ
質素

安売
理

高売

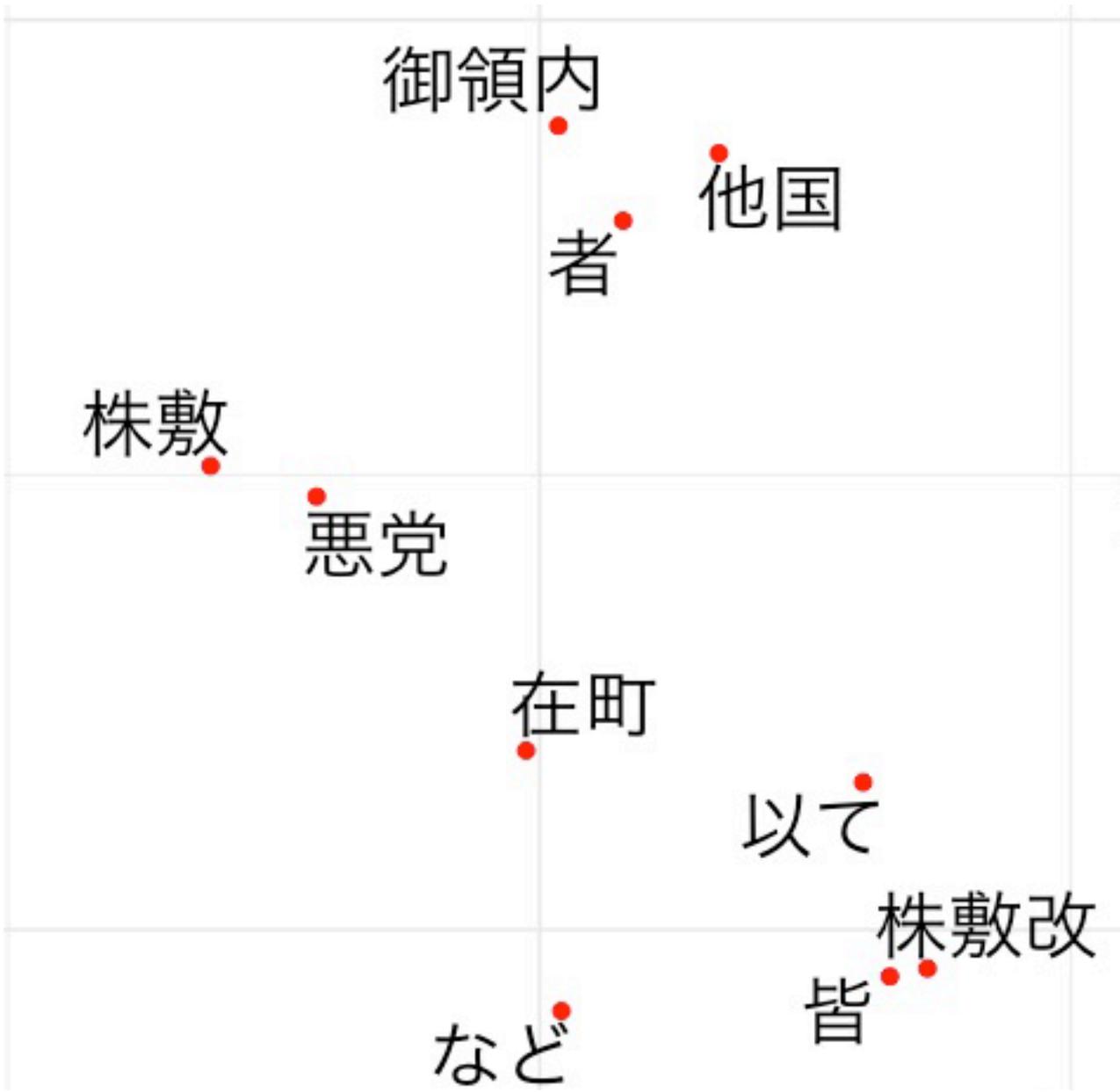
触

武士

諸色

品

結構



Rstudioによる分析 2. vector space

- **計量文献学とは**
- **文献の特徴を数値化し、統計学的手法を用いて文献の分析や比較を行う方法**
- **文献の著者の異同や時代の推定、同一著者の著作内容・思想と文体との関係の解析**

計量文献学～著者分析

- 「明六雑誌」 vs. 青陵の著作

Rstudio: 分析結果

The screenshot displays the RStudio interface with two main panels. The left panel shows R code for plotting authors based on MDS data. The right panel shows the Environment pane with a list of data frames and a scatter plot of the authors.

```
geom_point(size=2) +
1 geom_text(aes(label=authors, family="HiraKakuProN-W3"), hjust="inward", vjust="inward",
2 theme(axis.text.x=element_blank(),
3 axis.ticks.x=element_blank(),
4 axis.text.y=element_blank(),
5 axis.ticks.y=element_blank(),
6 legend.title=element_blank()) +
7 xlab("Coordinate 1") +
8 ylab("Coordinate 2")
9
0
1 #import dendrogram library
2 library(ggdendro)
3
4 #run the HCA function on our distance matrix, using the "ave" method
5
6 (Top Level) ↕
```

Environment

Object	Size
mds_data	53 obs. of 4 variables
Meikai.df	53 obs. of 5 variables
Mei roku50.df	50 obs. of 5 variables
nostop.norm.df	53 obs. of 10540 variables
nostop.tf idf.df	53 obs. of 10540 variables
ordered.freq.df	10556 obs. of 2 variables
reduced.norm.df	53 obs. of 3095 variables

Files **Plots** **Packages** **Help** **Viewer**

Zoom Export Publish

Coordinate 2

Coordinate 1

```
sole ~/Episteme/Digital Humanities/Sara/Austen_Texts/Chapter_Files/ ↗
<- fit$points[,1]
<- fit$points[,2]
s_data <- data.frame(cbind(x, y, authors, titles))
brary(ggplot2)
plot(data=mds_data, aes(x = x, y = y)) +
geom_point(size=2) +
geom_text(aes(label=authors, family="HiraKakuProN-W3"), hjust="inward", vjust="inward", size
+
theme(axis.text.x=element_blank(),
axis.ticks.x=element_blank(),
axis.text.y=element_blank(),
axis.ticks.y=element_blank(),
legend.title=element_blank()) +
xlab("Coordinate 1") +
ylab("Coordinate 2")
```



```
hca_data <- hclust(euc_dist_matrix, "ave")
#add our metadata labels
hca_data$labels <- authors

#set the font first
theme_set(theme_gray(base_size=12, base_family="HiraKakuProN-W3"))

#now plot with font size = 2 and rotate the plot 90 degrees
ggdendrogram(hca_data, rotate = TRUE, size = 2)

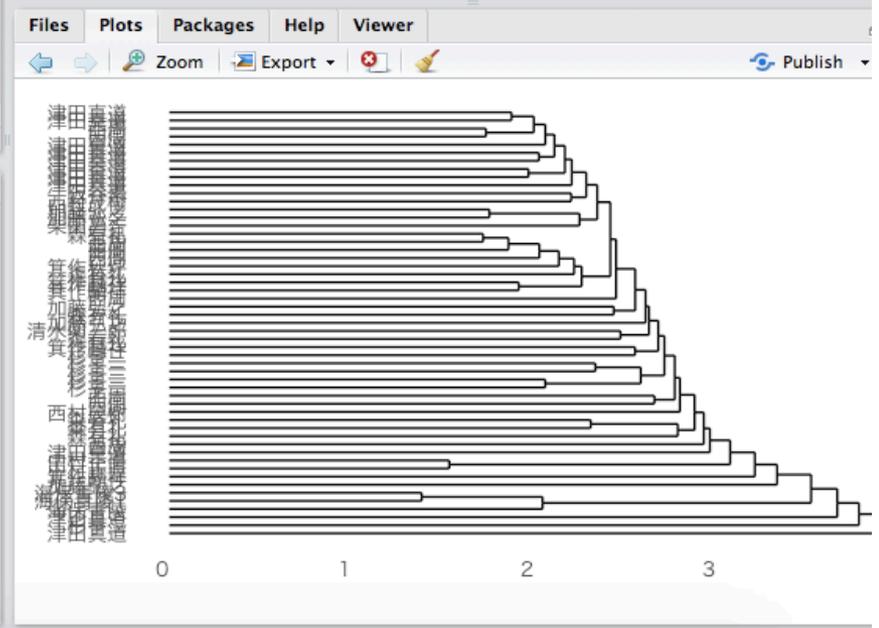
#cut the dendrogram into 2 subtrees; try with other values of k
plot(hca_data)
rect.hclust(hca_data, k=2)
```

```
geom_point(size=2) +
geom_text(aes(label=authors, family="HiraKakuProN-W3"), hjust="inward", vjust="inward", size
+
theme(axis.text.x=element_blank(),
axis.ticks.x=element_blank(),
axis.text.y=element_blank(),
axis.ticks.y=element_blank(),
legend.title=element_blank()) +
labs("Coordinate 1") +
ylab("Coordinate 2")
library(ggdendro)
hca_data <- hclust(euc_dist_matrix, "ave")
hca_data$labels <- authors
theme_set(theme_gray(base_size=12, base_family="HiraKakuProN-W3"))
dendrogram(hca_data, rotate = TRUE, size = 2)
```

Environment History

Global Environment

Value	Class
authors	chr [1:53] "西周" "西村茂樹" "加藤弘之" "森有礼" "津田真道"
Complete.meikai	"洋字を以て國語を書するの論 西周 吾輩 日常 二三"
Complete.meikai.spl...	Large character (96045 elements, 1.2 Mb)
euc_dist_matrix	Class 'dist' atomic [1:1378] 1.9 2.3 2.12 3.48 1.74 ..
fit	List of 5
hca_data	List of 7



RStudio

Project: (None)

Environment History

Global Environment

tt 10 obs. of 3487 variables

Values

authors	chr [1:53]	"西周" "西村茂樹" "加藤弘之" "森有礼" "津田真道"...
Complete.meikai		"洋字を以て國語を書するの論 西周 吾輩 日常 二三 ..."
Complete.meikai.spl...	chr [1:15351]	"洋字" "を" "以" "て" "國語" "を" "書する" ...
euc_dist_matrix	Class 'dist' atomic [1:45]	2.08 2.31 2.38 2.94 1.59 ...
fit	List of 5	
hca_data	List of 7	

Files Plots Packages Help Viewer

Zoom Export Publish

```

159 hca_data <- hclust(euc_dist_matrix, "ave")
160 #add our metadata labels
161 hca_data$labels <- authors
162
163 #set the font first
164 theme_set(theme_gray(base_size=12, base_family="HiraKakuProN-W3"))
165
166 #now plot with font size = 2 and rotate the plot 90 degrees
167 gg dendrogram(hca_data, rotate = TRUE, size = 2)
168
169
170 #cut the dendrogram into 2 subtrees; try with other values of k
171 plot(hca_data)
172 rect.hclust(hca_data, k=2)
173
174
170:1 (Top Level)

```

Console ~/Episteme/Digital Humanities/Sara/Austen_Texts/Chapter_Files/

```

ze=3) +
+ theme(axis.text.x=element_blank(),
+ axis.ticks.x=element_blank(),
+ axis.text.y=element_blank(),
+ axis.ticks.y=element_blank(),
+ legend.title=element_blank()) +
+ xlab("Coordinate 1") +
+ ylab("Coordinate 2")
警告メッセージ:
Removed 42 rows containing missing values (geom_text).
> library(ggdendro)
> hca_data <- hclust(euc_dist_matrix, "ave")
> hca_data$labels <- authors
> theme_set(theme_gray(base_size=12, base_family="HiraKakuProN-W3"))
> gg dendrogram(hca_data, rotate = TRUE, size = 2)
>

```

パロディーの発見 一つの手がかり KWICを抽出

• 例 とんと

The screenshot shows the KWIC concordance software interface. At the top, the title bar reads 'KWICコンコーダンス'. Below it is a search entry area with the following fields: '抽出語: とんと', '品詞: タグ', '活用形: ', and a '追加条件' button. Below these are sorting options: 'ソート1: 出現順', 'ソート2: 出現順', 'ソート3: 出現順', and '(前後 24 語を表示)'. A '検索' button is on the right, and the status 'Ready.' is at the bottom right.

The main area is titled 'Result' and displays a list of concordance lines. Each line consists of a left column of text, a middle column of the search term 'とんと', and a right column of text. The text is from classical Japanese sources, discussing various topics like land, governance, and social hierarchy. The search term 'とんと' appears in various contexts, often as a particle or a specific term.

At the bottom of the window, there is a control bar with buttons for 'コピー', '文書表示', '表示単位: H5', '前200', '次200', 'ヒット数: 19, 表示: 1-19', '保存', and '集計'.

今後の方針

「思想」・「思想家」を考え直す

- 内容 対 形式
- 思想 対 レトリク

>>>

- Sentiment analysis の可能性を探る

Sentiment analysis 文本情感分析

- **Opinion mining** (sometimes known as **sentiment analysis** or **emotion AI**) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and **study affective states** and **subjective information**.
- 意見マイニング（感情分析または感情AIとも呼ばれる）は、自然言語処理、テキスト分析、計算言語学、およびバイオメトリクスを使用して、感情的な状態と主観的な情報を体系的に識別、抽出、定量化、および調査することを指します。

- „I'd really truly love going out in this weather!“
- A basic task in sentiment analysis is classifying the *polarity* of a given text at the document, sentence, or feature/aspect level—whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry", "sad", and "happy".

- Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored.
- More sophisticated methods try to detect the holder of a sentiment (i.e., the person who maintains that affective state) and the target (i.e., the entity about which the affect is felt).