テキスト分析

福田真人

“Digital History and Argument”

“The Differences between Digital Humanities and Digital History”

「計量テキスト分析および*KH Coder*の利用状況と展望」

Digital History and Argument ①

テキスト分析:広く利用されるデジタルヒストリー手法

アルゴリズム依存の非価値中立的手法

Topic modeling：トピックの出現率・分布・トピックの傾向

Word-embedded modeling：単語・語群を空間上に表現

ネットワーク分析と関連

Identifying text reuse：引用などの特定

cf)百科全書

Digital History and Argument ②

“Stir the archives”

優位点：テキストの連続性と非連続性

パターンと例外

⇒時代区分論

長期・大量のテキスト分析

イギリス議会文書・裁判記録などの事例

Digital History and Argument ②

“Stir the archives”

優位点：テキストの連続性と非連続性

パターンと例外

⇒時代区分論

長期・大量のテキスト分析

イギリス議会文書・裁判記録などの事例

The Differences between Digital Humanities and Digital History ①

テキスト分析：一定の利用 but 歴史分野では限定的な面も

仮説なしボトムアップ法という画期性（？）

研究者が読まない単語をチェック可能

テキストデータ量の制約　検索用データの傾向

文学に比べ非印刷物の利用度合いが高い

API化やオープンデータ化の欠如　商業ソフト依存

文脈によらない　階層理解の欠如　サーチ単語の妥当性問題

The Differences between Digital Humanities and Digital History ②

新聞データベース　Chronicling　America　project

利用規約自由　データに欠損（新聞の種類・時期）

著作権やOCRの正確性といった問題

新聞以外の少数のテキスト分析（外交・ジェンダー・投票等）

歴史学者の方法論・デジタルソースへの信用が

研究の進展・データの拡充に重要

歴史学者の理解を得るためには人文情報学ではなく、

細分化されたデジタルヒストリーを軸にすべき

計量テキスト分析およびKH Coderの利用状況と展望　①

テキスト分析の目的：データ探索・信頼性の向上

事例：Twitter　雑誌（アサヒ芸能等）

　　アンケート　インタビュー　会議録

共通の枠組みの必要性（時期・年代の曖昧性？）

比較できる論点

今後の展望：利用例の提示

　　オープンアクセスの維持

計量テキスト分析およびKH Coderの利用状況と展望　②

『性・メディア・風俗――週刊誌『アサヒ芸能』からみる風俗としての性』

雑誌の目次タイトルから性風俗の分析

コーディングルールの仮説・検証

目録からの研究可能性

計量テキスト分析およびKH Coderの利用状況と展望　③

「ポスト保革イデオロギー時代における日本政治の対立軸――『保革溶解』の逸脱事例としての沖縄を中心に」

基本は歴史学的分析で、補完的に計量文献学を利用

仮説からコーディングルールを整備

議事録（全文テキスト化済）から分析

発話者別の分析で階層化に成功している