

#NDL全文使ってみた

近現代日本文学
研究の場合

第24回図書館総合展
2022.11.1 @ZOOM Webinar

フォーラム
#NDL全文使ってみた～「次世代デジタル
ライブラリー」&「NDL Ngram Viewer」

日比 嘉高 HIBI Yoshitaka
名古屋大学 hibi@nagoya-u.jp



1

はじめに

自己紹介とこの報告の概要

自己紹介と この報告について



日比 嘉高 HIBI Yoshitaka

名古屋大学人文学研究科教授
日本の近現代文学・文化論

■『プライヴァシーの誕生 モデル小説のトラブル史』新曜社2020

■「科研費採択課題を対象とした研究課題の計量テキスト分析——日本文学の場合」『社会文学』52、2020、pp.89-100

■「《巻頭言》日本近現代文学研究者はコンピュータを使って何をしたいのか。したくないのか。」『人文情報学月報』121、2021

ほか

この報告で話すこと

1. はじめに
2. 全文デジタルデータを用いた日本文学研究、とりわけ近現代文学研究の動向
3. 報告者・日比の、全文デジタルデータを用いた研究の実践報告
4. まとめと課題、提言

2

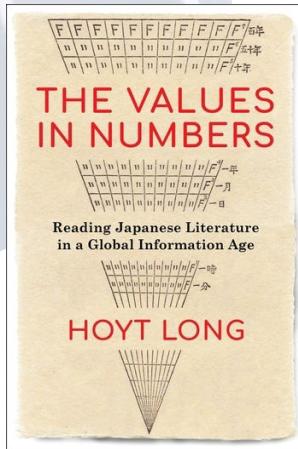
全文デジタルデータを用いた 日本文学研究の動向 とりわけ近現代文学研究について

古典文学研究の 場合



- ① 本文について著作権上の問題がないため、比較的早くから研究、活用が開始
- ② 国文学研究資料館の事例
 - i. 本文テキストデータの検索 古典選集本文DB、日本古典文学大系本文DB、嘶本大系本文DBを対象
 - ii. 画像による本文データ 日本古典籍データセット、江戸料理レシピデータセット
- ③ 研究の事例
 - i. 近藤みゆきによるNgramを用いた古典和歌の研究。詩句の新たな定型性の発見やジェンダー的偏差の問題の指摘。
『古代後期和歌文学の研究』(風間書房、2005)
『王朝和歌研究の方法』(笠間書院、2015)

近代文学研究の場合



- ① 近年、全文データベースが急速に充実
 - i. 国会図書館デジタルコレクション
+ 次世代デジタルライブラリー
 - ii. 青空文庫
 - iii. Maruzen eBook Library
 - iv. Google Books
 - v. 慶應義塾大学図書館蔵書のHathiTrustへの登載
- ② 上記(とくに i、ii)は、研究・教育で使われているが、方法論的な革新はさほど起こっていない
- ③ 革新的な研究の事例
 - i. Hoyt Longの近著。語りのパターンを析出するほか、計量分析、統計分析を用いて、作品のジャンル判定や、キーワードの使用頻度を解析。

The Values in Numbers, Columbia University Press, 2021.



3

全文デジタルデータを用いた研究 の実践報告

松尾芭蕉全発句の受容の状況を探る



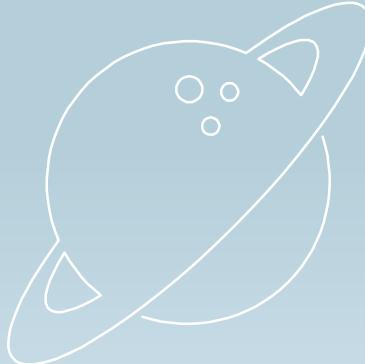
松尾芭蕉の俳句のうち、
戦前に最も言及された句は、どれ？

約336,000[☆]件

NDL 次世代デジタルライブラリー

約1,000句超

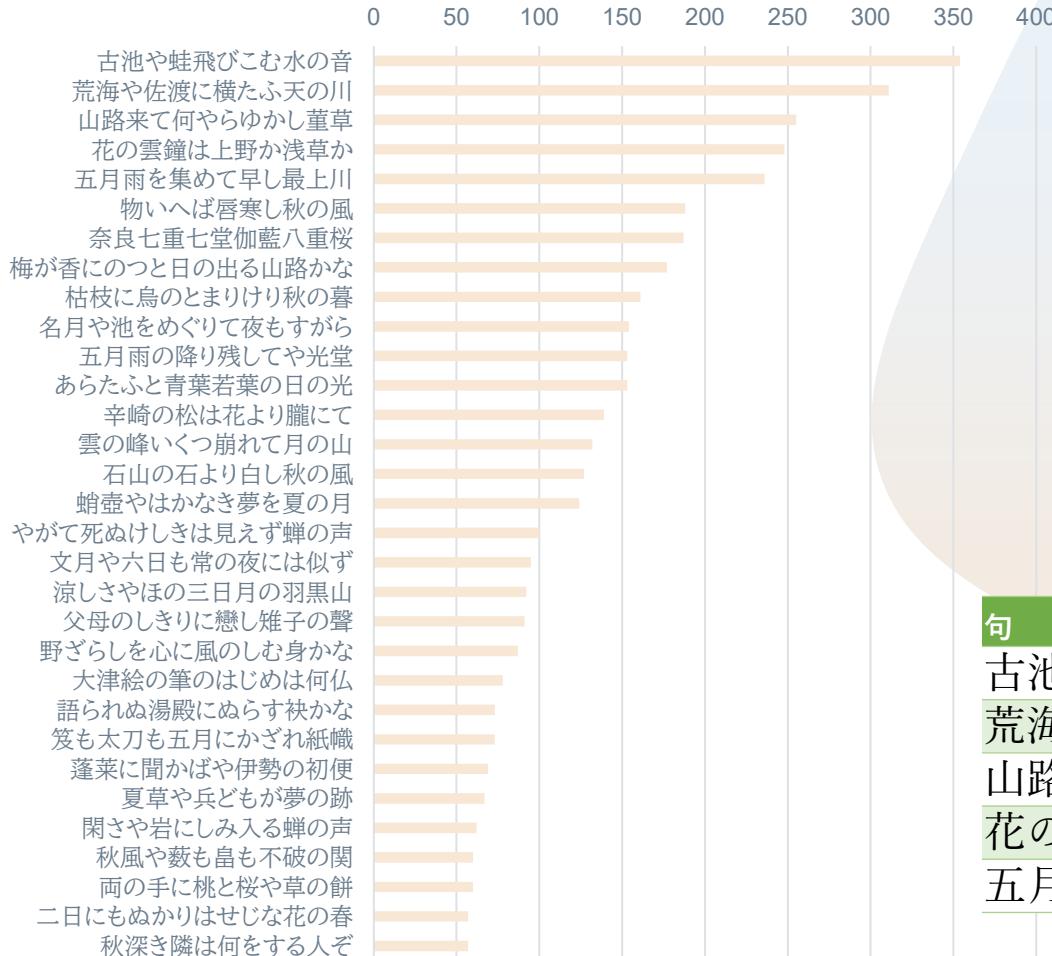
松尾芭蕉の全発句



文学研究的に言い直すと――

第二次世界大戦以前の文献において
引用された芭蕉の発句を、
計量的に分析し、
近代における芭蕉受容の様相を
再検討する。

引用頻度順30位



引用頻度順 30位と上位5句

- NDL次世代デジタルライブラリーの全文検索を、各句で検索
- 表記揺れを考慮するため、全句データは3つのバージョンを用意
- 検索期間は1868-1945

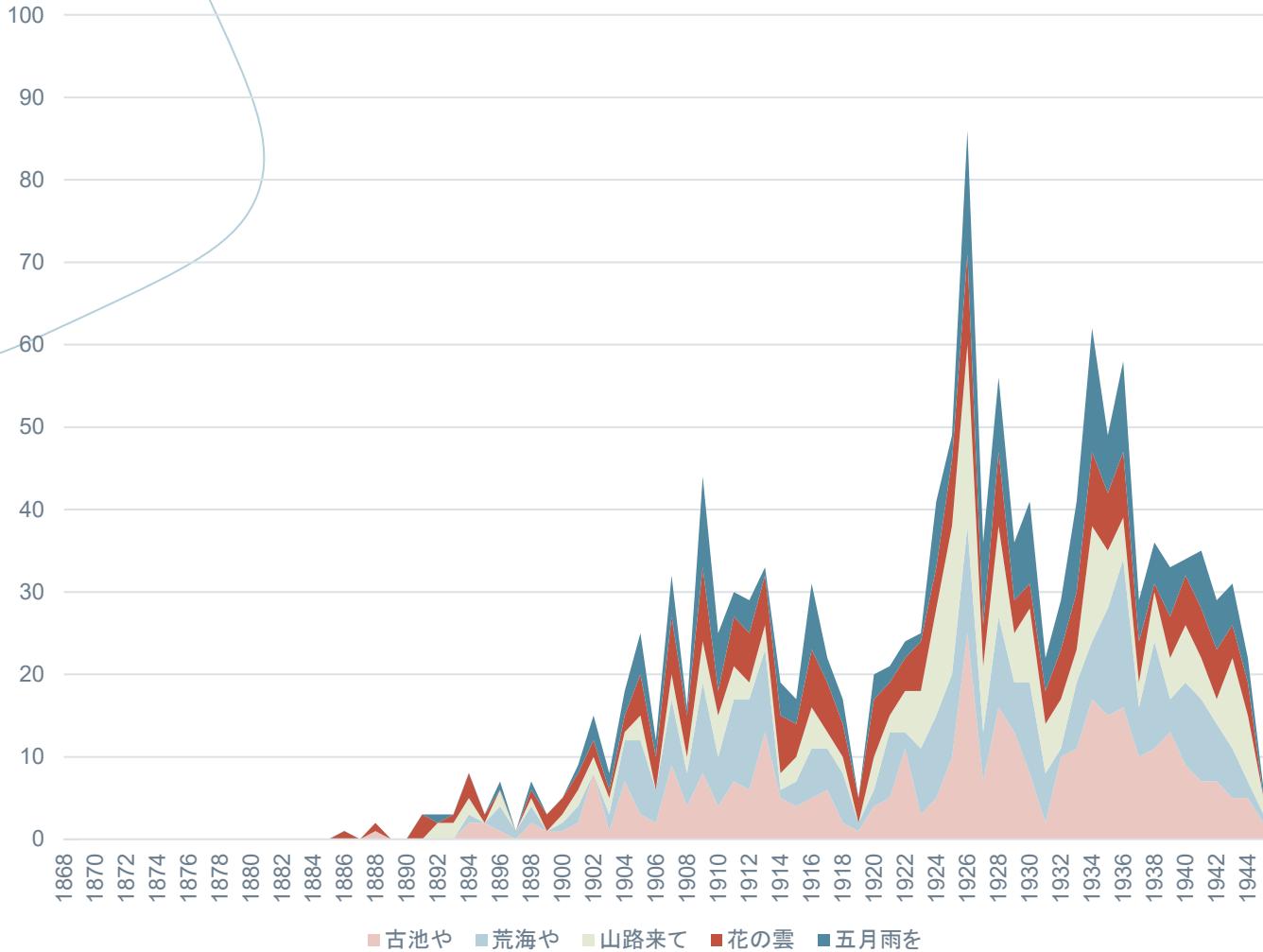
引用頻度上位5句

句	合計	頻度順位
古池や蛙飛びこむ水の音	354	1
荒海や佐渡に横たふ天の川	311	2
山路来て何やらゆかし董草	255	3
花の雲鐘は上野か浅草か	248	4
五月雨を集めて早し最上川	236	5

引用頻度上位5句の出現年グラフ



引用頻度 上位5句の 積み上げ面 グラフ





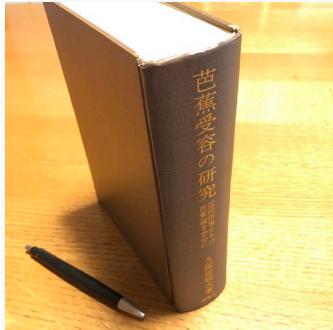
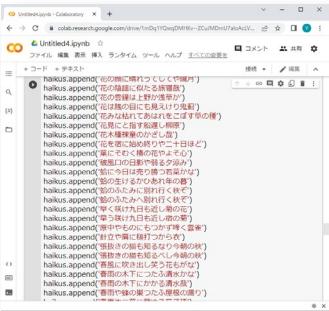
ヒトの行う受容研究

久保田晴次『芭蕉受容の研究 近代作家たちの芭蕉論を中心に』(桜楓社、1974)

- A5版、全874頁。
- 「巻末の一章「芭蕉に関わる近代作家たちの文献について」は、[...]合計二百八十七編に及ぶ。これによっても、この一書の背後にある著者の博捜の努力が窺われよう。」(*)
- **久保田の受容研究の目標:** 「いつ誰がどんな芭蕉論を発表したというには芭蕉論の解説で、それを年次を追って整理してもそれは芭蕉論の文献解説集にすぎません。本書のもくろみはその芭蕉論のあまたがどのような意味をなっているか、その芭蕉論から芭蕉論への流れの変化にどのような意味がみいだせるかをうかがおうとするにありました。」(ii～iii頁)
- 久保田は、前著『脱出の文学』(桜楓社1968)でも芭蕉の受容研究に取り組む。前著で引き出した主題は「脱出」、今回は「実存」。

* 山下一海「久保田晴次著『芭蕉受容の研究』」『連歌俳諧研究』50号、1976、pp.74-76.

ヒトとマシン、
それぞれの強み、
その協働。



■ 圧倒的な作業量と速度

- 「既存の価値づけ」を無視し、「量」に置き換える。「文脈」を無視する

■ 出力結果の解釈は、人の手に委ねられる（解釈がないと、「で？」となる）

■ 作業量と速度の限界

■ 文学者の重要度や文学史的価値から、作業をあらかじめ絞り込む。予断の存在

■「文脈」の理解ができる。作業によってさらにそれを深め、更新する。

4

まとめと課題と提言 研究で全文デジタルデータを使うために

まとめと課題と提言

- a. マシン・リーディングの驚異的な作業量と速度
これまでにない研究が生まれる可能性
- b. ただし、現状では全文データからからもたらされるのは、あくまで発想のヒント。それをもとに、ヒトがどう伸ばす・深めるか
 - i) 「掛け合わせ」—— **全文DB × 何か** —— は、一つの鍵
「芭蕉の全句」「〈私小説性〉データベース」…など
 - ii) 研究で「全ジャンル」「全時代」が必要なことは、実は多くない。
むしろ、特定のジャンルや時期、媒体などで適切に絞り込み、
「意味を持った資料体」にすることが大事
 - iii) 研究者も参与した各種全文データセットの構築・増加が課題
- C. マークアップによるテキストの構造化(**TEI**)。分析可能性の拡大
- d. 図書館側が用意する分類以外に、利用者による個別資料のタグ付けとその共有の仕組みを用意しては？ 「絞り込み」「掛け合わせ」のツールとして。 タグの例)「女性作家」「富山県出身」「変身」…