

# 社会・人間プロジェクト s3 発表会

八代集に見られる名づけの選好の変化

Xudong Chen  
22D50456

xchen@shs.ens.titech.ac.jp  
山元研究室

2022/08/07

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

- **目的**：八代集の歌ことばにおける同一の概念を表す語（名づけ）の選好の時代的属性の分析を例に、そのための方法論的整備
- **方法論**：
  1. 変異理論 (variationist linguistics)
  2. 語彙計量論 (lectometric approach)
- **検証**：
  1. 量的検証：分類語彙表の概念分類と照合、マッチング個数の数え上げ
  2. 質的検証：名づけのグループ候補の関係性の整理
  3. 事例：いままで注目されなかった要素の有無
- **結果**：
  1. 工学的には精度が悪い
  2. 抽出する関係性は、単なる類義関係ではない
  3. 人間・分類語彙表で特定できない項目の特定に有効

→ 分類語彙表、WordNet のような知識ベースなアノテーションセットとの併用

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

- ✓ 文献調査
- ✓ アルゴリズム実装
- ✓ 結果の算出とまとめ
- ✗ 細かい考察
- ✗ 論文執筆

反省

- 試行回数を増やすコードの書き方
- 仕事の一貫性；放置したことで理解しなおすこともある
- 研究ノートを取る：データ分析より常に短い文書を書く

## 背景：研究目的の位置づけ

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

### 語本位

- 使用頻度と残存率の変化 (frequency change) ← 井上・半沢 (2021)
- 連語の変化 (collocational/phrasological change) ← 石井 (2019)
- 指示内容の変化 (semasiological change)
  - 顕在的な意味の変化 (denotational change) ← 田野村 (2016)
  - 暗示的な意味の変化 (connotational change) ← 山元 (2007)
- 機能の変化 (functional change; grammaticalization) ← 新国他 (2019)

### 概念本位

- 名づけの変化 (onomasiological change) ← 本発表のねらい

## 背景:「名づけ」の選好という視座

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

- 同じことに対する異なる言い方
- 認知パターン (highlighting / backgrounding) を反映 (De Pascale 2019)
- 語彙の使用域・社会属性 → 使用域・スピーチグループの文化と認知の差
- 研究体系 が整えつつあるが、古代語・日本語の語彙研究が盲点

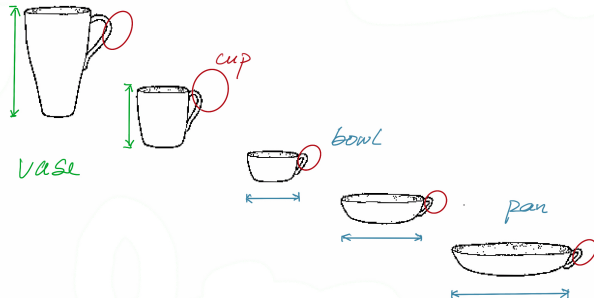


図 1: 同一の容器の異なる呼び方 (cf. Labov 1973)

背景：関連研究

Xudong Chen  
22D50456

概要  
達成度

背景

位置づけ  
「名づけ」  
関連研究  
問題意識  
研究方法  
変異理論と語彙計量論  
ワークフロー  
検証  
ケース・スタディー

結果

メタコードとの照合  
質的分析  
ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

- 変異理論に関連づけた研究は特定概念を対象とする
  - 非特定概念を対象とする研究は人手で選定
- 選定の漏れ・バイアスが生じる

文献	名づけ	概念	言語単位	変種基準	変異理論の志向
山田 (2021)	a. **される b. **なさる c. ご**になる d. ご**される e. ご**なさる	敬意	構文	時代 (江戸-昭和)	あり
松田 (2019)	a. 足りない b. 足らない	「足る」の否定	統語	時代 (話者の生年) 地域	あり
久屋 (2016)	外来語 / 既存語 e.g. ニーズ / 必要性	非特定	語	時代	あり
田中・山元 (2014)	固い表現 / 柔らかい表現 e.g. 参らす / 奉る	非特定	語	文体 / レジスター	なし

表 1: 方法論的に関連する日本語の研究

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

- 変異理論志向の語彙計量論フレームワークとワークフローの応用
- 古語資料における有効性 → 和歌資料のケースで確認
  1. 類義語群の特定の質と量
  2. 知識ベースのアノテーションとのマッチング
  3. いままで注目されなかった事例の有無

# 方法：変異理論と語彙計量論 (cf. De Pascale 2019)

概要  
達成度  
背景  
位置づけ  
「名づけ」  
関連研究  
問題意識

研究方法  
変異理論と語彙計量論  
ワークフロー  
検証  
ケース・スタディー  
結果  
メタコードとの照合  
質的分析  
ケース・スタディー

結論  
参考文献

付録  
ハイパーパラメータの影響

用語：	変種	変数	変異形
解釈：	言語の使用域・使用者の分類	概念	同一概念の類義語 (名づけ)
例：	地理的：イギリス英語・アメリカ英語 文体的：和漢混淆文・和文	「地下鉄」 「女の子」	underground; subway をんなご; めのわらわ

表 2: 変異理論の枠組みにおける語彙変異の捉え方 (cf. De Pascale 2019)

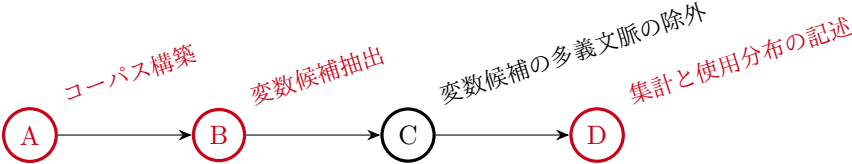


図 2: 語彙計量論ワークフロー (cf. De Pascale 2019, De Pascale and Zhang 2021)



## 方法：ワークフローのアダプテーション

## A. 八代集語彙データベース (Yamamoto and Hodošček 2021)

01:000001:0015 A00 BG-02-3120-01-0100 02 いは 言ふ いふ 言は いは

01:000001:0016 A00 BG-03-3012-03-2600 02 ん む む む む

01:000001:0016 A10 BG-09-0010-02-0100 02 ん む む む む

anthology ID	-
poem ID	-
token sequence ID	-
token type	-
general ID	-
POS ID	-
semantic group ID	-
semantic field ID	-
specific ID	-
POS number	-
surface form	-
lemma kanji	-
lemma kana	-
conjugation kanji	-
conjugation kana	-

## B. 変数候補抽出 : Cluster of Committees (De Pascale 2019: 174–177)

- 特徴量用の文脈語は Aida et al. (2021) で特定
- hyperparameter の調整

D. 使用分布傾向の記述 : external uniformity value ( $U$ )<sup>1</sup> (Ruetten et al. 2014)<sup>1</sup>低いほど変種間の選択傾向が異なる

- 概要
- 達成度
- 背景
  - 位置づけ
  - 「名づけ」
  - 関連研究
  - 問題意識
- 研究方法
  - 変異理論と語彙計量論
  - ワークフロー
- 検証
  - ケース・スタディー
- 結果
  - メタコードとの照合
  - 質的分析
  - ケース・スタディー
- 結論
- 参考文献
- 付録
  - ハイパーパラメータの影響

## 方法：検証

量的 抽出した候補同士のマッチングをメタコードと照合、マッチした個数をカウント  
質的 抽出した候補同士の意味関係の整理

not match		
BG-01-5530-10-0305	紅葉葉	
BG-02-5020-01-1201	紅葉づ	
POS match		
BG-01-1632-01-0100	月	
BG-01-5010-07-0700	月影	
group match		
BG-01-5260-03-0600	沖	
BG-01-5260-08-0100	浦	
field match		
BG-01-5240-16-1001	浅茅	
BG-01-5240-16-1101	浅茅生	

図 3: メタコードによる  
マッチングのレベ  
ル：漏れがある

- 抽出候補のなか適切と判断した語彙変数を対象に、分布傾向を  $U^2$  で記述

	声 (こゑ)	音 (ね)	
後拾遺集以前 (−1086)	33.3%	66.7%	100%
金葉集以降 (ca.1125−)	80%	20%	100%

表 3: 2 変種 × 2 変異形の場合の external uniformity value ( $U$ ) の計算例 (数値は仮定): 各セルにある数値は、語の相対頻度である;  $U$  は各々の変種の最小相対頻度の合計である

$$U = 33.3\% + 20\% = 53.3\%$$

- 時代性という社会的属性を、変異形がもっているかを考察
- 既存文献での言及を確認

<sup>2</sup>低いほど変種間の選択傾向が異なる

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

	POS match	group match	field match
#	522	132	92
%	54.21	13.71	9.55

表 4: 抽出候補の下位変異形のメタコードのマッチング

考察

- 抽出結果と知識ベースなメタコードアノテーションとの一致が少ない

→ 知識ベースなデータとともに応用するか semi-automatic な手法として応用する

# 結果：質的分析

	WordNet 準拠の関係性	解釈	例	候補数
a.	hypernym-hyponym holonym-meronym synonymy	上位-下位 部分-全体 類義	花-桜 水-川 千代-千世	35
b.	antononym	対義	夢-現	3
c.	entail cause	含意 因果	絞る-乾く	1
d.	similar also	類似 相関	更級-姨捨山	23
e.	participle, pertainym derivation	動詞化、形容詞化 派生	紅葉葉-紅葉づ	1

表 5: 抽出語彙変数候補の下位変異形の意味関係のカテゴリー：平均類似度上位 150 から

## 考察

- カテゴリ a のうち、同義異音 2、同義異表記 10 が含まれる → 機能はしている
- WordNet の意味関係のカテゴリーにマッピングが可能

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

Lexical variable

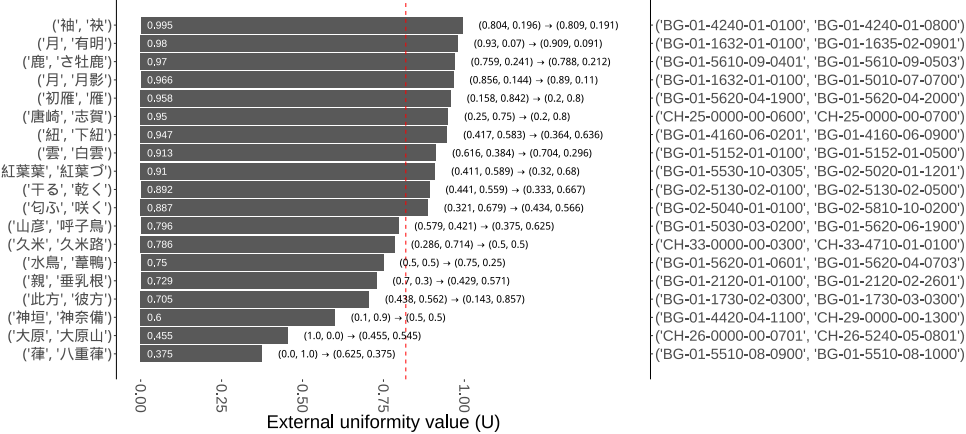


図 4: 抽出語彙変数 19 個それぞれの U : 値の低い語彙変数は、その下位変異形 (名づけ) の選好が 2 時代によって大きく異なり、時代性がある<sup>a</sup>

<sup>a</sup>上野 (1976) に基づき八代集を後拾遺集を転換点に 2 時期に分けた。

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

小町谷: 八重葎 後拾遺集 葎の門、葎の宿

Xudong Chen  
22D50456

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

1. 工学的には精度が悪い
  2. 抽出する関係性は、単なる類義関係ではない
  3. 人間・分類語彙表で特定できない項目の特定に有効
- 分類語彙表、WordNet のような知識ベースなアノテーションセットとの併用
- ほかの古語資料への適用



Xudong Chen  
22D50456

概要  
達成度  
背景  
位置づけ  
「名づけ」  
関連研究  
問題意識  
研究方法  
変異理論と語彙計量論  
ワークフロー  
検証  
ケース・スタディー

結果  
メタコードとの照合  
質的分析  
ケース・スタディー

結論

参考文献

付録  
ハイパーパラメータの影響

Aida, Taichi, Mamoru Komachi, Toshinobu Ogiso, Hiroya Takamura, and Daichi Mochihashi (2021) “A Comprehensive Analysis of PMI-based Models for Measuring Semantic Differences,” in *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, pp. 21–31, Shanghai, China: Association for Computational Linguistics.

De Pascale, Stefano (2019) “Token-Based Vector Space Models as Semantic Control in Lexical Sociolectometry,” Ph.D. dissertation, KU Leuven, Leuven.

—— and Weiwei Zhang (2021) “Scoring with Token-based Models,” in Kristiansen, Gitte, Karlien Franco, Stefano De Pascale, Laura Rosseel, and Weiwei Zhang eds. *Cognitive Sociolinguistics Revisited*, pp. 186–200: De Gruyter, DOI: [10.1515/9783110733945-016](https://doi.org/10.1515/9783110733945-016).

井上史雄・半沢康 (2021) “方言衰退の語彙論的過程,” 日本語の研究, Vol. 17, No. 1, pp. 1–18, DOI: [10.20666/nihongonokenkyu.17.1\\_1](https://doi.org/10.20666/nihongonokenkyu.17.1_1).

石井正彦 (2019) “コロケーションの成立と変化に関する事例的検討: 新聞「デフレ+動詞」句の通時的頻度調査から,” 現代日本語研究, Vol. 11, pp. 107–126, DOI: [10.18910/73343](https://doi.org/10.18910/73343).

久屋愛実 (2016) “見かけ上の時間を利用した外来語使用意識の通時変化予測,” 日本語の研究, Vol. 12, No. 4, pp. 69–85, DOI: [10.20666/nihongonokenkyu.12.4\\_69](https://doi.org/10.20666/nihongonokenkyu.12.4_69).

Labov, William (1973) “The boundaries of words and their meanings,” in Bailey, Charles-James and Roger W. Shuy eds. *New Ways of Analyzing Variation in English*, pp. 340–371, Washington, DC: Georgetown University Press.

松田謙次郎 (2019) “岡崎敬語調査に見る「足りない」～「足らない」の変異と変化,” 計量国語学, Vol. 32, No. 2, pp. 66–81, DOI: [10.24701/mathling.32.2\\_66](https://doi.org/10.24701/mathling.32.2_66).

新国佳祐・和田裕一・小宮智也・小川芳樹 (2019) “形式名詞「はず」の文法化と属格主語の容認性における世代間差,” in 小川芳樹 (編) コーパスからわかる言語変化・変異と言語理論 2, pp. 261–288, 東京: 開拓社.

Ruette, Tom, Dirk Geeraerts, Yves Peirsman, and Dirk Speelman (2014) “Semantic Weighting Mechanisms in Scalable Lexical Sociolectometry,” in *Semantic Weighting Mechanisms in Scalable Lexical Sociolectometry*, pp. 205–230: De Gruyter, DOI: [10.1515/9783110317558.205](https://doi.org/10.1515/9783110317558.205).

田中牧郎・山元啓史 (2014) “『今昔物語集』と『宇治拾遺物語』同文説話における語の対応 -語の文体的価値の記述-,” 日本語の研究, Vol. 10, No. 1, pp. 16–31.

田野村忠温 (2016) “「科学」の語史: 漸次的・段階的変貌と普及の様相,” 大阪大学大学院文学研究科紀要, Vol. 56, pp. 123–181, DOI: [10.18910/56924](https://doi.org/10.18910/56924).

山田彬亮 (2021) “江戸期から昭和期にかけての尊敬語のバリエーション: 多項混合効果ロジスティック回帰によるベイズ統計学的アプローチ,” 言語文化共同研究プロジェクト, No. 2020, pp. 5–25.

山元啓史 (2007) “モデリングによる歌ことばの変遷と分析―八代集・歌ことばソーラスの開発―,” じんもんこん 2007 論文集, Vol. 2017, pp. 163–170.

Yamamoto, Hilofumi and Bor Hodošček (2021) “Hachidaishu vocabulary dataset,” DOI: [10.5281/zenodo.4744170](https://doi.org/10.5281/zenodo.4744170).

上野理 (1976) 後拾遺集前後: 笠間書院, URL: <https://ci.nii.ac.jp/ncid/BN05497557>.

小町谷照彦 “むぐら.”

# 結果：ハイパーパラメーター操作による語彙変数抽出結果の変化

概要

達成度

背景

位置づけ

「名づけ」

関連研究

問題意識

研究方法

変異理論と語彙計量論

ワークフロー

検証

ケース・スタディー

結果

メタコードとの照合

質的分析

ケース・スタディー

結論

参考文献

付録

ハイパーパラメータの影響

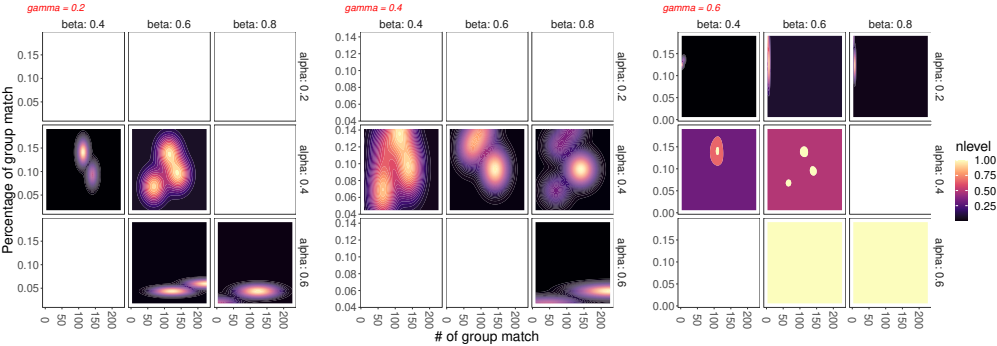


図 5: ハイパーパラメーター操作による抽出変数におけるグループマッチングの個数と比例の分布

→  $\alpha = 0.4$ ;  $\beta = 0.8$ ;  $\gamma = 0.4$  を用いる。