

文体の数量分析に基づく
川端康成の影武者問題の解明

孫 昊

同志社大学

文化情報学研究科

2018/06/09

目次

1. 研究背景

- 川端康成の影武者問題
- 本研究の対象

2. 研究方法

- コーパス
- 文体特徴量
- 識別方法

3. 結果と課題

研究背景 1/3

■ 川端康成

- ノーベル文学賞 (1968年)
- 日本を代表する文豪
- 代表作『雪国』, 『山の音』など
- 孤児根性→精神不安定
- ガス自殺 (1972年)

■ 川端康成の影武者問題

- 川端康成は遅筆だったため, 何人もの代作者をかかえていた (矢崎, 2003)

研究背景 2/3

■ 川端康成の影武者たち

- 中里恒子

『乙女の港』, 『花日記』, 『コスモスの友』

- 三島由紀夫

『眠れる美女』, 『山の音』, 『古都』

- 北條誠

『古都』

- 澤野久雄

『古都』

- 内田憲太郎

『空の片仮名』

研究背景 2/3

■ 川端康成の影武者たち

- ・ 中里恒子

『乙女の港』, 『花日記』, 『コスモスの友』

- ・ 三島由紀夫

『眠れる美女』, 『山の音』, 『古都』

- ・ 北條誠

『古都』

- ・ 澤野久雄

『古都』

- ・ 内田憲太郎

『空の片仮名』



川端康成

研究背景 2/3

■ 川端康成の影武者たち

- 中里恒子

『乙女の港』, 『花日記』, 『コスモスの友』

- 三島由紀夫

『眠れる美女』, 『山の音』, 『古都』

- 北條誠

『古都』

- 澤野久雄

『古都』

- **内田憲太郎**

『空の片仮名』

川端康成

研究背景 3/3

■ 空の片仮名

- 川端康成全集第21巻収録
- 龍胆寺雄は内田憲太郎の代筆と指摘
- 川端康成は否認，龍胆寺雄は追放された
- 川端康成の代筆問題のうち，文壇では誰も触れていないのは『空の片仮名』である (小谷野, 2013)
- 先行研究皆無

研究背景 3/3

■ 空の片仮名

- 川端康成全集第21巻収録
- 龍胆寺雄は内田憲太郎の代筆と指摘
- 川端康成は否認，龍胆寺雄は追放された
- 川端康成の代筆問題のうち，文壇では誰も触れていないのは『空の片仮名』である (小谷野, 2013)
- 先行研究皆無

文章の数量解析に基づいて
『空の片仮名』の影武者問題を解明

目次

1. 研究背景

- 川端康成の影武者問題
- 本研究の対象

2. 研究方法

- 著者識別モデル
- コーパス
- 文体特徴量
- 識別方法

3. 結果と考察

研究方法 1/6

■ 著者識別モデル

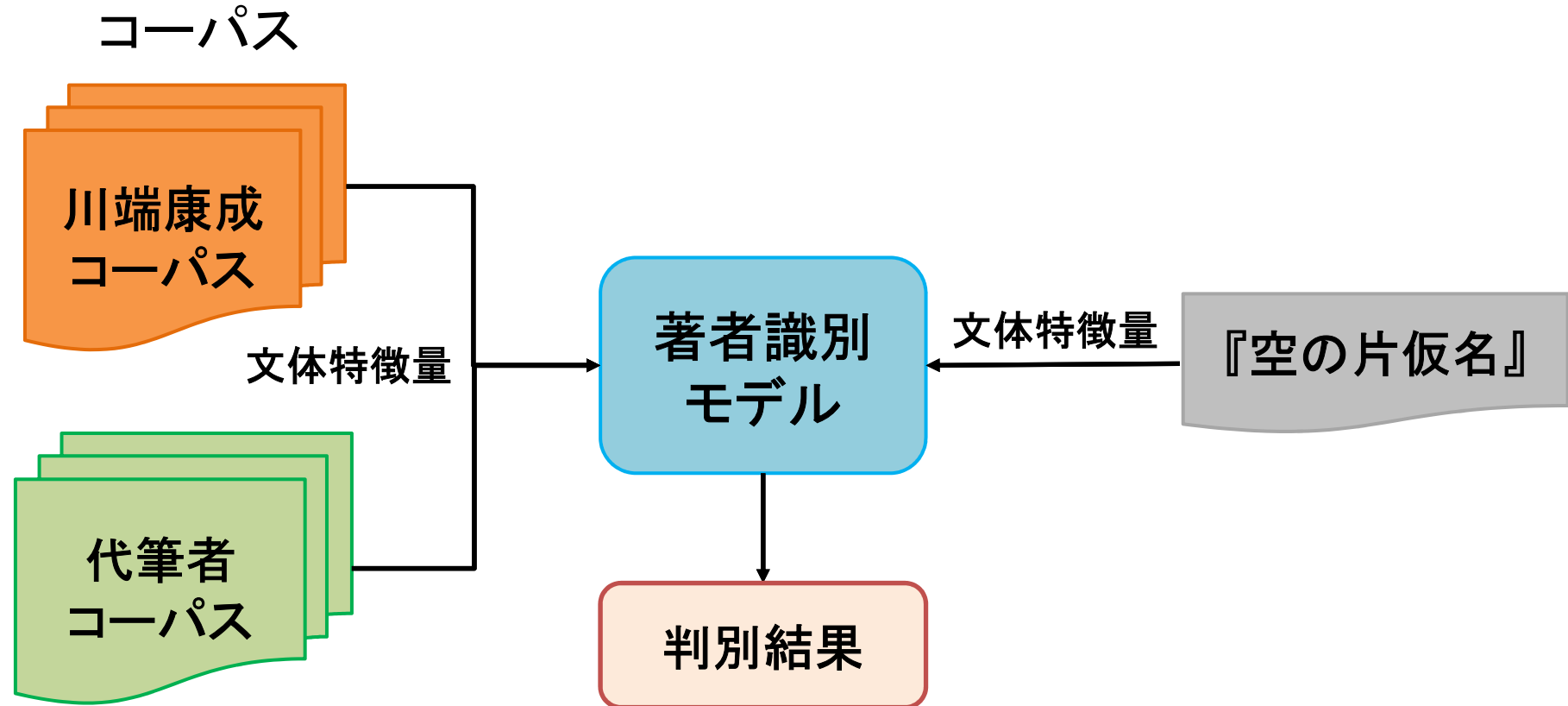


図1. 著者識別 (authorship attribution) モデル

研究方法 1/6

■ 著者識別モデル

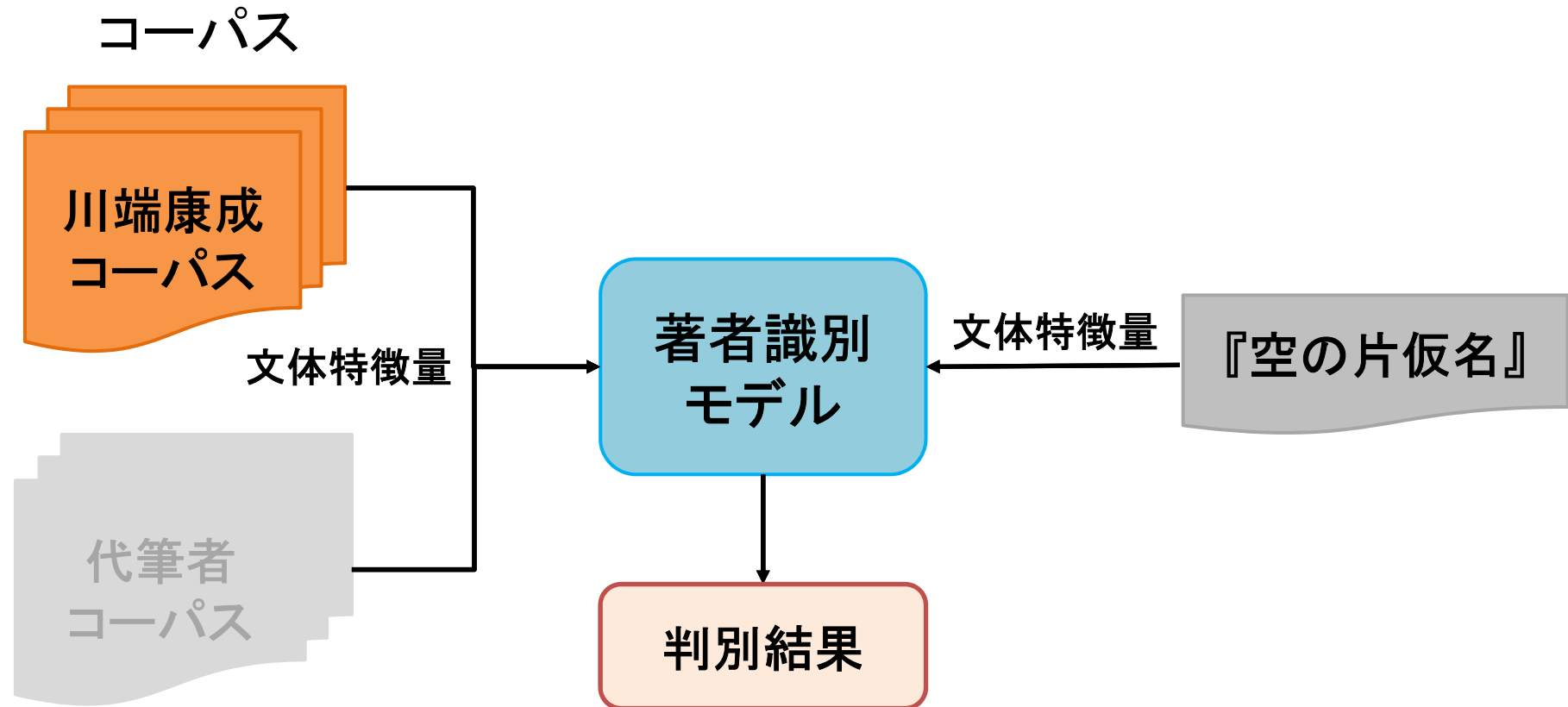


図1. 著者識別 (authorship attribution) モデル

研究方法 2/6

■ 対照作家の導入

- ・ 保険金殺人手紙分析 (金, 2009)

■ 解決策

- ・ 対照作家の導入

■ 対照作家の選定 (寺田, 1949)

- ・ 泉鏡花
- ・ 徳田秋聲
- ・ 横光利一

研究方法 3/6

- 川端康成コーパス (20)
たまゆら、みづうみ、雨の日、横町など
- 泉鏡花コーパス (20)
瓜の涙、縁結び、絵本の春、外科室など
- 徳田秋声コーパス (20)
チビの魂、のらもの、仮想人物、花が咲くなど
- 横光利一コーパス (20)
火、花園の思想、街の底、機械など

研究方法 4/6

■ 読点の打ち方

読点とその前の仮名:「は、」

■ 形態素タグのbi-gram

形態素解析器 (MeCab)でテキストを形態素に分割隣接している品詞タグの

例: 学会で話す.

表1. 内容語を除いたタグ付き形態素の例

形態素	タグ	タグのbi-gram
学会	名詞	名詞_助詞 助詞_動詞 動詞_記号
で	助詞	
話す	動詞	
.	記号	

研究方法 5/6

■ 文節パターン

文を実際の言語として不自然でない程度に区切った最小単位
日本語文を文節にわけ (CaboCha)
文節内の助詞と記号の原型とほかの品詞情報

例: 学会で話す.

表2. 文節パターンの例

文節情報		品詞	文節パターン
第1文節	学会	名詞	名詞_で
	で	助詞	
第2文節	話す	動詞	動詞_.
	.	記号	

研究方法 6/6

■ 識別モデル

- 教師なしの手法
 - 対応分析 (CA)
 - 階層的クラスタ分析 (HCA)
- 教師ありの手法
 - エイダブースト (AdaBoost)
 - 高次元判別分析 (HDDA)
 - ロジスティックモデルツリー (LMT)
 - ランダムフォレスト (RF)
 - サポートベクターマシン (SVM)

目次

1. 研究背景

- ・ 川端康成の影武者問題
- ・ 本研究の対象

2. 研究方法

- ・ コーパス
- ・ 文体特徴量
- ・ 識別方法

3. 結果と考察

結果分析 3/11

■ 文節パターンへの対応分析

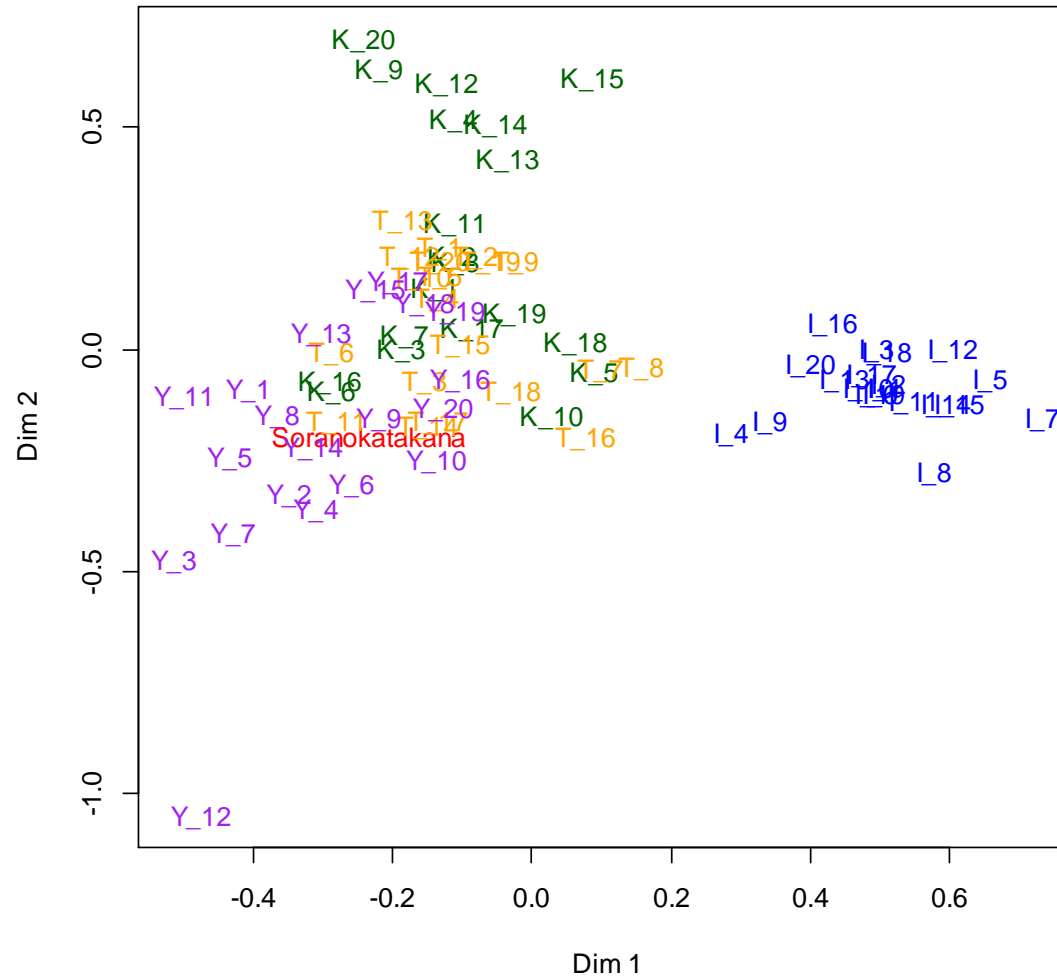


図4. 文節パターンへの対応分析個体スコア第1, 2成分散布図

結果分析 4/11

■ 読点の打ち方のクラスター分析

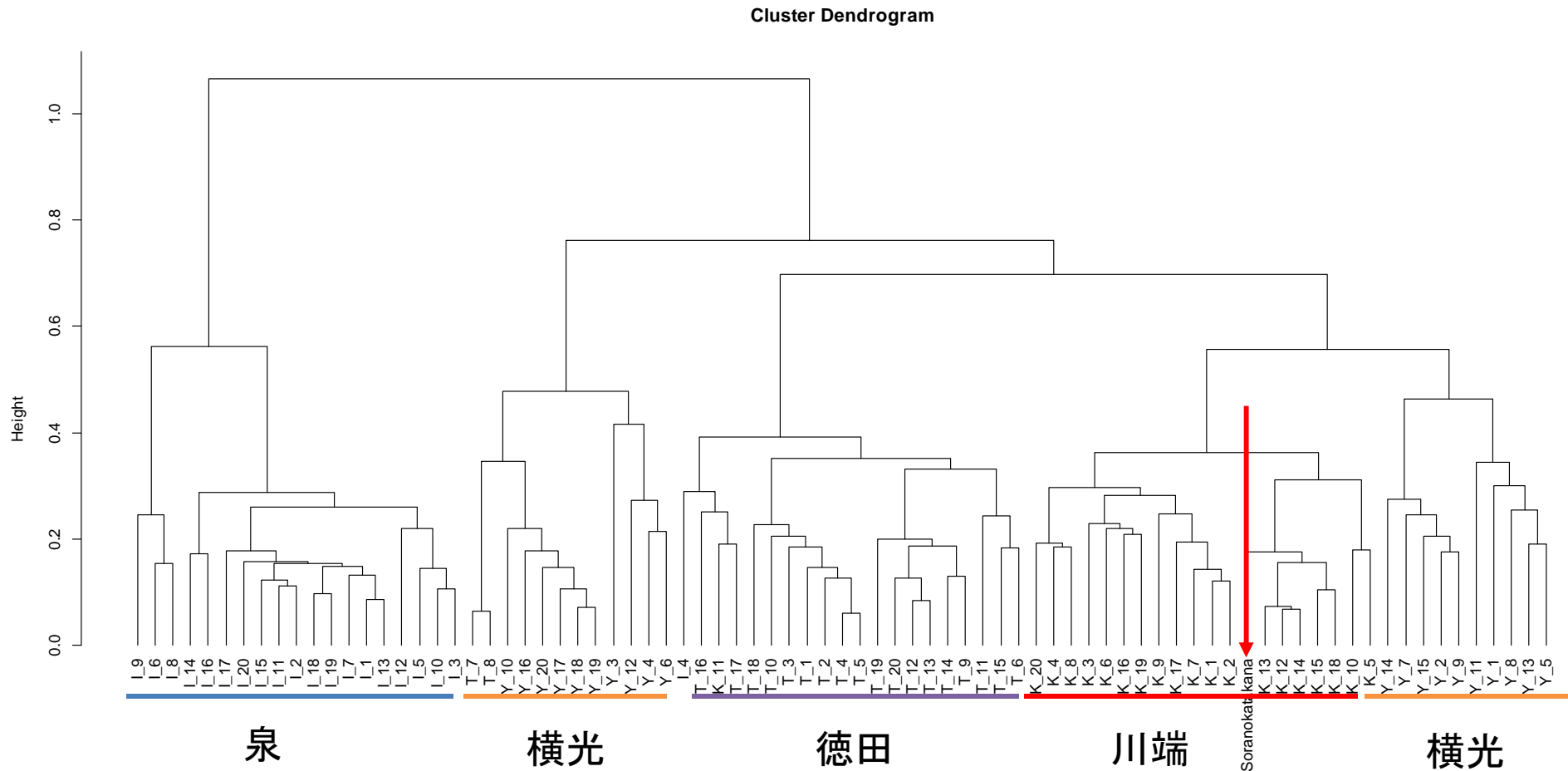


図5. 読点の打ち方の階層的クラスター分析 (ward法, SKLD距離)樹形図

結果分析 5/11

■ タグのbigramのクラスター分析

Cluster Dendrogram

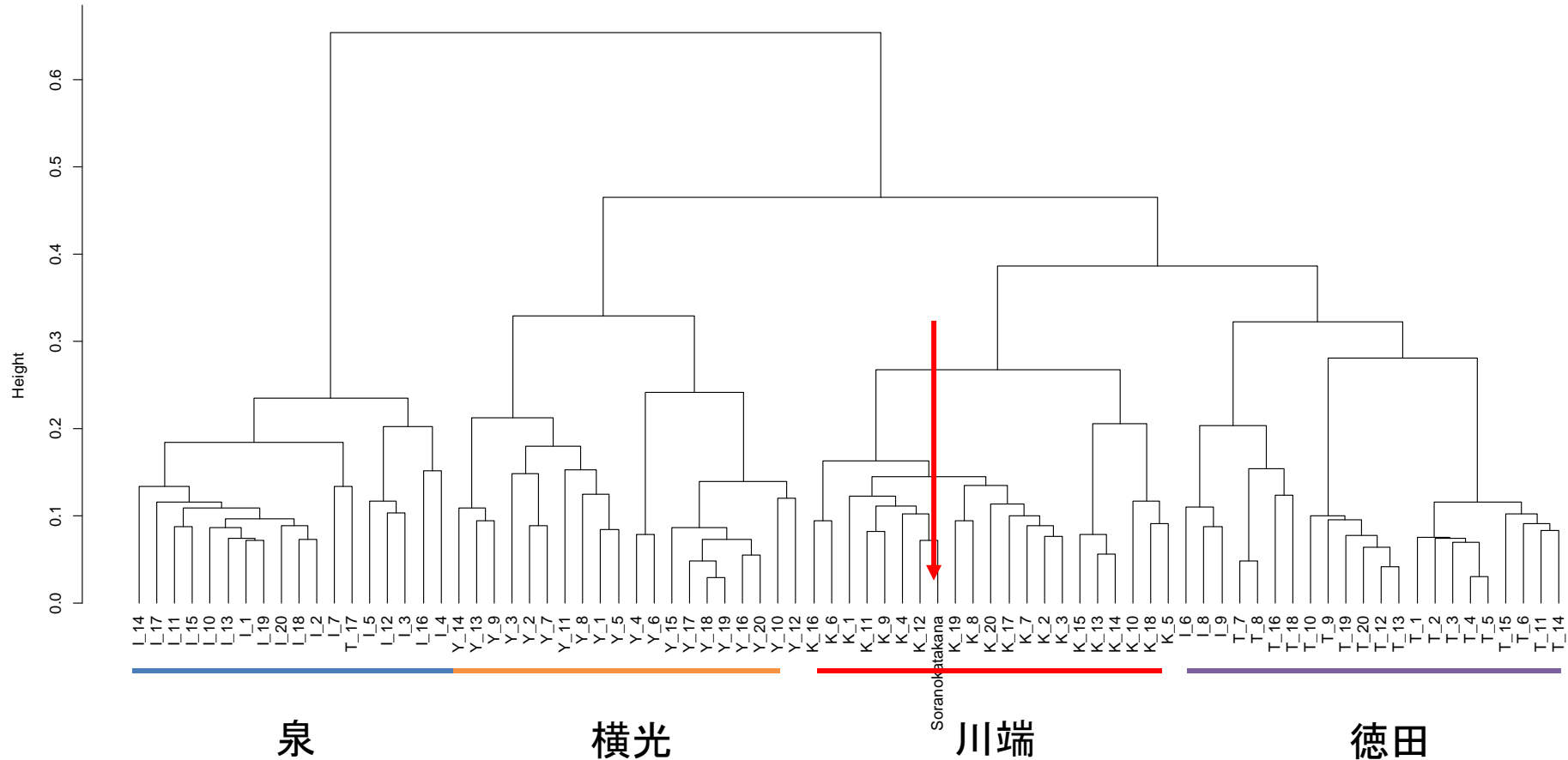


図6. タグのbigramの階層的クラスター分析(ward法, SKLD距離)樹形図

結果分析 6/11

■ 文節パターンのクラスター分析

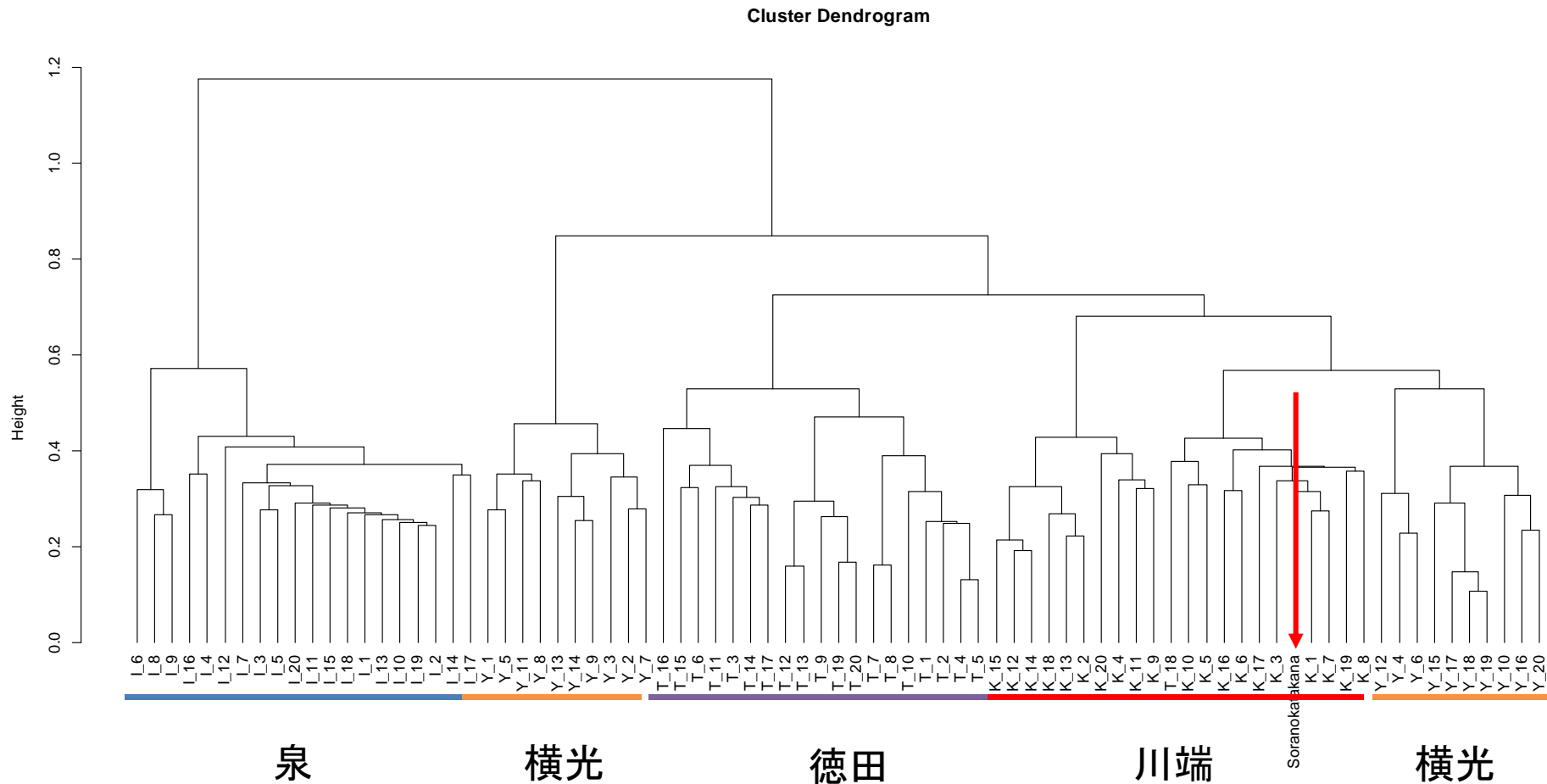


図7. 文節パターンの階層的クラスター分析(ward法, SKLD距離)樹形図

結果分析 10/11

■ 教師なし手法の結果まとめ

表3. 『空の片仮名』の教師なし手法による分類結果

空の片仮名	CA	HCA
読点の打ち方	川端	川端
タグbi-gram	川端・徳田・横光	川端
文節パターン	徳田・横光	川端

結果分析 11/11

■ 教師あり手法による分類結果

表4. 『空の片仮名』の教師あり手法による分類結果

空の片仮名	AdaBoost	HDDA	LMT	RF	SVM
読点の打ち方	川端	横光	川端	川端	川端
タグbi-gram	徳田	川端	川端	川端	川端
文節パターン	川端	横光	川端	川端	徳田

結論

■ 4群判別の結果

- 川端康成のほかに徳田秋聲と横光利一の結果も
- 教師なしの手法では川端康成の文章から大きく外れていない
- 教師ありの手法では川端康成に判別されたケースが圧倒的に多い

■ 結論

- 比較対象作家と比べ『空の片仮名』川端康成による作品の可能性が大きい

課題 1/2

■ 2群判別

- 川端康成と徳田秋聲
- 川端康成と横光利一

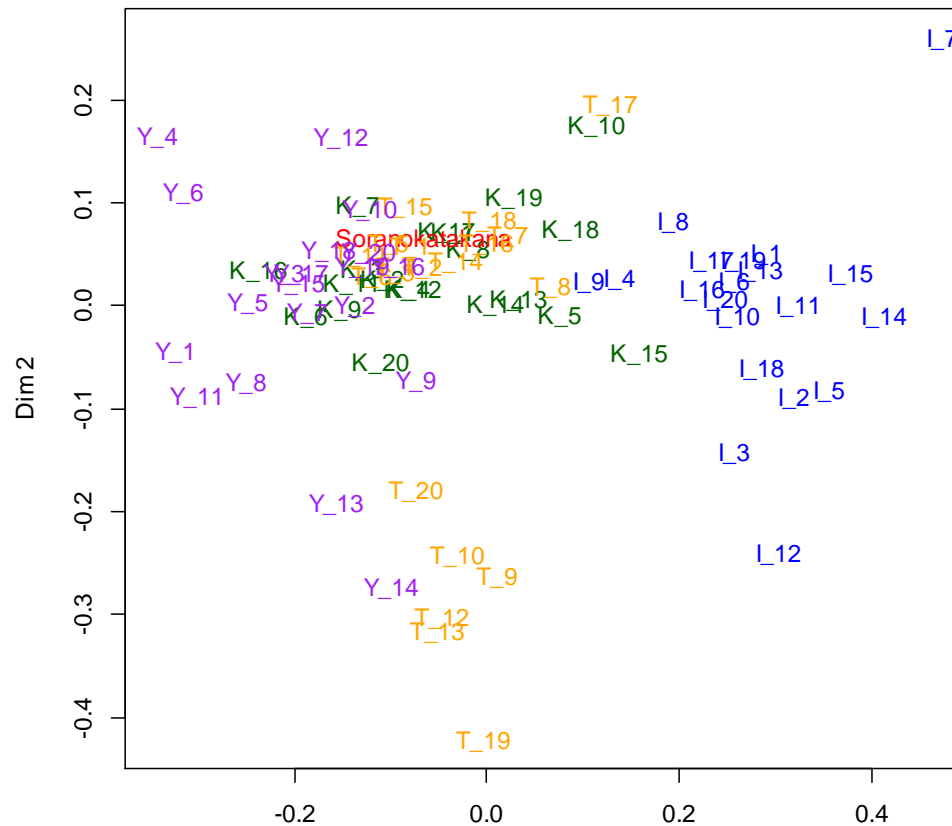


図3. タグbi-gramの対応分析個体スコア第1, 2成分散布図

参考文献

矢崎泰久 (2003). 『口きかんーわが心の菊池寛』. 飛鳥新社.

小谷野敦 (2013). 『川端康成一双面の人』.

金明哲 (2009). 『テキストデータの統計科学入門』. 岩波書店.

金明哲 (2013). 文節パターンに基づいた文章の書き手の識別. 行動計量学 40(1), 17-28.

金明哲 (2014). 統合的分類アルゴリズムを用いた文章の書き手識別. 行動計量学 41(1), 35-46.



ご清聴ありがとうございました。