



Text Mining Seminar 2018

計量文体学の観点からみた 宇野浩二の文体変化

劉 雪琴(博士後期課程) 金 明哲
同志社大学
文化情報学研究科

目次

- はじめに
 - 宇野浩二
 - 研究背景
- 研究目的
- データ
 - コーパス
 - 前処理
 - 特徴量
- 分析
 - 方法
 - 結果
- 結論と課題



宇野浩二



宇野浩二 (1891年7月26日 - 1961年9月21日)

- 日本の有名な作家
- 人生の哀感にユーモアを交えて描いた数多くの作品により、文壇の寵児

代表作：

- 「苦の世界」
- 「枯木のある風景」
- 「思ひ川」

受賞歴：

- 菊池寛賞 (1940)
- 読売文学賞 (1950)



3

宇野の創作と生涯

執筆の空白期：1927年-1933年 1944年-1946年



- 「蔵の中」(1919年)を発表し、饒舌な説話体の文章により有名。1927年頃から精神病を患い、執筆活動を中断して療養生活



- 「枯木のある風景」(1933年)の発表により、文壇に復帰。文体が**別人のように変化**



- 敗戦前後(1945年)の不安定な社会状況や妻の看病のため、再び執筆を中断。「青春期」(1946年)から執筆を再開、余韻のある文体



4

先行研究

- 精神病を患う前後
 - 文体の特徴は異なる(名詞・読点の増加など)
 - 梶谷(1971), 渋川(1974), 劉・金(2017a)等
- 変化が現れる時期
 - 形式名詞, 時制, 助詞や記号の連続使用が病後の作品に類似し, 宇野浩二の文体が入院する前に変わり始めた
 - 劉・金(2017b)
- 敗戦前後
 - 読点の多用が戦後の作品の独特な特徴
 - 梶谷(1971), 篠田(1972), 劉・金(2018)等



5

研究目的

- 他人に比べる際, 宇野浩二の三つの時期の作品における相違は明確に現れるのか?
- 文体が別人のように変化したと言われた宇野浩二の文体はどのように変わったのか?

本研究は, 同時期の文学者と比較し, 宇野浩二の文体変化の様子について, 品詞・記号・文節の側面から計量的分析を試みる



6

コーパス(201編)

宇野浩二：66編 (病前:27編； 病後:25編； 戦後:14編)
他の作家：各15編

番号	作家	生没年	作品	番号	作家	生没年	作品
1	泉鏡花	1873-1939	1920-1939	6	江戸川乱歩	1894-1965	1923-1938
2	永井荷風	1879-1959	1909-1956	7	横光利一	1898-1947	1919-1948
3	菊池寛	1888-1948	1916-1947	8	井伏鱒二	1898-1993	1925-1951
4	室生犀星	1889-1962	1919-1959	9	堀辰雄	1904-1953	1924-1942
5	宇野浩二	1891-1961	1921-1961	10	太宰治	1909-1948	1933-1948



7

前処理

宇野浩二：66編 (病前:27編； 病後:25編； 戦後:24編)
他の作家：各15編

- 小説の電子化・収集
 - ルビ削除
 - 新字・新仮名の統一
 - 会話文の削除
- 形態素解析・構文解析
 - MeCab・CaboCha
- データの集計
 - MTMineR



8

特徴量

- 形態素タグのunigram：品詞・記号の構成率
- 形態素タグのbigram：品詞・記号の接続関係
- 文節パターン：文節単位の情報

① 私 も 、 何気なく 、 立ち上がっ て 。

代名詞 助詞 記号 形容詞 記号 動詞 助詞 記号

② 私 も 、 何気なく 、 立ち上がっ て 。

代名詞_助詞 助詞_記号 記号_形容詞 形容詞_記号
記号_動詞 動詞_助詞 助詞_記号

③ 私も、 何気なく、 立ち上がって。

代名詞_も_、 形容詞_、 動詞_て_。



9

分析方法

- 教師なし機械学習法
 - 階層的クラスター分析(ward.2D 法 SKLD距離)
 - 個体間の類似度あるいは距離に基づいて、最も似ている個体から順次に集めてクラスターを作っていく方法
 - 主成分分析法(分散共分散行列)
 - 多次元データのもつ情報をできるだけ失わずに低次元空間に情報を縮約し、特徴の視覚化により元データを解釈する分析方法



10

分析の結果

- 形態素タグのunigram
- 形態素タグのbigram
- 文節パターン



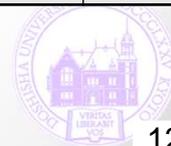
11

形態素タグのbigram

データ：201×118

118項目

タグのbigram	名詞_助詞	助詞_動詞	助詞_名詞	...	接続詞_形容詞	記号_接頭辞
宇野(A)_枯木のある風景	1636	784	881	...	1	2
宇野(A)_枯野の夢	5391	2533	2954	...	3	4
201個作品
菊池_身投げ救助業	765	436	368	...	2	0
菊池_青木の出京	2114	1051	990	...	0	0



12

文節パターン

データ：201×350

350項目

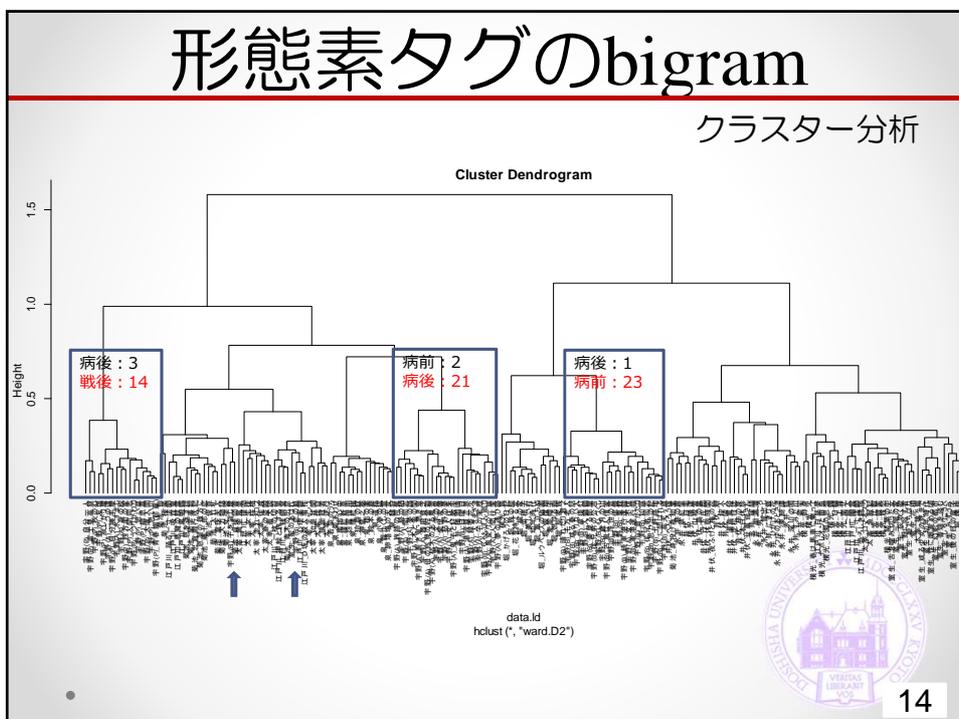
文節パターン	名詞_の	名詞_を	名詞_に	...	動詞_ながら_も_	動詞_たり_動詞_助動詞_。
宇野(A)_枯木のある風景	314	202	157	...	0	0
宇野(A)_枯野の夢	950	610	582	...	0	1
201個作品
菊池_身投げ救助業	109	120	81	...	0	0
菊池_青木の出京	386	276	186	...	1	0



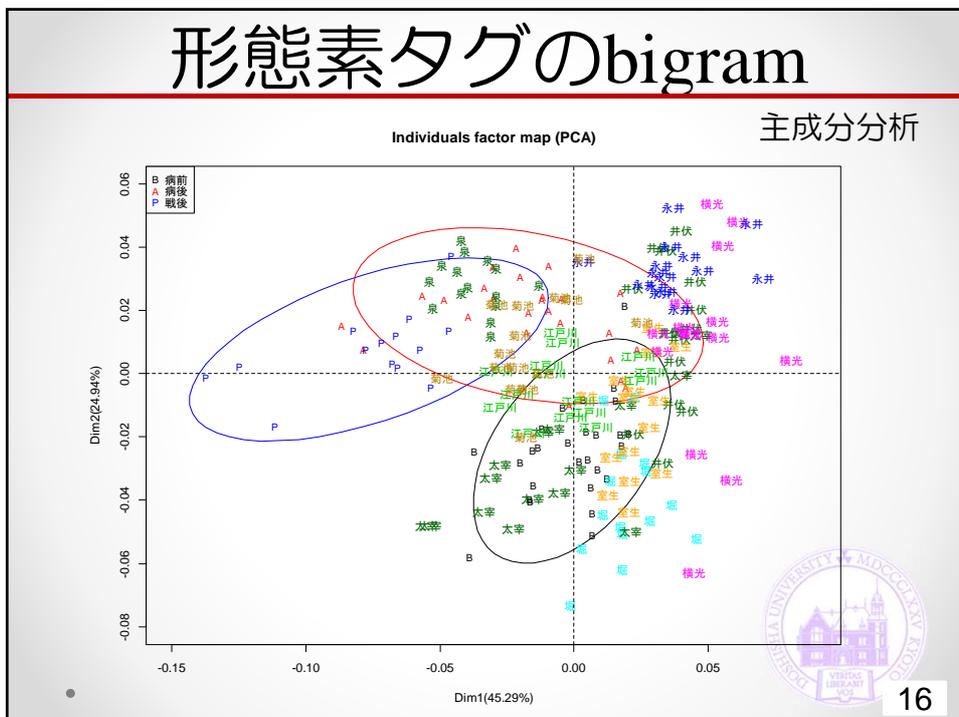
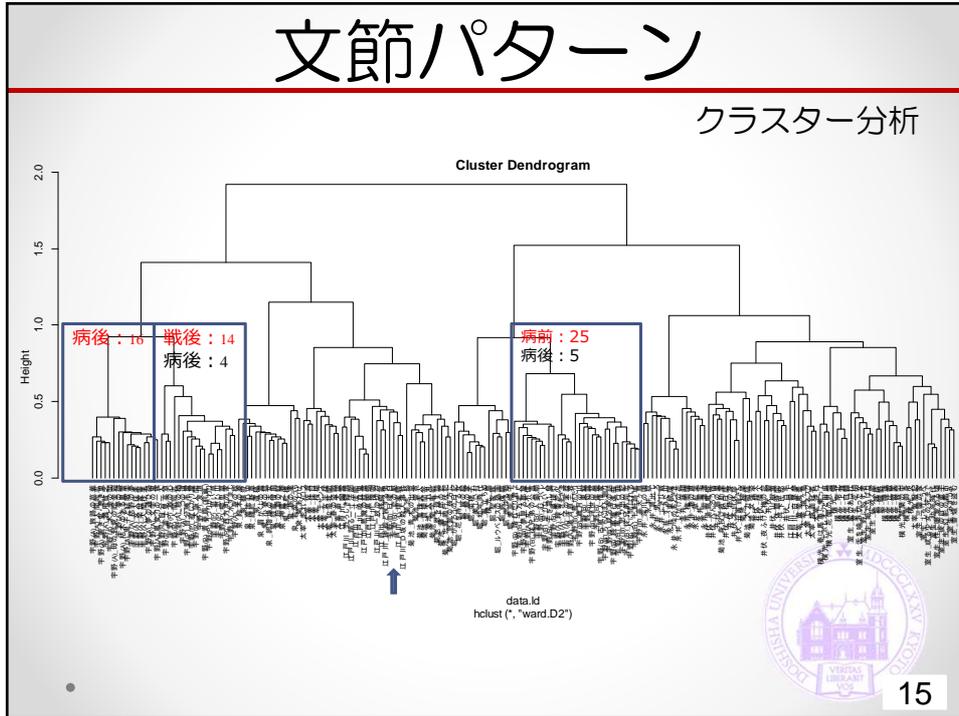
13

形態素タグのbigram

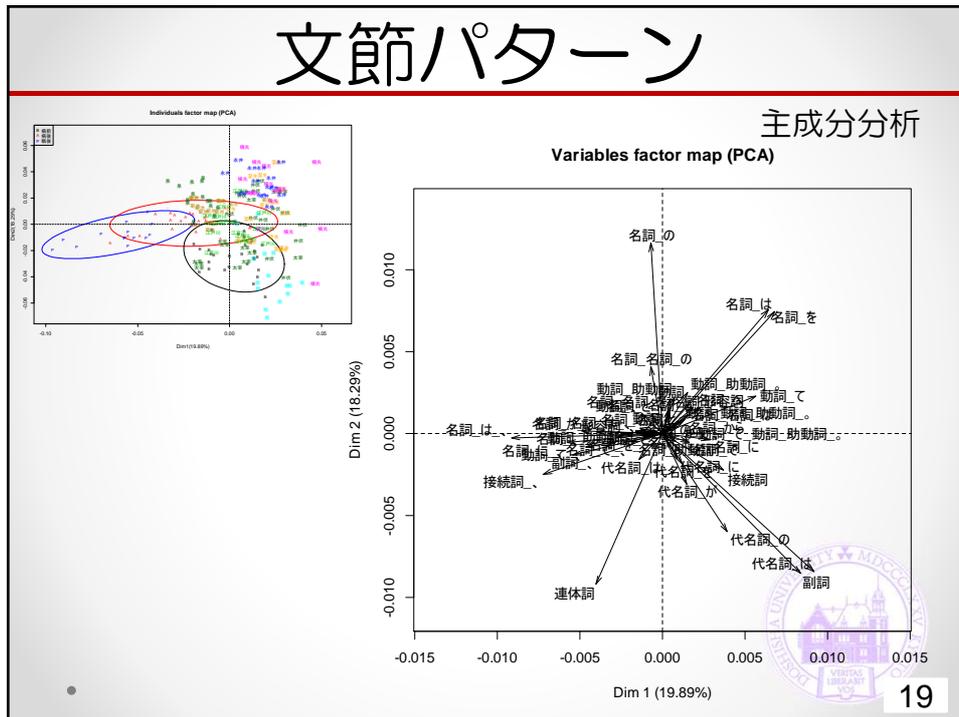
クラスター分析



14



文節パターン



まとめ

文体が別人のように変化したと言われた宇野浩二の文体はどのように変わったのか？

同時期の文学者と比較し、宇野浩二の文体変化の様子について品詞・記号・文節の側面から計量的分析(階層的クラスタ分析・主成分分析)

- 病前，病後，戦後の時の流れとともに変化
 - 病前：代名詞
 - 病後：名詞
 - 戦後：読点



結論

- 宇野浩二の作品における特徴が病前，病後，戦後の時の流れとともに変化(別人のように変化したといっても過言ではない)
- 戦後の作品群では異色な作品が多い
- 課題
 - 質的な研究の深め
 - 文体の他の側面の探索



21

参考文献

1. 篠田一士：「夢見る部屋の構図」すばる(10): 90-105. 集英社, 1972.
2. 春原千秋, 梶谷哲男：『現代文学者の病蹟—創作と狂気の謎—』, 新宿書房, 1971.
3. 金明哲：文節パターンに基づいた文章の書き手の識別, 行動計量学会, 40(1),17-28, 2013.
4. 金明哲：統合的分類アルゴリズムを用いた文章の書き手の識別, 行動計量学会, 41(1), pp.35-46, 2014.
5. 劉雪琴, 金明哲：「宇野浩二の病気前後の文体変化に関する計量的分析」, 計量国語学会, Vol.31, No.2, pp.128-143, 2017.
6. 劉雪琴, 金明哲：「入院する前に宇野浩二の文体は既に変わっていたのか」, 情報知識学会, No.27, pp.245-260, 2017.



19

ご清聴ありがとうございました



距離

本稿ではSKLD距離 (Kullback-Leibler Divergence) を用いる。その計算式を下記のように示す。

$$SKLD(X, Y)^2 = \frac{1}{2} \left\{ \sum_{i=1}^n x_i \log \frac{2x_i}{x_i + y_i} + \sum_{i=1}^n y_i \log \frac{2y_i}{x_i + y_i} \right\}$$



8