

小説における文体印象解析の試み

松本良太

- 要旨
- 動機
- 方法
- 使用データ
- 実験と結果
- 終わりに

要旨

小説を読んだ後に文章から感じる
「その作者らしさ、雰囲気」= 文体印象



客観的に示すことが難しい

これを、テキストデータを数量化・分析することで考察していく。

分析の際の着目点

①短文中の読点の数

②読みでの文字数

・・・訓読みで測定

③文のリズム

・・・短文における読点数の推移

動機

計量文献学における従来研究では「句読点」「文字数」「多頻出言語」などをパターンとして用いた研究がされてきた。



作者の文体から形成される文体印象についてはいまだ言及されていない。



そこで、新手法として作者の文体印象の特徴付け方法を提案する。

方法

①読点

理由

- ・作者ごとに特徴が出る
- ・読点数が読者に与える感覚の違い
多い→読みづらい
少ない(または0)→意味がわかりにくい

②読みでの文字数

理由

- ・作者の頭の中で最初に想起される語句の長さが作者の好む文の長さである
- ・小説のような文章の場合、論文などのような「型」の影響を受けないので、作者独自の判断で文字数が決まる。

・測定方法

漢字交じりの文字数と読みの文字数では、差がある場合が多いため、訓読みで数えたときの「読みの文字数」で計測する。



作者が頭の中で想起した言葉の長さを測定する。

③リズム

理由

- ・ある単文に用いられる読点数は、その前の単文に使用した読点数に何らかの影響を受けていると考えられる。

・測定方法

単文中の読点数を小説の始めから終わりまで順番に並べ、前後とのペアリングを行う。このペアリングの種類と頻度を測定する。

それと、あるペアリングは全体のペアリングの種類の中でどれだけ出現するか割合を算出する。(→読点数の推移確率)

使用データ

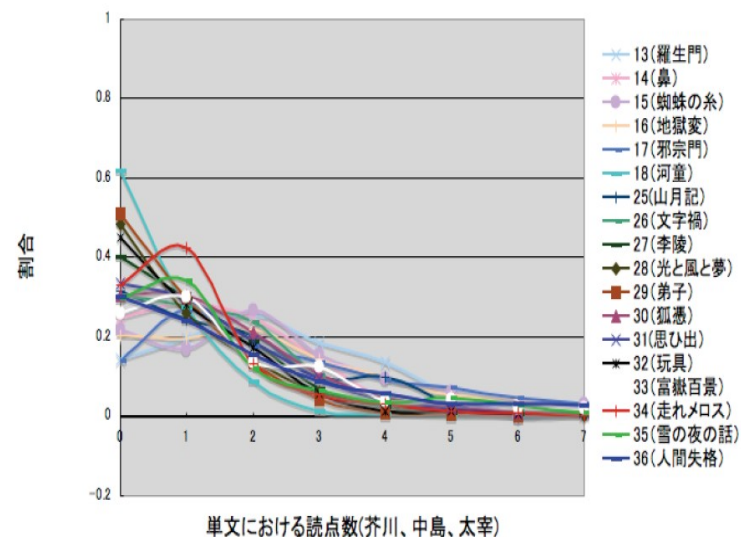
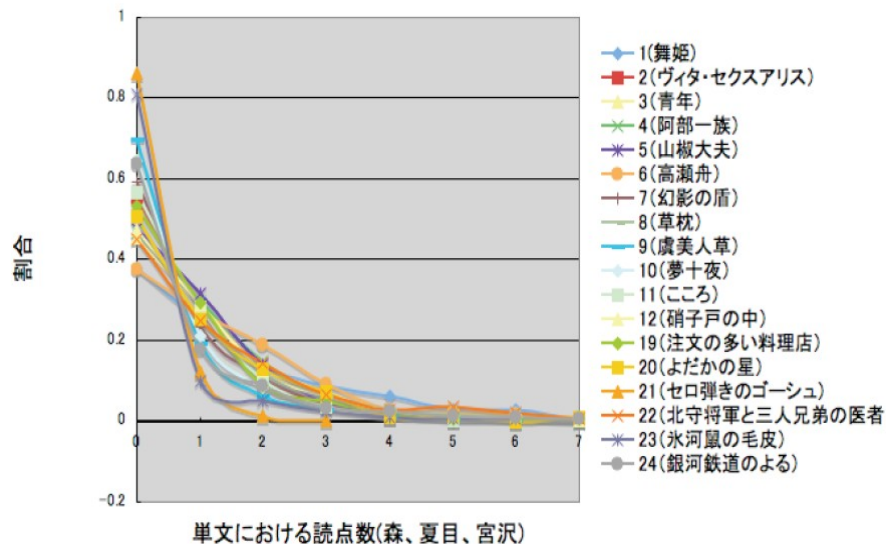
条件

- ・6名の著者による各6作品、計36作品を使用。
(森鷗外、夏目漱石、芥川龍之介、
宮沢賢治、中島敦、太宰治)
- ・ある作者の生存した時代に必ず他の比較対象者がいることを条件とする。
- ・各作者ごとに長編・短編をそれぞれ含み、かつ、書かれた時期が前期・中期・後期のものとして含まれる。

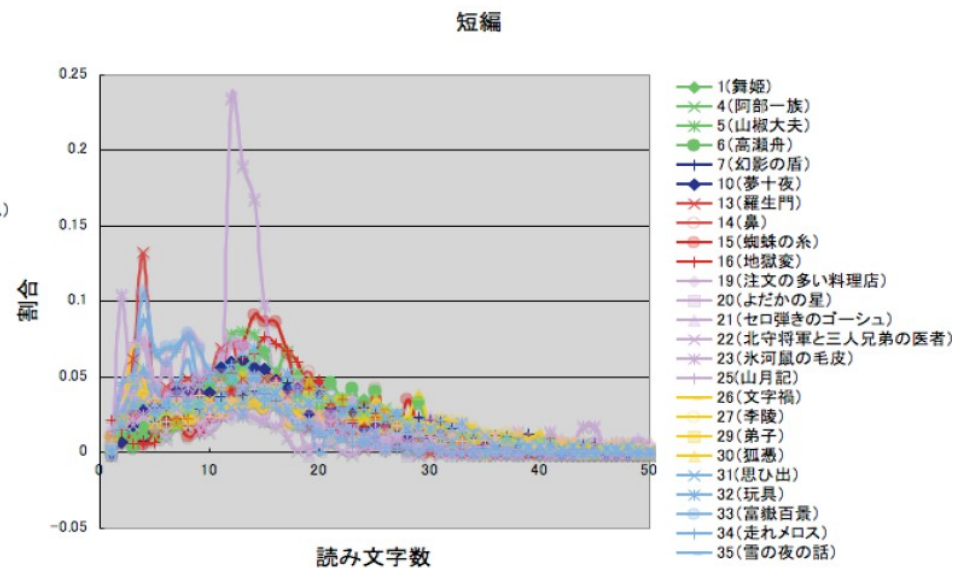
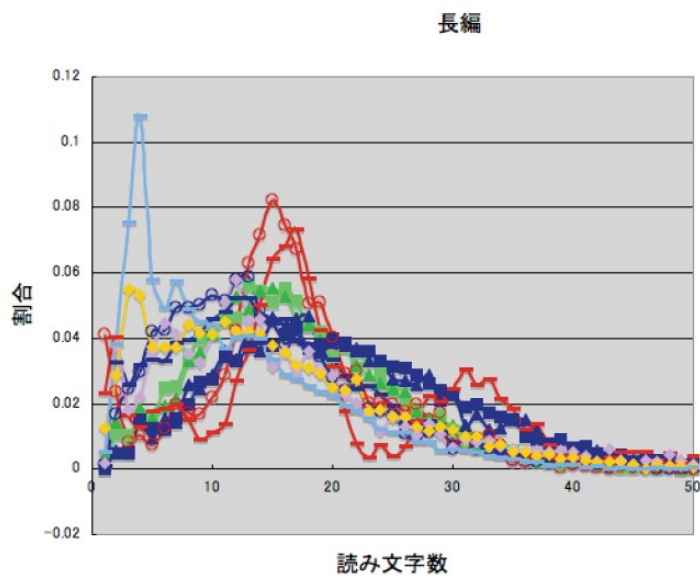
森鷗外 (1862-1922)	「舞姫」(1890)	宮沢賢治 (1892-1933)	「注文の多い料理店」(1921)
	「キタ・セクスアリス」(1909)		「よだかの星」
	「青年」(1910)		「セロ弾きのゴーシュ」
	「阿部一族」(1913)		「北守将軍と三人の医者」
	「山椒大夫」(1915)		「氷河鼠の毛皮」(1923)
	「高瀬舟」(1918)		「銀河鉄道の夜」(1927)
夏目漱石 (1867-1916)	「幻影の盾」(1905)	中島敦 (1909-1942)	「山月記」(1942)
	「草枕」(1906)		「文字禍」
	「虞美人草」(1908)		「李陵」(1942)
	「夢十夜」(1909)		「光と風と夢」
	「こころ」(1914)		「弟子」
	「硝子戸の中」(1915)		「狐憑」
芥川龍之介 (1892-1927)	「羅生門」(1915)	太宰治 (1909-1948)	「思ひ出」(1933)
	「鼻」(1916)		「玩具」(1935)
	「蜘蛛の糸」(1918)		「富嶽百景」(1939)
	「地獄変」(1918)		「走れメロス」(1940)
	「邪宗門」(1918)		「雪の夜の話」(1944)
	「河童」(1927)		「人間失格」(1948)

実験と結果

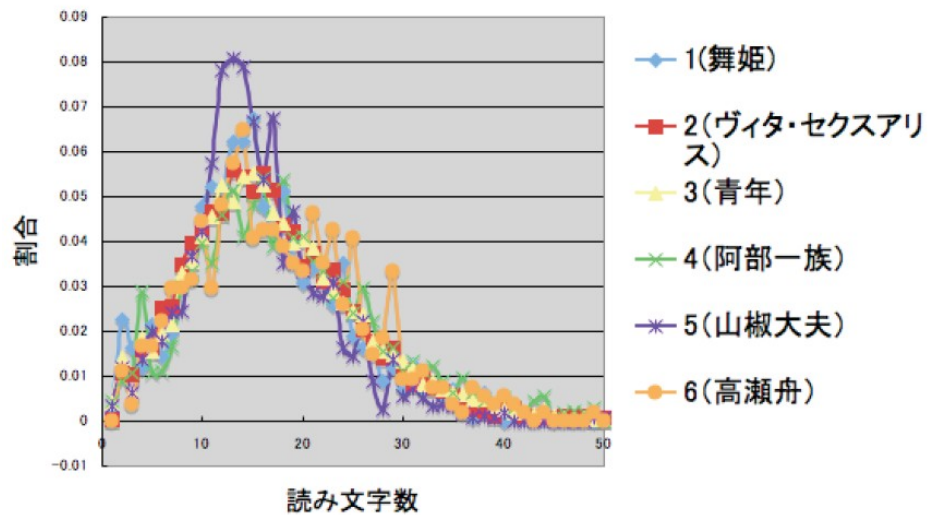
①読点



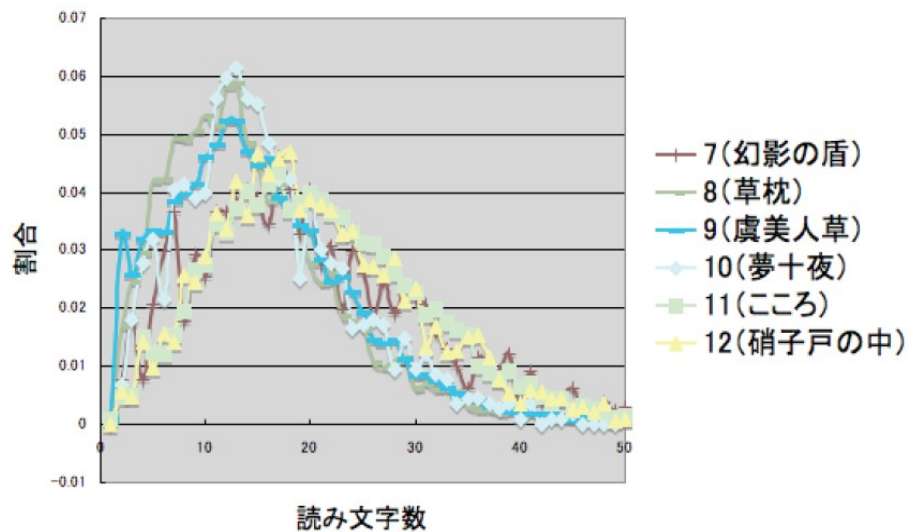
②読みでの文字数



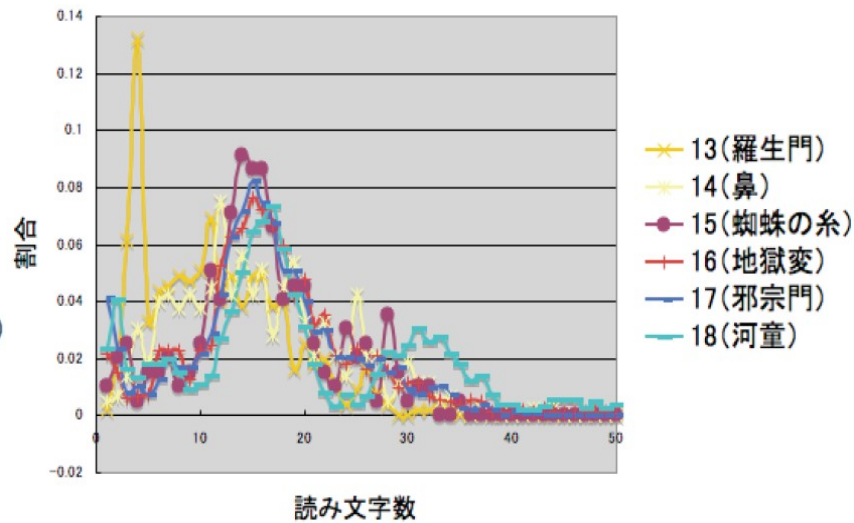
森鷗外



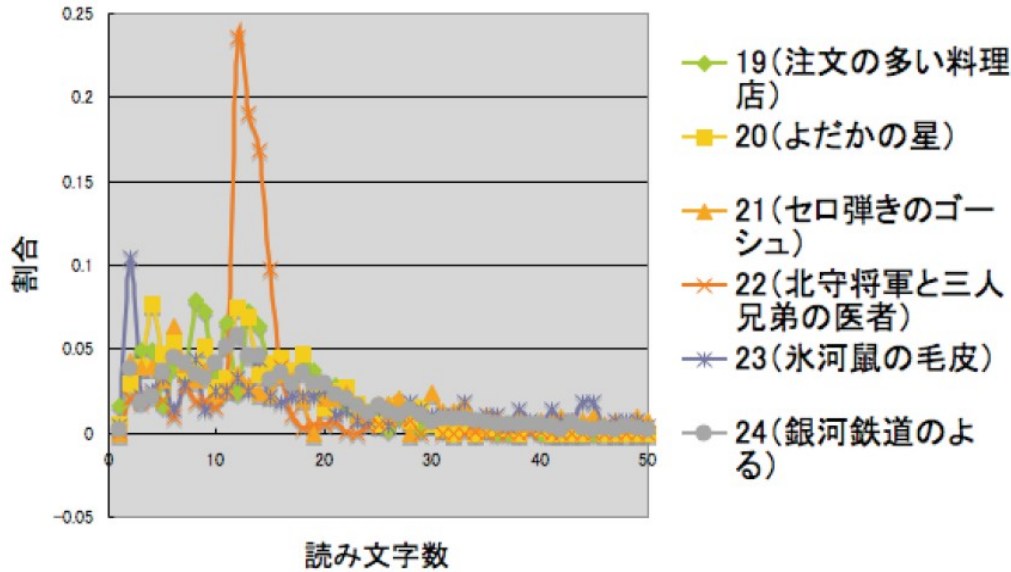
夏目漱石



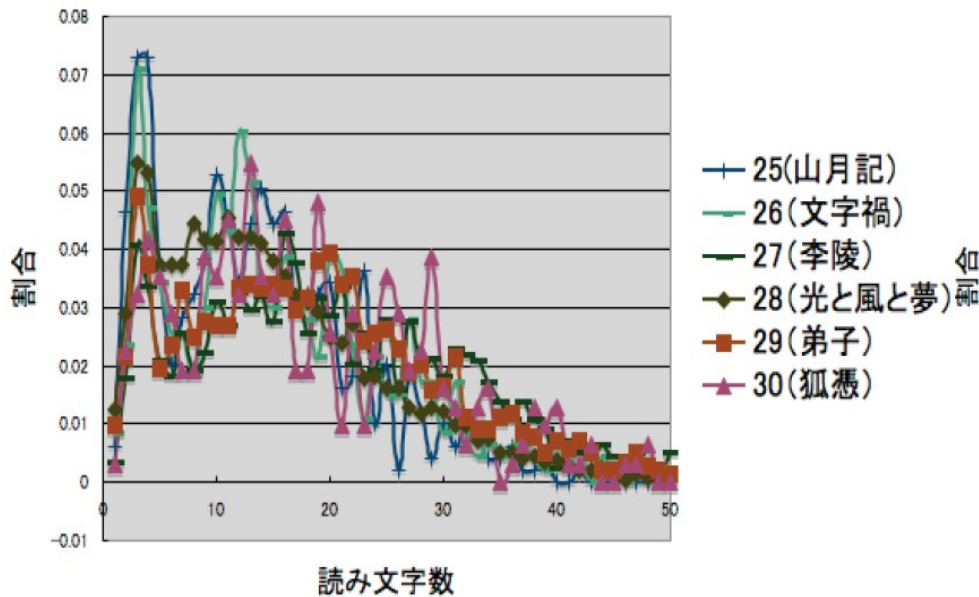
芥川龍之介



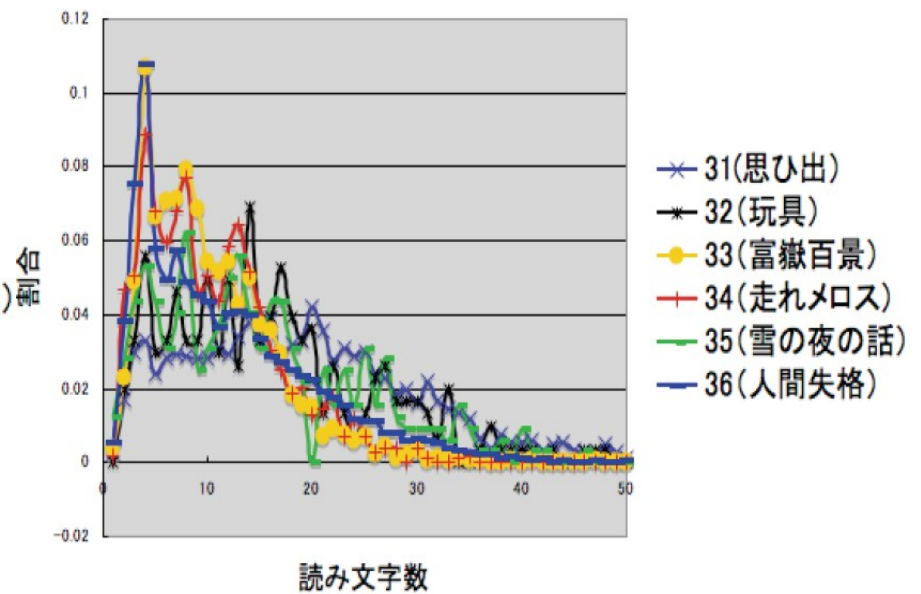
宮沢賢治



中島敦



太宰治



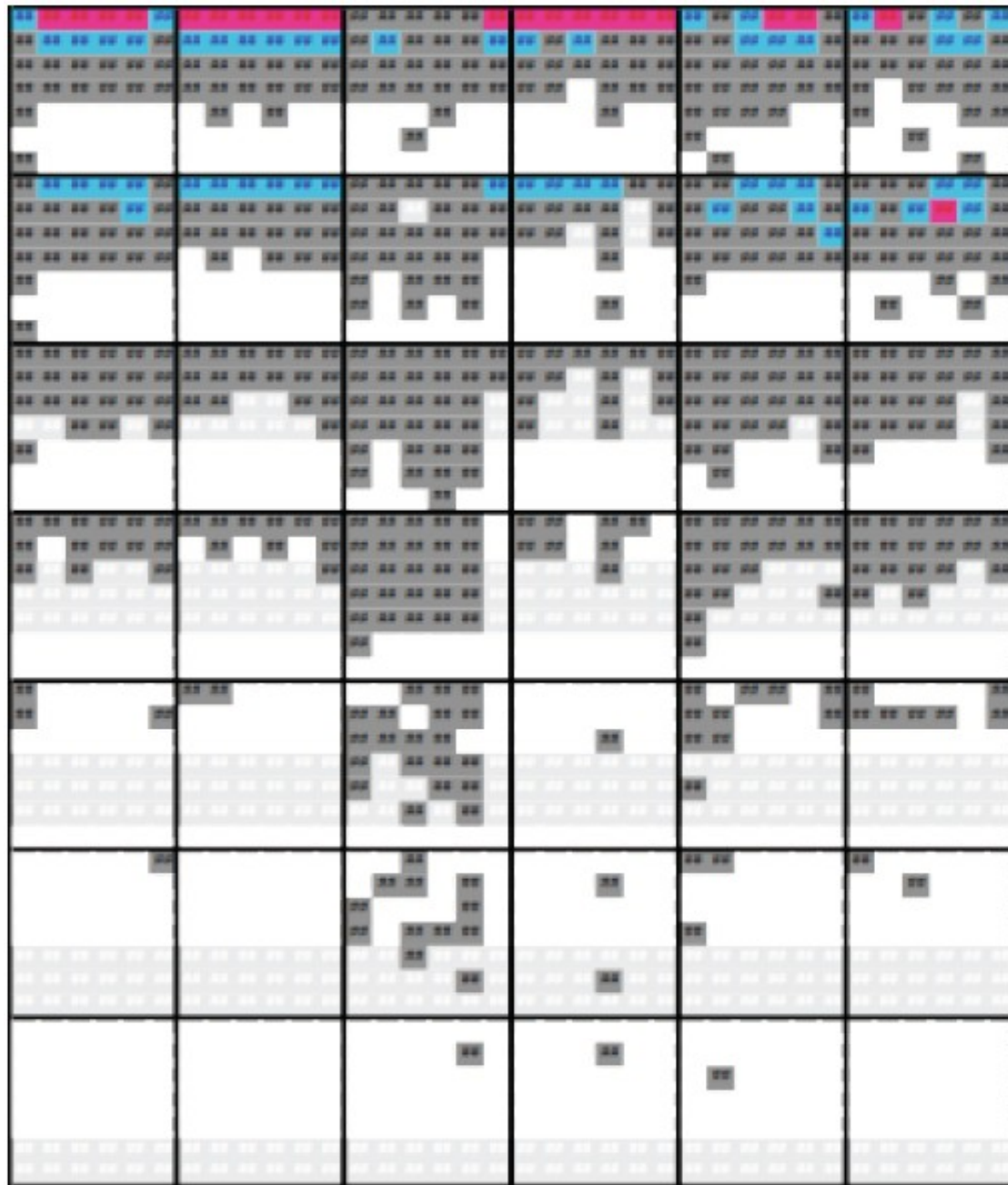
③リズム

白・・・0.001以下

青・・・0.1以上

0.2以下

赤・・・0.2より高い



終わりに

- ・分析の結果

今回の3つの観点のうち、特に作者によって読み文字数に差があることがわかった。

- ・今後の展望

読み文字数における推移確率(リズム)の算出
結果における心理学的検証