

統計を利用した小説文の解析 —芥川賞と直木賞の特徴—

広島大学附属高等学校 数学研究班2年
影本千晴・田中奈々・深水文乃・三宅菜月

▽動機

文藝春秋社が主催する日本を代表する文学賞、芥川龍之介賞・直木三十五賞（以下芥川賞・直木賞）は、両賞の境界が非常にあいまいで、度々受賞作品に関して議論が起こる。

この両賞受賞作品の文体に着目し、統計を用いて、それぞれの賞の受賞作に共通する客観的な特徴を見出したいと考えた。

▽解析対象

	作品名	作者	受賞年度
芥川賞	①暢気眼鏡	尾崎一雄	1937年上半期
	②忍ぶ川	三浦哲郎	1960年下半期
	③スティル・ライフ	池澤夏樹	1987年下半期
	④熊の敷石	堀江敏幸	2000年下半期
直木賞	⑤人生の阿呆	木々高太郎	1936年下半期
	⑥錯乱	池波正太郎	1960年上半期
	⑦ソウル・ミュージック・ラバーズ・オンリー	山田詠美	1987年上半期
	⑧虹の谷の五月	船戸与一	2000年上半期

▽仮説

・ひらがなの出現率

標準偏差 芥川賞>直木賞

・会話文の出現率

平均値 直木賞>芥川賞

・句読点間の平均文字数

作者ごとの個性が最も如実に表れると考えたため、両賞の特徴も出やすい。

・「」内（会話文）の平均文字数

平均値 直木賞>芥川賞

・一文の長さ

平均値 直木賞>芥川賞

・漢字間の平均文字数

標準偏差 芥川賞>直木賞

・1ページあたりの空白数

標準偏差 芥川賞>直木賞

解析対象は、芥川賞の受賞作品の平均的な文字数とされる2万文字に揃え、全文が2万文字に満たない作品はそのまま利用した。

▽考察

●会話文の出現率

標準偏差 直木賞>芥川賞

直木賞は「」を用いず会話を展開したり、地の文によって会話文で伝えられることを表現するなど、会話文の表現方法に作品ごとの個性がみられるためだと考えられる。

●句読点間の平均文字数

直木賞『錯乱』による平均値・標準偏差への影響が大きい
ため、外れ値として除外して考える。

平均値・標準偏差 芥川賞>直木賞

●一文の長さ

芥川賞『熊の敷石』による平均値・標準偏差への影響が大きい
ため外れ値として除外して考える。

平均値 芥川賞>直木賞

標準偏差 直木賞>芥川賞

最大値の平均値 芥川賞>直木賞

芥川賞のほうが文章中に長文を用いやすい傾向にあると思われる。

●漢字間の平均文字数

平均値・標準偏差に大きな差異はみられなかった。

最大値の平均値・標準偏差 芥川賞>直木賞

芥川賞のほうが形容詞や名詞などをあえてひらがなやカタカナで用いやすい傾向にあると思われる。

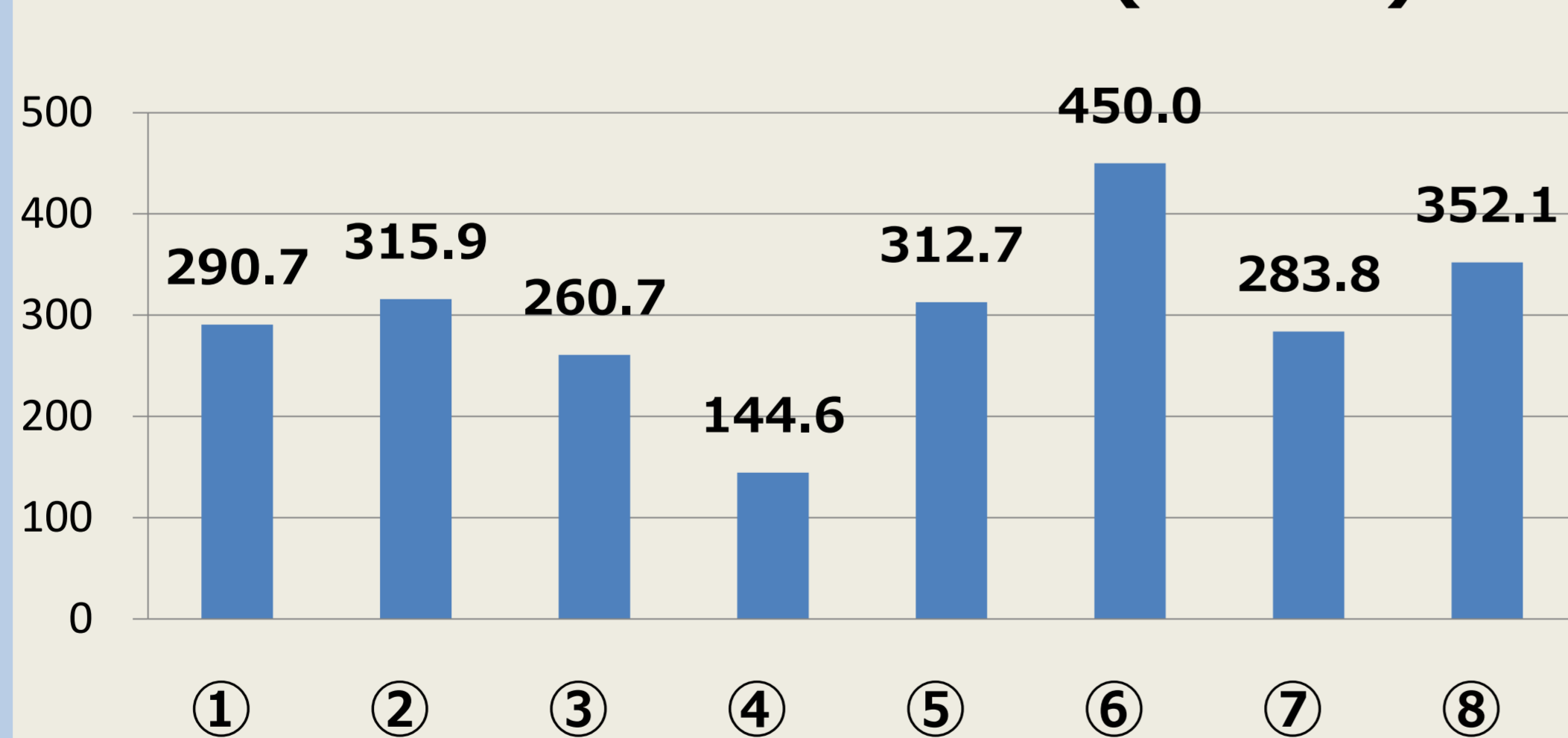
●1ページあたりの空白数

平均値 直木賞>芥川賞

標準偏差 芥川賞>直木賞

直木賞のほうが改行が多く、一段落が短いのではないかと
思われる。

1ページあたりの空白数(文字数)



▽結論

賞の特徴というよりむしろジャンルや、作者ごとの特徴があらわれているといえる項目があったが、ひらがなの出現率・会話文の出現率・句読点間の平均文字数・一文の長さ・漢字間の平均文字数・1ページの平均空白数の6項目において芥川賞・直木賞の特徴といえると考えられる違いがあらわれた。

●芥川賞

- ・句点をあまり用いず、読点を多用し、文章中に長文を用いやすい傾向にある。
- ・形容詞や名詞をあえて漢字で表記せず、ひらがなやカタカナにして用いる傾向にある。

●直木賞

- ・会話文の表現が多様な傾向にある。
- ・改行が多く、一段落が短くなりやすい傾向にある。

Analysis of Novels Using Statistics

Hiroshima University High School Second Grade Mathematical research team
Chiharu Kagemoto Nana Tanaka Ayano Fukamizu Natsuki Miyake

▽ Motive

Akutagawa Ryunosuke Prize and Naoki Sanjugo Prize are ones of the most famous novel prizes in Japan.

It is said there are some differences between these, but those differences are so ambiguous that there have been some discussion about them.

We will examine novels which won these prizes to define the characteristics through mathematical approaches.

▽ Objects of Analysis

	Title	Author	Winning Year
Akutagawa Prize	① Nonki megane	Kazuo Ozaki	1937
	② Shinobukawa	Tetsuro Miura	1960
	③ Sutiru-raifu	Natsuki Ikezawa	1987
	④ Kuma no shikiishi	Toshiyuki Horie	2000
Naoki Prize	⑤ Jinsei no aho	Takatarou Kigi	1936
	⑥ Sakuran	Shotarou Ikenami	1960
	⑦ Soul Music Lovers Only	Eimi Yamada	1987
	⑧ Niji no tani no gogatsu	Yoichi Funado	2000

▽ Hypothesis

● Ratio of appearance of hiragana

Akutagawa's standard deviation > Naoki's

● Ratio of appearance of conversation

Naoki's average > Akutagawa's

● Average between punctuation marks

It reflects author's characteristics clearly

● Average letters of conversation

Naoki's average > Akutagawa's

● Length of sentences

Naoki's average > Akutagawa's

● Average letters between kanji

Akutagawa's standard deviation > Naoki's

● Amount of blanks per page

Akutagawa's standard deviation > Naoki's

We use about **20,000 letters** from each books.
※If the books originally have less than 20,000 letters, we use the whole text.

▽ Consideration

● Ratio of appearance of conversation

Naoki's standard deviation > Akutagawa's

● Average between punctuation marks

Akutagawa's average

&

> the other

Akutagawa's standard deviation

※Data of ⑥ 『Sakuran』 are excluded because they are outlier.

● Length of sentences

• Akutagawa's average > Naoki's

• Naoki's standard deviation > Akutagawa's

※Data of ④ 『Kuma no shikiishi』 are excluded because they are outlier.

● Average letters between kanji

Average & standard deviation

→No differences(≠the other)

Average of largest number

&

Standard deviation of largest number

→Akutagawa's > Naoki's

= Akutagawa's books tend to use hiragana, not kanji for nouns and adjectives

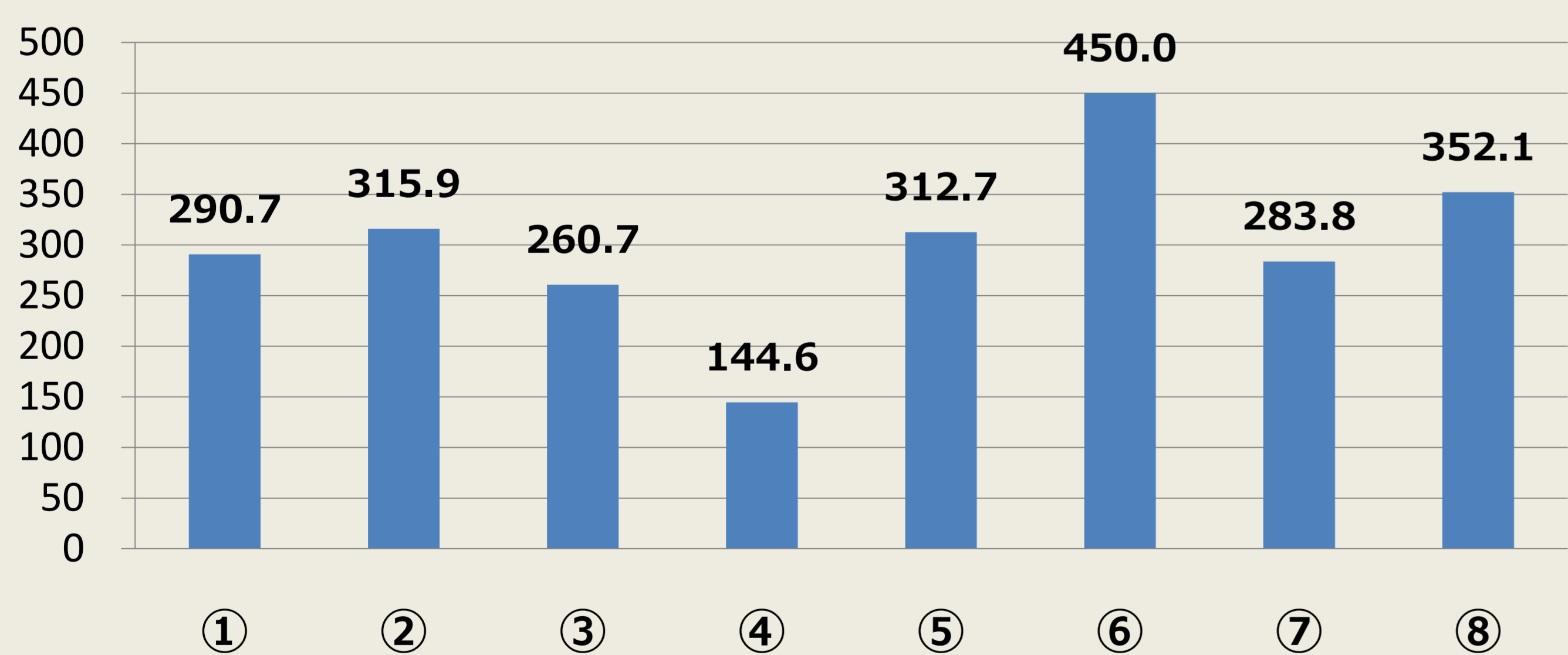
● Amount of blanks per page

• Naoki's average > Akutagawa's

• Akutagawa's standard deviation > Naoki's

= Naoki's books have shorter paragraph.

Amount of blanks per page (Letters)



▽ Conclusion

Six items are characterized by the differences of each prize: ratio of appearance of hiragana, Ratio of appearance of conversation, average between punctuation marks, length of sentences, average letters between kanji, and amount of blanks per page.

Akutagawa Prize

Tend not to use period; longer sentences; Tend to use Hiragana for noun and adjective intentionally

Naoki Prize

Various styles in conversation; shorter paragraphs