

落語にみるジップの法則

:回帰分析とフィラー

大井雄大

はじめに

ジップの法則とは、1949年にZipfにより紹介された経験則であり、出現順位を r として、その修験頻度を f 、定数を C としたとき、

$$f \times r = C$$

であるとしたものである。

授業で扱ったものは、テキストの中の語の出現頻度であった。

テキストとして出力するのと、音声として、出力するのでは明らかに使う身体的部位が違う。これには少なからず差異が生じると考えた。そこで、今回は発話内容として「落語」をとってきて、これを対象としてジップの法則を考察することにした。これに関して、佐野幸恵(2020)での「2.2 Zip則とベキ関数」で述べられている。反比例の式をベキ関数に一般化したものに着想を得て、両辺に対数を取ることによって、直線回帰分析が可能であると考えた。

方法

落語「死神」の公演音声 Word のディクテーション機能を用いて文字起こしし、そのテキストデータを MeCab を使って形態素解析した。解析では、語意外の要素 (句読点など) を除外して、各語の出現数を数値化した。これらのデータをもとに、出現数の順位と出現数をグラフに表した。さらに、その出現数に対して自然対数を取り、それを用いて回帰分析を行った。(ただし、出現数が自然数に限定されることから、出現数1の語は初出部分から一定値にとどまるため、性質上除くことにした。)この分析により、落語「死神」における語の使用頻度やその特徴を定量化した。

結果

出現した語、1226語を出現数順に並べた結果、図1の様になった。

また、回帰分析の結果、出現順位に自然対数を取った変数の係数が-1.5となり、定数項が10.12となった。

結論：話すからには身体的制限がさ らにかかる

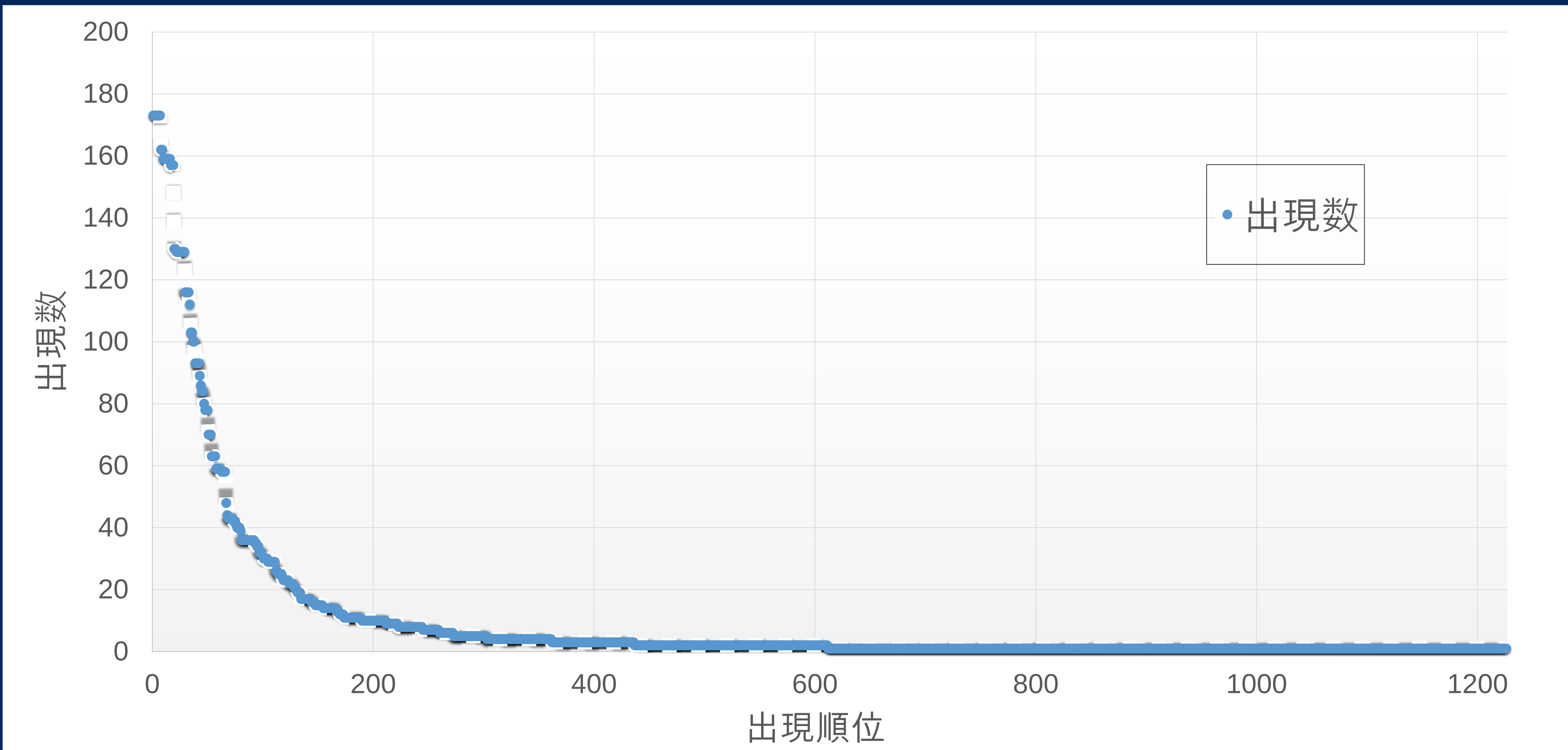


図1: 落語「死神」にみる各語の出現順位とその出現数

書字形

ん

な

で

だ

ね

た

て

の

よ

か

品詞

助動詞/助詞/感動詞

助動詞/助詞

助動詞/助詞/接続詞

助動詞

助動詞/助詞

助動詞

助動詞/助詞/接続詞

助詞

助詞

助詞

表1:落語「死神」にみる語の出現順位上位10位

$$f(r) \approx \frac{C}{r^{3/2}}$$

* r は出現順位

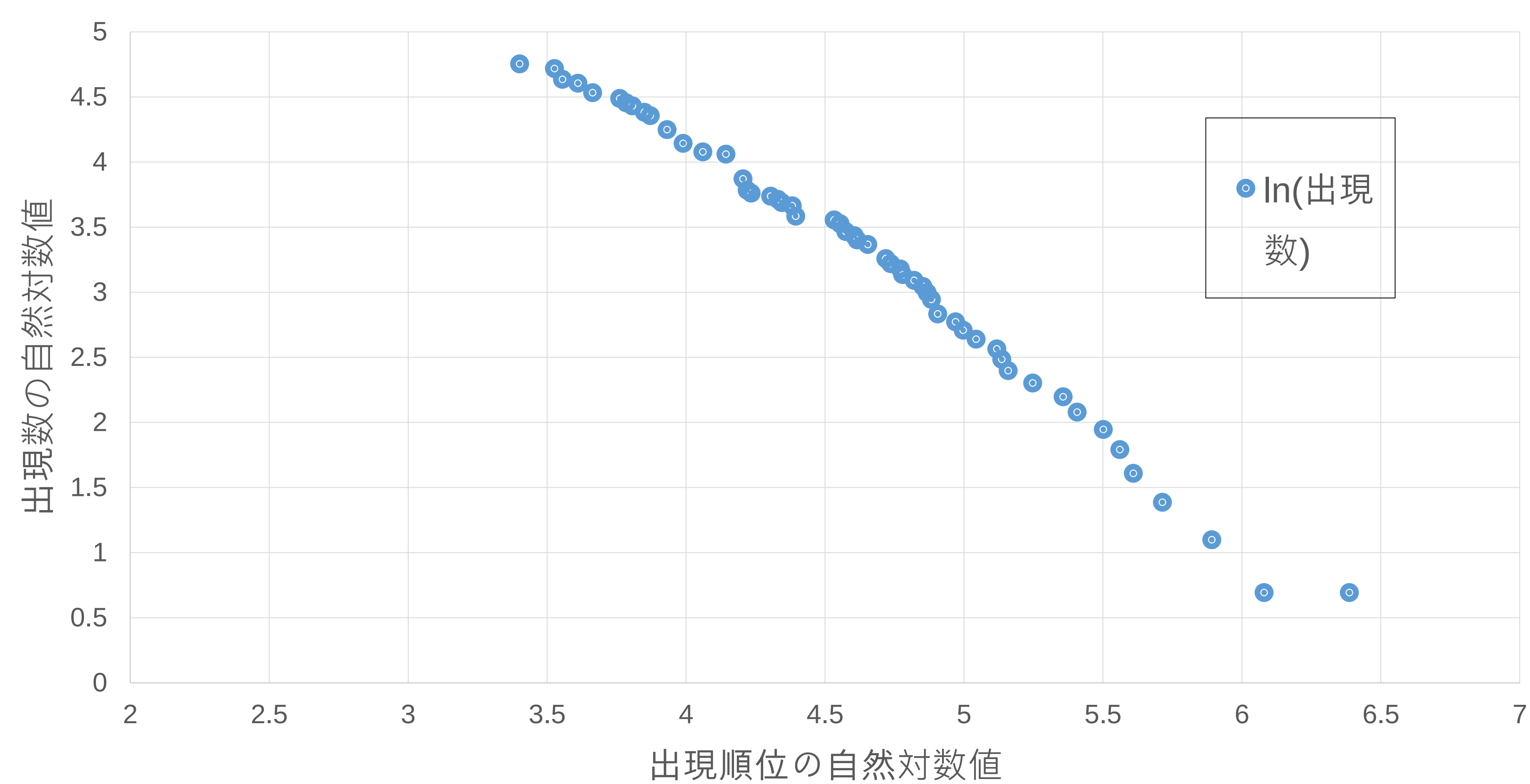


図2: 出現数2以上の値で両辺対数取った場合の関係

(結果の続き)

また、集計した語の上位10位をまとめたのが表1になる。

補足として、図2に従属変数である出現数と独立変数としての出現順位それぞれ自然対数を取った場合のグラフを表示している。

考察

結果をうけ、まず一つに図1を概観すると、授業内で扱った「最小努力の法則」は適応されるのだろうと考えた。

これに加えて、回帰分析による係数の推定によって、佐野幸恵(2020)で言及されたベキ関数として一般化したものとして、反比例ではなく、より出現順位に対して減少するような振る舞いをしていることが推定できる。これは、「はじめに」にて述べた、身体性というものに関わっているというものを考えた。

というのも、表1で示した上位10位の内、一位に感動詞としての「ん」が含まれていることにある。これ以外にも、19位には「えー」などの感動詞-フィーラが散見されている。これは話し言葉特有のものとして、少なくとも身体性の影響を垣間見ることができるだろうと考えた。

おわりに

考察をまとめると、今回の操作によって、話し言葉がテキストよりも身体性の影響があり、その影響は「最小努力の法則」を強める方に働いていると考えることができる。

文献

- ・佐野幸恵(2020)「Zipf 則・Heaps 則とその周辺」、情報の科学と技術、70巻2号・87～89
- ・桃月庵白酒(2015)、「毎日新聞落語会 桃月庵白酒3」
- ・George Kingsley Zipf (1949)「Human behavior and the principle of least Effort」、22～27、Addison-Wesley