

# 計量国語学会

特別講義

「『データで学ぶ日本語学入門』  
.....私ならこう読む」

荻野綱男

2018.9.29

計量国語学会(編)(2017)『データで学ぶ日本語学入門』  
朝倉書店 をどう読むか。

荻野は日本大学文理学部でこの本を教科書として使って  
講義を行った。

その経験を基に「私ならこう読む」をまとめて述べる。

# 第1章 音声・音韻

## (1) 図1.1

「どの音がよく使用されているの？」

一見すると、②が多いように見えるが、それは正しいか。

① 5種類 11.5% → 1種類あたり平均 =  $11.5/5 = 2.3\%$

② 131種類 72.0% → 1種類あたり平均 =  $72.0/131 = 0.55\%$

③ 3種類 16.5% → 1種類あたり平均 =  $16.5/3 = 5.5\%$

# 第1章 音声・音韻

## (2) 母音と子音・半母音

p.3 l.8 子音の数は、なぜ研究者によって数え方が異なるのか。  
音声学的に単音として数える vs. 音韻論的に音素として数える

(本書は音声学的立場)

音声学的な違いをどこまで区別するべきか  
音韻論で本当に音素の数が確定できるか

p.4 l.6 子音が比較的少ない=15~18種類

子音が平均的=19~25種類

# 第1章 音声・音韻

## (3) モーラ

p.6 研究者によって(さらに時期によって)モーラの種類数の数え方が違っているとき、他言語と比べることは可能か。

p.6 l.-3 なぜ英語は音節数の種類が多いか。

母音の前と後ろのそれぞれに CCCVCCCC のような子音連続がある。

p.6 l.-2 なぜ中国語は音節数の種類が多いか。

CVC のような閉音節がある。四声という声調の区別がある。

日本語ではモーラを用い、他言語では音節を用いて、同じ基準で比較しているといえるか？

# 第1章 音声・音韻

## (4) 音声言語に出現したモーラと書記言語に出現したモーラ

p.9 図1.2 を p.1 図1.1 と比較すると、モーラの使用率はほぼ共通している。

しかし、(たとえば)図1.1 の 11.5% と、図1.2 の 10.9% は「ほぼ同じ」とみていいのか。

$11.5\% = 1,652,337 / 14,368,149$   $10.9\% = 5,402 / 49,556$  を、とある手法で有意差計算すると、0.1%水準で有意差ありとなる。11.5%のほうが10.9%よりも有意に大きい。

(統計的な意味での)「有意差」と、(日本語学としての)「意味のある差」は別物である。

# 第1章 音声・音韻

## (4) 音声言語に出現したモーラと書記言語に出現したモーラ

p.9 図1.3 と図1.4 を比較して、モーラの使用率に差があるといっ  
ていいか。→いい。

では、なぜ差がある(差ができる)のか。

本書中では「解釈」が述べられていない。

→p.11 1.4「まとめ」で語種の違いだという解釈が載っている。

どうすれば、語種の違いのためだと確認できるのか。

## 第2章 文字・表記

(1) 図2.1 の解釈(なぜこのような結果になったか)

Q1. なぜカタカナが増えたのか

Q2. なぜローマ字が増えたのか

Q3. なぜ「そのほか」(数字や記号)が増えたのか

Q4. 漢字とひらがなを区別して論じるべきか、その必要はないか

35.8:56.0 $\div$ 26.9:35.7 あるいは 35.8:26.9 $\div$ 56.0:35.7 と見るか、ひらがなは漢字よりも減少していると見るか。

p.14 用語の問題:「ポイント」とは 正式には「パーセントポイント」

50%の比率のものが10%増えたら、60%? 55%?



## 第2章 文字・表記

### (2) 延べと異なり

1966新聞が「円一圓」を一つの字種として扱ったのはなぜか。

漢字入力がまだ普通には行えない状況で、区別がむずかしかった。

国立国語研究所の独自のコード(漢字テレタイプ)

そこにはない文字は特殊な入力方法を採用

漢字のJISコードが制定されるのは1978年が最初。

## 第2章 文字・表記

### (3) 漢字含有率

Q5. 「漢字」の使用量は、なぜ新聞より雑誌が低いのか。

Q6. 「そのほか」の使用量は、なぜ新聞より雑誌が高いのか。

## 第2章 文字・表記

### (4) 文字表記に関する人間心理

p.18 ネット調査(Web 調査)とは、ネット調査会社に委託して、事前に募集してある数十万～数百万人からデータを集める質問調査法。

p.20 図2.5 と図2.6で年齢差があるか。

図2.6 のデータを有意差検定してみると次ページのようになる。

## 第2章 文字・表記

### (4) 文字表記に関する人間心理

	20歳代	30歳代	40歳代	50歳代
20歳代	--	なし	5%	0.1%
30歳代	なし	--	なし	5%
40歳代	5%	なし	--	なし
50歳代	0.1%	5%	なし	--

隣接する年齢層間の有意差はないが、総体的には年齢差がある。

二つ離れた年齢層間には有意差がある

一貫した年齢差の傾向が認められること

それにしても、100人くらいのデータだと、目で見ても差があることがわかって「有意差」がないことがあるので要注意。

## 第3章 語彙

### (1) 図3.1

図3.1 の縦軸は単語数であるが、これは分類語彙表の語数を数えたものであるから、収録された単語数であり、言語作品で言われるような「延べ語数、異なり語数」という概念は当てはまらない。

分類語彙表やすべての国語辞典の見出し語の語数は、出現度数(頻度)の情報を持ち合わせていないので、語彙調査でいう場合の異なり語数のような意味合いを持っている。

# 第3章 語彙

## (2) 計量語彙論

p.24 Ⅰ-9 短単位と長単位

短単位:

型紙 | どおり | に | 裁断 | し | て | 外出 | 着 | を | 作り | まし | た。

長単位:

型紙どおり | に | 裁断し | て | 外出着 | を | 作り | ました | 。

p.25 Ⅱ.3-4 「TTR は、一般的に延べ語数が増えるほどその値が低くなる」のはなぜか。

## 第3章 語彙

### (3) 語彙量の実態

p.27 表3.3 表3.4 では、語種の分類において、人名・地名を別扱いしている。なぜそのようにするのか。

# 第3章 語彙

## (4) 基本語彙と特徴語

p.29 特化係数の考え方を理解しておくこと  
ただし、特化係数は万能ではない。

表3.6 では、全体の延べ語数は 134 であり、まあ大丈夫であるが、これが小さい(数十以下?)と不安定な結果になる。



## 第4章 文法・意味

### (1) 時代による変遷、同時代の使用のゆれ

時代による変遷を知るには、書かれた(話された)日付がはっきりわかることが必要。

雑誌・新聞は刊行年月日、国会会議録は会議開催日

ラジオ・テレビ番組も使えるはずだが、あまり使われない。  
(録画が残っていない。)

質問調査では、年齢差が時代による変遷に関わることがある。

## 第4章 文法・意味

(2) 分野による相違、使用者による相違

p.39 図4.3 の白書と知恵袋の違いは何か。

本書では、II.8-13 で「デ」に着目しているが、……。

むしろ①「ノ」の比率、②助詞の総量に着目すべき

p.39 図4.4 では学会講演(A)と自由会話(D)の違いが大きい。

本書では、p.40 I.1 で「ニを伴う複合辞」と「トを伴う複合辞」に着目しているが、……。

p.41 図4.7 は、かなり問題がある。10名約300例では(1人分が30例程度しかなく)データが少なすぎる。

## 第4章 文法・意味

### (3) 文法現象を数値化することの意味

数値化することに意味があるなら、本文中でも数値を示すべきではないか。

林(1982)の示したデータは、同書 p.365 によると、以下のようなものである。  
(グラフは省略した)

- A. ~が~で~を(「一郎が台所で酒を飲む」型) 10
- B. ~が~を~で(「一郎が酒を台所で飲む」型) 3
- C. ~で~が~を(「台所で一郎が酒を飲む」型) 3
- D. ~で~を~が(「台所で酒を一郎が飲む」型) 0
- E. ~を~が~で(「酒を一郎が台所で飲む」型) 0
- F. ~を~で~が(「酒を台所で一郎が飲む」型) 1

確かに10例は3例よりも多いが、……。

サインテストの計算をすると、10例と3例であれば、前者が多いといえる。間違える確率は 4.6%。しかし、3例が1例よりも多いとはいえない。

# 第5章 文章・文体

## (1) 図 5.1 表 5.1

品詞構成比率を求めるときは、延べ語数か、異なり語数か。

なぜ8品詞でなく4分類で見ていくのか。

p.45 l.-6 なぜ助詞・助動詞はカウントの対象外なのか。

素材テキストのサイズが大きく異なる。

自立語の総数を見ると、最大と最小で約10倍の差  
これでいいのか(揃える必要はないか)。

# 第5章 文章・文体

## (2) MVR

p.50 II.7-8 形容詞文はありさま描写、動詞文は動き描写としている。

動作を意味する形容詞・形容動詞はどちらに属すべきか。

活発な、すばしこい、おてんばな、俊敏な状態を意味する動詞はどちらに属すべきか。

彼は一郎の弟に当たる。

AとBは異なる(一致する)。

お前はすでに死んでいる。

お金が落ちている。

ビルがそびえている。

雨の中をじっと立ち尽くしている。

形容詞と動詞でほぼ同義の表現がある(「悲しい」と「悲しんでいる」)

# 第5章 文章・文体

## (2) MVR

p.52 名詞比率の大小とMVRの大小を組み合わせると、四つに分類される(はずである)。

④名詞比率Nが大きく、MVRが大きい文章

実際、そういうものがあるのか。

あるとすれば、p.53 図5.4 の右上に位置するはずである。