# 反復語の使用実態から見る話し言葉と書き言葉の連続性 −コーパスを用いた定量的分析を通して−

鯨井 綾希（東北大学大学院文学研究科）†

# Continuity between Spoken and Written Language as Seen from the Use of Repeated Words; A Quantitative Corpus Analysis

Ayaki Kujirai (Graduate School of Arts and Letters, Tohoku University)

## １．本発表の目的

本発表では、『日本語話し言葉コーパス』（以下CSJ）と『現代日本語書き言葉均衡コーパス』（以下BCCWJ）を利用して、文章や談話のまとまりや意味的連続性を形成するのに重要な役割を担う同一名詞の反復（以下これを反復語と記述）の使用実態を定量的な側面から明らかにし、その上で話し言葉や書き言葉といった言語表現上のバリエーションを考察に加え、反復語の量的側面から見たときのそれらの関係性を見出すことを目的とする。

## ２．調査上の対象・資料・方法

### ２．１ 調査対象

本発表で対象とするのは、以下の波線部で示したような、文章・談話内における内容を表す名詞の反復使用（以下これを反復語と表記）である。

（1）産業技術が極めて幼稚なうちは、思い込みで適当なことをやっていても一見通用するように見える時代がある。しかし、どのような分野でも産業技術は次第に高度化する。その結果、学問に基づいた本物の技術しか通用しないということがいろいろな産業分野でいま始まっている。
（大見忠弘2004『復活！日本の半導体産業-未来を拓く志-実力を磨いて世の中の役に立とう！-』財界研究所、『BCCWJ』サンプルID:PB45_00024）

（2）R : (F ええ)で猿だったら(F えーとー)聞こえんのか聞こえないのかっていうので(F えーっとですねー)(F えー)何だろう例えばピアノの音とかだと
L : (F はい)
R : (F えー)色んな音の成分がたくさん入ってるんですね
L : (F はい)
R : んでその(D い)一番低い音の成分から(F そのー)高い音の成分まで色々入ってるんだけど （『CSJ』ID:D04M0056を見やすく整形）

† donguri-no-stability@hotmail.co.jp

（1）も（2）も、「産業」「技術」「音」という名詞が反復的に使用されることでそれを中心とした内容展開が構築されている。このような反復語の使用は、文章・談話中において文の意味的連鎖やコミュニケーション上の機能に関わるものとして、従来から話し言葉・書き言葉の双方において定性的な研究が行われてきた（中田 1991、田中 1997、塩澤 2005、馬場 2006 など）。ただ、そもそも反復語が文章・談話内でどの程度現れる存在なのかという定量的側面に注目した研究は管見の限り見られない。よって本発表ではコーパスを利用して話し言葉・書き言葉双方に見られる反復語の定量的分析を行う。

## ２．２ 調査資料

扱う資料は、CSJ および BCCWJ である。CSJ は話し言葉のデータであり、本発表では対話データを収録したデータと、学会講演と模擬講演という二つの独話データとを取り上げる。対話データは、厳密には講演のインタビュー・課題指向対話・自由対話が含まれるが、それぞれのファイル数の少なさから、一括して扱う。書き言葉は BCCWJ の中の白書・新聞・書籍・雑誌の各サブコーパスを用いる。なお、BCCWJ のデータは、詳細な情報が付与されているコアデータのみ取り上げる[1]。調査資料の大きさは表 1 に示した[2]。

**表 1：分析資料の概略**

| | 白書 | | 新聞 | | 書籍 | | 雑誌 | |
|---|---|---|---|---|---|---|---|---|
| | 延べ語数 | 異なり語数 | 延べ語数 | 異なり語数 | 延べ語数 | 異なり語数 | 延べ語数 | 異なり語数 |
| 最小値 | 216 | 123 | 127 | 97 | 126 | 64 | 134 | 102 |
| 第1四分位値 | 701.5 | 238 | 230 | 144 | 254 | 144.5 | 273 | 177.8 |
| 中央値 | 981.5 | 303 | 275 | 169 | 438 | 215 | 521.5 | 291 |
| 平均値 | 1182.1 | 339.1 | 305.1 | 184.7 | 548.5 | 262 | 598.8 | 307.3 |
| 第3四分位値 | 1594.2 | 441.2 | 342.2 | 206.2 | 722 | 367.5 | 834.5 | 381.5 |
| 最大値 | 2925 | 713 | 1203 | 571 | 1593 | 719 | 1954 | 876 |
| ファイル数 | 62 | | 340 | | 83 | | 86 | |

| | 学会講演 | | 模擬講演 | | 対話 | |
|---|---|---|---|---|---|---|
| | 延べ語数 | 異なり語数 | 延べ語数 | 異なり語数 | 延べ語数 | 異なり語数 |
| 最小値 | 119 | 82 | 193 | 97 | 176 | 82 |
| 第1四分位値 | 590 | 175 | 278.5 | 139.5 | 291.5 | 119.2 |
| 中央値 | 717 | 207 | 337 | 164 | 342 | 146.5 |
| 平均値 | 778 | 226.1 | 345 | 168.1 | 365.1 | 151.2 |
| 第3四分位値 | 882 | 256 | 403 | 195 | 404 | 171.8 |
| 最大値 | 3509 | 917 | 659 | 280 | 922 | 268 |
| ファイル数 | 987 | | 107 | | 58 | |

[1] コアデータにはその他にも Web 上のブログや知恵袋という質問サイトのデータも存在するが、データ内に存在する各ファイルが対象にできる大きさに至っていないため、本発表では扱わない。

[2] ここで単位となる「語」の認定基準は後述。また、同じく後述するが、本発表では 50 語ごとの反復語の使用率とそのばらつきを分析に取り入れるため、延べ語数 100 語未満のファイルは分析対象から外した。

CSJ と BCCWJ は、意思伝達の媒体が音声言語か書記言語かという大きな違いがある。これは話し言葉か書き言葉かという区別を行う上で最も大きな要素である。基本的には、その違いによって「話し言葉的」か「書き言葉的」かの性質が決定されると考えられる。また、複数の話し手が話す・聞くという関係を相互に行う点で、CSJ における対話は書き言葉と決定的に異なり、典型的な話し言葉として認められる。一方、白書のような公的機関による書き言葉は、話し言葉的要素をあまり導入しない点で書き言葉の典型と認めうる。

CSJ の学会講演や模擬講演は一人の話し手によってなされ、筋書きが大方決まっている点で書き言葉の音声化とも呼べる。したがって、この二つは対話に比べて書き言葉に近い存在である。また、音声言語と書記言語のそれぞれでも、改まりの程度に違いがあり、白書と学会講演は共に公的な場でのものであるという点で改まりの程度が高い。書籍と模擬講演は、必ずしも改まりの場として捉えなくて良いという点で白書や学会講演よりも改まりの程度が低い。雑誌はそれらに比べるとさらに改まりの程度の低いものであると言える。新聞は改まりの程度としては白書同様の高さであると考えられるが、多くの情報を限られた字数内に収めなければならないという字数制限による表現上の制約が強く、その点で他のサブコーパスに対して特異な性質を持つ。

以上から、本発表では基本的に音声言語・書記言語、独話型・対話型、改まり度の高低、字数制限という四つの基準を設けることができる。これを模式的に表したものが図 1 である[3]。図 1 より、話し言葉と書き言葉は種々の性質の中で連続性を持っていることが分かる。

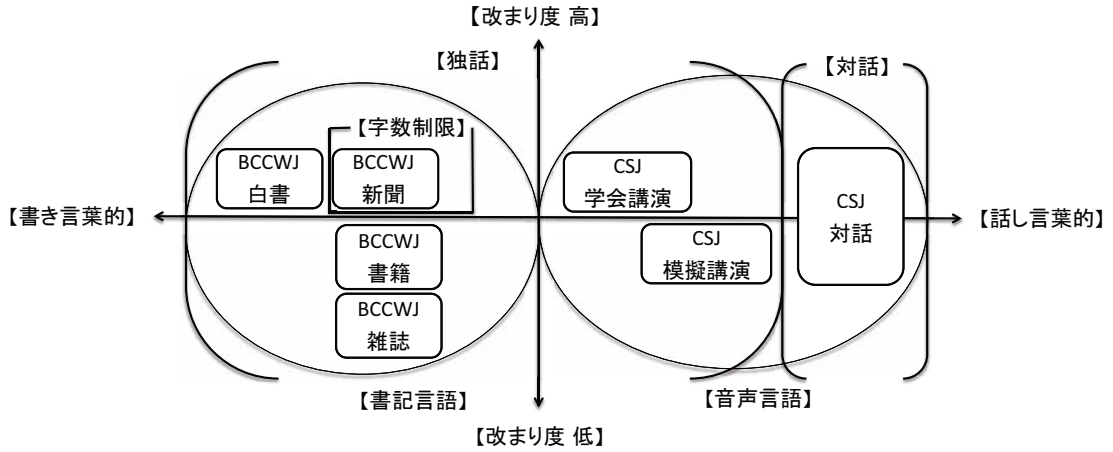


**図 1：資料間の関係性と連続性**

### ２．３ 調査方法

定量的分析に際しては、名詞 50 語あたりの反復語の平均使用率・使用率のばらつき、反復語 1 語あたりの平均使用頻度・平均反復間隔という四つの側面に注目した。本発表では各資料における計算結果の記述を行った上で、各資料間の計量結果の横断的に分析するこ

[3] 図 1 において学会講演が模擬講演よりも書き言葉的であるとしているが、これは典型的な書き言葉である白書と、改まりの度合いにおいて類似性が高いのが学会講演であると判断したことによる。

とで、反復語が文章・談話中においてどのように現れ、どのように使用されているのかを多角的に明らかにする。各資料の分析には、「語」相当の集計単位として「短単位」による形態論情報を設定した。CSJ と BCCWJ コアデータには形態論情報が付与されているため、それをそのまま利用した[4]。

## ３．調査結果

### ３．１　平均使用率

反復語が文章・談話中でどの程度現れるのかを明らかにするために、始めに名詞 50 語あたりの反復語の使用率の平均値を計算した。反復語は、延べ語数に対する 2 回目以降に使用された名詞の比率を求めれば良い。

本発表で 50 語ごとに計算したのは、文章の長さに影響を受けない形で対象ファイルにおける反復語の使用率を算出するためである[5]。この方法に基づいた対象資料ごとの各ファイルにおける反復語の平均使用率は表 2 に、その箱ひげ図を図 2 に示した[6]。

表 2：平均使用率(%)

| | 白書 | 新聞 | 書籍 | 雑誌 | 学会講演 | 模擬講演 | 対話 |
|---|---|---|---|---|---|---|---|
| 最小値 | 21.81 | 6 | 12.67 | 5 | 19.62 | 15.67 | 21.67 |
| 第1四分位値 | 26.93 | 16.4 | 21.02 | 16.34 | 31.83 | 26.33 | 31.8 |
| 中央値 | 30.65 | 19.38 | 25.6 | 19.59 | 35.87 | 29 | 34.2 |
| 平均値 | 31.1 | 19.74 | 25.56 | 20.07 | 35.7 | 29.44 | 35.19 |
| 第3四分位値 | 34.27 | 22.89 | 30 | 23.2 | 39.57 | 32.73 | 38.96 |
| 最大値 | 45.47 | 36 | 48.8 | 40.55 | 53.09 | 43 | 55.2 |

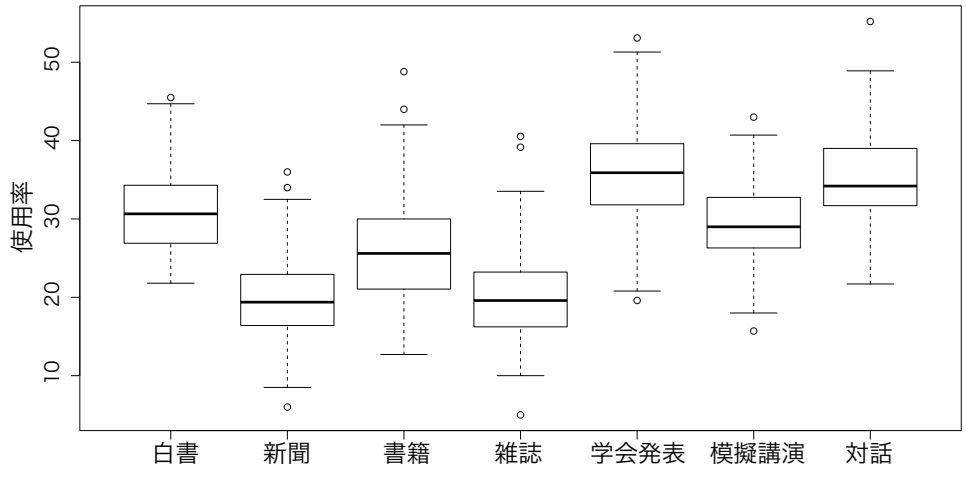


図 2：平均使用率(%)

[4] 短単位については小椋（2006）、小椋（2008）、小椋ほか（2011）を参照。なお、それを用いた集計や計算、作図はスクリプト言語 Perl と統計解析環境 R を利用している。

[5] 50 語ごとに集計した際の余りは切り捨てた。

[6] 図 2 の箱ひげ図は、箱内の線が中央値であり、下辺と上辺がそれぞれ第 1 四分位値と第 3 四分位値を指す。上下の丸は外れ値、上下の点線の先は外れ値を除いた場合の最大値と最小値である。

50 語あたりの反復語の使用率の平均は、書き言葉に比べて話し言葉の方が総じて高い値を示す。また、書き言葉内では、白書の方が書籍よりも使用率が高く、書籍は雑誌よりも使用率が高い。また、話し言葉のうち、学会講演が模擬講演や対話に比べて高い値を取る。この点で、反復語は、話し言葉的であり、かつ改まり度が高い場合において使用率が高まると言える。ただし、新聞の使用率が最も低くなっていることには注意が必要である。これは字数制限により同一語の反復を極力避けることと関係していると考えられる。

### ３．２　50 語ごとの使用率のばらつき

前節では各テキストで用いられる名詞 50 語ごとの反復語の使用率の平均を計ったが、50 語単位でくり返し計測しているため、計算ごとの結果にはばらつきが生じる。そこで、各データ内に含まれるファイルごとで、50 語ごとの使用率にどの程度のばらつきが見られるのか調査した結果を表 3 と図 3 に示す。なお本発表のばらつきは変動係数によって表した。

表 3：平均使用率のばらつき方

| | 白書 | 新聞 | 書籍 | 雑誌 | 学会講演 | 模擬講演 | 対話 |
|---|---|---|---|---|---|---|---|
| 最小値 | 0.144 | 0.055 | 0 | 0.065 | 0.046 | 0.056 | 0.032 |
| 第1四分位値 | 0.244 | 0.273 | 0.209 | 0.285 | 0.197 | 0.167 | 0.169 |
| 中央値 | 0.281 | 0.362 | 0.295 | 0.346 | 0.231 | 0.223 | 0.207 |
| 平均値 | 0.293 | 0.382 | 0.29 | 0.36 | 0.234 | 0.228 | 0.204 |
| 第3四分位値 | 0.338 | 0.467 | 0.357 | 0.417 | 0.269 | 0.282 | 0.243 |
| 最大値 | 0.481 | 1.324 | 0.609 | 1.006 | 0.434 | 0.487 | 0.362 |

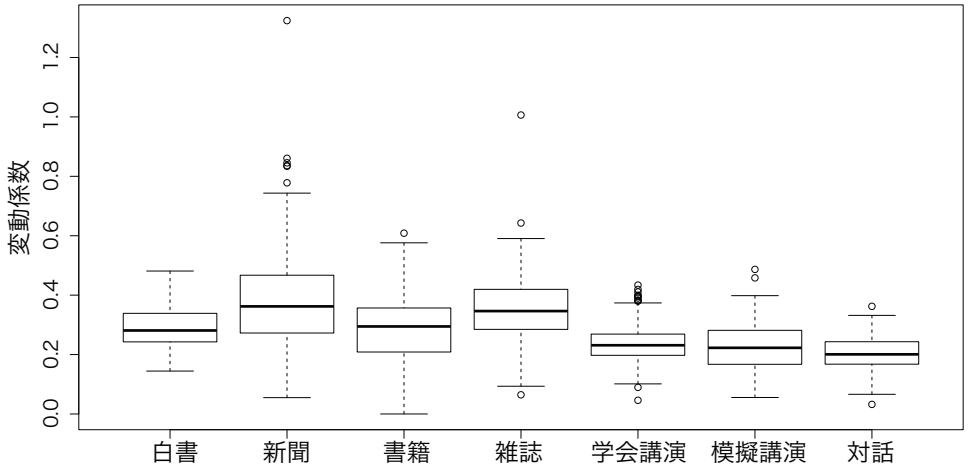


図 3：平均使用率のばらつき方

書き言葉のうち、白書を除く BCCWJ の各サブコーパスは最大値と最小値の差が大きく、ファイルごとの反復語の使用率のばらつきそのものが大きい。つまり、書き言葉には反復語の使用・不使用に関する一貫性が必ずしも見られず、量的に多様な様相を見せると考えられる。

また、変動係数の値の幅は、話し言葉の方が書き言葉よりも総じて小さい。したがって、

話し言葉の方が場面に拘らず反復語を用い、書き言葉の方が反復語を使ったり使わなかったりする傾向にあることが分かる。

### ３．３　１語あたりの平均反復頻度

反復語は 2 回目以降使用された語が全て含まれるため、反復語に含まれる見出し語自体は多様である。したがって、反復語の使用実態を記述するために、その一つひとつの語がどの程度反復されているのかを把握することも有意義であると考えられる。反復語に含まれる見出し語 1 語あたりの頻度を平均した結果が以下の表 4 と図 4 である。

**表 4：反復語 1 語あたりの平均反復頻度**

| | 白書 | 新聞 | 書籍 | 雑誌 | 学会講演 | 模擬講演 | 対話 |
|---|---|---|---|---|---|---|---|
| 最小値 | 3.77 | 2.35 | 2.9 | 2.36 | 2.83 | 2.99 | 3.24 |
| 第1四分位値 | 4.97 | 2.92 | 3.47 | 3.4 | 5.27 | 3.68 | 3.92 |
| 中央値 | 5.73 | 3.17 | 4.04 | 3.93 | 5.84 | 3.98 | 4.25 |
| 平均値 | 6.21 | 3.27 | 4.21 | 4.1 | 6 | 4.06 | 4.42 |
| 第3四分位値 | 7.1 | 3.51 | 4.85 | 4.46 | 6.58 | 4.35 | 4.69 |
| 最大値 | 12.68 | 9.04 | 6.47 | 9.82 | 11.9 | 5.37 | 7.18 |

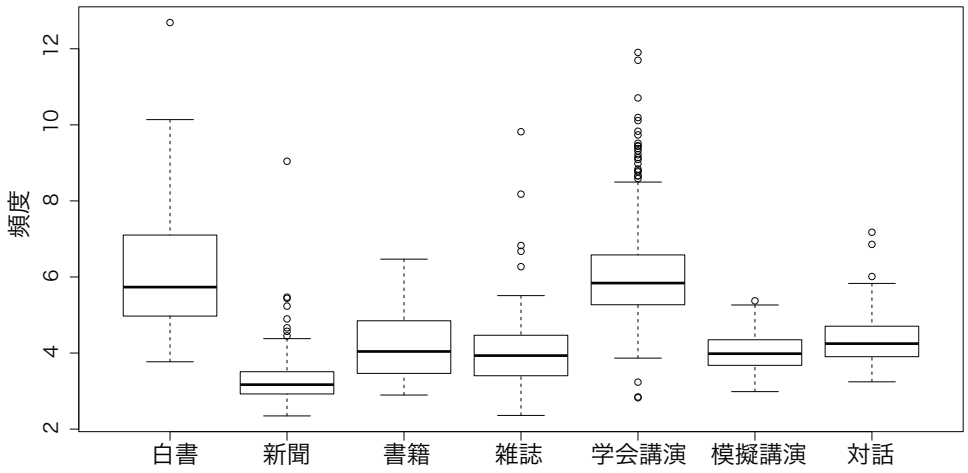


**図 4：反復語 1 語あたりの平均反復頻度**

学会講演に外れ値が多くあるが、全体としては学会講演と白書の二つにおいて反復語の頻度が高い。つまり、反復語の一つひとつの頻度は、書き言葉・話し言葉の別なく、改まりの程度が高くなると増加してくると考えられる。また、新聞が値の幅、値そのものの両方においてそれ以外のデータよりも小さくなっている。ここでも字数制限による反復語の使用の抑制が伺える。

### ３．４　１語あたりの平均反復間隔

反復語は文章・談話中で複数回用いられる語の集合である。そのため、反復語となった

語は、それぞれがある間隔を置いて改めて使用されており、前出の語と再使用された語との間には間隔が生まれる。反復語に位置づけられる語のそれぞれにおいて、どの程度の間隔で反復が行われているのかを調査し、その平均間隔を計算した結果が表 5 と図 5 ならびに表 6 と図 6 である。表 5・図 5 では名詞をもとにして距離を計測し、表 6・図 6 では品詞に拘らず全ての語によって距離を計測した。始めに、名詞の語数を利用した結果である表 5 と図 5 を示す。

**表 5：反復語 1 語あたりの平均反復間隔(名詞での間隔)**

| | 白書 | 新聞 | 書籍 | 雑誌 | 学会講演 | 模擬講演 | 対話 |
|---|---|---|---|---|---|---|---|
| 最小値 | 13.33 | 11.77 | 8.18 | 12.38 | 8.13 | 8.02 | 5.09 |
| 第1四分位値 | 28.09 | 22.12 | 18.77 | 25.27 | 17.87 | 14.5 | 12.84 |
| 中央値 | 37.88 | 27.31 | 29.96 | 38.03 | 21.63 | 18.32 | 15.73 |
| 平均値 | 38.39 | 28.44 | 32.31 | 40.34 | 23.43 | 18.96 | 15.53 |
| 第3四分位値 | 48.77 | 32.9 | 42.81 | 49.24 | 27.42 | 22.67 | 17.91 |
| 最大値 | 82.9 | 71.56 | 76.44 | 94.52 | 67.78 | 34.1 | 27.74 |

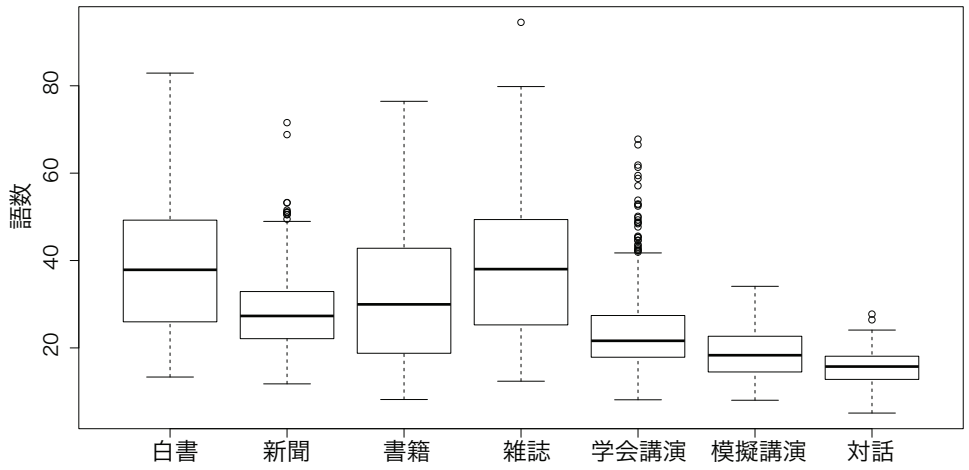


**図 5：反復語 1 語あたりの平均反復間隔(名詞での間隔)**

名詞間の測定では、BCCWJ と CSJ では、CSJ の方が明らかに間隔が短い。よって、書き言葉よりも話し言葉の方が短い間隔で語が反復されていると言える。また、話し言葉の中でも、話し手と聞き手の交替によって特徴付けられる対話は最も反復の間隔が短くなる。

一方、書き言葉の中で見た場合、特に最大値において新聞が他の書き言葉と明らかに異なった値を示している。他のデータとの比較から、字数制限による影響と考えられるが、その具体的な関係性については、内実をより詳細に観察する必要がある。

次に、全ての語を利用して反復の間隔を測定したものが以下に示した表 6・図 6 である。

表 6：反復語 1 語あたりの平均反復間隔(全語での間隔)

| | 白書 | 新聞 | 書籍 | 雑誌 | 学会講演 | 模擬講演 | 対話 |
|---|---|---|---|---|---|---|---|
| 最小値 | 37.22 | 30.52 | 34.47 | 48.47 | 26.21 | 46.18 | 29.87 |
| 第1四分位値 | 62.34 | 59.87 | 79.82 | 89.9 | 63.43 | 78.86 | 69.33 |
| 中央値 | 90.45 | 77.04 | 113.18 | 140.19 | 83.78 | 96.07 | 95.09 |
| 平均値 | 96.79 | 82.04 | 140.51 | 152.94 | 91.04 | 99.43 | 95.07 |
| 第3四分位値 | 123.89 | 100.19 | 203.26 | 191.78 | 110.66 | 116.51 | 115.69 |
| 最大値 | 216.15 | 187.66 | 368.78 | 379.94 | 317.15 | 169.95 | 163.32 |

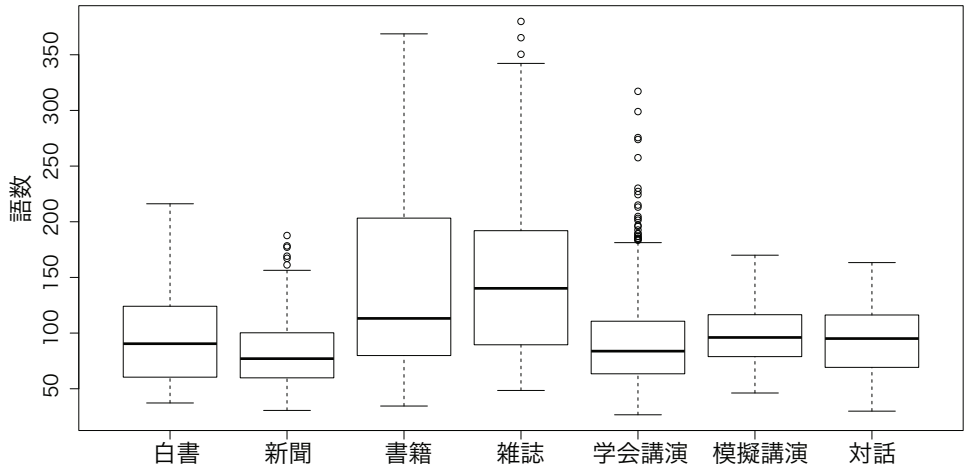


図 6：反復語 1 語あたりの平均反復間隔(全語での間隔)

表 6 と図 6 では、雑誌を除く各データの中央値が概ね 100 語前後の値となっている。つまり、同一の名詞を反復させる間隔そのものは、話し言葉・書き言葉に拘らず、基本的に 100 語前後が平均的であると考えられる。ただし、雑誌の中央値は 140 語と他に比べて間隔が大きい。また、全体としては、書籍や雑誌に見られるように書き言葉は最大値の値が大きくなる傾向にあり、反復語の使用間隔を大きくとることに対しての許容度が高いと考えられる。

新聞については、表 5・図 5 と同様、書き言葉の中では例外的に最大値が小さくなっており、字数制限の影響が存在すると考えられる。また、白書は名詞のみに注目した場合の間隔においては書籍や雑誌と大差がなかったが、全ての語を測定に利用すると、書籍・雑誌に比べて最大間隔が小さくなる。

話し言葉においては、最大値のみを見れば対話から白書にかけて間隔が増加していることが見て取れる。ここでも表 5・図 5 と同様に、典型的な話し言葉である対話においては、反復語の間隔を短くする傾向にあることが分かる。

## 4．考察とまとめ

本発表では、話し言葉のコーパスとして CSJ、書き言葉のコーパスとして BCCWJ を取り上げ、音声言語を基礎とする話し言葉と、書記言語を基礎とする書き言葉という二つの軸を設定するとともに、それぞれに含まれるサブコーパスを、独話と対話、改まり度の高低

といった諸特性によって分けた。そうした各種の枠組みを設定することを通して、「話し言葉的–書き言葉的」という構図を連続性を持った形で捉えられるようにした。その上で、反復語の使用実態から「話し言葉的–書き言葉的」という典型的な対比構図の様相を観察した。

反復語の使用実態から見た場合、そうした対比構図がよく表されているのは平均使用間隔（表 5 と 6・図 5 と 6）であると考えられる。基本的には、平均使用間隔は書き言葉において大きな値を取ることが可能で、話し言葉的になるほど反復の間隔が短くなる。なお、平均使用間隔については、測定に用いる単位を全ての語に広げると（表 6・図 6)、書き言葉・書き言葉に拘らず概ね 100 語程度が平均的な反復間隔となる。

「話し言葉的–書き言葉的」という観点での連続性は、反復語の使用の一定性・ばらつき（表 3・図 3）についてもある程度までは当てはまると考えられる。この場合、話し言葉的になるほど、反復語を一貫して用いるようになり、書き言葉においては、反復語を使用したりしなかったりといったばらつきが大きくなる。

反復語の使用率（表 2・図 2）についても、全体としては話し言葉の方が書き言葉よりも使用率が上がっており、同様の対比の中で捉えられる。しかし、反復語の使用率では、そうした対比に加えて、改まりの程度差が決定要因として関わっており、音声言語か書記言語かという典型的な話し言葉と書き言葉の対比構図の中だけでは捉えにくくなる。

反復語となる語の平均頻度（表 4・図 4）においては、話し言葉的・書き言葉的という連続性はあまり大きな意味を持たず、むしろ改まりの程度によってその量を変動させている。

また、全体を通して新聞サブコーパスが特異な値を示すことが多く、反復語においては、字数制限による使用制約が、諸要因の中でも強い力を持って作用することが多いと考えられる。

以上の影響関係を概略的に示せば、次の表 7 のようになる。

**表 7：反復語の使用実態に関わる諸要因**

| | 話し言葉–書き言葉の影響 | 改まりの高低の影響 | 対話–独話の影響 | 字数制限の影響 |
|---|---|---|---|---|
| 反復語の使用率 | 有り | 有り | 無し | 有り |
| | 話し言葉 ≧ 書き言葉 | 改まり高 ≧ 改まり低 | φ | 新聞 ≦ その他 |
| 反復語使用のばらつき | 有り | 無し | 無し | 無し |
| | 話し言葉 ≦ 書き言葉 | φ | φ | φ |
| 反復語1語あたりの頻度 | 無し | 有り | 無し | 有り |
| | φ | 改まり高 ≧ 改まり低 | φ | 新聞 ≦ その他 |
| 反復語の使用間隔 | 有り | 無し | 有り | 有り |
| | 話し言葉 ≦ 書き言葉 | φ | 対話 ≧ 独話 | 新聞 ≦ その他 |

また、この結果から、本発表では話し言葉と書き言葉の関係性について、以下の二つの事実を指摘できる。一つは、話し言葉と書き言葉は必ずしもある決定的要因によって性質を分離させているのではなく、典型的話し言葉から典型的書き言葉にかけて大きなグラデーションを描きながら変化させている点であり、これは量的構造から見たときの両者の連

続性を意味している。もう一つは、反復語の 1 語あたりの頻度や、字数制限を受ける新聞というサブコーパスの値において見られたように、状況次第では話し言葉と書き言葉という区分自体が有効に機能しなくなり、それ以外の要因、本発表であれば改まり度や字数制限のような要因によって表現の性質が決定されることがあるという点である。

なお、調査結果の違いはあくまでも量的差異による概略的なものであるため、それが具体的にどのような質的違いに基づいて現れたものなのかという点については、テキストの内実の観察を通した分析により達成されなければならない。この点は今後の課題としたい。

## 文　献

井上次夫（2011）「書き言葉らしさの判断と測定」『特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ（研究成果報告)』予稿集』, pp.89-96

小磯花絵・田中弥生・小木曽智信・近藤明日子（2011）「テキストの多様性をとらえる分類指標の構築を目指して」『特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ（研究成果報告）予稿集』, pp.431-442

小椋秀樹（2006）「第 3 章 形態論情報」『国立国語研究所報告 124 日本語話し言葉コーパスの構築法』, pp.133-186

小椋秀樹（2008）「『日本語話し言葉コーパス』の言語単位」『日本語学』27.5, pp.72-81

小椋秀樹・小磯花絵・冨士池優美・宮内佐夜香・小西光・原裕（2011）『『現代日本語書き言葉均衡コーパス』形態論情報規定集 第 4 版（上）（下)』国立国語研究所内部報告集 LR-CCG-10-05-01, LR-CCG-10-05-02（特定領域研究「日本語コーパス」研究成果報告 DVD 所収）

塩澤和子（2005)「コラムに観察されるくり返しの機能」『文藝言語研究 言語篇』47, pp.15-31

田中妙子（1997）「会話における〈くりかえし〉−テレビ番組を資料として−」『早稲田大学日本語研究教育センター紀要』9, pp.47-67

中田智子（1991）「会話にあらわれるくり返しの発話」『日本語学』10.10, pp52-62

馬場俊臣（2006）『日本語の文連接表現−指示・接続詞・反復−』おうふう

山崎誠（2010）「語の平均使用頻度に現れるテキストの特徴」『特定領域研究「日本語コーパス」平成 21 年度公開ワークショップ（研究成果報告会）予稿集』, pp.5-14

## 引用資料

「日本語話し言葉コーパス」（科学技術振興調整費開放的融合研究『話し言葉の言語的・パラ言語的構造の解明に基づく「話し言葉工学」の構築』）

「現代日本語書き言葉均衡コーパス・コアデータ」（特定領域研究「日本語コーパス」データ班，特定領域研究「日本語コーパス」研究成果報告 DVD 所収）

## 関連 URL

| | |
|---|---|
| The Perl Programming Language | http://www.perl.org/ |
| The R Project for Statistical Computing | http://www.r-project.org/ |