The University of Osaka
Institutional Knowledge Archive

Title	新聞社説の叙述系基本語彙
Author(s)	石井, 正彦
Citation	現代日本語研究. 2021, 13, p. 65-84
Version Type	VoR
URL	https://doi.org/10.18910/88322
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

新聞社説の叙述系基本語彙

A Quantitative Study on the Basic Vocabulary for the Description of Newspaper Editorials

石井　正彦
ISHII Masahiko

キーワード：基本語彙，書きことば，文章ジャンル，話題と叙述，語彙散布度

要　旨

書きことばの基本語彙を，文章の（話題面ではなく）叙述面を基幹的に支える語の集合と規定した上で，それら「叙述系の基本語彙」の機能や性質を具体的に把握するには，（多様な文章ジャンルを覆う基本語彙を一時に求めるのではなく）叙述の仕方が明確で安定している単一の文章ジャンルを対象とした調査を積み重ねる必要があることを論じ，その手始めとして，新聞社説の叙述系基本語彙の抽出を試みた。2010 年の『毎日新聞』の社説から 120 編の社説を抜き，それらを話題分野別の５区分に層別したコーパスに『Web 茶まめ』による語彙調査を施して，得られた 6300 余語のうち５層全てに出現する 512 語から，使用率及び均等度がそれぞれの中央値より大きい（接辞・数詞等を除く）110 語を候補として抽出した。これら候補語が真に叙述系基本語彙たり得ているかについては，続稿にて検討・報告する。

１．はじめに

計量語彙論でいう基本語彙，とりわけ書きことばのそれは，話題分野別に層別されたコーパスにおいて，より多くの層にわたってより均等に用いられる高頻度語（＝高使用率の語）の集合，すなわち，文章の話題面ではなく叙述面を基幹的に担う語の集合として求められるものと考えられる。

この場合，基本語彙の有用性を高めようとすると，より多様な文章ジャンルから構成される大規模なコーパスを対象に語彙調査を行い，さらに多くの層に

わたってより均等に用いられる高頻度語を求めていくことになるが，層別が多様になればなるほど,そこには話題分野の違いだけでなく叙述の仕方(叙述法)の違いも反映されるから，得られた高頻度語が真に叙述系基本語彙であるか否かを確認すること，言い換えれば，それらがなぜ多様な文章ジャンルにわたって広範に多用されるのか，その理由を各ジャンルの叙述面と関連付けて説明することが難しくなる。書きことばとして基本度のより高い語（＝異なる文章ジャンルに幅広く用いられる基本語）を求めようとすればするほど，それがなぜ基本語であるかを説明することが難しくなるという，矛盾した状況が生まれるのである。

この問題を解決するには，大規模な語彙調査などに比べて迂遠ではあるが，まずは叙述の仕方が明確で安定している単一の文章ジャンルを対象としてその叙述系基本語彙を特定するという作業を行い，それを多くの単一文章ジャンルについて積み重ねていくことによって，書きことばを広く覆う叙述系基本語彙に接近していく，という方法が有効であると考えられる。

本稿では，こうした問題意識から，上の作業の手始めとして，新聞の社説という単一の文章ジャンルを対象に，その叙述系基本語彙の取り出しを試みる。なお，紙幅の都合上，本稿では候補語の抽出までを行い，それらが真に社説の叙述系基本語彙であるか否かを確認する作業の結果については，続稿において報告することにしたい。

２．前提的議論

２．１．基本語彙

基本語彙については質的な規定と量的な規定とが可能であるが，計量語彙論は(1)のような量的な規定から「基本語彙」に接近しようとする[1)]。

(1) 様々の言語表現によく現われる（見出し）語を組として考えたものが基本語彙である。(中略)「よく現われる語」だという事は，基本語彙という集合Ｓに属する元である任意の見出し語ｕと，Ｓに属しない見出し語ｖとについて，前者の使用率が後者のそれより（概して）大きい事を含意する。(中略)「様々の言語表現に現われる語」だという事は，特殊な分野にだけよく使われるのでなく,(概して)広く満遍なく使われるよ

うな語である事を含意する。　（水谷 1964:7-8）

この「広く満遍なく使われる」度合いを「均等度」と呼ぶことにすると，量的に規定される基本語彙とは，要するに「使用率が大きく，かつ，均等度が大きい語の集合」ということになる。語の使用率と均等度は語彙調査によって知ることができるが，とりわけ均等度を求めるには層別[2)]を施した語彙調査が必要になる。これについては，次のような記述がわかりやすい。

(2) 基本語彙とは，使用率が大きく，しかも対象とする言語作品あるいは言語体系の中に幾つかの層を設けて考えることができる場合（たとえば雑誌であれば，実用記事・文芸作品・趣味など掲載する内容別の層を設け，また平安時代物語であれば作品別に層を設けることができる），できるだけ多くの層にわたって出現する語の集合をいう。（樺島 1980:345）

このように，いくつかに分けられた異なる層に（概して）均等に現れる（使用率の大きな）語があれば，それを基本語彙のメンバーとし得るわけであるが，ここには少し複雑な問題が潜んでいる。その一つは，層別はいろいろな観点から行うことができるから，どのような観点から層別するかによって語の均等度が変わる（＝基本語彙のメンバーが変わる）という問題であり，いま一つは，層別というものは基本的に層の間の違いに注目し，各層の特徴語を見出すために行うものである（＝層の違いに左右されず均等に現れる語の抽出をめざして行われるものではない）から，たとえ均等度の大きな語を得たとしても，それらの語の特徴や性質を積極的に規定することが難しいという問題である。

たとえば，国立国語研究所の「新聞三紙の語彙調査」（1966 年）では，話題分野別（政治，経済，社会，文化，スポーツ，家庭，娯楽などの 12 区分），文種別（ニュース，解説，社説・コラム，評論，実用読み物，探訪ルポなどの 17 区分），情報源記載形式別（一般無記名記事，通信社記事，冒頭記名記事，末尾記名記事などの 10 区分），記事内位置別（見出し，標題・欄名，リード，本文などの８区分）の４つの観点から層別が行われている（林 1969）が，語の均等度はこれら４種の層別ごとに求めることができるから，それぞれの結果から異なる４種の基本語彙が得られることになる。しかも，それぞれの基本語彙は，「話題の違いにかかわらず高使用率で使われる語」とか「文種の違いにかかわらず高使用率で使われる語」とかいうように，消極的にしか規定できない。

２．２．叙述系の基本語彙

上の第一の問題に対して，これまでの語彙調査，とりわけ書きことばの語彙調査のほとんどは，上でいう話題分野の違いによる層別にもとづいて基本語彙を求めようとしてきた。そこには，書きことばすなわち文章を組み立てる（使用率の大きな）語彙を，以下のように大きく二分してとらえる見方があるものと考えられる。

(3) グループ１　いろいろな分野で共通に使われる，文章表現の土台を形作る語

グループ２　その文章，あるいは文章群のテーマやムードに強く関わる語　（樺島 2004:63）

いま，文章というものを「話題について何事かを述べる＝叙述する」ものと考えれば，グループ１は文章の叙述（文章表現）の側面にかかわる語，グループ２は文章の話題（テーマやムード）の側面にかかわる語ととらえることができる。そして，この２種の語群がある程度分離するものと仮定するなら，話題分野の違いによる層別の場合は，グループ１は各層に共通して使われるために均等度が大きくなり，グループ２は話題によって使われるかどうかが異なるために均等度が小さくなる。一方，新聞調査の「文種」のように叙述のタイプや方法の違いによる層別の場合は，逆に，グループ１は叙述の違いに応じて使われるかどうかが異なるために均等度が小さくなり，グループ２は叙述の違いに関係なく使われ得るために均等度が大きくなる，ということが考えられる。

要するに，書きことばの基本語彙は，グループ１の「叙述系の基本語彙」である場合もグループ２の「話題系の基本語彙」である場合もあり得るのだが，これまでの語彙調査では前者を追究してきた，ということである。これは，書きことばについて，どのような叙述を行うことが多いかという問題と，何を話題とすることが多いかという問題の，どちらが重要であるかを考えれば，ごく自然な選択であろう。書きことばの基本語彙とは基本的には叙述系の基本語彙である，ということを確認しておく必要がある。

２．３．単一ジャンルの叙述系基本語彙

その上で，先の第二の問題については，叙述の仕方（叙述法）が定まってい

る単一の文章ジャンルを対象とすることによって，その叙述系基本語彙を積極的に規定すること，すなわち，取り出された個々の基本語が当該文章ジャンルのどのような叙述法をどのように担っているのかを説明することが可能になるものと考える。

上述したように，話題分野別の層別で均等度の大きな語を得ても，それらに対しては，特定の話題に偏って用いられることのない語（＝グループ２ではない）という消極的な規定しかできない。ただし，使用率の大きな語彙が叙述系のグループ１と話題系のグループ２とに分かれるのなら,それらは結果的に(グループ２ではない以上)叙述系のグループ１であろうと考えられるわけである。このことは大筋として正しいとしても，叙述系の基本語彙と標榜する以上は，それらが文章のどのような叙述法をどのように担うことによって基本語彙たり得ているのかを積極的に規定することが求められよう。

(2)で樺島(1980)が「対象とする言語作品あるいは言語体系の中に幾つかの層を設けて考えることができる場合」と述べるように，基本語彙は（少なくとも語彙調査が施せる程度の言語量をもっている必要があるが）様々な「使用範囲（range)」の語彙から取り出すことができる。いま仮に，使用範囲の異なる文章ジャンルとして，1)社説，2)新聞，3)書きことばの３種を設定し，それぞれから叙述系の基本語彙を取り出すことを考えてみよう。

まず，1)社説には，話題の違いにかかわらず，その本文の構成に「まず，その日のニュースなど時事的な話題からテーマを紹介し，次いで，そのテーマに関して世論を指導する立場からの見解を述べ，最後に，望ましい解決に向けて関係者や当局への要請や要望を述べる」（小宮 2011:221）という明確で安定した叙述の型があるため，話題の違いによる層別を施してもそこに叙述法の違いは関与せず，したがって，各層に共通して（より均等に）使われる語は各層に共通する叙述法を担うものである可能性が高い[3)]。

しかし，これが 2)新聞一般となると，社説は論説文，社会面記事は報道文，家庭面記事は説明文というように，紙面や記事の違いは話題だけではなく叙述面にも及ぶから，話題の違いによる層別を行ったとしてもそこには叙述法の違いも混入してしまい，各層に共通して（より均等に）使われる語を得たとしても，それらを同じ叙述法を担う語として積極的に規定することは難しくなる。

さらに，3)書きことば全体ということになると，たとえば国立国語研究所の『現代日本語書き言葉均衡コーパス』(BCCWJ，2011）のような，書籍・雑誌・新聞・白書・教科書・ベストセラーなど様々な文章ジャンルから構成されるコーパスの場合，それを話題別に層別したとしても，そこには様々な文章ジャンルの叙述法の違いがより大規模に混入してしまい，各層に共通して（より均等に）使われる語を得たとしても，それらが担っているはずの叙述法を見出すことはもっと難しくなるだろう。

このように，叙述の仕方が単純で型にはまっている単一の文章ジャンルの場合は,話題の違いによる層別に叙述法の違いがほとんどかかわってこないため,各層に共通して（より均等に）使われる語が当該文章ジャンルの叙述法を担っている可能性が高くなる。一方，複数ないし多数の異なる文章ジャンルから構成される，より高次（広範囲）の文章ジャンルになると，話題の違いによる層別に叙述法の違いも混入してしまうため，各層に共通して（より均等に）使われる語が担っているはずの叙述法を簡単に見出すことが難しくなる。

要するに，基本語彙を求めようとする使用範囲が広くなればなるほど，異なる文章ジャンルの異なる叙述法が混在してくるため，たとえ話題の違いによる層別を施し，各層に共通して（より均等に）使われる語を得（て，それらを叙述系基本語彙と標榜し）たとしても，それらが具体的にどのような叙述法を担っているかを特定することはより難しくなる，ということである。

もちろん，使用範囲が広くとも，使用率と均等度を基準として「量的な基本語彙」を取り出すことは技術的に可能である。そして，そのようにして取り出した基本語彙にも実用的な価値はある。たとえば言語教育の領域などでは，1)→2)→3)の順に実用的な価値は高まっていくはずである。しかし，基本語彙の語彙論的な研究においては，それらがなぜ基本語彙たり得ているのかを積極的に規定し，説明することがきわめて重要になる。そのためには，まずは，2)や3)よりも1)の，単一の文章ジャンルの叙述系基本語彙から検討を始める必要があるだろう。

３．本稿の目的

以上の議論を踏まえ，本稿では，書きことばの基本語彙研究の一環として，

新聞の社説に用いられる語彙から叙述系の基本語彙を取り出すことを試みる。

社説の語彙を対象とするのは，上述したように，社説が明確で安定した叙述の型をもつ単一の文章ジャンルであるためであり，また，後述するように，語の層別における均等度を測るためには各層のサイズ（語彙量）が等しいことが求められるが，社説は１編の長さがほぼ決まっていることから，こうした層のサイズ調整が容易であるためである。

４．調査対象とデータ

2010 年１年分の『毎日新聞』の社説から 120 編を抜き出し，その見出しと本文を語彙調査の対象とする。層別は話題分野別とし，〈国内〉〈国際〉〈外交〉〈経済〉〈社会〉の５層に分類する。分類にあたっては，橋本(2010:93-96）の，〈国内（国内の政治全般)〉〈国際（外国の情勢)〉〈経済（国内外の経済や金融政策)〉〈その他（身の回りの様々な事象)〉の４分類を参考に，〈その他〉を〈社会〉とし，また，〈国内〉に含められていた〈外交〉を独立させて５分類とした。各層のサイズ（語彙量）をほぼ等しくするために，あらかじめ 2010 年奇数月の社説 321 編を５分類した上で，最も少ない〈外交〉に合わせて各層 24 編の社説をランダムに抽出し，語彙調査用のコーパスとした。なお，以上の作業は，『ＣＤ－毎日新聞 2010 データ集』を（毎日新聞社の許諾を得て）利用し，行ったものである。

語彙調査（語彙表の作成）には，国立国語研究所『Web 茶まめ』（堤・小木曽 2015）による形態素解析機能と，マイクロソフト社製“Excel　2019”の表計算機能を用いた。各層の全文を『Web 茶まめ』によって「短単位」[4)]に解析し，その結果を“Excel”にインポートして，ピボットテーブル機能を使って語彙表にまとめた。後述する使用率と語彙散布度の計算も，この“Excel”上の語彙表の中で行った。語彙表作成にあたっては，各短単位の語彙素読み・語彙素・品詞大分類・語種がすべて一致するものを同語（一つの見出し語）と認め，また，品詞大分類が「助詞」「助動詞」「補助記号」であるものは除いた。語彙調査による各層の語彙量は，表１のとおり。

表１：『毎日新聞』社説各層の語彙量（2010 年 120 編，短単位）

	〈国内〉	〈国際〉	〈外交〉	〈経済〉	〈社会〉	計
延べ語数	9469	9284	9523	9091	9300	46667
異なり語数	2356	2467	2304	2352	2729	6308

５．均等度（語彙散布度）の測定

前述したように，量的に規定される基本語彙とは「使用率が大きく，かつ，均等度が大きい語の集合」であると規定できる。使用率と均等度のうち，その測定法が問題となるのは後者である。実は，ここまで「均等度」としてきた概念は，統計手法としてはそれと表裏の関係にある「散布度」（散らばり度）という形で計算されることが多い。すなわち，「均等度が大きい」とは「散布度が小さい」ということであり，逆に，「均等度が小さい」ということは「散布度が大きい」ということである。

均等度はこのように散布度の裏返しであるから，一般的な統計指標である範囲（レンジ），最大値最小値比，標準偏差，χ^2値などによって測ることもできるが，計量語彙論では，Juilland & Chang-Rodriguez(1964)や水谷(1964)をはじめとして独自の計算法が提案されてきた（これらを一般的な統計指標としての「散布度」と区別して「語彙散布度」と呼ぶことがある）。たとえば，Juilland & Chang-Rodriguez(1964)の均等度Ｄ（ただし，散布度の１の補数）は注目する語がとる各層の使用頻度の変動係数を層の数で割ったものを基本にしているし，水谷(1964)の散らばり度ＳＣは層間分散を最大層間分散で割ったもの（木村・山田 1994:5）である。他にもいくつかの計算法が提案されているが，いまだ定まった方法はないようである。

このように，語彙散布度については様々な考え方・計算法が提案されているが，本稿では，Gries(2008)が提案するＤＰ（deviation of proportions）という指標を採用する。Gries(2008:414-415)によれば，この指標は，(i)語彙項目の散布度を定量化することができ，(ii)コーパスの部分（層）のサイズが均等であるという不当な仮定に依存せず，(iii)感度が高すぎず低すぎず，(iv)統計的有意性の指標ではないため，仮説検定パラダイムの理論的問題を回避でき，(v)理論的には少なくとも０から１の範囲をとるという利点がある[5)]。

いま，表２のような語彙表があるとき，語WのＤＰは次のように求められる。すなわち，ある層 S_i での，語Wの構成比 f_i/F（実測比率）と語彙Ｖの構成比 t_i/T（期待比率）との偏差の絶対値 $|f_i/F-t_i/T|$ を求め（表３），それらの総和（表３の「計」）を 2 で割った値（$[\Sigma|f_i/F-t_i/T|]/2$）である。ＤＰの値は 0 以上 1 以下の範囲をとり，0 に近いほど散布度が小さく（＝均等度が大きく），1 に近いほど散布度が大きい（＝均等度が小さい）。このようにＤＰは，ある層で観測された語の割合と，その層が全体（コーパス）に占める割合の違い（絶対差）をベースにしているため，各層のサイズ（語彙量）が異なるデータでも扱うことができるとされる（本稿のデータは各層のサイズをほぼ等しくしている）。

表２：語彙表における語Wと語彙Ｖの頻度

語＼層	S_1	S_2	…	S_n	計
： ： W ： ：	f_1	f_2	…	f_n	： ： F ： ：
V	t_1	t_2	…	t_n	T

表３：語Wの比率偏差（絶対値）とその総和

語＼層	S_1	S_2	…	S_n	計
： ： W ： ：	$\lvert f_1/F-t_1/T\rvert$	$\lvert f_2/F-t_2/T\rvert$	…	$\lvert f_n/F-t_n/T\rvert$	$\Sigma\lvert f_i/F-t_i/T\rvert$

６．社説の叙述系基本語彙

６．１．候補語の抽出

以下，４節で作成した語彙表のすべての語（表１の異なり 6308 語）について，その使用率とＤＰを計算し，社説の叙述系基本語彙にふさわしい「使用率が大きく，かつ，均等度が大きい（＝ＤＰ値が小さい）語」を抽出する。

こうした場合，計量語彙論では，使用率と均等度を変数とする関数として「基

本度」(「使用度」「有用度」などとも）を設定し，語を基本度の大きさの順に一次元に並べて，その上位にあるものを基本語彙のメンバーとする，という方法を採ることが多い。たとえば，Juilland & Chang-Rodriguez(1964)は，使用頻度Ｆと均等度Ｄの積（Ｕ＝Ｆ×Ｄ）を語の一般的な使用度（Usage：Ｕ）とし，水谷(1964)は，使用率 P と散らばり度ＳＣを変数とする線形判別関数（Z=－6.0226＋1.5825logP－0.4181logSC[6]）を基本度としている。

ただ，このように基本度にまとめてしまうと，たとえば，均等度が小さくても使用率が極端に大きい語があれば，それら（話題系の基本語彙の可能性もあるもの）も基本度の上位に位置することになるから，叙述系の基本語彙だけを的確に取り出すことが難しくなってしまう。そこで，本稿では，研究がなお探索的な段階にあることを考慮して，

(4) 使用率と均等度についてそれぞれ基準となる値を決め，どちらの基準値をも上回る語を叙述系基本語彙の候補語とする

という単純な方法を採ることにする。ただし，候補語を配列する場合には，使用率より均等度を重視する（＝使用率を必要条件，均等度を十分条件とする）立場を採って，均等度の大きな語から降順に配列する。要するに，まずは一定以上の使用率をもつ語に絞り込み，それらについては（使用率の違いを問題とせず）均等度をそのまま基本度とみなすという考え方である（今回は，候補語の中の均等度の違いは問題としない)。

ただ，この方法でも，使用率と均等度それぞれの基準値をどのように決めるかという問題がある。これについても，本稿では，以下のような探索的な方法を採ることにした。

(5) ５層すべてに出現する（＝使用範囲が５である）語のうち，使用率と均等度がいずれもそれぞれの中央値を上回るものを叙述系基本語彙の候補語とする

抽出の対象を５層すべてに出現する語に限るのは，図１の散布図に示すように，語のＤＰと使用範囲（出現層数）とが逆相関の関係にあって，ＤＰ値の小さい（＝均等度の大きい）語は使用範囲５の語に限られてしまうからである。もっとも，叙述系基本語彙の規定からして，それに属する基本語がすべての層に出現する（出現しない層がない）のは当然のことであろう。

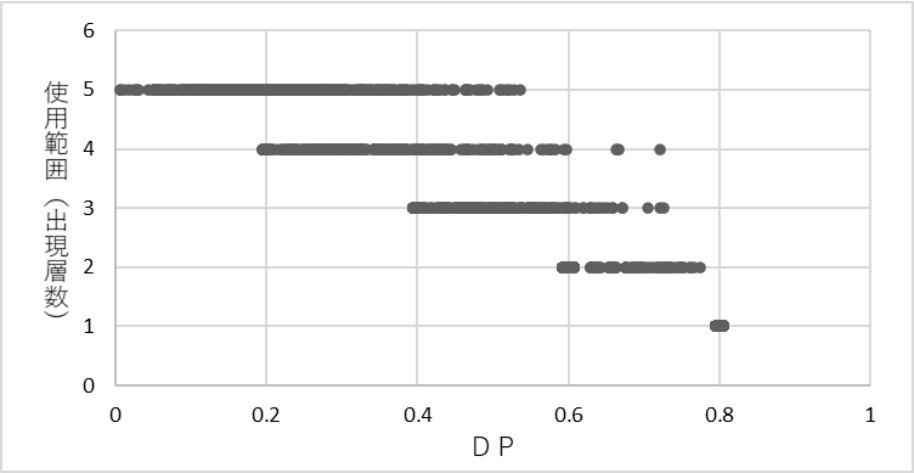

図１：社説語彙（6308 語）の使用範囲とＤＰの散布図

また，使用率・均等度とも中央値を基準とするのは，探索的な段階にあって，平均値よりも抵抗性（渡辺[他]1985:2）が高い（逸脱したデータの影響を受けにくい）中央値を採用することが適当と考えたからである。他の分位数ではなく中央値としたのは，最終的に得られる候補語が 100 語程度になることを見積もったからであり，積極的な理由はない。

以上を踏まえ，具体的には，

①５層すべてに現れる（＝使用範囲５の）512 語のうち，

②いずれも①の中央値である，使用率 0.5‰（＝全体度数 24）超，ＤＰ0.222 未満の 143 語から，

③接頭辞（「第」「不」「大（ダイ）」「約」），接尾辞（「化」「上（ジョウ）」「的」「内（ナイ）」「方（かた）」「後（ゴ）」「等（ら）」「力（リョク）」「性」「人（ニン）」「中（チュウ）」「つ」「者（シャ）」），数詞（「零」「一」「二」「五」「三」「八」「六」「七」「九」「万」「四」），名詞-助数詞可能（「月」「度」「年」「日」，ただし自立し得る「年間」「点」を除く），形状詞-助動詞語幹（「様（ヨウ）」）の計 33 語を除いた

110 語を，社説の叙述系基本語彙の候補とする。表４に，得られた候補語をＤＰの昇順（均等度の降順）に示す。

表４で，均等度の最も大きい（＝ＤＰ値の最も小さい）語は「社説」であった。これは，社説各編の見出しに使われる（＝本文にはほとんど使われない）語，つまり，基本的に社説１編につき１回だけ使われる語であり，そのために

表４：社説の叙述系基本語彙の候補語（使用率 0.5‰超かつＤＰ0.222 未満）

No.	語彙素	品詞	語種	全体度数	国内	国際	外交	経済	社会	ＤＰ	Ａ	Ｂ
1	社説	名詞	漢	120	24	24	24	24	24	0.007		
2	する	動詞	和	2546	515	482	549	495	505	0.012		○
3	出来る	動詞	和	128	25	25	28	24	26	0.019	○	
4	内容	名詞	漢	26	5	5	6	5	5	0.027		
5	有る	動詞	和	635	137	131	135	114	118	0.029	○	
6	言う	動詞	和	405	82	101	78	79	65	0.051	○	
7	今	名詞	和	47	12	9	9	9	8	0.052	○	
8	欲しい	形容詞	和	44	11	9	9	7	8	0.053		
9	責任	名詞	漢	62	13	9	14	13	13	0.054		
10	得る	動詞	和	64	14	13	15	13	9	0.059		○
11	当然	形状詞	漢	31	5	6	6	6	8	0.059		
12	取る	動詞	和	67	13	15	16	13	10	0.060		
13	ない	形容詞	和	344	58	76	61	67	82	0.061		○
14	来る	動詞	和	126	29	20	26	29	22	0.065		○
15	必要	名詞	漢	107	17	20	26	24	20	0.068		○
16	また	接続詞	和	40	10	6	9	7	8	0.069	○	
17	その	連体詞	和	182	46	36	26	39	35	0.069	○	
18	含める	動詞	和	34	6	7	6	6	9	0.072		
19	求める	動詞	和	68	15	14	16	15	8	0.082		
20	始まる	動詞	和	29	4	6	7	5	7	0.087		
21	深刻	形状詞	漢	27	6	6	6	6	3	0.088		
22	居る（いる）	動詞	和	669	99	110	138	165	157	0.089		○
23	それ	代名詞	和	97	21	20	11	25	20	0.091	○	
24	感	名詞	漢	33	8	6	7	8	4	0.095		
25	物	名詞	和	129	34	19	28	28	20	0.096	○	
26	はず	名詞	和	38	10	7	6	6	9	0.098		
27	因る	動詞	和	124	13	28	25	27	31	0.101	○	
28	結果	名詞	漢	50	10	6	10	9	15	0.101		
29	ため	名詞	和	126	29	23	35	20	19	0.101	○	
30	国	名詞	和	72	12	20	13	11	16	0.102		
31	続く	動詞	和	40	4	8	10	9	9	0.103		
32	事	名詞	和	450	64	70	116	103	97	0.104	○	
33	出る	動詞	和	37	9	7	4	8	9	0.106		
34	状況	名詞	漢	38	8	7	8	11	4	0.109		
35	理由	名詞	漢	27	7	4	7	4	5	0.112		
36	厳しい	形容詞	和	35	8	8	6	4	9	0.113		
37	十分	形状詞	漢	28	7	6	3	5	7	0.113	○	
38	行う	動詞	和	45	6	11	10	11	7	0.113		
39	付ける	動詞	和	32	6	8	7	3	8	0.116		○
40	目指す	動詞	和	37	9	9	8	8	3	0.118		
41	成る	動詞	和	377	60	59	63	114	81	0.123	○	
42	見る	動詞	和	79	7	20	21	16	15	0.124	○	
43	しかし	接続詞	和	99	19	18	24	28	10	0.126	○	
44	これ	代名詞	和	102	18	27	27	20	10	0.128	○	
45	主張	名詞	漢	41	10	8	12	4	7	0.130		
46	示す	動詞	和	62	18	10	7	12	15	0.130		
47	信頼	名詞	漢	26	3	4	7	6	6	0.133		
48	可能	形状詞	漢	56	8	12	17	12	7	0.134		
49	この	連体詞	和	113	29	32	20	10	22	0.138	○	
50	以上	名詞	漢	37	11	9	7	5	5	0.139	○	
51	対策	名詞	漢	43	8	4	8	12	11	0.141		
52	今後	名詞	漢	33	3	11	7	6	6	0.142		
53	考える	動詞	和	41	6	7	6	12	10	0.142		○

54	つながる	動詞	和	26	3	6	5	8	4	0.145		
55	今回	名詞	漢	68	7	15	13	11	22	0.146		
56	前	名詞	和	66	12	20	10	8	16	0.147	○	
57	理解	名詞	漢	27	9	2	6	5	5	0.149		
58	事態	名詞	漢	29	7	3	8	4	7	0.152		
59	重要	形状詞	漢	36	7	8	12	4	5	0.153		
60	年間	名詞	漢	25	7	4	7	3	4	0.153		
61	期待	名詞	漢	42	8	10	3	13	8	0.154		
62	対応	名詞	漢	77	10	9	24	16	18	0.155		
63	基本	名詞	漢	32	11	3	7	6	5	0.156		
64	続ける	動詞	和	40	6	13	7	9	5	0.156		
65	上	名詞	和	27	3	6	5	4	9	0.157		
66	持つ	動詞	和	45	8	16	6	6	9	0.157	○	
67	保障	名詞	漢	46	10	8	16	7	5	0.158		
68	影響	名詞	漢	38	8	5	12	9	4	0.161		
69	つく	動詞	和	109	32	11	30	15	21	0.162		○
70	昨年	名詞	漢	41	15	7	6	8	5	0.163		
71	事実	名詞	漢	30	6	9	8	3	4	0.164		
72	受ける	動詞	和	37	2	9	7	11	8	0.164		○
73	事業	名詞	漢	29	7	1	6	9	6	0.164		
74	居る（おる）	動詞	和	39	11	7	6	11	4	0.166		○
75	指摘	名詞	漢	53	11	4	16	8	14	0.167		
76	姿勢	名詞	漢	40	11	11	9	1	8	0.170		
77	避ける	動詞	和	27	6	8	7	2	4	0.172		
78	取り組む	動詞	和	26	5	3	9	3	6	0.174		
79	法	名詞	漢	58	8	9	8	12	21	0.175		
80	大きな	連体詞	和	31	3	10	7	7	4	0.176		
81	課題	名詞	漢	40	7	2	14	8	9	0.177		
82	こう	副詞	和	37	8	9	12	4	4	0.178		○
83	巡る	動詞	和	51	16	8	12	3	12	0.178		
84	地域	名詞	漢	56	13	6	20	6	11	0.182		
85	削減	名詞	漢	29	8	6	9	5	1	0.187		
86	そう	副詞	和	43	7	5	6	15	10	0.187		○
87	拡大	名詞	漢	38	9	4	9	12	4	0.188		
88	難しい	形容詞	和	27	8	1	5	5	8	0.190		
89	思う	動詞	和	34	8	8	4	3	11	0.193	○	
90	一方	名詞	漢	30	10	4	8	3	5	0.193		
91	動き	名詞	和	27	5	4	6	10	2	0.194		
92	過ぎる	動詞	和	27	6	4	4	10	3	0.195		
93	点	名詞	漢	43	14	5	12	6	6	0.198		○
94	社会	名詞	漢	70	9	12	11	10	28	0.201		
95	行く	動詞	和	71	19	4	17	21	10	0.201	○	
96	対する	動詞	混	75	12	12	25	20	6	0.201	○	
97	機関	名詞	漢	35	6	12	2	6	9	0.202		
98	評価	名詞	漢	35	5	8	8	12	2	0.202		
99	強い	形容詞	和	56	7	15	16	14	4	0.206	○	
100	説明	名詞	漢	31	12	4	7	4	4	0.206		
101	新た	形状詞	和	36	5	8	14	6	3	0.208		
102	中	名詞	和	56	16	4	7	18	11	0.209	○	
103	同時	名詞	漢	31	4	11	8	5	3	0.210		
104	今年	名詞	和	26	1	7	9	5	4	0.212		
105	活動	名詞	漢	27	1	8	8	4	6	0.213		
106	時代	名詞	漢	42	12	10	4	12	4	0.213		
107	方針	名詞	漢	38	12	5	10	9	2	0.214		
108	負担	名詞	漢	45	16	2	10	6	11	0.216		
109	発表	名詞	漢	26	5	3	8	8	2	0.217		
110	反対	名詞	漢	33	14	6	6	6	1	0.221		

使用頻度が５層すべてで（各層の標本数と同じ）24となり，各層での使用率がほぼ等しくなってＤＰの値が最も小さくなったものである。ＤＰが均等度を測る指標として有効に働いていることを示す例と言えよう。

ＤＰの有効性を確認するために，均等度が小さい（＝ＤＰ値が大きい）語も見ておこう。いま，上と同様に「使用範囲５の512語のうちで使用率が中央値0.5‰より大きい語」から，ＤＰが0.3超の43語（接頭辞，接尾辞，助数詞を除く）をとってみると，表５のようになる。

表５：均等度の小さい語（使用率0.5‰超かつＤＰ0.3超）

No.	語彙素	品詞	語種	全体度数	国内	国際	外交	経済	社会	ＤＰ
1	中国	名詞	固	130	2	32	90	5	1	0.535
2	国会	名詞	漢	78	57	1	6	9	5	0.528
3	予算	名詞	漢	36	26	2	4	2	2	0.519
4	企業	名詞	漢	68	4	2	3	48	11	0.511
5	再建	名詞	漢	27	4	2	1	19	1	0.509
6	成長	名詞	漢	48	5	6	1	33	3	0.493
7	ハトヤマ	名詞	固	66	36	2	23	2	3	0.487
8	人	名詞	和	47	4	1	1	10	31	0.478
9	首相	名詞	漢	155	80	6	56	11	2	0.470
10	太平（～洋）	名詞	固	25	1	3	16	4	1	0.436
11	会議	名詞	漢	66	5	37	18	3	3	0.430
12	米国	名詞	固	123	2	77	22	19	3	0.427
13	金融	名詞	漢	71	11	17	1	41	1	0.423
14	年度	名詞	漢	24	15	2	1	4	2	0.422
15	案	名詞	漢	60	37	1	10	11	1	0.414
16	政策	名詞	漢	85	34	8	6	34	3	0.402
17	アジア	名詞	固	33	2	5	20	5	1	0.402
18	自由	名詞	漢	30	1	3	16	8	2	0.401
19	財政	名詞	漢	58	17	4	6	29	2	0.395
20	地	名詞	漢	24	4	3	1	14	2	0.389
21	代表	名詞	漢	42	23	10	3	1	5	0.384
22	世界	名詞	漢	97	3	44	10	31	9	0.379
23	回復	名詞	漢	28	3	3	5	16	1	0.377
24	判断	名詞	漢	33	2	3	5	4	19	0.376
25	仕舞う	動詞	和	27	5	2	1	15	4	0.361
26	管理	名詞	漢	38	7	1	9	1	20	0.360
27	資金	名詞	漢	41	10	1	2	21	7	0.358
28	計画	名詞	漢	32	1	11	4	13	3	0.356
29	消費	名詞	漢	44	6	1	8	24	5	0.351
30	共同	名詞	漢	38	7	6	21	3	1	0.349
31	政治	名詞	漢	89	35	14	3	31	6	0.344
32	発展	名詞	漢	24	2	4	13	1	4	0.338
33	安全	名詞	漢	41	3	11	19	2	6	0.329
34	経済	名詞	漢	121	13	15	25	63	5	0.328
35	実現	名詞	漢	36	19	7	4	5	1	0.325
36	安定	名詞	漢	38	2	9	8	18	1	0.323
37	形	名詞	和	25	10	1	3	8	3	0.322
38	認める	動詞	和	29	2	4	7	2	14	0.321
39	強化	名詞	漢	35	2	5	13	12	3	0.315
40	維持	名詞	漢	29	1	8	2	6	12	0.304
41	制度	名詞	漢	60	26	3	3	12	16	0.303
42	民主	名詞	漢	98	43	26	7	14	8	0.302
43	関する	動詞	混	27	1	6	13	3	4	0.301

これを見ると，多くが特定の話題分野に偏って使われる語であることがわかる。ＤＰ値が最も大きい（＝均等度が最も小さい）「中国」は〈外交〉〈国際〉に，それに続いて大きい「国会」「予算」は〈国内〉に，それぞれ大きく偏っている。一方，和語名詞の「人」(No.8) は〈社会〉に，和語動詞の「仕舞う」(No.25) は〈経済〉に偏っている。これらは叙述系の語とも考えられるが，社説全体ではなく特定の層に偏ってその叙述に多用されている可能性がある。なお，表５の 43 語はあくまで使用範囲 5 の語であり，ここに使用範囲 4 以下の語は含まれていないから，これらが均等度の小さい語の全体を代表するわけではない。

６．２．候補語の構成（概観）

２節および３節で，社説は明確で安定した叙述の型をもつ単一の文章ジャンルであると述べた。上で抽出した 110 語は，そうした社説の叙述法を支えるという点で共通する機能・性質をもっていると予想される。ただし，社説の叙述法というものは，社説独自のものだけで構成されているわけではなく，社説が所属するより上位の文章ジャンル，たとえば，広くは（話しことばに対する）書きことばの叙述法，狭くは論説文や意見文などの叙述法といったものと重なるところをもつはずである。したがって，社説の叙述系基本語彙にも，社説独自の，あるいは，社説により特徴的なものと，書きことばや論説文・意見文などと共通するものとが混在している可能性がある。いま，そのことについて詳しい検討を行うことはできないが，以下では，それにつながる予備的な検討を行ってみたい。

表４の右端の「Ａ」列は，林(1971)が，以下の国立国語研究所の語彙調査４種のすべてで頻度順上位 500 位以内にあることから「基幹度最高の語」とした 48 語中にあるもの，同じく「Ｂ」列は，４種のうち１）を除く３種において 500 位以内にある「現代だけの３種間のまたがり」56 語の中にあるものに，それぞれ〇印を付けたものである。

１）明治初期文献の語彙調査（明治 10～20 年）

２）婦人雑誌の語彙調査（昭和 25 年）

３）総合雑誌の語彙調査（昭和 28～29 年）

４）新聞の語彙調査（昭和 41 年）

これら，A列25語，B列14語の計39語は，時代を超え，ジャンルを超えて幅広く多用される「書きことばの叙述系基本語彙」に所属している可能性がある。もちろん，林(1971)の分類は，林自身がことわっているように，調査単位も異なる語彙調査を対象に「基幹語彙を求める手続きをたどってみ」る試行であり（林1971:11），確定的なものではない。そのことを踏まえた上で，これら39語をⅠ類，残りの71語をⅡ類とし，さらに，それらを110語における均等度の中央値（ＤＰ＝0.147）を境にして上位半数に位置するものと下位半数に位置するものとにそれぞれ二分して,『分類語彙表』(国立国語研究所2004）の「中項目」別に整理したのが表６である。

表６：社説の叙述系基本語彙（候補語）の構成

『分類語彙表』中項目	Ⅰ類		Ⅱ類	
	均等度上位	均等度下位	均等度上位	均等度下位
1.10　事柄	(25)物 (32)事 (23)それ (44)これ			(58)事態 (71)事実
1.11　類	(29)ため	(102)中	(28)結果 (35)理由 (26)はず	(63)基本 (68)影響 (62)対応 (110)反対 (65)上
1.13　様相			(34)状況 (4)内容	
1.15　作用				(105)活動 (85)削減 (87)拡大 (91)動き
1.16　時間	(7)今	(56)前	(55)今回 (52)今後	(103)同時 (60)年間 (106)時代 (70)昨年 (104)今年
1.17　空間		(93)点		(84)地域
1.18　形				(76)姿勢
1.19　量	(50)以上			
1.25　公私			(30)国	
1.26　社会				(94)社会
1.27　機関				(97)機関

1.30　心			(24)感 (47)信頼 (9)責任 (1)社説 (45)主張 (51)対策	(61)期待 (57)理解 (98)評価 (107)方針 (81)課題 (79)法 (75)指摘
1.31　言語				(100)説明 (109)発表
1.34　行為				(108)負担
1.35　交わり				(67)保障
1.37　経済	(15)必要			
1.38　事業				(73)事業
2.11　類	(27)因る	(96)対する	(54)つながる (18)含める	
2.12　存在	(5)有る (22)居る (3)出来る (39)付ける (41)成る	(74)居る	(46)示す (33)出る (12)取る	
2.15　作用	(14)来る	(72)受ける (95)行く (69)つく	(20)始まる (31)続く	(64)続ける (83)巡る (92)過ぎる (77)避ける
2.30　心	(53)考える (42)見る	(89)思う	(19)求める (40)目指す	(78)取り組む
2.31　言語	(6)言う			
2.34　行為	(2)する		(38)行う	
2.37　経済	(10)得る	(66)持つ		
3.10　真偽	(17)その (49)この	(82)こう (86)そう		(59)重要
3.11　類			(11)当然	
3.12　存在	(13)ない			
3.13　様相			(21)深刻 (48)可能 (36)厳しい	(88)難しい
3.14　力		(99)強い		
3.16　時間				(101)新た
3.19　量	(37)十分			(80)大きな
3.30　心			(8)欲しい	
4.11　接続	(16)また (43)しかし			(90)一方
計	26	13	29	42

あくまで一つの試みに過ぎないが，Ⅰ類は社説以外の書きことばでも叙述系の基本語として働いている可能性のある語，Ⅱ類は（社説独自とは言えないかもしれないが少なくとも）社説に特徴的な叙述系の基本語彙である可能性がある。Ⅰ類に和語や指示詞の類が多いことなどはこの予想に適うものであるが，注目されるのは意味分野（中項目）における両者の分布の違いである。動詞などに大きな違いはないが，名詞にはⅡ類が多く，なかでも「1.11 類」「1.13 様相」「1.15 作用」「1.16 時間」「1.30 心」「1.31 言語」はほぼⅡ類で占められている。また，同じ中項目でも，Ⅰ類とⅡ類とで語の性質が異なるものも注目される。たとえば，「1.10 事柄」は，Ⅰ類が「物」「事」という和語形式名詞，Ⅱ類が「事態」「事実」という漢語抽象名詞だが，後者の方が前者より意味範囲が狭い。「2.30 心」は，Ⅰ類が「考える」「見る」「思う」，Ⅱ類が「求める」「目指す」「取り組む」で，前者が思考や判断に関する動詞，後者が指向や意志に関する動詞であるという違いがある。このようにⅡ類の語は社説に特徴的な叙述を支えるために働いている可能性があるが，それらの候補語が真に社説の叙述系基本語彙であるか否かを確認する作業は，続稿において行うことにしたい。

注

1）「基本語彙」の質的な規定には，たとえば森岡(1977:215)の次のような記述がある。

> 現在，基本語彙というのは，使用度数によってきめられがちであるが，語の構造や性質から考察することがより重要ではないかと思う。すなわち，これらの一次名が基本語彙の資格を備えていると考えるのであるが，その理由は，
>
> ⑴一語基からなりそれ以下の単位に分解できない。
>
> ⑵知識体系の第一次的枠組みを構成している。
>
> ⑶造語機能が高く，これを基準にして多くの語を生産する。
>
> ⑷長い歴史を通して受継がれた語彙で新たに造語することができない。
>
> などであって，一語基の和語名は特にその性質を備えている。

2）「層別」とは，本来，統計調査の標本抽出において対象とする母集団をいくつかの層に分けることを意味する専門用語であるが，ここでは，語彙調査の

対象となる文章群をいくつかのカテゴリー（層）に分類するという一般的な意味で用いている。

３）橋本(2010:110-113) は,「数多くの社説に出現する語」を,「社説の話題を指し示す」「時事用語」と,「社説における『説明』や『論述』に関わる部分で用いられる」「論説用語」とに二分できるとする。前者はグループ２の話題系基本語彙，後者はグループ１の叙述系基本語彙と重なるものであろう。

４）国立国語研究所が『現代日本語書き言葉均衡コーパス』(BCCWJ) に用いた調査単位。「新聞三紙の語彙調査」の「短単位」とは異なるので注意を要する。基本語彙を求めようとする場合，調査単位として形態素相当の短い単位と単語（自立語）相当の長い単位のどちらが適当かという問題については議論の余地があるが，今回は大量のコーパスデータを安定的に解析できる『Web 茶まめ』の利用を優先し,「短単位」を採用することとした。なお，同システム（unidic-mecab 1.3.12 ）の新聞を対象とした解析精度は 98～99％以上とされている（国立国語研究所コーパス開発センター(2011:65)）。

５）Egbert, Burch & Biber (2020)は，語彙散布度の測定と解釈が言語的に意味のある層別にもとづいて行われるべきであると主張する中で，Gries(2008)のＤＰを各層の大きさが不揃いのコーパスに対する有効な測定法の一つであると評価している。また，ＤＰを日本語に適用した研究としては，日本語教育のための読解基本語彙の選定に用いた本田(2019)がある。

６）この展開式は，木村・山田(1994:5)による。

引用文献

樺島忠夫(1980)「語彙」国語学会編『国語学大辞典』東京堂出版，pp.344-346

樺島忠夫(2004)『日本語探検――過去から未来へ』角川書店（角川選書 361）

木村睦子・山田雅一(1994)「基本度関数について」『国立国語研究所報告 107　研究報告集 15』秀英出版，pp.1-34

国立国語研究所(2004)『国立国語研究所資料集 14　分類語彙表－増補改訂版－』大日本図書

国立国語研究所コーパス開発センター(2011)『現代日本語書き言葉均衡コーパス』マニュアル，国立国語研究所コーパス開発センター

小宮千鶴子(2011)「新聞の文体」中村明[他]編『日本語文章・文体・表現事典』朝倉書店，pp.218-224

堤 智昭・小木曽智信（2015)「歴史的資料を対象とした複数のUniDic辞書による形態素解析支援ツール『Web 茶まめ』」『じんもんこん 2015 論文集』，pp.179-184

橋本和佳(2010)『現代日本語における外来語の量的推移に関する研究』ひつじ書房

林　四郎(1969)「新聞語彙調査における層別とその意味」国立国語研究所報34『電子計算機による国語研究　Ⅱ』秀英出版，pp.1-15

林　四郎(1971)「語彙調査と基本語彙」国立国語研究所報告39『電子計算機による国語研究　Ⅲ』秀英出版，pp.1-35

本田ゆかり(2019)「コーパスに基づく『読解基本語彙１万語』の選定」『日本語教育』172，pp.118-133

水谷静夫(1964)「語の基本度」国立国語研究所報告25『現代雑誌九十種の用語用字　第三分冊　分析』秀英出版，pp.7-51

森岡健二(1977)「命名論」大野晋・柴田武[編]『岩波講座日本語２　言語生活』岩波書店，pp.203-248

渡部洋・鈴木規夫・山田文康・大塚雄作(1985)『探索的データ解析入門－データの構造を探る－』朝倉書店

Egbert, J., Burch, B., and Biber, D. (2020). Lexical dispersion and corpus design. *International Journal of Corpus Linguistics*, 25(1), 89-115.

Gries, S. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403-437.

Juilland, A. and Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words,* The Hague: Mouton.

付記　本稿は，JSPS科研費 JP18K00612の助成を受けたものです。

（文学研究科教授）