The University of Osaka
Institutional Knowledge Archive

Title	大規模な電子資料に見る現代日本語の動態
Author(s)	田野村, 忠温
Citation	待兼山論叢. 文化動態論篇. 2008, 42, p. 55-77
Version Type	VoR
URL	https://hdl.handle.net/11094/9791
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

大規模な電子資料に見る現代日本語の動態

田野村　忠温

１．はじめに

電子媒体の言語資料すなわちコーパスに基づく日本語研究をめぐる状況はこの数年来急速な展開を見せている。そのことについては拙論（2009）であらましを述べたが、利用可能なコーパスの面では、言語研究のために設計された狭義コーパスの開発・公開が進むとともに、それぞれに固有の特徴を有するいくつかの大規模な広義コーパスの利用が可能になってきた。

コーパスの単純な利用法は、所与のコーパスをそのまま単一の資料として扱うというものである。すなわち、コーパス内部の異質性には目を向けず、むしろ異質性を含む全体をそのまま利用することでそこにコーパスの代表性を期待する用法である。

それに対して、コーパスの“区分対照用法”とでも呼ぶべき利用法が考えられる。それは、コーパスの内部を何らかの基準に基づいて区分して扱い、それぞれの区分されたデータから得られる情報を比較対照することにより、区分間の差異や関係を明らかにしようとするものである。コーパスを区分する基準としては、ジャンルやレジスター、書かれた（話された）時期、書き手（話し手）の属性など、さまざまなものが考えられる。コーパスの規模が小さければそれを区分すると各部分はもはや統計的に信頼し得る情報を提供しなくなるが、大規模なコーパスは区分された部分もそれぞれに相当の規模を有し得るため、独立した資料として使用に耐えることが期待される。

従来、日本語研究におけるコーパスの利用は非区分的な用法が中心的であったが、このたびコーパスの区分対照用法の試みとして、２種類の大規

模な日本語の電子資料（広義コーパス）――国会会議録のデータと新聞記事データ―― を時間の要素に基づいて区分して用いることにより、現代日本語の変化や変動の様相を明らかにする可能性を探ってみた。本小論はその試行結果の報告である。

２．現代日本語の文法変化 ――国会会議録データに基づく分析

日本語文法の通時的研究と言えば、通常、一世紀を超えるような時間幅における文法の変化を対象とするものであった。遠い過去の日本語の場合、利用可能な資料が限定されていることから、短期間の文法の変化を精密に観察することはそもそも不可能である。現代日本語の場合には出版物を中心とする膨大な資料があるが、文法の細かい動きを確認するためには大規模な統計的調査が必要であり、旧来の紙媒体の資料に基づいてそれを行うことは現実的に困難であった。

ここでは、最近利用可能になった国会会議録のデータを利用して、現代日本語における文法の変化の様相を探ってみる。国会会議録は 1947 年以来の国会の本会議・各種委員会の議事を文字化した資料で、分量は 2008 年 10 月現在で約 7 ギガ（＝ 70 億）バイト、文字数にして約 35 億字である。これを 1940 年代から始めて 10 年ごとに区分して利用する（1940 年代と 2000 年代は 10 年分に満たない）。

国会の場で用いられる日本語はいかにも古臭く形式ばった表現に満ちあふれているという印象があるが、そのようなものが現代日本語の通時的な変化を明らかにする資料として果たして役に立つだろうか？[1)]

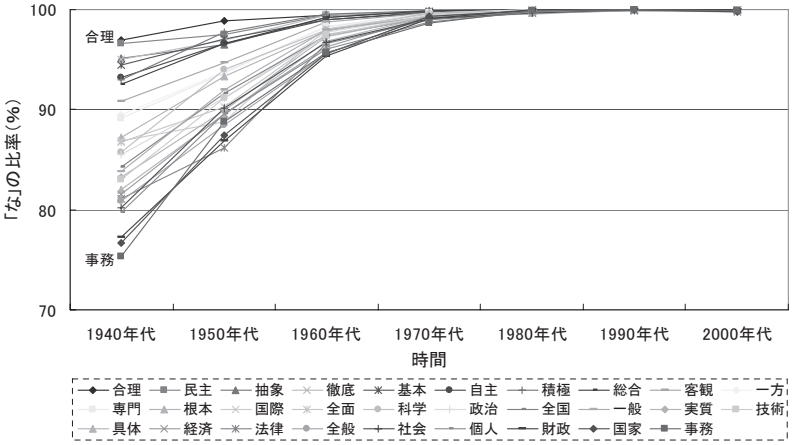
２．１　形容動詞類の連体形における「な」と「の」の選択

日本語のコピュラ述語におけるコピュラの異形態の分布はさまざまな条件に依存し、しかも、ゆれの要素を含む複雑な問題である（拙論（2006））。

ここでは、国会会議録データを使って「～な」と「～の」という形をした言い回しの使い分けの状況の一端を観察するが、この「な」と「の」の関係も最終的にはコピュラ異形態の分布という広い文脈に位置付けられるべきものである。なお、「有名な作家」と「無名の作家」という対比に見るような「有○」「無○」という対を成す形容動詞における「な」と「の」の使い分けを支配している意味的な要因については、新聞記事データを用いて若干の考察を行ったことがある（拙論(2002)）。

まず、単純な事例から見る。（図１）は、「～的」という形の表現が連体的に使われるときの「な」と「の」の関係について調べた結果である。「～的な」と「～的の」の用例全体において「～的な」の用例の占める比率の変化を、10年単位のいずれの年代においても両形の用例数の合計が500件以上あるものについてのみ示している（用例数の合計の多いものは約40,000例ある）。図が小さく、モノクロ印刷ということもあってグラフの各線が見分けづらいがご容赦願いたい。[2)]

（図１）「～的な／～的の」から「～的な」への統一



「な」の比率が時間の経過とともに高まっていき、1980年代までには「な」への移行・統一が完了したことを容易に見て取ることができる。また、1930年代以前においては「の」の比率がさらに高かったと推定して間違いないであろう。

この図は、「な」と「の」の使い分けという個別の問題を超えて、文法の変化に関わる一般的な事実を示唆している。それは、

(1) 一群の形式に関わる文法事象が変化するとき、しばしば、個々の形式や言語的文脈による変化の遅速の差が見られる。

ということである。これが多くの文法変化に通用する一般化であることは、以下で見るほかの事例からも確かめられる。目下の事例に即して言えば、「～的」の形の表現全体としては1940年代から1980年代にかけて「な」が「の」に取って代わる変化が進行したと言えるわけであるが、「合理的」「民主的」「抽象的」「徹底的」のように1940年代に「な」の比率がすでに95％を超えていたものもあれば、「個人的」「財政的」「国家的」「事務的」のように同時期には「な」の比率が80％未満だったものもある。

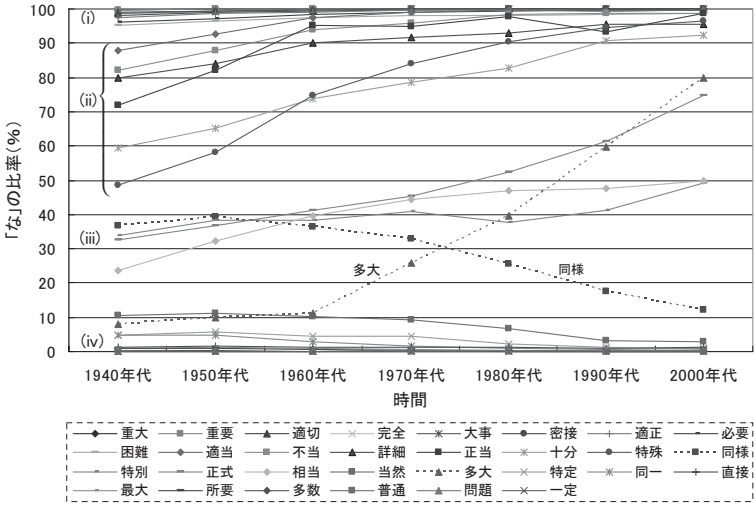
表現によるこの変化の遅速の差の原因は然るべき検討を要する問題であるが、上記の8語に限って見れば、「合理」「民主」「抽象」「徹底」のようにあまり名詞として使われない語の場合は「な」への移行が早かったのに対し、それを「個人」「財政」「国家」「事務」のような普通の名詞の場合が後追いする形になっていると言える。

なお、今や1940年代に至るまでの時期において「な」と「の」の使い分けが具体的にどのような経緯をたどったかということも知りたいところであるが、残念ながら目下の資料ではそれを確かめることができない。

さて、次に（図２）に示すのは、「～的」という形の表現以外の、一般の形容動詞の類における「な」と「の」の使い分けの変化の様子である。

ここでは、「～な」と「～の」の用例数の合計がすべての年代において 1,000 件以上あるものについてのみ示している（用例数の合計の多いものは約 44,000 例）。

（図２）「～な／～の」における「～な」の比率の変化 ──形容動詞一般



（図２）に見る「～な」と「～の」の関係の変化の様相は（図１）の場合に比べて複雑で、(1)の一般化をいっそう明確に具現している。まず、「な」の比率がすべて 70％～ 100％の範囲に収まっていた（図１）と異なり、ここではその比率が 0％から 100％の全範囲にわたっている。また、一部の例外 ──破線で示した「多大」と「同様」── を除けば、形容動詞は「な」と「の」の使い分けの変化に関して次のような３つないし４つの類に分けて捉え得ることが分かる。

(2)（ｉ）1940 年代の段階ですでに「な」の比率が 95％を超えており、その後完全に 100％に達したもの（「重大」「重要」「適切」「完

全」など）

（ⅱ）1940 年代には「な」の比率がほぼ 50 ～ 90％の範囲にあり、その後比率が高まって、現在では 90％を超えたもの（「適当」「不当」「詳細」「正当」など）

（ⅲ）「な」の比率が 1940 年代における 20 ～ 40％の範囲から漸次高まり、現在では 50 ～ 80％の程度に達しているもの（「特別」「正式」「相当」）

（ⅳ）全期間を通して「な」の比率が概ね 10％以下であり、どちらかと言えば全体的に比率が低下する傾向にあるもの（「同一」「直接」「最大」「所要」など）

「の」から「な」へと比重を移す全体的な動きのかげで、実は（ⅳ）のようにそれに逆行する場合もあることが興味を引く。この種のことが日本語の文法変化においてどれほど見られるか今は明らかではないが、一般に

（3）文法変化の相対的に大きな動きのかげで、それには一致しない目立たない動きが進行していることがある。

と言える可能性がある。もっとも、「な」の比率がある場合は高まり、ある場合は低下するという矛盾した現象と捉えるとその本質が見えてこないが、これは「の」と「な」の混用の状態から両者が区別される状態へと向かう両極化の変化として一元的に解釈できる可能性がある。すなわち、例えば拙論（2002）で述べたような「な」と「の」の使い分けに関わる意味的な要因に基づいて、すべての形容動詞が相互排他的な２つの類に編成されつつあるという可能性である。

また、（図２）では「多大」「同様」の２語が他と異なる振る舞いを示しているが、このことから推測でき、実際ほかの事例の分析においても繰り返し確かめることのできる、文法変化に関わるもう１つの一般的事実は、

(4) 往々にして、文法変化の動きに逆行さえするような個別的な例外が観察される。

ということである。こうした例外的現象の持つ意味合いはおそらく場合によりさまざまであろう。すなわち、体系が新しい論理に基づく調整を被る際にある項目の位置付けが変更されるのが例外に見えるという可能性もあるだろうし、単に特定の人物・マスコミなどの影響によって真に例外的な言い回しが普及・定着するということもあり得ると思われる。ともあれ、（図２）における当該の２語の例外的な動きに対して循環論に陥らない解釈を与えるにはさらなる考察が必要である。

なお、先に見た「～的」における「な」と「の」の関係の変化は、形容動詞類全般に関わる今の分類にきれいには収まらないが、(i) に準じるものと言える。

２．２　引用句におけるコピュラの顕在化傾向

コピュラの異形態の複雑な分布の様相はいくつかの局面においてゆれを伴う。「な」と「の」の使い分けと並ぶ問題の１つが、引用の「と」に先行する位置でのコピュラの潜在である。「問題だと思う」「重要だと思う」に対して、「問題φと思う」「重要φと思う」とも言うことができる。

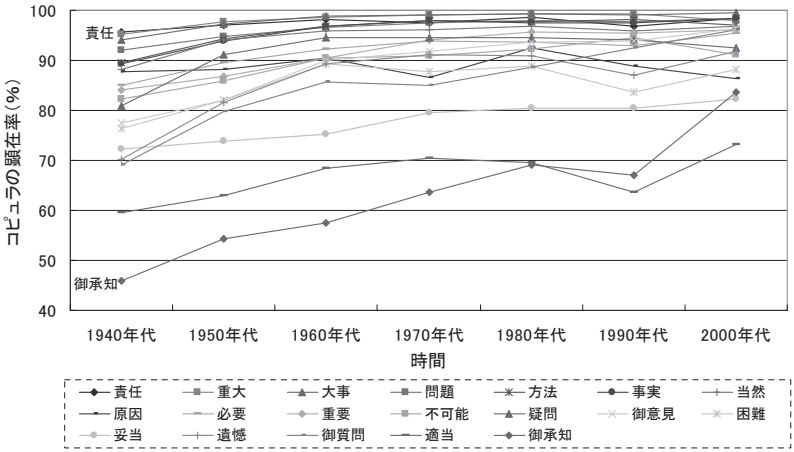
共時的な観点からコピュラ異形態の分布を考察した拙論（2006）においては、現代語の素朴な感覚に基づき、特に迷うこともなくそうしたコピュラの“潜在”という見方を採った。すなわち、「問題だと思う」のように明示的なコピュラを伴う表現が元来の形であり、「問題φと思う」はそこにあったコピュラが端折られた表現であろうと解釈した。

しかし、国会会議録データに基づいてコピュラの使用状況を多少調べたところによれば、日本語の事実はそうではなかった可能性がある。「～と

思う」におけるコピュラの“潜在”は時間の経過とともに増加しておらず、むしろゆるやかな減少の傾向を示しているのである。

その様子を示したのが（図３）である。ここではコピュラの潜在率ではなくその顕在率に基づいてグラフ化してある。潜在率と顕在率のいずれによって作図しても同じことであるが、“無（＝コピュラの不在）”の比率より“有”の比率のほうが分かりやすいように思われるので顕在率を採用した。この図におけるコピュラの顕在率とは、「～{だ／である／φ}と思う」という３通りの言い回しの用例中に占める「～{だ／である}と思う」の用例の比率である。「思う」はすべての活用形を含む。（図３）に示したのは、「～{だ／である／φ}と思う」の用例数の合計がすべての年代において 100 以上の漢字表記語に関わる統計である（用例数の合計が多いものは約 11,000 例）。

（図３）引用句におけるコピュラの顕在化傾向

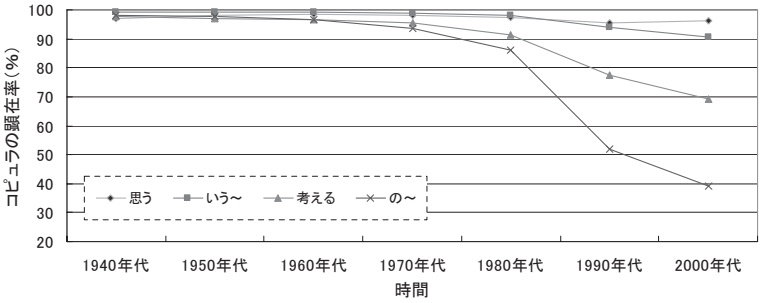


この図から、引用句におけるコピュラの顕在率が全体に徐々に高まっており、コピュラの省略化が進行しているわけではないことが確かめられる。

もっとも、引用の「と」に「思う」以外の述語が続く場合や、顕在・潜

在のコピュラに漢字表記語以外の表現が先行する場合にはまたそれぞれに異なる状況が観察され、（図３）に示した調査結果だけをもって目下の問題の考察材料とすることはできない。例えば、昨今「～べきφとの意見」のような表現をよく耳にするが、「～べき{だ／である／φ}と～」における「だ」の顕在率は、（図４）に見るように最近急速に低下している。ここでは「と」の４種類の後続文脈 ──「（～と）思う」「（～と）いう～」「（～と）考える」「（～と）の～」── の場合ごとの比率の変化を示している。用例数はプロットされた個々の点につき 300 から 30,000 の範囲にある。

（図４）「～べき＿と～」の空所位置におけるコピュラの省略化



「（～と）の～」、次いで「（～と）考える」の場合に、「べき」に続くコピュラが省略される傾向が強まっていることが分かる。ただ、後続文脈の違いによってコピュラの顕在率にかなりの開きが生じていることの理由については、解釈の可能性も思い当たらない。

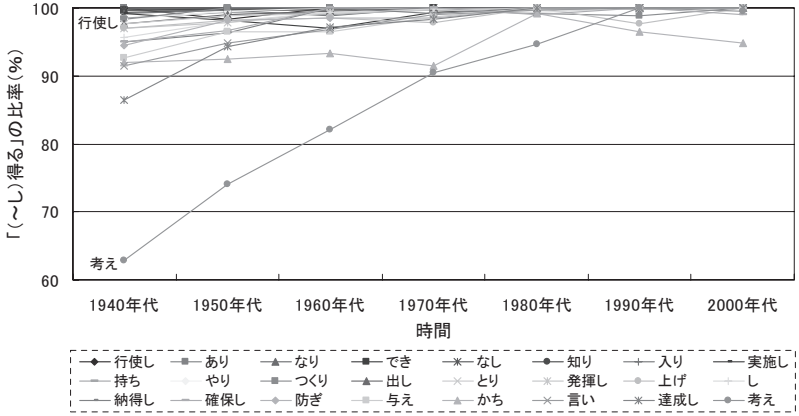
もし（図３）が示唆するようにコピュラの顕在化が引用句におけるコピュラに関する変化の全体的な方向だと言えるとすれば、それは一般に言われる日本語の分析化の動きの１つの現れと考え得るかも知れない。また、「べき」に続く位置でのコピュラの省略傾向はそれに反するものであり、先の一般化（４）に該当する例とも言うことができる。

２．３　可能表現「～し得られる」の衰退と「～できる」の一般化

今日の日本語の感覚からすれば、「～し得られる」という可能表現は、可能を表す「得る」と「られる」を同時に含んでいていかにも冗長である。ところが、国会会議録データの古い部分にはこの「～し得られる」という表現がかなりの頻度で出現する。「いろいろな原因が考え得られる」「自由なる発展が期し得られる」「~~世界の信頼はかち得られない~~」76頁追記参照といった調子である。中には、「生活ができ得られる賃金」「とうてい解決はでき得られない」のように現代の感覚では３重に冗長と言える「でき得られる」という表現もある。

（図５）は、ともに「得る」を含む可能表現「～し得られる」と「～し得る」の関係を調べてみた結果である。「～し得られる／～し得る」全体において、冗長性のない「～し得る」の占める割合が時間の経過とともに増していった様子が分かる。用例数の合計は 50 ～ 30,000 例の範囲にある。「得られる」「得る」に先行する動詞の認定は、形態素解析ソフトウェア MeCab を利用して行った。[3)]

（図５）「(～し) 得られる」から「(～し) 得る」への収束

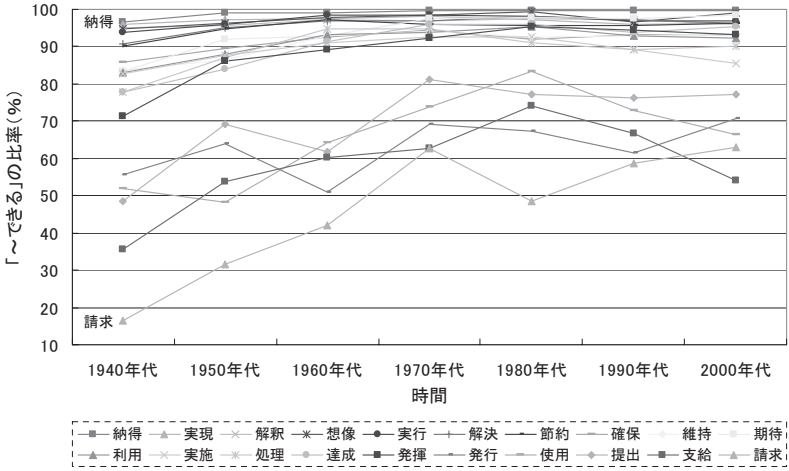


動詞によって「～し得られる」の比率が異なり、特に「考える」の場合に「考え得られる」の比率が高いが、そのあたりの理由は不明である。

なお、「～し得られる」という言い回しは現代の語感では誤用と言ってもよいほど不自然に感じられるが、単独の「得られる」が“入手できる”という可能の意味を表し得る普通の表現であることを考え併せるに、古い時代の「～し得られる」に含まれる「得る」は可能というよりむしろ入手、実現といった意味を表すものとして捉えられていたのかも知れない。とすれば、「得る」に「られる」が続くことは当時の感覚では特に冗長性を伴わなかったことになる。とは言え、少なくとも「でき得られる」はいずれにしても可能表現の重複と言わざるを得ないわけであるが、これとて「できうれば」「できうるかぎり」といった表現が今もときに使われることを思えば、実は現代の感覚との隔たりはさほど大きくもないのかも知れない。

（図６）は、今もともに普通に使われる可能表現である「～することができる」と「～できる」の関係を調査した結果である。「～することができる」と「～できる」の用例全体に占める「～できる」の用例の割合が時間とともに増加している様子が分かる。「～」の部分に漢語が入った用例のうち、用例数の合計が 100 以上のものの統計を示している（最大は 6,000 例余り）。

（図６）「～することができる」から「～できる」への移行



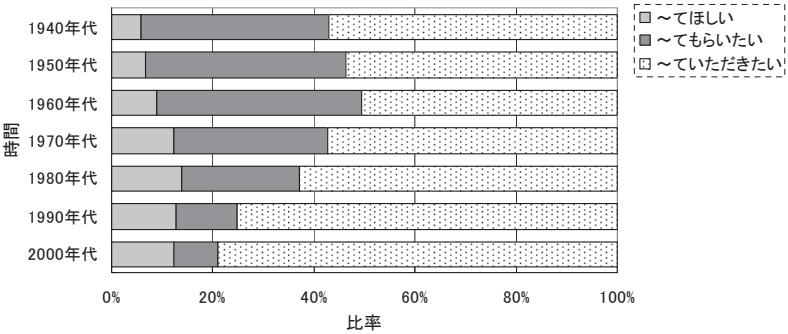
語によって「～できる」の比率に大きな開きがあり、比率の大小はある程度語の意味的な特性と相関しているように思えるが、確かなことは分からない。「飲むことができる」「見ることができる」のような和語動詞に基づく表現 ──それらは主に「飲める」「見（ら）れる」との関係が問題となる── の使用状況も併せて検討してみる必要がある。

２．４　要望・依頼の表現

国会会議録データによる現代日本語の文法変化の試行的な調査の結果をもう２つ簡単に示す。

まず、（図７）は、話し手の要望を表す「～してほしい」「～してもらいたい」「～していただきたい」の比率の年代による違いを調べてみた結果である。統計は「欲しい」「貰いたい」「頂きたい」「戴きたい」のような漢字表記の用例も含んでいる。用例数の合計は約 50,000 ～ 200,000 の範囲にある。

（図７）話し手の要望を表す表現の比率の推移



この図から、「～してほしい」の比率が 1940 年代から 1970 年代ごろにかけて高まっていること、そして、「～してもらいたい」の比率が過去 30 ～ 40 年のあいだに急速に下がってきて「～していただきたい」の使用が増していることが分かる。後者の変化は、ちょうど動詞「やる」の表す配慮の度合いが「あげる」との関係において相対的に低下した現象と同様に、「もらう」の表す配慮の度合いが濫用気味の「いただく」との関係において低下していることを意味するものかも知れない。

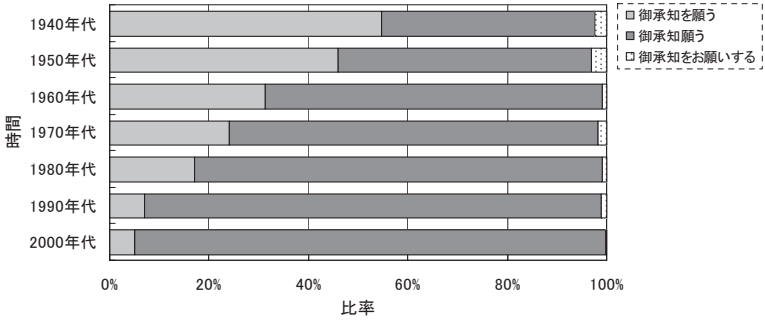
もう１つの事例として、次のような言い回しは現代の感覚で見てどうであろうか。

そのように御承知を願いたいと思います
御意見のある方はお述べを願いたいと思います

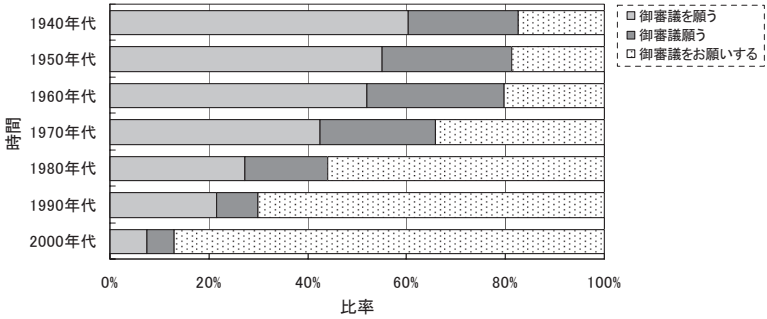
筆者なら「ご承知願いたい」「お述べ願いたい」と直したいところであるが ――と言っても、それらの言い回しも自ら使うことはめったにしないのだが―― 、国会会議録データの相対的に古い部分にはこうした言い回しの用例が頻繁に現れる。ほかにも「ご承知をお願いする（お願いいたす)」のような言い回しも見られ、少なくとも「御（お/ご）～を願う」「御～願う」「御～をお願いする」という３つの句型の使い分けが問題となる（これらの変異形が

まだほかにもある)。国会会議録データで依頼を表すこの３句型の使い分けの状況を調べてみると、「～」の部分に入る語によって様子が大きく異なることが分かる。ここでは「承知」と「審議」の２語の場合だけについて各句型の使用率の変化を（図８)、(図９）に示す。いずれも用例数の合計は約 800 ～ 8,000 の範囲にある。

（図８）依頼３句型の使用率の推移（1)「承知」の場合



（図９）依頼３句型の使用率の推移（2)「審議」の場合



「承知」と「審議」は、それらの表す行為・活動が内面的か外面的かという点においても、活動の時間的なあり方という点においても異なり、その違いが句型の使い分けの傾向に差をもたらしているものと見られる。ただ、一見気紛れに見える多数の語のさまざまな振る舞いを統一的に解釈す

ることは容易ではなく、また、「お述べを願う」のような和語動詞の場合は漢語動詞の場合とはまた違った振る舞いを見せる。（図８）、（図９）にとりあえず例示したのは、そうした複雑な状況のほんの一面に過ぎない。

２．５　言語変化の普及浸透に関する推測

古臭く形式ばった印象の強い国会の日本語から、現代日本語の文法変化の様相をこれほどまでに多様かつ明瞭に観察することができるということは筆者の事前の予想にはなかった。

国会会議録データの時間に基づく区分対照用法を試してみて事前の予想が良い意味で裏切られたわけであるが、この結果を受けて考えてみるに、形式ばった言葉遣いであってもそこには２通りの様式で一般の日本語の新しい潮流が流れ込むということではなかろうか。

その１つは無意識的な影響である。言語の表現には話し手が意識的に制御しやすい部分とそうでない部分とがある。同じことを言うにも複数の表現があるというとき、その選択のあるものは意識的に、あるものは無意識的に行われる。例えば、待遇に関わるような性質の選択はかなり意識的に行われる。「わたし」「わたくし」「おれ」、「だ」「です」「でございます」といった選択肢からの選択は状況に応じて話し手が判断する。選択を誤れば人間関係上の問題も生じ得る。それに対して、「特別な」と言うか「特別の」と言うか、「重要と思う」と言うか「重要だと思う」と言うかといったことは通常無意識の選択に委ねられ、あるいは、そもそも当の話し手にとって選択の余地なく習慣的に一方に決まっている。どの表現を使ってもさしたる問題が生じることもない。せいぜい例えばラ抜き言葉のような社会的に烙印を押された表現を使うと人から指摘を受ける恐れがあるという程度のことである。言語のこうした無意識的な面での日本語の変化はあらたまった言葉遣いにも容易に忍び込むものと思われる。

また、意識的な表現の選択による影響の可能性も考えられる。待遇に関わる表現は意識的に選ばれる面が強いわけであるが、世間で「もらう」はぞんざいで「いただく」のほうが礼儀にかなっているという感覚が普及すれば、国会の場での発言者の多くもそれに従いたい心境になるであろう。もちろん、「いただく」を使うことが個人的にも習慣化し、それが国会の場でも無意識に口を突いて出て来るということもあり得るから、意識的か無意識的かという区別は絶対的なものではない。

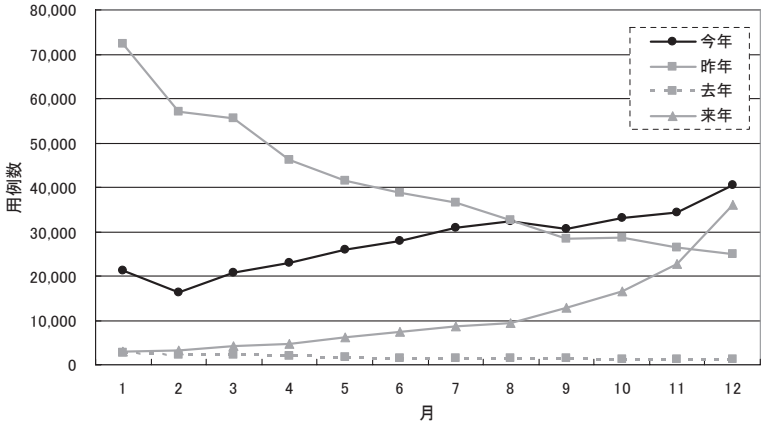
以上のような事情から、日本語の変化、新しい動きが一見旧態依然としているようにも感じられる形式ばった言葉遣いの中にも取り入れられ、浸透していくことになるのではないかと推測される。

３．語彙使用の季節変動 ──新聞記事データに基づく分析

大規模なコーパスの時間に基づく区分対照用法のもう１つの試みとして、朝日新聞 1987 ～ 1997 年、毎日新聞 1991 ～ 2005 年の延べ 26 年分の新聞記事データを月ごとに区分して利用する可能性を試してみる。データの総量は約 3 ギガ（= 30 億）バイト、文字数にして約 15 億字である。

いろいろな表現の組み合わせで調査してみた結果の一部を以下に示す。まず、（図 10）に示すのは、「今年」「昨年」「去年」「来年」の用例数の年間の推移である。「今年度」などのように「～年度」の形をしたものは除外している。[4)]

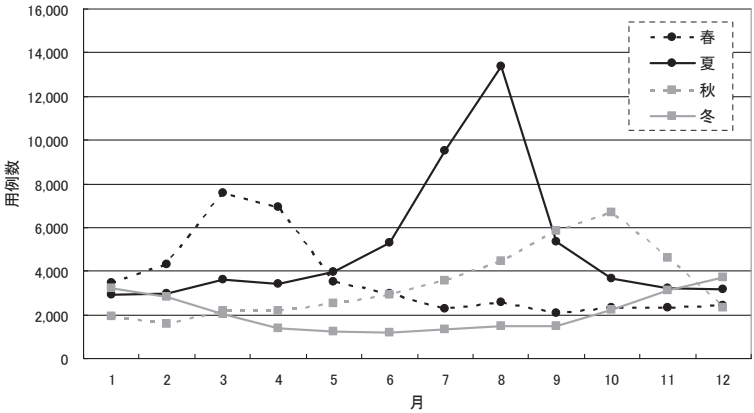
（図 10）「今年」「昨年」「去年」「来年」の用例数の推移



１月から時間の経過とともに使用頻度が下がっていく「昨年」と、年末に向かって使用頻度の増えていく「来年」の用例数の推移の傾向は事前の予想にも合致した自然なものである。「今年」の用例数が１月に比べて２月は多少減り、それ以後 12 月にかけて増加していくということは事前の予想にはなかったが、年の始めに「今年」のことを語る機会が多いということであろうから、結果を見て考えれば納得は行く。ちなみに、「昨年」が多く「去年」が少ないのは新聞記事という資料の性格によるものであろう。

（図 11）は、四季を表す「春」「夏」「秋」「冬」の用例数の年間の推移である。調査の都合上、漢字が前接ないし後接した用例は除外している。

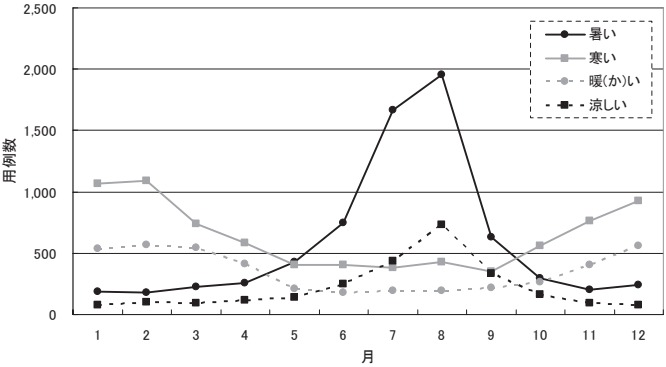
（図 11）「春」「夏」「秋」「冬」の用例数の推移



4 語のうち「夏」の使用が際立って多く、逆に「冬」が少ないということの理由は不明であるが、用例数の推移の様子自体は、「今年」「来年」などの場合と同様、大まかには常識的に予想できるところに一致している。

これに対し、（図 12）に示す寒暖を表す形容詞「暑い」「寒い」「暖かい」「涼しい」の用例数の推移は少々様相が異なる。調査の都合上、連体・終止形と連用形だけを数えている。

（図 12）「暑い」「寒い」「暖かい」「涼しい」の用例数の推移



四季を表す4語の場合とは異なり、「暖かい」「暑い」「涼しい」「寒い」は季節の移り変わりに従って順次使用のピークを迎えるわけではなく、「暑い」と「涼しい」、「寒い」と「暖かい」の用例数の推移がそれぞれ同期している。

これは次のように解釈することができよう。すなわち、暑さという不快を前提としてその不在の状態を言うのが「涼しい」であり、寒さという不快を前提としてその不在の状態を言うのが「暖かい」である。一般に「暑い」の反対語は「寒い」、「暖かい」の反対語は「涼しい」だと考えられ、それはある意味では誤りではないが、言語使用上の意識としては通常むしろ「暑い」と「涼しい」、「寒い」と「暖かい」がそれぞれ対立を成す。例えば、「(この部屋は) 寒い？」は“寒いか暖かいか”を問うものであって、“寒いか暑いか”という問いではない。同様に、「暖かい？」は“暖かいか寒いか”という問いであって、“暖かいか涼しいか”の問いではない。こうしたことから、「涼しい」は暑い季節にこそ使われ、「暖かい」は寒い季節にこそ使われやすいことになるのであろう。

以上のような月ごとの用例数の推移の調査以外にも、曜日ごと、あるいは、平日か休日かといった観点から同様の分析をしてみることも考えられる。この種の調査によって日本語研究上注目に値する知見がどれほど得られるものかは定かでないが、ともかくさらにいろいろと試してみる価値はある。[5)]

なお、使用した新聞記事データの月ごとの記事量には多少ばらつきがあるので本来それを考慮して統計を補正する必要があるが、今回のような粗い観察に影響するほどの違いを生じないので(図10)～(図12)において用例数の調整は行っていない。

４．おわりに

大規模なコーパスの時間に基づく区分対照用法の可能性を、国会会議録と新聞記事という２種類の電子資料に基づいて試してみた。特に、国会会議録のデータについては、これが現代日本語の通時変化を明らかにするという目的にとって今後大きな役割を果たし得るものであることが分かった。また、若干の事例の試行的な分析だけからでも、文法変化に広く通用すると思われる一般的な事実が見えてきた。国会会議録という膨大な量のデータは現代日本語の変化を跡付けるための材料のまさに宝庫と言ってよいだろう。

付言すれば、現在国立国語研究所で構築が進められている『現代日本語書き言葉均衡コーパス』では、収録される日本語の各サンプルに対して書籍のジャンル、刊行年、著者の生年・性別といった情報が付加される。そうした情報に基づいてコーパスを区分対照して用いることにより、従来の言語資料、従来のコーパスでは望むことのできなかった種類・水準の日本語の調査が可能になるはずである。

注

１）国会会議録のデータは国立国会図書館の Web サイトに設置された国会会議録検索システムのページ（http://kokkai.ndl.go.jp/）から取得することができる。単純な文字列検索は同検索システムを通して行うことができるが、検索を含む各種の処理を自在に行うためにはデータを手元のコンピュータ上に置いて使う必要がある。

この試行的な小論では国会会議録のデータを差し当たり完全に“信頼”して文法の変化を論じるが、同データを日本語研究資料として見るとき注意を要する問題は多い。何より重要なのは、国会での言葉遣いは明らかに“日本語の平均像”――そのようなものを厳密に特定することはそもそも無理であろうが―― とは考えられないということである。と言っても、国会会議録データで観察される文法の変化が一般の日本語におけるそれと隔絶されたものだと考える必要もない。両者の関係については本文で後に触れる（２.５）。

また、国会会議録データが国会の場における実際の発話の完全に忠実な記録とは限らないという問題もある。一般に話しことばを書き起こした文章を言語研究の資料として使う際には実際の発話と記録のあいだのずれの可能性に注意する必要がある。このことについては、最近出版された松田編（2008）に、国会会議録における“整文”などの問題に関する詳細な分析が収められていて興味深く参考になる。

2）調査の便宜上、「～的な」「～的の」の直後に漢字が続く用例だけを調べた。そのようにしたのは、後ろに形式名詞が続く用例を除外するのが目的である。例えば、形式名詞の「の」に先行する位置では「の」の現れが強く制約される。「無名の作家」のように通常は「の」を選ぶ傾向の強い「無名」も、形式名詞「の」に先行する位置では「×彼が無名ののは」ではなく「彼が無名なのは」となる。そうした異質な要因の干渉を避けるための措置である。

本来具体的な用例数をすべて掲げるのが望ましいが、特定の問題を綿密に分析するという文脈でもなく、紙幅の制約もあるので省略に従う。以後の例についても同様とする。事例ごとに関連の用例数の規模を示すので、統計の信頼性の程度はそこから推定できよう。

3）MeCab は http://mecab.sourceforge.net/ で公開されている。

4）単なる文字列として検索しただけなので、統計がゴミを多少含む可能性がある。以後の例についても同様である。

5）山崎誠氏のお勧めにより、月の内部における「今月」「来月」「先月」の使用頻度の推移や、週の内部における「今週」「来週」「先週」の使用頻度の推移などについても調べてみたところ、「今年」「来年」「昨年」の場合に共通しつつも、それとは違うところもある結果が得られた。

文献

田野村忠温（2002）「形容動詞連体形における『な／の』選択の一要因 ──『有名な』と『無名の』── 」『計量国語学』第 23 巻第 4 号

田野村忠温（2006）「コピュラ再考」藤田保幸・山崎誠編『複合辞研究の現在』（和泉書院）

田野村忠温（2009）「コーパスと文法研究」『国文学解釈と鑑賞』第 74 巻第 1 号

松田謙次郎編（2008）『国会会議録を使った日本語研究』（ひつじ書房）

付記

本稿は文部科学省科学研究費補助金特定領域研究「日本語コーパス」の日本語学班研究会（千里朝日阪急ビル、2008 年 2 月 17 日）での発表内容の一部にそ

の後の調査事例を加えてまとめ直したものである。発表時に最も詳しく観察した一字漢語複合サ変動詞の活用の変化については、その後行った他の資料に基づく分析と併せて別途論じたいと考えている。

（文学研究科教授）

追記　２．３に挙げた「世界の信頼はかち得られない」という用例は不適切であった。「勝ち得る」は「勝つ」の可能表現ではないからである。「強行手段も避け得られなかった」という用例に入れ替えてお読み願いたい。また、同じ理由により、（図５）の「かち」のグラフ線を抹消する必要がある。当該のグラフ線はやや例外的な振る舞いを示しており、その抹消により図全体のグラフ線の動きが斉一になる。（2009年1月）

SUMMARY

A Corpus-Based Analysis of Some Time-Related Aspects of Contemporary Japanese

Tadaharu TANOMURA

The past several years have seen rapid progress in the conditions surrounding corpus-based research of the Japanese language. With respect to available Japanese corpora, several corpora specifically designed for use in language studies as well as several kinds of huge electronic texts with distinctive characteristics have become publically available.

This paper reports some of the results of the author's attempts to utilize two kinds of electronic texts in the investigation of time-related aspects of the language. In the first half, diachronic changes in Contemporary Japanese will be analyzed based upon the texts of *Kokkai Kaigiroku*, the minutes of the National Diet of Japan, covering the period 1947-2008. It will be shown that the *Kokkai Kaigiroku* data enables us to uncover hitherto unknown recent diachronic changes of Japanese grammar and morphology. The topics dealt with include changes in the choice between *na* and *no* as the suffixes of nominal adjectives, changes in the frequency of copula ellipsis in quotative clauses, and changes in speakers' preference for some idiomatic phrase patterns expressing possibility and ability.

In the second half of the paper, seasonal changes in the use frequencies of a few sets of nouns and adjectives will be analyzed using newspaper texts. We will see, among other things, that the use frequencies of *atatakai* (warm) and *suzusii* (cool) synchronize with those of *samui* (cold) and *atui* (hot) respectively, attesting the validity of an independent semantic analysis of the four terms.

Keywords: corpus linguistics, Japanese grammar, diachronic change, seasonal change in the use frequency of words