The University of Osaka
Institutional Knowledge Archive

Title	類義語を共起関係から紐解く ： コーパス分析を用いた指導の有用性
Author(s)	久保, 圭
Citation	大阪大学日本語日本文化教育センター授業研究. 2014, 12, p. 23-31
Version Type	VoR
URL	https://doi.org/10.18910/50787
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

類義語を共起関係から紐解く
—コーパス分析を用いた指導の有用性—

久保　圭

【要旨】

日本語には、似た意味をもつさまざまな言葉がある。たとえば、「誤りを直す」という意味をもつ日本語のサ変名詞には、修正、是正、訂正、改竄などがある。しかし、その意味的なニュアンスや用法（どのような対象における誤りを直すのか、など）は一律ではなく、すべて異なっていると考えられる。これは、上記に挙げた類義語がすべて同じ意味をもつのであれば、これほど多様な表現は必要ないと考えられるためである。

上記のような類義語の使い分けの習得に関心を寄せる留学生は少なくないように感じられる。しかし、日本語の言語直感をもたない彼らにとって、その学習は困難をともなうものであるため、類似した意味の言葉を教授する際には、効果的な説明を準備しなくてはならないだろう。

本論文では、日本語学習者に向けた教育の現場において、類似した意味をもついくつかの語の意味と使い分けを説明する際に、大規模コーパスから得られる、各類義語の共起関係に関する調査データの活用が有効であることを示唆する。

1．はじめに

日本語における類義語の使い分けは、日本語学習者が、その学習過程において最も困難に感じるものの一つであると思われる。しかし、類義語を適切に使い分けるための知識は、彼らの多くが目標とする「日本人が使う日本語」がもつ母語直感に近いものを習得するために必要な要素であると思われる。事実、筆者が担当する文法や作文の授業においても、受講者から類義語の使い分けに関する質問を受けることは少なくない。

本論文では、日本語学習者が類義語を習得する際の困難を緩和すべく、その有効な説明法についての提案を試みる。具体的には、まず、日本語の大規模コーパスから抽出した、共起関係に関する調査データを活用することで、共通して「誤りを直す」という意味をもつ類義語「修正」「是正」「訂正」「改竄」の意味的・用法的差異を浮き彫りにする。また、日本語コーパスによる調査データを日本語教育に導入することの有用性と問題点について述べる。

2．類義語とは

三省堂『大辞林 第三版』によれば、類義語は「同一の言語体系のなかで、語形は異なっていても意味の似かよった二つ以上の語」と解説があり、その一例として「ホテル／旅館／宿屋」などが挙げられている。この三つの語は、すべて「宿泊のための施設」を指す言葉であるが、日本語の母語話者であれば、「ホテル」は洋風の宿泊施設であり、それに対して「旅館」は和風の宿泊施設であること、そして「宿屋」は少し古風な言いかたであり、現代ではあまり用いられない言葉であるということは直感的にわかるのではないだろうか。

しかし、この「日本語の機微に対する直感」は日本語の母語話者に特有のものであり、日本語学習者が体得するには非常に難しいものといえる。つまり、彼ら日本語学習者にとっての学習とは、「日本語母語話者がもつ直感に限りなく近づく」ためのプロセスであり、それは超級の

日本語学習者にとっても同様と考えられる。

当然のことながら、日本語教育の場においては、この「なんとなくわかる、類義語の細やかな意味の違い」を明示的に指導しなくてはならない。では、説明をおこなう際にはどのような手段を用いることが効果的、つまり「わかりやすく、実践的」であるだろうか。

そこで、本論文では類義語の一例として「誤りを直す」という意味をもつ日本語、とりわけ「修正」とその類義語として位置づけられる「是正」「訂正」「改竄」というサ変名詞を分析対象として扱い、コーパス検索によって収集した共起関係の調査データを分析する。また、この調査データが各類義語の意味的差異と用法を説明する際に有用であることを示す。

3．辞書の問題点

どのような学習者が、どのような言語を学ぶにしても、ある言葉の意味がわからない場合や、おおよその意味はわかるけれども、用法がわからないといった場合には、まずはその言葉を辞書で調べてみるのが一般的な解決策であろう。しかし、辞書を引いたからといって、必ずしもその言葉の意味や用法が把握できるわけではない。特に、いくつかの類義語を調べるといった際には、それぞれの解説が非常に似通っていることもしばしばであり、その意味や用法の差異が理解できないということも多々あると考えられる。たとえば、本論文の分析対象である「修正」と、その類義語である「是正」「訂正」「改竄」を『大辞林 第三版』で調べてみると、以下のような解説が載っている。

（1）【修正】まちがっていたり、不十分であるところを直して正しくすること。
（2）【是正】誤っている点をなおして正しくすること。
（3）【訂正】言葉や文章の誤っている部分を正しく直すこと。
（4）【改竄】文書の字句などを書き直してしまうこと。普通，悪用する場合にいう。

上記のそれぞれの解説をくらべてみると、まず、（1）は「修正をおこなう対象がどのような性質のものであるか」について明らかではない。この点については（2）も同様であり、「直すべき誤りがどこにあるのか」が不明である。（3）の解説では「言葉や文章の誤っている部分」と「誤りの在り処」について簡潔に記載されているが、この解説も、さらに踏み込めば「どのような性質の言葉や文章なのか」が明らかではないと考えられる。また、（4）の解説には「悪用する場合にいう」という意味的ニュアンスに関する文面があり、上記の解説のなかでは、他の語との差異が感じられるものではあるが、やはり「どのような性質の文書を書き直すのか」について読み取ることはできない。また、「改竄」という行為において直される「誤り」の内容は、その他の類義語におけるものとは性質が異なっているようにも感じられる。

以上のように、辞書に記載されている解説から各類義語の意味的差異を理解しようとすることには困難が生じる場合がある。当然、上記における語の意味的なニュアンスや用法（どのような対象における誤りを直すのか）はすべて異なっていると考えられる。

もし、上記に挙げた類義語がすべて同じ意味をもつのであれば、表現を多様にする必要はなく、いずれか一つの言葉に定めて、それにあらゆる意味を担わせてしまえば事足りるからである。この「類義語のそれぞれの意味には共通する部分がみられるが、何かしらのところでは異

なりがある」という考えに立脚しない場合、各類義語の意味は「すべて同じ」ということになってしまい、日本語学習者に「このなかのどれを使っても、意味は同じで正しい」という誤解を与えてしまいかねないだろう。その意味的差異があまりに小さく、日本語母語話者による使用においても使い分けがなされていないということも確かにあり得るが、そのような場合を除いたとしても、各類義語の意味と用法については慎重な指導が望まれるところである。

4．分析

4.1　本論文で用いるコーパス：Sketch Engine

近年、インターネットの普及によって、我々は大量の情報を取得したり、処理したりすることが比較的容易にできるようになった。また、言語処理の発達とともに、必要な情報を抽出する際の技術も成熟しつつあり、さまざまな目的に合わせた利用が可能となっている。この有用性は言語学や言語教育、その他の分野においても非常に注目されており、その大規模に蓄積された言語資源（＝コーパス）を利活用した研究が国内外において盛んにおこなわれている。

本論文で用いるSketch Engineは、さまざまな言語のウェブデータをコーパス化したものであり、ウェブ上においてテキストデータと検索ツールを利用することができるシステムとなっている。現在、Sketch Engineに実装されている日本語コーパスは、jpTenTen11とJpWaCの二種類があり、jpTenTen11には約103億語、JpWaCには約４億語が収載されている。

なお、本論文における各類義語の調査は、jpTenTen11を用いておこなったものである。

4.2　調査から得られたデータ

本論文では、Sketch Engineを用いたコーパス検索によって、「修正」「是正」「訂正」「改竄」のそれぞれが「どのような対象」における誤りを正す語であるのかについて、その共起関係と使用頻度に関する調査をおこなった。

それぞれの類義語が対象（以下、必要に応じてＸとする）を明示する場合に用いられるフレーズには、さまざまなものが考えられる（Ｘを修正する／Ｘが修正される／Ｘの修正をする／Ｘに修正を加える、など）が、本論文ではヲ格を用いて対象をあらわすフレーズに限定して調査をおこなう。つまり、「Ｘを修正／是正／訂正／改竄（する／した、など）」というフレーズのＸについて、それぞれの語で調査をおこなったということである。

表１は「修正／是正／訂正／改竄」の各語の左側に共起する一つ目の形態素を頻度順に並べたものであるが、この内容から、ヲ格を用いた表現がどの類義語についても一般的なものであることがわかる。

また、表２-５は、Sketch Engineによるコーパス検索の調査結果である。各表には、それぞれの類義語の対象（上記のとおり、ヲ格で表示されるものに限る）の上位30件を頻度順で示してある。たとえば、表２は「修正」の対象をまとめたものであるが、これは「Ｘを修正する」のＸに入る言葉として、「問題」が最も高頻度（ヲ格を用いた文章54,265件のうち4,457件が該当）で用いられているということを示している。

表２内には「の」や「それ」、そして「点」など、その指示対象が具体的ではない言葉も含まれているが、これは「ヲ格のすぐ左側に共起する一つ目の形態素を抽出する」という検索方法によるものである。これについては、さらに左側にある形態素を調査することで、具体的な対

表１：各類義語の左側にあらわれる一つ目の形態素（頻度順、上位５件）

「修正」の１語左側	頻　度
無	103,957
を	**54,265**
の	42,239
に	25,120
で	15,220

「是正」の１語左側	頻　度
を	**5,992**
の	5,457
格差	2,890
に	1,397
が	950

「訂正」の１語左側	頻　度
に	7,713
の	6,996
を	**6,593**
と	4,238
で	4,111

「改竄」の１語左側	頻　度
の	3,667
を	**3,492**
に	1,409
が	1,192
データ	820

表２：ヲ格によってあらわれる「修正」の対象（全体数：54,265）

対象	頻度
問題	4,457
具合	3,833
バグ	3,642
の	3,132
それ	1,365
点	836
部分	807
これ	588
現象	543
内容	505

対象	頻度
エラー	497
ファイル	456
性	424
ところ	394
位置	373
ミス	371
データ	365
プログラム	346
箇所	320
方	312

対象	頻度
など	303
軌道	292
コード	288
間違い	280
ズレ	276
金額	262
もの	262
情報	248
そこ	243
記事	209

象が明らかになる場合もある（たとえば「表記が誤っていたのを修正した」など）が、本論文ではＸにあらわれる大まかな形態素の傾向を把握したいという目的に沿って、考察に必要であると考えられる場合を除き、ヲ格よりも二つ以上左に位置する形態素について、詳細には触れないこととする。また、表２-５についての考察は次節で述べる。

４.３　観察結果のまとめ

まず、表２の「修正」については、「問題」「点」「部分」「内容」「ミス」「間違い」など、どのような対象についても広く用いられる（つまり、「～の問題」「～の点」などは、～に入る言葉がかなり多様であると考えられる）表現との共起が多くみられた。

その他に特徴的だったのは、「バグ」「エラー」「ファイル」「データ」「プログラム」「コード」など、パーソナル・コンピュータなどをはじめとする、電子機器や情報機器に関わりのある言葉との共起が目立っていることである。なお、二番目に頻度の高い「具合」については、文脈を確認したところ、ほぼ「不具合」の「具合」が接頭辞の「不」と分離した形であらわれたものであった。以下の（５－９）に、収集した文章の一例を挙げる。

表3：ヲ格によってあらわれる「是正」の対象（全体数：5,992）

対象	頻度	対象	頻度	対象	頻度
格差	736	状況	67	方	44
それ	446	欠陥	66	こと	41
これ	324	平等	60	行為	40
均衡	176	差別	55	過ぎ	39
問題	106	偏り	53	偏在	39
状態	82	現状	52	バランス	39
公平	75	点	50	ら	39
円高	74	の	46	差	36
歪み	72	集中	45	さ	36
社会	71	弊害	45	構造	34

表4：ヲ格によってあらわれる「訂正」の対象（全体数：6,593）

対象	頻度	対象	頻度	対象	頻度
間違い	416	の	101	名前	50
それ	313	言葉	99	誤解	48
金額	277	情報	99	方	48
送料	269	こと	93	誤字	47
誤り	221	価格	81	など	47
発言	131	ミス	79	一部	44
内容	128	これ	74	勘違い	43
部分	110	記録	64	文	41
書	104	文章	61	もの	40
記事	101	箇所	54	事	37

表5：ヲ格によってあらわれる「改竄」の対象（全体数：3,492）

対象	頻度	対象	頻度	対象	頻度
データ	340	日付	56	日	31
歴史	187	記事	47	数値	29
記憶	175	資料	43	など	29
情報	133	結果	41	事実	27
証拠	111	カルテ	40	それ	23
記録	96	ファイル	39	時間	22
書	91	書類	37	文書	20
内容	86	ページ	37	成績	20
期限	74	過去	36	プログラム	19
サイト	70	数字	32	システム	19

（5）サーバーとの接続が切断されてしまう問題を修正しました。
（6）文庫の挿絵が数ページずれて配置される不具合を修正しました。
（7）文書を新規作成したときにレイアウトがくずれるバグを修正しました。
（8）複数のパッチでファイルを修正するのは避けてください。
（9）メールで指摘された間違いを修正しました。

次に、表3の「是正」については、最も頻度の高かった「格差」をはじめとして、「均衡」「公平」「円高」「平等」「差別」「差」「構造」など、社会や社会問題に関連のありそうな言葉が対象として多くみられた。なお、「均衡」については、文脈を確認したところ、すべて「不均衡」の「均衡」が接頭辞の「不」と分離した形であらわれたものであった。

また、「均衡」「偏り」「バランス」などの言葉と共起していることからもわかるように、「社会的なアンバランスをあるべき方向に直す」といった意味で用いられることが多く、この意味で用いられているものは調査結果のほとんどを占めている。以下の（10－14）に、収集した文章の一例を挙げる。

（10）大企業と中小企業の格差を是正する。
（11）暴力的手段によって経済的不均衡を是正しようとすることになる。
（12）日本経済の歪みを是正することができる。
（13）そうした状況を是正する課税制度を考えてもいいのではないか。
（14）経済の二重構造を是正していくための改革措置が必要だと強調している。

そして、表4の「訂正」については、大きく目立つところはみられなかったが、やや特徴的と思われるのは、「金額」「送料」「価格」など、金銭に関する言葉が上位に含まれているということである。また、「発言」や「言葉」といった、「言及や記述」に関するものもみられる。以下の（15－19）に、収集した文章の一例を挙げる。

（15）受注確認メールにて、ご請求金額を訂正させていただきます。
（16）この場をお借りして、先の発言を訂正し、お詫びしたい。
（17）言い訳に言い訳を重ね、記事を訂正することは遂に無かった。
（18）相手のミスを訂正しなきゃ気がすまない。
（19）誤解を招く恐れがあるという事で文章を訂正する事になりました。

最後に、表4の「改竄」については、『大辞林 第三版』にも「悪用する場合にいう」という記載があったように、その意味的ニュアンスが他の類義語とは大きく異なっている。この「改竄」は「誤りを直す」といった意味ではなく、逆に「正しいものを誤りにする」という意味である。「改竄」は「修正」の類義語ではあるが、その意味的な共通点は「是正」や「訂正」の場合と違い、さらに抽象的な「あることを別のことに書き換える」くらいに留まるだろう。

また、共起する語の特徴としては、まず、「データ」をはじめとして、「歴史」「記憶」「資料」「カルテ」など、「記録すること」に関連のある言葉が多い。次に、「日付」「数字」「数値」「成績」など、「数」に関わるものも多くみられた。一般的に、カルテや日付を書き仕損じることは、書き手や当事者にとって大きな不利益を招く可能性が高い。よって、これらはなるべく正しく記録されることが望まれるものであるといえる。そのようにして正しく書かれたものを「別のことに書き換える」理由とは、たいていは「正しく書かれた内容、つまり事実が、自分や当事者にとって不利益なものである」という、本来とはまったく別の意図によるものであろう。「改竄」が「悪用する場合にいう」と用法が定められているのは、上記のようなプロセスに起因す

ると考えられる。以下の（20－24）に、収集した文章の一例を挙げる。

（20）パソコンのデータを改竄した上、伝票の一部を廃棄していた可能性がある。
（21）ターゲットは記憶を改竄され、妹がいることを忘れた状態にある。
（22）押収したフロッピーディスクの日付を改竄した疑いを報じていた。
（23）はずれ券の数字を改竄して100万円の当選券を偽造。
（24）従業員の残業時間を改竄して、時間外労働賃金を支払っていなかった。

上記の内容について、以下の（25－28）にまとめる。次節では、コーパスからの調査データを用いる有用性と問題点について述べる。

（25）【修正】 幅広い対象について用いることができるが、「電子機器や情報機器」に関わりのある言葉とよく共起することが特徴的である。
（26）【是正】「社会や社会問題」と関連がある言葉との共起関係が強くみられる。「社会的なバランスを整える」という行為を指す。
（27）【訂正】「修正」と同様、幅広い対象について用いることができる。「金銭」に関する言葉、また、「言及や記述」に関する語との共起が特徴的である。
（28）【改竄】「記録」に関連のある言葉との共起が多く、行為者あるいは当事者にとって「都合の良くないもの」を書き換えるという行為を指す。

5．コーパスからの調査データを用いる利点と問題点

前節において、「修正」とその類義語である「是正」「訂正」「改竄」の共起関係について、Sketch Engineを用いた調査をおこない、その結果から、各類義語の意味的差異について分析をおこなった。では、大規模コーパスから得られる調査データを日本語教育に活かすことについて、他にどのような利点があるだろうか。また、考慮すべき問題点は何だろうか。

コーパスから収集されたデータを用いる利点としては、大きく以下の三点が挙げられる。

利点１：共起関係を含め、実際の使用（＝使用実態）に基づいたデータが得られる

まず、コーパスから得られるデータは、使用実態に基づいたものであるという点が大きな特徴として考えられる。ある語について、基本的な意味や用法には大きな変化がみられないことがほとんどであるが、言語の創造的側面を考慮に入れるならば、逸脱的用法が新たに生まれることもまた稀有なことではない。辞書に記載される解説は、基本的に正用としてのものに限定されるが、コーパスを活用することによって、そのような新規的・逸脱的用法についても把握することが可能となる。

利点２：共起関係の傾向から、類義語間の細やかな意味的差異を把握し得る

次に、共起関係の傾向をみることで、類義語間における細やかな意味的差異を捉えることが可能となる。前節の内容についていえば、たとえば、「修正」と「訂正」は意味的にかなり近いと考えられ、実際、多くの場合においてパラフレーズが可能である。しかし、実際の使用にお

いては、「修正」は電子機器や情報機器に関わりのある語との共起が目立ち、一方、「訂正」は金銭に関係のある語や、言及や記述に関係のある語との共起が特徴的である。これにより、これらの類義語における共起関係の傾向には異なりがみられることがわかる。一見、その意味がほぼ同じように感じられる言葉の細やかな違いを捉えようとする際にも、コーパスを用いた共起関係に関する調査データは有用であると考えられる。

利点3：日本語学習者にとって、より理解しやすく、実践的な例文を提示できる

そして、これは上記の二点と関連するが、コーパスによって得られるデータは、共起関係についてのものだけではない。コーパス検索によってまず得られるものは、検索した語句を含む「文章」のデータであり、共起関係に関するデータは、むしろその文章から得られる副産物であるといえる。その副産物である共起関係のデータから、各類義語の意味的・用法的特徴を浮き彫りにすることが可能である点は前述の通りであるが、コーパス検索の段階で得られた文章は、日本語学習者に類義語を説明する際の例文として有効活用することができる。各類義語の使用実態と意味的差異を反映した「わかりやすく、すぐに使うことのできる」例文を提示することで、日本語学習者にとっての学びはより実りのあるものになると考えられる。

以上において、コーパスからの調査データを用いる利点について述べたが、考慮すべき問題点もいくつかある。

問題点1：コーパスにはノイズとなるテキストが含まれている

本論文で用いたSketch Engineに実装されている日本語コーパスjpTenTen11についても、ウェブ上のテキストを集積してつくられたものであるため、個人のブログなどにおける文章も多く収録されている。その文章内には誤字・脱字がみられることも多々あるため、このようなノイズとなる文章をなるべく取り除くように留意すべきである。このノイズをめぐる問題点については、『青空文庫コーパス』や国立国語研究所が提供する『現代日本語書き言葉均衡コーパス』などの書き言葉のみを扱ったコーパスにも同様のことがいえる。

問題点2：それぞれのコーパスが備える性質について吟味する必要がある

本論文で用いたSketch Engineに実装されている日本語コーパスjpTenTen11は、前述の通り、ウェブ上のテキストを集積してつくられたものである。それゆえ、話し言葉に近い書き言葉を用いた文章が非常に多くみられる。この問題点については、書き言葉のみを扱ったコーパス、あるいは話し言葉のみを扱ったコーパスを用いることで、部分的な解消が可能であると思われるが、スタイルに偏りが生じてしまうという新たな問題点が生じる。

問題点3：正例／新規的・逸脱的用法／誤用の区別が困難になる場合がある

コーパス検索をおこなうと、ある言葉の新規的・逸脱的用法が把握できる反面、単なる誤用と思われる用法によって書かれた文章についても収集してしまうため、日本語学習者に提示する例文を検討する際には、少なくともそのような誤用による文章は除外しなくてはならない。また、場合によっては、新規的・逸脱的用法が検索の上位にあらわれることも考えられるため、

これを正例とみなしてしまわないよう吟味が必要である。

どのようなコーパスを用いる場合にも、目的とするデータを得るための検索手法や、調査データにとってノイズとなるテキストの除去について留意し、また、収集した結果の捉えかたにも間違いや偏りがないように心がける必要がある。

6．おわりに

本論文では、日本語学習者に向けた教育の現場において、類似した意味をもついくつかの語の意味と使い分けを説明する際に、大規模コーパスから得られる、各類義語の共起関係に関する調査データの使用が有効であることを示唆した。

まず、日本語の大規模コーパスSketch Engineを用いて、「誤りを直す」という意味をもつ類義語「修正」「是正」「訂正」「改竄」の共起関係に関するデータを抽出し、それぞれの意味的・用法的差異について分析した。また、日本語コーパスによる調査データを日本語教育の現場に導入することの有用性と、コーパスデータを用いる際の問題点について述べた。

本論文では「修正」「是正」「訂正」「改竄」の四つについて分析をおこなったが、「誤りを直す」という意味をもつ言葉はこれがすべてではなく、この他にも「校訂」「改訂」「修訂」「校閲」などがある。今後、これらについてもコーパス検索による調査・分析をおこなうことで、各類義語の意味的・用法的差異がさらに深くまで明らかになるかもしれない。

辞書

松村明（編）（2006）『大辞林 第三版』，三省堂.

コーパス検索に用いたツール

Sketch Engine（http://www.sketchengine.co.uk/）

参考文献

岩佐靖夫（2011）「日本語教育における類義語指導の一考察：系統的な指導原理に向けての提言」『尚美学園大学総合政策研究紀要』20，尚美学園大学，17-24.

スルダノヴィッチ・イレーナ，仁科喜久子（2008）「コーパス検索ツールSketch Engineの日本語版とその利用方法」『日本語科学』23，国書刊行会，59-80.

新屋映子（2010）「類義語『状況』『状態』の統語的分析：コーパスによる数量的比較」『計量国語学』27（5），計量国語学会，173-193.

Erjavec, Irena Srdanovic, Tomaz Erjavec and Adam Kilgarriff (2008). A Web Corpus and Word Sketches for Japanese. *Information and Media Technologies* 3(3), 529-551.

（くぼ　けい　本センター非常勤講師）