

全文 テキストにおける語彙の分布と文章構造 成果報告書

著者	山崎 誠, 内山 清子, 江田 すみれ, 小森 理, 清水 まさ子, 高崎 みどり, 馬場 俊臣, 馬場 康維, 村 田 年
ページ	1-152
発行年	2013-03-25
シリーズ	国立国語研究所共同研究報告 ; 12-06
URL	http://doi.org/10.15084/00002701

テキストにおける語彙の分布と文章構造 成果報告書

山崎誠・内山清子・江田すみれ・小森理・清水まさ子・高崎みどり・
馬場俊臣・馬場康維・村田年

2013年3月

テキストにおける語彙の分布と文章構造 成果報告書

目次

本報告書について (山崎誠)	1
論文の論理構造における分野基礎用語に関する分析 (内山清子)	3
テキストの違いと受身文の違い	
— 会話・ブログ・新書の受身の使われ方をもとに (江田すみれ)	13
学術論文における問題提起疑問文とそれに対する答え方 (清水まさ子)	31
文章中の語彙の機能について— “テキスト構成機能” という観点から— (高崎みどり)	41
接続表現の二重使用と文章ジャンル	
— 『現代日本語書き言葉均衡コーパス』を利用して— (馬場俊臣)	67
状態空間表現を用いた文章の特徴付け (馬場康維・小森理)	89
「手」の慣用句を指標とした文章ジャンルの判別	
— 現代日本語書き言葉均衡コーパスを用いて— (村田年・山崎誠)	102
自然科学系書籍における複合動詞の使用傾向	
— 後項動詞を指標として— (村田年・山崎誠)	115
共起語率の分布からみるテキストの語彙的特徴 (山崎誠)	137
段落間の類似度を利用したテキストの結束性の測定 (山崎誠)	145

本報告書について

山崎 誠

この報告書は、国立国語研究所萌芽・発掘型共同研究プロジェクト「テキストにおける語彙の分布と文章構造」（2009年10月～2012年9月。プロジェクトリーダー：山崎誠）の研究成果の一部を論集として掲載するものである。本共同研究の趣旨は、テキストの産出過程とともに形成される動的な語彙を文章構造との観点から定量的な手法で分析することであった。具体的には、『現代日本語書き言葉均衡コーパス』に含まれるひとまとまりの完結したテキストあるいは学術論文等を用いて、語の使用頻度と出現状況との関係、とくに文章構造と語（内容語、機能語）の出現状況との関係を語彙的結束性の現れという観点から探った。また、当該テキストの持つ特性（表現意図、ジャンル、文体等）との相関を調査・分析し、語彙に内包された文章構成機能を明らかにすることを目的とした。

共同研究員は、次のとおり（五十音順）。内山清子（国立情報学研究所）、大塚みさ（実践女子短期大学）、金明哲（同志社大学）、江田すみれ（日本女子大学）、小森理（統計数理研究所）、清水まさ子（国際交流基金日本語国際センター）、高崎みどり（お茶の水女子大学）、馬場俊臣（北海道教育大学）、馬場康維（統計数理研究所）、村田年（慶應義塾大学）

本報告書に収めた論文を簡単に紹介すると、専門用語の出現傾向から論文の構造との関係を分析した内山論文、受け身文の現れ方がジャンルによって異なることを明らかにした江田論文、学術論文における問題提起疑問文の機能を分析した清水論文、談話構成語や指示語などのテキスト構成機能を持つ語の詳細な分析を行った高崎論文、接続表現の二重使用の分布と組み合わせについて分析した馬場（俊）論文、学術論文を品詞を手掛かりに状態空間表現という概念を用いて分析した馬場（康）・小森論文、慣用句をもとにして学術論文のジャンルを判別した村田・山崎論文、同じく複合動詞の使用傾向から学術論文のジャンルを判別した村田・山崎論文、隣接段落間で共通して現れる語の計量的分析から文章のジャンルや文章構造との関係を記述した山崎論文、同じく文章中のすべての段落の類似度をもとに文章の結束性を観察した山崎論文の10論文である。これらの多くは後述の共同研究発表会で発表したもの、あるいは、共同研究の成果として「コーパス日本学ワークショップ」で発表したものをまとめたものである。

本共同研究では以下の日程で共同研究発表会を開催した。

第1回：2010年 2月25日 国立国語研究所

- ・テキストにおける多義語の意味分布と語彙的結束性（山崎誠）
- ・科学的な書物における「ている」の使われ方ー「運動長期」「パーフェクト」の果たす「話題提供」「結論」の機能についてー（江田すみれ）

第2回：2010年 8月23日 同志社大学文化情報学部

- ・出現間隔と意味的距離から見た多義語の意味分布（山崎誠）
- ・社会科学系論文の本論における構成要素間の結びつき（清水まさ子）

第3回：2010年12月 5日 日本女子大・目白キャンパス

- ・学術論文の本論における論の展開—構成要素のつながりから見る—（清水まさ子）
- ・「させる」文の文脈の違いによる使用状況について—会話・小説・科学的入門書のコーパスによる調査結果—（江田すみれ）
- ・テキストにおける語彙的連鎖（山崎誠）

第4回：2011年 3月 6日 国立国語研究所

- ・「手」の慣用句を指標とした文章の所属ジャンル判別の可能性—現代日本語書き言葉均衡コーパスを用いて（村田年）
- ・接続表現の二重使用と文章ジャンル（馬場俊臣）
- ・文節の語彙属性パターンに基づいた文体分析（金明哲）

第5回：2011年 6月26日 国立国語研究所

- ・語彙の分布の視覚化（山崎誠）
- ・論文の構成要素の分布から学術論文のタイプを見る（清水まさ子）
- ・文章と語彙の関わりについて（高崎みどり）

第6回：2011年 9月24日 北海道教育大学札幌駅前サテライト

- ・文章における共起語率の分布（山崎誠）

第7回：2012年 7月 1日 お茶の水女子大学

- ・学術論文における専門用語の分野基礎性に関する一考察（内山清子）
- ・状態空間表現を用いた文章の特徴付けの試み（馬場康維・小森理）

第8回：2012年10月28日 国立情報学研究所

- ・接続表現二重使用のジャンルによる偏り—BCCWJ「中納言」検索結果に基づいて—（馬場俊臣）
- ・学術論文の分野・タイプによって構成要素の出現はどう変わるのか— 一編の論文が持つ異なる属性に注目して—（清水まさ子）
- ・明示的な構成を持つテキストの語彙的特徴（山崎誠）

なお、上記とは別に名古屋大学で開かれた2012年日本語教育国際研究大会で以下のパネルセッションを行った。

パネルセッション「構成要素の出現状況に基づく文章・談話の構造分析」（2012年8月19日）

- ・テキストのタイプ・構成と語彙的結束性との関係（山崎誠）
- ・多義語の出現環境と意味実現との関連（大塚みさ）
- ・『ていた』のテキストの違いによる機能の違い（江田すみれ）
- ・論証型論文における構成要素の特徴とは何か（清水まさ子）

論文の論理構造における分野基礎用語に関する分析

内山 清子 (国立情報学研究所) [†]

An Analysis of Domain-Specific Introductory Terms in Logical Structure of Scholarly Papers

Kiyoko Uchiyama (National Institute of Informatics)

要旨

本論文は、学術論文に含まれる多くの専門用語の中から、分野において必須で重要な用語を分野基礎用語と定義し、その用語の出現傾向について分析を行う。分野基礎用語は特定分野の研究をこれから学ぶような学部の学生は、専門が異なる研究者などに対して、効率的に分野の論文を理解するために、最低限知っておくべき用語を提示することを提案する。この分野基礎用語をどのように選定すべきであるのかについて、様々な観点を想定し、その観点を実際の文章に当てはめて分析を行った。また分野基礎用語が、論文中にどのような出現傾向を示すのか、特に文章の論理構造においてどのような役割を果たしているのかについて分析と考察を行う。

キーワード： 専門用語、分野基礎性、分野基礎用語、論理構造

1. はじめに

学術論文には分野で使われる専門用語や、著者が自分の研究を特徴づけるために作り出す独自の専門的な複合語などが数多く含まれる。これらの語は、分野の初心者にとって初めて遭遇する用語であり、その用語の意味を理解した上で論文を読み進めることが必要となる。しかし、分野初心者にとって、専門用語はすべて未知の語であり、どの語が重要な語であり最初に学ぶべき用語であるのか、また対象論文の研究内容の手がかり語となる用語であるのかなどの区別ができない。こうした専門用語に対して優先度を示すことにより、分野初心者が論文を読んで理解するための支援になるのではないかと考えた。そこで、本研究では、対象分野において最初に必ず学ばなければならない語、その分野における基礎的・必須である専門用語を分野基礎用語と呼び、分野基礎用語の選定方法を検討し、論文の論理構造における出現分布について分析を行う。

2. 関連研究と分野基礎用語の位置づけ

従来、分野の用語（専門用語）については、専門性や重要性といった指標や関連用語収集などのテーマで研究がおこなわれてきた。まず、専門度を推定する研究として、専門外の人に対して専門用語を使わずに平易な用語に置き換えるために、専門外の人から見て比較的専門的な用語か、かなり専門的な用語かの2段階に分けたものがある。次に用語の重要性については、複合語を構成している単語の種類や隣接する単語の数をベースにして用語らしさとしての重要性を計算する手法が提案されてきた。また、関連用語収集として、複数の書籍に共通する用語をシードワード

[†] kiyoko_at_nii.ac.jp

に設定して、その用語から関連する用語を自動的に収集する研究が行われた。この研究におけるシードワードは、本研究における分野基礎用語と一部一致している。

本研究において、論文を理解するために効率的な用語として分野基礎用語を位置づけるために、分野基礎用語から始まり専門性・難易度が高い用語に至る学習段階を想定し、自分の知識と目標レベルに応じた以下の4段階の知識・学習レベルを設定した。

- (1) 一般、大学学部生、他の研究分野の研究者
- (2) 大学学部生（その分野を専門に学びたい学生）
- (3) 大学院修士（修士論文テーマ探し）
- (4) 大学院博士、研究者（博士論文、研究論文テーマ探し）

まず、第一段階の一般、大学学部生、他の研究分野の研究者に対しては、分野知識を持っていないことを前提として、分野の全体的な概略を説明した解説文や理解しやすい教科書などに掲載されている用語を提示することが有効であると考えられる。次は学部3年生を想定して、卒業論文をまとめるために必要な分野の成り立ちも含めた詳細な概要を把握する必要がある。この段階では分野でよく利用される用語の理解を深めることが重要となる。第3段階は、大学院修士の学生が自分の修士論文のテーマを探すために、その分野の最新動向も踏まえて、興味のあるトピックに関する論文を読む必要性が出てくる。この段階では、論文を読むために、よく使われる用語に関連した専門性の高い用語を学ぶ。最後の段階では、大学院博士課程の学生や研究者として、過去の詳細な研究成果も含めた狭く深い情報が重要となってくる。この段階では、分野の中の特定のトピックに対する専門家が使っている専門性と難易度の高い知識を持っていることが前提となる。本論文では、このような4つの知識・学習段階を考えた中で、分野初心者に必要な最初のレベル（1と2）に必要な用語を分野基礎用語と位置付ける。

3. 基礎性判定の観点と尺度

3.1 優先度

学問を学ぶ時の学ぶ優先順位がある程度決まってくる。自然言語処理の場合は、形態素解析を学んだ後で、構文解析、構文解析を学ぶといった優先度のことである。こうした学習において共通して初期の段階に優先的に教えられる項目は特に重要で、分野基礎性が最も高い用語として絶対的な尺度であると考えられる。たとえば、複数の教科書に共通する用語や同じ研究分野の大学講義等で複数の先生が共通して初期に教える用語などは優先度が高い語であると考えられる。

3.2 経年推移度

昔は論文等で頻繁に使われていた用語が年数を経るごとに頻度が減ってきたり、その反対に増えてきたりする用語がある。分野基礎性が高い語は、年数経っても平均してある一定以上の頻度を保って出現し、分野基礎性が低い語は突然爆発的に使われたとしてもある時期に落ち着いて、以後使われなくなったりする等、出現頻度に安定性がないと考えられる。

3.3 親密度（頻度）

重要な語は、文章で繰り返し使われる語であり、様々な指標に頻度情報が含まれているのはその理由からである。論文や書籍などで頻繁に使われる用語は重要であるから繰り返され、繰り返されることによってその語に親しみを感じ、より使われるようになる。分野基礎性においても同様のことが言え、分野において重要な概念であるため繰り返し出現する用語は、その分野の研究者が親しみ、馴染みのある語として頻繁に用い繰り返される。頻度が高い語は、分野において親密度が高い語であると言える。この傾向から、頻度情報は分野基礎性においても重要な観点であると考えられる。

3.4 下位分野偏り度

同じ分野でもいくつかの下位カテゴリに分類することができ、言語学の場合は形態論、統語論、意味論、語用論などが下位カテゴリとなる。自然言語処理の場合は、人工知能の他、人工知能の一分野としての機械学習、自然言語処理の応用システム（情報検索、機械翻訳、質問応答）、認知科学等複数の分野が関連している。特に言語学は、ある言語現象について自然言語処理を用いて検証することがあるため、関連が深い。また、応用分野では自然言語処理技術を用いていることから、これらの応用にも基礎的な部分が共通している。

複合領域の分野および下位カテゴリの用語をどの程度分野基礎性が高い用語に含めるかも難しい問題である。特定研究分野全体に広く一定した頻度で用いられている用語は分野基礎性が高く、一方で、下位カテゴリの中でしか使われない用語は、分野基礎性はそれほど高くないと考えられる。

教える側から考えると、隣接領域の中でも基礎的な用語について万遍なくカバーして提示するのが必要である。一方、学ぶ側では、背景知識を持った人間であれば、たとえば、言語学を専攻している学生は、言語学の知識があるのでそれは知る必要がなく、その他の領域の知識を探す必要がある。その場合、自分の知識に欠けている下位分野の用語の中から基礎性が高いものから学ぶのと効率が良い。その場合、対象とする分野と下位分野の関連性強さを基準にして考える必要がある。言語学であれば形態論、統語論、生成文法のうち、自然言語処理との関連の強い用語が自然言語処理の分野基礎性の高い用語に含まれることになる。つまり、下位分野単独における基礎専門性ではなく、上位分野との関連性が強い語が重要な語となる。

3.5 語構成度

分野基礎性が高い用語はその用語単独でも多く出現するが、前や後ろに様々な語が接続して多くの新規複合語を構成していることが予測できる。たとえば、「機械翻訳」という用語の場合、後ろに「システム」が結合して「機械翻訳システム」、前に「統計的」が結合して「統計的機械翻訳システム」など、様々な派生の専門用語および新しい複合語（いずれは専門用語として認識され

るものも含む)を生成することができる。この基準は重要度計算の時でも利用されているが、どれだけの語と接続する可能性があるのかで、その基となる用語の重要性が計算できる。ある用語が新規の専門用語を構成している数が多ければその用語の概念は重要であると考えられる。

3.6 定義明確度

定義明確度として、分野基礎性が高い用語はまず定義を明確に述べるが行われる。たとえば、「形態素解析とは、与えられた文を形態素の単位に分割し、その文法機能（一般には品詞および活用情報）を同定する処理を言う。」というように手掛かり語「とは」を用いて定義付けを行う。このように「AとはB」のAに当たる用語は分野基礎性が高いと考えられる。上記の6つの観点のうち、定義名義度を除いた5つの観点とその尺度について表1に示す。個々の尺度は、それぞれが相互に関連しているものとする。

4. 分析

4.1 対象データと抽出単位

3章における、基礎性判定における観点について、指標の妥当性を検証するために、さまざまなリソースにおける出現頻度、重複度について分析を行った。できるだけ電子的に利用可能なリソースを収集し、分野基礎用語の各リソースにおける出現パターンを調べた。

リソースとして実験的に、教科書や講義資料、事典、論文を題材とした。まず教科書には必ず基礎的な用語が含まれており、単語の頻度ではなく各資料に共通した用語が重要であると仮定する。事典については、特定分野の知識として必須情報を獲得することが可能で、各章の出現頻度やどの章に出現したのかが重要となってくる。一方、論文については、より専門的となり、研究における基礎的な用語や研究動向、流行り廃れなどについて観察することが可能であるとする。ここでは、出現頻度と出現年数が重要であるとする。

このように各リソースでは特徴も重要な観点も異なるため、リソース別の抽出単位と頻度計算の違いを分けた。文書頻度については、各教科書、講義資料で共通する単語の頻度とした。教科書では、索引語の教科書間、講義資料では、講義資料間の重複を調べた。出現頻度では、テキスト中における出現頻度として、論文においては、抄録に出現する単語数、事典では全テキスト中に出現する単語数について調べた。

4.2 分析結果

まず、優先度分析では、教科書や書籍、講義資料で共通して用いられる、あるいは最初に説明される(出現する)用語は基礎性が高いものが多いという仮説のもと、1996年から2007年に出版された自然言語処理の書籍の索引語と、講義資料は自然言語処理関連講義から3講義の資料を使ったが、講義資料を入手するのが困難であるのと、著者や講義者によって重点的に説明する箇所が違っているなど、共通事項が少なかった(表1)。

次に経年推移度では、長期間出現し、頻度が高い語は分野基礎性が高いと想定し、情報処理学

会自然言語処理研究会で 1993 年から 2006 年まで 14 年間の抄録データ 1 MB 分について分析を行った。結果としてはほぼ仮説のとおりであった (表 2)。

親密度、頻度については、言語処理学事典と論文抄録で調査を行ったが、出現年数が長いと累計することで頻度が高くなり、出現年数で頻度を割るとはやりの語が交じることがあった。

表 1：講義資料と教科書の上位頻度語

講義資料の用語	教科書の用語
形態素解析	意味ネットワーク
構文解析	格フレーム
格文法	形態素解析
文脈自由文法	シソーラス
機械翻訳	格文法
LFG	形態素
GPSG	深層格
HPSG	表層格
機械翻訳システム	接辞
情報検索	自然言語処理

表 2：経年推移度

用語	頻度	年数
コーパス	477	14
機械翻訳	197	14
形態素解析	188	14
類似度	149	14
文字列	129	14
構文解析	126	14
情報抽出	118	14
翻訳システム	108	14
情報検索	108	14
再現率	98	14

語構成の分析では、多くの専門用語を構成している語は分野基礎性が高いという仮説のもとで、言語処理学事典の索引語を調査した。抽出単位は索引語を構成している 2 グラムの頻度計算を実施した。その結果として索引語は統制されたリストのため、あまり複合語のバリエーションが登録されていなかった (表 3)。本文中では基礎的用語に様々な形式で複合語を合成していたため、語構成調査は、本文のデータが必要であることがわかった。

定義明確度では、定義文で説明される用語は分野基礎性が高いと予測して、言語処理学事典の本文から、手がかり語「～とは」を使った定義文の～に該当する 194 語を抽出した。事典では、重要な用語に対して定義が必ずされていると思ったが、実際はその例が少なく (表 4 に例を示す)、また定義をしていたとしても、「～とは」という定型表現で定義付けをしているわけではなかった。自動抽出にひっかからなかった例も多かった。また、定義文を論文に適用してみることを少し試みたが、論文ではほとんど定義をしていないことがわかった。定義していたとしても、それは自分の研究で重要な個別の用語であるなど、特殊な例が多く、基礎的な用語に対して、定義付けをしていないことがわかった。この結果から、論文の場合は、分野知識を共有していることを前提としているため、あえて基本的な用語に対して定義することがないことがわかった。

以上のように、各基礎性判定に関わるであろう、観点にしたがって、様々なリソースを使って

分析を行った。この分析は、自動抽出が可能であるかどうか視野にいれながら行ったが、実際分析をした結果としては、仮説通りの結果であるにしても、観点が多いことや、その観点を自動抽出のスコアリングにどう反映すれば良いのかが非常に難しい。今回は、分野基礎用語をまずは決めてしまい、その用語がこれまで調べてきたリソースの中でどのように出現するのかを詳細に分析することに集中することとした。次の章からは、分野基礎用語の正解をどう決めたのか、また決定した基礎用語が論文中にどのように語られているかなどについて分析と考察をおこなっていく。

表 3 分野基礎用語をベースとした複合語の例

用語	生成頻度	例
コーパス	32	均衡コーパス, 話し言葉コーパス
意味論	25	語彙意味論, フレーム意味論
機械翻訳	12	機械翻訳システム, 統計的機械翻訳
アルゴリズム	12	EM アルゴリズム, ブースティングアルゴリズム
n グラム	9	n グラムモデル, 単語 n グラムモデル
句構造	9	句構造文法, 主辞駆動句構造文法
言語学	9	メタ言語学, 計算言語学
主要部	9	右側主要部規則, 主要部先行型
曖昧性	9	曖昧性解消, 構造的曖昧性

表 4 : 定義文の例

機械学習	1960 年あたりから人工知能の一分野として研究が始まった分野であり、一般に、過去の事例をもとに、それらの中に潜む構造を見出したり、将来の事例についての予測を行ったりするための技術を開発することを目指す。
機械翻訳	コンピュータ・プログラムで、テキストをある言語（原言語という）から別の言語（目的言語という）に翻訳することを指し、自動翻訳や言語翻訳と呼ばれることもある。
固有表現	人名、組織名、地名といった固有の名前を持つ対象を指す表現のことである。

4.3 分野基礎用語の選定

これまで、このように様々な観点から分野基礎用語の出現傾向を分析し、自動抽出を試みてきたが、理想的な選定方法としては、専門家に分野基礎用語を選定してもらい、多くの専門家が共通して選定した用語は分野基礎用語であると決定することが考えられる。しかし、専門家の意見を数多く集めることが難しいため、専門家の判断と同等であると見なせる客観的な基準を検討した。

そこで分野基礎用語を抽出する対象として、分析と同様に教科書、事典、論文の3種類を用意した。用語は、形態素解析を行い品詞が名詞あるいは名詞の連続であるものを抽出した。この3種類とも専門家が執筆したものであるため、これらのリソースから抽出した用語は複数の専門家の判断と同等であると考えられる。詳細は以下の通りである。

- (1) 教科書：「自然言語処理」分野の日本語の教科書 39 冊の目次に出現する用語（異なり語数 694 語）
- (2) 事典：「言語処理学事典」の目次に出現する用語（異なり語数 463 語）
- (3) 論文：情報処理学会自然言語処理研究会で発表された論文のタイトル、抄録、キーワードに含まれる用語（異なり語数 13493 語）

教科書と事典の目次に出現する用語に着目した理由として、目次は初心者にもわかりやすい表題および学んでほしい用語を必ず著者が選定する、つまり著者が考える分野基礎用語は目次に含まれると考えたためである。この3種類のリソースに共通して出現する用語は90語であり、この90語を分野基礎用語と選定した。

5. 論文の論理構造における分野基礎用語

論文の論理構造において、分野基礎用語がどのような出現パターンを示すのかを調べた。本論文における論理構造とは、「抄録」、「はじめに」、「関連研究」といった論文を構成している章に関連している意味のあるまとまりのことを指している。分析対象の論文コーパスは、分野基礎用語の選定時に利用した論文とは異なり、情報処理学会の論文誌に掲載された自然言語処理分野の論文の中から抄録で「実験」、「評価」、「精度」、精度の数値「%」などを含んでいる100論文を選んで論文コーパスとした。実験を扱った論文に絞ったのは、論理構造が比較的わかりやすく、論文の流れもある程度パターン化できるのではないかと仮定したためである。本論文では、論理構造の要素を「抄録」「はじめに」「実験」「関連研究」「おわりに」「その他」の6種類に分けた。「その他」は多くの場合、「関連研究」の記述の後から、「実験」記述の前までのまとまりを指している。

分析対象の論文コーパスを論理構造の要素に分割し、それぞれの要素の中における分野基礎用語の出現傾向を分析した。表5に出現頻度100以上の用語について、論理構造別の出現頻度を示す。なお、ある用語が別の用語の部分文字列となっている場合は（「文字」「文字列」など）、重複している数を差し引いて数えている。

最も出現頻度が高い「意味」は、一般的な文章にも使われる単語であるため、用語と見なすことが難しいが、実際に出現している文を読むと、「意味」が他の分野基礎用語と共に出現するなど、重要な役割を果たしていることがわかった。自然言語処理において「意味」を理解することが目的でもあるため、本論文では用語と扱うことに意義があると考えた。このように表1のリストを見ると、分野初心者でも意味がわかるような「品詞」「辞書」「文字」などの単語が並んでいる。これらは分野基礎用語の定義である、「必ず学ばなければならない語、その分野における基

礎的・必須である専門用語」という基準からはずれることになる。しかし、これらの単語は、研究の背景など導入部分を記述するためには必須の語、および重要な手がかり語の役割をはたしていることがわかった。

表 5 論文の論理構造における分野基礎用語の出現頻度

	抄録	はじめに	実験	関連研究	おわりに	その他	合計
意味	54	231	360	93	49	561	1348
コーパス	64	160	448	79	59	330	810
品詞	33	116	339	28	34	361	550
辞書	30	103	310	43	36	239	522
日本語	40	136	182	45	38	225	441
生成	19	80	186	46	36	324	367
未知語	15	50	167	22	20	160	274
知識	28	101	88	16	37	105	270
言い換え	17	93	99	25	29	185	263
形態素解析	25	60	131	14	20	122	250
文字	7	26	89	24	9	65	220
シソーラス	9	39	89	34	16	39	187
アルゴリズム	16	43	73	14	17	252	163
照応	7	36	65	40	6	68	154
固有表現	5	14	108	4	7	91	138
形態素	6	27	87	2	10	124	132
文字列	8	22	82	13	4	186	129
クラスタリング	7	30	66	14	7	23	124
語義	7	35	57	7	16	45	122
機械学習	16	43	25	15	13	34	112
構文解析	7	45	35	13	9	46	109
機械翻訳	14	51	22	13	8	27	108
言語処理	16	59	9	12	10	31	106
決定木	11	34	40	11	9	74	105
言語モデル	10	19	53	12	10	70	104

次に、分野基礎用語が出現する文が全体のどのくらいの割合を占めているのかを調べ、表 6 に

示す。分野基礎用語が一つの文に複数出現することもあるため、文単位での傾向を分析した。その結果、「抄録」、「はじめに」の論理構造の要素では、全体の半分以上を占めていることがわかった。次いで「おわりに」「関連研究」の要素で4割以上に分野基礎用語が含まれている。これは分野基礎用語の90語のうち頻度0を除いた74語が、「抄録」や「はじめに」などの論文の重要な部分を説明する文章に半分以上含まれるということになる。この結果を見ると、「抄録」や「はじめに」に多く出現する用語が分野基礎用語なのではないかと予測されるが、これまで行ってきた実験では「抄録」の中で高頻度な用語が、分野基礎用語にはなっていなかった。今回はこれまでと正解セットや分析対象コーパスが異なっているため、単純に比較することはできない。しかし、今回の対象コーパスが論文誌に採択された実験論文であるため、論理構造がはっきりしていることや、用語の使い方や表現も推敲を重ねるなど、質の高い文章であることから、分野基礎用語の出現傾向が特徴的になったのだと考えられる。

これまで、分野基礎用語は分野特有の専門用語で、分野初心者がその分野を理解する上で必ず学ばなければならない用語と考えていた。しかし、客観的な指標による分野基礎用語の選定および実際の論文中に出現する傾向を分析すると、必ずしもその用語自体を学ぶ必要はなく、むしろその用語が手がかり語となって周辺用語との関連により、その分野の理解を深める役割を果たしていた。つまり、分野基礎用語をベースとして、周辺用語との関連を示してあげることにより、分野初心者への論理解を手助けすることができるのではないかと考えられる。

表6 論文の論理構造における分野基礎性用語を含む文の割合

	文数	分野基礎性用語を含む文数	割合
抄録	656	362	0.552
はじめに	2448	1284	0.525
実験	8931	2701	0.302
関連研究	1222	542	0.444
おわりに	805	394	0.489
その他	11965	3439	0.287
合計	26027	8722	0.376

6. まとめ

本論文では、その分野で必ず学ぶべき用語や手がかり語となる分野基礎用語の選定基準と、実際の論文における出現パターンの分析を行った。選定の基準は、多くの専門家が執筆した本や事典の目次、論文のタイトル、抄録、キーワードの中から共通して出現するものとした。この客観的な基準に従って抽出した分野基礎用語が論文の論理構造の要素別に出現する頻度に基づいて分析を行った。

分析の結果から、今後は分野基礎用語が出現する文が研究のどのような内容を表現しているの

か（研究の背景、動機、既存研究の比較など）をさらに詳しく分析し、分野基礎用語と共起する用語との文法的関係（主語、目的語、補語、修飾語など）と意味的關係（目的、手法、対象など）を付与するなど、論文の内容理解の支援をする表現方法を検討していく。

文 献

中川裕志、森辰則、湯本紘彰(2003)「出現頻度と連接頻度に基づく専門用語抽出」、自然言語処理、Vol.10 No.1、pp.27-4

佐々木靖弘、佐藤理史、宇津呂武仁(2006)「関連用語収集問題とその解法自然言語処理」、Vol.13 No.3、pp.151-175

千田恭子、篠原靖志、奥村学(2005)「技術成果を効果的に伝える表題作成支援手法：開発と評価」、情報処理学会論文誌、Vol.46 No.11、pp.2728-2743

内山清子(2010)「専門用語の分野基礎性に関する一考察」、情報処理学会自然言語処理研究会報告、2010-NL-199(15)、pp.1-6

Kiyoko Uchiyama(2011)、「A Study for Identifying Domain-Specific Introductory Terms in Research Papers」、Proceeding of the 9th Terminology and Artificial Intelligence、pp.147-150

自然言語処理学会、『言語処理学事典』(2010)、共立出版株式会社

テキストの違いと受身文の違い

—会話・ブログ・新書の受身の使われ方をもとに—

江田すみれ（日本女子大学）

1. はじめに

近年コーパスを用いて文法項目の調査をした結果がいろいろ発表されている(森・庵 2011、堀、中俣、プラシャント 2012 ほか)。本稿は受身について、会話・ブログ・新書での使われ方を調査したものである。前田(2011)が、シナリオを資料として受身が単文末・複文末・従属節末のどこにあらわれたかを調査し、非文末が 3/4 を占め、特に連用節末が半数を超えていることを述べているが、本稿は調査対象を 3 種類のコーパスにするとともに、調査項目も増やし、会話と書き言葉での受身の違いを考察した。

2. 受身についての先行研究

前田(2011)はシナリオを資料として受身が単文末、複文末、従属節末のどこにあらわれたかを調査し、非文末が 3/4 を占め、特に連用節末が半数を超えていること、複文末に受身が来る場合の従属節では「て・たら」の節が多いこと、連用節末に受身が来る場合は「て」節が非常に多いことを明らかにした。また、文末の形式は「た」だけでなく「る」も用いられており、受身は単体でなく様々なモダリティ表現とともに出現していると、受身が実際にどのように使われているかを述べている。

しかし、前田(2011)はシナリオだけを資料としており、資料の種類が少ない。前田は自然会話は場面に支えられて成り立つ不完全な文によって構成されるものであること、話者によってバリエーションがあり、どの学習者にとってもモデルとなる自然談話を探すことは難しいこと、の理由により(p.69)、制御された話し言葉であるシナリオを会話資料として使う、と述べている。それもひとつの考え方であろう。しかし、学習者がその中で自己表現をしなければならないのは、その不完全な文によって成り立つ自然会話である。そこで本稿は自然会話を資料とし、前田論文とどのような違いがでるか調査してみた。また、新書・ブログの 2 種類のコーパスも用いて会話と書き言葉との違いを見てもみることにした。

小川・安藤(1999)は、受身は影響を受けた側からの事実の認識を述べる文であり、利害は文の構造と関係するのではなく、運用レベルで表れるとしている。指導にあたっては、常に能動文と対応させる必要はない。初級では出来事の文を扱い、ガ格で表される影響の受け手は「私」つまり話し手に固定してよいとしている。初級で直接・間接・所有物受身を提示する。中級では、コト・モノが(人に)～られることを傍らから描写した文を教える。二格は「関係者」「当事者」「担当者」などであるが、明示されないことが多い。受身は項を減らし簡潔な表現をするため、視点の統一のために用いられることを教えるとしている。

本稿は初級での教育は影響の受け手を「私」に固定してよいとする主張に疑問を持った。本当に「私」だけでいいのであろうか。また、二格が明示されることが少ないというのはどのテキストでも同様なのか、調べたい。

本稿は会話・ブログ・新書の3種類のコーパスを用いて前田調査と同様に受身の出現状況の調査、小川・安藤の論文を通して持つにいたった疑問を明らかにする調査を行い、会話と書き言葉で受身の使い方が同じかどうかを調べる。

3. 本稿の疑問

受身は実際の文脈でどのように用いられるのだろうか。具体的に言うと、直接受身・間接受身などの受身の用法はテキストの種類の違いによって使われ方が異なるであろうか。受身の使われる文中の位置はシナリオでは非文末が多いとのことだが、ほかのテキストではどうなのだろうか。複文の場合はどのような節とともに使われるのだろうか。モダリティ表現・終助詞との関係はどうであるか、ガ格名詞・行為者・作用者名詞の出現状況はどうであるか。これらについて、会話・ブログ・新書のコーパスではどのような結果が出るだろうか。テキスト¹の種類の違いによって受身の使われ方は違うだろうか、同じだろうか。

以上の調査をする中で、初級では「私が～られる」で指導することが提案されているが、すべて「私が」でいいのだろうか、初級で直接・間接・所有物受身すべてを出していいのか、という疑問に対する答えも得たいと考える。

4. 本調査の方法

4.1 コーパスについて

堀・李・江田は学習項目解析システム「はごろも」を構築しネット上で公開している。この学習項目解析システムは、使用者が入力した文章を解析すると、その中に使われている文法表現が「はごろも」プロジェクトによる文法表にそって、6段階にレベル分けされて表示されるものである。現在、<http://130.158.168.228/Checker/>で公開されている。今後は、また複数のコーパスから抽出した実際の例文も表示する計画である。

今回はその資料として用いた新書（CASTEL/J）・ブログ・会話のコーパスを使い、受身文の意味的な用法、文中の位置、文末表現、節との関係、ガ格の名詞、二格の名詞を調査し、違う種類のテキストでの受身文の出現状況を知ることが目的としている。

資料として使ったコーパスは以下のとおりである。

新書

日本語教育支援システム研究会 『CASTEL/J』の新書部分

104.8 万形態素

¹ 用語について説明する。コーパスは資料として用いた電子化されたテキスト、テキストはその内容と関係させて文章について述べる場合に用いた。

ブログ

京都大学情報学研究科・N T Tコミュニケーション科学基礎研究所共同研究ユニットによるブログ記事のデータ。

4つのテーマ（京都観光、携帯電話、スポーツ、グルメ）のブログ記事をデータ化したもの。249記事、4,186文。10.4万形態素

会話

宇佐美まゆみほか『BTS（Basic Transcription System）による多言語話し言葉コーパス』1 日本人同士の会話。89.8万形態素

4.2 受身文の分類基準

「はごろも」の文法項目の分類に従い、直接受身、直接受身で行為者が不特定多数のもの、持ち主・身体部分の受身、間接受身に分類した²。分類基準は以下のとおりである。

A 直接受身

能動文が復元できる文、つまり「対応する能動文においてヲ格名詞やニ格名詞などの補語として表される名詞を主語とし、それに伴って、能動文の主語名詞を主語以外の項として表現する受身文（日本語記述文法研究会 2009：219）」である。

本稿では、行為者が多数であっても特定できる場合は直接受身とすることとした。

(1) 悩みなさそうとかよく言われる。(会話)

(1)は自分が自分の周囲の人に言われることを述べている。この場合、誰が言ったかは話し手は知っている。こうした例は直接受身に分類した。(1)の例は主語を中心に述べており、主語の前景化(日本語記述文法研究会 2009：227)を目的としている。

B 直接受身で行為者が不特定多数のもの

行為者不明、または一般の人々であって特定できない場合を不特定多数とする。あるいは行為者を問題にする必要がない場合も行為者不特定多数とした。

(2) んー、海外市場も、あの一、えーっと、ほ、香港の市場が、えーっと、緩和されて、はい。(会話)

(3) ジュニアモードとかティーンズモードが機能として搭載されているのもうなずけます。(ブログ)

(2)(3)ともに行為者が不明である。(2)は香港の中央銀行が緩和を決定するのであろうか。(3)のように携帯にジュニアモードを搭載することは誰が行為者となるのであろうか。設計者した技術者であろうか、搭載することを判断した会社であろうか、このような場合を行為者不特定多数とした。

² 「はごろも」の文法項目は『日本語能力試験出題基準』『日本語文型辞典』『現代語複合辞用例集』『日本語表現文型：用例中心・複合辞の意味と用法』『現代語の助詞・助動詞—用法と実例—』の2冊以上に掲載されている項目を中心に選んでいる堀（2012）。

(4)から(6)のように、行為者でなく影響を与えられる側に焦点をおいて表現する、能動主体の背景化（日本語記述文法研究会 2009:231）を目的としている文もここに分類した。

(4) 本論において言及された日本社会の諸現象の多くは、断片的には、多くの人々がすでに指摘したり、また十分経験したりしていることである。（新書）

(5) お土産コーナーに1枚4000円する金板百人一首が販売されていて、買おうか一瞬迷ったが、いったいこれをどうしようか？（ブログ）

(6) 余談ではありますが、家に帰ってフル充電された携帯電話をチェックしてみるとメールが3通しか来ておらず、ちょっとした儚さを感じたことなど誰にも言えません。（ブログ）

(4)では言及したのは引用した論文の筆者であって実際には行為者が特定できる可能性がある。しかし、この文の筆者は社会現象に焦点をおき、言及した人間には関心が無い。(5)も、きちんと考えれば販売する店の人が行為者となろう。しかし、これも、受身を使うことによって行為者を背景化する述べ方をしている。(6)は自分が充電した携帯について語っているが、携帯のほうに焦点があるため受身になっている。このような例を直接受身で行為者が不明なものに分類した。これらの文では項がひとつ減少する。

C 所有物受身

ヲ格名詞やニ格名詞の持ち物や身体部分を、持ち主を主語にして述べる受身である。

(7) われわれのほうも、憲法が掲げる夢にばかり眼を奪われてはいられない。（新書）

(8) そして、SIMカードを使っているから、それをほかの人に盗まれたら、大変なことになります。（ブログ）

(7)は「憲法が掲げる夢が我々の眼を奪う」という文の受身、(8)は「ほかの人が私のカードを盗んだら」と述べている文であり、両者とも所有物受身と読める。

D 間接受身

日本語記述文法研究会(2009)は間接受身について「対応する能動文の表す事態には直接的に関わっていない人物を主語とし、話し手がその人物と事態を主観的に関係づけ、事態と間接的な関係をもったものとして表現する受身文」と定義している(p.217)。事態に直接関わらない人が影響を受けたと述べることから、能動文が復元できない文ということもできる。

(9)政治に騙されてはならないように、憲法に関しても表面的な意味以上のものを読みとらなくてはならない。（新書）

(10)携帯電話の話人ごとながら、どうしてこんな人間に買われてしまったのだろうか、気の毒になる。（ブログ）

(9)は憲法について述べている。政治家はだまそうとして政治をするのではないだろうが、この文章の筆者は政治が人をだます、人は政治にだまされるという捉え方をしており、政治と人との関係を主観的に関係づけていることから間接受身と分類した。(10)は、人は携帯のために買うという動作をしたのではないのに、携帯側では人に買われたと述べており、動作主の動作が相手に

向かった動作ではないという点で間接受身に分類した。

5. 調査結果

5.1 テキスト別受身の種類

表の出現数は、それぞれのコーパスの大きさが同じではないため（4.1 参照）、単純に比較することはできないが、約 90 万形態素の会話中の受身文 66 件に対し、約 100 万形態素の新書中のそれが 698 件ということから、新書での受身文の多さが読み取れる。

表 1 によると、会話は直接受身が多いが、新書は 80% が不特定多数による受身であり、口語的な書き言葉であるブログでも 70% が不特定多数による受身であった。会話とブログ・新書での受身の使われ方の違いが見られる。

	会話		ブログ		新書	
直接受け身	51	77.3%	35	16.5%	77	11.0%
直接受け身: 不特定多数	11	16.7%	149	70.3%	560	80.2%
持ち主・身体 部分の受け 身	1	1.5%	6	2.8%	23	3.3%
間接受け身	3	4.5%	22	10.4%	38	5.4%
合計	66	100.0%	212	100.0%	698	100.0%

間接受身は会話・新書で少ないが、ブログでは 10% を越えている。所有物受身はどのコーパスでも少ない。間接受身を初級から教えることについては、会話での使用頻度の低さを見ると、検討の余地があると言えるだろう。

5.2 ガ格の名詞

小川・安藤(1999)によって初級ではガ格名詞を「私」に固定して指導することが提案されている。本稿では会話・ブログ・新書においてガ格名詞がどのようなになっているか調べた。

表 2 は「私」関係の名詞がどの程度ガ格に用いられているかを示している。

	会話		ブログ		新書	
私、私/人、 我々	56	84.8%	56	26.4%	24	3.4%
それ以外	10	15.2%	156	73.6%	674	96.6%
合計	66	100.0%	212	100.0%	698	100.0%

表 2 では、(11) のように個人的な「私」がガ格に使われるもの、(12) のように「私」あるいは「人」一般と読める例、(13) のように個人的な「私」ではなく一般的な人としての「我々」を表現する例をすべて「私」関連の名詞として採用した。

(11) もっと、ていうかねー、" もっと早めに連絡しなさい " (うん) って言われたく
2人笑い>。(会話)

(12) いつでもどこでも人を呼び出せることができるというのは、いつでもどこでも人

から呼び出されるということなのです。(ブログ)

(13) そうすれば論旨は通るのだが、どちらの立場をとってもひどく架空な結論に導かれる。(新書)

表2は「私」あるいは「我々」を含む広い「私」関連の名詞の出現結果であるが、表3は「私」だけを選んだものである。個人的な「私」の視点で事態を述べている文は表3のような出現状況であった。

	会話		ブログ		新書	
私	55	83.3%	45	21.2%	14	2.0%
それ以外	11	16.7%	167	78.8%	684	98.0%
合計	66	100.0%	212	100.0%	698	100.0%

ガ格名詞に個人的な「私」が占める割合は、会話では83%、ブログでは20%強、新書では2%程度である。初級日本語は日常会話ができるようになることを目的とすることを考えると、小川・安藤(1999)が初級で「私」に視点をおいた受身文を教えることを提案するのは実際の用例によって支持されたといえる。

しかし、「私」の述べ方は会話と新書で多少異なるようである。会話の受身では、「私」関連の主語の文66例中55例が個人的な「私」を主語としている。

(14) やー、だから、普通に話し掛けられればさー (あーはいはいはいはい)、先輩とか普通にしゃべるじゃん。(会話)

(15) 一部始終撮られてんの<笑い>。(会話)

しかし、ブログ・新書ではこうした個人的な「私」の視点で物事を述べている文は、ブログでは「私」の例56例中45例とある程度あるが、新書では24例中14例と半数強であった。それ以外は以下の例のように「私たち」「我々」などがガ格と読める例であった。

(16) 道端で、片言の日本語、あるいは英語で道を尋ねられることもわりと経験しうる光景です。(ブログ)

(17) ここに現われる生命・自由・幸福追求が、独立宣言からのまる写しであるから、起草者としては基本的人権のつもりだったのだろうが、われわれとしてはそうした思惑に拘束される必要はない。(新書)

新書では「私」関連の名詞を主語とする文が少ないだけでなく、「私」の視点で述べる文はかなり少ないという点で会話と異なる。

表4はガ格名詞が有情物であるか無情物であるかを調べたものである。これによると、会話は90%が有情物を主語とする受身、新書は90%近くが無情物を主語とする受身であることがわかる。

	会話		ブログ		新書	
有情	58	87.9%	71	33.5%	96	13.8%
無情	8	12.1%	141	66.5%	602	86.2%
合計	66	100.0%	212	100.0%	698	100.0%

会話はほとんどが「私」を主語とする受身であるのに対し、新書は有情物をガ格とする受身文が全体の1割強で、コト・モノをガ格とする受身文が一般的に用いられているといえる。小川・安藤(1999)が中級では「コト・モノが人に〜られる」ことを傍らから描写した文を教えることを提案しているのは、実際の用例によって支持された。

5.3 単文・複文の出現状況

次に受身が文のどの部分で使われているかを調査した。

前田(2011)にならい、単文末、複文末、従属節内と分類した。

単文の出現状況は出現割合の高い会話でも20%程度で、ブログ・新書は約10%、複文が80-90%となった³。会話の複文の割合80%と比較すると、ブログ・新書では受身文は複文の割合がやや高いといえるかもしれない。

複文中では、3者とも従属節内の使用が60-70%と、文末より従属節内での使用が高い⁴。

会話では「〜って言われて。」「〜って言われたんだけど。」「去年散々言われたし。」のように「て」「し」などの従属節で言い終わる形式が多く見られた。表5はこれらを従属節内と分類した結果である。会話では従属節内の例47例中、従属節で言いさす文が18例であった。

		会話		ブログ		新書	
単文		13	19.7%	21	9.9%	73	10.5%
複文	従属節	47	71.2%	149	70.3%	445	63.8%
	複文末	6	9.1%	42	19.8%	179	25.6%
不明		0	0.0%	0	0.0%	1	0.1%
合計		66	100.0%	212	100.0%	698	100.0%

前田(2011)は日本語教科書の受身の例文がどのような形で提示されているか調査した結果を以下のように述べている。

日本語教科書 『みんなの日本語』 82例中単文末が81例、連体節末が1例

『教師と学習者のための日本語文型辞典』 単文末42%、連用節末38% (pp.76-79)

『文型辞典』は比較的現実の受身の使用状況と似ているが、初級の教科書は形を教えることに集

³ 新書の文章は元来複文が多いと考えられる。そこで、『CASTEL/J』の新書の中から人文科学・社会科学・自然科学の3作品をランダムに選び、それぞれから1万字を抽出し、合計3万字のテキストを作り単文、複文の割合を調べてみた。すると

単文 121文 (19.7%)

複文 494文 (80.3%)

で、複文は約80%となった。この結果から新書の文章では複文が多いことが推定できる。

⁴ 表中の「不明」というのは次の文である。

・ベッドの下に置いてあったアルバムまで引っ張り出され、…… これは、私がどこに改行の印を入れたか、たぶんみなさんお分りでしょう。

上のように、例文として文章中に出された文で、文自体が完成していないものであり、意図がわからなかったので、「不明」とした。

中しているためか、ほとんどが単文で示されており、中級あたりで現実の使用状況に近い使われ方を見せ、文型を「拡大」していく必要があるとしている。本稿の調査からも同様のことが言える。

5.4 文末の形

単文末、複文末における述語の形とテンス・アスペクト、補助動詞、モダリティ、終助詞を調べた。

文末に受身が使われているものを「述語のみ」「補助動詞・モダリティ・終助詞つき」「従属節による言い終わり」の3種類に分類した。

新書では表5の「単文」と「複文末」を合計した数が表6の「合計」にあたる。しかし、会話とブログは表5の単文・複文末の合計とは異なっている。従属節で言い終わる形になっているものがあり、それらを「従属節内による言い終わり」という文末の形として採用したためである。

	会話		ブログ		新書	
述語のみ	8	21.6%	42	55.3%	144	57.1%
補助動詞・モダリティ・終助詞つき	11	29.7%	17	22.4%	108	42.9%
従属節による言い終わり	18	48.6%	17	22.4%	0	0.0%
合計	37	100.0%	76	100.0%	252	100.0%

(18) カバに水かけられたけど。 (ブログ)

前田(2011)の結果同様、会話では述語のみの形は少数派である。補助動詞・モダリティ・終助詞の付いた形がある程度あり、それより多いのが従属節による言い終わりの形であった。

(19) 去年散々言われてたし。(会話)

一方、書き言葉を見ると、述語のみはブログで55%、新書で57%であり、書き言葉では述語のみで終わる形は受身文の半数以上ある。会話と書き言葉では述語のみの割合は異なることがわかる。しかし、補助動詞・モダリティつきの文は新書で43%となり、これらの表現も重要性が高い。ブログでは従属節による言い終わりが補助動詞・モダリティつきと同程度使われている。

次にテンス・アスペクトについて見る。述語のみで終わっている文末のテンスを「る」「た」「ている」「ていた」で分類した。

今回資料とした会話では「る」と「た」が出現した。しかし、述語で言い切る形が少ないため、これだけの例から結論を出すのは難しい。しかし、日本語教科書の例のように「た」が多いということはないようである⁵。

⁵ 前田(2011)では『みんなの日本語』の例文82例中文末が「た」の例は60例とのことである。

	会話		ブログ		新書	
述語のみ(る)	4	10.8%	6	7.9%	53	21.0%
述語のみ(た)	4	10.8%	11	14.5%	15	6.0%
述語のみ(ている)	0	0.0%	19	25.0%	67	26.6%
述語のみ(ていた)	0	0.0%	6	7.9%	9	3.6%
補助動詞・モダリティ・終助詞つき、	11	29.7%	17	22.4%	108	42.9%
従属節による言い終わり	18	48.6%	17	22.4%	0	0.0%
合計	37	100.0%	76	100.0%	252	100.0%

新書は「る」が23%「た」が6%と「る」の割合が比較的高い。科学的なテキストでは「る」が主体(仁田 2009、江田 2013)と言われるが、その性質がここでも見られる。引用あるいは過去の事態を現在に関係させる働きをもつ「ている」(庵 2001、江田 2013)もよく使われている。

(20) ヨーロッパの場合は精神的姿勢が外側を向いているので、こんな場合どうしても外部のほうがより強く、しかも具体的相貌をもって認識される。(新書)

(21) ところで、今日の最初にお話しした保田与重郎の「日本の橋」には、朝鮮の慶州あたりに伝わる或る橋の伝説も誌されています。(新書)

(20) はヨーロッパの姿勢について一般化して述べる例である。(21) は引用の例である。

ブログは「ている」がやや多く、次いで「た」、そのほかは「る」「ていた」が同じ程度に使われている。

(22) 栗や秋刀魚、きのこなど旬の味覚がどこのスーパーでも安く売られている。(ブログ)

(23) 実際こういう観光名所に行ってみると、間違いなく同じガイド本を手にした人で埋め尽くされている。(ブログ)

ブログでの「ている」の文は(22)(23)のようにそのときの状態を描写する文が多い。

会話・ブログで約30%、新書で約40%を占める、補助動詞やモダリティ形式の付いた形はたとえば以下のようなものである。

(24) カンフォラに友人とランチを食べに行った際、レジ際に置いてある栗を何気なく触って、見事に刺されてしまったのである。(ブログ)

(25) するとなぜ、神社神道を国教化したことが、日本国憲法では否定されるにいたったのだろうか。(新書)

(24)は「てしまう」「のだ」が、(25)は「にいたる」「のだ」「だろう」が接続している。

表8は文末によく用いられるモダリティ・終助詞をあげたものである。「のだ」はどのテキストでも用いられており、頻度も高い。前田(2011)ではシナリオでは95の文末表現のうち10例が「てしまう」だったと述べられている(p.71)が、今回の会話では「てしまう」「ちゃう」は1

会話		ブログ		新書	
て。	12	のだ	8	のだ	35
のだ	5	てしまう	4	ことになる	17
けど。	4	ようだ	4	てしまう	7
たら。	3	だろう	2	なくてはならない	6
よ	3	らしい	2	からだ	5
し。	2			てくる	4
ね	2			ものだ	4

例と少なかった。しかし、ブログ・新書では用いられていた。新書では「ことになる」「なくてはならない」「からだ」「てくる」「ものだ」がよく用いられていた。「ことになる」「なくてはならない」の新書の例を下に挙げた。

(26) したがって、天皇制においても、古来国民が純粋理念の高みで天皇に求めたものが適切に表現されたことになるというのである。(新書)

(27) 財政処理の権限についても同様であるが、これについては、国会が中心かと思わせる条文の印象にまどわされないようにしなくてはならない。(新書)

文末の表現は会話・ブログ・新書では「のだ」は共通だったが、そのほかの表現はテキストによって述べ方に違いがあるようである。江田・小西(2007)で、会話は相手との関係を考慮した表現、新書は論理的な表現が用いられると述べているが、今回の調査でも新書と同様の傾向が見られた。

前田(2011)が指摘するように、単純に受身だけ学ぶのではなく、補助動詞やモダリティ要素とともに使えるようになることが必要と言えるが、テキストによって強調すべき形が少しずつ異なるので、教える際にはジャンルの違うテキストを使うなどの工夫があると効果が上がるであろう。

文末に補助動詞やモダリティがつくのは受身に特有なのか、文末表現一般的なのかはさらに調べてみなければならない。

5.5 従属節の種類

複文の従属節を、前田(2011)の方法にならい、連体節、連用節、疑問節、引用節と分類した。シナリオでは、文末に受身が使われる場合の従属節は「～て」「～たら」などの連用節の割合が高い(前田2011)と報告されているが、今回の調査対象の会話・ブログ・新書ではどうであろう。受身が複文末に来る場合の従属節の場合と、受身が従属節として使われている場合に分けて述べる。

5.5.1 文末に受身が使われている複文中の従属節

新書では複数の節が組み合わされた文が多い。どのような節が受身と共に使われるか、には、受身と意味的に直接関わる節を選んだ。

(28) 理念が市民革命のままなので、現実との通り路を見失い、人権には公共の福祉に

よる制約ばかりでなく、予算上の制約もついてまわることが忘れられているのである。

(新書)

例えば (28) では「理念が市民革命のままなので」「現実との通い路を失い」「人権には公共の福祉ばかりでなく」の連用節と、「予算上の制約もついてまわることが」の連体節が組み合わさった文である。これらの節のうち「忘れられている」に最も関係する節は「予算上の制約もついてまわることが」であろう。そこで、この文では受身と関わる節として「連体節」を採用した。この

	会話		ブログ		新書	
連用節	1	16.7%	30	71.4%	120	67.0%
連体節	0	0.0%	5	11.9%	36	20.1%
引用節	5	83.3%	7	16.7%	11	6.1%
疑問節	0	0.0%	0	0.0%	12	6.7%
合計	6	100.0%	42	100.0%	179	100.0%

ように整理したものが表9である。

会話は複文末の例が少なく6例しかないが、今回の会話では80%以上が「言われる/言われた」に前接する引用節であった。

(29) 絶対注意した方がいいよとかいって言われた。 (会話)

ブログ・新書は70%ほどが連用節で、新書は20%、ブログは12%が連体節であった。表10でよく使われる形を見ると、連用節では、ブログと新書は「連用形・て・と・ば」など、上位のものがある程度共通だった。

会話	ブログ	新書
て	1 連用形	8 連用形
	て	8 て
	と	3 が
	が	3 から
	ので	3 ため
	ば	2 と
	から	1 ば

(30) 大人にはやるべきことが生じ、それぞれのコミュニティから様々な要求が容赦なく突きつけられるのです。(ブログ)

(31) 本文に挙げた定義がアメリカの通説であり、日本でも採用されているところである。(新書)

連用節が多いのは視点を統一することが目的となっているためといえよう。(30) では「大人」に「やるべきことが生じる」ことと「大人」は「要求をつきつけられる」ことが述べられており、視点を統一するため、主節の行為者を背景化した受身が用いられている。

また、新書では連体節の割合が比較的高いといえる。

(32) するとなぜ、神社神道を国教化したことが、日本国憲法では否定されるにいたったのだろうか。(新書)

(33) そこでは規則を制定したり、さまざまな措置を決定したり、紛争解決のための判

断を下したりすることがおこなわれる。 (新書)

(34) 財政処理の権限についても同様であるが、これについては、国会が中心かと思わせる条文の印象にまどわされないようにしなくてはならない。 (新書)

(32) (33) は連体節が主語となり事柄を表している。(34) はガ格が人、連体節が二格である。(32) (33) のようなガ格に無情物が来て行為者が有情物である受身について、日本語記述文法研究会 (2009) は「通常は話し手が視点を重ね合わせにくい無情物をあえて主語とし、視点を重ねやすい有情物はあえて主語からはずす」(p.229) 形の受身文であり、こうした文では能動主体の背景化が重要であると述べている。この「コト・モノが人に～られる」文の中で、ガ格となっている無情物を連体節によって詳しく表現しているのが (32) (33) の形と言える。通常は視点を重ねにくい無情物であるが、詳しく述べることによって意識化して主語としていると見ることが可能であろう。

従属節の表れ方は、会話はあまりにも例が少なく比較できるかどうか分からないが、今回の会話では引用節が多かったことが特色であろう。ブログ・新書を比較すると、今回は両者に共通に連用節が多く見られた。共通する連用節は「連用形」「て」「と」などであった。また、新書の特色は連体節の多さと複数の節の組み合わせられた文の多さであった。

5.5.2 受身が従属節として使われる場合

受身が従属節として使われる場合の受身の節は表 11 のようであった。会話・ブログ・新書の三者に共通して連用節が多い。会話では連用節が 85%なので、ほとんどが連用節といっても過言で

	会話		ブログ		新書	
連用節	40	85.1%	71	47.7%	205	46.1%
連体節	4	8.5%	73	49.0%	221	49.7%
引用節	2	4.3%	3	2.0%	12	2.7%
疑問節	1	2.1%	2	1.3%	7	1.6%
合計	47	100.0%	149	100.0%	445	100.0%

会話では連用節19例・疑問節1例が節によるいい終わり

はない。ブログ・新書では連用節・連体節がそれぞれ 50%弱を占めているという点が会話との違いである。

(35) 何 (あ)、 " 何でスペイン語やってるんだ " っていわれても、やっぱり、答えられないけど…。 (会話)

(36) [グルメ] 河原町の居酒屋この間先輩たちにつれられ、河原町の居酒屋へ行ってきました。 (ブログ)

(37) この日本列島における基本的文化の共通性は、とくに江戸時代以降の中央集権的政治権力にもとづく行政網の発達によって、いやが上にも助長され、強い社会的単一性が形成されてきたのである。 (新書)

従属節に受身が使われるのは、先行研究 (日本語記述文法研究会 2009、前田 2011) で言われ

てきたように、視点の統一、主語の統一という点が大きい。特に会話・ブログでは話者・書き手の視点から物事を述べるということが例 (35) (36) などに表れている。新書では (37) に見られるように主語の統一ということが大きく作用している。

連用節に使われる形の頻度の高いものを表 12 にあげた。ブログと新書は共通のものが使われ

会話		ブログ		新書	
て	21	て	15	連用形	56
から	3	連用形	14	て	29
と	2	が	5	が	26
けど	4	と	4	から	10
たら	3	ので	4	ば	10
ても	3	けど	4	ため	9
し	2	たら	4	ても	8
ば	1	ば	3	と	7
		ても	3	ので	7

ているが、会話では連用形・「が」がなく、「たら」が使われるなど話し言葉は多少違うことがわかる。

次に連体節を見る。

(38) 海外市場縮小って言うのも、わからんし、あの一、えーっと、香港の市場が緩和されたというのもよくわからない。(会話)

(39) 市内の美観を守るために、犬用のトイレまで設置されているのは中国大陸では考えられない、(<笑い>) (会話)

今回の会話では連体節は 4 例しかなく、明確なことは言えないが、具体的な名詞が主名詞になっている例はなく、「書かれていないというのがあって」「現地人に言われたほうが傷つく」のように、4 例が形式名詞であった。

(40) 音楽ダウンロードといえば、a u は内臓されている専用のミュージックプレイヤーで音楽を聴こうと思ったら、音楽のデータをサイトから購入するかCDから取り込まねばならない。(ブログ)

(41) 日本国民は、正当に選挙された国会における代表者を通じて行動し(中略)わが国全土にわたって自由のもたらす恵沢を確保し、政府の行為によつて再び戦争の惨禍が起ることのないやうにすることを決意し、ここに主権が国民に存することを宣言し、この憲法を確定する。(新書)

(42) プラスーというのは二十年ばかり前に発見されたピス語という言語だということですから、その後また新しく発見された言語があればさらにいくつか増えます。(新書)

ブログ・新書での連体節は、文中の名詞を受身によって修飾し、情報の多い文をすっきりと整理して表現するために用いられる場合がある。会話での連体節が、それがないと文が成り立たない連体節であるのに対し、ブログ・新書での連体節は連は修飾成分として働いている場合が見られたという点が今回の資料での会話と新書の連体節の違いと言える。会話ではそうした複雑な連体節

を含む文は使わず、2文にして述べるのであろう。

5.6 行為者の明示・非明示

行為者が明示されているか否かを調査した。表13がその結果である。無情物が二格に来る場合は正確には行為者とは言えないが、用語を統一するため、行為者という語を使う。一方、会話・

	会話		ブログ		新書	
明示	7	10.6%	32	15.1%	58	8.3%
非明示	59	89.4%	180	84.9%	640	91.7%
合計	66	100.0%	212	100.0%	698	100.0%

ブログ・新書、どのテキストにおいても行為者が表示される文は8%から15%であり、行為者が明示される文はかなり少ない。行為者が表示されないという点では3種のテキストは共通している。

しかし、表示されない理由は書き言葉と会話では異なっていた。表1で見たように、新書は行為者不明の受身が多く、会話は直接受身が多い。新書では行為者を問う必要のない受身、行為者を降格する受身が多いため、行為者が明示されない受身文が用いられる。

会話では表3の話し手の部分で見たように「私」の視点で述べる直接受身がよく用いられる。そして、会話では、話し手聞き手の間で了解されている事項は省略されるため、以下の例のように行為者が明示されない。

(43) もっと、ていうかねー、"もっと早めに連絡しなさい" (うん) って言われたく2人笑い>。(会話)

(44) んー、だから、なんか、何て言うかな、私も、すごく振り回されちゃった部分がある<しー> {<}、, <うーん>。(会話)

(43)(44)ともに、話し手の間では誰が言うのか、誰が振り回すのか、了解されている。

現象としては、会話と書き言葉で同様に行為者が明示されないが、その原因は異なるのである。

5.7 行為者の格

行為者を表現する際の格はどのようなものであるか、明示された行為者について調べた。

表14で分かるように、今回調査した会話では「に」が100%であった。それに対し、新書は

	会話		ブログ		新書	
に	7	100.0%	23	71.9%	17	29.3%
によって	0	0.0%	0	0.0%	24	41.4%
から	0	0.0%	3	9.4%	8	13.8%
で	0	0.0%	6	18.8%	7	12.1%
の間で	0	0.0%	0	0.0%	2	3.4%
	7	100.0%	32	100.0%	58	100.0%

「によって」が最も多く、次いで「に」「から」「で」「の間で」など多くの形が用いられている。ブログでは「に」と「で」「から」が見られたが、「によって」は使われていなかった。

「によって」を見てみよう。

(45) 要するに、天皇によって制定されたものが天皇によって改正されて日本国憲法になったという筋道である。(新書)

(46) 日本語は数からいっても一億人以上もの人がつかっている世界でも指折りの大言語ですし、また、長い歴史のなかで鍛えられ、明治以後は欧米語との接触によってさらに磨きをかけられ、豊かになっています。(新書)

(47) するとこれらの権利については、一二条のような憲法によって保障された権利という限定がないから、基本的人権には限られないことになる。(新書)

「によって」は(45)のように有情物が行為者である場合もあるが、(46)(47)のように無情物が行為者あるいは作用者として表現されている場合もある。日本語記述文法研究会(2009)では産出動詞を使う文、能動文に二格がある文、かたい文体で用いられると説明されている(p.223)。本稿の用例では、無情物が行為者・作用者である場合が多い、ということを加えることができる。

「で」は新書・ブログで見られた形である。

(48) けれども、憲法草案が国民の知らない間に作られたことについての不満はその帝国議会でしばしば表明され、なぜこうことを急ぐのかと質問が呈せられた。(新書)

(49) 栗や秋刀魚、きのこなど旬の味覚がどこのスーパーでも安く売られている。(ブログ)

(50) ここの名物ラーメンは最近雑誌なんかでも取り上げられている「ど根性ラーメン」!(ブログ)

「で」は日本語記述文法研究会(2009)では原因・手段などの意味と関係しながら能動主体を表現できるとしている(p.225)。本稿の集めた例では、(48)では機関、(49)では場所ともとれる。(50)は手段に近い用例であろう。「で」は場所・手段と重なる意味が強いといえる。

行為者の格は会話では「に」であったが新書では「によって」「に」など多くの助詞が用いられている点が異なっていた。

6. まとめ

以上の結果をまとめる。

1 受身の用法

- ・ 会話では直接受身が多いのに対し、新書・ブログでは行為者が不特定多数の受身が多い。
- ・ 三者共通に、所有物受身・間接受身は使用頻度が高くない。
- ・ 会話に比べて新書では受身の使用率が高い。

2 ガ格名詞について

- ・ 会話ではガ格が「私」の文が大半を占める。
- ・ 新書ではガ格に「私」が来る文は少ない。有情物がガ格に来る文も少ない。
- ・ 会話では「私」の視点で事柄を述べるガ格前景化、新書・ブログでは行為者の背景化が受身を用いる主な理由であるといえる。

3 単文・複文について

- ・ 会話・ブログ・新書、三者共通に、受身は複文で用いられることが多い。
- ・ 会話・ブログ・新書、三者共通に従属節内での使用が多く、会話では「～されて。」「～されるし。」のように従属節で言い終わる文が見られる点が特色である。

4 文末表現について

- ・ 会話では述語のみによる文末は全体の20%で、80%は補助動詞・モダリティや終助詞つき、あるいは従属節による言い終わりであった。
- ・ ブログ・新書では述語のみで終わる文は55-60%で少ないとはいえない。
- ・ ブログでの補助動詞・モダリティつきあるいは従属節による言い終わりは45%であった。
- ・ 新書での補助動詞・モダリティつきは40%であった。
- ・ 述語のみの場合、新書は「る」「ている」が多く、ブログは「た」「ている」が多かった。
- ・ 新書は事態を一般的なこととして述べるため「る」、引用のため「ている」が用いられ、ブログでは事態の描写の「た」、状態の描写の「ている」が用いられる。

5 従属節の形

5.1 受身が複文末に用いられる場合

- ・ 会話は複文末の例が少なかったが、引用節が大半であった（例 ～と言われた）。
- ・ ブログ・新書は連用節が多く、新書では連体節もある程度見られた。

5.2 受身が従属節として用いられる場合

- ・ 会話はほとんどが連用節であったのに対し、ブログ・新書は連用節と連体節が見られた。
- ・ 連用節は主語の統一、視点の統一のために用いられていた。
- ・ 新書では従属節に受身が使われる場合、連体節を構成することが見られた。簡潔な文に多くの情報を盛り込み、さらに視点を統一するために用いられているのであろう。

6 行為者の表示・非表示

- ・ 3種類のテキストすべてにおいて行為者が表示されない文が80%を超えていた。
- ・ 行為者を表示しないのは、新書では行為者を背景化するため、会話では話者間での共通理解事項は省略されるためである。

7 行為者を示す助詞

- ・ 行為者を表示する助詞は、会話では「に」であった。
- ・ 新書で最も多かったのは「によって」であり、そのほか「に」「で」「から」など多くの種類の助詞が用いられていた。
- ・ ブログでは「に」「で」が用いられていた。
- ・ 「によって」「で」は行為者が無情物の場合に用いられていた。

会話と書き言葉では受身の使われ方が以下のような性格を持っていることがあきらかになった。会話は「私」の視点から出来事を述べ、主語を前景化することが重要であるのに対し、書き言葉では行為者を背景化して事柄を述べることが重要である。会話では受身は個人的な内容を述

べるのに用いられているが、新書では、一般的な事態を述べる、行為者不明の受身が用いられているという点が大きな違いである。

会話もブログ・新書も、受身が複文で用いられることは共通しているが、会話では従属節による言い終わりが多い点が書き言葉と異なる。

文末表現は、会話は補助動詞・モダリティつきが基本と言えるが、書き言葉では述語のみ、モダリティつきが、ほぼ同じくらいの頻度で用いられる。従属節との関係では、三者共に連用節と共起するが、新書は連体節と共に使われる点が特色である。

行為者は三者共に表示されないことが非常に多い。会話とブログ、新書は行為者を表示する助詞に違いが見られた。

7. 日本語教育への応用

今回の調査結果を日本語教育に応用するために、以下の点を提案する。

会話と書き言葉は受身を使う目的が異なる。会話は「私」の視点で個人的な事柄を述べるために使い、新書では行為者を背景化するために用いられる。目的が異なれば表す内容も異なる。新書では中立的な、感情を含まない受身が用いられる。また、新書では連用節・連体節と共に用いられる複雑な文の中で使う受身を理解すること、あるいは複雑な構文の中で受身を使うことが求められる。受身と補助動詞・モダリティの組み合わせられた形も中級でしっかり取り上げる必要がある。実際の文脈では行為者は一般的に表示されない。中級に入った段階で行為者を表示しない文を理解することに慣れていく必要がある。

間接受身・所有物受身は使用頻度が低いため、中級での学習項目にまわしても十分であろう。

資料

- ・京都大学情報学研究科・NTT コミュニケーション科学基礎研究所『ブログデータ』
- ・日本語教育支援システム研究会『CASTEL/J』
- ・宇佐美まゆみ監修『BTS による多言語話し言葉コーパスー日本語会話 1』東京外国語大学大学院地域文化研究科 21 世紀 COE プロジェクト「言語運用を基盤とする言語情報学拠点」

単文複文調査の資料 各 1 万字

- 井上忠司 (1982) 『まなざしの人間関係』
- 中根千枝 (1967) 『タテ社会の人間関係』
- 米山正信 (1991) 『化学とんち問答』

参考文献

- ・庵功雄 (2001) 「テイル形、テイタ形の意味の捉え方に関する一試案」『一橋大学留学生センター紀要』第 4 号
- ・小川誉子美・安藤節子 (1999) 「文法項目の段階的シラバス化ー受身の場合ー」

『世界の日本語教育』9 国際交流基金 pp1 - 13

- ・江田すみれ・小西まどか(2007)「3種類のコーパスを用いた3級4級文法項目の使用頻度調査」『日本女子大学 紀要 文学部』57
- ・江田すみれ(2013)『「ている」「ていた」「ていない」のAspect』くろしお出版
- ・中俣尚己・清水由貴子・建石始(2012)「実質語との共起に着目するコーパスを用いた文法研究」『日本語教育国際大会 2012』
- ・仁田義雄(2009)『日本語のモダリティとその周辺』ひつじ書房
- ・日本語記述文法研究会(2009)『現代日本語文法 2』くろしお出版
- ・日本語記述文法研究会(2007)『現代日本語文法 3 Aspect・テンス・肯否』くろしお出版
- ・日本語記述文法研究会(2003)『現代日本語文法 4 モダリティ』くろしお出版
- ・日本語記述文法研究会(2008)『現代日本語文法 6 複文』くろしお出版
- ・プラシャント・パルデシ・今井新悟・李在鎬・砂川有里子・赤瀬川史朗・今村泰也(2012)「日本語教育につながるコーパス研究—現状と今後の展望—」『日本語教育国際大会 2012』
- ・堀恵子・江田すみれ・山田ボヒネック頼子・母育新(2012)「Web ツールを通して世界とつながる日本語教育—文法用例文検索システム「はごろも」と国内外での可能性—」『日本語教育国際大会 2012』
- ・堀恵子(2012)「文法検索システム『はごろも』の文法リスト作成」シンポジウム「Web でつながる日本語教育」日本女子大学 配布資料
- ・前田直子(2011)「受動表現の使用と「拡大文型」の試み」『日本語/日本語教育研究』2 日本語/日本語教育研究会 pp67 - 84
- ・森篤嗣・庵功雄(2011)『日本語教育文法のための多様なアプローチ』ひつじ書房

学術論文における問題提起疑問文とそれに対する答え方

清水まさ子 (独) 国際交流基金日本語国際センター)

1. はじめに

論文やレポートを書く際には、「なぜOOは発生しなかったのだろうか」や「OOは果たして影響しているのだろうか」といった問題提起の機能を持つ疑問文(以下、問題提起疑問文と呼ぶ)を用いることがある。論文の書き方参考書でも問題提起疑問文が紹介されているが、その多くは問題提起疑問文そのものだけを紹介しており、提起された問題がどのように答えられているかまでは記述されていない。

そこで本論文では、学術論文中で提起された問題に対して、どのように答えが出されているのかということも含めた問題提起疑問文の用法について考察する。

2. 先行研究

2. 1. 本研究における問題提起疑問文とは何か

疑問文とは終助詞「か」で終わる文(例:雨が降っていますか。),および「か」は用いられていないが、文意を変えることなく「か」を補える文(例:雨がふっている?)を指す(田野村1988, p123)。本研究は学術論文を対象にしているため、前者の疑問文、つまり終助詞「か」を伴う疑問文を調査対象とする。

同じ終助詞「か」をもつ疑問文でも、その疑問文が表す機能は異なる。宮崎他(2002)では、疑問文の機能を<質問>と<疑い>に分けた。まず、質問の機能から見てみる。

(1)A: 誰がここに来ましたか?

B: Cさんです。

(1)は「誰」が来たかという情報が抜けている状態のAがBに対して情報を求めている場面である。このように、「聞き手から情報を引き出そうとする機能」を<質問>と呼んでいる。(同上, p175) また次のような疑問文もこの<質問>の機能を有していると言われる。

(2)A: 今日、彼来るかな。

B: 来るんじゃない?

傍線部は「来るのではないか」という文が口語体になった文であるが、これは「来る」という肯定への答えを含んでいる。(2)のような疑問文を安達(1999)は「傾き」をもつ疑問文と呼んだ。「傾き」をもつ疑問文も、<質問>の機能の一つだとされる(宮崎他2002, p181-182)。

次に<疑い>の機能にうつる。<疑い>は、<質問>機能とは異なり、相手に対して問いかけで答えを出そうとせず、相手が存在しない状況や話し手自身の「独話的な環境」(宮崎2005, p184)で用いられているという。論理的なテキストでは、この<疑い>の機能の独話的な用法である<思考過程>や<疑念>から派生した<問題提起>機能が用いられているという(同上, p197-198)。

次の(3)のような例である。

(3) 今まで、納豆の歴史に関して述べてきた。それでは、納豆はこれからどのように発展していくだろうか。ここでは、納豆の将来について考えてみたい。

以上みてきたように、疑問文には<質問>と<疑い>の機能があるが、本調査で問題提起疑問文と呼ぶ文は、上記で述べた疑問文の機能のうち、<疑い>機能の中の<問題提起機能>を持つ文を指す。

2.2. 問題提起疑問文に関する先行研究

文章中における問題提起について述べた先行研究に西田 (1986)、樺島 (1980)、野村 (2000) がある。

野村 (同上) は問題提起文は情報を要求し、その次に来る文がパラグラフを開始させる効果を持っていると述べた。この野村の説は、例えば「①なぜ日本人はあまり休みをとらないのか。②ここでは、日本人の働き方について考察する」といった文章がある場合、①の問題提起文によって話題をとりあげ、②の文によってパラグラフを開始させていることになる。ここで問題提起文は、パラグラフ (=任意の話題にかかわるまとまりである部分) (同上; p 123) を開始する前に、その話題を読み手に導入する役割を持っていると言える。

次に西田 (1986)、樺島 (1980) は、問題提起文と、それに対応する解答について述べた。まず西田は文章中における疑問文 (=「問いかけの文」(p62)) によって行われる問題提起は、必ずその次に答えとなる解決・説明を要求すると述べた (西田 1986, p 62)。この問題提起とその答えの提示のし方についてパターンを示したのが樺島である。樺島はまず、「理解を高める型」というパターンを挙げた。これは文章全体のトピックを最初に提示し、すぐ後の文で答えを提示するというものである。また次に「問題解決の型」というパターンを挙げた。これは、問題提起の後にすぐ答えが用意されておらず、答えは「後回しに」されるものである。

以上のように、文章中における問題提起文は、その文章の流れに影響を与える機能を持ち、また問題提起文には、それに応じた解答が存在することがわかった。

3. 先行研究に関する問題点と本調査の目的

先行研究で疑問文には<質問>と<疑い>という機能があり、論理的な文章では<疑い>の機能から派生した<問題提起>の機能が用いられていることがわかった。また、文章中の<問題提起>は、必ずそれに対応する解答や説明がつき、またその<問題提起>と解答や解説との間には、いくつかの提示パターンがあることがわかった。しかしながら、実際に<問題提起>とその解答がどのような関係になっているのか、またどのような用法が多く用いられているのかについて、調査した先行研究はない。特に、<問題提起>は論文やレポート等で用いられる表現の一つであるが、それらの答え方まで含めた説明は、まだあまり調査されていないようである。

そこで本研究では、学術論文における問題提起疑問文で提示された問題が、どのように解答さ

れているのかについて調査する。

4. 調査方法

調査は次のように行った。

- 1) 疑問文全体の機能の割合について調査する。
- 2) 問題提起疑問文に対する答え方について調査する。
- 3) 問題提起疑問文の出現箇所について調査する
- 4) 1) ～ 3) の複合調査をする。

5. 結果

5.1. 疑問文の機能の割合

それではまず、学術論文全体において用いられている疑問文が<質問>機能であるか<疑い>機能であるのかについて調査した。それが以下の表1である。

表1 疑問文の機能

	用例数	割合
<質問>	46	40.7%
<疑い>	67	59.3%
合計	113	100.0%

表1は疑問文の機能について調査したものだが、疑問文全113例中、<質問>機能は46例、<疑い>機能は67例と、<疑い>機能のほうが多かった。<質問>機能46例は、すべて傾きをもつ疑問文であり、相手に直接問いかけるものではなかった。

この調査から学術論文内の疑問文は、<質問>機能中の、いわゆる傾きをもつ疑問文よりも<疑い>機能の疑問文のほうが多く出現していることがわかる。

5.2. 問題提起疑問文に対する答え方

次に、問題提起疑問文に対する答えがどのように出現しているのか調査した。樺島(1980)は、問題提起疑問文に対する答え方を、「直後に来るか」「直後に来ないか」という観点から調査した。本研究でも疑問文に対して、その答えが直後に出ているのか、または直後に出ていないかで分類した。本研究で述べる答えが直後に出現している例とは、次のような例を指す。

(4) これはパズルなのだろうか。これは必ずしもパズルではない。(経済学)

(5) 日本の場合は、アメリカタイプかイタリアタイプのどちらにより近いであろうか。筆者はどちらかといえばイタリアにより近いと考えている。(経済学)

(4)も(5)も、傍線部の疑問文に対して、その後続する文で答えを述べている。このように、明確に答えが疑問文の後に提示されている例を、「直後に答えが出ている例」として採用した。

一方、疑問文の直後に答えが出ない例は、次のような文である。

(6) 憑依現象について心理学的にはどのような理解が可能であろうか。2事例という限界はある

が、両事例の症状や経過、臨床像について心理学検査の所見と擦りあわせて論じることとする。(心理学)

(7) いったいなにが隠され、なにが歪められているのか。情報操作がおこなわれているこ

とを意識するにつれ、真相を知りたいという焦燥もますますつのる。(英文学)

(6)では「どのような理解が可能であろうか」という疑問に対して、その直後にはその問いに対する答えは書かれていない。また(7)も傍線部の疑問文に対して、その答えは直後に書かれていない。このような例を「直後に答えが出ていない例」として採用した。

以上の2種類の疑問文の答え方を調査した結果が、次の表2である。

表2 問題提起疑問文に対する答え方

答え方	用例数	割合
直後に答え	20	29.9%
直後に答えない	40	59.7%
疑問文が連続	5	7.5%
不明	2	3.0%
合計	67	100.0%

表2の「疑問文が連続」というのは、次のように疑問文の後に疑問文が連続している例をさす。

(8) 通常兵器ガヴァナンスが今後どのように発展していくのか。また、そうした発展が安全保障分野のグローバルガヴァナンスにいかなる影響を与えるのか。(政治学)

表2を見ると、疑問文の直後に答えが表れない場合は59.7%、疑問文の直後に答えが現れるのは29.9%となり、疑問文の直後に答えが表れない場合ほうが多いことがわかった。

5.3. 問題提起疑問文の出現箇所

それでは、前節でみた直後に答えがある疑問文と直後に答えがない疑問文は、どのような場所で出現しているのか。次に1節の中でどの部分に出現しているのか見る。ここでいう節とは、1つのトピックとしてまとめられているものである。例えば次の(例)は、「新しいヨーロッパ」という大見出しのものと、「(1) 地域としての「ヨーロッパ」と、「(2) 「新しいヨーロッパ」の論理と構造」という2つの節が書かれている。この1節内において疑問文は「節の冒頭段落」、「節の中」、「節の最終段落」部分のどこに表れているのか見る。

(例) — 「新しいヨーロッパ」の展開 / (1) 地域としての「ヨーロッパ」

カーのヨーロッパ観は、同時代の国際関係をめぐる多様な問題を反映しながら動的に形作られた。それは特に、ヨーロッパを一つの地域として認識する過程、そしてその地域における多文化・多民族的な政治・経済単位の必要性および重要性を認識する過程から成り立っていた。

第一次世界大戦中に外交官としての第一歩を踏み出したカーのヨーロッパ観は、極めて限定されていた。… (途中略)

(2) 「新しいヨーロッパ」の論理と構造 大戦の激化を背景として、「新しいヨーロッパ」論は、多文化・多民族的な政治・経済単位の考察の段階に入る。・・・ (政治学)

表3 節内での問題提起疑問文の位置

	節の冒頭段落	節の中	節の最終段落	その他	合計
用例数	18	34	12	3	67
割合	26.9%	50.7%	17.9%	4.5%	100.0%

表3は問題提起疑問文全体の出現位置である。表3をみると、最も多く出現したのは「節の中」で34例、次に「節の冒頭段落」部分で18例、そして最も少なかったのは節の最終段落部分で12例となった。論理的なテキストで用いられる疑問文は問題提起機能があることは先の先行研究でも述べられていた。よって問題提起疑問文を位置させるのは節の冒頭段落のほうが多いと考えていたが、それよりも節の中に出現する疑問文のほうが多かった。

5.4 疑問文の出現箇所と答え方の関係

今までの調査を合わせて、問題提起疑問文の出現位置と、その答え方を合わせてみた結果が、表4である。

表4 問題提起疑問文の解答パターンごとの出現位置

	節の冒頭段落	節の中	節の最終段落	その他	合計
直後に答え	2	17	1	0	20
	10.0%	85.0%	5.0%	0.0%	100.0%
直後に答えない	15	14	8	3	40
	38.5%	33.3%	20.5%	7.7%	100.0%
疑問文が連続	1	2	1	0	4
	25.0%	50.0%	25.0%	0.0%	100.0%
不明	0	1	2	0	3
	0.0%	33.3%	66.7%	0.0%	100.0%

表4を見ると、直後に答えがある疑問文は全20例あったが、そのうち17例が節の中であり、この答え方をする疑問文はほとんどが節の中にあることがわかった。一方、直後に答えない疑問文の場合は、「節の冒頭段落」と「節の中」の両箇所に多く出現していることがわかった。

このことから、直後に答えがある疑問文は「節の中」にほとんど限定されているが、直後に答えない疑問文は、「節の中」のみならず「節の冒頭段落」にもあるので、比較すると用例数が多くなっていることがわかった。

6. 考察

次に直後に答えがない問題提起疑問文の解答パターンと、直後に答えがある問題提起疑問文の解答パターンが実際にどのように用いられているのか概観し、文章中における問題提起疑問文の用法について考察する。

6.1 直後に答えがある問題提起疑問文の用法

まず直後に答えがある問題提起疑問文は節の中に出現するものが最も多かったが、これらの例

は次のようなものである。

- (9) しかし、三浦半島の繊維土器は、繊維を多量に混入する代わりに砂礫を減じているので、相対的に非繊維の土器よりも軽いといえなくはない。繊維混入の意図が、より文化的・宗教的なものだと考えた場合はどうだろうか。繊維に装飾効果を求めるという(D1)の仮説は、外面から見える繊維の痕跡を条痕やナデによって消そうという意識が認められることから、棄却するのが妥当である。また、(D2)は(*省略)今回は保留する。(考古学)
- (10) ハンバートは自分のニンフェット幻想を求めるばかりで、ロリータが何者なのかということに思いをめぐらせないが、そもそもロリータの出自はどこなのだろうか。ナボコフは彼女の血統について、あとがきで“a dash of Irish blood” (312)と述べていた。ロリータの父親ハロルド(Harold Haze)がアイルランド系であることはいくつかの証拠が作中にあり明らかだが6)、母親シャーロットの出身はいったいどこなのだろうか。それについてナボコフはあとがきをロシア語訳した際に「ドイツとアイルランドの血の混合」(378)という付け足しをおこなっている。(英文学)
- (11) CBTを用いたセラピーのなかで、「人を信じられるようになりたい」というクライエントの一つの主訴は改善された。この際、A子が述べる「信じることができない人」とは誰を指しているのだろうか。もちろん、A子は男性から暴力を受け続ける被害に遭ったため、当然男性全般は“信じられない”対象であったといえる。しかしそれだけでなく、援助する側にある人たち、つまり、家族、主治医、セラピストも“信じられない”対象であったと思われる。A子は援助者らをまったく信用していない訳ではなかったが、自分のことをすべて受け入れてもらえるのかという不安から、完全に信頼することはできないでいた。(心理学)
- (12) このサイズ変数を用いて、(*中略)有意差がみられた。なぜ報告書で打製石斧と認定された石器の大半を筆者は横刃型石器と認定したのか。その理由は、出土打製石器に認められた摩耗痕跡を、報告者が「使用痕」としたのに対し、筆者はその一部を「風化浸食痕」とみたという違いに集約される。本資料中、主に頁岩壘円礫表面にみられる筋状痕跡は、(1)条線が非常に細かく浅く、長く連続したり、礫面の凹凸に沿って滑らかな曲線を描いたりする(図5)。(考古学)

上の4用例は、いずれも直後に答えがあると考えられるが、まずその特徴としては、答え方のバリエーションがいくつかあることである。今回観察された中では、一つ目は、(9)や(11)の波線部のように疑問文の後に「～と言える」「～妥当である」といった論者の判断を表すものである。二つ目としては、(12)のように疑問文の後に「その理由は～である」といったメタ言語表現を用いるもの、そして3つ目としては、(10)のように他資料からの情報を載せるものである。

また特徴のもう一つとして、問題提起→答え、という1つの形式でも、それが文章中に与える影響の違いである。(9)や(10)は、問題提起→答え、という短い展開となっている。例えば(9)は「繊維混入の意図」について、「文化的・宗教的なものだと考えた場合はどうだろうか」という問題提起を呈するが、それに対して「棄却するのが妥当である」と答え、次の話題に移っている。ま

た(10)も、「ロリータの出自はどこなのか」という問いに対して、ナボコフは「“a dash of Irish blood” (312)と述べていた」という答えを出し、次に今度は母親の出自に関する問いを出している。このように、疑問文で欠けている情報を答えによって補う、というパターンが見られた。

(11)や(12)は、1つの疑問文の後にすぐ答えが書かれているが、答えただけで話が終わらず、答えの後にもその答えを展開させるような話が続いている。例えば(11)では、疑問文で「A子が述べる信じられない人は誰か」という問題提起が提出される。それに対して次の文では、「男性全般」が信じられない対象であることが述べられる。しかし、それでは終わりではなく、実は援助する人たちのことも信じられていない、という話が続く。また(12)では、「なぜ筆者は石器に対して報告書とは異なる認定をしたのか」という内容の疑問文に対して、「その理由は～に集約される」という答えが出される。そして、その詳細な理由に関して後文に続く。このように、(11)や(12)は、問題提起疑問文が「問題提起→答え」だけではなく、「問題提起→答え→答え」に関する話の展開、という構造になっている。

以上のように、直後に答えがある問題提起疑問文は節の中で用いられるが、その答え方にはバリエーションが見られた。また後文の談話の展開に影響を与えているものもあった。

6.2. 直後に答えがない問題提起疑問文の用法

直後に答えがない問題提起疑問文は節の冒頭段落に出現するものと、節の中に出現するものがあった。ここではそれぞれにおける用法の特徴を記す。

6.2.1. 節の冒頭段落で用いられる場合

節の冒頭段落部分に出現する疑問文の問題提起において特徴的なのは、次の例のように疑問文の後に「論をどのように論述していくか」といったメタ言語表現文が用いられることである。

(13)では、「私」を詩人に敵対する説教師的ペルソナと捉えたとき、“The World”の最終連はどのような意味を帯びてくるのだろうか。以下、最終連は省略をせず順次見ていくことにする。(英文学)

(14)ではなぜ1990年代後半にM2の実体経済に対する予測力が消滅したのだろうか。この問いに答えるため、まずM2の構成要素の中でどの部分が予測力を喪失させたのかを検証する。(経済学)

(13)は英文学論文におけるある章の冒頭部分であり、これからどのように論を進めるのかについては「以下、～見ていくこととする」という文の後に書かれている。また(14)は「M2の予測力はなぜ～消滅したのか」という経済学論文のある節であるが、後文ではこの問いに答えるためにどのような論の展開を行うのが書かれている。この例のように、節の冒頭段落において問題提起を行い、その後の論の述べ方の詳細を後文で述べるという用法が見られた。これらの例は、野村(2000)で述べられていた「情報を要求する文」(同上：p133-134)であると言える。野村は、問題提起部分を「情報を要求する文」とし、その後続く論の述べ方を示した部分を「パラ

グラフを開始させる効果」をもつ部分とした。

しかし次のように、問題提起の後にメタ言語表現文を示さずに、すぐに話題が続くこともあった。

- (15) 4. 『荒地』における群衆 エリオットは「ロイド」論において（*省略）受身的な集団が発生することを嘆いていた。では、「ロイド」論で語られた「原形質の状態」と『荒地』は、どのように関わっているのか。

『荒地』には草稿も含め、群衆の姿がたびたび描かれている。第1部「死者の埋葬」“The Burial of the Dead”には、（*省略）“crowds of people, walking round in a ring” (1.56) の姿がある。（*省略）つまり、（*省略）『荒地』第5部で描かれる群衆は、「ロイド」論で語られた“protoplasm”の状態に陥った人々と同様に、原始の状態へと立ち戻っているのである。（英文学）

(15)はこの疑問文に入る前まで、T. S. エリオットの「マリー・ロイド論」における「protoplasm」、つまり「原形質」の状態とは何かについて述べ、この段落から同じ作者の「荒地」という作品における原形質の状態について論じ始めている。そして、この疑問文において「『原形質の状態』と『荒地』は、どのように関わっているのか」という疑問文の後、まずは「荒地」の群衆に関して話題が始まり、次段落まで話が続き、その終わりに「つまり、『荒地』第5部で描かれる群衆は、…」という文で、最初の疑問文に対する答えを出している。

これらの用例は、問題提起疑問文が節全体の問題提起として用いられる場合であると言える。

6.2.2 節の中に用いられる場合

それでは次に、節の中に用いられている問題提起疑問文の例を出す。

- (16) ウィルソン (Harold Wilson) 政権が欧州での影響力拡大のため、関係改善に向けて動き出した国が、（*中略）西ドイツであった。アデナウアー (Konrad Adenauer) 陣により、英独関係を改善させる機も熟していた(5)。

イギリスは何故、英独関係を重視したのだろうか。西ドイツは冷戦による欧州の分断を象徴する国家であり、「鉄のカーテン」をすぐ東側に臨む冷戦の前哨国家であった。その西ドイツは英ソ関係、英米関係の中心に位置する要でもあった。ドイツ問題は東西間交渉の最重要課題であり、（*省略）ソ連の主要交渉相手国である西ドイツに対する影響力を高めることで、冷戦にも効果的に関与できる。（政治学）

- (17) これらの結果から、M1 部分ではなく定期性預金の動きが M2 の役割低下の主要因であることが判明した。

それでは、定期性預金が予測力を喪失した理由とは何なのだろうか。一般に、M2 の変動と銀行貸出の動きとは密接な関わりがあることが指摘される。もしそれが事実なら、銀行貸出の実体経済への予測力を調べることで、われわれは新たな示唆を得るかもしれない。（経済学）

(16)の例の問題提起疑問文が用いられる前までは、「ウィルソン政権が西ドイツを重視していた」ことが書かれており、疑問文で「なぜ」重視されているのか、問題提起によって前文までの事実に対する「理由」に焦点が当てられ、問題提起以降、その理由を解明する説明がなされている。また(17)では、まず「定期性預金の動きが M2 の役割低下の主要因であることが判明した。」という事実が述べられ、問題提起文によって、その事実が起きた「理由」に焦点をあて、その理由を解明する説明がなされている。

これらは、問題提起疑問文を節の中で用いることによって、今までの議論に関してはいるが、新たな視点からの議論を始めるきっかけとなっていることが分かる。

7. まとめ

以上の研究から、以下のことがわかった。

- 1) 学術論文内における疑問文の機能は、<質問>よりも<疑い>が多い。いわゆる傾きのある疑問文よりも多く用いられている。
- 2) 問題提起疑問文に対する答えは、直後に出る場合よりも直後に出ない場合のほうが多い。
- 3) 直後に答えがある問題提起疑問文は「節の中」に出現する 경우가ほとんどであるが、直後に答えがない問題提起疑問文は、「節の中」のみならず「節の冒頭段落」にも出現している。
- 4) 「節の中」に出現する直後に答えがある問題提起疑問文は、問題提起→答え、という短い展開となっているものや、また、疑問文の後にすぐ答えが書かれているが、答えただけで話が終わらず、答えの後にもその答えを展開させるような話が続いている場合がある。
- 5) 「節の冒頭段落」で用いられる直後に答えがない疑問文は、節全体への問いかけは、問題提起後にくこの論は次にこうなる>といったメタ言語表現と、それを伴わずに話題が始まる場合がある。
- 6) 「節の中」で用いられる直後に答えがない疑問文は、問題提起疑問文を節の中で用いることによって、今まで述べてきた議論に関してはいるが、新たな視点からの議論を始めるきっかけとなっていることが分かる。

以上のことをまとめると、学術論文における問題提起疑問文は、①問題提起→答え、と解答がすぐさま出てくる問題提起疑問文（使用箇所：節の中）、②問題提起→答え（話題開始のきっかけ）→話題／問題提起（話題開始のきっかけ）→話題、とそれまでの述べられてきた議論に関係させつつ、新たな視点から議論を始める働きをする問題提起疑問文、③節の冒頭段落においてその節全体への問題提起を行う問題提起疑問文、といくつかの用法があるのではないかと考えられた。

[参考文献]

安達太郎(1999)『日本語疑問文における判断の諸相』くろしお出版

- 樺島忠夫（1980）『講談社現代新書 文章構成法』講談社
- 田野村忠温（1988）「否定疑問文小考」『国語学』152, 123-109
- 西田直敏（1986）「文の連節について」『日本語学』5-10, 57-66
- 野村眞木夫（2000）『日本語のテキスト—関係・効果・様相—』ひつじ書房
- 宮崎和人・安達太郎・野田春美・高梨信乃（2002）『新日本語文法選書4モダリティ』くろしお出版
- 宮崎和人（2005）『現代日本語の疑問表現—疑いと確認要求—』ひつじ書房

文章中の語彙の機能について—“テキスト構成機能”という観点から—

高崎みどり（お茶の水女子大学大学院人間文化創成科学研究科）

要旨：文章の中で語彙がどのような機能を持つかについて、テキスト構成機能と結束性という二種類の観点から考察した。材料として学術書コーパスを用いて、いくつかの用例を検討し、実際のテキストの中で、語彙がどのように機能して、テキストを作り上げているのかを見た。その結果、テキスト構成を受けもつ語彙の中心は、名詞であり、ほかの語彙に比べて相対的に抽象的上位にある漢語が多くその働きをするのではないかと考えた。語彙の結束性に関しては、テキスト全体やテキストの意味的構成部分である分節の一貫性に貢献していること、また、談話構成語となる語とそれに組み合わさる分節内部の語彙とは、何らかの結束的な関係性を持っているのではないかと、ということ考察した。そして、談話構成しているという合図や、ひとまとまりの分節を談話構成語に組み合わせることに関しては指示語の貢献があることもわかった。ここで具体的に観察される談話構成語の相対的抽象性や、語どうしの結束性というものは、語彙論において理論的に得られる体系とか、類義や上位下位関係を必ずしも反映してはいない、ということもわかった。

キーワード：語彙 テキスト構成 結束性 分節 指示語

Key Word: vocabulary text-organizing cohesion segments demonstratives

1. はじめに

2. “テキスト構成機能”とは何か

2. 1 テキスト構成機能と“分節”生成——語彙に関して、文章の中で実際に起きていること
2. 2 どのような語がテキスト構成機能をはたすのか
2. 3 名詞以外のテキスト構成機能の可能性について
3. テキスト構成補助機能—指示語に関して
 3. 1 「指示語句」という考え方について
 3. 2 コ・ソ・ド系の機能の差異について
4. テキスト構成補助機能—語彙的結束関係について
5. まとめ

1. はじめに

本稿では、文章中の語彙の機能について、“テキスト構成機能”という考え方をを用いて考察してみたい。

先行研究では、マッカーシー（1995）が、「議論の内容や分野を伝えることではなく、議論に

構成と構造を与える (to organize and structure) 働きを持つ語を「談話構成語 discourse-organizing words」と呼んで論じている。こうした、テキストに構成と構造を与えるような働きについて、本稿では“テキスト構成機能” (text-organizing function) ¹と呼ぶことにする。

マッカーシー (1995) は「談話構成語」がどのようなものであるかについて、語のタイプを文法語 (grammar words) と語彙語 (lexical words) ²とに区別した上で、その中間にあるような機能を持つ語であるとし、テキスト分析の方法として注目すべきものとする。すなわち“issue” “problem” “dilemma”のような語がその例とされ、「これらの語は、テキストの分節の代わりをしているのである (ちょうど代名詞のように)。分節は、1つの文である場合もあるし、数個の文、パラグラフ全体、あるいは、それよりも広い範囲である場合もある。」 (p.106) と、説明されている。すなわち、テキストの中で使用されるとき、“issue (争点)”がテキストの内容のうちどここの部分をさすのか、“dilemma (ジレンマ)”とは何と何をさすのか、といったその語が及ぶ範囲が一種の“分節”となることその他、より大きなテキストパターン (問題—解決など) を示し、談話の全体像を予測させる働きを持つ、とする。

このマッカーシーの、“代名詞のように” (just as pronouns can) は、談話構成語となる言語形式が、具体的な物事を代用・代表するような簡約な形や意味内容を持っていることを示唆する。意味内容としても、「もの」や「こと」のような、いわゆる形式名詞を最も抽象度の高い形式とすると、語彙の中でのその抽象度・一般性は、相対的に高いものがあるのではないかと、思われる。

また分節 (segments) とは、そうしたテキスト構成語に組み合わせられる (match the words with the segments) テキスト内容のことであり、テキスト内部において、明確に区切りをつけることが可能となるような援助がその中に用意されているものと思われる。その援助とは、テキスト構成語と分節の中の語彙の結束関係や、テキスト構成語に付される指示語や修飾・限定語句なのではないかと考える。

ここでは、テキストの中のどのような言語形式が、テキスト構成の機能を果たすのか、また、テキスト構成語が、テキストの中でどのように“分節”の範囲に及んでいくのか、についてコーパスを利用して観察する。また、テキスト構成を援助すると思われるいくつかの言語形式についても、考察する。

なお、使用コーパスは、国立国語研究所「文章における語彙の分布と文章構造」プロジェクト (プロジェクトリーダー：山崎誠) 作成の“学術入門書コーパス”から、『政治学入門』(阿部齋 岩波テキストボックス)、『日本外交史講義』(井上寿一 岩波テキストボックス)、『アメリカの経済 第2版』(春田素夫・鈴木直次 岩波テキストボックス)、『刑法原論』(内藤謙 岩波テキス

¹ 高崎 (2001) および高崎 (2001) においては、「談話構成語」というマッカーシーの用語をそのまま使用したが、本稿では、書かれたもの (テキスト) を考察対象にしているということを明確に示すために、「談話構成語」と同様の内容をもつ概念を「テキスト構成語」と呼び、談話を構成する機能も「テキスト構成機能」と呼ぶことにする。「談話」と「テキスト」という用語自体の吟味も複雑で、諸説あるため、ここでは単純な説明の仕方を採用する。

² マッカーシー (1995 p105) は、「この区別は、ほかに機能語 (function words) と内容語 (content words) とか、虚語 (empty words) と実語 (full words) とか呼ばれることもある。この区別が役に立つのは、言語の閉じた体系 (closed systems) に属して文法的意味を担う語と、開いた体系 (open systems) に属して、名詞、動詞、形容詞、副詞という主要な語類に属する語とを区別することが可能になるからである。」と述べている。

トボックス)、の4種、計976ページ、約19万4千字を使用した。

2. “テキスト構成機能”とは何か

2-1, “テキスト構成”と“分節”生成——語彙に関して、文章の中で実際に起きていること

「テキスト構成語」(以下「」を外してこの語を用いる)を論じるにあたって、語がテキスト構成にあずかる、とはどんなことを指しているのか、について考えてみたい。

ひとつづきのテキスト全体は、形式的には段落や文で区切られて示される。それもテキスト構成といえる。しかしながらここで考えたいのは、語彙とテキストとの関係においてテキスト構成機能が現前するケースで、そうした考え方は、結束性(語彙的結束性や文法的結束性)や、接続詞の機能に注目して談話の意味的内容的区切りを見ていく先行研究に多く見られるものである。それらによって区切られる意味のまとまりが、分節となる。それは文の一部や、段落の一部、数個の文、または数個の段落、であるかもしれないし、あるいは文脈からテキスト構成語の指定する特定の命題や話題を抽出して得られるものであるかもしれない。ここで思い起こされるのは、「テキストは、結局、形式の単位ではなく、意味の単位である」(ハリディ/ハッサン 1991 p.157)という捉え方である。テキストを構成するものは、意味の分節であり、また、テキストの意図を実現すべく構成された分節の組み合わせや包含関係などの相互関係である。そして長大な学術的テキストを読み込むには、段落ごとに“筆者の言いたいこと”をまとめるのではなく、いくつかの合図になるような語から、大小の意味の分節を作り出し、それらを対応させて関係付け、プロットとして全体構成することが必要になってくる。

今回のコーパスの中にそうした例を求め、少し長いが、『アメリカ経済』「第4章 企業経営と経営革新」の「第2節 巨大企業体制と近代的経営システム」から「(3) 大企業経営の硬直化」全文を引いてみる。そして、まず概説的に、本稿で取り上げたいことのあらましを述べたあと、「2-2、どのような語がテキスト構成機能をはたすのか」以降でそれらについて詳述する。

(例1)

(3) 大企業経営の硬直化

このように(g)アメリカの大企業は戦後しばらくの間、繁栄を謳歌したが、1970年代に入ると、一転して、企業活力の低下、経営の硬直化などと呼ばれる問題に直面するようになった。第2章で見たように、鉄鋼や家庭電器、自動車など成熟した基幹的な製造業では、新製品開発や生産工程の改良で海外の競争者に立ち遅れ、国際競争力の低下が明らかとなった。しかも、米企業は海外の競争相手との正面对決を最初から避け、高収益の高価格製品へ特化したり、海外の低賃金国へ工場を移転したりしているように見えた。大企業はまた、この頃に高まった安全性や公害の防止、雇用の平等を求める市民運動や政府規制の強化にも速やかに対応できず、「反企業」の風潮をもまねいた。石油危機の到来とも相まって、生産性上昇率は大きく鈍化し、利益率も低下した。

では、なぜ^a大企業の活力は低下したのであろうか^a。きわめて多様な原因^aが考えられるが、ごく単純化すれば^b、成功を支えた経営管理システム^eが戦後の繁栄のなかで次第に機能不全に陥り^c、同時に、経営者^fに「覇者のおごり」が蔓延したこと^dをあげることができる。これらは、程度の差はあれ、「大企業病」に陥った外国の成功した企業でも見られたものだった。

まず第1に、戦後、アメリカの大企業では、大規模化と経営多角化が進んだが、これに伴って分業と専門化、階層的な調整という伝統的な経営管理システム^eは強化される一方となった。経営のトップから現場に至る各担当者の責任と権限、その命令系統はいっそう明確に定義され、意思決定は直属の上司からその部下へトップダウン式に下される傾向もますます強まった。これらは多様な資質をもち、比較的頻繁に移動する従業員をごく短い期間で有能な社員へと育成する上で有効なシステム^eだったが、分業と専門化の程度が高まり、管理が複雑になると、その調整コストは高まり、社内での迅速なコミュニケーションやスピーディな意思決定は困難になっていた。

第2に、成熟産業の大企業を中心に、マーケットシェアの拡大や新製品開発より投資収益率、株価の上昇などを重視する経営姿勢^gが強まったことがあげられる。戦後しばらくの間、成熟産業の大企業は、経済の繁栄と強固な寡占体制、絶大な国際競争力を背景に、安定した利益をあげていたが、その一方^h、市場は成熟化し、大きな技術革新の波もおさまったかのように思われた。このような環境ⁱのもとでは、経営者^fの関心は、成果が現れるまでに長い期間と大きなコストを要する研究開発や設備投資よりは、目先を変えただけの新製品の開発や経費節減、手っ取り早く財務指標を改善できるM&Aなどへと移った。経営多角化と事業部の増大もこのような経営姿勢^jを促進した。多角化した企業の経営者^fは、各部門の技術に精通していなかったから、投資収益率などの財務指標によって業績を評価し、社内の資源配分と人事上の処遇を決定した。このため^k、現業部門でも財務指標が優先されリスクの回避志向が広がった。

このような経営姿勢の変化^lは、社内における財務スタッフの地位と発言力を高め、ビジネス・スクール出身のMBA（経営学修士）が会社のトップへと上りつめる割合を上昇させた。しかし、技術に無関心な彼らの台頭はますます財務統制を会社内に広め、少なくとも製造業部門では衰退を促す一因となった。こうして^m米国大企業は1970-80年代に大きな転機を迎えるのである。この点ⁿではさらに、株主の果たした役割も大きい。これについては後でふれよう。

(p 1 2 6 - 1 2 8)

いろいろな下線や、□囲み、記号付け等が施してあり、煩わしくなっているがお許し願いたい。これらの印を使って、テキストの中に、“テキスト構成機能”やその補助機能をもついくつかの言語要素を抽出し考察してみることとする。

1、「原因」について

まず2段落目の第2文の□囲みした「原因」という語に注目する。このたった1語の名詞が、引用テキストの大部分を“(大企業の活力が低下した)原因”という概念でまとめる力、“その原因によって起こったこと”と“その原因”の二つにテキストを分節化する機能を持っている

る。以下でそのテキスト構成の力を詳述するが、本稿ではこのような語を「テキスト構成語」と考えてたい。

この語が重要なテキスト構成を果たす上での補助機能を担う言語形式も多く顕在している。

まず、「なぜ^a～か^a」という疑問形式は、「原因」にあたる内容を答えとする質問のモダリティ形式を提示して、以下にその答えとなる内容が来ることを予告し、さらに「原因」という語が次に来ることによって今後の展開が急押しされる。モダリティを表す機能的な言語形式から「原因」という内容語寄りの語へと次元が転換される現象がみられる。

2、「ごく単純化すれば^b」と「きわめて多様」

この文脈中では、「ごく単純化すれば^b」の「単純」という語と直前の「多様」という語が対義関係にあり、ここに語彙的結束関係が生じている。それによって「原因」についても、単純化された原因と、多様な原因との対比という2種類の分節ができて、小規模ではあるが対立的テキストの構成が成立する。

3、「原因」と組み合わされる分節の提示

そしてここで“多様な原因を単純化した原因”を述べる下線部 c と d が来て、「原因」と組み合わせられる分節として提示される。その手がかりは「経営管理システム^e」と「経営者^f」が各々作る語彙的結束性であり、ここから取り出される分節の命題は“経営管理システムの機能不全”と“経営者の「覇者のおごり」”である。その中心となる名詞「経営管理システム^e」と「経営者^f」は、以後の反復語句となって、語彙的結束性を現出させているのである。それらから取り出される共通の「経営」という語が、このテキスト部分の最初から繰り返し反復され（太字部分）、後半では「経営姿勢」という複合語となって反復されている。このように語彙的結束が、大小の入り組んだ分節を何重にも形成し、それがテキストの意味的な構成となっている。そして、終わりから三行目の「一因」という「原因」の類義語でこの分節は区切られることとなる。

4、「原因」を限定・修飾する「きわめて多様な」

「テキスト構成語」は、単独で示される場合（「列車の到着が遅れた。原因は信号機故障であった。」）と、指示語がつく場合（「この原因は」）や今回の例「きわめて多様な原因が」のように、限定・修飾する語句が付随する場合がある。長大なテキストの場合、こうした限定・修飾する語句があったほうが、分節の特定化が迅速に進む場合が多い。

以上のように「原因」は文脈から種々の援助を受けながら、たった1語でも長大な言語量のテキスト部分と対峙し、受け止めて、それらと組み合わせさせてテキストの目的を実現することがわかる。

5、指示語のテキスト補助機能について

次に、テキスト構成機能を援助する言語要素のひとつとして、指示語³に注目してみたい。

先の例について、□囲み (g) ~ (n) の8カ所がそれにあたり、すべて指示語か、指示語と名

³ 本稿では「指示語」の定義は特にしないが、名詞（「これ」など）連体詞（「この」など）・副詞（「こう」など）・形容動詞（「こんな」など）のほか、「こうして」「このような」など指示語を含む連語等も広く含むこととする。

詞句等との組み合わせになっている。以下のとおりである。

- このように(g)
- その一方(h)
- このような環境(i)
- このような経営姿勢(j)
- このため(k)
- このような経営姿勢の変化(l)
- こうして(m)
- この点(n)

これらのうち、

- このような環境(i)
- このような経営姿勢(j)
- このような経営姿勢の変化(l)

に注目する。これらは、文脈の中で見ると、いずれもその前の部分（イタリック波線部）と“組み合わせる（match the words with the segments）”ことによって内容が理解される。指示語は一見すると、前にそれがあるという合図以上の機能は果たしていないように見えるが、そのあとの「環境」「経営姿勢」「経営姿勢の変化」という名詞ないしは名詞相当の連語によって、その前のどの部分が分節として組み合わせられるかが明確となるのである。すなわちこれらの「環境」「経営姿勢」「経営姿勢の変化」は、それぞれ、指示語に補助されてこそ、分節化された部分を大きく括ってテキスト構成の働きが可能になると言ってよいだろう。

特に(j)「経営姿勢」は、このパラグラフのはじめのほうにある「投資収益率、株価の上昇などを重視する経営姿勢」が予告的に働くこともあって、(i)「環境」と組み合わせられる分節が(j)「経営姿勢」の分節に吸収され、さらに(l)「経営姿勢の変化」に吸収されるという、情報の蓄積と事柄同士の関係性が見て取れる。すなわち、「環境」「経営姿勢」「経営姿勢の変化」は、このテキストのなかで、それぞれに応じた意味の分節を作って受け、そうすることによってテキストの意味内容を構造化し、分節部分を要約したり、名付けたりしてさらなるテキストの展開をはかるといふテキスト構成の機能を果たしているものといえよう。

また、「その一方」(h)、「このため」(k)、「この点(では)」(n)は、上記の(i)(j)(l)と同様の形式であるが「一方」「ため」「点」の各語の意味は、上記の3つと比較すると、かなり形式化しているといえよう。「一方」や「ため」「点」は、特定の方向や、目標、形を意味するというよりも、指示語と組み合わせさせたかたちで、対比や原因・結果、焦点化・補足といったパターン化した関係づけをなしており、前を分節化するだけではなく、その後ろをも分節化しているところに、より高度なテキスト構成機能をみることができ、むしろ接続詞の働きに近いものといえよう。

また、「点」については、辞書にある第一義の「大きさがなく位置だけをもつ図形」（『日本国

語大辞典』以下同) といった幾何学的定義でなく、意味ブランチで第2番目の「さし示す事柄。箇所。」という意義が利用されている。第一義が、自立的な意味といえるのに対し、第二義は、文章の存在を前提とした、文脈の中での機能であって意味とは言い難い。しかしながらコーパス中に第一義での使用はほとんど無く⁴、第二義での使用が多いのだが、それは、第一義からの“広がり”の無さ” “小ささ” が、簡単さ、あるいは絞られた集約的な感じを連想させ、論点の簡約化、集約化というテキスト構成を、言述のストラテジーとしたいときに、メタファー的に拡張的使用がなされたものと考えられる。

以上、実際の用例から、テキスト構成が起きていると思われる箇所に注目してその概要を示した。要するにテキスト構成とは、単独で特定の語がその働きを専ら担うというよりも、文脈の援助や干渉を受けて、その語自体も自立的意味よりは文脈的な機能を優先させて働き、また、テキストを構成的に提示する分節の特定化のために、場合によっては指示語や修飾部分を伴って選択的・指定的に働くものであると言えるだろう。

語彙論の立場からもこうした現象について、一定の興味が示されている。

斎藤 (2011) は、高崎 (2011) をふまえ、「語と文章との関連性を成立させる機縁は何か」という観点から、「語意の機能性」ということを

「語意の機能性」というのは、ある語が、その有する意味から、文章中において必然的に一定の機能を果たしたり、あるいは、その文章の内容との関わりから結果的に特別な機能を担わせられたりする場合が存するという点である、前者の例としては、高崎氏が挙げている「談話構成語」や固有名詞などが相当し、後者の例としては、キーワードや主題、タイトル (としての語) が相当するが、重要なのは、前者の場合はその語の意味が本来的に有する特殊な性格に基づく機能であるのに対し、後者の場合はあくまでも文章全体の内容との関わりで有する機能であるという点である。そういう意味では、語彙論的には、前者の方がより興味深い。(p271)

と述べて、テキスト構成語として働く語には本来的に有する特殊な性格がある、と指摘する。

以下では、この分析にもとづき、どんな語がテキスト構成語になるのかについて検討してゆきたい。

2-2、どのような語がテキスト構成をはたすのか

“テキスト構成語”として考えられる典型は、前述の“代名詞のように”が顕在的に現れた形として、指示語をともないかつ意味が希薄化した語彙語という現れ方である。後述するように指示語は必須要素ではないが、指示語自体も代名詞と同様に文法的結束性を有しているため、

⁴ 今回使用したコーパス全体で「点」は539回使用されていたが、第一義での用法は無かった。

なんらかの指示語が付された方が、テキスト構成語としての機能が発揮しやすいのである。

語彙語と文法語の中間、というマッカーシーの位置づけのしかたを考えると、当然、どちらかにより接近するものであったり、揺れたり、ということがありうる。精密なテキスト構成のためには、語彙語的な意味のベースを保持しつつ希薄化し、かつ文法語にもなりきっていない語がたくさん必要なのである。例文の中で「環境」「経営姿勢」「経営姿勢の変化」は、やや語彙語寄りであり、「一方」「ため」は文法語寄りといえようか。

「点」は、その真ん中にあり、テキスト構成語の典型ではないかと思われる。また、この語は指示語を伴わなくとも、テキスト構成の機能を発揮することができ、「～という点から見て」「～のは…点である」といった形で分節化に関与し、テキスト構成をなしている。

また、

(例2)

資金循環の説明から外れるが、重要な政策問題として2点触れておこう。

(1) 連邦免許の株式公開企業であるファニーとフレディは、債券を長期金融市場に売出して資金を調達し、自己の資産として商業銀行などの金融機関から住宅抵当を買い取って保有する。(中略) その業務の健全性や破綻した場合の危険性が問題化されるに至った。

(2) 1986年の税制改革が、消費抑制・貯蓄促進的な意図もあって、消費者ローン利子の所得控除を全般的に廃止し、住宅ローン利子のみ(持ち家2軒まで)控除を認めたことから、消費者金融の新たな発展を促した。(中略) しかし、消費者の債務依存を不健全に拡大する可能性も大きくなった

(「アメリカの経済」p86)

という例で、(1)(2)の箇条書きにして、企業の健全性や住宅ローンについての課題が述べられている部分が、“重要な政策問題”と名付けられた分節となっている。この中では「問題」という語も、「点」とほぼ同じ部分を分節化するテキスト構成の機能を果たしている。「問題」と「点」の機能分担としては、この場合、「問題」はその分節の内容を端的に表す“名付け”であり、「点」は、あまり大きくない言語量の、内容的にまとまった分節がいくつあるかという単位として働く、ということになるだろう。なお、「点」は、「観点」「係争点」「焦点」「転換点」「論点」等々の熟語となって、それがさらにテキスト構成語として機能している場合も少なくない。

一方「問題」については、高崎(2012)でも触れたように、今回のコーパスの中では729回使用されている。辞書的に言えば「問題」という語の語義のブランチのうち、多くの辞書で第一義とされるのは、「答えを求めるための問い。解答や教えを要求する問い。質問。」(『日本国語大辞典』)と解説される意味で、古語辞典類にも用例があり、現代でも「試験の問題は難しかった」のように使用されている。しかし、この第一義で使用されているのは今回のコーパス中では「研究の最前線の動向を反映した応用問題を考えるパートです。」(『日本外交史講義』p.i)といった例があったが、これをを含め4回のみであった。それ以外は、第二義以降の「批判や論争、または研究の対象となる事柄。解決しなければならない事柄」「心にとめて考

えるべき事柄。注目すべき点」(いずれも『日本国語大辞典』)の方で使用されているのである。

もちろん、これらがすべてテキスト構成語として文脈を形成しているわけではなく、「問題」という語がただちにテキスト構成語として働くとは言えない。それはあくまでも特定のテキスト内での出来事なのであるが、高頻度で使用される抽象度の高い語が、テキスト構成語となる可能性はかなり高いことは、以下の用例を見ても理解できるだろう。

(例3)

さらに刑法解釈学の犯罪総論では、すべての犯罪に共通する一般的成立要件を検討するとき、きわめて詳細な論議が展開されている。そのような論議が展開されるのは何のためであろうか。この「問題意識」から、本書は、構成要件該当性、違法性、責任、未遂犯、共犯などの「問題」を検討した。その際、刑法の現実の運用を示す判例の動向については、とくに留意している。

そして刑罰論でも、応報刑論と目的刑論の対立を基軸に、死刑の存廃をはじめさまざまな論議が従来から展開されてきた。しかも現在、監獄法の全面改正が立法上の「問題」になっている。この状況のなかで「問題解決」の方向をどのように考えたらよいであろうか。本書が、刑罰論の基本「問題」、刑事政策の現代的「問題」などをとりあげているのは、このような「問題意識」からである。

犯罪は、個別的にみれば、殺人罪、窃盗罪、文書偽造罪、公務執行妨害罪などである。本書は、犯罪各論で、このような個々の犯罪の成立要件と、それに対する刑罰についても、脳死、安楽死・尊厳死、コンピュータ犯罪などの現代的「問題」を視野に入れて概観した。判例の動向に留意していることは、犯罪総論の場合と同じである。現行刑法の解釈について学ぼうとする読者は、犯罪総論と犯罪各論から読んでいただいてもよい。

(『刑法原論』「はしがき」)

ここに使用されている「問題」という語はすべて前述の第二義の意味で使用されているのだが、引用部分が「はしがき」であることから、今後の展開で扱う内容を端的に列挙して、それらどうしに関係付けを提示するのに「問題」という語が便利に使われている。そういう用法の中から、下線部の「問題意識」「現代的問題」のようにテキストのある部分を指定して分節化をはかっているテキスト構成の用法も生じてくるのである。この語がテキスト構成を果たす上での補助機能を担う言語形式も「この」「このような」という指示語、そして、やはり「原因」のところを見たような「何のためであろうか」「どのように考えたらよいであろうか」という疑問のモダリティ形式を先行提示して、モダリティを表す機能的な言語形式から「問題意識」という、より語彙語に近い語へと次元転換される現象がみられる。

もうひとつ例をみてみよう。例1のところ、テキスト構成語であると指摘した「経営姿勢」

という語の中に含まれる「姿勢」という語である。この語も、第一義は「からだの構えかた。」(『日本国語大辞典』)であり、第二義が「(比喩的に)物事に対するときの心の持ち方。ゆきかた。態度。」(同上)である。この例1の文脈の中で、「～などを重視する」や「関心は～移った」という部分をうまく簡潔にまとめて、次の文脈の展開につなげるのに適した語が「姿勢」なのである。これも規模は小さいかもしれないが、テキスト構成を担っていると考えても良いだろう。この場合は、「姿勢」をさらに明確にテキスト構成を担わせるために「経営」という限定を加えて明確化しているのだ、とも考えられる。

以上のように今までみてきた、大小を問わずテキスト構成の機能が認められる語は、「環境」「点」「問題」「姿勢」など、語彙語である漢語であるか、「経営姿勢」「経営姿勢の変化」のようにそれらを含む漢語の複合語や連語、または「一方」「ため」など形式名詞に近いような文法語寄りのものか、であった。ちなみに、先には触れなかったが、「環境」も国語辞典の第一義は、あまり馴染みがないが「四方のさかい。周囲の境界。まわり」(『日本国語大辞典』)ということであったようだ。すなわち、テキスト構成機能を果たしていた語は、語彙語であっても、たとえば辞典の第二義が利用されていたことからわかるように、文脈のなかで拡張した機能を果たして、やや具体から離れて、文法語と語彙語との中間にあるような存在であることが確認できた。

このことを「文法化」という概念を使って言い換えれば、河上(1996 p179)にあるように「『文法化』はもともと内容語だったものが、次第に機能語としての文法的な特質、役割を担うようになる現象をいう。」ということがテキスト内で生起しているのだと言えよう。なお、この引用の場合、注iiにも記したが「内容語」がマッカーシーの「語彙語」にあたり、「機能語」が「文法語」にあたる。また、「文法化されていく語は一般的な語、つまり基本語やもともと意味自体が希薄な語であることが多い。」(同 p182)という指摘もある。

実際のテキストを観察すると、本稿でいうテキスト構成語は、文脈が必要に迫られて、既存の語から何がしかの使える部分を直感的に選択し、それぞれ語彙的意味を希薄化したり、第二義を作ったりして使用し続けていった結果だといえよう。

こうしたことから、国語辞典の意味ブランチの並べ方の順序についてのルールが存在にも、テキストの中での語彙の機能ということが関連してくると考えざるをえない。

先に述べた「問題」「点」「姿勢」の場合は、確かに第二義に相当していたが、「原因」という語は一義のみであり、高崎(2012)で検討した外来語「アプローチ」という語は、第一義が「学問研究において、対象にせまること。またその方法。研究法。おもに社会科学についていう。」であり第二義以下が「敷地の入口や門から特定の場所や建物に通じる小道」「スキーのジャンプ競技、走り幅跳び、走り高跳びなどで、～」「ゴルフで～」等となっていて、第二義が談話構成とは言えない。また、辞書によってもその順序のルールは異なっているのである。

語が構文上の理由で活用したり助辞をとまったりして構文の中で機能するように、語の意味も文脈の中で構成に寄与するために多義のブランチにわたって“活用”し、意味限定のために修飾語が付随したりするのである。今、テキストの中で起こっている語彙語のふるまいは、

野村（2003）の指摘するように、「文法も語彙項目と同様に、形式と意味の慣習的な結び付きである「記号」であり、意味を表すために存在するといえよう。語彙項目と文法の違いは、記号の形態の複雑度、記号の意味の抽象度の差にすぎず、従来行われてきたように語彙と文法とはまったく性格の違うものとして2分されるものではなく、連続したものであるということになる。」（p55）ということを実感させる様相を呈しているのである。

2-3、名詞以外のテキスト構成機能の可能性について

このことについて考えるために、先の例文1に戻るが、指示語を談話構成機能のてがかりとすると、

- ・ このように(g)アメリカの大企業は戦後しばらくの間、繁栄を謳歌したが、1970年代に入ると、一転して、企業活力の低下、**経営**の硬直化などと呼ばれる問題に直面するようになった。
- ・ こうして(m) 米国大企業は1970-80年代に大きな転機を迎えるのである。

の2箇所が問題になってくる。

この例1のテキストは、前述したように、『アメリカ経済』「第4章 企業経営と経営革新」の「第2節 巨大企業体制と近代的経営システム」から「(3) 大企業経営の硬直化」の項の全文を引いている。はじめの文は(g)の「このように」で始まっているが、これは、前の項の部分、すなわち「(2) 戦後の大企業経営の発展」という項の内容を受けているということを示す指示語である。そして続く「アメリカの大企業は戦後しばらくの間、繁栄を謳歌したが、1970年代に入ると、一転して、企業活力の低下、**経営**の硬直化などと呼ばれる問題に直面するようになった。」はそれらの内容を要約して、提示している。これもかなり長いテキスト(=「(2) 戦後の大企業経営の発展」1項目分)を「直面する」という端的な語を中心としていくつかの格や修飾語のついた語句に言い換えているといえる。このような場合、テキスト構成機能はたしかに働いているものと思われるが、“テキスト構成語”を特定するのは難しい。だが、たしかにこの文1文で、前の部分の内容を取り込む働きをしており、テキストを構造化しているテキスト構成の働きを認めることができるし、この文の述語は「直面する(ようになった)」であるため、形式的にも「このように」という連用形のかかっていく先としては「直面する」であると考え、これをテキスト構成語と考えても良いかと思う。しかしながら、先に述べた「このような」の場合と比べると、「このように」では、名詞ではない語がテキスト構成語となっており、しかもかなり長い修飾・限定部分を先行させているので、テキスト構成語として抽出するのはかなり難しいように思われる。

さらに(m)の「こうして」も、同様に考えれば、かかっていく先は「転機を迎える」になる。上記に出した、この例文全体のはじまり「このように(g)アメリカの大企業は戦後しばらくの間、繁栄を謳歌したが、1970年代に入ると、一転して、企業活力の低下、**経営**の硬直化などと呼ばれる問題に直面するようになった。」から以降のすべての、かなり長い部分を分節化して、端的に「転機を迎える」と要約することでテキスト構成をしている、といちおうは考えられる。このよ

うに、名詞以外のテキスト構成語というのは、それだけを取り出すのは難しいと言わねばならない。

しかし、上記の場合に限って言えば、いずれにしても「直面」「転機」という名詞に収斂させていると見ることも可能であることを考えると、テキスト構成という機能は、本来名詞が中心になって働くものであるという可能性が高いのではないか。名詞以外の可能性はないわけではないが、焦点がぼやけてしまい、加えて、用言であるといろいろな格や修飾語が付随しやすく、端的なテキスト構成を見ることが難しいのかもしれない。

田中・深谷（1998）では、名詞概念の形成と動作動詞の概念形成に関して論じる中で「概念を分類操作するためには、動詞的概念を名詞的概念として処理するほかないということである。分類操作をするには、連続的な動作で概念的に掴み取り、対象化する必要があるが、掴み取った瞬間に、それは名詞化されるのである。」（p136）と指摘しているが、同様のことがテキスト構成時に起こるのではないか。すなわち、分節化とは、文脈の描写・叙述の流れをひとまず止めて概念化することであるからである。

3. テキスト構成補助機能—指示語に関して

3.1 「指示語句」という考え方について

先述したように、“テキスト構成語”として考えられる典型は、“代名詞のように”が、顕在的に顕れた形で、指示語をともなった、意味が希薄化した語彙語という現れ方であると思われる。指示語は必須要素ではないが、テキスト構成補助機能を有しているために、なんらかの指示語が付された方が、テキスト構成語としての機能が際立つのである。先のマッカーシー（1995）でも「指示詞 demonstratives」は「閉じた体系」であるとして、「文法語」（機能語）の範疇に入れられているが、この場合、文法語は、語彙語が語の語彙的意味よりも機能面を発揮することを助けおり、文法の方へ引き寄せているともいえそうである。

逆に言えば、コーパス検索でテキスト構成語をさがすのであれば、指示語を検索のてがかり語とすることができそうである。

ここで、高崎（1988）で提案した「指示語句」の考え方を援用したい。

「指示語句」とは、先の2-1で示した例1を使って考察した「テキスト構成機能を援助する言語要素」のところであげたうちの その一方(h) このような環境(i) このような経営姿勢(j) このため(k) このような経営姿勢の変化(l) この点(n)、以上の□で囲まれた全体を「指示語句」とよぶのである。先程は、組み合わせられる分節に焦点をあてて見ていたのだが、この指示語句の場合は、前方にあるかなりの量の叙述を、このような環境(i)なら「環境」という観点で選択して分節として捉え直し、「このような環境のもとでは、～」と連用修飾の形で後の述語部分に受け継いでいく、その機能の方に焦点をあてて見ている。すなわち、文脈から特定の部分や命題を取り出す「指示」の働きをしているのは、指示語だけではなく、それに続く語句「環境」も寄与していると考えて、[指示語+後に続くなんらかの要素]まで含めて「指示語句」と呼んだのである。この“後に続くなんらかの要素”は、名詞で、先行文脈の繰り返しではなく、量的にも

要約的になり、意味的にも抽象化したり、比喩的な言い換えになったりするなどして文脈を展開するために少しずつの変形・変容を加えている、と考えるわけである。高崎（1988）ではこれを「指示語句は、指示語のみの場合と比べて、単なる指示・代用ではなく、前の叙述を繰り返すようにみえながらその実微妙にずれて、“変容”⁵するところにその意味があると言うことができる」とし、ここに“文章展開の様相”、本稿で言えばテキスト構成機能をみているのである。

そして同じく高崎（1988）においては、その“変容”のパターンを①要約・敷衍⁶②名づけ、③比喩、④次元変換、⑤形式化（抽象化）⁷、の5種類に分けて見ている。これが、本稿の言い方であれば指示語を合図とする名詞によるテキスト構成のパターンとなり、ある程度の大きさを持った分節に対して、テキスト構成語がどのようなパターンで組み合わさっているのかを示したものとなる。

今回のデータから、指示語「このような」を手掛かりとして例を引くと、①要約については、たとえば先の例1では、「このような経営姿勢の変化①は」がそれにあたる。これはすぐ前の「第2に」からはじまる段落が、ある期間における経営者たちの経営に対する関心や評価について述べて、それが「強まった」「移った」「広がった」などという、前後における何らかの“変化”の面から捉えられて記述されている部分を受けて端的に要約した語句となっている。

②名づけは、本コーパス中の『日本の外交史講義』「第8章経済成長による外交の変容」⁴ 『高度経済成長』下の対外政策の統合（p182）における

（例4） 日米安保条約の軍事的機能が、憲法第9条によって制限されたことは、この条約による日本の軍事協力が、基地貸与と駐留米軍の経費負担という間接的なものに止まることを意味しました。このような間接「貢献」は、アメリカから許容されています。

この「間接『貢献』」は、「要約」というよりも、むしろ書き手の物事に対する把握の仕方が提示されており、「いわゆる」とか「名づけて」といった前置きを暗黙のうちに含むことが多く、専門語として示されることもある。また、かぎっこなどで区切られて皮肉や揶揄など批判的なニュアンスを暗示することもある。

③比喩は、同書「第2章 〈帝国〉日本の対外膨張」¹ 『脱亜』への転換（p31）の

（例5） もちろんこの場合の「内政改革」は、内政干渉に限りなく近いものでした。しかし清国と共同で行うこと、また朝鮮政府内にも「内政改革」を志向する政治勢力があったこと、これらを前提とすれば、たとえ内政干渉であっても、朝鮮側に受け入れる余地があり、清国と

⁵ この高崎（1988）では、金岡孝（1963）からの「展開ということは、ある事がら、その事がらとなんらかの関連をもつ他の事がらに変容することをいう」（p49）を引いて、「変容」の説明としている。

⁶ 高崎（1988）では、単に「要約」とのみしていたものに、本稿では「敷衍」を付加する。

⁷ 高崎（1988）では単に「形式化」とのみしていたものに、本稿では「抽象化」を付加する。

共同で行うのであれば、「内政改革」によって、朝鮮が自立する。そうなれば、朝鮮も西欧国家体系の下で、永世中立国となる基礎的条件が作られる。このようなシナリオにおいて、6月2日の派兵決定は、朝鮮永世中立化構想実現の第1歩となるものと位置づけられていたのです。

のような例で、映画や演劇と無関係の文脈で「シナリオ」という語が用いられるのは、その前の内容と、“あらかじめ決められた物事の展開”という部分で共通すると捉えて、比喩が成立するからである。ただし、この場合の比喩は隠喩であり、“モノからコトへ”というカテゴリー転換を含むものとなる。

④次元変換は、『刑法原論』「I 犯罪現象と法」「2 犯罪現象の基本動向」(p12)の中の

(例6) (1) 窃盗・遺失物等横領の増加傾向とその内容

窃盗の認知件数は、1986年の137万5096件からもほぼ一貫して増加傾向にあり、95年には、157万件を超えて戦後第2位となっている。しかし、このような窃盗の認知件数の激増は、86年の認知件数に対比して、主として、自転車盗、オートバイ盗、車上ねらいなど手口が単純である比較的軽微な事犯が約21万件、自動販売機荒らしが約7万件、それぞれ増加したことによるものであり、悪質な侵入盗(空き巣ねらい、忍び込みなど)は、約6万件減って減少傾向を示している。

という例である。ここでは、数字や現象の動態が、「激増」という名詞に品詞に転換されてコト化しているという面で次元転換をしていると考えられる。この次元転換は、種々のタイプが考えられるが、高崎(1988)では、数字や観察などの客観的記述から主観的評価や意味付けへ、という変容や、あるいは品詞的な転換は、上記のように名詞化ばかりでなく、辞的なものから詞的なものへ(たとえば「～ねばならない」から「義務」へ)という段階もありうるとし、またいくつかの概念を複合語へとまとめあげる場合もありうるとしている。

最後の⑤形式化(抽象化)の例としては、『日本の外交史講義』「第7章 冷戦と戦後国際秩序の模索」(p142)において

(例7) 冷戦というと、日本は何か傍観していたような気がしませんか？たとえば米ソはたしかに冷戦を戦ったのでしょうか。では日本も冷戦を戦ったといえるのでしょうか？あるいは冷戦は日本に何をもたらしたと思いますか？このようなことを考えながら、日本にとって冷戦とは何だったのかをまとめておきましょう。

のような場合で、指示語に続く言語形式「こと」が抽象化・形式化の度合いが強く、要約や名づけなどの機能が希薄になっている。「こと」のほかにも「中」「かたち」「もの」などが見られ、また、それらよりは抽象化の度合いはやや低いものの、「面」「点」や「意味」「立場」「状況」「見地」「関係」などがこれに含まれると考えられる。

以上であるが、ここで手掛かりとして選んだ指示語の「このような」は、「この」だけの場合と比べて、前方を大きく漠然とさしたり、他に類例のあることを含みとして持って、曖昧さを増す方向に働いたりして、かなり大きな分節を構成できる可能性を持っているといえよう。

①～⑤までの“変容”のありかた、すなわちテキスト構成機能のバラエティを指摘してきたが、実際のテキスト内部では、これらは排他的な分類として、つねに独立事象として観察されるわけではない。ひとつの指示語句、テキスト構成の中に、複数の変容のあり方として見出されることのほうが多いのである。もちろんどのあり方が強く出ているか、という差異はあっても、次元転換は談話構成の基本的な機能であり、また要約も分節を受け止めてテキスト構成するというという目的であれば、より短く端的な言語形式が来ることが多いのは自明なことといえよう。そして、比喩も、この学術書の場合、文学的な一回性の表現ではなく、専門家のテキストの中で繰り返し使用されることで、専門分野の思弁・論理の展開になくはならぬ役割を負っていくようになるというプロセスがありうるのだと考えられよう。

ここでは、いわば前をどう受けるかということに焦点をあてた指示語句の変容のパターンを、談話構成のパターンに置き換えて見てきたが、実際には談話構成は指示語が伴わない場合もあって、もっと多くのパターンが見いだせそうである。ひきつづき追究したい。

3.2 コ・ソ・ド系の機能の差異について

3.2.1 「このように」「こうして」の場合

さて、ここまででは「このような」をテキスト構成のてがかりとしたが、これらはいわゆる連体形であり、あとには名詞ないし名詞句がきて、それがテキスト構成語となっていた。それでは連用形である「このように」「こうして」は、テキスト構成の補助にどのように関わるのであろうか。

先に「2-3、名詞以外のテキスト構成機能の可能性について」のところで触れた

- ・ このように(g)アメリカの大企業は戦後しばらくの間、繁栄を謳歌したが、1970年代に入ると、一転して、企業活力の低下、経営の硬直化などと呼ばれる問題に直面するようになった。
- ・ こうして(m)米国大企業は1970-80年代に大きな転機を迎えるのである。

の2例を再度検討すると、指示語「このように」「こうして」とテキスト構成語と判断される語（「直面する」「転機を迎える」）の間が、「このような」（例：このような環境—例1）、「こうした」（例：こうした状況—『アメリカ経済』p197）の類よりも距離が長く、いろんな語句があいだに入ってきていることがわかる。先の分類で言えば①要約・敷衍にあたる変容のしかただが、要約というよりはむしろ、敷衍的な言い換えとでも言ったほうが適切かもしれない。

このような、前の部分を取り込んでこれから述べていく部分の根拠にしつつ、展開させていく方法は、こうした専門書のような長大なテキストには欠かせないもので、種々の方法での“自己引用”が図られる。「このように」「こうして」が連用形であるため、作用性や状態性をもつ用言

がそのあとに来るので、程度や様態などの修飾語、様々な格的成分を要求して記述が長大になるものと思われる。

しかしながら「このように」「こうして」がつねに文相当の長いテキスト構成語を伴うわけではない。短いものの例として、『刑法原論』の「IV 刑罰論」章の節「19 刑罰論の基本問題」の「(2) 刑罰の根拠」項全文を掲げる。

(例8)

(2) 刑罰の根拠

刑罰は、たしかに、過去に罪を犯したことを前提条件にして、犯罪に均衡する反作用として犯罪者に科せられるし、刑罰の内容は利益剥奪という苦痛である。そのことを「応報」というとき、「応報」の要素は、現実の刑罰に存在している。しかし、そのような「応報」は、経験的事実であり、刑罰という概念の要素であって、それによって刑罰を正当化し、「根拠」づけることはできない。さらにまた、犯罪に対して国家的・道義的非難を加えることを「応報」といい、それを正当化根拠として刑罰を加えることも、すでに検討した国家刑罰権の根拠と限界を超えている。国家刑罰権の根拠と限界についての本書の立場からみれば、刑罰は、犯罪を行ったことを前提条件とし、犯罪に均衡する反作用として科せられる利益剥奪（苦痛・害悪）であるから、刑罰の概念には「応報」の要素があるが、そのような刑罰が正当化され、「根拠」づけられるためには、その刑罰が犯罪防止による生活利益保護の効果と必要性をもつものでなければならない。国家が犯罪防止による生活利益保護の効果も必要もないのに刑罰を加える権能をもつといえるかは疑問である。日本においても、起訴猶予、微罪処分、交通反則通告制度、犯罪少年に対する保護処分のような犯罪の非刑罰的処理（犯罪であっても、事件を刑罰以外の措置で終了させる制度）がかなり広い範囲で行われているのは、その場合には、刑罰を科すことが、そのレッテル貼り作用によって、対象者の社会復帰を困難にし、さらには、その犯罪傾向を固定・増進させて、対象者の再犯予防という特別予防目的にとって逆効果となり、また、一般予防のためにも刑罰を科す必要はないと解されていることが大きな理由であろう（⇒33頁）。

このように考えるとき、刑罰の正当化「根拠」は、前述の意味の「応報」の要素をもつところの刑罰を手段として犯罪を防止することにより生活利益を保護することの必要にある。その意味での「相対的応報刑論」が妥当であると思われる。したがって、犯罪防止のための「一般予防」の効果と必要、および「特別予防」の効果と必要は重要な意味をもっているのである。（『刑法原論』「IV 刑罰論」「19 刑罰論の基本問題」p146）

この「このように考えるとき」の場合、例文の項の最初の「刑罰は」から、直前の「大きな理由であろう」までをすべて「考える」で分節化している。また、「（⇒33頁）」という、参照ペー

ジ（「I 犯罪現象と法」の「3 犯罪現象の法的処理過程（1）—警察・検察庁の段階—」）まで含むと、分節部分は非常に長大になる。しかし、大きい割には、「考える」という語が具体的な動作ではない、形式的抽象的な意味で使用されているため、要約や名づけ等の変容はうかがえず、単に順接的な“そうだとすると”“だから”などの接続詞・接続表現に限りなく近い機能をはたしていることとなる。ほかにも「このように理解されるとき」「このようにみえてくると」「このように行われた」等、何ら変容させることなく、そのまま展開させていく場合も少なくない。

すなわち「このように」がテキスト構成を援助するときには、該当分節に対して、あとに続くテキスト構成する語句・文相当の部分が、要約か言い換え・敷衍かあるいは抽象化形式化か、または接続詞化か、という選択肢が、「このような」よりも広いといえよう。「こうした」に対する「こうして」も同様のことが言えるものと思われる。

（例9）

自民党から失われた票がすべて社会党に流れれば、いずれは社会党が多数党になることもありえたであろう。自民党の内部できえ、そのような予測が立てられていた。しかし実際には、1960年代に入るとともに、社会党も含めた野党の多党化が起こり、自民党から流出した票は各野党に分散することになった。 こうして、社会、公明、民社、共産などの野党が、与党の自民党と対立する形がみられることになる。 ただ、こうした多党化にもかかわらず、自民党の優位それ自体は持続しており、一党優位政党制は崩れていない。（『政治学入門』「VII むすび」p205）

この例で、連用形指示語（「こうして」）に続くテキスト構成する部分は、用言部分（「みられることになる」）を含む文相当の点線部で、先行する分節（波線部）を敷衍している。一方、「こうした」は直後に「多党化」がきている。これに組み合わせられる分節は、先行する波線部および点線部を合わせた部分で、「多党化」というテキスト構成語で要約しているものと考えられる。

3.2.2 ソ系の場合

今まであげてきた談話構成を補助する指示語句はすべてコ系であった。ソ系の指示語にその機能がないかという点、もちろんそんなことは無い。上の例11における3行目の「そのような予測」がそれにあたる。小規模なテキスト構成ではあるが、「予測」というテキスト構成語が働いている。「自民党から失われた票がすべて社会党に流れれば、いずれは社会党が多数党になることもありえたであろう」の部分で、「～ば～であろう」という“辞的な次元”から「予測」という“詞的な次元”へのテキスト構成の転換も含んで、分節化している。このようなソ系の指示語がテキスト構成の補助を行っている場合も少なくない。

しかしながら、テキスト構成補助の機能の有無は問わないとして、4 資料全体でコ系の〈このような・このように・こうした・こうして〉の延べ語数合計が819であるのに対して、ソ系〈そのような・そのように・そうした・そうして〉の延べ語数合計は64と、十分の一以下であったこ

とからも推定できるように、ソ系の指示語がテキスト構成補助機能を発揮する場面は、コ系のそれと比較すると質・量ともに少ないのである。

この例にもコ系とソ系の、よく指摘される主観・判断、と客観・文脈指示という差も見て取れるが、それについては、機会を改めて、論じてみたい。

3.2.3 ド系の場合

本稿では、従来あまり注目されてこなかったド系に焦点を当てたいと考える。ド系は、先に挙げたコ・ソ系と同形式の〈どのような・どのように・どうして〉⁸⁾は、延べ語数合計で141出現しており、ソ系合計より多い。

これらは必ず後の方に向けて分節を形成する点がコ系、ソ系の指示語と異なる点であろう。たとえば、『政治学入門』の「IV 社会集団と政治」章の「14 女性の政治参加」という節における「積極的優遇措置」という項で、

(例10)

積極的優遇措置

こうした現状に接して、それは女性が管理能力や決定能力において、本来的に男性に劣るからであると考えたとすれば、それは女性に対して根拠のない偏見をもつことになるであろう。少なくとも大学までの教育の過程においてみるかぎりには、男性の本来的優位を裏づけるような事実はまったく存在しない。たしかに現実の社会においては、機会の乏しさのゆえに、女性が管理能力を十分に発揮できないでいることは少なくないであろう。しかし、それは女性が管理的職務につくことが稀であることの結果であり、その原因ではないと考えられる。それゆえ、ともかくも決定や管理にたずさわる職務にできるだけ多くの女性を登用することが目下の急務であるといわなければならない。そのためには、具体的にどのような方策が考えられるであろうか。

まず考えられるのは、政府が管理的職務につく女性を急増させるために必要な措置を講ずることである。こうした措置として参考になるのは、アメリカが少数派の格差是正のためにとってきた積極的優遇措置（アファーマティブ・アクション）であろう。積極的優遇措置とは、女性やアフリカ系アメリカ人など、これまで不利な立場に置かれてきた人々に雇用や教育の機会を保障するためにとられる措置を指している。具体的には、不当な差別を受けた者に対して法廷が下す救済命令、連邦政府と契約関係にある企業に対する大統領の行政命令などにより推進されてきた。この措置の結果、多くの企業は女性やアフリカ系アメリカ人の雇用比率を高めるためにクォータ（割当率）を設定し、その枠内で女性やアフリカ系アメリカ人を優先的に採用することになった。1970年代以降のアメリカで、女性の職場進出、職域拡大、管理職増加などがめざましいのは、この積極的優遇措置によるところ大であったといわれている。

⁸⁾ 「どうした」は形としてはありうるが、本稿で着目する談話構成としてのド系では用例が無い。

わが国では、目下のところこうした方策がとられる見込みは薄い。最近では、アメリカでも、保守化の傾向が強まるとともに、積極的優遇措置に対する批判的な議論が勢いを得ている。しかし、差別されている少数派（女性は数の上では少数派ではないが、その社会的実勢からいえば、少数派といてよい）の地位を急速に向上させるためには、政府による何らかの優遇策が必要なことも確かである。積極的優遇措置は、こうした優遇策の一つとして十分に考慮に値するものではなかろうか。たしかに、一挙に男女の格差をゼロにするような過激な積極的優遇措置は強い反発を招いて、かえって逆効果に終わるかもしれない。しかし、男女比を漸進的に均等化する方策もありうる。たとえば、さしあたって管理職における女性の比率を2割とか3割にする目標を設定して、積極的優遇措置をとることは、決して無理とはいえないであろう。

『政治学入門』「IV 社会集団と政治」章の「14 女性の政治参加」節 p129)

この例では、「どのような方策」が、以下で「方策」の内容が具体化されることを予想させ、その具体的な提示が終了する「こうした方策」の前までをひとまとまりとして分節となす。

【ド系の指示語＋～疑問詞：か】というのは、読み手に対する働きかけ表現という面もあるが、「このような」等と逆に後方に分節化が向かう機能に注目したい。この場合だと「方策」について述べられている後方の部分までを分節化することになり、かつ後方でそれらが確実に述べられていることを保証し予告する。

すなわち不定のド系の指示語で投げかけられた語句は、その不定が特定の部分となって呼応し分節化が完了するまで、ずっとペンディングとなるという、かなり強力なテキスト構成機能を持っているといえるだろう。

「どのように」はどうであろうか。さきほど検討したコ系では、「このような」と「このように」は異なる振る舞いをしていたが、やはり「～ような」と「～ように」では異なっているようである。

たとえば、

(例 11)

このような意味を持つロシア革命に、日本はどのように対応したのでしょうか？ これは史実を知らなくても、容易に想像できるでしょう。先にみたように、清朝崩壊後の辛亥革命に伴う中国の政治的混乱に乗じて、日本は21カ条要求を突きつけました。ロシア革命に際しても、日本はこの革命に武力で干渉します。これがシベリア出兵です。日本は米英など7カ国と共同出兵しました。目的は、反革命勢力を支援しシベリアに親日政権を確立することにあります。その後、他国が順次撤兵していったのは対照的に、派兵した国のなかで最大規模の駐兵を継続し、撤兵したのはもっとも遅れて1922年のことでした。

ところがその3年後の1925年1月に日本は日ソ基本条約を締結し、国交を樹立しています。イギリスがいったんは承認しながらすぐに断交したことや、1933年まで国交がなかったアメリカと比較すると、革命干渉戦争後の日本の対ソ関係改善は著しかったといえます。なぜでしょう？

ソ連社会主義への理解が深まったからでないことは、明らかです。コミンテルンの日本支部として1922年に結成された日本共産党は、ほどなくして治安警察法違反で一斉に検挙されています。また日ソ基本条約が調印されたのと同じ1925年の4月、治安維持法が公布されています。日ソ国交樹立に伴うソ連社会主義の日本国内へ及ぼす影響を考慮して、あらかじめ用意周到な準備がなされていたのです。

しかしそうだからこそ、対ソ関係の改善が進みました。国内の治安維持体制が強化されたこともあって、1920年代の労働組合の組織率は、わずか5、6パーセント。普通選挙制度下の総選挙で無産政党が獲得した議席数は、つねにひと桁。実際の政治過程での社会主義の影響力は、この程度のものでしたのです。このことがかえって、国家としてのソ連との外交関係改善のハードルを比較的安くしました。当初の「世界革命」の夢が破れ、「一国社会主義」体制へ移行したソ連に対して、日本外交はイデオロギーとは異なるレベルで接近します。日ソ基本条約の付属議定書に基づいて、両国は1925年12月、石油・石炭利権契約を結びました。ここに示されているように、資源開発や貿易への関心から、対ソ関係を改善して、経済外交を展開しようとしていたのです。体制の違いを超えて、日本はソ連と、戦後世界経済のネットワークのなかで、外交関係を設定しようとしたといい換えることもできます。ロシア革命という構造変動に、日本は以上のように対応しました。（『日本外交史講義』「第3章 国際強調の受容」「1 第1次世界大戦と国際政治の構造変動」の「C ロシア革命」の項 p55）

上の例は、ロシア革命という出来事に対する日本の対応の仕方について、「どのように〜か」の形で投げかけられて、「対応した」がテキスト構成の働きをしている。続く波線部から“革命に武力で干渉”“革命干渉後の対ソ関係改善は著しい”“イデオロギーとは異なるレベルで接近”“戦後世界経済のネットワークのなかで、外交関係を設定”などを「対応」として拾い上げる。そして、“革命には干渉したが、その後は世界経済ネットワークの中で対応した”という分節を作り上げて、「対応」の「どのように」の部分に定めるものとしている。この場合は、何がしかの活動として、用言的な語句の連続が「どのように」に定めるものとなっているため、「このように」の場合と同様に、分節の対象となっている部分は長大にわたることも少なくなかった。

今回の資料に限ってしか言えないが、たとえば、『刑法原論』の「1、はしがき」の第1文は「本書は、犯罪現象と法がどのようににかかわるかを、近代刑法の展開という観点から考えようとするものである。」となっている。つまりこの一冊の本全体が、「どのように」という不定詞を満たすことで成立しているとも言えるくらい、この指示語は長大な部分に「かかわる」という語を関係付ける力がある。

『政治学入門』の「まえがき」でも「本書では、現代の政治や社会にみられるさまざまな事象をできるだけ説明の素材として用いるようにしたが、それは、今日の政治学が現実政治の課題をどのように捉えているかを明らかにしたいためであるといつてよい。」となっており、また、『日本外交史講義』の「序章」でも「いい換えるとこの本の目的は、国民国家としての日本が、国際公法のルールの下、他の国民国家との間で勢力均衡のパワーゲームをどのように繰り広げたのか

を再現することにあります。」という「どのように」の使用例がある。

以上、指示語のテキスト構成補助の様相をみてきたが、やはり、テキスト構成のために、ある程度まとまりをつけるためには、名詞に収斂させる形である方が展開させやすいものと思われる。「～ような」「～ように」を比較すると、コ系ソ系ド系いずれも「～ような」のかたちの方が多く、理論的にはありうる「～ようだ」「～ようで」のかたちが殆ど無いのも、テキスト構成補助機能としては、テキスト構成語となるのが名詞の場合が多いことと連動しているのだと考えられる。

また、コ系の「このような」「このように」等は読解の現場に読み手の意識を引き付ける“現場指示”としてテキスト構成の援助に利用されており、ド系「どのような」「どのように」は疑問終助詞「か」を伴って、読み手の中に“解”を知りたいという欲望を喚起しようとする形式として利用されているように思う。とくに、ド系で掲げられたテキスト構成語に関しては、読み手は完全な“解”が得られたという確信が持てるまでテキストに没頭するに違いない。

もちろんこれらの指示語がすべてテキスト構成補助機能を発揮しているわけではない。ただ、前述した「指示語句」の考え方として、指示語は、テキストの変容、すなわちテキスト構成のマーカ―となることが多く、また、テキスト構成語だけではテキスト構成機能を十分に発揮できない場合もあるということはあると思う。

4. テキスト構成補助機能—語彙的結束の関係

テキストの文脈のなかで観察すると、分節の中の語彙と、談話構成語となる語、あるいは分節内の語同士のあいだには、同語反復や、いわゆる類義的な言い換え語の関係が見られることが多くある。今までみてきた例文にも、この関係を見出すことができる。

たとえば、先に挙げた例10では、「どのような」という指示語が補助をして「方策」が談話構成語として、以下にそれに組み合わされる分節のあることを予告していた。すなわち「方策」の内容が具体化されることを予想させ、その具体的な提示が終了するまでを、ひとまとまりとして分節となしていた。テキスト構成としては波線部がそれに相当する分節になるが、既に見出しに「積極的優遇措置」という語句が掲げられ、それを含めて「措置」という語が複合語になっているものまで含めると5回反復して出現する。すなわち分節内部は「措置」という語が反復されることで結束性を有し、ひとまとまりをなしていると言ってよい。

ハリディ／ハッサン(1997)では、「結束性は、テキスト内のある要素と、その要素の解釈に欠くことのできない他の要素との間の意味的な関係である。」(p.9)とし、「語彙的結束性」の言語体系における表示は「再叙⁹(語彙的指示の同一性)」や「コロケーション(語彙環境の類似)」によってなされる、とする。一方「文法的結束性」は、「指示、代用、省略、接続」によってなされる、とする。

この「方策」によって組み合わせられる分節の中に反復する「措置」は、一見「類義語」に見

⁹「再叙(reiteration)」の内容は、同一語の繰り返し、同義語や近似同義語、上位語、一般名詞(people, stuff, move などのような一般的指示をもつ名詞類)、人称指示語、だという。「コロケーション(collocation)」は、「同じ語彙的環境を共有すること」を意味し、類似した文脈に現れる傾向のある2つの語彙項目あるいは「長い結束性の連鎖」が構築されることとなると述べる。

えるのだが、語彙論的にいうと必ずしもそういう捉え方はできない。たとえば、『分類語彙表』では、

方策—1. 3084—08 [1. 3084 は計画・案、 同じ行の類義語は策・一策・万策・秘策]

措置—1. 3850—11 [1. 3850 は技術・設備・修理、同じ行の類義語は対処・善処・臨機応変]

となっており、「1. 3 人間活動—精神および行為」までしか共通ではない。念の為に『三省堂 類語新辞典』をひくと、

方策—L 活動—L1 思考—j 方法 措置—L 活動—L5 行為—g 整備・安定

というようにやはり「L 活動」までしか共通ではない。「方策」は思考であり「措置」は活動であるということで語彙論的には異なる範疇に属するようである。このように、テキスト中では、言い換えや具体化の中で相互に交換できるような意味で使用されていても、辞典やシソーラスにおける類義語や上位語・下位語等の体系への所属の仕方とは相違がみられることが少なくない。

またもう一つ注目したいこととして、石井（2011）では、

コーパス言語学とも関連の深いテキスト言語学は、語彙がそのテキスト構成機能によって構造化されていることに注目する。(p287)

として、単語のテキスト構成機能とは、M. A. K. ハリディ/R. ハッサン（1997）の「再叙」やマッカーシー、マイケル（1995）の「談話構成語」の機能である、と述べている。さらに続けて、

このうち、再叙にはいくつかの方策があるが、同義（類義）関係や上位・下位関係にある単語はその主要なメンバーである。これは、これらの語彙的な関係をもつ単語が、まさにそうした関係をもつがゆえに、再叙という機能を果たしつつテキストの中に発現したと考えることもできるが、逆に、以下のように再叙というテキスト構成の機能を果たすために、そうした語彙的な関係の語群が用意されている（語彙として構造化されている）と解することもできる。(p287)

として、マッカーシー（1995）の以下のような記述を引く。

つまり、同義語というものは、単に授業に出てくる新しい単語を理解する方法でもなければ、辞書や類語辞典を作るための抽象的な概念でもなく、他の言語的手段と同様に、自然な談話を作成するために存在しているのだ。(p94)

このことを、「語彙がなぜそのような用意され、構造化されているのかという問いに対するテキスト言語学からの解答がある。」とまで踏み込んで述べているほどである。ゆえに我々テキスト言語学の人間は、そこまで言ってもらえるのなら、やはりテキストが語の意味を作り出し、性格付けをし、テンポラリーに機能を与え続けたことで、語にテキスト自体を構成する力を持たせているのだ、と考えてもよからう。

しかしながら、石井（2011）からは「再叙としてはたらく一般名詞（general noun）や談話構成語としてはたらく単語群がどのような意味分野に多いか、すなわち、語彙の中にどのように位置づけられているかといった問いにも、それらの談話構成機能の観点から何らかの解答が期待できる。」という課題を出されてもいる。

語彙的結束性という観点から、この例文10「積極的優遇措置」全文をもう一度見てみると、「女性」という語の全体を通じた反復とともに、「管理」という語が単独で、あるいは「管理能力」「管理的職務」「管理職」といった複合語に含まれる要素として反復されていることがわかる。

このように、単純な1語の同語反復・類語反復ということだけではなく、日本語独特の、漢字1字が1語を表す表語文字という性格もあって、文脈形成のなかでその展開を反映して漢語どうしの何重もの複合・合成や、またその逆の“解体”された部分が現出することも少なくない。この例文だと、項目の題名にもなっている「積極的優遇措置」という合成語が、そのまま反復されたり、ばらけさせて、「措置」だけを取り出して反復させたり、「優遇策」という、「優遇」を取り出して「方策」と合体させるような語にしたりして、その反復や合成と解体の様相が語彙的な結束を作り上げている。石井（1999）ではこうしたテキストにおける語彙の様相を、「臨時一語化」と「脱臨時一語化」として考察している。

すなわち語彙の「テキスト構成機能」とは、ある語がテキスト構成語となって分節と組み合わせられて働くだけでなく、語彙が再叙されること（同語反復や語彙論的に上位下位同義・類義の語による言い換えなど）による結束性によってテキスト全体や、テキストを構成する分節を構成することにもある。そのなかではもちろん、語彙論的な関係だけでなく、当該テキスト限りの臨時的な結束関係もありうる。

そうした現象は、分節を作る目安となり、さらに今回のような長大なテキストであると、いくつかの専門用語などがテキスト全体で繰り返し出現し再叙されていく様相が観察される。

たとえば、『政治学入門』には「政党」という語がテキスト全体に、偏ることなく337回反復使用されている。『日本外交史講義』には「冷戦」が193回、『アメリカの経済』には「市場」が206回、『刑法原論』には「刑法」が588回反復使用されている。

加えて専門用語ではない、「問題」や「原則」「傾向」「状況」「変化」というような語が、その都度別個の内容を意味しつつ、テキスト構成語として頻用されて、結果的にテキスト全体に“同語反復”され、長い結束のつながりを形成しているかのように見えることもある。一般的な語が、分節とつながりを持ち、組み合わせられてテキスト構成語となってテキストが意味展開している様相も観察される。語彙的結束性自体が、テキスト構成機能と深く関係するといえるだろう。

5. まとめにかえて

以上、文章中の語彙の機能をみてきた。語彙は、テキストの中でテキストから意味を限定され、文法的に制限されつつ、また一方で辞書的自立的意味と、文脈的干渉を受けた文脈的意味とのほごまで揺れつつ、おのずからテキスト構成のような機能を持つようになったのだと思われる。「姿

勢」とか、「動き」とか、「潮流」とか、思いもかけないような具体性のある語が、大きな分節と組み合わせる姿をみることができた。また、「情況」「変化」「原則」といった典型的な一般名詞、「りんご」とか「空」などと比べて抽象度の相対的に高い語が、学術書テキストでなくてはならないテキスト構成語として頻度高く使用される様相も見ることができた。そしてテキスト構成語となる語には漢語が多いことから、日本語における漢語の使いこなしが進んでいることを感じさせられた。そのことは、助詞助動詞や副詞、接続詞、感動詞などには和語しかないと言われてきたが、漢語が文法語のほうに振れてくるということの意味しているのか、とも思われる。

テキスト構成語に組み合わせられる意味の分節という単位を仮定すると、分節の中は結束性を示す語彙が現れてまとまりを示し、またテキスト構成語と当該分節内の語彙もまた結束性を持ちうることもわかった。それが語彙論的結束性と一致しない、臨時的な結束性である場合も観察された。語彙的結束性でつながった語彙の身元確認として参照するシソーラスの蔵の棚の整理の仕方が、現実のテキストにおける語彙同士の関係性と合致しないことや、テキスト構成語で発揮される“辞書の第二義”というものの意味づけも考えさせられた。

テキスト構成語のある合図や補助としての指示語の働きにも注目し、その前後で文脈が展開・変容するパターンを旧稿“指示語句”の考え方を援用してさぐってみた。特にD系は予告的に働いて、かなり強力にテキスト構成に関与するということがわかった。

テキスト性とはテキストを単なる語の集合でなく、意味を持った存在たらしめるもので、意図が伝わるように語彙体系から来た語が再配置された様相を示すものであり、テキストとは語が機能面を発揮して、テキストの意味が正確に伝達されるような機能というものを示す場であることが実感された。テキスト内には、テキスト性を担うものはただ一つでなく、いくつもの結束や構成の手がかりが用意され、カバーし合って、テキスト内で働いているということもわかってきた。

こう見てくると、テキストの構成は文の構造よりもずっと複雑で、そのテキストの意味実現、伝達目的に少しでもかなうように、いろいろな規模の単位が組み合わせたり、入れ子のようになっはめこまれたりしつつ、時間的線条的一方的で流動的かつ言語量多量という条件を、平面的結果的雙方向的定着的かつ簡約に把握するという操作をしやすくするように調整する機能が存在している。

実際のテキストの中での語彙の働きをいろいろと見てきたが、実は、すべての語が、文章のなかで、内容と機能の両面でテキストを構成し、テキストに意味を持たせることに貢献しているわけである。たった1語でテキスト全体を受け止める（「題名」）こともあれば、1語以下の「こういう」の「こ」だけで、ある概念部分を分節化して取り出して次に展開させるという合図機能も果たせる自在さがテキストの中の語彙の様相にはあるのである。

ここで主として見たテキスト中の語彙の機能は、テキスト構成と結束性だけである。可能性としては、ほかにももっと機能があるに違いない。もちろん両者は実際のテキストの中では密接な関係性もちつつ、テキストを、意味的に一貫性を持った存在として、かつ内部において無数の分節から構成された存在として、すなわちテキストたらしめるために機能していた。コーパスで、

テキスト構成語を観察してみると、こうした意味・用法を持つ語が必要だったのにそれまで欠けていたので、文脈の中で生まれ、育っていった言語現象が定着してテキスト構成語になったのだというふうに考えられる。ちょうど上代に漢文を訓読するときに接続詞的なものが必要で、無理に他品詞から転成させてきて間に合わせたように、近代的な分析的文章のために比喩的でも欧米語直訳にしても、苦し紛れに引っ張ってきた語の使用が慣習化し、カスタマイズされて、その語の第二義として辞書にも載るようになったのであろう。ほかにもテキストの中でこそ生じた語彙のいろいろな側面があるに違いない。

ひとつやり残したことを述べたい。テキスト構成機能補助に指示語が活躍するのは、接続詞に指示語を含むものが多いことに関係があるものと思われる。テキスト構成に働く語は文法語に近く、形式化した語が来やすいことと、接続詞には形式化した「いう」「する」などがふくまれることも共通する。語彙語と文法語の中間的存在ということは、テキスト構成語としての文法語方向への揺れは、接続詞化する機能の可能性も含んでのことではないか。そのあたりを今後の課題としたい。

文献

- M. A. K. ハリディ/R. ハッサン (1991) 『機能文法のすすめ』 笈寿雄訳 大修館書店
- M. A. K. ハリディ/R. ハッサン (1997) 『テキストはどのように構成されるか』 安藤貞雄他訳 ひつじ書房
- 石井正彦 (1999) 「文章における『臨時一語化』と『脱臨時一語化』—脱臨時一語化の形式を中心に」『日本語研究』19:1-15 東京都立大学国語学研究室
- 石井正彦 (2011) 「隣接諸分野の語彙研究と『これからの語彙論』」 斎藤倫明・石井正彦編『これからの語彙論』275-291 ひつじ書房
- 金岡孝 (1963) 「主題と構成」 森岡健二・永野賢・宮地裕・市川孝編『講座 現代語第3巻 読解と鑑賞』36-56 明治書院
- 河上誓作編著 (1996) 『認知言語学の基礎』 研究社
- マッカーシー、マイケル (1995) 『語学教師のための談話分析』 安藤貞雄・加藤克美訳 大修館書店
- 野村益寛 (2003) 「認知言語学の史的・理論的背景」 辻幸夫編『認知言語学への招待』17-62 大修館書店
- 斎藤倫明 (2011) 「日本語学・言語学の諸分野とこれからの語彙論」 斎藤倫明・石井正彦編『これからの語彙論』255-274 ひつじ書房
- 高崎みどり (1988) 「文章展開における“指示語句”の機能」『国文学 言語と文芸』103:67-88 大塚国語国文学会
- 高崎みどり (2011) 「文章論・文体論と語彙」 斎藤倫明・石井正彦編『これからの語彙論』113-124 ひつじ書房
- 高崎みどり (2012) 「テキストの結束性に与る語彙とその機能について」『第2回コーパス日本語

学ワークショップ』予稿集 国立国語研究所言語資源系・コーパス開発センター
田中茂範・深谷昌弘 (1998) 『〈意味づけ論〉の展開 情況編成・コトバ・会話』紀伊國屋書店
*辞典類

『日本国語大辞典第2版』小学館 2001年

『三省堂 類語新辞典』中村明主幹 三省堂 2005年

『分類語彙表 増補改訂版』国立国語研究所 大日本図書 2004年

謝辞

本稿は2012年9月6日 国立国語研究所言語資源系・コーパス開発センター主催の『第2回コーパス日本語学ワークショップ』において、「文章における語彙の分布と文章構造」プロジェクト(プロジェクトリーダー:山崎誠)の一員として、発表した「テキストの結束性に与る語彙とその機能について」の内容を加筆修正したものに基づいております。当日ご質問やコメントを寄せくださった会場の先生方、プロジェクトのメンバーの皆様には感謝いたします。また、ワークショップ発表の機会および本稿執筆の機会を与えてくださったプロジェクトのリーダー、山崎誠先生に深く感謝いたします。

接続表現の二重使用と文章ジャンル

— 『現代日本語書き言葉均衡コーパス』を利用して—

馬場俊臣（北海道教育大学）

要旨

接続表現（接続詞及び接続詞的機能を果たす連語等）は、文章構造の分析（特に文の接続関係の分析）の重要な指標である。本稿では、接続表現の二重使用（特に異種併用）の出現の偏りと文章ジャンルとの関係の有無について試行的・探索的な調査分析を行った。二重使用に関する先行研究の成果を踏まえたうえで、BCCWJ（『現代日本語書き言葉均衡コーパス』モニター公開データ（2009年度版））を利用し、「書籍」（NDC 第一次区分別）「国会会議録」「白書」「Yahoo! 知恵袋」の文章ジャンルに分けて調査分析を行った。その結果、二重使用は、全般的に「国会会議録」での使用の比率が高いなど接続表現の二重使用の使用率に偏りがあること、二重使用の組み合わせの種類（異なり）は多いが頻用される組み合わせはある程度限られることなどが明らかとなった。

キーワード：接続表現、接続詞、二重使用、BCCWJ、文章ジャンル

1. はじめに

文章は文の連続として成立しており、連続する隣り合った二文の間の広義の論理的関係である「文の接続関係」を分析することは、文章の展開的構造を把握するためにきわめて重要である。文の接続関係を捉える指標として、主に接続表現¹（接続詞及び接続詞と類似の接続機能を持つ副詞や名詞、連語や文相当の表現なども含む）が利用される。

文の接続関係の分類は主に接続表現の意味用法の分類を手がかりとして行われてきており、原則的には、文と文との間に接続表現を想定し接続関係を判定することが多い。しかし、接続表現を想定できない二文があること、接続表現の二重使用があることなど、接続関係の分析に当たった問題がある。

本稿では、接続表現の二重使用に焦点を当て、接続表現の二重使用（特に異種併用）の出現の偏りと文章ジャンルとの関係の有無について試行的・探索的な調査分析を行い、文章構造の分析の指標となる接続表現に関する基礎的知見を得ることを目的とする。

まず、2章で二重使用に関する先行研究の成果及び文章ジャンルと接続表現の出現傾向の関係に関する先行研究の成果を概観する。次に、3章及び4章でそれらの成果を踏まえたうえで、BCCWJ（『現代日本語書き言葉均衡コーパス』モニター公開データ（2009年度版））を利用し、「書籍」（NDC 第一次区分別）「国会会議録」「白書」「Yahoo! 知恵袋」の文章ジャンルに基づいて接続表現の二重使用の実態の調査分析を行う。最後に、5章でその結果を示し今後の展望

をまとめる。

2. 先行研究の概観

2.1 接続表現の二重使用

接続表現の二重使用には、(1)の「だがしかし」(逆接型+逆接型)のように、同じ接続類型の接続表現が重ねて用いられる「同種併用」と、(2)の「そして一方」(添加型+対比型)のように、異なる接続類型の接続表現が重ねて用いられる「異種併用」とがある。

(1) …三十八, 三十九, 四十, 四十一と小坊主が柄杓で茶を汲むたびごとに、境内から声があがるようになった。たしかに不思議な釜だ。まだ茶が尽きない。だがしかし, こんな小坊主が光泉寺にいたというのも不思議である。(井上ひさし「腹鼓記」)(BCCWJ)²

(2) 福江島一つとっても、堂崎の教会や水の浦の教会など、歴史も古く堂々たるものであった。そして一方, 宮原教会のように、ほとんど民家のような家の上に、ぽつんと十字を乗せたものまである。(中西進「日本のかたち ころの風景から」)(BCCWJ)

同種併用の場合は、文の接続関係の認定は問題ない。しかし、二重使用の理由や二重使用の可能な接続類型やその際の順序などの検討が必要である。

異種併用の場合は、接続関係の認定をどう扱うかが大きな問題となる。二重使用による文脈展開のあり方を検討する必要がある。また、異種併用が可能な接続類型やその際の順序などの検討も必要である。³

2.2 二重使用の実態・傾向に関する研究成果

接続表現の二重使用の実例の提示及びその研究の必要性は、早くは、市川(1957)(1963)(1978)や森田(1958)などで示されてきた。その後、内省や少数の用例に基づいた部分的な指摘や分析は行われてはいたが、大量の用例に基づいて二重使用の可能な接続類型の組み合わせや二重使用の際の順序などを明らかにする研究は2000年以降にならないと現れない。⁴

本節では、二重使用に関する実態・傾向を分析した研究の成果を概観する。

まず、馬場(2003)(2006)では、市川(1978)の7種の接続類型⁵に基づき、すべての組み合わせの二重使用(同種併用・異種併用)の実例を可能な限り収集し、二重使用の可能な接続類型の組み合わせの実態と特徴を明らかにしている。市川(1978)で例示している「接続語句」90種類の組み合わせ(90種類×90種類=8100組)を網羅的に調べている。調査範囲(コーパス)は、CD-ROM版『新潮文庫の100冊』(翻訳作品は除外),『朝日新聞 天声人語・社説1985-1991増補改訂版』,『朝日新聞 一面記事1989-1993秋』,『CD-毎日新聞'93』,電子ブック版『日本大百科全書 改訂第2版』である。

二重使用の可能な組み合わせを表1のようにまとめている。なお、「×」(実例を見つけていない組み合わせ)は、あくまでも実例を見つけていない組み合わせであり、必ずしも不自然な組み合わせであることを示しているわけではないことに留意しておく必要がある。

全体的な傾向として、同種併用に関しては、逆接型、添加型、対比型、転換型の同種併用は見られたが、順接型、同列型、補足型は見られなかったこと、異種併用に関しては、大きく「逆接型 転換型」「添加型 対比型」「同列型」の順で先行しやすいこと、順接型、補足型の異種併用は極めてまれであることを指摘している。

表1 馬場 (2003: 16) (2006: 140)の「接続詞の二重使用—接続類型別のまとめ」

先行 後続	順接型	逆接型	添加型	対比型	転換型	同列型	補足型
順接型	×	×	△	×	△	△	×
逆接型	△	○	×	×	×	×	×
添加型	×	○	○	○	○	×	×
対比型	○	○	○	○	○	×	×
転換型	×	△	×	×	○	×	×
同列型	○	○	○	○	○	×	○
補足型	×	×	×	×	×	×	×

注 「○」は実例を見つけている組み合わせ。
「×」は実例を見つけていない組み合わせ。
「△」は稀な用例あるいは限られた用例であることを示す。

次に、石黒 (2005) は、新聞における接続詞の二重使用の実態を定量的に分析し、二重使用のタイプや使用理由、表現効果を考察している。調査対象語は、市川 (1978) の 89 種類⁶の接続語句の出現頻度調査を行った結果、使用頻度 1000 以上となった 14 語の組み合わせ (14 語×14 語=196 組)である (「だから」「しかし」「だが」「でも」「ところが」「そして」「さらに」「しかも」「また」「一方」「さて」「たとえば」「ただ」「もっとも」の 14 語)。『毎日新聞 CD-ROM』2001 年度版を使用して調査を行っている。考察の結果、二重使用の際の相互承接の順序の原則は「転換型>逆接型>添加型>順接型・対比型>同列型」となること、また二重使用による表現効果として「接続詞の意味の限定・強調 (類似の意味の接続詞を重ねる)」「(そしてまた、だがしかし、しかし一方等)」「複線的文脈の提示 (前後の文脈の接続に複数の見方があることを示す)」「(しかしさらに、しかしまた、しかしだから等)」「多層的構造の提示 (後続の文脈の入れ子型構造を示す)」「(しかしたとえば、しかしだからこそ等)の三つが挙げられること、さらに「二重使用された接続詞は、書き手がとくに明確な表現意図をもって用いるものなので、使用頻度こそ少ないものの、とりわけ特徴的な文体分析の指標になる」ことを示している。

さらに、多田 (2010) は、「複合接続詞の全体像」(一語からなる接続詞との共通性、文頭における副詞との語順、文法化の程度、文頭用法の語順など)を考察しており、調査分析の一部に接続表現の二重使用の調査が含まれている。「BCCWJ2008 年度版の書籍データ」を使用し、まず「BCCWJ 全体を Himawari で検索して語順の先後を調べる」ことを行い、そのうえで「これに複合接続詞を加え」て二重使用の傾向を示している。取り上げているのは、「が、そして、し

かし、しかも、じゃあ、あるいは、なお（猶）、さらに（更に）、かつ、また（又）、ただ、いっぽう（一方）」の12語、及び「だが、でも、さて、ところが、ただし、ところで、まあ、そこで、そうすると、そのときは、つまり、だから、それから、そうして、それで、だけど、で」の17語である。

先行しやすいのは「逆接」の接続詞と「そして」と「また・ただ」であること、後続しやすいのは「時間的な関係を表す接続詞」（それから等）と「ことがらを並べて比較するような接続詞」（その一方等）であること、形態的な制約もあること（「と」「だ・です・な」などで始まる複合接続詞は他の接続詞が先行しにくい）こと、「接続詞性の強い接続詞は文頭に位置し、構成的な意味が残る接続詞は文の内側へ語順が下がる」ことなどを指摘している。

以上のように、接続表現の二重使用に関する大量の用例に基づいた実証的な研究の成果が着実に示されてきているが、文章ジャンルによる出現傾向の違いについては明らかにされていない。

2.3 文章ジャンルと接続表現の出現傾向の関係に関する研究成果

本節では、文章ジャンルの違いと接続表現の出現傾向の違いとの関係を分析した近年の研究成果を簡潔に示す。

まず、村田（2000）（2002）（2007）では、詳細な統計的分析に基づいて、論述的文章（経済学教科書、物理学・工学・文学論文）に共通する語句として「したがって、すなわち、たとえば」等が挙げられること、文学論文は「つまり」「たとえば」などの頻度が高く、理工学論文では「したがって」などの頻度が高いことなどを指摘している。

高澤（2002）では、論説文では「しかし、だが、しかも」が多用され、随筆では「しかし、そして、だから」が多用されることを指摘している。

高橋（2005）では、日常談話では「で」「でも」が多用され、講義では「で」が多用されること、また、講義では「そうすると、そして」「つまり、たとえば」が多用されるのに対し、討論では「だから」「ただ」が多用されることを指摘している。

清水（2006）では、新聞コラム（主観的な談話）での接続詞の出現率は報道文（客観的な談話）の2倍以上であること、報道文では「逆接」と「添加」の接続詞が多くコラムでは「逆接」の接続詞が多いことなどを指摘している。

石川（2010）では、「現代日本語書き言葉均衡コーパス（検索デモンストレーション版）の書籍サブコーパス」を利用し、「NDC（日本十進分類法）に基づく10種類の内容ジャンル」別の「したがって」「だから」「こうして」「このため」「ゆえに」の5種類の接続表現の使用頻度に基づいた因子分析を行い、「状況的論理接続因子と主観的論理接続因子が抽出され、前者には「このため」「したがって」「こうして」が、後者には「だから」「ゆえに」が含まれることが示された。また、前者は社会科学などの論説性の高い内容ジャンルと、後者は文学や哲学などの随筆的内容ジャンルとそれぞれ関係することも示された。」としている。

以上のように、単独の接続表現に関しては、文章ジャンルの違いにより接続表現の出現傾向の違いがあることが明らかにされている。本稿の調査研究は、二重使用された接続表現を対象とし

て、文章ジャンルの違いによる出現傾向の違いを明らかにしようとするものである。

3. 文章ジャンルによる偏りの分析(1)——高頻度組み合わせ調査

3.1 目的・使用コーパス

本研究では、接続表現の二重使用の文章ジャンルによる偏りの実態を明らかにするために、「高頻度組み合わせ調査」及び「高頻度語調査」と名付けた2種類の調査を行った。本章では、使用頻度が高いと予想される組み合わせを対象とした「高頻度組み合わせ調査」の結果を報告する。コーパスはBCCWJ(『現代日本語書き言葉均衡コーパス』モニター公開データ(2009年度版))を使用した(全文検索システム『ひまわり』使用)。なお、文頭(前文脈「。」)のみの検索を行った。また、「再び」の意味の「また」や「一層」の意味の「さらに」などのような副詞用法は目視により除外した。さらに、「しかし、だから害鳥である、ということにはならない。」のような包摂関係⁷にある例も除外した。

3.2 対象語の選定・調査方法

使用頻度が高いと予想される組み合わせの選定には、石黒(2005)及び多田(2010)の調査結果を利用した。石黒(2005)の調査結果の使用頻度上位5組(しかし一方、だが一方、しかしたとえば、そしてまた、ただ一方)及び多田(2010)の調査結果の使用頻度上位5組(あるいはまた、そしてまた、なおかつ、しかしまた、そしてさらに)の合計9組(1組は重複)をまず選んだ(「なおかつ」は副詞として扱われるが、ここでは多田に従い対象語に含める)。ただし、この9組では、すべての接続タイプの組み合わせが考慮されていないため、独自の調査を行い、「また一方」「ただししかし」の2組を追加し、計11組を選定した(表2参照)。

独自の調査の方法を簡潔に示す。まず、BCCWJで市川(1978)に示されている90種類の接続表現の頻度調査を行い、7種の接続タイプのそれぞれ上位1語(しかし、だから、一方、また⁸、ところで、つまり、ただ)を取り出し(表3参照)、その7語の組み合わせ(7語×7語=49組)の二重使用の頻度調査を行った(表4参照)。この二重使用の頻度調査の結果、相対的に頻度が高い「また一方」「ただししかし」の2組を追加した。なお、「しかし一方」「しかしまた」「ただ一方」は既に調査対象に含まれている。

表2 高頻度組み合わせの選定

組み合わせ	石黒(2005)		多田(2010)		本調査		類型
	頻度	順位	頻度	順位	頻度	順位	
しかし一方	56	1	6	26	106	2	逆接型+対比型
だが一方	36	2	--	--	17	10	逆接型+対比型
しかしたとえば	21	3	-	-	32	7	逆接型+同列型
そしてまた	18	4	363	2	243	1	添加型+添加型
ただ一方	11	5	0		11	11	補足型+対比型
しかしまた	10	6	106	4	41	6	逆接型+添加型
そしてさらに	9	7	96	5	59	4	添加型+添加型
あるいはまた	-	-	560	1	92	3	対比型+添加型
なおかつ	-	-	332	3	18	9	(連語相当・添加)
また一方(追加)	3	15	2	37	47	5	添加型+対比型
ただししかし(追加)	0		22	15	19	8	補足型+逆接型

注 「-」は対象外 「--」は詳細不明

表3 市川(1978)掲載の接続表現の頻度調査結果

類型	接続表現	頻度	全体順位
逆接型	しかし	21026	1
順接型	だから	6682	5
対比型	一方	2051	23
添加型	また	18237	2
転換型	ところで	563	46
同列型	つまり	4749	10
補足型	ただ	5591	7

表4 高頻度7語の組み合わせによる頻度調査結果

		先行						
		しかし	だから	一方	また	ところで	つまり	ただ
後続	しかし	-	0	1	1	0	0	19
	だから	8(だからこそ)	-	0	2(だからこそ)	0	0	0
	一方	106	0	-	47	0	0	11
	また	41	1	6	-	2	0	2
	ところで	0	0	0	0	-	0	0
	つまり	0	4	0	0	0	-	0
	ただ	1	0	0	0	0	0	-

3.3 使用頻度の調査結果

11組の二重使用の使用頻度の集計結果を示す。

まず、各組の4ジャンル別の出現頻度は表5の通りである。なお、表中の「BK」「MD」「WR」「YC」は、それぞれ「書籍」「国会会議録」「白書」「Yahoo! 知恵袋」を表す。また、「書籍」のNDC第一次区分別の出現頻度は表6の通りである。

表5 高頻度組み合わせ 11 組の出現頻度 (ジャンル別)

組み合わせ	BK	MD	WR	YC	合計
そしてまた	93	149	1	0	243
しかし一方	59	26	18	3	106
あるいはまた	30	62	0	0	92
そしてさらに	43	15	1	0	59
また一方 (追加)	14	25	8	0	47
しかしまた	30	9	0	2	41
しかしたとえば	27	3	1	1	32
ただしかし (追加)	2	17	0	0	19
なおかつ	6	10	0	2	18
だが一方	17	0	0	0	17
ただ一方	4	7	0	0	11

表6 高頻度組み合わせ 11 組の出現頻度 (書籍 NDC 第一次区分別)

組み合わせ	0 総記	1 哲学	2 歴史	3 社会科学	4 自然科学	5 技術・工学	6 産業	7 芸術・美術	8 言語	9 文学	分類なし	BK 集計
そしてまた	5	13	8	21	9	1	0	3	4	27	2	93
しかし一方	1	4	6	17	4	2	1	4	3	15	2	59
あるいはまた	0	4	2	11	2	1	0	1	2	6	1	30
そしてさらに	0	8	7	7	2	1	2	1	1	13	1	43
また一方 (追加)	0	0	2	8	0	0	0	1	0	3	0	14
しかしまた	0	7	3	8	0	0	0	1	2	8	1	30
しかしたとえば	0	5	4	6	1	3	1	1	2	4	0	27
ただしかし (追加)	0	1	0	0	0	0	0	0	0	1	0	2
なおかつ	1	1	1	2	0	0	0	0	0	1	0	6
だが一方	0	2	0	3	0	1	0	1	0	10	0	17
ただ一方	0	0	0	2	0	1	0	0	1	0	0	4

3.4 分析方法

「そしてまた」を例にして、分析方法を示す。

まず、「使用率」を次の2種類に分けて算出する。

一つは、先行接続表現「そして」に対する使用率（「先行表現比」）であり、「二重使用の組み合わせの使用頻度 ÷ 先行接続表現の使用頻度 × 100」で算出する。もう一つは、後続接続表現「また」に対する使用率（「後続表現比」）であり、「二重使用の組み合わせの使用頻度 ÷ 後続接続表現の使用頻度 × 100」で算出する。

「合計」（平均）の2倍以上の使用率を示す箇所を「偏り」と見做し注目する。表では、「合計」（平均）の2倍以上の使用率の箇所を網掛けで、2.5倍以上を網掛け・斜体で示す。ただ、使用頻度が極端に少ない場合は使用率が高くても偏りがあるかどうかの判断が難しい場合も

あるので、使用頻度 10 以上の組み合わせには下線を付す（表 7、表 8 参照）。

表 7 「そしてまた」の先行表現比・後続表現比（ジャンル別）

	BK	MD	WR	YC	合計
そしてまた	<u>93</u>	<u>149</u>	1	0	243
そして	12987	1746	149	590	15472
また	8759	2163	5688	1627	18237
先行表現比	0.72	8.53	0.67	0.00	1.57
後続表現比	1.06	6.89	0.02	0.00	1.33

表 8 「そしてまた」の先行表現比・後続表現比（書籍 NDC 第一次区分別）

	0 総記	1 哲学	2 歴史	3 社会科学	4 自然科学	5 技術・工学	6 産業	7 芸術・美術	8 言語	9 文学	分類なし	BK 集計
そしてまた	5	<u>13</u>	8	<u>21</u>	9	1	0	3	4	<u>27</u>	2	93
そして	231	1179	1211	2808	644	472	314	664	198	4980	286	12987
また	221	640	953	3250	783	604	412	379	178	1232	107	8759
先行表現比	2.16	1.10	0.66	0.75	1.40	0.21	0.00	0.45	2.02	0.54	0.70	0.72
後続表現比	2.26	2.03	0.84	0.65	1.15	0.17	0.00	0.79	2.25	2.19	1.87	1.06

3.5 調査結果——「合計」（平均）の 2 倍以上の使用率を示す偏り

以下、表 9～表 18 に、「合計」（平均）の 2 倍以上の使用率を示す偏りが認められる組み合わせの集計結果を示す。なお、表中の「合計」は 4 ジャンル（書籍、国会会議録、白書、Yahoo! 知恵袋）の集計結果であり、「BK 集計」は「書籍」のみの合計である。

表 9 「しかし一方」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	WR	合計	8 言語	分類なし	BK 集計
しかし一方	<u>26</u>	<u>18</u>	106	3	2	59
しかし	2155	944	21026	311	213	16885
一方	238	765	2193	20	11	1009
先行表現比	1.21	1.91	0.50	0.96	0.94	0.35
後続表現比	10.92	2.35	4.83	15.00	18.18	5.85

表 10 「あるいはまた」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	合計	8 言語	分類なし	BK 集計
あるいはまた	<u>62</u>	92	2	1	30
あるいは	378	1371	24	11	952
また	2163	18237	178	107	8759
先行表現比	16.40	6.71	8.33	9.09	3.15
後続表現比	2.87	0.50	1.12	0.93	0.34

表 11 「そしてさらに」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	合計	1 哲学	分類なし	BK 集計
そしてさらに	<u>15</u>	59	8	1	43
そして	1746	15472	1179	286	12987
さらに	596	4169	160	24	2435
先行表現比	0.86	0.38	0.68	0.35	0.33
後続表現比	2.52	1.42	5.00	4.17	1.77

表 12 「また一方」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	合計	BK 集計
また一方	<u>25</u>	47	14
また	2163	18237	8759
一方	238	2193	1009
先行表現比	1.16	0.26	0.16
後続表現比	10.50	2.14	1.39

表 13 「しかしまた」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	合計	1 哲学	8 言語	分類なし	BK 集計
しかしまた	9	41	7	2	1	30
しかし	2155	21026	1421	311	213	16885
また	2163	18237	640	178	107	8759
先行表現比	0.42	0.19	0.49	0.64	0.47	0.18
後続表現比	0.42	0.22	1.09	1.12	0.93	0.34

表 14 「しかしたとえば」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	合計	1 哲学	2 歴史	BK 集計
しかしたとえば	32	5	4	27
しかし	21026	1421	1706	16885
たとえば	4738	279	235	3302
先行表現比	0.15	0.35	0.23	0.16
後続表現比	0.68	1.79	1.70	0.82

表 15 「ただししかし」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	合計	1 哲学	9 文学	BK 集計
ただししかし	<u>17</u>	19	1	1	2
ただ	1563	5591	177	1312	2791
しかし	2155	21026	1421	4971	16885
先行表現比	1.09	0.34	0.56	0.08	0.07
後続表現比	0.79	0.09	0.07	0.02	0.01

表 16 「なおかつ」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	YC	合計	0 総記	1 哲学	2 歴史	3 社会科学	BK 集計
なおかつ	<u>10</u>	2	18	1	1	1	2	6
なお	647	182	2829	25	63	151	451	1057
かつ	23	3	55		2	2	4	27
先行表現比	1.55	1.10	0.64	4.00	1.59	0.66	0.44	0.57
後続表現比	43.48	66.67	32.73	0.00	50.00	50.00	50.00	22.22

表 17 「だが一方」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	合計	1 哲学	5 技術工学	9 文学	BK 集計
だが一方	17	2	1	10	17
だが	5128	231	118	3022	5097
一方	2193	61	71	153	1009
先行表現比	0.33	0.87	0.85	0.33	0.33
後続表現比	0.78	3.28	1.41	6.54	1.68

表 18 「ただ一方」の先行表現比・後続表現比（ジャンル別・書籍 NDC 第一次区分別）

	MD	合計	3 社会科学	5 技術工学	8 言語	BK 集計
ただ一方	7	11	2	1	1	4
ただ	1563	5591	541	80	44	2791
一方	238	2193	382	71	20	1009
先行表現比	0.45	0.20	0.37	1.25	2.27	0.14
後続表現比	2.94	0.50	0.52	1.41	5.00	0.40

3.6 文章ジャンルにより偏りが認められる組み合わせ

文章ジャンルにより偏りが認められる組み合わせについてまとめると次のようになる。

- ① 「国会会議録」で偏って使用される組み合わせが多い。これらの組み合わせは、全体で「添加型」「対比型」「補足型」の機能を果たしていると見られる。
 - ・頻度 10 以上 2.5 倍以上
「そしてまた」（添加型+添加型）、「あるいはまた」（対比型+添加型）、「また一方」（添加型+対比型）、「ただししかし」（補足型+逆接型）
 - ・頻度 10 以上 2 倍以上
「しかし一方」（逆接型+対比型）、「そしてさらに」（添加型+添加型）、「なおかつ」（補足型+添加型）（全体で慣用的）
 - ・頻度 9 以下 2.5 倍以上
「ただ一方」（補足型+対比型）
 - ・頻度 9 以下 2 倍以上
「しかしまた」（逆接型+添加型）
- ② 白書では、「しかし一方」が多用されている。
- ③ 書籍 NDC 第一次区分別では、「1 哲学」（「そしてさらに」「しかしまた」など）「8 言語」（「そしてまた」など）で偏りが見られるが、いずれも頻度が 10 以下である。たまたま使用されたとも考えられるが、他の区分では使用率が低いことは留意しておく必要がある。

4. 文章ジャンルによる偏りの分析(2)——高頻度語調査

4.1 目的・使用コーパス

本章では、予め二重使用の組み合わせを限定しないで用例を調査する「高頻度語調査」の結果を報告する。この調査は使用頻度が高い接続詞が先行する組み合わせを対象とした調査である。使用コーパス（BCCWJ）や検索条件は前章の「高頻度組み合わせ調査」と同じである。

4.2 対象語の選定・調査方法

先行表現となる高頻度語（対象語）の選定は次の手順で行った。

まず、市川（1978）の90種類の接続表現の頻度調査をBCCWJで行い上位5語を選定した。「しかし」（逆接型）（1位）（頻度 21026）、「また」（添加型）（2位）（頻度 18237）⁹、「そして」（添加型）（3位）（頻度 15472）、「でも」（逆接型）（4位）（頻度 7549）、「だから」（順接型）（5位）（頻度 6682）の5語である。

この5語には、対比型、転換型、同列型、補足型の接続表現が含まれていないため、それぞれの接続類型の使用頻度上位1語の計4語を追加した。「一方」（対比型）（全体順位 23位）（頻度 2051）、「ところで」（転換型）（同 46位）（頻度 563）、「つまり」（同列型）（同 10位）（頻度 4749）、「ただ」（補足型）（同 7位）（頻度 5591）の4語である。¹⁰

さらに、「ところで」の二重使用例が極めて少ないために「さて」（転換型）（同 51位）（頻度 479）を追加した。また、逆接型の接続表現の比較のために「だが」（逆接型）（同 8位）（頻度 5128）を追加した。

以上の計11語を対象語として選定し、これらの語が先行表現となる二重使用の用例を収集した。後続表現は目視で抽出していった。その際、市川（1978）に示されている90種類の接続表現（以下「市川」と略称）以外にも、広く接続表現（多くは複合的）と認められる表現（以下「市川以外」と略称）も含めた。

なお、「しかし」を先行表現とする二重使用については「書籍」「国会会議録」「白書」「Yahoo! 知恵袋」の4分類での偏りとともにさらに書籍 NDC 第一次区分での偏りも調査したが、「しかし」以外を先行表現とする二重使用については「書籍」「国会会議録」「白書」「Yahoo! 知恵袋」の4分類での偏りのみを調査した。

4.3 調査結果・分析

4.3.1 「しかし」を先行表現とする二重使用の文章ジャンルによる偏り

調査結果及び分析方法の具体的な例示として、「しかし」を先行表現とする二重使用の場合を以下に示す。

まず、二重使用の組み合わせ（後続表現が「市川」の場合）の異なり組数は40組あり、そのうち、使用頻度（合計）10以上の組み合わせを示すと表19のようになる。

また、二重使用の組み合わせ（後続表現が「市川以外」の場合）の異なり組数は44組あり、そのうち、使用頻度（合計）10以上の組み合わせを示すと表20のようになる。

文章ジャンルによる偏りを明らかにするために、「しかし」の出現頻度（総数）に対する二重使用の比率を算出した。

表21の「しかし総数」は「しかし」の出現頻度、「しかし+市川」「しかし+市川以外」はそれぞれ「市川」「市川以外」を後続表現とする二重使用の出現頻度である。「しかし+市川 比率」は「 $\frac{\text{「しかし+市川」}}{\text{「しかし総数」}} \times 100$ 」, 「しかし+市川以外 比率」は「 $\frac{\text{「しかし+市川以外」}}{\text{「しかし総数」}} \times 100$ 」で求めた比率であり、「合計 比率」は「しか

し+市川 比率」と「しかし+市川以外 比率」の合計である。偏りを視覚的に明示するために、「合計」の比率の2倍以上となる場合は網掛け、2.5倍以上となる場合は網掛け・斜体にしてある（表 21 には該当箇所はない。表 25 以降参照。）。

「しかし」は、「国会会議録」「白書」での比率が、「合計」の2倍以上にはなっていないが高くはなっており、多く使用されていることが分かる。表 22 に、「国会会議録」「白書」での使用頻度の高い組み合わせを示した。「しかし一方」「しかし同時に」などが多用されていることが分かる。

表 23 は、書籍 NDC 第一次区分別の表である。「1 哲学」「8 言語」での「合計 比率」が高くなっている。表 24 に、「1 哲学」「8 言語」での使用頻度の高い組み合わせを示した。「しかし同時に」「しかしまた」などの使用頻度が高い。「1 哲学」では「しかしだからといって」「しかしたとえば」などが多いことなど、「国会会議録」「白書」とはやや異なった傾向にあることが分かる。

「しかし」以外の他の先行表現についても同様の分析を行った。ただし、書籍 NDC 第一区分で細分類すると各欄の値が小さくなりすぎるため、「書籍」「国会会議録」「白書」「Yahoo! 知恵袋」の4分類での偏りのみを見ていく。

表 19 先行表現「しかし」の二重使用の使用頻度（合計）10以上の組み合わせ（後続表現が「市川」の場合）

しかし+市川	BK	MD	WR	YC	合計
しかし一方	59	26	18	3	106
しかしまた	30	9	0	2	41
しかしたとえば	27	3	1	1	32
しかし他方	17	7	7	0	31
しかしそのとき	22	0	0	1	23
しかし少なくとも	15	4	0	0	19
しかしやがて	17	0	0	0	17
しかしそれにもかかわらず	15	1	0	0	16
しかし逆に	12	1	1	1	15
しかしそのためには	8	2	2	0	12
しかしそもそも	10	0	0	1	11
しかしでは	7	4	0	0	11
しかしそれなら	9	2	0	0	11

表 20 先行表現「しかし」の二重使用の使用頻度（合計）10以上の組み合わせ（後続表現が「市川以外」の場合）

しかし+市川以外	BK	MD	WR	YC	合計
しかし同時に	54	22	4	0	80
しかしそれでも	45	3	1	3	52
しかしいずれにしても	22	20	0	0	42
しかしその一方	28	6	5	0	39
しかしだからといって	34	2	1	1	38
しかし結局	28	4	0	1	33
しかしそれにしても	18	5	0	0	23
しかし反面	7	5	1	0	13
しかしその反面	10	1	1	0	12
しかしそれと同時に	8	2	1	0	11
しかしそうはいつでも	10	0	0	0	10
しかしそれでもなお	9	0	0	1	10

表 21 先行表現「しかし」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
しかし総数	16885	2155	944	1042	21026
しかし+市川	319	75	29	9	432
しかし+市川以外	316	82	16	7	421
しかし+市川 比率	1.89	3.48	3.07	0.86	2.05
しかし+市川以外 比率	1.87	3.81	1.69	0.67	2.00
合計 比率	3.76	7.29	4.77	1.54	4.06

表 22 先行表現「しかし」の使用頻度の高い組み合わせ（ジャンル別）

	MD	WR
しかし一方	26	18
しかしまた	9	0
しかし他方	7	7
しかし同時に	22	4
しかしいずれにしても	20	0

表 23 先行表現「しかし」の出現頻度（総数）に対する二重使用の比率（書籍 NDC 第一次区分別）

	0 総記	1 哲学	2 歴史	3 社会科学	4 自然科学	5 技術・工学	6 産業	7 芸術・美術	8 言語	9 文学	分類なし	BK 集計
しかし総数	384	1421	1706	4870	1060	728	489	732	311	4971	213	1688 5
しかし+市川	5	42	33	87	14	10	8	13	10	91	6	319
しかし+市川以外	6	37	31	105	10	12	7	12	7	86	3	316
しかし+市川 比率	1.30	2.96	1.93	1.79	1.32	1.37	1.64	1.78	3.22	1.83	2.82	1.89
しかし+市川以外 比率	1.56	2.60	1.82	2.16	0.94	1.65	1.43	1.64	2.25	1.73	1.41	1.87
合計 比率	2.86	5.56	3.75	3.94	2.26	3.02	3.07	3.42	5.47	3.56	4.23	3.76

表 24 先行表現「しかし」の使用頻度の高い組み合わせ（書籍 NDC 第一次区分別）

	1 哲学	8 言語
しかしまた	7	2
しかしたとえば	5	2
しかし他方	5	0
しかしそのとき	4	0
しかし一方	4	3
しかし同時に	8	3
しかしだからといって	11	0

4.3.2 「また」を先行表現とする二重使用の文章ジャンルによる偏り（表 25 参照）

「国会会議録」での比率が高い。使用頻度が高い組み合わせは、「また一方」（添加型+対比型）（書籍 14 回）（国会会議録 25 回），「また同時に」（添加型+添加型）（書籍 24 回）（国会会議録 15 回），「またさらに」（添加型+添加型）（書籍 12 回）（国会会議録 14 回）などである。

4.3.3 「そして」を先行表現とする二重使用の文章ジャンルによる偏り（表 26 参照）

「国会会議録」での比率が高い。使用頻度が高い組み合わせは、「そしてまた」（添加型+添加型）（書籍 93 回）（国会会議録 149 回），「そしてさらに」（添加型+添加型）（書籍 43 回）（国会会議録 15 回）などである。

4.3.4 「でも」を先行表現とする二重使用の文章ジャンルによる偏り（表 27 参照）

使用頻度は「書籍」に多い。「国会会議録」の比率は高いが、用例数は 1 例であり、見かけ上高くなっているだけである。

4.3.5 「だが」を先行表現とする二重使用の文章ジャンルによる偏り（表 28 参照）

使用頻度は、「書籍」に多い。「国会会議録」の比率は高いが、用例数は 8 例であり、見かけ

上高くなっているだけである。8例すべてが「だがしかし」（逆接型＋逆接型）である。

逆接型の接続表現の「しかし」「でも」「だが」については、国会会議録では「しかし」は多用されているが、「でも」「だが」はほとんど使われていないことが分かる。

4.3.6 「だから」を先行表現とする二重使用の文章ジャンルによる偏り（表 29 参照）

使用頻度は「書籍」に多い。「国会会議録」での二重使用は 21 例であり、比率は見かけ上高くなっているだけである。使用頻度が相対的に高い組み合わせは、「だからたとえば」（順接型＋同列型）（書籍 9 回）（国会会議録 5 回）、「だからむしろ」（順接型＋対比型）（書籍 7 回）（国会会議録 2 回）などである。

4.3.7 「一方」を先行表現とする二重使用の文章ジャンルによる偏り（表 30 参照）

使用頻度は合計 8 でありごく少ない。「一方また」（対比型＋添加型）（書籍 3 回）（国会会議録 3 回）などがある程度である。

4.3.8 「ところで」を先行表現とする二重使用 の文章ジャンルによる偏り（表 31 参照）

使用頻度は合計 8 でありごく少ない。「ところでいったい」（転換型＋転換型）（書籍 3 回）（国会会議録 1 回）などがある程度である。

4.3.9 「さて」を先行表現とする二重使用の文章ジャンルによる偏り（表 32 参照）

「国会会議録」は二重使用は 26 例であり、比率は見かけ上高くなっているとも見られる。使用頻度が相対的に高い組み合わせは、「さてそこで」（転換型＋順接型）（書籍 1 回）（国会会議録 11 回）、「さてそれでは」（転換型＋転換型）（書籍 1 回）（国会会議録 6 回）、「さて次に」（転換型＋添加型）（書籍 0 回）（国会会議録 4 回）などである。

4.3.10 「つまり」を先行表現とする二重使用の文章ジャンルによる偏り（表 33 参照）

使用頻度は合計 8 でありごく少ない。「つまりたとえば」（同列型＋同列型）（書籍 1 回）（国会会議録 2 回）などがある程度である。

4.3.11 「ただ」を先行表現とする二重使用の文章ジャンルによる偏り（表 34 参照）

「国会会議録」での比率が高い。使用頻度が高い組み合わせは、「ただしかし」（補足型＋逆接型）（書籍 2 回）（国会会議録 17 回）、「ただ一方」（補足型＋対比型）（書籍 4 回）（国会会議録 7 回）、「ただそうはいつても」（補足型＋逆接型）（書籍 1 回）（国会会議録 5 回）、「ただたとえば」（補足型＋同列型）（書籍 0 回）（国会会議録 5 回）などである。

表 25 先行表現「また」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
また総数	8759	2163	5688	1627	18237
また+市川	69	56	15	11	151
また+市川以外	41	21	4	2	68
また+市川 比率	0.79	2.59	0.26	0.68	0.83
また+市川以外 比率	0.47	0.97	0.07	0.12	0.37
合計 比率	1.26	3.56	0.33	0.80	1.20

表 26 先行表現「そして」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
そして総数	12987	1746	149	590	15472
そして+市川	229	206	3	3	441
そして+市川以外	44	19	1	0	64
そして+市川 比率	1.76	11.80	2.01	0.51	2.85
そして+市川以外 比率	0.34	1.09	0.67	0.00	0.41
合計 比率	2.10	12.89	2.68	0.51	3.26

表 27 先行表現「でも」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
でも総数	4611	63	1	2874	7549
でも+市川	55	0	0	7	60
でも+市川以外	39	1	0	7	46
でも+市川 比率	1.19	0.00	0.00	0.24	0.79
でも+市川以外 比率	0.85	1.59	0.00	0.24	0.61
合計 比率	2.04	1.59	0.00	0.49	1.40

表 28 先行表現「だが」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
だが総数	5097	20	2	9	5128
だが+市川	90	8	0	1	99
だが+市川以外	89	0	0	0	89
だが+市川 比率	1.77	40.00	0.00	11.11	1.93
だが+市川以外 比率	1.75	0.00	0.00	0.00	1.74
合計 比率	3.51	40.00	0.00	11.11	3.67

表 29 先行表現「だから」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
だから総数	5095	1009	1	577	6682
だから+市川	39	19	0	2	60
だから+市川以外	1	2	0	0	3
だから+市川 比率	0.77	1.88	0.00	0.35	0.90
だから+市川以外 比率	0.02	0.20	0.00	0.00	0.04
合計 比率	0.79	2.08	0.00	0.35	0.94

表 30 先行表現「一方」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
一方総数	1009	232	764	46	2051
一方+市川	5	3	0	0	8
一方+市川以外	0	0	0	0	0
一方+市川 比率	0.50	1.29	0.00	0.00	0.39
一方+市川以外 比率	0.00	0.00	0.00	0.00	0.00
合計 比率	0.50	1.29	0.00	0.00	0.39

表 31 先行表現「ところで」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
ところで総数	363	105	8	87	563
ところで+市川	4	3	0	0	7
ところで+市川以外	1	0	0	0	1
ところで+市川 比率	1.10	2.86	0.00	0.00	1.24
ところで+市川以外 比率	0.28	0.00	0.00	0.00	0.18
合計 比率	1.38	2.86	0.00	0.00	1.42

表 32 先行表現「さて」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	YC	YC	合計
さて総数	528	336	94	0	958
さて+市川	8	26	1	0	35
さて+市川以外	3	0	0	0	3
さて+市川 比率	1.52	7.74	1.06	0.00	3.65
さて+市川以外 比率	0.57	0.00	0.00	0.00	0.31
合計 比率	2.08	7.74	1.06	0.00	3.97

表 33 先行表現「つまり」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
つまり総数	4003	428	33	285	4749
つまり+市川	4	2	0	0	6
つまり+市川以外	0	2	0	0	2
つまり+市川 比率	0.10	0.47	0.00	0.00	0.13
つまり+市川以外 比率	0.00	0.47	0.00	0.00	0.04
合計 比率	0.10	0.93	0.00	0.00	0.17

表 34 先行表現「ただ」の出現頻度（総数）に対する二重使用の比率（ジャンル別）

	BK	MD	WR	YC	合計
ただ総数	2791	1563	41	1196	5591
ただ+市川	11	40	0	1	52
ただ+市川以外	17	20	0	2	39
ただ+市川 比率	0.39	2.56	0.00	0.08	0.93
ただ+市川以外 比率	0.61	1.28	0.00	0.17	0.70
合計 比率	1.00	3.84	0.00	0.25	1.63

5. まとめ

調査結果及び分析をまとめると、次のようになる。

- ① 接続表現の二重使用は文章ジャンルによる偏りが見られる。
- ② 二重使用は、全般的に「国会会議録」での使用の比率が高い。
- ③ 「国会会議録」での二重使用の組み合わせの種類（異なり）は多いが、頻用される組み合わせはある程度限られる。頻用される組み合わせは、全体で「添加型」「逆接型」「対比型」の機能を果たす場合が多い。

「そしてまた」（添加型+添加型）、「あるいはまた」（対比型+添加型）、「また一方」（添加型+対比型）、「ただししかし」（補足型+逆接型）、「しかし一方」（逆接型+対比型）、「そしてさらに」（添加型+添加型）、「なおかつ」（補足型+添加型）、「しかし同時に」（逆接型+添加型）などが頻用されている。

- ④ 「白書」では、「しかし一方」が多用されている。
- ⑤ 書籍 NDC 第一次区分別では、「1 哲学」（「そしてさらに」「しかしまた」など）「8 言語」（「そしてまた」など）で使用率が高い傾向がある。ただし、使用頻度はいずれも低い。

今後の課題として、接続表現の単独使用の場合も含めて、接続表現の文章ジャンルによる偏りの要因を、文章ジャンルの違いによる文章構造の違いと関連させながら考察する必要がある。また、接続表現の単独使用と二重使用とでの接続関係の差異の検討に基づいた多層的な文章構造の解明が必要になってくる。さらに、文体（話し言葉的・書き言葉的）や文脈展開などの面からも接続表現の単独使用と二重使用の共通点と相違点を考察する必要がある。

注

- 注1 本研究では、一単語としての接続詞だけではなく、接続詞的な機能を持つ連語や副詞類も含めて広く対象に含めるために、「接続詞」ではなく「接続表現」という用語を用いる。
- 注2 用例は、すべて BCCWJ (『現代日本語書き言葉均衡コーパス』モニター公開データ (2009 年度版)) を利用している。
- 注3 ごくまれに、下記の「しかし、また一方」や「そしてまた、たとえば」などのように、三重使用 (以上) の例もある。

「○山口 (和) 政府委員先ほど申し上げました審議会の意見具申の際に、許可制を採用してはどうかという御意見が、ただいま先生から御指摘ございましたように出ておったことは事実でございます。しかし、また一方、規制強化は絶対に反対であるというような意見が、消費者の方々あるいは大型店の方々から出されておまして、そういった点を総合勘案いたしまして届け出制でいってはどうかというように、ただいまのところ考え」 (衆議院) (BCCWJ)

「(略) ビデオテープを含めまして、放送教材、印刷教材、いずれにいたしましてもこれはもちろん一般に市販をし普及を図って、国民各層の広範な教育需要にこたえるというようなことももちろん対応をする事柄でございます。そしてまた、たとえばこの放送大学がそれぞれ地方の国立大学と相提携をいたしまして、この放送大学の教育内容で、たとえば公開講座というような形で、それぞれ地方の地域の方々の御要望の強いような科目を取り上げまして、(略)」 (参議院) (BCCWJ)

- 注4 接続表現 (接続詞) 及び接続表現 (接続詞) の二重使用に関わる研究の流れについては、馬場(編著) (2010) 及び馬場 (2011) で概要をまとめている。
- 注5 市川 (1978: 89-93) は、8 種の接続類型を示しているが、8 種のうちの「連鎖型」は「接続語句は普通用いられない」型であるため、それを除いた 7 種類を取り上げている。市川が挙げている各類型で使われる「接続語句のおもなもの」の詳細を次に示す。□ 内は下位分類である。

順接型 [順当] = だから・ですから・それで・したがって・そこで・そのため・そういうわけで
それなら・とすると・してみれば・では (以上, 仮定的な意) [きっかけ] = すると・と・そうしたら
ら [結果] = かくて・こうして・その結果 [目的] = それには・そのためには

逆接型 [反対, 単純な逆接] = しかし・けれども・だが・でも・が」といっても・だとしても (以上, 仮定的な意) [背反・くいちがい] = それなのに・しかるに・そのくせ・それにもかかわらず [意外・へだたり] = ところが・それが

添加型 [累加, 単純な添加] = そして・そうして [序列] = ついで・つぎに [追加] = それから・そのうえ・それに・さらに・しかも [並列] = また・と同時に [継起] = そのとき・そこへ・次の瞬間

対比型 [比較] = というより・むしろ (以上, 比較してあとのほうをとる) 」まして・いわんや (以上, 比較されるものをふまえて, 著しい場合に及ぶ) [対立] = 一方・他方・それに対し (以上, 対照的な対立) 」逆に・かえって (以上, 逆の関係での対立) 」そのかわり (交換条件) [選択] = それとも・あるいは・または

転換型 [転移] =ところで・ときに・はなしかわって [推移] =やがて・そのうちに [課題] =さて(主要な話題を持ち出す)「そもそも・いったい(以上, 原本的な事柄を持ち出す) [区分] =それでは・では [放任] =ともあれ・それはそれとして

同列型 [反復] =すなわち・つまり・要するに・換言すれば・言い換えれば(以上, 詳述・要約・換言) [限定] =たとえば・現に(以上, 例示・例証)「とりわけ・わけても(以上, 抽出)」せめて・少なくとも(以上, 最小限度)

補足型 [根拠づけ] =なぜなら・なんとすれば・というのは [制約] =ただし・もつとも・ただ [補充] =なお・ちなみに

注6 順接型の「では」と転換型の「では」を1種類として調査していると思われる。

注7 「しかし, だから害鳥である, ということにはならない。」では「しかし～ということにはならない」の「～」の部分に「だから害鳥である」が入っている。

注8 文頭に使われた副詞の「また」(「再び」の意)を排除しきれていない。「また」の出現頻度(総数)には副詞の「また」も含まれている。

注9 注8に同じ。

注10 転換型は, 「やがて」「それでは」「いったい」の3語の方が, 「ところで」よりも使用頻度が高いが, 「やがて」「いったい」は副詞とも見做されること, また「それでは」は指示詞の用法も含まれている可能性があることから, この3語は除外した。

参 照 文 献

馬場俊臣(2003)「接続詞の二重使用の分析—用例と各接続類型の特徴—」『北海道教育大学紀要 人文科学・社会科学編』53(2): 1-17. (馬場(2006)収録: 110-142).

馬場俊臣(2006)『日本語の文連接表現—指示・接続・反復—』東京: おうふう.

馬場俊臣(編著)(2010)『現代日本語接続詞研究—文献目録・概要及び研究概観—』東京: おうふう.

馬場俊臣(2011)「接続詞の二重使用に関わる研究について」『語学文学』49: 1-10.

馬場俊臣(印刷中)「接続表現の二重使用と文章ジャンルについて」『北海道教育大学紀要 人文・社会科学編』63(2). (2013年2月発行予定).

石黒圭(2005)「接続詞の二重使用とその表現効果」中村明・野村雅昭・佐久間まゆみ・小宮千鶴子(編)『表現と文体』: 160-169. 東京: 明治書院.

石川慎一郎(2010)「第9章 因子分析: データに隠れた要因を探る」石川慎一郎・前田忠彦・山崎誠(編)『言語研究のための統計入門』: 219-244. 東京: くろしお出版.

市川孝(1957)「文章の構造」岩淵悦太郎(編者代表)『講座 現代国語学Ⅱ ことばの体系』: 279-306. 東京: 筑摩書房.

市川孝(1963)「文章論」『国語シリーズ No.57 文章表現の問題』: 7-46. 東京: 教育図書.

市川孝(1978)『国語教育のための文章論概説』東京: 教育出版.

森田良行(1958)「文章論と文章法」『国語学』32: 91-105.

- 村田年 (2000) 「多変量解析による文章の所属ジャンルの判別—論理展開を支える接続語句・助詞相当句を指標として—」『統計数理』48(2): 311-326.
- 村田年 (2002) 「論理展開を支える機能語句—接続助詞, 助詞相当句による文章のジャンル判別を通じて—」『計量国語学』23(4): 185-206.
- 村田年 (2007) 「専門日本語教育における論述文指導のための接続語句・助詞相当句の研究」『統計数理』55(2): 269-284.
- 清水まさ子 (2006) 「談話のジャンルと接続表現との関係—新聞の報道文とコラムを比較して—」『東アジア日本語教育・日本文化研究』9: 55-75.
- 多田知子 (2010) 「複合接続詞——文の文頭部分の階層性——」『国文論叢』42: 52-39, (要旨) 68.
- 高橋淑郎 (2005) 「大学講義を対象とした類型的文体分析の試み」中村明・野村雅昭・佐久間まゆみ・小宮千鶴子(編)『表現と文体』: 35-46. 東京: 明治書院.
- 高澤信子 (2002) 「接続詞と接続表現との関係について—文章理解における有効性—」『立教大学日本語研究』9: 12-37.

付 記

本稿は, 国立国語研究所共同研究「テキストにおける語彙の分布と文章構造」研究発表会 (2011 年 3 月 6 日, 国立国語研究所) における研究発表「接続表現の二重使用と文章ジャンル」の内容を加筆修正したものである。
2 章及び 4 章の内容の一部は, 馬場俊臣 (印刷中) に発表予定である。

状態空間表現を用いた文章の特徴付け

馬場康維 (統計数理研究所)

小森 理 (統計数理研究所)

Feature Extraction of Sentence Structure based on State Space Representation Model

Yasumasa Baba (The Institute of Statistical Mathematics)

Osamu Komori (The Institute of Statistical Mathematics)

1. はじめに

様々な目的で文章をデータとした解析が行われている。文章の類型化、著者推定、ウェブ上の文章からのトピックの抽出など分野も目的も用いられる手法も様々である。たとえば、著者推定の問題では、品詞の出現比率や特定の単語の出現比率がしばしば用いられる。この方法は品詞や特定の単語の静的な分布を用いて文章の特徴を抽出することによる解析である。ところで、文章は文の連なりから成っており、文章の要素である文は一連の語や記号の系列で成り立っている。即ち、文や文章は形態素の系列で成り立っている。したがって文章の特徴を用いて何らかの解析をするには、単に形態素の静的な分布のみならず形態素の出現順序を考慮した動的な解析が有効であろう。そこで、一連の語や記号の連なりである文章を“品詞”という状態を推移する系列とみなし、文章の構造をこの系列の構造としてとらえることで文章構造の解析ができるのではないかと考えたのがこの研究の発端である。

“状態”の定義は分析の対象・目的によって変わる。形態素解析を利用してテキストデータを品詞の系列で表現する場合には、名詞、動詞、助詞などの品詞が状態に対応する。文の構造を抽出する際には名詞句、動詞句といった品詞の結合した状態を用いた方が形態素のままの状態で文を表現するよりは文の構造が把握しやすく、構造の分析には適している。一方、より詳細な構造を分析の対象にするならば、名詞を名詞の種別に分割した状態を考えるというように状態の分割も必要である。さらに文章全体を構造化してとらえるには段落の状態を考慮する必要がある。このように“状態”は分析の場面、場面に応じてフレキシブルに定義されるものである。

この報告では、文章構造のモデル化の基礎的な研究として文を状態空間で表現し時系列としてとらえる試みについて述べる。ここで用いたデータは国立国語研究所共同研究プロジェクト「文章における語彙の分布と文章構造」により作成されたテキストデータの一部である。文章あるいは文の解析にはまず文法的なモデルを用意し単語の意味を考慮するというような方法があるが、ここでは、データから得られる情報をもとに文の構造的な把握をするというプロセスによって文章構造のモデル化を図る。具体例として上記の名詞句、動詞句等の推移確率を計算し、それにもとづいた主成分分析の結果を紹介する。これは名詞、動詞、助詞等の頻度による解析、つまり静的な解析とは異なり、文章のつながりを考慮している点で、既存の解析法とは異なっている。また文章の特徴抽出をより細密化する試みとして、名詞句、動詞句等をさらにいくつかの「パター

ン」にまとめることができる例も紹介する。このパターン抽出は文章を一つ一つ読んで探していく作業であるためかなりの労力を費やすが、基本的なパターンはそれほど多くはないと予想されるため、一旦辞書のようなデータベースを構築できれば、さまざまな分野の作品またその著者たちの特徴をより鮮明に捉えることができると思われる。

ここで述べるモデル化の方法はまだ完成されたものではないが、機械学習を利用したデータの自動的な収集、文章の類型化による文章の分類等様々な応用が考えられる。

2. 品詞による表現—形態素解析の利用

最も基礎的で素朴な品詞状態による表現の例を示す。ここで例示に用いたデータは、近藤和敬、“ヒルベルトの数学における公理的方法からカヴァイエスの概念の哲学へ”（以下、近藤論文と呼ぶ）をテキスト化し形態素解析を行って得られた品詞データである。テキストデータには、段落のタグがついており、“論文のタイトル+著者名+所属”は一つの段落として扱われている。表1は形態素解析の結果を示している。最も基礎的なこのタイプのデータを時系列的に表現しただけでも文の特徴が見いだせる。形態素解析の品詞のカテゴリーが異なった状態になるように（句読点0，名詞句10，動詞句20，形容詞句30，副詞40，連体詞50，接続詞60，その他-10）数値を対応させた。この数値は便宜上割り振ったもので何らかの最適化をしたものではない。この数値を割り振られた品詞の状態空間を用いて文章の一部を表現してみると図1、図2のようになる。図1は、近藤論文の“タイトル+著者名+所属部分”である。図2は最初の文の時系列である。図1には句読点がないこと、動詞+句点で終わっていないこと等、タイトルであることが類推できる特徴が存在する。一方、図2では文末が動詞+句点という典型的な連結で終わっている。このことから、状態空間表示により時系列を表現することで、文の特徴抽出が可能になることが推察される。

表1 文の形態素解析

文字	品詞	品詞	文節
		(詳細)	ID
数学	名詞	一般	1
基礎	名詞	一般	1
論	名詞	接尾	1
の	助詞	連体化	1
論争	名詞	サ変接続	2
の	助詞	連体化	2
結果	名詞	副詞可能	3
,	記号	読点	0
数学	名詞	一般	1
的	名詞	接尾	1
認識	名詞	サ変接続	1
の	助詞	連体化	1
確実	名詞	形容動詞語幹	2
性	名詞	接尾	2
の	助詞	連体化	2
アブリオリ	名詞	一般	3
な	助動詞	*	3
基礎	名詞	一般	4
付け	名詞	接尾	4
が	助詞	格助詞	4
不可能	名詞	形容動詞語幹	5
で	助動詞	*	5
ある	助動詞	*	5
こと	名詞	非自立	6
から	助詞	格助詞	6
,	記号	読点	0
合理	名詞	一般	1
論	名詞	接尾	1
的	名詞	接尾	1
な	助動詞	*	1
認識	名詞	サ変接続	2
論	名詞	接尾	2
は	助詞	係助詞	2
その	連体詞	*	3
説得	名詞	サ変接続	4
力	名詞	接尾	4
を	助詞	格助詞	4
半減	名詞	サ変接続	5
さ	動詞	自立	5
せ	動詞	接尾	5
た	助動詞	*	5
よう	名詞	非自立	5
に	助詞	副詞化	5
思わ	動詞	自立	6
れる	動詞	接尾	6
。	記号	句点	-1

図1 タイトルの品詞による時系列表現

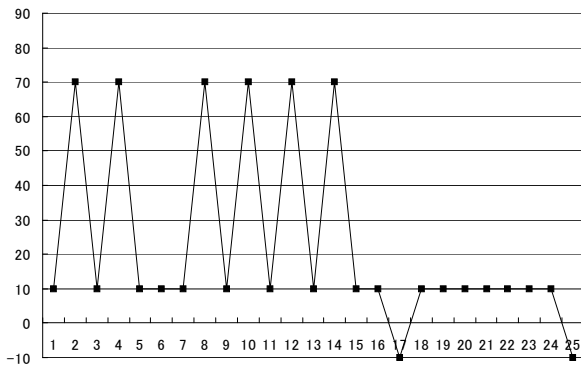
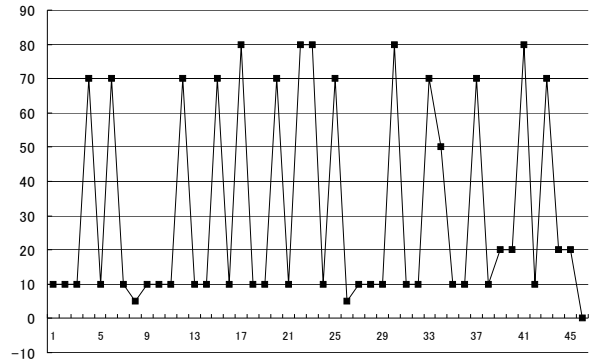


図2 文の品詞による時系列表現



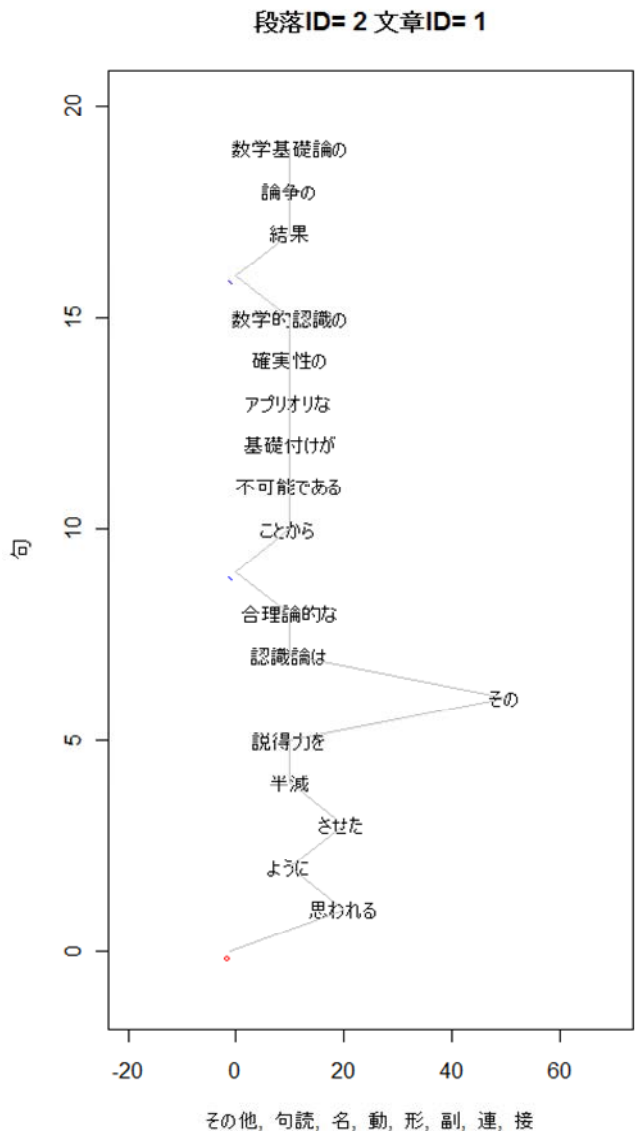
3. 状態の縮約

一連の文章に形態素解析を行うと、例えば、“数学基礎論”は“数学”、“基礎”、“論”というように3つの名詞に分解され、名詞状態が連続して出現する。上述のグラフ表現では“名詞”状態が3時点続くことになり、時系列を観察すると、状態“名詞”のフラットな直線が現れる。品詞の状態による表現空間での時系列を観測すると、数学+基礎+論のように名詞のみが連続する場合はそれらを一つの単語（便宜上名詞と呼ぶ）として扱うことが可能な場合がほとんどであることが分かる。これを踏まえ、第1ステップとして、連続した“名詞”状態を一つの“名詞”状態として縮約を行った。

さらに、例えば、“数学基礎論+の”が次の“論争”を修飾している。名詞+助詞で一つのかたまりと考える方が文の構造の表現には便利である。即ち、品詞による状態表現を特定の結びつきを示す品詞と品詞の状態に縮約した方が構造解析には都合がよい。そこで、品詞で表現した状態を縮約し助詞を中心にまとめた状態で文を表現したものが図3である。

このように助詞を中心にして状態をまとめ

図3 助詞の違う両表表現 (右) レアな文の時系列表



てみると助詞の種類によりそれぞれの役割があることが分かる。そこで、名詞+助詞を一つの状態とみなし、近藤論文のデータから推移確率を作ったものが表4である。表中、“名詞の”や“名詞を”はそれぞれ、名詞+助詞(の)や名詞+助詞(を)を表している。つまり名詞と助詞の連結ごとに状態を割り振ったことになる。これが第2段階の状態空間の縮約である。表5は表4の項目に関する出現頻度をまとめたものである。

表4 助詞による状態表現の推移確率(%) (近藤論文)

	句点	名詞を	名詞に	名詞が	名詞な	名詞の	名詞は	動詞	読点	その他
句点	0	3.6	3.6	8.0	4.4	11.1	12.9	0	0	55.7
名詞を	0	0	10.1	0.5	3.5	3.5	0.5	41.4	7.1	33.1
名詞に	0	1.6	3.2	1.6	6.4	4.3	0	41.7	7.5	33.2
名詞が	0	4.0	9.0	0	4.5	5.6	0	26.0	6.8	44.8
名詞な	0	20.7	3.7	9.8	2.4	18.3	7.3	0	1.8	35.5
名詞の	0	13.6	7.1	11.3	7.3	18.6	6.8	0.5	0	35.5
名詞は	0	7.0	4.2	0.7	5.6	9.9	0.7	3.5	47.2	21.0
動詞	26.2	5.5	9.4	7.5	2.8	9.7	7.2	0	7.2	25.1
読点	0	7.2	6.3	6.6	9.6	26.2	3.9	1.5	0	38.4

表5 句等の出現頻度 (近藤論文)

	度数	パーセント
句点	226	6.9
名詞を	198	6.1
名詞に	187	5.7
名詞が	177	5.4
名詞な	164	5.0
名詞の	382	11.7
名詞は	142	4.3
動詞	362	11.1
読点	332	10.2
その他	1096	31.4

さらに“名詞の”+“名詞”は合わせて“名詞”とみなしても良い状態である。名詞についてはこれが第2段階の縮約になる。この外に、すぐに目につく状態の縮約には、“名詞”+“する”がある。この状態は、その機能上から、“名詞”と“する”を合わせた状態を“動詞”状態とみなすことができる。これらのプロセスを簡単に示すと下記のようなになる。

名詞+名詞 ⇒ 名詞
 名詞+助詞(の) ⇒ 名詞(の)
 名詞(の)+名詞 ⇒ 名詞
 名詞+する ⇒ 動詞

このように、次々に状態をまとめていくことにより、階層構造を持った状態空間が構成できるこ

とになる。

4. 論文の特徴の比較

比較のために他の論文データを用いて集計を行った。比較に用いたデータは、横地徳広“認識論的転回の地平を求めて－ハイデガーとカント『純粹理性批判』”(以下、横地論文と呼ぶ)、である。推移確率を表6に、各句等の状態の出現頻度を表7に示してある。この推移確率の状態(項目)も表4と同じものを用いている。

表6 助詞による状態表現の推移確率(%) (横地論文)

	句点	名詞を	名詞に	名詞が	名詞な	名詞の	名詞は	動詞	読点	その他
句点	0	6.1	3.0	0.8	2.3	5.3	23.5	0	0	59.5
名詞を	0	0	13.3	0.6	0	1.9	3.2	32.9	0.6	47.1
名詞に	0	3.4	1.7	3.4	0.8	1.7	0.8	35.6	7.6	44.3
名詞が	0	6.4	7.7	0	0	9.0	0	23.1	5.1	48.9
名詞な	0	6.1	3.0	9.1	3.0	21.2	12.1	0	0	45.3
名詞の	0	22.0	11.0	10.4	4.4	4.4	6.6	11.0	0	39.0
名詞は	0	8.6	5.2	3.4	1.7	15.5	0	1.7	31.9	32.4
動詞	28.8	8.0	8.4	1.8	2.2	6.6	8.4	0	10.2	25.0
読点	0	8.2	6.5	8.2	2.2	22.9	9.5	0.4	0	41.7

表7 句等の出現頻度 (横地論文)

	度数	パーセント
句点	133	6.6
名詞を	158	7.8
名詞に	118	5.8
名詞が	78	3.9
名詞な	33	1.6
名詞の	182	9.0
名詞は	116	5.7
動詞	226	11.2
読点	231	11.4
その他	749	33.8

表4、表5、表6、表7から、2つの論文の表現の比較が可能になる。いずれの場合でも、動詞の次に続くのは句点である確率が高く、句点の次には文章のはじまりである名詞が続く確率が高いというような傾向があることが分かる。これらの中で主語の役割を担うものは主に“名詞は”と“名詞が”であるが、句点からの推移確率が2つの論文で大きく異なることも分かる。文章のスタイルの違いが推移確率に反映されていると言える。ここでは、2つの論文のみを比較したが、多くの論文について推移確率の比較あるいは頻度の比較を行うことにより文献間の距離が算出で

きる。したがってそこから文献の類型化ができるであろう。

5. 主成分分析の適用例

前節で二つの論文の比較をした。もう少し論文を増やして論文間の比較を試みる。ここでは上記の二つに以下の3論文を追加して分析を試みた。手塚博「ミシェル・フーコーの権力分析における真理の概念- 権力の行使としての反省」(手塚論文), 小島優子「ヘーゲルにおける「罪責」と「犯罪」—『精神現象学』を中心に」(小島論文), 山田圭一「最晩年ウィトゲンシュタインの連続性テーゼが意味するもの」(山田論文)。

まずそれぞれの論文で前節の表5のように句等の出現頻度を計算し、少なくとも一つの論文中に5パーセント以上出現する句等を選び出した。横地論文では「名詞が」が3.9パーセントであるが、近藤論文では5.4パーセントであるため、推移確率の項目の一つに選ばれている。このようにして抽出した状態をベースにして、状態の頻度の分布と推移確率を求めたものが表8、表9、表10、表11、表12、表13である。それぞれの特徴が読み取れる。

表8 助詞による状態表現の推移確率(%) (手塚論文)

	句点	名詞を	名詞に	名詞が	名詞な	名詞の	名詞は	動詞	読点	その他
句点	0	2.9	3.4	5.8	2.4	12.5	13.5	0	0	59.9
名詞を	0	0	12.0	0.5	0	0.5	0.5	39.3	2.6	44
名詞に	0	4.0	1.1	0	2.9	1.1	0	44.8	10.3	35.7
名詞が	0	4.6	13.2	0	2.6	10.5	0	15.8	11.8	42
名詞な	0	15.3	0.9	10.8	2.7	10.8	3.6	0	1.8	54
名詞の	0	16.5	8.3	8.9	6.7	10.2	7.9	0.6	0.3	40.5
名詞は	0	3.1	5.5	2.3	4.7	8.6	0	2.3	42.2	31.7
動詞	21.7	4.5	6.1	8.0	2.6	6.7	3.5	0	7.3	39.2
読点	0	7.2	4.1	7.8	3.8	23.5	6.6	0.3	0	46.1

表9 句等の出現頻度 (手塚論文)

	度数	パーセント
句点	209	6.6
名詞を	191	6.1
名詞に	174	5.5
名詞が	152	4.8
名詞な	111	3.5
名詞の	315	10.0
名詞は	128	4.1
動詞	313	10.0
読点	319	10.1
その他	1232	36.8

表10 助詞による状態表現の推移確率(%) (小島論文)

	句点	名詞を	名詞に	名詞が	名詞な	名詞の	名詞は	動詞	読点	その他
句点	0	2.6	15.1	5.7	1.0	8.3	12.5	0	0	54.6
名詞を	0	0	5.6	2.2	0	5.6	2.2	43.0	6.1	35.7
名詞に	0	9.2	1.5	0	1.0	2.4	1.5	40.8	18.0	25.8
名詞が	0	6.1	11.4	0	0.8	12.1	0	21.2	6.1	43.1
名詞な	0	18.6	9.3	7.0	0	7.0	7.0	0	0	51.2
名詞の	0	11.8	11.8	6.7	2.0	10.2	7.1	1.6	0	49.2
名詞は	0	6.0	6.0	0.6	0	6.0	0	9.6	46.4	25.2
動詞	27.2	5.7	6.5	3.4	0.8	5.1	9.6	0	4.2	37.5
読点	0	7.3	8.1	10.2	3.1	20.9	9.2	2.9	0	38.5

表11 句等の出現頻度 (小島論文)

	度数	パーセント
句点	193	6.5
名詞を	179	6.0
名詞に	206	6.9
名詞が	132	4.4
名詞な	43	1.4
名詞の	254	8.6
名詞は	166	5.6
動詞	353	11.9
読点	382	12.9
その他	1060	34.4

表12 助詞による状態表現の推移確率(%) (山田論文)

	句点	名詞を	名詞に	名詞が	名詞な	名詞の	名詞は	動詞	読点	その他
句点	0	0.5	7.9	2.6	2.1	11.5	6.8	0	0	68.2
名詞を	0	0	9.3	0.5	0.5	8.2	1.6	27.5	1.1	50.3
名詞に	0	1.9	1.3	1.3	1.3	4.4	2.5	32.1	6.9	48.2
名詞が	0	2.0	6.5	0	5.2	13.7	0	14.4	3.9	54.9
名詞な	0	13.6	5.8	10.7	7.8	20.4	1.9	0	0	40.1
名詞の	0	16.9	7.1	10.3	3.4	14.9	8.1	1.0	0	37.7
名詞は	0	3.9	0.6	2.6	0.6	12.9	0	4.5	27.7	46.2
動詞	26.6	2.7	4.4	5.1	4.8	9.9	6.1	0	9.6	30.0
読点	0	6.2	4.7	8.0	4.7	27.9	9.8	0	0	39.5

表 1 3 句等の出現頻度 (山田論文)

	度数	パーセント
句点	192	5.9
名詞を	182	5.6
名詞に	159	4.9
名詞が	153	4.7
名詞な	103	3.2
名詞の	409	12.6
名詞は	155	4.8
動詞	293	9.0
読点	276	8.5
その他	1335	39.9

この5つの論文の状態空間での推移確率をデータとして主成分分析を行った。まず 9×10 の推移確率の行列を行ごとにベクトル化し要素数90のベクトルとした。そのベクトルに対し主成分分析をし、第1主成分と第2主成分を表示すると図4の散布図が得られる。(山田論文, 横地論文), (手塚論文, 小島論文) と (近藤論文) に大きく3つに分かれることが分かる。第1主成分の寄与率は0.51, 第2主成分の寄与率は0.21となり第3主成分までの累積寄与率は0.88となる。第1主成分、第2主成分、第3主成分の因子負荷量を計算し、図示したものが図5, 図6, 図7である。なお因子負荷量の図では、

句点⇒句, 読点⇒読, 名詞が⇒名が, 名詞の⇒名の

といった省略表記をしている。たとえば、状態間の推移を表す「句点名詞の」は「句名の」というように省略した表現が用いられている。

図4 推移確率から求めた5つの論文の特徴づけ

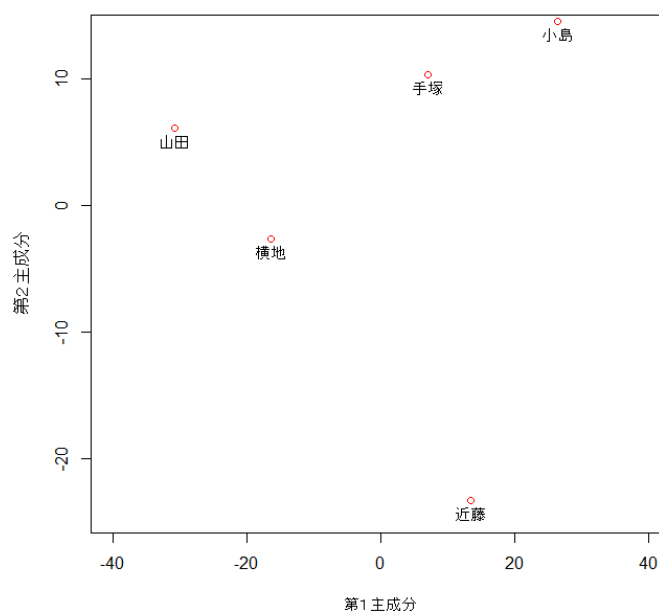


図5 第1主成分に対する因子負荷量

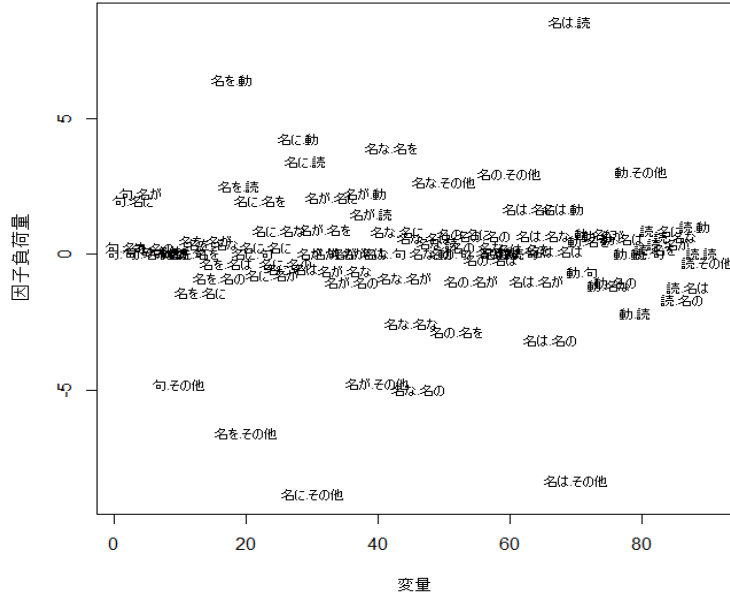


図6 第2主成分に対する因子負荷量

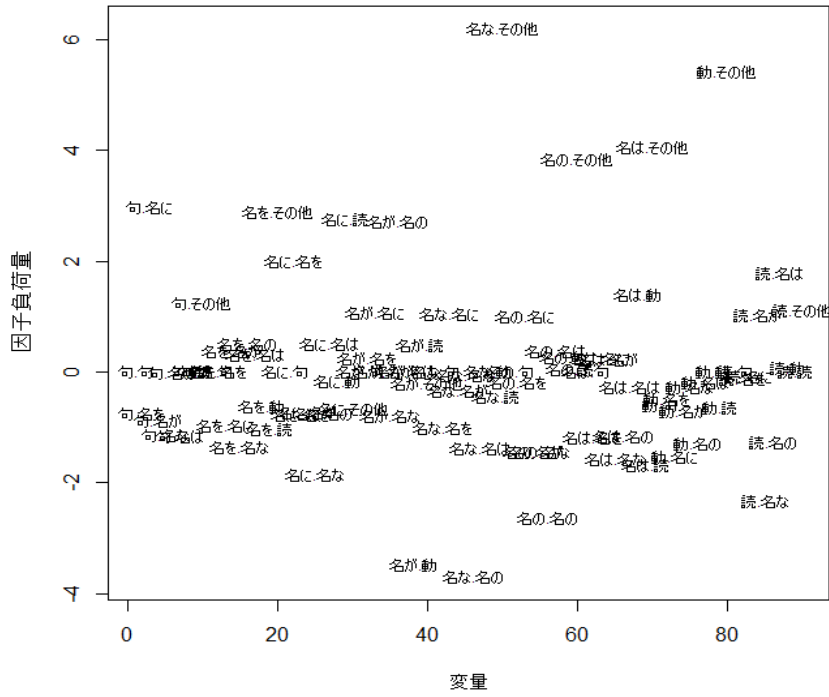
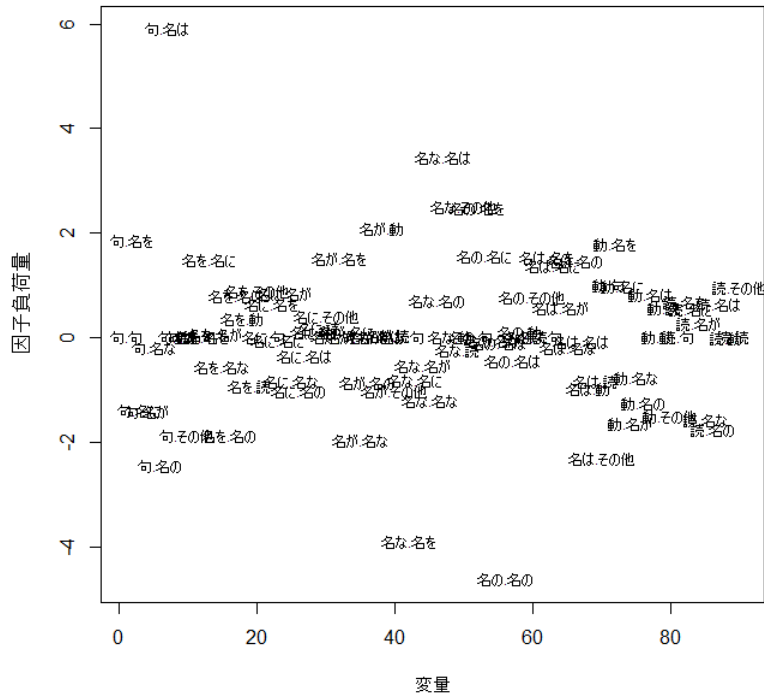


図7 第3主成分に対する因子負荷量

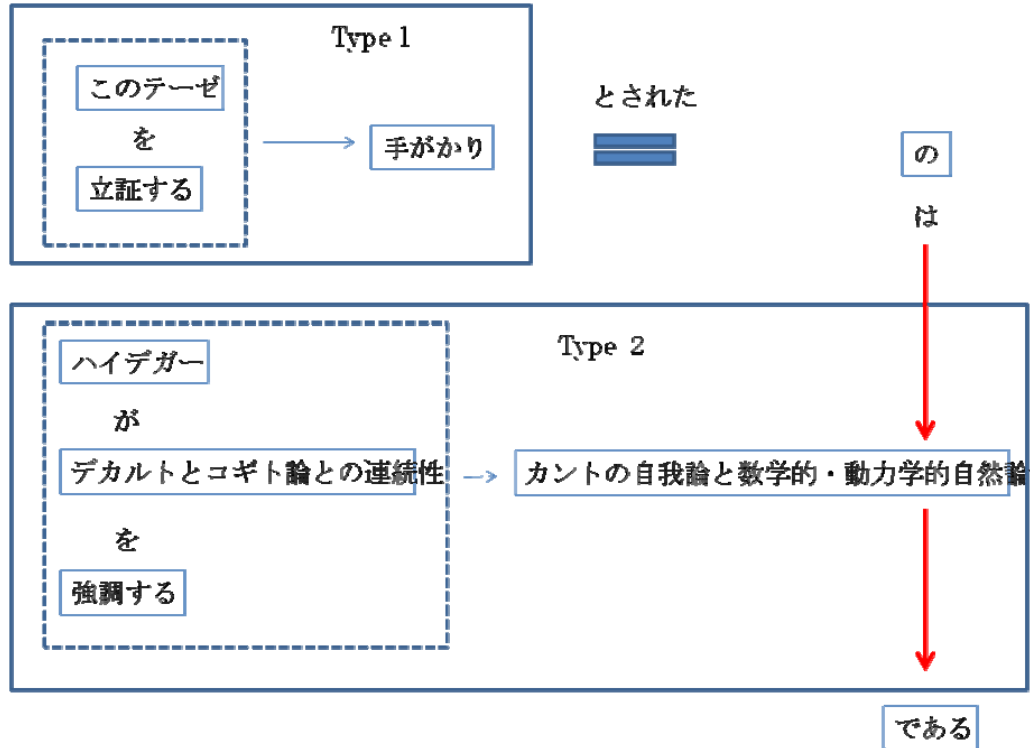


第1主成分に大きく寄与する推移確率の因子として「名詞は→読点」, 「名詞を→動詞」, 「名詞に→その他の品詞」, 「名詞は→その他の品詞」のようなものがあり, これは文の基本的な構造を規定する因子と考えられる。第2主成分に大きく寄与する因子としては「名詞な→その他」, 「動詞→その他」, 「名詞が→動詞」, 「名詞な→名詞の」などの修飾関係に関する因子が見られる。近藤論文はその他の4つの論文と比較すると, 第二主成分に大きな違いがあるため, 修飾関係にその違いがあると推察される。このように文章を状態空間とみなし, それを推移確率で表現することにより, 単なる品詞の頻度の集計からでは読み取れない動的な文章の特徴を捉える事ができる。

6. より詳細な文章の特徴づけ

上記までの議論では品詞同士の結合（「名詞+名詞」）と品詞と助詞との結合（「名詞+は」）に注目し, 文構造の探索を行った。しかしながら文の特徴をより詳しく捉えようとするならば, さらに上位の文構造まで考慮する必要がある。一例として横地論文の次の一文を考察する。「このテーゼを立証する手がかりとされたのは, ハイデガーがデカルトのコギト論との連続性を強調するカントの自我論と数学的・動力的自然論である」。一見複雑な構造をもつ文章であるが, 文の構造に注目すると図8のようになる。

図8 より詳細な文構造



この文章の骨格は矢印で示された「AはBである」という至って単純な構造である。この文章を複雑に見せているものは名詞の修飾関係である。これも大きくわけて Type1 と Type2 があることが分かる。Type1 では「手がかり」という名詞を「目的語（このテーゼ）+を+動詞（立証する）」という語が修飾している。Type2 も同様な構造であるが「ハイデガー」という主語が新たに加わった修飾語になっている。このような文の「パターン」に注目し文章または作品全体の特徴づけを行うことにより、より詳細な文構造の分類が可能になる。

7. おわりに

品詞や句による状態空間表現について状態の縮約のプロセスを示した。文は語がつながった一連の系列から成っておりいわば時系列的な表現が必要である。この観点からすると状態空間表現は系列事象としての文あるいは文章を表現するのに適した表現であると考えられる。

これまでの文章の構文解析は、単語の出現割合に注目したものが多かった。これはいわゆる静的な文章解析である。静的な文章解析では、語の連なりは無視される。語や品詞の出現頻度のみが用いられており、文章のなかの語の出現の順序という情報、即ち時系列的な情報は用いられていない。時系列的な表現という立場からすると静的な解析は時間軸に沿った情報をまとめて状態

空間を占める割合だけに注目しているといえる。時間と状態という二つの情報のうち時間についてまとめて状態空間上だけの分布をみているというのが静的な解析である。

一方ここで提案した状態空間による表現は語の連なりといういわば文章の動的な要素を考慮したものとなっており、図3または推移確率に示したような文章の動的な表現を分析の対象としたものである。この状態空間表現による分析は、状態をどう定義するかによって様々な対象の分析に応用できる。名文と呼ばれる文章には読者に訴えるリズムがあり、これが文章の内容理解を深める。名文と駄文の違いは状態空間でどう表現されるのか。理路整然とした文章と理解しがたい文章は状態空間表現でどう違うのかなど、様々な解析に応用できる。この動的要素が色濃く出るものが詩や音楽の世界の歌詞である。このような対象にも状態空間表現が応用できる可能性がある。

今回の試みはまだ試行錯誤の段階であり、状態空間の構成も定まったものではない。今後大量のテキストデータの分析を積み上げることによって、文章の背後にある時系列的な状態の推移の様々なパターンの把握と類型化を試みたい。また大量データの処理にあたり、機械学習に適した構造モデルを構築することも考えている。

参考文献

- 小島優子 (2007) 「ヘーゲルにおける「罪責」と「犯罪」—『精神現象学』を中心に」哲学, Vol.58, pp.177-190.
- 近藤和敬 (2009) 「ヒルベルトの数学における公理的方法からカヴァイエスの概念の哲学へ」哲学, Vol.60, pp.169-184.
- 田中章夫 (1974) 「句のエントロピーに基づく構文合成」言葉の研究第5集, pp.125-146.
- 手塚博 (2009) 「ミシェル・フーコーの権力分析における真理の概念— 権力の行使としての反省」哲学, Vol.60, pp.217-232.
- 中野洋 (1974) 「自動項分解の構想」言葉の研究第5集, pp.147-157.
- 町田健 (2011) 「言語構造基礎論：文の意味と構造」勁草書房.
- 山田圭一 (2008) 「最晩年ウィトゲンシュタインの連続性テーゼが意味するもの」哲学, Vol.59, pp.309-325.
- 横地徳広 (2005) 「認識論的転回の地平を求めて—ハイデガーとカント『純粹理性批判』」哲学, Vol.56, pp.270-282.

データの出处

国立国語研究所共同研究プロジェクト「文章における語彙の分布と文章構造」(チームリーダー 山崎誠)

研究成果のまとめ

村田 年（慶應義塾大学日本語・日本文化教育センター）

山崎 誠（国立国語研究所）

要旨

現代書き言葉均衡コーパス（BCCWJ）の文章資料を対象に、慣用句、複合動詞を指標として、テキスト中に現れる句や語の分布がジャンルによって異なることを明らかにした。具体的な研究成果は以下の二つである。一つは、慣用句を指標とした分析である。BCCWJの書籍コーパス中、ジャンルが明示的なテキスト9種を「人文科学系」「社会科学系」「自然科学系」の三つに分類し、「手」のつく慣用句74を指標として正準判別分析を行った。その結果、判別に有効な5つの慣用句が選択され、三つのジャンルが明確に分離された。もう一つは、BCCWJ書籍コーパスの中の、自然科学系ジャンルのテキストにおける複合動詞の後項動詞の使用傾向の分析である。選択した26後項動詞を指標として調査した結果、「だす」「こむ」「つける」「あげる」「あう」によって構成される複合動詞の使用が全体の50%以上を占めていることがわかった。この調査結果を踏まえ、上記5語を含む複合動詞のテキストにおける具体的な使用例を、意味特徴から分類し、さらに結合する名詞をまとめて提示することによって専門日本語教育の現場に資するデータを提供した。

キーワード：

テキスト、ジャンル、現代日本語書き言葉均衡コーパス、BCCWJ、慣用句、複合動詞、後項動詞、判別分析、専門日本語教育

本プロジェクトの研究成果を以下の通り、発表した。

1. 村田年・山崎誠（2011）「「手」の慣用句を指標とした文章のジャンル判別—現代日本語書き言葉均衡コーパスを用いて—」『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター第39号
2. 村田年・山崎誠（2012）「自然科学系書籍における複合動詞の使用傾向—後項動詞を指標として—」『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター第40号

以下に転載する。

「手」の慣用句を指標とした文章ジャンルの判別

—現代日本語書き言葉均衡コーパスを用いて—

村田 年 (慶應義塾大学日本語・日本文化教育センター)

山崎 誠 (国立国語研究所)

1. はじめに

専門分野における学習・研究を目標とする日本語学習者が中・上級レベルに到達し、その後どのように学習を進めるかということについて考える時、文章のジャンル¹⁾は一つの重要な視点であると言えよう。日本語学習者が文章ジャンルの違いをより一層意識化し、各ジャンルにおいて特徴的な表現を学んでいくことは中・上級レベル以降の学習の効率化につながると考えられる。

村田 (2007) では論述的な文章に特徴的な接続語句と助詞相当句を選択し、ジャンルによって異なる文章の特徴がそれらの語句の使用傾向の違いに反映されることを実証した。また村田 (2008) では文章ジャンルを判別するための指標として複合動詞の後項動詞を取り上げ、その使用傾向が文章ジャンルによって異なることを明らかにした。本稿では、文章のジャンルを判別するための新たな指標の可能性として慣用句を取り上げたい。従来は個人レベルで収集した文章資料の範囲で分析を行ってきたが、今回は、現代日本語書き言葉均衡コーパスの使用が可能となったので、それを対象資料とする。分析のための指標としてすべての慣用句を取り上げることは数の多さから難しいので、まず「手」を含む動詞慣用句と形容詞慣用句に絞って調査を行った。

2. 分析に用いた文章資料

2.1 『現代日本語書き言葉均衡コーパス』

調査に用いた資料は、『現代日本語書き言葉均衡コーパス』モニター公開データ (2009 年度版)』である。このデータは、国立国語研究所を中心に構築している大規模コーパス『現代日本語書き言葉均衡コーパス』の一部で、2009 年 7 月現在、著作権者から許諾が得られたデータ約 4,300 万語を収録している。データの内訳は表 1 のとおりである。この調査では、そのうち書籍に該当するデータ (表 1 の (1)~(3)) を使用した。書籍のデータは暫定的なものであり、最終的にはデータ量は約 2 倍以上になる。『現代日本語書き言葉均衡コーパス』については、前川 (2008) 同 (2009) を、この調査で使用した書籍の部分については山崎 (2009) を参照されたい。

表1 モニター公開データ（2009年度版）の内訳

データ名	語数
(1) 出版サブコーパス・書籍	1,300 万語
(2) 図書館サブコーパス・書籍	1,500 万語
(3) 特定目的サブコーパス・ベストセラー	230 万語
(4) 特定目的サブコーパス・白書	480 万語
(5) 特定目的サブコーパス・Yahoo!知恵袋	520 万語
(6) 特定目的サブコーパス・国会会議録	490 万語

語数については注の2) 参照。

2.2 文章資料データの母集団

この調査で使用した書籍のデータは、3種類のデータから成っている。同じ書籍であるが、それぞれ母集団が異なる。1つ目は、「出版サブコーパス・書籍」で、2001年～2005年に日本国内で出版された全書籍からコーパス収録条件³⁾で絞り込んだ約32万冊が母集団である。2つ目は、「図書館サブコーパス・書籍」で、東京都内の自治体ごとにISBNにより管理されている蔵書データを利用して、13自治体以上に共通して所蔵されている書籍からコーパス収録条件で絞り込んだ約38万冊が母集団である。ここでは、ISBNの付与が普及した1986年～2005年刊行の書籍を対象とした。3つ目は、1976年～2005年版の『出版年鑑』及び『出版指標年報』に掲載されたベストセラーリスト上位20位に挙げられた書籍約950冊を対象としている。

2.3 文章資料サンプルの抽出方法

書籍の1つのサンプルは1万字を超えない範囲のひとまとまりの文章（1章、1節など外的な構成を基準として判断したまとまり）である。母集団の全ページをサンプリングフレームと考え、そこからランダムに抽出したページの中の1点をランダムに指定し、その点を基準にテキストを抽出している。

2.4 文章資料のジャンル

書籍については以下の11に分類され、語数は以下の通りである。

表2 書籍内の分類

分類	語数
0 総記	642,351
1 哲学	1,799,241
2 歴史	2,674,253
3 社会科学	6,854,656

4 自然科学	1,631,154
5 技術・工学	1,363,963
6 産業	878,954
7 芸術・美術	1,311,554
8 言語	492,255
9 文学	11,267,012
分類なし	478,536

本稿で分析対象とする文章資料は、書籍の中ではジャンルが明示的な1哲学から9文学までとし、0総記と分類なしは除くことにする。また、1から9までを大きく3つのジャンルに分け、それぞれ「人文科学系」「社会科学系」「自然科学系」として以下のようにまとめた。

人文科学系：哲学，歴史，芸術・美術，言語，文学

社会科学系：社会科学，産業

自然科学系：自然科学，技術・工学

「歴史」を社会科学系に入れる考え方もあるが、専門日本語教育の観点から、大学での学科編成も考慮に入れて本稿では人文科学系に分類した。また、「芸術・美術」についても書籍の内容に芸術理論，美学，美術史，音楽学などが含まれているため、人文科学系に分類した。

3. 分析

3.1 指標としての「手」の慣用句の抽出方法

この調査で対象とした慣用句は、『基本慣用句五種対照表』（佐藤理史編 2007：以下、「対照表」と略す）に基づいている。この対照表は5種類の資料に現れる慣用句を一覧したものである。5種類の資料は以下のとおりである。

- ①金田一春彦，金田一秀穂監修『新レインボー小学国語辞典改訂第3版』学研，2005
- ②金田一京助編『小学館学習国語新辞典全訂第二版』小学館，2006
- ③宮地裕編『慣用句の意味と用法』明治書院，1982
- ④米川明彦，大谷伊都子編『日本語慣用句辞典』東京堂出版，2005
- ⑤金田一秀穂監修『小学生のまんが慣用句辞典』学研，2005

対照表に現れる慣用句は3,628句で、この中から「手が～」「手を～」「手に～」を中心に、「手」が使われている動詞慣用句と形容詞慣用句を抽出し、その使用頻度を調査した。検索項目は合計91項目となった。検索結果については、活用形の違い、漢字の異同等を同定する作業を行った。例えば、「手がつけられない」「手のつけようがない」は「手をつける」の項目にまとめた。「手を

合わせる」は「手を合わす」と同じ項目として扱った。また、「手の裏を返す」が2例あったが「手のひらを返す」の項目にまとめた。このような方法で検索結果を整理したところ、分析対象の手の慣用句は全部で74項目となった。

3.2 分析指標としての「手」の慣用句

以下に指標74項目のリストを挙げる。

手が上がる、手が空く、手が後ろに回る、手がかかる、手が切れる、手が込む、手が足りる、手がつく、手が出る、手が届く、手がない、手が伸びる、手が入る、手が離れる、手が早い、手がふさがる、手が回る、手が焼ける、手に汗を握る、手に余る、手に入れる、手に負えない、手に落ちる、手に掛かる、手に掛ける、手にする、手につく、手に手を取る、手に取る、手になる、手にのる、手に入る、手に渡る、手のひら／裏を返す、手も足も出ない、手をあける、手をあげる、手を合わせる／合わす、手を入れる、手を打つ、手をかえる、手をかける、手を貸す、手を借りる、手を切る、手を下す、手を組む、手を加える、手をこまねく／こまぬく、手を差し伸べる、手を染める、手を出す、手をつかねる、手を尽くす、手をつける、手を取り合う、手を取る、手を握る、手を抜く、手を濡らす、手を引く、手を広げる、手を施す、手を回す、手を結ぶ、手を焼く、手を休める、赤子の手をひねる（ような）、飼い犬に手をかまれる、猫の手も借りたい、胸に手をあてる、手をのばす、手を離す、手をわずらわせる／わざらわす

3.3 調査結果

2. で述べた文章資料を対象に、3.2で抽出した「手」の慣用句74項目の出現回数を調査した。その調査結果を表3に示す。表中の代表項目は3.2で挙げた第1番目の項目によって代表する。

表3 各文章資料における「手」の慣用句の出現数

	代表項目	1 哲学	2 歴史	3 社会科学	4 自然科学	5 技術・工学	6 産業	7 芸術・美術	8 言語	9 文学
1	手が上がる	0	0	4	0	0	2	0	0	9
2	手が空く	1	0	2	0	0	1	0	0	9
3	手が後ろに回る	0	0	0	0	0	0	0	0	3
4	手がかかる	1	2	8	2	8	2	0	0	14
5	手が切れる	0	1	0	0	0	0	0	0	2
6	手が込む	1	5	3	0	2	1	4	0	23
7	手が足りる	0	0	3	0	0	0	0	0	8
8	手がつく	1	0	2	1	0	0	0	0	5
9	手が出る	1	1	5	1	4	1	1	0	22

10	手が届く	1	4	13	3	2	0	2	2	37
11	手がない	2	3	5	3	1	1	2	2	16
12	手が伸びる	0	3	3	0	0	1	2	0	35
13	手が入る	0	0	1	0	1	2	0	1	9
14	手が離れる	2	0	9	1	0	0	0	0	7
15	手が早い	0	0	2	0	0	0	0	0	4
16	手がふさがる	0	0	0	0	0	0	0	0	1
17	手が回る	1	1	4	0	3	2	0	0	16
18	手が焼ける	0	0	0	0	0	0	0	0	1
19	手に汗を握る	0	1	0	0	0	0	0	0	8
20	手に余る	5	1	6	0	0	0	1	2	16
21	手に入れる	56	74	158	30	48	19	31	7	371
22	手に負えない	0	2	3	0	0	0	0	1	10
23	手に落ちる	1	7	4	0	1	0	0	0	16
24	手に掛かる	1	8	2	0	0	0	5	0	32
25	手に掛ける	0	1	1	0	0	0	0	0	17
26	手にする	55	43	103	18	24	10	47	6	550
27	手につく	3	3	6	1	1	0	0	0	33
28	手に手を取る	0	0	1	0	0	0	0	0	8
29	手に取る	13	21	37	13	14	4	14	9	307
30	手になる	4	11	7	1	2	1	16	0	16
31	手にのる	1	0	0	0	0	0	0	0	9
32	手に入る	18	18	78	24	23	20	10	5	154
33	手に渡る	1	8	10	0	3	3	2	1	27
34	手のひらを返す	1	0	5	1	1	1	1	0	14
35	手も足も出ない	1	2	3	0	2	2	1	0	16
36	手をあげる	0	0	1	0	0	0	0	0	0
37	手をあげる	7	5	45	4	1	2	6	3	164
38	手を合わせる	10	8	15	3	0	0	1	0	76

39	手を入れる	7	4	10	3	8	3	12	2	95
40	手を打つ	5	26	35	2	5	0	18	1	102
41	手をかえる	1	1	4	0	3	1	2	1	8
42	手をかける	1	11	14	1	14	2	3	0	204
43	手を貸す	3	8	15	2	1	2	2	1	82
44	手を借りる	1	1	6	0	0	0	1	0	25
45	手を切る	1	4	5	3	1	0	1	0	28
46	手を下す	2	3	4	1	1	0	0	0	34
47	手を組む	3	7	15	1	2	0	2	1	37
48	手を加える	1	5	14	4	8	6	10	1	31
49	手をこまねく	3	3	13	2	1	3	1	1	18
50	手を差し伸べる	12	7	17	7	2	0	2	2	65
51	手を染める	6	2	11	3	3	1	4	3	21
52	手を出す	13	17	73	9	12	3	11	1	197
53	手をつかねる	0	0	0	0	0	0	0	0	4
54	手を尽くす	1	2	4	1	1	1	2	0	26
55	手をつける	12	8	43	7	10	2	7	9	110
56	手を取り合う	1	1	3	0	1	0	3	1	17
57	手を取る	5	7	12	3	2	1	8	0	172
58	手を握る	4	12	10	6	3	0	7	0	140
59	手を抜く	12	1	11	2	3	5	5	1	18
60	手を濡らす	0	0	0	0	0	0	0	0	4
61	手を引く	5	10	25	3	4	2	4	2	119
62	手を広げる	1	4	12	0	3	3	2	0	26
63	手を施す	1	0	1	1	1	0	0	0	8
64	手を回す	0	7	5	0	3	0	0	0	56
65	手を結ぶ	3	9	10	1	1	1	0	0	23
66	手を焼く	1	3	9	1	6	0	0	0	21
67	手を休める	0	0	1	0	0	0	1	0	4

68	赤子の手をひねる	0	0	0	0	0	1	1	0	5
69	飼い犬に手をかまれる	1	0	0	0	0	0	0	0	2
70	猫の手も借りたい	1	0	1	1	0	0	0	0	3
71	胸に手をあてる	1	2	2	0	0	0	2	0	11
72	手をのばす	11	11	18	4	7	1	8	7	347
73	手を離す	1	2	19	3	3	1	3	1	121
74	手をわずらわせる	3	0	2	1	1	0	1	0	9
	合計	311	411	973	178	251	114	269	74	4258

3.4 分析方法

「手」の慣用句74項目の中から文章のジャンル判別に特に有効な項目を選択するために、多変量解析の一手法である正準判別分析のステップワイズ法を用いて分析を行い、判別に寄与する項目を選択した。なお、分析の際には、2.1で述べたように文章資料の語数がそれぞれ異なるので、「手」の慣用句の出現数を100万語当たりの出現頻度に換算し直した出現率を用いた。

4. 分析結果と考察

3.4の分析方法により、74項目の「手」の慣用句を説明変数とし、文章資料グループ（以下、ジャンル）を基準変数としてステップワイズ法を用いて判別分析を行った。その結果、逐次的に5個の説明変数が予測的に組み込まれ、その手続き内で削除された変数もなく、2.4で述べた3つのジャンルの判別に有効な、以下の5つの慣用句が選択された。

- ①手に余る ②手を打つ ③手をこまねく／こまぬく
④手をあける ⑤手を取り合う

対象とした文章資料グループは3つのため、判別関数は2つ算出された。記述的指標としてウィルキンスの Λ を用い、 Λ に基づく χ^2 値については正規性の仮定が満たされないため、目安としてのみ用いる。これらに関する指標を示すと表4のようになる。 Λ の値は判別関数1と2の判別に大きく寄与する情報が含まれていることを示している。関数1のみで寄与率は97.8%である。

次に判別空間におけるジャンル間の関係について検討する。選択された5慣用句による判別関数平面での各文章資料の判別得点とジャンルの重心をプロットしたものを図1に示す。

表4 判別関数の固有値等

判別関数	固有値	寄与率	p 値	Λ	χ^2
関数 1	1,115.956	97.8	0.000	0.000	41.144
関数 2	25.248	2.2	0.011	0.038	13.070

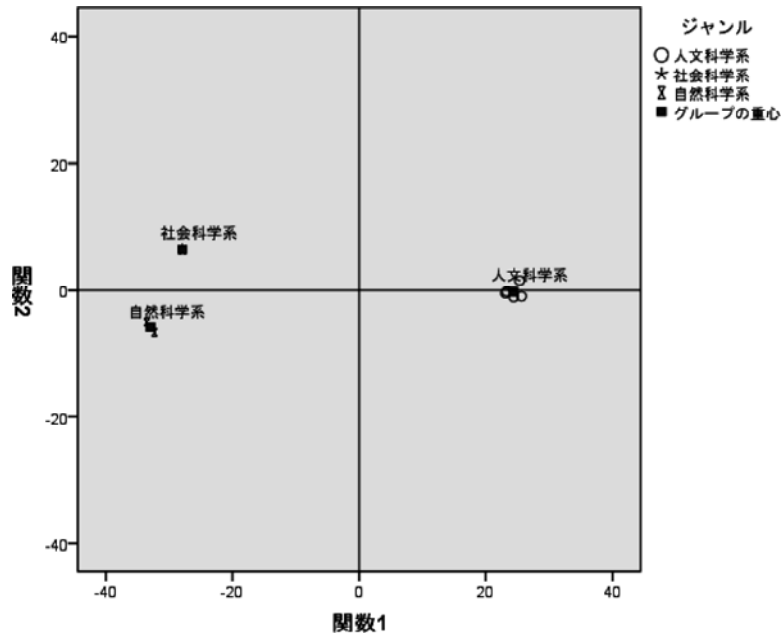


図1. 判別分析による9資料の重心および個体のプロット

図1を見るとわかるように、上記5慣用句によって3つの文章グループがはっきりと分離されているのが分かる。判別関数平面における各文章資料グループの重心の値は表5の通りである。

表5 判別空間における各文章資料グループの重心

ジャンル	関数 1	関数 2
人文科学系	24.351	-0.224
社会科学系	-27.937	6.423
自然科学系	-32.939	-5.862

関数1ではまず人文科学系が自然科学系と社会科学系から分離され、関数2では社会科学系が自然科学系から分離されている。

次に構造係数と判別空間における各ジャンルの重心の関係から、選択された5慣用句がどの文章資料グループを分離するのに有効かを考察する。表6に5慣用句の構造係数を示す。

表6 選択された5慣用句の構造係数

代表項目	関数1	関数2
手に余る	*0.024	0.019
手を打つ	*0.021	-0.005
手をこまねく	-0.010	*0.233
手をあける	-0.014	*0.147
手を取り合う	0.026	*-0.026

* 有意な係数

表6から関数1では「手に余る」「手を打つ」が人文科学系と自然科学系・社会科学系を分離し、関数2では「手をこまねく／こまぬく」「手をあける」「手を取り合う」が社会科学系と自然科学系を分離する特徴的な慣用句となっていることがわかる。ここで判別の可否を評価するための「見かけの的中率」をクロス集計で求めたところ、その結果は100%となった。さらに交差妥当性を検討したところ、その結果も同様に100%となった。

以上の結果から、選択された5つの「手」の慣用句を指標として、三つの文章資料のジャンルが判別できることが実証されたと考えられる。

最後に日本語教育の観点から、個々の「手」の慣用句の出現率の結果が教育現場の資料としては有意義だと考えられるので、表7にその結果を示す。

表7 各慣用句のジャンルごとの100万語当たりの出現頻度

	代表項目	人文科学系	社会科学系	自然科学系
1	手が上がる.	0.81	2.88	0.00
2	手が空く	1.37	1.44	0.00
3	手が後ろに回る	0.27	0.00	0.00
4	手がかかる	2.56	3.48	7.06
5	手が切れる	0.55	0.00	0.00
6	手が込む	7.52	1.59	1.46
7	手が足りる	0.72	0.45	0.00
8	手がつく	1.01	0.30	0.61
9	手が出る	3.67	1.89	3.53
10	手が届く	10.95	1.95	3.29

11	手がない	9.25	1.89	2.56
12	手が伸びる	5.78	1.59	0.00
13	手が入る	2.84	2.43	0.73
14	手が離れる	1.75	1.35	0.61
15	手が早い	0.36	0.30	0.00
16	手がふさがる	0.09	0.00	0.00
17	手が回る	2.37	2.88	2.19
18	手が焼ける	0.09	0.00	0.00
19	手に汗を握る	1.09	0.00	0.00
20	手に余る	9.43	0.90	0.00
21	手に入れる	129.90	45.36	53.34
22	手に負えない	3.67	0.45	0.00
23	手に落ちる	4.59	0.60	0.73
24	手に掛かる	10.20	0.30	0.00
25	手に掛ける	1.90	0.15	0.00
26	手にする	144.11	26.85	28.50
27	手につく	5.76	0.90	1.34
28	手に手を取る	0.72	0.15	0.00
29	手に取る	71.59	10.11	18.15
30	手になる	19.91	2.19	2.07
31	手にのる	1.37	0.00	0.00
32	手に入る	48.35	34.50	31.43
33	手に渡る	9.50	4.92	2.19
34	手のひらを返す	2.58	1.89	1.34
35	手も足も出ない	3.50	2.73	1.46
36	手をあける	0.00	0.15	0.00
37	手をあげる	31.18	9.03	3.17
38	手を合わせる	16.16	2.25	1.83
39	手を入れる	27.13	4.92	7.67

40	手を打つ	37.31	5.25	4.87
41	手をかえる	5.20	1.74	2.19
42	手をかける	25.27	4.38	10.83
43	手を貸す	15.57	4.53	1.95
44	手を借りる	3.94	0.90	0.00
45	手を切る	5.32	0.75	2.56
46	手を下す	5.29	0.60	1.34
47	手を組む	11.15	2.25	2.07
48	手を加える	14.83	8.94	8.28
49	手をこまねく	7.20	5.37	1.95
50	手を差し伸べる	20.74	2.55	5.73
51	手を染める	15.12	2.79	4.02
52	手を出す	41.69	14.37	14.25
53	手をつかぬる	0.36	0.00	0.00
54	手を尽くす	5.16	1.74	1.34
55	手をつける	43.17	8.73	11.57
56	手を取り合う	6.77	0.45	0.73
57	手を取る	26.95	2.94	3.29
58	手を握る	24.60	1.50	5.85
59	手を抜く	14.54	7.35	3.41
60	手を濡らす	0.36	0.00	0.00
61	手を引く	24.31	6.03	4.75
62	手を広げる	5.90	5.22	2.19
63	手を施す	1.28	0.15	1.34
64	手を回す	7.63	0.75	2.19
65	手を結ぶ	7.08	2.64	1.34
66	手を焼く	3.56	1.35	4.99
67	手を休める	1.12	0.15	0.00
68	赤子の手をひねる	1.21	1.14	0.00

69	飼い犬に手をかまれる	0.74	0.00	0.00
70	猫の手も借りたい	0.83	0.15	0.61
71	胸に手をあてる	3.81	0.30	0.00
72	手をのばす	61.75	3.84	7.55
73	手を離す	16.50	3.99	4.02
74	手をわずらわせる	3.25	0.30	1.34
	合計	1064.11	275.91	291.81

表7の結果は、日本語学習者が人文科学、社会科学、自然科学の各分野の文章において、どのような「手」の慣用句に触れる可能性が高いのかという一つの可能性を示していると言えよう。

5. おわりに

本研究では、多義性を持つ慣用句について特に意味上の分類は行わず一つの指標として扱って分析を行った。今後の課題としては、多義性を持つ「手」の慣用句について意味分類を行い、その意味機能情報を含む指標を用いて文章ジャンルを判別することの可能性を探っていきたいと思う。

本稿は文部科学省科学研究費基盤研究C（課題番号 20520429 研究代表者山崎誠）の補助を受けて行った研究成果の一部である。

注

- 1) ここではジャンルという語を個々人の持つ文体的特徴を超えたところに存在するある特徴パターンを持った文章グループと定義する。
- 2) 「語数」は、コーパスの構築に際して使われている解析の言語単位である「短単位」で数えたもの。空白、補助記号（句読点など）、記号（A, B, C, ア, イ, ウなど）を含まない数である。
- 3) 書き言葉コーパスの設計と目的に照らして、以下のような書籍は対象外とした。40ページ以下の書籍、ページ数情報のない書籍（ランダムサンプリングの際にページ情報を利用するため）、官公庁刊行物のうち流通していないもの、学習試験図書、電子資料、地図資料、写真集、漫画などである。

謝 辞

本稿の慣用句データの整理については本塾大学院文学研究科国文学科日本語教育学分野修士2年生顧翌清さんの協力を得た。ここに記して感謝の意を表したい。

参 考 文 献

- 宮地 裕 (1982) 『慣用句の意味と用法』 明治書院。
 前川喜久雄 (2008), 「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発」, 日本語の研究, 4(1), pp. 82-95.
 前川喜久雄 (2009) 「代表性を有する大規模日本語書き言葉コーパスの構築」 人工知能学会誌, 24(5), pp. 616-622.

- 山崎 誠 (2009)「代表性を有する現代日本語書籍コーパスの構築」人工知能学会誌, 24(5), pp. 623-631.
- 村田 年 (2007)「専門日本語教育における論述文指導のための接続語句・助詞相当句の研究」『統計数理 (特集「文化を科学する」)』統計数理研究所 Vol. 55, No. 2, pp. 269-284.
- 村田 年 (2008)「文章と複合動詞—論述文ジャンルを特徴づける新たな指標を探して—」『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター36号 pp. 1-33.

関連 URL

「KOTONOHA 国立国語研究所言語コーパス整備計画」

<http://www.ninjal.ac.jp/kotonoha/>

「ことば不思議箱」

<http://kotoba.nuee.nagoya-u.ac.jp/>

自然科学系書籍における複合動詞の使用傾向

—後項動詞を指標として—

村田 年 (慶應義塾大学日本語・日本文化教育センター)

山崎 誠 (国立国語研究所)

1. はじめに

専門日本語教育における論文をはじめとする論述的文章指導のためには、日本語学習者に対して文章ジャンルの特徴パターンの違いを言語要素の具体的な指標を用いて示すことが有効だと考えられる。

村田 (2008) では、文章のジャンル判別のための新たな指標を求め、複合動詞の後項動詞をその指標候補として取り上げた。ここで言う後項動詞とは複合動詞において連用形の前項動詞に続く後要素の動詞を意味する。312編の資料(経済学入門書、経済学論文、工学論文、物理学論文、文学論文、新聞社説、近代小説、現代短編小説)を対象に、論述的文章ジャンルにおける複合動詞の使用傾向を見るために、選択した26個の後項動詞の使用頻度を調べ、それを小説・社説ジャンルの場合と比較した。その結果、「だす」「こむ」は造語力が非常に強く、複合動詞として多用されていること、また複合動詞全体の使用については小説・社説ジャンルの資料(新聞社説、近代小説、現代短編小説)のほうが論述的文章ジャンルの資料(経済学入門書、経済学論文、工学論文、物理学論文、文学論文)より多いことが明らかとなった。さらに非常に限られた資料の範囲ではあるが、論述的文章ジャンルでは、「あげる」が「取り上げる」「引き上げる」の形で使用頻度が高く、「たつ」も「成り立つ」の形で多用され、小説・社説ジャンルにおける「あげる」「たつ」の使用頻度の平均値を上回っていた。このパイロット研究によって、複合動詞の後項動詞が文章のジャンル判別のために有効な指標となり得ることが示唆されたので、より多くの文章資料を対象に実証分析を行いたいと考えた。

2. 研究目的

複合動詞の後項動詞が文章のジャンル判別の指標となり得るかどうかを実証的に分析することを目的として、本研究ではまず自然科学系ジャンルの文章における複合動詞の使用傾向を明らかにする。今回は、『現代日本語書き言葉均衡コーパス』(BCCWJ)の書籍コーパス内の自然科学系ジャンルである「自然科学」と「技術・工学」の文章を対象に後項動詞の使用頻度調査を行うことにした。自然科学系ジャンルを選んだ理由は、論述的文章ジャンルの中でも特に自然科学分野は強い論理性が求められ、そこで多用される複合動詞は論述的文章に現れる複合動詞の雛形と捉えることができるのではないかと考えたことによる。

3. 調査対象

3.1 調査対象としての後項動詞

調査対象の後項動詞は、村田 (2008) と同様、姫野 (1999) を参考に選択した 22 の動詞のほか、アスペクトを表す「始める」「続ける」「終わる」と過剰・過度を表す「すぎる」の 4 動詞を含めた 26 動詞である。姫野 (1999) が指摘するように、アスペクトを表す 3 動詞については、「終わる」以外は時間と関係するほとんどの動詞と結合する可能性があり、「終わる」も意志的行為の終了を表すだけで造語力は低い。また「すぎる」も過剰・過度を表せる多くの動詞と結合できる。そのため、この 4 動詞の使用頻度を調べることで、あまり意味がないという側面があることは確かである。しかし、これらの 4 動詞は日本語教育においては初・中級レベルでその造語力を学ぶ、使用頻度の高い動詞である。筆者は専門日本語教育を行う立場から、これらの 4 動詞が自然科学系ジャンルの文章で実際にどのような前項動詞と結合し、どのぐらいの頻度で用いられているかという複合動詞の使用実態についても関心があるので、本調査の対象語に加えることにした。

以下に 26 動詞を挙げる。

<26 後項動詞> (アイウエオ順)

あう、あがる、あげる、あわせる、いる、いれる、おわる、かかる、かける、きる、こむ、こめる、すぎる、だす、たつ、たてる、つく、つける、つづける、でる、とおす、なおす、なおる、ぬく、はじめる、まくる

3.2 調査対象としての文章資料

3.2.1 『現代日本語書き言葉均衡コーパス』

調査に用いた資料は、『現代日本語書き言葉均衡コーパス』(以下 BCCWJ) の現時点での最新版のデータである。検索はオンライン検索ツール「中納言」を利用し、2011 年 9~12 月に検索を行った。なお、BCCWJ のデータは 2012 年 3 月頃に微修正を加えたものにバージョンアップされる予定であるが、本稿で扱った書籍及び白書のデータについてはその差異はきわめてわずかである。検索を行った時点での BCCWJ のデータ量は表 1 のとおりである。なお、BCCWJ については、前川 (2008, 2009) を、この調査で使用した書籍の部分については山崎 (2009) を参照されたい。

表 1 BCCWJ のデータ量

サブコーパス	ジャンル	サンプル数	語数 (万)
	書籍	10,105	2,849
出版サブコーパス	雑誌	1,989	443
	新聞	1,479	138
図書館サブコーパス	書籍	10,461	3,011
	白書	1,500	488
	教科書	412	93

	広報紙	354	400
特定目的サブコーパス	ベストセラー	1,377	371
	Yahoo!知恵袋	91,445	1,028
	Yahoo!ブログ	52,680	1,027
	韻文	253	23
	法律	346	100
	国会会議録	159	510
合計		172,560	10,481

(注) 「語数」は、コーパスの構築に際して使われている解析の言語単位である「短単位」で数えたもの。空白、補助記号(句読点など)、記号(A、B、C、ア、イ、ウなど)を含まない数である。

3.2.2 データの母集団

この調査で使用した書籍データは3種類のデータから成っている。同じ書籍であるが、それぞれ母集団が異なる。一つ目は「出版サブコーパス・書籍」で、2001～2005年に日本国内で出版された全書籍からコーパス収録条件で絞り込んだ約32万冊が母集団である。コーパス収録条件とは、書き言葉コーパスの設計と目的に照らして適切かどうかの条件であり、これにより、40ページ以下の書籍、ページ数情報のない書籍(ランダムサンプリングの際にページ情報を利用するため)、官公庁刊行物のうち流通していないもの、学習試験図書、電子資料、地図資料、写真集、漫画などは対象外とした。二つ目は「図書館サブコーパス・書籍」で、東京都内の自治体ごとにISBNにより管理されている蔵書データを利用して、13自治体以上に共通して所蔵されている書籍からコーパス収録条件で絞り込んだ約38万冊が母集団である。ここでは、ISBNの付与が普及した1986～2005年刊行の書籍を対象とした。三つ目は、1976～2005年版の『出版年鑑』及び『出版指標年報』に掲載されたベストセラーリスト上位20位に挙げられた書籍約950冊を対象としている。書籍内は11に分類され、その内訳は「総記、哲学、歴史、社会科学、自然科学、技術・工学、産業、芸術・美術、言語、文学、なし」である。

3.2.3 サンプルの抽出方法

書籍のサンプルは1万字を超えない範囲のひとまとまりの文章(1章、1節など外的な構成を基準として判断したまとまり)である。母集団の全ページをサンプリングフレームと考え、そこからランダムに抽出したページの中の1点をランダムに指定し、その点を基準にテキストを抽出している。

4. 調査方法

4.1 複合動詞の抽出方法

複合動詞の抽出に当たっては後項動詞となる26動詞を対象として検索を行った。辞書の見出

しに相当する語彙素を手がかりにしたが、若干の調整を加えた。その理由は、BCCWJ の形態素解析の元になっている解析用辞書 UniDic の語彙素と本研究で扱う後項動詞とが必ずしも 1 対 1 で対応していないためである。例えば「あう」は「会う」「合う」の 2 つの語彙素で検索した。これは、BCCWJ では「会う」と「合う」が別の動詞として扱われているためである。この区別は意味的なものであり、用字法に拠っているものではない。同様にして、検索対象となった動詞の語彙素を () 内に示す。

あう (会う、合う)、あがる (上がる)、あげる (上げる、揚げる)、あわせる (合わせる)、いる (入る)、入れる (入れる)、おわる (終わる)、かかる (掛かる、懸かる)、かける (掛ける)、きる (切る)、こむ (込む、こむ)、こめる (込める)、すぎる (過ぎる)、だす (出す)、たつ (立つ)、たてる (立てる)、つく (付く、着く、衝く、突く、漬く、つく)、つける (付ける、着ける、点ける、突ける、憑ける、つける)、つづける (続ける)、でる (出る)、とおす (通す)、なおす (直す)、なおる (直す)、ぬく (抜く)、はじめる (始める)、まくる (捲る)

これらの中には検索結果が 1 例 (揚げる、憑ける、点ける) というものもあるが、データの再現性を重視して別立てにしてある。検索条件は、長単位検索を用い、以下の条件を指定した。以下の例は、後項動詞「まくる」の場合である。

キー：(語彙素 LIKE "_%捲る") WITH OPTIONS unit
="2" AND tglWords="100" AND tglKugiri="" AND tglFixVariable="2"

なお、BCCWJ の形態素解析の精度は約 98% であり、約 2% の誤解析が含まれている。検索された例における誤解析は手作業で確認し外したが、検索されなかった例についてはフォローすることができないため、そのままデータとして用いていることを明記しておく。

4.2 調査結果の整理—複合動詞の選択方法—

4.1 節の調査結果を見ると、用字の異なりをはじめ、前項動詞の音便形の存在などいくつかの問題が出てきた。そこで、本調査では以下に述べる方法で見出し語としての複合動詞を選択した。

- (a) 各後項動詞の用字の差異 (例：あう／合う／会う) ならびに動詞の活用変化の形 (例：あわない、あいます、あう、あえば、あおう、あって、あった等) は同定して同じ動詞項目として扱った。
- (b) 複合動詞の前項動詞については、同音で用字が異なるものはコーパスの語彙素項目の用字に合わせて整理した。例えば、「書き出す」は「搔き出す」とは別の見出し語になっているが、「探し出す」は「捜し出す」として一つの見出し語となっている。このように本調査の結果としてまとめた表内の複合動詞の代表見出し項目については、異なる用字の可能性も含まれている。
- (c) 本調査では複合動詞の動詞としての用法を対象とするので、語彙素項目が動詞でも、実際には名詞、副詞として用いられているものは使用頻度からは外した。例えば「思い切る」「張り切る」の見出し語で「思い切って実施する」「張り切った声」のように用いられている場合や「届け出る」の見出し語で「届け出」のように用いられている場合である。

- (d) 前項動詞が音便形になっている語については、斎藤 (1992) が指摘するように、前項動詞が音便形になる語が必ずしも非音便形を持つわけではなく、音便形と非音便形の意味関係も同じとは限らない。例えば、辞書¹⁾で「ぶつける」は「ぶつつけるの転」という説明があるが、実際には各語の意味するところは同一とは言えない。このように複雑な意味関係を持つため別の見出し語として立てた。
- (e) 原文で例えば「追いつける」「隣り合わせる」のように出現し、意味から考えて対応する「追いつく」「隣り合う」の可能形、使役形として考えられるものは元の形の複合動詞として数えた。しかし、「落ち着ける」のように「腰/身を落ち着ける」の形が「腰/身が落ち着く」の態の変化と捉えられないものはそのまま「落ち着ける」の項目として扱った。
- (f) 動詞が三つ続く場合は最後尾の動詞を後項動詞とした。
- 以上の方法でデータを整理した。

5. 調査結果

5.1 結果の概要

書籍コーパスの自然科学系ジャンル（自然科学、技術・工学）における 26 の後項動詞による複合動詞の総延べ語数は 29254 語で、約 3 万語に上った。各後項動詞の使用頻度（延べ語数）は表 2 のとおりである。

表 2 で使用頻度が圧倒的に高いのは、4500 語を超える「だす」「こむ」の 2 動詞で、これらを後項動詞とする複合動詞は、本調査における全複合動詞の約 32% を占めている。次に造語力の高い 2000 語前後の動詞は「つける」「あげる」「あう」の 3 動詞で、これらから成る複合動詞を上記の結果と合わせれば、全複合動詞の約 54% を占めることになる。さらに、1200～1300 語台の「かける」「はじめる」「つく」「あわせる」「あがる」「いれる」までの 11 動詞による複合動詞を合わせれば全体の約 81%（「はじめる」を除くと約 76%）となる。つまり、自然科学系ジャンルの書籍に出現する約 3 万語の複合動詞の中で、本調査で対象とした 11 の後項動詞から成る複合動詞でその 8 割以上がカバーされるということである。

次に、同じ複合動詞の中での異なり語数を後項動詞別に見ていく。表 3 は各後項動詞に結合する前項動詞の異なり語数を示したものである。

表 2 26 後項動詞使用頻度順

	後項動詞	延べ語数
1	だす	4683
2	こむ	4609
3	つける	2484
4	あげる	2238

表 3 26 後項動詞の異なり語頻度順

	後項動詞	異なり語数
1	はじめる	471
2	つづける	342
3	あう	237
4	すぎる	220

5	あう	1907
6	かける	1321
7	はじめる	1319
8	つく	1275
9	あわせる	1274
10	あがる	1249
11	いれる	1216
12	つづける	1087
13	すぎる	797
14	たつ	665
15	きる	613
16	なおす	572
17	たてる	560
18	でる	386
19	かかる	272
20	ぬく	224
21	いる	125
22	おわる	123
23	こめる	106
24	まくる	58
25	とおす	53
26	なおる	38
	合計	29254

5	だす	217
6	こむ	179
7	なおす	128
8	あげる	103
9	きる	102
10	かける	94
11	つける	88
12	おわる	60
13	あがる	57
14	つく	54
15	いれる	52
16	あわせる	41
17	たてる	38
18	でる	38
19	まくる	38
20	かかる	35
21	ぬく	28
22	いる	25
23	たつ	21
24	とおす	13
25	こめる	9
26	なおる	4
	合計	2694

表3で上位に「はじめる」「つづける」「すぎる」が来ているのは、3.1節で述べたように、結合の可能性を持つ動詞が多数あるため、当然の結果だと考えられる。延べ語数が多く使用頻度が高い後項動詞を見ていくと、異なり語数も多いという傾向が見られる。しかし、使用頻度が高ければ必ず結合する前項動詞の種類も多いというわけではない。例えば、延べ語数 2000 語程度の

表4 26 後項動詞による使用頻度 10 回以上の複合動詞

だす(4683*217)				こむ(4609*179)							
1	取り出す	562	56	怒り出す	13	1	取り込む	309	56	信じ込む	19
2	作り出す	455	57	繰り出す	12	2	組み込む	226	57	座り込む	18
3	生み出す	396	58	剥き出す	12	3	入り込む	190	58	そそぎ込む	18
4	思い出す	351	59	湧き出す	12	4	持ち込む	152	59	呼び込む	18
5	引き出す	268	60	洗い出す	11	5	差し込む	146	60	つき込む	17
6	飛び出す	124	61	起き出す	11	6	煮込む	146	61	考え込む	16
7	送り出す	111	62	曝け出す	11	7	書き込む	142	62	炊き込む	16
8	吐き出す	89	63	成り出す	11	8	まき込む	119	63	引っ込む	16
9	呼び出す	86	64	這い出す	11	9	飛び込む	117	64	買い込む	15
10	押し出す	80	65	貸し出す	10	10	落ち込む	108	65	切り込む	15
11	打ち出す	76	66	駆り出す	10	11	吸い込む	107	66	しゃがみ込む	15
12	溶け出す	68				12	染み込む	106	67	滑り込む	15
13	乗り出す	67		でる(386*38)		13	思い込む	104	68	使い込む	15
14	持ち出す	65	1	届け出る	49	14	流れ込む	95	69	揉み込む	15
15	言い出す	63	2	申し出る	48	15	埋め込む	91	70	入れ込む	14
16	突き出す	63	3	流れ出る	45	16	飲み込む	90	71	積み込む	14
17	考え出す	59	4	にじみ出る	31	17	読み込む	82	72	挟み込む	14
18	切り出す	55	5	突き出る	26	18	送り込む	80	73	しまい込む	13
19	見付け出す	55	6	湧き出る	20	19	つけ込む	73	74	放り込む	13
20	動き出す	51	7	はみ出る	17	20	見込む	71	75	売り込む	12
21	醸し出す	48	8	浮き出る	16	21	打ち込む	68	76	教え込む	12
22	抜け出す	48	9	溢れ出る	14	22	盛り込む	63	77	決め込む	12
23	捜し出す	47	10	生まれ出る	13	23	突っ込む	61	78	寝込む	12
24	踏み出す	47	11	飛び出る	13	24	押し込む	58	79	焼き込む	12
25	食み出す	46	12	抜け出る	13	25	溶け込む	57	80	着込む	11
26	追い出す	45	13	吹き出る	10	26	追い込む	56	81	冷え込む	11
27	逃げ出す	41				27	踏み込む	55	82	回り込む	11
28	絞り出す	39		たつ(665*21)		28	し込む	51	83	頼み込む	10
29	導き出す	39	1	成り立つ	352	29	絞り込む	48	84	溶かし込む	10
30	吹き出す	38	2	煮立つ	129	30	刷り込む	47	85	投げ込む	10
31	差し出す	37	3	飛び立つ	35	31	詰め込む	46	86	塗り込む	10
32	流れ出す	37	4	思い立つ	31	32	乗り込む	46		こめる(106*9)	
33	読み出す	33	5	引き立つ	28	33	吹き込む	46	1	閉じ込める	62
34	映し出す	31	6	下り立つ	12	34	包み込む	45	2	立ち込める	15
35	描き出す	31	7	浮き立つ	11	35	流し込む	44			
36	売り出す	29	8	そそり立つ	11	36	引き込む	35		なおす(572*128)	
37	選り出す	29	9	沸き立つ	10	37	潜り込む	35	1	見直す	176
38	煮出す	29		たてる(560*38)		38	抱え込む	34	2	遣り直す	49
39	張り出す	29	1	組み立てる	137	39	運び込む	34	3	書き直す	29
40	掘り出す	27	2	煮立てる	93	40	覗き込む	32	4	考え直す	29
41	書き出す	26	3	見立てる	49	41	申し込む	29	5	作り直す	24
42	聞き出す	25	4	仕立てる	41	42	ため込む	28	6	問い直す	16
43	割り出す	24	5	埋め立てる	38	43	押さえ込む	27	7	立て直す	14
44	掻き出す	22	6	引き立てる	37	44	駆け込む	26	8	見詰め直す	13
45	走り出す	21	7	掻き立てる	29	45	くい込む	26	9	取り直す	12
46	編み出す	20	8	駆り立てる	18	46	割り込む	25	10	持ち直す	10
47	汲み出す	18	9	打ち立てる	17	47	沈み込む	24		なおる(38*4)	
48	投げ出す	18	10	申し立てる	13	48	はめ込む	24	1	立ち直る	21
49	引っ張り出す	18	11	飾り立てる	11	49	織り込む	23	2	開き直る	12
50	放り出す	17	12	積み立てる	10	50	混ぜ込む	23			
51	溢れ出す	16				51	逃げ込む	22		ぬく(224*28)	
52	連れ出す	14		とおす(53*13)		52	のめり込む	21	1	くり抜く	56
53	泣き出す	14	1	見通す	21	53	折り込む	20	2	引き抜く	56
54	抜き出す	14		まぐる(58*38)		54	住み込む	20	3	見抜く	22
55	弾き出す	14			0	55	刻み込む	19	4	生き抜く	16

つける(2484*88)				あげる(2238*103)							
1	見付ける	544	14	焼き付く	22	1	取り上げる	417	17	湧き上がる	20
2	取り付ける	413	15	噛み付く	18	2	仕上げる	245	18	跳ね上がる	17
3	結び付ける	143	16	吸い付く	18	3	作り上げる	183	19	腫れ上がる	14
4	張り付ける	139	17	絡み付く	16	4	引き上げる	156	20	燃え上がる	14
5	盛り付ける	131	18	纏わり付く	16	5	持ち上げる	153	21	晴れ上がる	12
6	巻き付ける	88	19	流れつく	15	6	申し上げる	105			
7	押し付ける	86	20	染み付く	12	7	積み上げる	73		あう(1907*237)	
8	受け付ける	60	21	抱き付く	12	8	押し上げる	66	1	出会う	305
9	引き付ける	60	22	生まれ付く	11	9	たち上げる	65	2	付き合う	159
10	縫い付ける	53	23	食い付く	10	10	打ち上げる	64	3	話し合う	133
11	駆け付ける	47				11	見上げる	63	4	見合う	73
12	ぶつける	45		いれる(1216*52)		12	吊り上げる	32	5	似合う	69
13	締め付ける	43	1	受け入れる	380	13	焼き上げる	28	6	向き合う	67
14	決め付ける	36	2	取り入れる	365	14	築き上げる	27	7	絡み合う	58
15	吹き付ける	28	3	流し入れる	87	15	差し上げる	26	8	重なり合う	55
16	突き付ける	27	4	仕入れる	39	16	汲み上げる	23	9	知り合う	43
17	押さえ付ける	26	5	組み入れる	27	17	盛り上げる	22	9	立ち会う	43
18	焼き付ける	26	6	溶き入れる	22	18	書き上げる	20	11	向かい合う	37
19	遣っ付ける	26	7	振り入れる	21	19	吸い上げる	18	12	触れ合う	35
20	こすり付ける	25	8	申し入れる	21	20	突き上げる	18	13	噛み合う	31
21	割り付ける	24	9	回し入れる	20	21	吹き上げる	17	14	釣り合う	30
22	塗り付ける	21	10	注ぎ入れる	17	22	練り上げる	16	15	混ざり合う	25
23	生み付ける	19	11	踏み入れる	17	23	敷え上げる	15	16	助け合う	23
24	括り付ける	17	12	並べ入れる	16	24	炊き上げる	15	17	混じり合う	23
25	寄せ付ける	17	13	迎え入れる	13	25	拾い上げる	15	18	巡り会う	23
26	辿り着ける	16	14	投げ入れる	12	26	切り上げる	14	19	引き合う	22
27	漕ぎ着ける	15	15	ふるい入れる	12	27	掴み上げる	14	20	競い合う	21
28	植え付ける	14	16	聞き入れる	11	28	抱き上げる	13	21	出し合う	19
29	打ち付ける	14	17	差し入れる	10	29	引っ張り上げる	13	22	取り合う	18
30	叩き付ける	14	18	通し入れる	10	30	纏め上げる	13	23	語り合う	17
31	落ち着ける	13	19	混ぜ入れる	10	31	育て上げる	12	24	協力し合う	16
32	縛り付ける	13	20	戻し入れる	10	32	降り上げる	12	25	隣り合う	16
33	据え付ける	13	21	割り入れる	10	33	磨き上げる	12	26	ぶつかり合う	16
34	踏み付ける	12		いる(125*25)		34	込み上げる	11	27	関わり合う	15
35	痛め付ける	10	1	立ち入る	33	35	巻き上げる	11	28	分かち合う	15
36	送り付ける	10	2	込み入る	14	36	読み上げる	11	29	分け合う	15
37	書き付ける	10	3	見入る	11	37	買い上げる	10	30	関連し合う	14
38	飾り付ける	10				38	組み上げる	10	31	溶け合う	12
39	組み付ける	10		きる(613*102)			あがる(1249*57)		32	支え合う	11
40	照り付ける	10	1	使い切る	57	1	出来上がる	280	33	交換し合う	10
			2	仕切る	55	2	立ち上がる	149	34	抱き合う	10
			3	乗り切る	52	3	仕上がる	114			
	つく(1275*54)		4	踏み切る	43	4	盛り上がる	83		あわせる(1274*41)	
1	落ち着く	240	5	断ち切る	35	5	召し上がる	61	1	組み合わせる	402
2	結び付く	233	6	言い切る	28	6	炊き上がる	59	2	混ぜ合わせる	363
3	辿り着く	92	7	打ち切る	25	7	浮かび上がる	47	3	間い合わせる	66
4	思い付く	91	8	割り切る	23	8	焼き上がる	46	4	縫い合わせる	63
5	住み着く	59	9	押し切る	16	9	浮き上がる	42	5	炒め合わせる	45
6	行き着く	53	10	思い切る	16	10	打ち上がる	31	6	重ね合わせる	37
7	取り付く	50	11	決まり切る	15	11	持ち上がる	31	7	盛り合わせる	27
8	追い付く	46	12	疲れ切る	15	12	膨れ上がる	29	8	繋ぎ合わせる	22
9	こびり付く	34	13	締め切る	14	13	飛び上がる	25	9	考え合わせる	20
10	張り付く	33	14	分かり切る	13	14	茹で上がる	25	10	照らし合わせる	19
11	考え付く	32	15	出し切る	10	15	起き上がる	24	11	掛け合わせる	17
12	しがみ付く	23	16	伸び切る	10	16	舞い上がる	22	12	すり合わせる	17
13	飛び付く	23									

		はじめる(1319*471)		つづける(1087*342)		
13	突き合わせる	16	1 出始める	45	1 飲み続ける	49
14	練り合わせる	16	2 飲み始める	40	2 増え続ける	48
15	張り合わせる	15	3 動き始める	27	3 生き続ける	43
16	取り合わせる	14	4 使い始める	25	4 持ち続ける	42
17	居合わせる	13	5 作り始める	25	5 有り続ける	35
18	食べ合わせる	12	6 なり始める	23	6 使い続ける	35
19	打ち合わせる	10	7 考え始める	22	7 言い続ける	21
20	こすり合わせる	10	8 思い始める	20	8 作り続ける	21
21	足し合わせる	10	9 話し始める	19	9 動き続ける	18
			10 増え始める	19	10 取り続ける	17
	かける(1321*94)		11 感じ始める	18	11 し続ける	16
1	出掛ける	318	12 書き始める	16	12 働き続ける	15
2	見掛ける	159	13 取り始める	16	13 食べ続ける	13
3	振り掛ける	124	14 普及し始める	16	14 生存し続ける	12
4	働き掛ける	94	15 現われ始める	15	15 塗り続ける	12
5	話し掛ける	91	16 見え始める	15	16 与え続ける	11
6	追い掛ける	57	17 歩き始める	14	17 歩き続ける	11
7	呼び掛ける	56	18 し始める	14	18 考え続ける	11
8	投げ掛ける	42	19 持ち始める	14	19 吸い続ける	11
9	仕掛ける	38	20 起こり始める	13	20 存在し続ける	11
10	引っ掛ける	36	21 食べ始める	13	21 こだわり続ける	10
11	語り掛ける	31	22 目立ち始める	13	22 住み続ける	10
12	回し掛ける	28	23 吸い始める	12	23 走り続ける	10
13	問い掛ける	23	24 出来始める	12	24 減り続ける	10
14	立て掛ける	21	25 見せ始める	12		
15	吹き掛ける	16	26 語り始める	11	すぎる(797*220)	
16	持ち掛ける	16	27 働き始める	11	1 取り過ぎる	68
17	押し掛ける	11	28 付き始める	10	2 食べ過ぎる	59
18	成り掛ける	11	29 取り組み始める	10	3 なり過ぎる	55
19	見せ掛ける	10			4 通し過ぎる	39
	かかる(272*35)		おわる(123*60)		5 飲み過ぎる	34
1	取り掛かる	62	1 食べ終わる	21	6 掛かり過ぎる	26
2	引っ掛かる	59	2 使い終わる	13	7 入れ過ぎる	20
3	差し掛かる	26			8 行き過ぎる	19
4	襲い掛かる	18			9 し過ぎる	19
5	降り懸かる	16			10 増え過ぎる	18
6	通り掛かる	15			11 遣り過ぎる	18
7	申し掛かる	13			12 使い過ぎる	17
8	殴り掛かる	10			13 有り過ぎる	11
					14 出過ぎる	11

「あう」の異なり語数は237語で、延べ語数4500語以上の「だす」と「こむ」を上回っている。また延べ語数が665語で26動詞中14位の「たつ」は異なり語数が21語であるが、延べ語数が613語の「きる」は異なり語数が102語、延べ語数572語の「なおす」は128語で、「たつ」の異なり語数の5~6倍となっている。この数値は、延べ語数が1200語台で8位から11位の「つく」「あわせる」「あがる」「いれる」の異なり語数(54語、41語、57語、52語)と比べても約2倍となっている。このように結合する前項動詞のバリエーションは後項動詞によって大きな差があることがわかる。

次に26の各後項動詞ごとに複合動詞の使用傾向を具体的に見ていく。ここでは対象となる複合動詞が約3万語と多いため、紙幅の関係で便宜上使用頻度が10回以上のものだけを表4に頻度順に挙げる。

表4では、各見出し後項動詞の横に、その後項動詞から成る複合動詞の総数に当たる延べ語数と異なり語数を(4683*217)のように記してある。まず、さまざまな動詞と結合する可能性を持つ「はじめる」「つづける」「おわる」「すぎる」の4動詞を除いた22の後項動詞について見てい

く。なお、この4語については5.1.6節で触れる。

表4を見ると、ほとんどの後項動詞で、使用頻度が非常に高い複合動詞は上位の限られた数の語であることに気づく。これら上位の複合動詞の使用は、当該後項動詞全体の中でどのぐらいの割合を占めているのだろうか。各後項動詞によって構成される複合動詞はその総数が異なるため、ここでは試みに延べ語数が1000語以上のものは上位4語の場合と8語の場合、1000語未満のものは上位2語の場合と4語の場合についてそれぞれ全体数における使用割合を調べた。その結果を表5に示す。

表5 22後項動詞の使用頻度上位語が全体に占める割合

	後項動詞	延べ語数	異なり語数	語数	A 使用割合	語数	B 使用割合
1	だす	4683	217	4	38%	8	50%
2	こむ	4609	179	4	19%	8	31%
3	つける	2484	88	4	50%	8	65%
4	あげる	2238	103	4	45%	8	62%
5	あう	1907	237	4	35%	8	48%
6	かける	1321	94	4	53%	8	71%
7	つく	1275	54	4	51%	8	68%
8	あわせる	1274	41	4	70%	8	80%
9	あがる	1249	57	4	50%	8	67%
10	いれる	1216	52	4	72%	8	79%
11	たつ	665	21	2	72%	4	82%
12	きる	613	102	2	18%	4	34%
13	なおす	572	128	2	39%	4	49%
14	たてる	560	38	2	41%	4	57%
15	でる	386	38	2	25%	4	45%
16	かかる	272	35	2	44%	4	61%
17	ぬく	224	28	2	50%	4	67%
18	いる	125	25	2	38%	4	54%
19	こめる	106	9	2	72%	4	88%
20	まくる	58	38	2	13%	4	24%

21	とおす	53	13	2	57%	4	79%
22	なおる	38	4	2	87%	4	100%

注：A 使用割合は上位 2 語と 4 語の場合で B 使用割合は上位 4 語と 8 語の場合である。

表 5 からわかるように、上位 2 語あるいは 4 語という非常に限られた数の語だけで使用割合が全体の 50%を超える後項動詞が 22 語中 11 語もある。範囲をもう少し広げ、上位 4 語あるいは 8 語までとすれば、それらの語の使用割合が全体の約 50%あるいは 50%を超える後項動詞は 22 語中 18 語 (49%の「なおす」、48%の「あう」も含める) に上り、後項動詞全体の約 8 割を超える。残りの 4 語は「こむ」「きる」「でる」「まくる」であるが、これらの語を後項動詞とする複合動詞は相対的に頻度の高い語が少なく、使用がばらけていることがわかる。

上記の結果が意味することは、自然科学系ジャンルの書籍の文章で用いられる複合動詞は、22 の後項動詞から見た場合、使用頻度が高い複合動詞は非常に限られた数の語だということである。すでに見たように複合動詞全体の使用頻度を見れば、26 の後項動詞のうち 5 つの動詞(「だす」「こむ」「つける」「あげる」「あう」) から構成される複合動詞の使用数のみで全体の約 54%と、5 割を超えている。この使用頻度の高い 5 つの後項動詞によって構成される複合動詞の中で頻度が高い複合動詞は、自然科学系ジャンルの書籍の文章に特徴的な複合動詞だと考えられる。

以下、上記 5 つの後項動詞から成る複合動詞の使用傾向について具体的に見ていく。なお、意味特徴については姫野(1999)の分類に依拠する。

5.1 個別結果

5.1.1 だす

「だす」は村田 (2008) のパイロット研究でも造語力は 26 語中 1 位であった。使用頻度上位 8 語のみで全体の 50%を占めている。最上位は 562 語の「取り出す」で、次が 455 語の「作り出す」、396 語の「生み出す」351 語の「思い出す」と続く。少し離れて 268 語の「引き出す」、さらに半減して 124 語の「飛び出す」、111 語の「送り出す」、89 語の「吐き出す」の順である。意味特徴を見ると 8 語すべてが語彙的複合動詞で、開始のアスペクトを表わす統語的複合動詞は入っていない。姫野(1999)も指摘するように、「出す」の第一義は「外部への移動」であり、それが転じて、出現や物事の顕在化を表わす。頻度上位語のうち、「取り出す」「引き出す」「飛び出す」「送り出す」「吐き出す」の「出す」は前項動詞の修飾関係から外部への移動の方法を表わしている。「取り出す」「引き出す」「送り出す」「吐き出す」は他動詞として、「飛び出す」は自動詞として働く。一方「作り出す」「生み出す」「思い出す」は顕在化の意味を持ち、「作り出す」「生み出す」は創出²⁾を表わし、「思い出す」は顕現³⁾を表わして 3 語とも他動詞として働く。

次に名詞との結びつきの観点から見ると、「思い出す」以外の動詞で、自然科学、技術・工学分野の名詞群との結びつきが強かった。具体例を以下に挙げる。

- ・取り出す：遺伝子、細胞、受精卵、DNA、脊髄、脳、酵素、リンパ球、細菌、微生物、エネルギー、電流、水素、肉、実、内臓、種など

- ・生み出す：原理、細胞、雑種、数学、解、種、重元素、超分子、ワイン、相対性理論、エネルギー、毒素、電力、風味など
- ・作り出す：重力場、物質、細胞、抗体、元素、有機物、エネルギー、化学物質、宇宙、電気、意味、味、環境、脳、音、空間など
- ・飛び出す：電子、中性子、X線光子、アルファ線、電波、弾丸、水など
- ・送り出す：血液、電気、信号など
- ・吐き出す：息、ナトリウム、二酸化炭素、黒煙など
- ・引き出す：(力学的な) 仕事、結論、説明、仮説、動き、動作、味、おいしさ、甘みなど

日本語教育では、通常、複合動詞としての「出す」を導入する場合、開始のアスペクトを表わす「はじめる」の用法と関連させ、統語的複合動詞としての機能に注目することが多いように思われる。「雨が降り出す」「車が走り出す」のようにアスペクトに焦点を当て、前項動詞を入れ換えながら練習することは、学習者に後項動詞「出す」の造語力を理解させるにはとても有効だと考えられる。しかし、本調査結果を見ると、専門日本語教育につなげていく場合、その使用頻度の高さから、学習者が早い時期から語彙的複合動詞に注目するよう指導する必要があると言える。例えば、「遺伝子を取り出す」「心臓が血液を送り出す」「息を吐き出す」「電子が飛び出す」「味を引き出す」などは日本人には耳慣れた表現でも、学習者にとっては一つ一つ具体的な場面とともに学ばないと定着が難しいと考えられるからである。

5.1.2 こむ

「出す」の次に使用頻度が高かった後項動詞は「こむ」である。表4を見るとわかるように、「こむ」を後項動詞とする複合動詞は、「出す」とは傾向が異なり、使用頻度が200語を超す語は2語のみで、100語台が11語、それ以下は徐々に頻度が落ちていく形で、個々の複合動詞の頻度が相対的にばらけているのがわかる。実際に使用割合が50%を超えるのは上位17番目の語のところである。17語は以下の通りである。

- 取り込む、組み込む、入り込む、持ち込む、差し込む、煮込む、書き込む、
- まき込む、飛び込む、落ち込む、吸い込む、染み込む、思い込む、流れ込む、
- 埋め込む、飲み込む、読み込む

この17語の意味特徴を整理する。姫野(1999)は「こむ」の意味を二つに分類し、一つは「内部移動」(主体あるいは対象がある領域の中へ移動すること)、もう一つは「程度進行」(動作・作用の程度が進行すること)と呼び、前者が全体の約8割を占めると述べている。17語中、「程度進行」の意味を持つのは「思い込む」「煮込む」「読み込む」の3語であった。特に「煮込む」は技術・工学分野である調理法で多用されている。この3語以外に「書き込む」も「程度進行」の意味で用いられる可能性を有しているが、本調査では「書き込む」はすべて「メモリーにデータを書き込む」のような「内部移動：対象の移動」の用法のみであった。

「内部移動」を意味する14語の具体例を見ていく。まず圧倒的に頻度が高いのは「取り込む」である。「対象を取って内部に移動させる」という意味の他動詞である。「DNAの一部は染色体に取り込まれ」「オキアミの群れを海水とともに一気に取りこむ」「電子は光のエネルギーを取りこ

んで」「空気を十分に取り込む」「栄養素を体に取り込む」「ファイルを取り込む」など具体物の移動を示す用法が大多数を占める。一方、「公共工事の発注に市場原理的な要素を取り込み」「主人公に近い世代を読者層に取り込み」「研究の成果を産業の方に取り込む」など抽象的な場面でも用いられている。使用頻度2位の「組み込む」も他動詞で「対象の移動」を示す。「温度の上昇に伴う力の変化を理論に組み込む」「無限は数学の構造の中にしっかりと組み込まれ」「遺伝子がヒトゲノムに組み込まれる」のように集合体や組織体の中への移動を表している。特に自然科学分野で名詞「遺伝子」と結合している例が多く見られた。

このほか「対象の移動」を意味する他動詞として「持ち込む」「差し込む」「まき込む」「吸い込む」「埋め込む」「飲み込む」がある。「持ち込む」は、「チラシを営業所に持ち込む」「企業がさまざまな技術を持ち込んで」「日本にマラリアを持ち込む」のように具体物、抽象物ともに「持ち込む」ことはイメージしやすいが、「訴訟に持ち込む」「消費者からの苦情が持ち込まれる」のような用法は日本語学習者にはイメージがつかみにくいと思われる。「差し込む」は「陽/日/光が差し込む」のように自動詞の用法もあるが、本調査では「メモリ/電極を差し込む」のような方法を示す他動詞の用法が多数を占めていた。「まき込む」は、姫野(1999)の指摘にあるように、対象自体の移動は問題にされず、枠組みそのものが動き、結果的に対象を内部にとりこんだ領域を形成するという意味を持つ。「水分が回りの砂を巻き込んで」「ピザにハムを巻き込む」「煙が循環流に巻き込まれる」のように実際に「巻いて中に入れる」状態を説明する例もあるが、「葛藤の渦に巻き込まれる」「民間人が戦争に巻き込まれる」「子どもが犯罪に巻き込まれる」「トラブルに巻き込まれる」のように受身形で用いられ、否定的な意味を表す例が多かった。「吸い込む」は、息、空気、煙などの気体以外に、アスベスト、粉塵、花粉なども吸い込む対象となっている。人がその対象を鼻孔を通じて体内に吸って入れるという意味ではイメージは難しくないだろう。しかし、「拡張期血圧」は、心臓が次の収縮のために血液を吸い込んだ瞬間の血圧である」「玄関の隙間が床下の空気を室内に吸い込んでしまう」「塩は水をどんどん吸い込んでいる」「光を吸い込むバックサテンという素材なら」のように、無情物が受身形ではなく能動形で「吸い込む」場合は、擬人化用法として注意が必要である。「埋め込む」は対象の個体への移動を表す。「無数の地雷が埋め込まれている」「レールを土に埋め込んだ併用軌道」のように実際に土中に具体物を埋めるほか、「体内に埋め込む体内型人工心臓」「インプラントを埋め込む」「HTML メールに埋め込まれた画像」「画像にスタンプを埋め込む」のように、医療技術、IT技術関連で多く用いられている。「飲み込む」は、「食物、水、唾液、消化物」のほか煙、炭粉などを実際に飲んで体内に入れるという意味での用法が多い。また、頻度は少ないが注意を要する用法としては、「消費文化/内紛の渦/無意識の渦にのみ込まれる」のような擬人化用法、「言葉をのみ込む」で言外に抑制を表す慣用的用法、「要点/事情をのみ込む」のように「のみ込む」が「理解する、十分心得る」の意味で用いられる用法である。

次に、自動詞として用いられ、主体の閉じた空間への移動を表すのは「落ち込む」「飛び込む」「入り込む」「差し込む」「流れ込む」である。このうち、「落ち込む」は自己の内部への移動(主体あるいは対象の一部が基底部に向かって沈下する)の意味でも用いられる。本調査結果では、

後者の精神的な作用を表す「(人が) 落ち込む」の方が、前者の「閉じた空間への移動」より使用頻度が約3割多かった。専門日本語教育の立場からは「需要／個人消費／売上高が半分の規模／5分の1に落ち込む」のような表現が重要であろう。「飛び込む」は「～が目／耳に飛び込む」のような表現のほか、自然科学系ジャンルに特徴的な表現として「ナトリウム原子から電子が一つ飛び出して塩素原子に飛び込む」のような表現があり、このような例では「飛び出す」と「飛び込む」を対で導入するのが効果的であろう。「入り込む」は、「化学物質が体内に／がん細胞がリンパ管や血管に／電荷を帯びたイオンが細胞内に／水虫菌は生体の中に／原因物質は肺の中に／宇宙船の中に放射線が／水平尾翼が後方乱流に／塩素原子が有機化合物の構造中に／入り込む」というように現象を説明する際によく用いられている。「差し込む」は前述したように自動詞としては「陽／日／光が差し込む」のように用いられていた。「流れ込む」は血液（血）、水、気体などの流動体と結び付き、「肝臓に流れ込んでいる血管」「糖類は血液へと流れ込み」「大動脈に血が流れ込み」などの生体内の描写によく用いられている。そのほか「雨が周辺の川から流れ込んで」「南からの湿ったあたたかい空気が流れ込んで」のような気象状況、「川は海へ流れ込み」「水は地球の中心に流れ込み」のように地理的な説明にも用いられている。同じく自動詞の「染み込む」は、「雨／雨水が土に染み込む」「繊維に染み込んだ汚れ」のような具体的な用法がほとんどで「歴史／人の意識／考え方が染み込む」のような抽象的な名詞との結合の用法はほとんどなかった。「染み込む」で特徴的なのは、「味が染み込む／味を染み込ませる」という調理に伴う表現で、味以外にも「風味」「香り」が用いられ、使用頻度は全体の約3分の1と高かった。

5.1.3 つける

「つける」を後項動詞とする複合動詞は、使用頻度上位4語の使用が全体の50%を占める。頻度順に「見つける」「取り付ける」「結び付ける」「張り付ける」である。「つける」は語彙的複合動詞と習慣を表す統語的複合動詞を持つが、本調査の上位語には統語的複合動詞は入っていない。

「見つける」は、この一語のみの使用で全体の約25%を占める高頻度複合動詞である。姫野(1999)では「対象の補足」に分類され、感覚動詞「見る」に「つける」がついて何かを認知することを表す。「見つける」と結合する名詞としては「解／方法／原因／因果関係／証拠／法則／薬／化合物／遺伝子／がん」などがあつた。

「取り付ける」は、姫野(1999)が「前項動詞が接辞化して「つける」の本義が最も強く生き、数の上でも多い」と指摘するように、第2位の使用頻度となっている。「取り付ける」の意味特徴は二つに分類でき、一つは「対象への接着・密着」、もう一つは「対人行為接触」である。本調査結果では、ほとんどが前者の意味で、「斜めの面に電気連結器を取り付け」「モーターにリード線を取り付ける」「橋脚の間に水車を取り付ける」「冷却装置を原子炉に取り付ける」のように具体的に「何かをある場所に接着・設置する」という意味で用いられている。特に技術・工学分野に用例が多く、用例数は自然科学分野の約8倍であつた。後者の用法は「諒解／同意／賛成／診察予約／取材／融資を取り付けた」のように用いられ、用例も非常に少ないが、この「取り付ける」は「得ることができる」を意味し、前者とは大きく異なる。指導の際に注意を要する点である。

「結び付ける」「張り付ける」の2語は、「見付ける」「取り付ける」に比べると大幅に頻度は落

ちる。意味特徴は「取り付ける」と同様、「対象への接着・密着」である。「結び付ける」は、「陽子と中性子を結び付ける」「公園や園芸と福祉とを結び付け」「デザインマネジメントを成果に結び付ける」「日常の風景に心象を結び付ける」「パソコン同士を結び付ける」のように用いられている。二つの対象同士の接着に重点が置かれることから、「パソコン同士」を目的語に取ることができる。また、「結び付ける」は、上記の例のように「AとBを」「AをBと」「AをBに」「AにBを」のほか、「AとBとを」とも言えるので、指導の際には助詞の整理が必要である。

「張り付ける」は、「表の中にグラフを張り付ける」「振動吸収版にビロードを貼りつける」「周囲に両面テープをはりつける」のように用いられ、技術的な方法を表している。「張り付ける」も「取り付ける」と同様、技術・工学分野の用例が多く、自然科学分野の約4倍であった。また、IT技術関連でお馴染みのコピーを意味する「貼り付け」の意味で、「書式／画像／値を貼り付ける」のように用いられている例も多かった。

また、使用頻度上位4語に入らなかったが続く5位に「盛り付ける」が入っている。具体例としては「器／皿に盛り付ける」のように調理の最後の段階を描写するものが非常に多いが、それ以外に彫刻の方法で「粘土／石膏を盛り付ける」のように用いられている例もあった。この「盛り付ける」も「張り付ける」「取り付ける」と同様、技術・工学分野の用例が多く、自然科学分野の約7倍強であった。このことから「つける」を後項動詞とする複合動詞の使用頻度上位語は、自然科学系ジャンルの中でも技術・工学分野に特徴的な複合動詞だと言えよう。

姫野 (1999)⁴⁾ は、後項動詞としての「つける」「つく」の両方に共通する前項動詞として、音便形を含めて10語を挙げている。上記3語「取り付ける」「結び付ける」「張り付ける」はいずれもこの10語の中に入り、それぞれ対応語として「取り付く」「結び付く」「張り付く」を持っている。ここで表4の「つく」の項目を見ると、それぞれの語の頻度は「結び付く」が233語で第2位、「取り付く」が50語で第7位、「張り付く」が33語で第10位で、「結び付く」の頻度が際立って高い。用例を見ていくと、「結び付く」は「結び付ける」と違って、自然科学分野の使用が技術・工学分野の約1.7倍であった。具体例としては自然科学分野では「原子と原子が結び付き」「アレルゲンがIgEに結びついて」「猿の体つきは森の生活に結びついて」「温暖化の問題は感染症に直接結びつく」、技術・工学分野では「舞台と映画という2つのメディアが結びつく」「先端技術に結び付くような基礎研究」「カレーと福神漬がかたく結び付いて」のように用いられている。「取り付く」は、対象にしっかり「つく」ことを意味し、「ヤドリバエはじめは寄主の体の一部分に取りついて」「人間に取りついたウィルス」「角質層に水虫菌は取り付き」のように用いられている。「取り付く」が受け身の「取り付かれる」になると、しっかり「つかれる」ことを意味するため、否定的なイメージにつながりやすく、实例でも「ウィルス／魔力／神がかった思想／思い込み／怨みに取りつかれる」のように名詞自体が否定的な意味合いを持つ語と結び付く例の方が、「現代ビジネス社会に取り付かれた大人」のように中立的な名詞と結合している例より多かった。「張り付く」は、「対象への接触・密着」を表し、「コレステロールは血管の内側にははりつき」「クモが足を広げて張り付き」「細胞質は薄い層状になって細胞周辺に張りつき」のように用いられ、自然科学分野でも技術・工学分野でもほぼ同数で用いられていた。

5.1.4 あげる

「あげる」を後項動詞とする複合動詞は、使用頻度上位 5 語による使用で全体の約 52%を占める。頻度順に「取り上げる」「仕上げる」「作り上げる」「引き上げる」「持ち上げる」である。「あげる」を後項動詞とする複合動詞はすべて語彙的複合動詞である。

「取り上げる」は、一語で使用頻度が全体の 19%、約 2 割を占めている。姫野 (1999) によれば、意味特徴は「上昇—全体的上昇—空間的上昇—対象の上昇」に分類される。具体例としては「大腸菌を取り上げることにしよう」「光メモリーへの光可逆反応の応用を取り上げて」「地球の温暖化の問題を取り上げておこななければならない」「国会でも過剰請求が取り上げられ」「三つの要因を取り上げる」のようにテーマ化を行う時に用いられている。論文の読解・作成でも必須の語と言える。

次に頻度が高い使用頻度 2 位の「仕上げる」と 3 位の「作り上げる」は、「完了・完成」の意を表し、人間の作業活動の終了に伴う完成品あるいは動作の完了そのものに重点が置かれる。具体例としては「仕上げる」は、「最終報告書を仕上げる」「低カロリーのサラダを仕上げる」「セーターを仕上げて」「毎週日曜にはミサ曲を仕上げなければ」のように用いられ、「作り上げる」は、「予測方式を過去のデータから作りあげる」「律令制度を作り上げ」「特殊相対性理論をつくりあげるにあたって」「決議案をつくりあげ」「力学の法則をつくり上げた」「良好な人間関係を作りあげる」のように用いられている。「仕上げる」は技術・工学分野で多用され、その使用頻度は自然科学分野の約 7 倍である。一方、「作り上げる」は技術・工学分野と自然科学分野で約半分ずつの使用であった。続く「引き上げる」「持ち上げる」は「取り上げる」と同様、「上昇—全体的上昇—空間的上昇—対象の上昇」の意味特徴を持つ。「注射器の内筒を少し引き上げ」「立方体をシャボン液につけて引き上げると」「ダンベルを引き上げ」「機首を 81 度まで引き上げ」のように具体物を空間的に上昇させる用例が非常に多かった。それ以外では「厚生年金の支給開始年齢を 65 歳に引き上げる」「教える内容を引き上げる」のように対象となる抽象物のレベルを上昇させる例や「賃金／価格／家賃／消費税／水道料金」のような名詞と結び付いて「金額を上昇させる」意味で用いられる例が見られた。指導の際には、単に「上げる」を用いる場合と異なり、「引き上げる」にはある種の抵抗感、負担感を超えて「上げる」という語感が含まれることを説明する必要がある。

5.1.5 あう

「あう」を後項動詞とする複合動詞は、使用頻度上位 9 位 (10 語) による使用が全体の約 53%を占める。使用頻度は「出会う」が 305 語と際立って多く、159 語の「付き合う」、133 語の「話し合う」がそれに続き、4 位の「見合う」は 3 位の「話し合う」の使用数の約半分、100 語を下回り、「似合う」「向き合う」「絡み合う」「重なり合う」と頻度は遞減する。「知り合う」と「立ち会う」は 43 語で同数の 9 位であった。

姫野 (1999) は、「あう」を後項動詞とする複合動詞を「対称関係、相互性」を成立させる統語的複合動詞として、その意味特徴を三つに分類している。一つは、互いを相手として働きかけあう「相互動作・作用」で、例としては「二人が抱き合う」という場合である。二つ目は、同一の

対象を相手とする「共同動作」で、例としては「子供たちが犬を抱き合う」という場合である。三つ目は、同一の場で同じ働きをする「並行動作・作用」で、例として「ねずみがもがき合う」という場合である。この三つの柱の下には動作が行われる時間を基準として「同時」「交互」「同時・交互」の下位分類があり、さらにその中を意味特徴によって分類している。ここでは実際の具体例をこの三つの分類に従って見ていきたい。

使用頻度第1位の「出会う」は、4節で述べたように「出合う」「出逢う」「出会う」「出あう」のすべての代表項目として挙げている。意味特徴は「相互動作・作用」の「同時に起こること―遭遇」で、一語化した無意志動詞として働く。結合する名詞の具体例を見ると「表現、マイナスの反応、複雑な事情、怪物、守護神、障害、鳥、不幸なケース、虫、いろいろな人、道、アクシデント、昔の恋人、場面、先生、宇宙人、胃酸、被害者、A型やB型の赤血球、光景、健康食品、解剖学、記述、細胞、流氷、電子、メス、文明、老木、母親」など、遭遇できる様々な語と結びついている。9位の「立ち会う」もこのグループに入る。具体例としては「彼らの死／お産／最期／診察／誕生／監査／契約／トラブルの現場／に立ち会う」のように、助詞「に」を取り「(人)が～に立ち会う」の形で用いられる。

2位の「付き合う」は、意味特徴は「相互動作・作用」の「精神的接触―交際・交流」である。同じグループに9位の「知り合う」も入る。まず「付き合う」について、「付き合う」相手は通常、人であるが、人以外の例として「ストレス、病気、リュウマチ、危険因子、五十肩、本やインターネット、動物、馬、ゾウ、キリン、イタチ、核分裂、摂食障害、再発ガン、筋肉痛、高尿酸血症、自分の中の細胞、からだ、うつ病、フェラーリ、病、プルトニウム、自然環境、車、風」などと結びついていて、相対的に「病気」の類と「付き合う」例が非常に多いことがわかる。また「付き合う」の特徴として「AがBと付き合う」の形ではなく「AがBに付き合う」形になると、対象との関係が対等ではなくなるため、「不本意ながら」の意味が生じると考えられる。具体例の中では「会話、寝つきの悪いムスメ、父の晩酌、買い物」のような語との結びつきが見られた。一方、「知り合う」は人のみと結び付いていて、「(人)と知り合う」の形で用いられ、「お互いを知り合う」という例も1例あった。

3位の「話し合う」は、意味特徴は「相互動作・作用」の「同時に、交互に起こること―社会的な働きかけ」である。「話し合う」と同様、社会的な働きかけの意味を持つ語にはほかに「語り合う」「助け合う」なども含まれるが、本調査では「話し合う」の頻度が圧倒的に多かった。第4位の「見合う」と5位の「似合う」は、意味特徴は「相互動作・作用」の「同時に起こること―接触―関係（ものごとの抽象的なかわりあい方）―バランス」を表わす。「見合う」の使用傾向には特徴があり、73例中71例が「AがBに見合う」の形で用いられていた。具体例としては「活動に見合った食事量」「技術革新に見合った法律」「人口の増加に見合う形で」「人間の感性にみあった意欲的な再開発」「最低一万円の代金に見合う演技」「診療報酬に見合っていない」「販売コストに見合わない」がある。一方、「AがBと見合う」の形で用いられているのは以下の2例のみであった。

- ・イギリスにおけるPCBおよびダイオキシンの食事摂取は他国と見合っており、心配には及ばない。

・米国の大学や国の試験研究機関をオープンにするのと見合って日本の民間の研究所をオープンにしるというも…。

「似合う」は、「AがBに似合う」の形と「AはBが似合う」の形で用いられていた。前者の具体例としては「この丸い服は彼女に似合っている」「瑞々しいカキにはキリリとした辛口の白がよく似合う」「優しい笑顔に似合う花モチーフのピアス」「自分が着たいと思った服が似合わない」「フランスは君に似合わない」「長い脚にスニーカーがよく似合う」「巨大な体に似合わない素早さ」があり、後者の例としては、「トシ子さんはタバコが似合う」「さやかはショートカットが実によく似合う」「ベリー柄のお皿はフランスの大地や太陽が似合う」「片口の酒器は秋口が似合う」「ハワイアンシャツが似合うヤツ」がある。後者の「AはBが似合う」の原形は「A(に)はBが似合う」だと考えられるが、格助詞「に」が落ちる例が多いことから、指導の際には「AはBが似合う」の形での練習が必要であろう。

6位から8位までの「向き合う」「絡み合う」「重なり合う」は、意味特徴としては基本はやはり「相互動作・作用」である。「向き合う」は「物理的接触(物理的な触れ合いに伴って生じる事柄)一隣接」を表す。具体例としては「病気/大自然/患者/根本的な原因/冷戦の時代/カメラ/に向き合う」「死/子どもたち/自らの疾病/職場の不正/と向き合う」のように用いられ、助詞は「に/と」の両方を取る。「絡み合う」は「物理的接触—接触」を表す。具体的には「低音部と高音部がからみあって」「新しい脳と古い脳がからみあったから」のように「AとBが絡み合う」の形で用いられるが、実際には「AとB」のように絡み合う主体を個々に取り上げずに「複雑に因果が絡み合った現象」「複数の因子が絡み合う」「二人の視線が絡み合い」「三つ又のトゲがからみあって」「さまざまな現象が複雑にからみあって」のように名詞自体が複数の場合も多かった。また「絡み合う」の特徴として副詞「複雑に」と結びつく例が多数あり、58例中20例に上っている。「重なり合う」は意味特徴としては「物理的接触—重複」である。具体例としては「聖と俗が重なりあって」「体の内部と外部の両方の危険因子が重なり合って」のように「AとBが重なり合う」の形で用いられるが、実際には「絡み合う」と同様、結合する名詞自体が複数を表わす例がほとんどであった。名詞と結びついて複数を表わすマークとなる「同士」「互いに」が用いられている用例が多く、具体的には「並行する紐同士が重なり合わないように」「金属原子の最外殻同士と重なり合う」「銀河どうしは互いに重なり合い」「互いに重なり合う配列を持つクローン」のように用いられている。「同士」「互いに」以外で複数を表わすマークとして働いている語としては「幾重にも」「いくつも」「いくつかのN」「多数のN」「両方のN」などがあつた。「重なり合う」を導入するときにはこうした複数を表わす語のバリエーションとともに練習を行うのが有効であろう。

以上、後項動詞「あう」から構成される複合動詞の中で頻度の高い10語については、意味特徴はすべて「相互動作・作用」に分類されるもののみで、有情物の意志的行為に限られる「共同動作」と「並行動作・作用」の意味特徴を持つ語は入っていなかった。

5.1.6 はじめる、つづける、おわる、すぎる

最後に「はじめる」「つづける」「おわる」「すぎる」の後項動詞としての使用傾向について少

し触れておく。日本語教育では、4級語彙「おわる」、3級語彙「はじめる」「つづける」「すぎる」は初級レベルで導入され、複合動詞としての働きについては、通常、初級後半～中級レベルで取り上げられる項目である。「はじめる」「つづける」は結びつく動詞が多いため、指導の際には統語的な働きの説明が中心になりやすいと言えよう。例えば「はじめる」の場合、初級レベルの学習者が統語的機能を理解し、ランダムに既習動詞と結び付けて「食べ始める」と同様、「寝始める」「乗り始める」「勉強をし始める」という文を作ったとしても、既習語彙が限られているため、語の使用環境について制限があることに言及するだけに留めざるを得ないであろう。「勉強をし始める」についても「勉強を始める」と言う方が自然であることは、教える側は知っていると考えられるが、初級レベルでは学習者の作例が文法的に誤りでない限り、そのままになりやすく、結果としてその不自然さが上級レベルまで持ち越されることもある。その意味でも、指導の際に、実際の使用頻度についての情報が含まれた用例集を資料として持つことは重要であろう。

表3を見るとわかるように、アスペクトを表わす「はじめる」「つづける」と過剰・過度を表わす「すぎる」は動詞との結合の自由度から異なり語数は非常に多い。一方、「おわる」は意志的行為の終了のみにしか使えないため異なり語数は少ない。表4でそれぞれ見ていくと、「はじめる」を後項動詞とする複合動詞が一番多く1319語で、次が1087語の「つづける」、その次が797語の「すぎる」で、最後に予想通り123語の「終わる」である。各後項動詞の中で頻度が高い複合動詞を見ていく。

「はじめる」は、第1位が「出始める」、第2位が「飲み始める」で使用頻度は40語台であった。「出始める」については、「症状／黄疸／湿疹／母乳／尿／痛み／血の塊／白煙／雄成蜂／油分／が出始める」のような具体的な名詞と結び付く例のほか、「疑問の声／チーム医療の動き／新しい美術史の研究／影響／噂／兆候／需要／が出始める」のような抽象名詞と結び付く例も多かった。「飲み始める」は酒類にも使われているが、健康食品、野草茶、免疫乳酸酵素、ビタミンB1、薬など健康関係の話題で多く用いられ、相対的に自然科学分野で多かった。また、上述した「勉強をし始める」の例は予想通り少なく1319例中14例であった。そのうち、「漢語＋を＋する動詞」と結び付く用例の動詞としては、「準備をする」「整理をする」「算段をする」「肺呼吸をする」の4例で、それ以外の例は「むき出しにする」「耳にする」「気がする」「ツンツンとする」「暮らしをする」「歩いたりする」「恋をする」のように、慣用表現のほか、和語、副詞、「～たりする」との結び付きであった。このように本調査では「漢語＋を＋する動詞」と「はじめる」が結合し、複合動詞となる例は非常に少なかった。

「つづける」は、40語台以上は頻度順に「飲み続ける」「増え続ける」「生き続ける」「持ち続ける」であった。「飲み続ける」と結び付く語は、酒類のほか「湧水／抗がん剤／免疫抑制剤／ピル／薬剤／ダイエット食品／サプリメント／薬剤／胃散／健康茶」などで、「飲み始める」と結合している語と重なっているものが多かった。「増え続ける」と結び付く語には、「人口／感染者／大腸ガン／アレルギー性の疾患／生活習慣病／自動車の数／地震の数／化学物質／喘息／高齢者／ウィルス／自動車／排出量／署名」などの具体的名詞のほか、「需要／投資／モノ」などの抽象的な名詞があった。「生き続ける」と結び付く語は、具体的な名詞としては「DNA／細胞／カプト

ガニ／ミミズ／ヒト／人間／東京駅／生き物／お菓子／辞書」などで、抽象的な名詞としては「自然や風土／視座／先人の知恵／多様な「農」／宗教」などがあつた。「持ち続ける」は、「種子／土地／発電設備」などの具体的名詞との結合も見られるが、ほとんどが「夢／気持ち／興味／関心／不安／疑い／疑問／権力／こだわり／影響力／願い／憧憬／夫婦の時間／人気／情熱／愛着／感謝／好奇心／姿勢／能力／意欲／価値／信念／優位性／熱意／自負／目標」のような抽象名詞と結合していた。

「すぎる」については、頻度順に「取り過ぎる」「食べ過ぎる」「なり過ぎる」である。実際の用法では形容詞・形容動詞と結びつく例が非常に多く、「なり過ぎる」についても「熱く／複雑に／なり過ぎる」のように形容詞、形容動詞と結合して変化を表す例が 55 例中 49 例であつた。「取り過ぎる」は、「甘み／カロリー／塩分／動物性蛋白質／砂糖／糖分／コレステロール／脂肪／栄養／水分／油／食塩／にがり／ナトリウム／ビタミン C／肉／夜食や間食」などと結合し、摂取の意味で用いられている例がほとんどである。それ以外の例では「場所／スペース／動物／肩のこり／年」の例があつた。

「おわる」については、意志的行為の終了のみを表すという制約があることから延べ語数自体が非常に少なく、その中で頻度第 1 位は 123 語中 21 語を占める「食べ終わる」、第 2 位は 13 語の「使い終わる」で、「使い終わる」と結合している名詞は「注射器／水／パソコン／顆粒タイプのだし」などであつた。

6. おわりに

本研究では、『現代日本語書き言葉均衡コーパス』（書籍コーパス）の自然科学系ジャンル（自然科学と技術・工学）の文章を対象に、複合動詞の後項動詞となる 26 動詞を検索した結果、延べ語数 29254 語、異なり語数 2694 語の複合動詞が抽出され、その動詞としての用法について使用傾向を調べた。その結果、後項動詞「だす」「こむ」「つける」「あげる」「あう」によって構成される複合動詞の使用が全体の 5 割以上を占めていることがわかつた。また、「はじめる」「つづける」「おわる」「すぎる」を除く 22 の各動詞について見ると、使用頻度の高い上位 2~4 語の使用が各複合動詞の総数の半分以上を占めるものが 11 動詞に上り、非常に限られた数の複合動詞が繰り返し用いられていることが明らかになった。今後の課題としては、自然科学系ジャンル以外の文章における後項動詞の使用傾向を調べ、本調査結果と合わせて実証分析につなげたいと思う。

最後に専門日本語教育への応用について触れておきたい。各文章ジャンルにおいて高い造語力を持つ個々の後項動詞がどのような複合動詞を造り、それがどのぐらいの頻度で用いられているのかというデータは、文章指導のための基礎資料として非常に重要だと言える。語彙教育の効率化を図るためには、造語力のある頻度の高い後項動詞を中心に、語彙を増強するのが有効であろう。本調査結果で明らかになったように、自然科学系ジャンルの書籍では、頻度が高い複合動詞は比較的限られた数の語彙的複合動詞で、その動詞が繰り返し用いられている。学習者がそのような頻度の高い複合動詞を先に学ぶことによって多用される動詞の存在を知り、複合動詞の学習が無限ではなく、ある程度の傾向を持っていることを知ることになれば、学習意欲も継続できる

であろう。また、専門分野では固有の名詞との結びつきにも注目し、それらを一つの表現として捉え、実際の使用環境に即した練習を行うことも重要である。さらに接頭辞「取り」と結合した「取り出す／取り込む／取り上げる」などはどれも頻度が非常に高い語なので、まとめて取り上げ、その意味を確認する方法も有効だと考えられる。

付 記

本研究は国立国語研究所の共同研究プロジェクト、「テキストにおける語彙の分布と文章構造」による研究成果の一部である。

注

- 1) 『大辞林』第三版三省堂、『言泉』第一版小学館。
- 2) 姫野 (1999) によると、「創出」は人が何らかの手段で無の状態から対象を生じせしめ、前項動詞はその方法を示し、創作活動、加工作業に関するものが主である。
- 3) 姫野 (1999) によると、「顕現」は対象がもともとそこにあつて、人の知覚に触れないでいたものが、変化が加わり、見えたり聞こえたりするようになって存在が明らかになることである。
- 4) 姫野 (1999) p. 119 参照。

参 考 文 献

- 斎藤倫明 (1992) 『現代日本語の語構成論的研究—語における形と意味—』 ひつじ書房
姫野昌子 (1999) 『複合動詞の構造と意味用法』 ひつじ書房
前川喜久雄 (2008) 「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発」, 日本語の研究, 4(1), pp. 82-95.
村田 年 (2008) 「文章と複合動詞—論述文ジャンルを特徴づける新たな指標を探して—」『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター36号 pp. 1-33.
前川喜久雄 (2009) 「代表性を有する大規模日本語書き言葉コーパスの構築」 人口知能学会誌, 24(5), pp. 616-622.
山崎 誠 (2009) 「代表性を有する現代日本語書籍コーパスの構築」 人口知能学会誌, 24(5), pp. 623-631.

関連 URL

- 「中納言」 <https://chunagon.ninjal.ac.jp/>
「KOTONOHA 国立国語研究所言語コーパス整備計画」
<http://www.ninjal.ac.jp/kotonoha/>

共起語率の分布からみるテキストの語彙的特徴¹

山崎 誠 (国立国語研究所言語資源研究系)

Lexical Characteristics of Text as Seen in the Distribution of Co-occurrence Rate

Makoto Yamazaki (Dept. Corpus Studies, NINJAL)

1. はじめに

「現代日本語書き言葉均衡コーパス」(Balanced Corpus of Contemporary Written Japanese、以下 BCCWJ と略す)が 2011 年に完成し、それを利用した日本語研究のさまざまな展開が期待されている。BCCWJ の特徴として、多様な日本語を収録していることやアノテーションの充実が挙げられる。それらを生かした研究が今後発多く発表されることと思われる。本発表では BCCWJ のアノテーション情報を利用してテキストの結束性に関する特徴を捉える試みを紹介する。

2. テキストにおける結束性

結束性 (cohesion) とは、文章をひとつの統一体としてまとめあげるために必要な性質のひとつとされる。結束性について最初に詳細に研究を行ったのは Halliday & Hasan(1976)である。それによると、結束性について次のように紹介されている。

「結束性が生じるのは、談話のある要素の解釈 (INTERPRITATION) が別の要素の解釈に依存する場合である。一方を効果的に解釈するためには他方に頼らなければならないという意味で、一方は他方を前提 (PRESUPPOSE) とする。こういうことが生じるとき、結束関係が成立する。その結果、前提語と被前提語という 2つの要素が、少なくとも潜在的には、統合されて 1つのテキストになるのである。」(邦訳 p.5)

庵(2007:12)によれば、結束性は推論にもとづくつながりである一貫性(coherence)の下位概念であるとされる。また、結束性には文法的結束性と語彙的結束性とがあり、前者の手段として「指示」「代用」「省略」が、後者には「再叙 (reiteration)」と「コロケーション」がある²。再叙には以下の4つのタイプがある。

- (a)同一語 (繰り返し)
- (b)同義語 (または近似同義語)
- (c)上位語
- (d)一般語

Károly(2002:162)によれば、英語の作文においては、(a)の同一語の繰り返しよりは(b)~(d)を合わせた「異なる語の繰り返し」の方が多く用いられるということだが、同義語 (類義語) や上位

¹ 本稿は第1回コーパス日本語学ワークショップ (2011年3月) で発表したものである。

² Halliday & Hasan(1976)では、文法的結束性と語彙的結束性の中間の性質を持つものとして「接続」が挙げられている。

語の判断を自動的に行うことが難しいため、本発表では(a)の同一語の繰り返しのみを観察対象とする。同一語の繰り返しは、本発表で用いた図書館書籍のデータでは、10,369 サンプル中同一語の繰り返し³が無かったサンプルは17個しかなかった。それらはいずれも延べ語数22語以下の小さなサンプルで、サンプルの短さがその原因である。ある程度の長さを持つテキストには必ず同一語の繰り返しがあると見てよいだろう。

3. データ

本発表では、2011年12月にリリースされた『現代日本語書き言葉均衡コーパス』のDVD版を使用した。Disk1のM-XMLフォルダに含まれるxmlファイルが対象である。このxmlファイルは可変長サンプルと固定長サンプルを統合したもので、短単位、長単位の形態論情報のタグのほか可変長部分には文章構造のタグを含んでいる⁴。

本発表ではこのxmlファイルにおいて<paragraph>というタグが付与された部分を対象にそこに含まれる短単位の形態論情報をもとに分析を行う。結束性を観察するには文も妥当な単位であるが、BCCWJに付与された文を表すタグ<sentence>は見出しや図表のキャプションにも付与されており、通常の本文との区別をしなければならないため、今回の調査では確実に本文部分を表している<paragraph>タグを対象とした。<paragraph>タグを含むサンプル数は表1のとおりである。

表1 対象サンプル数

媒体	全サンプル数	P サンプル数
出版書籍	10,117	9,742
雑誌	1,996	1,767
新聞	1,473	1,457
図書館書籍	10,551	10,369
白書	1,500	1,496
教科書	412	0
広報紙	354	354
ベストセラー	1,390	1,374
Yahoo!知恵袋	91,445	0
Yahoo!ブログ	52,680	0
韻文	252	0
法律	346	56
国会会議録	159	159
合計	172,675	26,774

³ ここでは同一語の繰り返しには、助詞・助動詞は含めていない。以下も同様。

⁴ タグの詳細については小木曾ほか(2011)を参照。

教科書、Yahoo!知恵袋、Yahoo!ブログ、韻文は<paragraph>タグを用いていないため、対象サンプル数はゼロである。なお、<paragraph>タグの問題点については西部ほか(2011:232)を参照されたい。

表2は、対象となったサンプルの延べ語数、段落数、1段落あたりの延べ語数、1段落あたりの異なり語数のそれぞれの平均値である。1段落あたりの延べ語数を見てみると国会会議録の値が大きい。これは国会会議録における段落の認定（1発言が1段落）が影響しているものである。なお、語数には補助記号、空白、助詞、助動詞は含まれていない。

表2 各媒体の延べ語数等の平均値

	サンプルの延べ語数	段落数	1段落の延べ語数	1段落の異なり語数
出版書籍	1,384.61	43.76	50.51	37.06
雑誌	891.17	29.81	40.05	33.27
新聞	334.33	9.28	38.78	33.33
図書館書籍	1,450.16	54.53	45.76	34.70
白書	1,793.10	29.32	64.74	44.33
広報紙	2,903.53	103.14	28.14	23.39
ベストセラー	1,404.46	69.30	29.52	24.28
法律	219.50	6.93	24.04	15.03
国会会議録	17,885.87	144.06	151.30	76.21

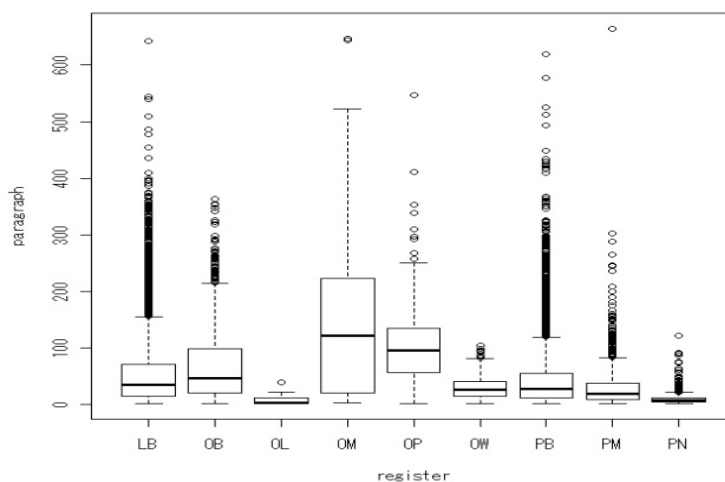


図1 段落数の分布

図1は、サンプルあたりの段落数の分布の様子を媒体ごとに表したものである。全体的に分布が右に（大きい方に）かたよっていることが分かる。また、図書館書籍と出版書籍はほぼ似たような分布を示している。

4. 結束性の算出方法

本発表では、ある段落とそれに隣接する段落との間で共通して現れる語の多寡に着目した。語の単純な繰り返しを扱うことのメリットは、他の結束性を表す現象と比べて正確な把握がしやすいこと、また、頻繁に起きる現象であるため、観察がしやすいことである。一方、デメリットとしては観察結果が「語」の単位認定基準に依拠してしまうこと及び同じ語か異なる語かだけの把握にとどまり、意味的な関係が把握できないことである。共通する語だけでなく、類義語等まで含めた計測方法として Hoey(1991)や Károly(2002)があるが、扱っているデータ量はさほど多くない。大量のデータを使って自動的に計測するには語の繰り返しがもっとも適していると思われる。

本発表では、以下の式により結束性の度合いを計り、共起語率と名付けた。

$$C(a, b) = \frac{F(a, b)}{N_a}$$

a, b : 段落番号(1~n)

C(a, b) : 段落 a の段落 b に対する共起語率。

F(a, b) : 段落 a と段落 b とで共通して現れる語の延べ語数を段落 a 内で数えた数。

N_a : 段落 a の延べ語数。

共起語率は、水谷(1980)の非対称類似度を利用した指標である。そのため、連続する 2 つの段落の間の共起語率に 2 つの値が存在する。後続の段落に対する共起語率と前接の段落に対する共起語率である。上述の式では、 $b=a+1$ のとき、後続段落に対する共起語率となり、 $b=a-1$ のとき、前節段落に対する共起語率となる。ただし、文章の冒頭の段落の前接段落及び最後の段落の後続段落は存在しないため、便宜的にその場合の共起語率は 0 とする。

この方法で共起語率を測るにはひとつ制約がある。それは、文章が 2 つ以上の段落から構成されていなければならないことである。そのため、表 1 で対象としたサンプルから 1 段落しかなかったサンプル 340 サンプルを除外した。

なお、計測対象からは言語表現とは見なさない補助記号、空白、及び文章の結束性には影響を及ぼさない助詞、助動詞を除外した。

5. 結果

表 3 は、段落あたりの共起語の数と共起語率の平均値である。後続段落との共起語率と前接段落との共起語率とはほぼ等しい値を示している。このことは、どの媒体もそれぞれ同程度の依存

関係でつながっていると解釈できる。個々に眺めてみると、法律、白書、国会会議録の共起語率が高く、新聞、ベストセラー、雑誌の共起語率が低いことが分かる。

表 3 共起語の数と共起語率

	後続段落との共起語数	後続段落との共起語率	前接段落との共起語数	前接段落との共起語率
出版書籍	12.98	0.22	12.74	0.22
雑誌	6.89	0.16	6.82	0.16
新聞	5.99	0.15	5.84	0.16
図書館書籍	10.49	0.19	10.36	0.19
白書	20.00	0.31	19.84	0.31
広報紙	5.19	0.18	5.13	0.17
ベストセラー	5.49	0.15	5.47	0.15
法律	12.16	0.48	12.31	0.47
国会会議録	40.45	0.30	39.01	0.30

表 4 NDC 別の共起語の数と共起語率

	後続段落との共起語数	後続段落との共起語率	前節段落との共起語数	前節段落との共起語率
0 総記	12.97	0.22	12.95	0.22
1 哲学	17.55	0.25	17.73	0.24
2 歴史	14.80	0.21	14.60	0.21
3 社会科学	15.02	0.24	14.84	0.24
4 自然科学	14.32	0.24	13.96	0.24
5 技術・工学	10.72	0.22	10.56	0.21
6 産業	11.03	0.21	10.82	0.21
7 芸術・美術	12.02	0.20	11.98	0.20
8 言語	10.40	0.21	10.17	0.20
9 文学	5.07	0.12	4.97	0.12
分類なし	3.46	0.13	3.45	0.13

表 4 は、図書館書籍のデータについて、NDC（日本十進分類法）別の共起語数と共起語率を算出したものである。図書館書籍全体では共起語率は 0.19 であったが、NDC 別に見ると「9 文学」

と「分類なし」の値が他と比べて低いことが分かる。「分類なし」についてはデータを見ていないので理由は分からないが、「9 文学」は会話文のような短い段落が多いため、共起語率が低くなつたと推測される（表 3 のベストセラーの値の低さもそれに起因しているであろう）。それを確かめるために、1 段落あたりの延べ語数の平均と共起語率の平均との相関を見てみよう。図 2 にその結果を示す。正の相関が認められ、決定係数は 0.799 と高い値を示した。

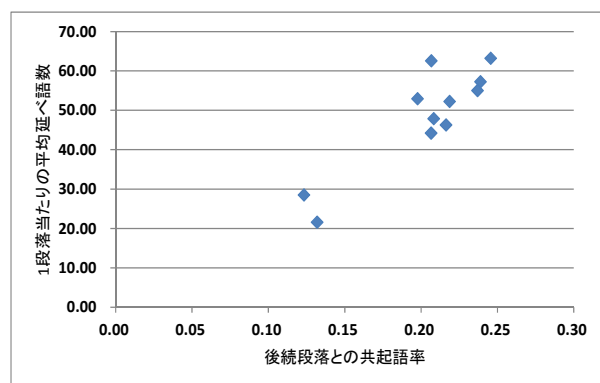


図 2 段落の延べ語数と共起語率との相関

6. 文章中の共起語率の推移

共起語率の値はひとつの文章中でどのような変化を示すのだろうか。白書の例を見てみよう。

図 3 は OW1X_00000（昭和 54 年版経済白書）というサンプルである。

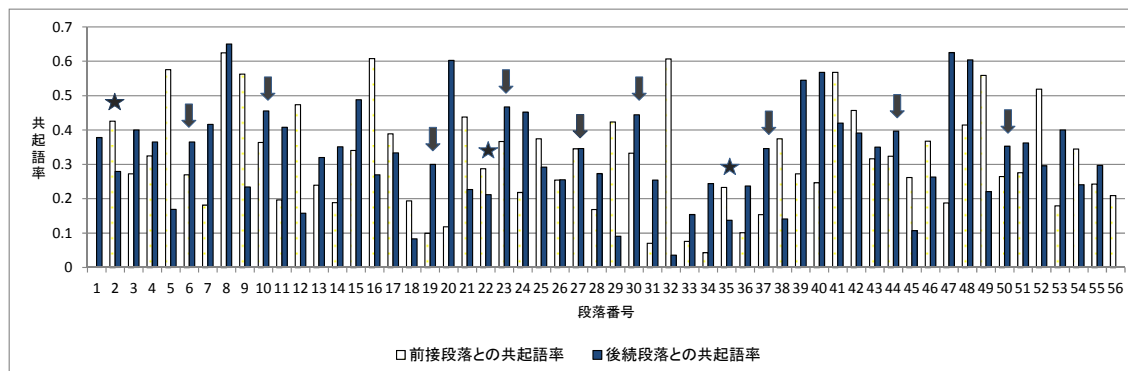


図 3 文章中の共起語率の推移

図 3 で、★を付けた 3 箇所は大きな節が開始する箇所、下向きの矢印を付した 9 箇所はその節の中で小見出しが立っている箇所である。矢印の部分における後続段落との共起語率（右側の棒）と前接段落との共起語率（左側の棒）とを比べてみると、9 箇所のうち 8 箇所が後続段落との共

起語率が前接段落との共起語率を上回っている（残りの1箇所は同じ値）。このことは、新規の内容になった最初の段落は、新しい話題を展開させるため、その次の段落との結束性が高くなっていると言えるのではないだろうか。

逆に矢印の直前の段落は、あるまとまりの最後の段落を意味する。この部分の後続段落と前接段落の共起語率はどうなっているかという、9箇所中6箇所前接段落との共起語率の値のほうが高い。これは一つの例にすぎないが、このような文章中での共起語率の推移を利用して段落のまとまりを自動的に推測することに応用出来る可能性がある。

7. まとめと今後の課題

本発表では非常に単純な指標である共起語率を用いて文章の結束性の度合いを観察した。その結果、法律、白書、国会会議録のように結束性の高い文章と新聞、ベストセラー、雑誌のように結束性の低い文章があることが分かった。NDC別に観察したデータでは、文学の結束性が低いという結果になった。これは文学に会話文が多く、その会話が1段落と認定されているというデータの特徴の現れである。

また、文章中の共起語率の推移をみることにより文章のセグメンテーションへの応用が考えられることを示した。

今後の課題として以下の3点を挙げる。これらを通じて文章における結束性について客観的な記述を目指したい。

(1)西部ほか(2011:232)によると、サンプルを構成する文がすべて段落に分割される訳でない指摘されている。また、<paragraph>の認定は行頭の空白をもとに自動的に認定しているとのことなので段落の実態を確認して分析に問題がないかどうか確認する必要がある。

(2)段落と文の両方を利用した結束性の測定の方法を探る。

(3)指示詞や接続詞など文法的結束性的手段との相関を調べること。

謝 辞

本研究は国立国語研究所の共同研究プロジェクト「テキストにおける語彙の分布と文章構造」による研究成果の一部である。データとして利用した BCCCWJ の書籍部分は、文部科学省科学研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」（平成18～22年度、領域代表者：前川喜久雄）による補助を得て構築したものである。

参考文献

Halliday, M.A.K. and Hasan, R.(1976) *Cohesion in English*. Longman (邦訳『テキストはどのように構成されるか』、大修館書店、1997刊)

Hoey, Michael.(1991) *Patterns of Lexis in Text*. Oxford University Press.

Károlyi, Krisztina.(2002) *Lexical Repetition in Text*. Peter Lang.

庵功雄(2007)『日本語におけるテキストの結束性の研究』、くろしお出版

小木曾智信、間淵洋子、前川喜久雄(2011)『『現代日本語書き言葉均衡コーパス』における形態論情報付き XML フォーマット』、言語処理学会第 17 回年次大会予稿集、pp.352-355.

西部みちる、大島一、間淵洋子、小林正行、田島孝治、高田智和、山口昌也(2011)『『現代日本語書き言葉均衡コーパス』における電子化テキストの構築』、国立国語研究所内部報告書 (LR-CCG-10-03)

水谷静夫(1980)「用語類似度による歌謡曲仕分『湯の町エレジー』『上海帰りのリル』及びその周辺」『計量国語学』12(4)、pp.145-161.

段落間の類似度を利用したテキストの結束性の測定¹

山崎 誠 (国立国語研究所言語資源研究系)

Measurement of Textual Cohesion Using Similarity between Paragraphs

Makoto Yamazaki (Dept. Corpus Studies, NINJAL)

1. はじめに

テキストの結束性 (cohesion) は、一貫性 (coherence) とともにテキストの基本的な性質とされる重要な概念である。とくに結束性は Halliday&Hasan (1976) 以来、多くの研究が行われている。本稿は、テキストを構成する段落間の類似度を使ってテキストの結束性を計量的に明らかにしようとするものである。

2. テキストにおける結束性とその現れ方

結束性とは、文章をひとつの統一体としてまとめあげるために必要な性質のひとつとされる。結束性について最初に詳細に研究を行ったのは Halliday&Hasan(1976)である。それによると、結束性について次のように紹介されている。

「結束性が生じるのは、談話のある要素の解釈 (INTERPRITATION) が別の要素の解釈に依存する場合である。一方を効果的に解釈するためには他方に頼らなければならないという意味で、一方は他方を前提 (PRESUPPOSE) とする。こういうことが生じるとき、結束関係が成立する。その結果、前提語と被前提語という 2つの要素が、少なくとも潜在的には、統合されて1つのテキストになるのである。」 (邦訳 p.5)

また、結束性には文法的結束性と語彙的結束性とがあり、前者の手段として「指示」「代用」「省略」「接続」が、後者には「再叙 (reiteration)」と「コロケーション」がある。再叙には以下の4つのタイプがある。

- (a)同一語 (繰り返し)
- (b)同義語 (または近似同義語)
- (c)上位語
- (d)一般語

平林 (2003:72) によれば、日本人高校生の英作文 (1 作文あたり平均約 100 語、10 文) の分析から 1 作文あたり平均 12.87 個の結束数が現れ、その内訳は、指示 4.9 個、接続 3.62 個、語彙的結束性 3.36 個、代用 1.00 個、省略 0 個だという²。

Károly (2002:162) では、英語の作文においては、(a)の同一語の繰り返しよりは(b)~(d)を合わせた「異なる語の繰り返し」の方が多く用いられるという報告がある。しかし、本稿では、同義語 (類義語) や上位語の判断を自動的に行うことが難しいため、(a)の同一語の繰り返しのみを観察対象とする。

¹ 本稿は第 2 回コーパス日本語学ワークショップ (2012 年 9 月) で発表したものである。

² 作文における結束性の割合はひとつの目安となるが、今回例として取り上げた白書のデータでは指示(1 個)、接続(1 個)よりも語彙的結束性(181 回:繰り返し使用された語における繰り返しの数) のほうが多く、やや異なる出現状況であった。

3. データと方法

3.1 データ

本発表では、2011年12月にリリースされた『現代日本語書き言葉均衡コーパス』のDVD版を使用した。Disk1のCORE/M-XMLフォルダに含まれる白書のxmlファイル62個が対象である。これらのxmlファイルは可変長サンプルと固定長サンプルを統合したもので、短単位、長単位の形態論情報のタグのほか可変長部分には文章構造のタグを含んでいる³。本稿では結束性を適切に観察するため、文章構造のタグを含まない固定長部分を対象外とし、可変長部分のみを用いる。

対象となるデータには文章構造のタグとして<paragraph>が使われており、このタグで囲まれた部分を一つの段落としてテキストの構成を考えることにする。なお、<title>のタグで囲まれた見出し部分や注記、図表のキャプションなどは対象外とした。

3.2 結束性の測定方法

結束性の測定方法は2節で挙げた語彙的結束性のタイプのうち(a)の同一語の繰り返しを利用する。各段落をひとつの語彙とみなし、語彙の類似度を利用して、各段落間の関連を観察する。

語彙の類似度を表す主な指標にはC(宮島(1970))とD(水谷(1980))とがあるが、本稿ではD(以降、単に「類似度」と言う)を用いる。この類似度は、非対称的であることに特徴があり、語彙aの語彙bに対する類似度と語彙bの語彙aに対する類似度がそれぞれ別の値をとることができる。見方を変えれば語彙の「依存度」(水谷(1980:147))と考えることもできるものであり、順を追って構成されるテキストの内容の関係をさぐる上で適切な指標となるものである。語彙aの語彙bに対する類似度は次の式で表される⁴。

$$Da|b \stackrel{\text{def}}{=} \sum_{M \in V_a \cap V_b} P_a(M) = \frac{1}{N_a} \sum_{M \in V_b} F_a(M)$$

V_x : 表現域 x の上の語彙

$P_x(M)$: 見出し語 M の表現域 x における使用率

N_x : 表現域 x の延べ語数

$F_x(M)$: 表現域 x での見出し語 M の使用度数

類似度の測定にあたっては、短単位を用い、品詞が空白、補助記号、助詞・助動詞であるものを除外した。これらは結束性を表しているわけではないからである。同様に、結束性への貢献が低い語群についても除外した。この語群の候補として田中(1973)の「無性格語」を利用した。無性格語とは、田中によれば「これらの単語は、どんな文章にも現われるようなものであって、ある特定の文章や文献の性格とか特徴とかを反映することは、ほとんどない。いわば無性格な語群であろう。」(田中(1973:157))とされるものである。田中(1973)では108語が無性格語としたリストアップされている。本稿ではこの無性格語を短単位での実現形に合わせて適宜修正して用いた⁵。また、田中(1973)の趣旨を汲み、リストに上がっていない数詞についても無性格語として処理した。

³ タグの詳細については小木曾ほか(2011)を参照。

⁴ 式は水谷(1983)より引用。

⁵ 本稿で用いた無性格語のリストを付表に掲げた。

4. 分析例

4.1 データ

表1は『平成16年度文部科学白書』（サンプルID:OW6X_00000）の可変長部分を段落に分けて示したものである⁶。見出し部分は本文とは別立てにして、その見出しが及ぶ範囲が明らかになるように示した。このサンプルは、接続詞が第5段落の「さらに」の1つのみ、指示詞が第3段落の「これ（まで）」の1つのみであり、文法的結束性が少ないという特徴を持っている。

表1 白書データ（OW6X_00000）の構造

見出し1	見出し2	見出し3	段落番号	テキスト
1 日本文化の発信による国際文化交流の推進	(1) 文化庁文化交流使事業	① 文化庁文化交流使事業	P1	文化庁文化交流使事業は、芸術家、文化人等、文化に携わる人々に、一定期間「文化交流使」として世界の人々の日本文化への理解の深化や、日本と外国の文化人のネットワークの形成・強化につながる活動を展開してもらうことを目的として、平成15年度から始めた事業です。
			P2	「文化交流使」の活動には、(i)日本在住の芸術家、文化人が海外に一定期間滞在し、日本の文化に関する講演、講習や実演などを行う「海外派遣型」、(ii)海外在住の日本文化に深い知見を持つ芸術家、文化人が、講演、講習、現地メディアへの投稿、出演等を行う「現地滞在型」、(iii)講演等で来日する諸外国の著名な芸術家が、日本滞在期間を利用して学校などを訪問して実演・講演等を行う「来日芸術家型」の三つの類型があります。
			P3	平成16年度は、「海外派遣型」文化交流使として11名、「現地滞在型」文化交流使として4名、「来日芸術家型」文化交流使として4組の指名を行いました。重要無形文化財保持者、写真家や音楽家など様々な分野で活躍中の方々の活動を通じて、日本文化のこれまで紹介されていなかった一面や、日本文化になじみの薄かった国や地域での日本文化の紹介などの活動を行っています。
		② 文化庁文化交流使活動報告会	P4	平成15年度に文化庁文化交流使として海外で活動した人々による報告会を、東京国立博物館平成館大講堂にて開催しました。
		P5	笑福亭鶴笑氏(落語家)、田中千世子氏(映画評論家)、バロン吉木氏(漫画家)、三浦尚之氏(福島学院大学教授)、渡辺洋一氏(和太鼓奏者)の5名が活動報告を行うとともに、国際文化交流について討論し、さらに笑福亭鶴笑氏によるパペット落語(笑福亭鶴笑氏が自ら考案した落語形式で、足や膝につけた人形を操りながら演じる。)の実演が行われました。	
	(2) 国際文化フォーラムの開催		P6	「国際文化フォーラム」は、国際的に著名な国内外の芸術家・文化人などを招聘し、座談会、講演などの形式により、世界の文化芸術の最新諸相や動向について語り合ってもらうことを目的として、平成15年度から開始した事業です。
			P7	平成16年度も15年度に引き続き、11月に関西地区で、「文化の多様性」の共通テーマの下に、「国際情勢における『文化の多様性』の意義」、「シルクロードと仏教文化」などについて話し合い、世界に向け、文化のメッセージを強く発信しました。
	(3) 国際芸術見本市			P8

⁶ 該当箇所は文部科学省のホームページでも確認することができる。URLは次のとおり。

http://www.mext.go.jp/b_menu/hakusho/html/hpab200401/hpab200401_2_277.html

4.2 各段落間の類似度

4.2.1 全体の傾向

OW6X_00000 を構成する 8 個の段落相互の類似度を表 2 に挙げた（同一段落どうしの類似度は必ず 1 になるので除く）。値は 0.0364～0.5614 の間に分布し、平均値は 0.280 である。類似度の分布の様子を図 1 に示した⁷。

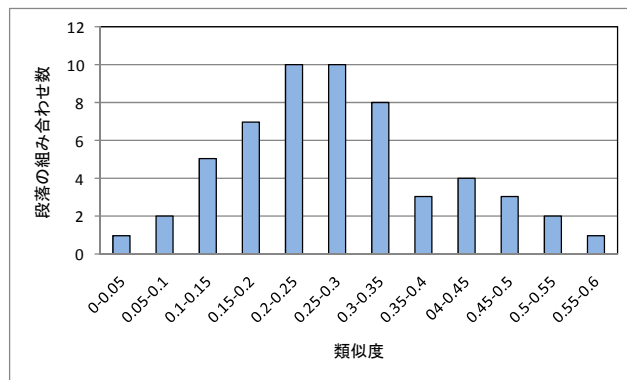


図 1 類似度の分布

表 2 はすべての段落間の類似度であるが、ここからある段落 a から他の段落 b への類似度において、対象とする相手方の段落 b との類似度が最も高い段落（表の太字のセル）がほとんど第 1 段落（P1）～第 3 段落（P3）に集中しており、このサンプルは前方の段落に依存する傾向があることが見て取れる。

表 2 段落間の類似度

	P1	P2	P3	P4	P5	P6	P7	P8	平均
P1		0.5128	0.4103	<u>0.4359</u>	0.2821	<u>0.4615</u>	0.2564	0.3077	0.3809
P2	0.3906		0.5156	0.1719	0.25	0.3125	0.0781	0.2344	0.2790
P3	0.4118	0.5686		0.3529	<u>0.3333</u>	0.2549	0.1765	0.3333	0.3473
P4	0.5	0.3	0.45		0.25	0.25	0.25	0.4	0.3428
P5	0.12	0.18	0.16	0.08		0.12	0.04	0.12	0.1171
P6	0.4815	0.3333	0.2963	0.1852	0.2593		<u>0.2963</u>	<u>0.3333</u>	0.3121
P7	0.2857	0.1429	0.25	0.25	0.1786	0.3214		0.2143	0.2347
P8	0.2368	0.2895	0.3158	0.1579	0.2105	0.2632	0.1053		0.2255
平均	0.3466	0.3324	0.3425	0.2334	0.2519	0.2833	0.1718	0.2775	

注

(1)縦の系列の段落が横の系列の段落に対してとる類似度の表。例えば、P3 の P2 に対する類似度は 0.5686（この値は P2 への P3 からの類似度と解することもできる）。

(2)太字は、当該段落から他の段落への類似度のうちもっとも値が高いもの。P4 の段落を例にとると、P4 の横の列（0.5,0.3,0.45,0.25,0.25,0.25,0.4）の中でいちばん高い値の 0.5 になる。

(3)下線は、他の段落から当該段落への類似度のうちもっとも値が高いもの。P5 の段落を例にとると、P5 の縦の列（0.2821,0.25,0.3333,0.25,0.2593,0.1786,0.2105）の中でいちばん高い値の 0.3333 に

⁷ 例えば、階級 0.1-0.15 は 0.1 より大きく 0.15 以下であることを示す。そのほかも同様。

なる。

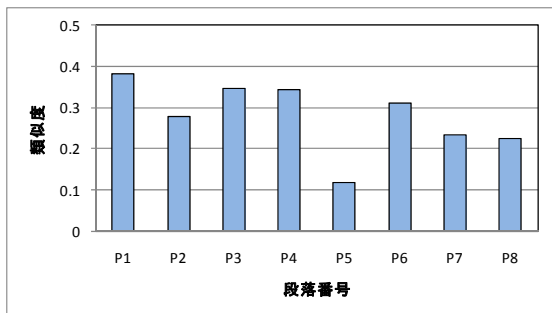


図2 他段落への類似度

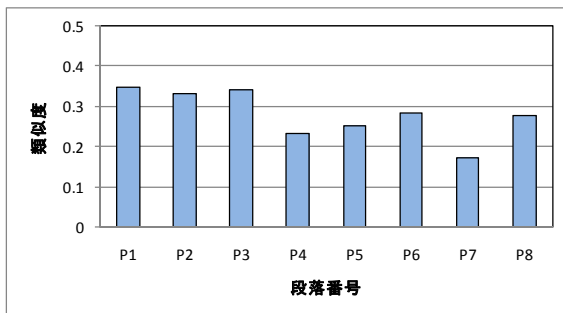


図3 他段落からの類似度

図2は、ある段落の他の段落に対する類似度の平均、図3はある段落の他の段落からの類似度の平均である。他段落からの類似度の平均は全体的に同じような値を示しているが、他段落への類似度の平均では、第5段落の値がほかとくらべて低くなっていることが分かる。このことは第5段落と他の段落とで共通して用いられる語について、第5段落での使用度数は少ないが、他の段落での使用度数が多いことを示唆する。例えば、第5段落と、(第5段落への類似度をもっと高い)第3段落の場合は「行(な)う、家(か)、活動、交流、文化、名(めい)」の6語が共通して現れる語であるが、「文化」は第5段落に1回しか使用されていないのに対して第3段落では7回使用されている。同様に「交流」は1回に対して3回、「行(な)う、名(めい)」はそれぞれ1回に対して2回であった。この共通出現語の使用の不均衡が類似度の非対称性に影響していると考えられる。このことを踏まえて第5段落と第3段落を比較してみると、テキストの表層的な構造では第5段落は第4段落に従属するものであるが、語の分布状況から見ると第3段落とも関係が深いことになる。

4.2.2 類似度からみた全体の構成

図4は、段落間の類似度の平均値0.280の1.5倍(0.420)以上の値を持つ段落の組み合わせを図示したものである。図の見方は、例えばP1→P2であればP1のP2に対する類似度が高かったことを表している。両矢印は相互に類似度の値が高かったことを示す。図4から、第1段落から第4段落までは結束性が強いことが伺える。一番多くの矢印が出入りしている第1段落がこのテキストの中心的な位置を占めていると言えよう。

第1段落と相互に類似度が高い2つの段落(第4段落と第6段落)のうち第6段落は、「～は、・・・てもらふことを目的として、平成15年度から開始した事業です。」という文構造であり、第1段落と骨格は同じである。このような「形式の類似性」も結束性に貢献していると考えられる。

この中では、第5、第7、第8段落が比較的独立性が高く他と分離されているが、後述のように第5段落は第4段落を具体的に展開したものであり、第7段落も事情は同じである。この関係は今回の分析からは読み取れない。

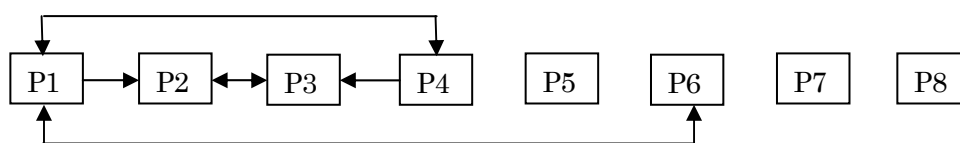


図4 類似度からみたテキストの構成

4.2.3 直前・直後の段落との類似度

類似度の値を利用して連続する段落間の切れつきについて考察する。山崎（2012）では直前の段落への類似度よりも直後の段落への類似度の方が大きいところが内容的な切れ目であり表層的なテキストの構成とも一致する場合が多いと指摘しているが、このサンプルではどうであろうか。結果を図5に示す。

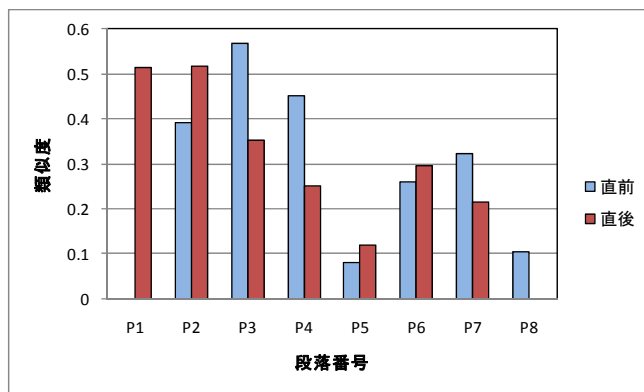


図5 直前・直後の段落との類似度

図5から第2段落、第5段落、第6段落が直後への類似度が高いことが分かる。表層的には第2段落は第1段落の続きであり、第1段落の内容を具体化しているものであるが、具体的な内容が多くなったために第1段落への類似度が相対的に低くなったものと思われる。同様の関係は第5段落にも見られる。第5段落も直前の第4段落の内容を具体化したものであるが、第4段落が第5段落の短いまとめの内容であることから直前の段落への類似度が相対的に低くなったものである。本稿の類似度の測定は同一語かどうかによっているので、このようなものごとを具体化して述べるようなつながりについては感度が弱い。

なお、第8段落は後続の段落がなく、上記の方法では観察できないが、直前の段落との類似度がかなり低いことからここも内容的な切れ目に相当する可能性が高いと思われる。

5. まとめと今後の課題

本稿では段落間の非対称的類似度を利用して、テキストの結束性のようすを概観した。今回扱ったデータは白書のサンプル1つのみであったが、すべての段落間の組み合わせを観察することにより、どの段落とどの段落とが関係が深いのか結束性の一端を伺うことができた。また、隣接した段落以外にも結束性の高い段落があり、それらの関係を利用したテキストの構成の分析への発展の可能性を示唆した。

本稿で利用した「無性格語」のリストは雑誌九十種調査の結果から作られたもので、異なるレジスターの分析に耐えるかどうかは検証が必要であろう⁸。例えばリストには固有名詞「日本」が含まれているが、白書の分析には「日本」は重要な話題として必要な語であり、必ずしも無性格とは言えないだろう。

今後の予定としては、指示詞や接続詞などのほかの結束性を表す手段との関連も視野に入れて語彙的結束性の現れ方を総合的に記述したいと考えている。

⁸ 今回使用したサンプルについては無性格語を排除しなくてもほとんど同じ結果であったが、どのような場合にこのリストが有効かは確認が必要である。

付表 本稿で用いた「無性格語」

語彙素	語彙素読み	品詞
余り	アマリ	副詞
余り	アマリ	形状詞一般
有る	アル	動詞-非自立可能
言う	イウ	動詞一般
行く	イク	動詞一般
一	イチ	名詞-数詞
今	イマ	名詞-普通名詞-副詞可能
居る	イル	動詞-非自立可能
内	ウチ	名詞-普通名詞-副詞可能
円	エン	名詞-普通名詞-助数詞可能
円	エン	名詞-普通名詞一般
御	オ	接頭辞
多い	オオイ	形容詞一般
大きい	オオキイ	形容詞一般
大きな	オオキナ	連体詞
置く	オク	動詞-非自立可能
於く	オク	動詞一般
同じ	オナジ	形状詞一般
同じ	オナジ	連体詞
思う	オモウ	動詞一般
居る	オル	動詞-非自立可能
会	カイ	名詞-普通名詞一般
方	カタ	接尾辞-名詞的一般
方	カタ	名詞-普通名詞-助数詞可能
月	ガツ	名詞-普通名詞-助数詞可能
彼	カレ	代名詞
考える	カンガエル	動詞一般
聞く	キク	動詞一般
九	キュウ	名詞-数詞
位	クライ	名詞-普通名詞-副詞可能
来る	クル	動詞-非自立可能
五	ゴ	名詞-数詞
こう	コウ	副詞
五十	ゴジュウ	名詞-数詞
事	コト	名詞-普通名詞一般
此の	コノ	連体詞
此れ	コレ	代名詞
三	サン	名詞-数詞
さん	サン	接尾辞-名詞的一般
三十	サンジュウ	名詞-数詞
氏	シ	接尾辞-名詞的一般
氏	シ	名詞-普通名詞一般
四	シ	名詞-数詞
然し	シカシ	接続詞
七	ナナ	名詞-数詞
七	シチ	名詞-数詞
自分	ジブン	名詞-普通名詞一般
仕舞う	シマウ	動詞-非自立可能
者	シャ	接尾辞-名詞的一般
知る	シル	動詞一般
十	ジュウ	名詞-数詞
十	トオ	名詞-数詞
為る	スル	動詞-非自立可能
生活	セイカツ	名詞-普通名詞-サ変可能

語彙素	語彙素読み	品詞
千	セン	名詞-数詞
そう	ソウ	副詞
そう	ソウ	名詞-助動詞語幹
そう	ソウ	形状詞-助動詞語幹
そして	ソシテ	接続詞
其の	ソノ	連体詞
其れ	ソレ	代名詞
第	ダイ	接頭辞
対する	タイスル	動詞一般
出す	ダス	動詞-非自立可能
達	タチ	接尾辞-名詞的一般
為	タメ	名詞-普通名詞-副詞可能
つく	ツク	動詞一般
付く	ツク	動詞-非自立可能
強い	ツヨイ	形容詞一般
的	テキ	接尾辞-形状詞的
出来る	デキル	動詞-非自立可能
出る	デル	動詞一般
度	ド	名詞-普通名詞-助数詞可能
どう	ドウ	副詞
時	トキ	名詞-普通名詞-副詞可能
所	トコロ	名詞-普通名詞-副詞可能
所	トコロ	名詞-普通名詞一般
共	トモ	名詞-普通名詞一般
共	トモ	接尾辞-名詞的-副詞可能
取る	トル	動詞一般
無い	ナイ	形容詞-非自立可能
中	ナカ	名詞-普通名詞-副詞可能
何	ナニ	代名詞
何	ナン	名詞-数詞
成る	ナル	動詞-非自立可能
二	ニ	名詞-数詞
日	ニチ	名詞-普通名詞-助数詞可能
二十	ニジュウ	名詞-数詞
日本	ニッポン	名詞-固有名詞-地名-国
人	ニン	接尾辞-名詞的一般
年	ネン	名詞-普通名詞-助数詞可能
はいる	ハイル	動詞一般
場合	バアイ	名詞-普通名詞-副詞可能
八	ハチ	名詞-数詞
日	ヒ	名詞-普通名詞-副詞可能
人	ヒト	名詞-普通名詞一般
一	ヒト	名詞-数詞
ひとり	ヒトリ	名詞-普通名詞-副詞可能
百	ヒャク	名詞-数詞
方	ホウ	名詞-普通名詞一般
僕	ボク	代名詞
程	ホド	名詞-普通名詞-副詞可能
前	マエ	名詞-普通名詞-副詞可能
また	マタ	接続詞
また	マタ	副詞
万	マン	名詞-数詞
見る	ミル	動詞-非自立可能
目	メ	名詞-普通名詞一般

語彙素	語彙素読み	品詞
目	メ	接尾辞-名詞的-一般
持つ	モツ	動詞-一般
物	モノ	名詞-普通名詞-一般
物	モノ	名詞-普通名詞-サ変可能
問題	モンダイ	名詞-普通名詞-一般
遣る	ヤル	動詞-非自立可能
行く	ユク	→行く(イク)
良い	ヨイ	形容詞-非自立可能
様	ヨウ	形状詞-助動詞語幹
因る	ヨル	動詞-一般

語彙素	語彙素読み	品詞
四	ヨン	名詞-数詞
四十	ヨンジュウ	名詞-数詞
等	ラ	接尾辞-名詞的-一般
零	レイ	名詞-数詞
六	ロク	名詞-数詞
分かる	ワカル	動詞-一般
訳	ワケ	名詞-普通名詞-一般
私	ワタクシ	代名詞
私	ワタシ	代名詞

付表の注

「余り」「円」「同じ」「方(かた)」「七」「十」「そう」「所」「共」「何」「まだ」「目」「物」「私」については、短単位での品詞が複数に渡っているため、該当するものを列挙した。

謝 辞

本研究は国立国語研究所の共同研究プロジェクト「テキストにおける語彙の分布と文章構造」による研究成果の一部である。データとして利用した BCCCWJ は、国立国語研究所のプロジェクト及び文部科学省科学研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」(平成18～22年度、領域代表者：前川喜久雄)による補助を得て構築したものである。

参考文献

- Halliday, M.A.K. and Hasan, R.(1976) *Cohesion in English*. Longman (邦訳『テキストはどのように構成されるか』、大修館書店、1997刊)
- Károly, Krisztina.(2002) *Lexical Repetition in Text*. Peter Lang.
- 小木曾智信、間淵洋子、前川喜久雄(2011)『現代日本語書き言葉均衡コーパス』における形態論情報付き XML フォーマット」、言語処理学会第17回年次大会予稿集、pp.352-355.
- 田中章夫(1973)「自動抄録処理におけるキー・ワードの性格」『電子計算機による国語研究V』秀英出版、pp.141-184.
- 平林健治(2003)「日本人初級学習者の英文ライティングの結束性の視点からみた分析」、愛知新城大谷短期大学研究紀要、2-4、pp.67-76.
- 水谷静夫(1980)「用語類似度による歌謡曲仕分『湯の町エレジー』『上海帰りのリル』及びその周辺」、計量国語学、12(4)、pp.145-161.
- 水谷静夫(1983)『朝倉日本語新講座 2 語彙』朝倉書店
- 宮島達夫(1970)「語いの類似度」、国語学、82、pp.42-64.
- 山崎誠(2012)「共起語率の分布からみるテキストの語彙的特徴」、第1回コーパス日本語学ワークショップ予稿集、国立国語研究所、pp.221.226.

執筆者一覧

山崎誠（国立国語研究所言語資源研究系准教授）

内山清子（国立情報学研究所学術コンテンツサービス研究開発センター特任准教授）

江田すみれ（日本女子大学文学部教授）

小森理（統計数理研究所統計思考院特任助教）

清水まさ子（独立行政法人国際交流基金日本語国際センター専任講師）

高崎みどり（お茶の水女子大学大学院人間文化創成科学研究科教授）

馬場俊臣（北海道教育大学教育学部札幌校教授）

馬場康維（統計数理研究所名誉教授／統計数理研究所統計思考院特命教授）

村田年（慶應義塾大学日本語・日本文化研究センター教授）

国立国語研究所共同研究報告 12-06

テキストにおける語彙の分布と文章構造 成果報告書

2013年3月25日発行

著者 山崎誠、内山清子、江田すみれ、小森理、清水まさ子、高崎みどり、馬場俊臣、
馬場康維、村田年

発行 大学共同利用機関法人人間文化研究機構国立国語研究所

〒190-8561 東京都立川市緑町10-2

電話 042-540-4300（代表）

<http://www.ninjal.ac.jp/>

©国立国語研究所

ISBN 978-4-906055-26-5

ISSN 2185-0127

Distribution of Vocabulary and Sentence Structures in Texts

Makoto Yamazaki, Kiyoko Uchiyama, Sumire Goda,
Osamu Komori, Masako Shimizu, Midori Takasaki,
Toshiomi Baba, Yasumasa Baba, Minori Murata

March 2013