



Title	＜最終講義＞探索的コーパス言語学・補考
Author(s)	石井, 正彦
Citation	現代日本語研究. 2024, 15, p. 1-23
Version Type	VoR
URL	<a href="https://doi.org/10.18910/98460">https://doi.org/10.18910/98460</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

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

The University of Osaka

# 〈最終講義〉探索的コーパス言語学・補考

Final Lecture: Supplementary Considerations on the Exploratory Corpus  
Linguistics

石井 正彦

ISHII Masahiko

**キーワード：**探索的コーパス言語学，探索的・確認的・記述的アプローチ，コーパス検証型・駆動型，探索的データ解析，映像KWIC

## 要 旨

石井(2019)が提唱した「探索的コーパス言語学」について，その理論と方法に関する補足的な考察を行い，探索的コーパス言語学が標榜する「探索的アプローチ」とはどのようなものを改めて問い直す。具体的には，探索的アプローチが（大規模標本コーパスによる）確認的アプローチや（伝統的な）記述的アプローチとどのように異なるのかを，関連する諸論考を引きながら検討するとともに，任意的で定式化されていないと批判される探索的アプローチの方法論・手法について，マルチメディア・コーパスにおける映像KWICをその一候補として提案・検討する。

## 1. 講義の目的

ご紹介いただきました石井です。まず，最終講義に向けてお力添えくださった基盤日本語学講座の，世話教員の三宅先生を初めとする先生方，助教の李さん，学生の皆さん，そのほかご関係の皆様に，御礼申し上げます。それから，今日一緒に最終講義を行うことをお許しくださった田野村先生にも御礼申し上げます。また，今日は遠方からも懐かしい顔たくさんおいでくださって，本当にありがとうございます。それでは，どうぞよろしくお願いします。

はじめに，私の主な研究歴を以下にお示しします。

- 1) 複合動詞, 複合名詞 (学術用語), 臨時一語を中心とした語構成の研究  
『現代日本語の複合語形成論』(ひつじ書房, 2007 年)
- 2) 国立国語研究所の教科書・テレビ放送の語彙調査を出発点とする計量的な語彙研究  
『これからの語彙論』(斎藤倫明氏との共編著)(ひつじ書房, 2011 年)  
『語彙の原理』(編著)(朝倉書店, 2019 年)
- 3) 新聞を中心とする基本語彙とその変化の研究  
「現代新聞語彙における“基本語化”と“非基本語化”」  
(『阪大日本語研究』34 号, 2022 年)
- 4) テレビ放送を中心とする言語行動 (言語と非言語行動との関係) の研究  
『マルチメディア・コーパス言語学』(孫栄爽氏との共著)  
(大阪大学出版会, 2013 年)
- 5) 探索的データ解析による日本語研究の可能性の研究  
『探索的コーパス言語学』(大阪大学出版会, 2019 年)  
*Exploratory Corpus Linguistics* (大阪大学出版会, 2024 年)

今日は、このうちの5番目、探索的コーパス言語学にかかわるお話をしてみたいと思います。この、2019年に大阪大学出版会から刊行した本ですが、『探索的コーパス言語学』というふうに銘打ってというか、仰々しく付けて出したにもかかわらず、私自身の中では、「探索的コーパス言語学」という考え方が本当に成立するのか、まだ確信を持ってないというか、詰め切れていないところがたくさんあって、少しでもそういう突き詰める作業をしていきたいと思っているのですが、今日はその一端といえましょうか、現時点で考えてみたことをお話したいと思います。

それで、講義の目的としては、「探索的コーパス言語学における探索的アプローチとはどのようなものか」ということを改めて問い直すということ、具体的には、「探索的アプローチは確認的アプローチとか記述的アプローチとはどう違うのか」、そして「探索的アプローチの新たな手法として何か考えられることがあるのか」ということを検討する、ということになろうかと思います。

## 2. 探索的コーパス言語学とは何か

### 2. 1. 確認的アプローチと探索的アプローチ

最初に、そもそもこの「探索的コーパス言語学」とは何かということについて、『探索的コーパス言語学』の冒頭に書いたところを(1)に抜きましたので、

ご覧ください。

(1) 「探索的コーパス言語学 (Exploratory Corpus Linguistics)」とは、コーパス言語学において、これまで主流であるところの「確認的なアプローチ」の研究に対し、それとは異なる「探索的なアプローチ」による言語研究を標榜して、筆者が仮に設けた名称である。ここで、「確認的 (confirmatory)」対「探索的 (exploratory)」とは、統計学の探索的データ解析における用語を借用したもので、「理論モデル主導型 (theoretical model oriented)」対「データ主導型 (data oriented)」という用語で呼ばれることもある。その詳細は後述するが、ごく簡単に言えば、確認的な（理論モデル主導型の）アプローチとは「データによって仮説を検証する」ものであり、探索的な（データ主導型の）アプローチとは「データによってデータそのものを説明する（データ自身に語らせる）」ものである。（『探索的コーパス言語学』, p. 1）

## 2. 2. 理論モデル主導型とデータ主導型

この「確認的アプローチ 対 探索的アプローチ」という考え方は、(2)に示すように、統計学における「理論モデル主導型 対 データ主導型」という考え方に重なるものです。

(2) ここで、理論モデル主導型とデータ主導型との違いをより具体的に理解するには、後者の端緒となった探索的データ解析の考え方が参考になる。探索的データ解析 (Exploratory Data Analysis ; EDA) とは、推測統計学に代表されるような、統計的仮説を検証するための「確認的データ解析 (Confirmatory Data Analysis ; CDA)」に対して、限られたデータからその構造や特徴に関する情報を探り出すことを目的として、アメリカの統計学者テューキー (J. W. Tukey) が中心となって主に 1970 年代に開発した一連のデータ解析の手法や考え方をさす (Tukey1977)。そこでは、CDA のように標本から母集団を推定するという、すなわち、「部分から全体への推論を明確な手段をもって行う」理論モデル主導型の方法は採らず、どのようなデータでも対象として、「データに含まれる混沌とした情報を整理し、データについての洞察を得、データの構造を探り、そうすることによってデータの説明を行おうとする」「いわばデータそれ自身に語らせる」(渡部[他]1985 : 170) ことが行われる。テューキー自身の比喻によれば、理論モデル主導型の CDA が得られた証拠から犯罪を判定し量刑を確定する検事や裁判官のような仕事であるのに対して、データ主導型の EDA は証拠を収集して立件していく刑事や警察官のような仕事であるという (吉田[編]1995 : 104)。

(『探索的コーパス言語学』, pp. 4-5)

探索的データ解析という統計学の一つの学派で言われている「理論モデル主

導型 対 データ主導型」という考え方をコーパス言語学にも当てはめれば、「確認的コーパス言語学 対 探索的コーパス言語学」という、そういう捉え方ができるのではないかということですね。

## 2. 3. コーパス検証型とコーパス駆動型

ただし、これに類する考え方は、実は以前からコーパス言語学の方でも唱えられています。(3)をご覧ください。

(3) コーパス検証型言語学という用語が示すのは、大規模コーパスを使用して言語研究の知見を得ることができなかった時代に提唱された言語理論や言語記述について、もっぱらそれらを説明・検証・例証する目的でコーパスを用いる方法論である……コーパス検証型研究手法を取る言語学者は、理論とデータの関係について「自信」を持っている。彼らは、もともと基本的に正しいと考えている言語モデルや言語記述を持ち込み、それらが提唱する範疇に基づいてコーパス分析を行ない、それに沿ってデータをふるいにかけるからだ。(中略)

コーパス駆動型研究手法では、言語学者の責任はデータ全体を一体的に扱うことにあり、コーパスから得られる言語的証拠に関して、包括的言語記述を行うことが目指される。それゆえコーパスは、既存の言語理論を支持するための単なる用例集以上のものとして、また、確立された既存の理論体系を拡張しうる手段以上のものとしてとらえられるのである。この手法では、コーパスに見られる言語的証拠と完全に整合し、かつ、それを直接的に反映した理論的説明がなされる…理論が証拠から独立して存在することはないし、理論化に至る一般的な方法論上の手続きも明白で、観察から仮説構築へと進み、一般化を経て理論的説明との統合に至るのである。

(Tognini-Bonelli 2001 : 84-87)

この「コーパス検証型 対 コーパス駆動型」という考え方ですが、(3)を読むだけではよくわかりません。コーパス検証型言語学の方はわかります。仮説を検証する。そのためにコーパスを見るんだ、用いるんだと。しかし、コーパス駆動型については何を言っているかよくわからない。それで、もう少しわかりやすく書かれたものを(4)に示します。

(4) 「コーパス検証型」の研究では、言語理論から導き出された言語構造の妥当性を前提とする。主な研究目標は、あらかじめ定義された言語的特徴の体系的な使用パターンを分析することである。したがって、コーパスに基づく定型言語研究では、研究者はあらかじめ定型表現を選択し、その表現がどのように使われているかを発見するた

めにコーパスを分析する。

これとは対照的に、「コーパス駆動型」の研究は、コーパスの分析から言語構成要素そのものが浮かび上がってくるような、より帰納的なものである。コーパス駆動型のアプローチは、言語学の標準的な実践とは異なり、分析に用いるべき言語構成要素について、先験的な仮定を最小限に抑える。最も基本的な形として、コーパス駆動型分析では単語の存在のみを仮定する。コーパス分析から発見された単語間の共起パターンが、その後の言語記述の基礎となる。(Biber 2009 : 276)

「コーパス駆動型の方では先験的な仮定を最小限に抑える」「最も基本的な形として、単語の存在のみを仮定する」というのは、語彙調査の用語で言えば、単位切りした単位語のみを用いて、それらを見出し語にまとめることはしない、ということです。それは、単位語を見出し語にまとめる時点で既に先験的な(同語異語判別の)理論がある。それは使っちゃいけませんよって言ってるんです、駆動型研究ではね。見出し語を使うともう先に理論があって、その理論に基づいてコーパスを探していくっていう検証型の研究になる。そうじゃなくて、単位切りはするけど、それをどう見出し語にまとめるとかまとめないっていうことはやらない。「単語のみを仮定する」というのはこういうことらしいです。

でも、まだちょっとよくわからないですね。一番わかりやすいというか端的に言っているものを(5)に紹介します。

(5) コーパス言語学が方法論なのか学問なのか、あるいはその両方なのかという議論は、「コーパス検証型」アプローチと「コーパス駆動型」アプローチのどちらが望ましいのかという問題と結びついている。(略)それは、コーパスの手法を「トップダウン」的に利用して、対象テキスト内のあらかじめ決められた言語的特徴を調査する研究と、コーパスの手法を利用して、対象テキストの言語を「ボトムアップ」的に調査し、「そこに何かがあるのか」を探る研究である。(Whiteside 2016 : 57)

つまり、あらかじめ自分でトップダウン的に設定した言語的特徴をコーパスに探しに行く、コーパスではどうなっているんだろうそれらは、という研究と、そういうことをせずに、コーパスをとにかくボトムアップ的に調査して、そこに一体何かがあるのか探っていこうとする研究ということ。こういう言い方が、コーパス検証型とコーパス駆動型の違いとして端的でわかりやすいと思うんですけど、このようなことがコーパス言語学の方でも言われていて、コーパス検証型は確認的コーパス言語学、それからコーパス駆動型が探索的コーパス言語

学に対応するか、あるいはそれに近いんじゃないかというふうに思っています。

ただ、(3)の引用の最後の方にある「理論化に至る一般的な方法論上の手続きも明白で、観察から仮説構築へと進む一般化を経て理論的説明との統合に至る」というような書き方は、探索的じゃないなと思うんですよね。その点では、コーパス駆動型と探索的コーパス言語学とがまったく同じっていうわけではないと思います。

### 3. 日本語のコーパス言語学

#### 3. 1. 確認的アプローチ

では、日本語のコーパス言語学ではどうなのかということですが、大規模コーパス言語学ではどうか。これにはまず確認的アプローチという側面があると思います。同じく『探索的コーパス言語学』の引用ですが、(6)をご覧ください。

(6) それから半世紀近く、同じ国語研究所といっても、もはや語彙調査という枠組みではなく、コーパス言語学という新しい枠組みにおいて、BCCWJは構築・公開された。そのデータ規模は延べ1億語(短単位)を上回るとされ、雑誌九十種調査のざっと200倍、同調査で母集団とされた1.4億語に迫る規模であり、統計的推測の精度は格段に向上しているはずである。同時に、均衡コーパスとして、雑誌だけでなく、書籍・新聞・白書・ブログなど異なるレジスターからもテキストを集め、書きことばの多様な変種とその実態をとらえることがめざされている。同コーパスの『利用の手引 第1.1版』によれば、BCCWJ設計の基本方針の第一には次のようにある。

##### (1) 現代日本語の縮図となるコーパス

従来、国立国語研究所が行ってきた語彙調査の手法を生かし、コーパスがその母集団の統計的な縮図になり、母集団に対し代表性(representativeness)を持つように設計する。これにより、母集団における言語的諸特性の分布が過不足なく表現できることになり、データの信頼性を高めることが出来る。

BCCWJは「母集団の統計的な縮図」としての「標本」コーパスであり、その大規模性・均衡性にもとづく「代表性」によって統計的推測の精度すなわち「信頼性」を高めているのである。このことは(日本語のコーパスとして)画期的であるが、しかしその基本的なアプローチが標本から母集団を推測するという理論モデル主導型のそれであることは明らかであり、それはまた半世紀前の語彙調査と基本的に変わらないものである。

(『探索的コーパス言語学』, p. 10)

日本語の大規模コーパス言語学の象徴ともいえるBCCWJ(『現代日本語書き言葉均衡コーパス』)は、理論モデル主導型、確認的コーパス言語学のためのコー

パスだということが、まずは言えるっていうことだと思うんですね。

### 3. 2. 記述的アプローチ

2 番目の記述的アプローチということですが、これは大規模コーパスを母集団や標本と捉えることをせず、大規模かつ標準的な言語データの集積と捉え、そこに見られる限りでの構造や特徴を記述しようとする伝統的な言語研究のアプローチということで、伝統的な言語研究が大規模コーパスを使っているという、そういうイメージですね。(7)は、宮島達夫さんの 2007 年の論文からの引用ですが、こういった立場を仮に記述的アプローチと今呼びたいと思います。

(7) 単にコーパスを利用して言語現象をしらべた、という研究を「コーパス言語学」とよぶ必要はない。コーパスの第 1 の価値は、膨大な用例の量にある。これからの研究には、当然それを利用すべきだが、それは「コーパス言語学」でも「用例言語学」でもない。大量の例文をしらべることは、まさに言語学の王道、限定語なしのザ・言語学だ、というのが、わたしの立場である。  
(宮島 2007 : 41)

BCCWJ を利用される方の多くは、この記述的アプローチだと思うんですね。BCCWJ を標本コーパスだと認識して利用している人はそんなにいないと思います。ですから、直感的にはこちらの方が主であって、確認的アプローチは従というか、二次的のように思えますけど、本来はそうじゃないんですね。BCCWJ は、そういう、母集団の縮図となるように設計するというのが基本思想であって、何のためにそうするかといえば、それは確認的アプローチをとっているから、それでもって仮説を検証することができるコーパスを目指しているから、ということですね。

### 3. 3. 探索的アプローチ（『探索的コーパス言語学』が提案した方法）

もう一つが探索的アプローチなんですけど、これを標榜するのが「探索的コーパス言語学」であるわけですが、最初に申しましたように、それについてはまだまだ突き詰めなきゃいけないことが残っているわけです。で、その探索的コーパス言語学というときの探索的アプローチとはどんなものかということについては、『探索的コーパス言語学』ではとりあえず以下の 3 点を考えました。



- (8) ①全文コーパスによる探索
- ②多様なコーパスによる探索
- ③探索的データ解析による探索

『探索的コーパス言語学』という本の中で私が探索的アプローチだろうと思ってやったことは、一つは全文コーパスというものを使っていろいろ探ってみるということです。これは先ほど BCCWJ は標本コーパスだって申しましたが、この全文コーパスというのは、対象となるテキストを丸ごとコーパスとして利用するというもので、私が主にやったのは、国語研究所の高校教科書の語彙調査、これは全数調査で全文コーパスを作っているんですが、そのうちの物理の教科書を1冊丸ごと全部入れたデータを使って、そこでたった1回しか出てこない単語は、一体なぜたった1回なのかっていうことを探った。これは多分標本コーパスではできないですね、たった1回ってわからないから。全文だから初めてたった1回だったってことがわかるので、いくら大規模でも、標本コーパスでサンプリングしてある場合は、偶然そこで1回だった単語を全文(母集団)の中でも1回だったとは言えないんですよ。標本から母集団を統計的に推測する研究ではできないだろうなっていう、そういう意味合いです。

それから多様なコーパスによる探索、これは(9)をご覧ください。

- (9) さらに、第3部「多様なコーパスによる探索」では、全文コーパス以外にもさまざまなコーパスの可能性を考えることによって、探索的コーパス言語学の可能性——コーパスを使って可能になるデータ主導型研究の領域——を広げていくことを試みる。具体的には、筆者が実際に試作したいくつかのコーパス(マルチレベル通時コーパス、単一言語パラレルコーパス、マルチメディア・コーパスなど)を紹介し、それぞれによる分析事例をあげて、それらが新たにどのような日本語研究を可能にするか、ということを検討する(第9章)。(『探索的コーパス言語学』, p. 15)

全文コーパスによる探索、そして(詳しい説明は省きますが)探索的データ解析による探索のほかに、多様なコーパスっていうか、これまであまりないようなコーパスを自分たちで実際に作ってみて、それでどんな研究が可能になるかということを追求していく、というのも探索的コーパス言語学の一つの手法なんじゃないかっていうことですね。

そういうことで、この①②③の方法を『探索的コーパス言語学』では探索的

アプローチとして提案したわけです。しかし、これは、あくまでその時点でそう考えたというだけであって、探索的コーパス言語学の方法論を原理的に突き詰めたものではないので、この辺りがまだまだ問題っていうことですね。

#### 4. 探索的アプローチに対する批判（と反論）

##### 4. 1. 「確認」と「探索」とは循環的（両者をことさら区別する必要はない）

探索的アプローチには批判もあります。それに対する反論的なものもあるんですけども、それらを合わせて紹介しておきたいと思います。

まず最初に、確認と探索とは循環的で、両者をことさら区別する必要はないという批判があります。これは、(10)の引用でマケナリーさんとハーディーさんの共著の中で言われているものです。

(10) コーパス検証型とコーパス駆動型の区別は、単純な二項対立がしばしばそうであるように、実際にはもう少し流動的なものである。コーパス検証型の研究者は既存理論の枠組みを使用しようとするが、その枠組みを実際のデータに適用してみた結果、うまくいかないとわかると、いわゆるコーパス駆動型と呼びうるような手法で元の枠組みの修正を行う。こうした過程は循環的なもので、言語学者全般、とくに計算言語学者は、以前よりこのことをよく理解していた。

(McEnery and Hardie 2012=2014 : 223)

つまり、コーパス駆動型なんて言ってるけれども、それはコーパス検証型とされている研究の中のある1段階、つまり仮説の修正とかですね、そういうところで行うものであって、従来から検証型の中に組み込まれてるっていう批判です。仮説の発見とその検証、そして仮説の修正、また仮説の再検証というふうに循環していく科学的アプローチの中に、探索的アプローチというのは組み込まれているんだから、それ自体は何かとりたてて仮説検証型と区別する必要はないと、このマケナリーさんとハーディーさんは言ってるんですね。「単純な二項対立がしばしばそうであるように」っていう、ちょっと意地悪な言い方してますから、嫌いなんでしょうね、こういう分け方がね。

それに対してホワイトサイドさんが、(11)で次のように言っています。

(11) コーパス駆動型とコーパス検証型の区別は、時に主張されるほど明確ではないかもしれないが、McEnery と Hardie の主張する「実質的な違いはない」というのは、反論としては行き過ぎである。言語理論に関する既存の知識から自由になることは明らかに不可能であるが、可能な限りパターンや特徴を出現させるように設計されたコーパス・アプローチをとることと、あらかじめ決められた言語的特徴を探し出すようなコーパス・アプローチをとることには、非常に大きな違いがある。このような理由から、コーパス駆動型アプローチとコーパス検証型アプローチの区別は、依然として極めて有用なものであると主張したい。(Whiteside 2016: 60)

あらかじめ何を調べるか決めておいて、それをコーパスの中に探しに行くというやり方と、何も決めずにコーパスの中に何があるんだろうって探るというやり方はやっぱり違うと。それを実質的には同じものだと言うのは、言い過ぎじゃないかっていうことです。

それから、(12)は私が『探索的コーパス言語学』の中で書いたものです。私も、探索的ということを仮説の発見とか修正というふうに決めつけることはよくないんじゃないか、探索ということには仮説発見などとは別の面があって、むしろそちらの方が大事ではないか、そういう考え方をしています。

(12) しかし、私見では、コーパス駆動型を「元の枠組み、すなわち仮説がうまくいかないとわかると、その修正を行う」といった、仮説検証過程と連続的・循環的に行われる仮説形成過程と断じることが適当でないように思われる。先の Tognini-Bonelli (2001) の「この手法では、コーパスに見られる言語的証拠と完全に整合し、かつ、それを直接的に反映した理論的説明がなされる…理論が証拠から独立して存在することはない」という説明は、コーパス駆動型が「データによって仮説を検証する」ものではなく、「データによってデータそのものを説明する(データ自身に語らせる)」ものであることを示しているようにも思える。いずれにせよ、「データに語らせる」というアプローチは、それ自身は検証されることを前提としていない。もちろん、そのようにして得た知見や洞察が、結果的に仮説として扱われ、検証過程に付されることはあっても、最初から仮説形成検証過程に組み込まれているとか、それを前提として存在しているとかと考えることは適切ではない。(『探索的コーパス言語学』, p. 8)

#### 4. 2. 完全に「探索的」(＝仮説をもたない)ということはありません

次に、完全に探索的ということはありません、つまり、仮説をまったく持たないということはありませんという批判ですね。(13)は、これは批判的に書

いているわけじゃないんですけど、バイバーさんの引用です。

(13) 実際には、コーパス駆動型研究の傘の下で、かなり広範な方法論が用いられてきた。これらの方法論はすべて、コーパスの帰納的分析を通じて新しい言語構成要素を発見するという研究目標の性質によって、純粋なコーパス検証型の研究と区別することができる。しかし、コーパス駆動型の研究は、コーパス検証型の手法を取り入れることが多いため、ハイブリッド型と考えるのが妥当かもしれない。

(略)

パターン文法の研究は、コーパス駆動型研究の最も発展した例としてしばしば引用されるため、ここでは参考になるが、実際には、コーパス駆動型とコーパス検証型の両方の方法論が用いられている。それは、各パターンの語彙的関連がコーパス分析によって発見されるという点ではコーパス駆動型の研究であるが、分析があらかじめ定義された言語カテゴリー（「名詞」や「動詞」といった基本的な文法カテゴリー、フレーズの種類、さらには構文構造など）によって決定される部分があるという点ではコーパス検証型の研究である。

(Biber 2009:278-279)

つまり、純粋なコーパス駆動型ってちょっと難しいと。コーパス駆動型研究と言われている研究も、多くはコーパス検証型の考え方、つまり言語的に先見的に設定される仮説とか前提とかというものを利用してやっていることが多い、ということですね。それで、完全なコーパス駆動型というのは実際にはなかなか難しいので、コーパス検証型も取り入れた「ハイブリッド型のアプローチ」と言った方がいいかもしれないと、バイバーさんはここで述べています。

また、パターン文法というのは、ここではフレーズオロジー (phraseology) とかそういう研究だと思うんですよ。そういう研究をしている人が品詞の概念をまったく抜きにしてコーパスを探索するとか、そういうことはあり得ないわけで、そこで品詞の概念を使っているってことは、コーパス検証型の面があるということですね。だから、完全な仮説なしっていうことは難しいということ言ってるんだと思うんです。

(14)は、荻野綱男さんの2021年のものです。『計量国語学』の「言語統計学入門」という連載シリーズの第1回の一節ですけど、「統計学の種類」の一つとして「探索統計学」つまり「探索的データ解析」を次のように説明しています。

(14) (探索統計学は) 混沌とした多様なデータから何か規則性を把握しようとすることが目的である。

考え方として、発見的な手法であり、最初の仮説がない(ふりをする)。この点が推測統計学と大きく異なる点である。推測統計学のように仮説-検定を重視すると、最初に仮説として考えていなかったものは、まったく扱えないことになる。それが不満だということで発展してきた考え方が探索統計学である。実際には、データを集める段階でどんなデータを集めればいいのか、事前に考えておくことが必要なので、仮説がまったくないということは考えにくいのだが、分析のしかたとしては、仮説が成り立つか否かを検定(判断)するのではなく、集めてきたデータをあれこれ分析し、試行錯誤し、その中から何かを発見するという手順になる。ある意味で柔軟な考え方なので、多くの分野に応用できる可能性がある。(荻野 2021: 35-36)

この「ふりをする」というのが、本当に言い得て妙というか、つまり、ふりをしていたら、ハイブリッド型ってことですね。まったくもって仮説がないっていうのはなかなか難しいんじゃないかっていうことですね、これも。

それに対して、この先見的な仮説や前提を持つか持たないかっていうことについては、そもそも探索的データ解析の先導者であるテューキーさんが、その代表的な著作の中で(15)のように言ってるんですね。

(15) 探索的データ解析は、探偵の仕事である——数値探偵の仕事、あるいは計数探偵の仕事、あるいはグラフ探偵の仕事である。

犯罪を捜査する探偵には、道具と理解の両方が必要である。もし彼が指紋の粉末を持っていなければ、物の表面に指紋を見つけることはほとんどできないだろう。犯人がどこに指を置いたかを理解していなければ、正しい場所を探すことはできない。同様に、データ解析者にも道具と理解の両方が必要である。(略)

理解にはさまざまな限界がある。多くの探偵小説が明らかにしているように、ロンドンのスラム街で、ウェールズの人里離れた村で、パリの貴族の間で、牧畜業を営む西部で、あるいはオーストラリアの奥地で犯罪者を発見するには、まったく異なる種類の詳細な理解が必要である。ロンドン警視庁の警官が牛泥棒の追跡をうまくやるとは思わないし、テキサスのレンジャーがバーミンガムの中心部で効果を発揮するとは思わない。同様に、地震に関するデータ、化学製造技術に関するデータ、サービス産業における企業の規模や利益に関するデータ、人間の聴力に関するデータ、自殺率に関するデータ、人口成長に関するデータ、恐竜の化石に関するデータ、ミバエの遺伝子に関するデータ、分子生物学の最新の研究成果に関するデータなどを扱う際に高い効果を発揮するには、まったく異なる詳細な理解が必要である。これらのどの分野でも、あるいは他の多くの分野でも、データ解析の完全な紹介をするには、われわれが

持っている時間よりもはるかに多くの時間が必要であろう。

しかし、ロンドン警視庁の刑事が、西部開拓時代や奥地でまったく役に立たないというわけではない。彼は従来の探偵の仕事について、どこでも役立つある種の一般的な理解を持っている。

データ解析にも同様の一般的な理解がある。

(Tukey 1977 : 1)

ここで言っている「理解」というのは、これは仮説とか前提とかいうのとはかなり重なると思うんです。まったく同じではないと思いますが、例えばコーパスを分析するときに品詞の概念を前提にするっていうのは、これはもう、仮説というよりも前提ですよ。チューキーさんはここで「理解」を“understanding”と言ってますけど、彼が「理解」と言ってるものがどういうものかというのは、この範囲だけではよくわからないんですが、そういう先見的な前提というものを含んでるんじゃないかと思いますね。ですから、探索的データ解析においても、当初からバイバーさんの言うハイブリッドのようなことは認められているというか、考えられているんじゃないかっていうふうに思います。これが正しい理解かどうかはなお検討しなきゃいけないと思いますけれども、今のところ私はそう思っています。ですから、ハイブリッドであるということが探索的な研究を解消してしまうとか、検証型と同じだっていうようなことにはならないんじゃないかっていうふうには少し考えております。

#### 4. 3. 「探索」だけでは研究の蓄積にならない

次の批判は、探索だけでは研究の蓄積にならないというものです。これは、先ほどの荻野綱男さんが2002年に国語学会（現、日本語学会）の機関誌『国語学』の「語彙研究の新分野」という特集の中で書かれたもので、「計量言語学の観点から見た語彙研究」というものですね。語彙研究にはテキストを扱うものと、人間を対象としたものがあるっていうことで、(16)は、そのうちの人間を対象とする語彙研究の中の一節です。

(16) しかし、これらを用いた論文も、研究の基本的な態度としては、調査した結果をデータとして、その中をあれこれ探りながら何かを引き出そうとしているスタイルであることには変わりがない。(略)

調査スタイルでないものという、文献[132, 134, 135]に見られる実験スタイル

ということになろう。こちらは事前に仮説があり、それを検証することを目的として実験計画を立案し、要因をコントロールする（要因以外は結果に影響がないように工夫する）実験をしている。したがって、最後には有意差をきちんと計算し、仮説の検証ができたかどうかははっきり示している。このスタイルは、心理学分野に多いようで（論文が発表される雑誌、その論文が引用する文献の分野から判断できる）、言語学分野には、あまり見られない。

こうしてみると、人間を対象にした研究の問題点が浮かび上がってくる。全体があまりに「調査」的だということである。（略）実験の場合は、研究仮説とその検証という形で目的や手段をはっきりと意識するのに対し、調査の場合は、そのような仮説や問題意識が少なしさらには、理論や法則、過去の研究の流れなどと自分が行おうとする研究との関連が意識されることが少ない。

もっとも、そのような調査に携わってきた研究者の言いたいことはよくわかる。いろいろな研究を行う第一歩として、まずは何かの調査をして、分析の手がかりや糸口をつかむ必要があったのだということである。新しい研究方法が開拓され、自分の課題に関しても可能性が広がりつつある。そういうときだからこそ、新しい方法を試したい。今までそんな研究は行われてこなかったではないか。そういう気持ちはよくわかる。しかし、そういうことを行ってきたために、多くの研究が成果の積み重ねにならず、そのたびに新しい地平で開始されなければならなかったのではないか。結果的に研究の進展や蓄積が不十分のままになってきたのではないか。（荻野 2002 : 106）

荻野さんのおっしゃる調査スタイル、つまり探索的なアプローチは、それだけで終わったら研究の蓄積にならない、それを解決するためには、実験的なスタイル、つまり確認的なアプローチ、仮説検証というスタイルでやらないと、研究が蓄積されず、次に発展していかないという、そういう批判です。

実は、探索的アプローチと確認的アプローチは、双方相補って進めていくものだっていうことが探索的データ解析の方でも言われていて、それとは別に、探索的データ解析の創始者のテューキーは(17)のように書いています。

(17) 刑事司法のプロセスは、証拠の探索——アングロサクソン系の国では警察やその他の捜査機関の責任——と、証拠の強さの評価——陪審員や裁判官の問題——とに明確に分かれている。データ解析においても同様の区別が役に立つ。探索的データ解析は刑事的性格を持つ。確認的データ解析は、司法または準司法的な性格を持つ。（略）

探偵が手がかりを見つけられない限り、裁判官や陪審員は何も検討することができない。探索的データ解析が兆候（通常は量的なもの）を発見しない限り、確認的データ解析が考慮すべきことは何もないであろう。

実験やある種の計画的な調査は、このルールにいくつかの例外や部分的な例外を与える。実験や調査の一部として一連のデータ解析が計画されているからである。しかし、このような場合でも、計画された分析に自己を限定すること、つまり探索を伴わないことで、最も興味深い結果を見失うことがあまりに頻繁に起こる。

すべての探偵小説が思い起こさせるように、犯罪を取り巻く状況の多くは偶然であったり、誤解を招いたりするものである。同様に、データから見出される兆候の多くも、偶然であったり、誤解を招くものであったりする。犯罪の発見においても、データ解析においても、すべての兆候を決定的なものとして受け入れることは、破壊的に愚かなことである。しかし、いくつかの、あるいは大部分は偶然にすぎないという理由で、すべての兆候を収集しないことは、適切な処罰に値する（そしてしばしば処罰を受ける）重大な過失である。

探索的データ解析がすべてであることはありえないが、最初の一步として、基礎となる石となるものは他にない。

(Tukey 1977 : 1-3)

実験計画に従って確認的なアプローチをとるにしても、探索的なアプローチと組み合わせてやっていかないと、仮説以外のことに目が行き届かないわけだから、大事なことを見失ってしまうことがあり得るってということで、荻野さんもおっしゃっていますが、調査スタイルと実験スタイルっていうのは、本来、このチューキーが書いているように、相互補完的に組み合わせられるものなんだろうと思います。

#### 4. 4. 探索的アプローチには定まった筋道・方法がない

次の批判は、探索的アプローチには定まった筋道、方法がないというもので、これも上で紹介した荻野さんの2002年の引用ですね、(18)です。

(18) 人間を対象とする調査・研究は、いろいろな面について長年の経験がある。だからこそ、きちんとした研究計画を立てて、今後の人が追試しなくてもすむような、決定的な成果を上げ、それを積み重ねていくことも必要なのではないか。ある意味で余分ともいえるような中途半端な「調査」を重ねるよりは、決定的な証拠を突きつけるような「実験」が求められているのではなからうか。(略)

調査的研究が不要になったわけではなく、まだまだ開拓されなければならないが、一方では、場当たりの「調査」だけでなく、もっとすっきりした形の実験的研究が盛んに行われなければならないだろう。そのようなことを考慮すると、現状の(日本語学分野の)大学教育のカリキュラムで、そのようなことにも対応できるのか、心配の種は尽きない。



ともかく、「実験」的なアプローチについて十分意識できるような人材が生まれ、育ち、次世代を教育するようにしないことには、明るい展望が開けないのではなかろうか。(人間を対象とする)計量語彙論にとっては、ここが一番の問題点であるように思う。  
(荻野 2002 : 106-108)

このまま探索的研究だけやっていると駄目だっていう、そういう話ですが、それは、(18)にあるように「すっきりした形の」定まった筋道、方法がないからです。ただ、これは探索的データ解析でも言われてるんですね。(19)の引用は、渡部洋さんほかの『探索的データ解析入門』の「はじめに」に書かれているものです。

(19) 以上のような観点から展開され発達してきた探索的データ解析は、統計的仮説検定などを中心としたいわゆる確認的データ解析と対立するものであるというよりは、Tukey 自身も強調しているように、むしろ相互に足りない所を補い合うという関係にある。探索的データ解析で得られた洞察は、別個のデータを用いた確認的な解析で実証される必要も多いであろうし、また探索的に得られた成果は研究を次の段階へ進めるのに役立てられなければならない。探索的データ解析の筋道は明確に定まっているものではなく、任意的であり、それだけに自由に利用できる便利さはあるが、用いられた手法の評価は、比較的長期の研究の流れの中で、またその領域で積み上げられた知見に照らし合わせてなされるべきものであろう。  
(渡部[他]1985 : 4)

探索的データ解析というものは、先ほどテューキーが、一般的な道具とか理解とか、そういったものをきちんと示しているように、定まった手法を何も提示していないってことではないんですよ。ないんですけど、推測統計学のように仮説を統計的に検証する筋道が、〇〇検定とか言ってきちっと決まってるわけじゃないんですよ。それが本当に「探索」として有効なのかどうかっていうのはまだわからない面があって、だからこそ任意的であり、自由だ。だけどそれがむしろ危ういっていうか、そういう批判にもなってるわけです。

ですから、今日のこの講義の目的にもあるように、探索的コーパス言語学についても、その筋道、方法を明確にしていくという作業が必要になるっていうことですね。これは、さっき言いましたように、全文コーパスを使うとか、いろんなコーパスを作ってその可能性を考えてみるとか、それからこの探索的データ解析の手法で探索的な統計をやってみるとかっていうことなんですけど、まだまだそんなことでは済まないっていうか、もっともっと突き詰めていかな

きやいけない。突き詰めてはいかなきやいけないんだけど、一体どんなものがあるんだろうかっていうことですね。

## 5. 探索的アプローチにおける新たな方法の模索

### 5. 1. 「驚くべき（面白い）事実」の発見

それについて最後に、探索的アプローチにおける新たな方法の模索ということで、「驚くべき事実の発見」っていう、ちょっと変な見出しなんですけど、(20)は、探索的コーパス言語学の16から17ページに注として書いたことです。

(20) データ主導型のアプローチは、仮説形成と密接にかかわるものの、仮説形成そのものではない。仮説形成の思考法についてはC・S・パースの「アブダクション」が知られているが、米盛(2007: 53-54)によれば、それは次のようなものである。

探究という科学的行為はある問題状況に直面し何らかの疑念を抱くようになるとき、その疑念に刺激されて生じます。つまり探究はわれわれの信念にそむくある意外な事実を観察したり、あるいはわれわれの期待の習慣に反する何らかの変則性に気づくことから始まります。そして探究の目的は、その意外な事実や変則性がなぜ起こったかということについて、その理由または説明を与えることによって、われわれの疑念を合理的に解決することにあります。このように、ある意外な事実や変則性の観察から出発して、その事実や変則性がなぜ起こったかについて説明を与える「説明仮説」(explanatory hypothesis)を形成する思惟または推論が、アブダクションです。このアブダクションの推論の形式を、パースはつぎのように定式化しています。

驚くべき事実Cが観察される、

しかしもしHが真であれば、Cは当然の事柄であろう、

よって、Hが真であると考えるべき理由がある。

ここで「驚くべき事実C」というのはわれわれの疑念と探究を引き起こすある意外な事実または変則性のことであり、「H」はその「驚くべき事実C」を説明するために考えられた「説明仮説」です。(米盛 2007: 53-54)

データ主導型のアプローチが、このこのアブダクションという推論形式の中に位置づけられるとすれば、それはその出発点である「ある意外な事実や変則性の観察」(パースの定式化では「驚くべき事実の観察」)にとどまり、「説明仮説」の形成にまで及ぶものではないと考えられる。ただし、科学哲学の世界では、こうした「事実の発見」は「偶然」によることが多く、法則や理論の発見のように「目的志向的な自覚的探究の結果」ではないとされるが(伊東 1981: 244)、データ主導型のアプローチにおける「探索」とは、どのようなデータであっても、そこにどのような構造が隠されているかを、データそれ自身に語らせるように探り出す行為であり、それは目的志向的・自覚的な「事実発見」の作業と言えるだろう。(『探索的コーパス言語学』, pp. 16-17)

探索的データ解析が、この仮説形成の推論の形式の中で、そこで発見された事実を仮説に組み上げて検証しようと思っていることは、結果的には十分ありうることなんだけど、探索的アプローチが最も中核としている部分っていうのは、この最初の驚くべき事実の観察っていうところですね。ここに注力しようっていうのが、多分探索的なアプローチであって、一般的には偶然によることが多いと言われていることを目的意識的に自覚的に行う、そういう方法を探索的データ解析は用意しようとしているんじゃないか、と思うわけです。

## 5. 2. マルチメディア・コーパスの映像KWIC

そして、探索的コーパス言語学もそのような方法が用意できたらというのが私の希望ですね。そこでどんなことが考えられるかっていうことの一つの例として、マルチメディア・コーパスの映像KWICというものを紹介したいと思います。

KWICに関しては、先ほどの田野村先生のご講義の中にも出てきていて、私も田野村先生がお作りになった“KWIC”という名前のソフトですね、これを使わせていただいている、私などは新聞のコーパス使うことが多いんですけども、例えば新聞1年分のコーパスのKWICをある単語とか形式とかについて作るっていうのは、非常に簡単ですね。田野村先生の“KWIC”を使うと、前文脈で並べたもの、後文脈で並べたもの、キーを含む後文脈で並べたものって、三つ出てくるんですよ。本当に便利で、例えば「まとめる」という動詞のKWICを作って前文脈で並べたものを見ると、「まとめる」の直前にどんな助詞が来るかっていうのが一覧できるわけですね、ザーッと並ぶから。そうすると、「〇〇をまとめる」っていうのはもちろん多いんだけど、「〇〇を〇〇にまとめる」っていうのもあるんですよ。「考えを論文にまとめる」とかそういうものがありますね。その中の一つに、新聞だったからなんですけど、「5回を2失点にまとめる」ってのがあったんですね。これは野球の話で、先発ピッチャーが最初の5回を失点2点に抑えて降板したという意味で、これをどうして「まとめる」っていうんだろう、面白いなと思うんです。あるいは野球などのスポーツ報道特有の言い方かもしれませんが、それでもやっぱり「〈対象〉を〈結果物／生産物〉にまとめる」という連語の一つのバリエーションな

んだとは思いますがね。

でも、私などはそういうことを考えていると、つい話が現実の雑な方に行ってしまうと、一体何失点に抑えたら「まとめる」と言えるのだろうかとか、何回まで抑えたら「まとめる」と言えるのだろうかとか、そんな方が気になって、やっぱり4回じゃちょっと「まとめた」とは言いにくいから5回か6回か。でも、8回9回までいっちゃうと「まとめる」とはちょっと言いにくくなっちゃうなとかね。1点だと上出来で、2点から4点ぐらいまではまとめているけど5点取られたらもう駄目、試合を壊したなとか、そういう現実の方に行ってしまうので、肝心なその構文の分析が全然できなくなっちゃって本当に我ながら情けないんですけど。

いや、言いたかったのは、KWICっていうのがそういうことをいろいろ思いついたり発見したりするのにとてもいいものなんですよ。本当に探索的な発見のツールとして、とてもよくできてると思います。これは従来のコーパス言語学でも使われている定番の方法だと思いますね。それに類するものとして「映像KWIC」というものを、もうだいぶ前になりますけど、今日もいらしてくださってる孫栄爽（ソン・ヨンソク）さんと2人で共同で発表したことがあるので、そのことをちょっとご紹介したいと思います。2016年に日本語学会でブース発表したんですけど、(21)はそのときの予稿集からの引用ですね。

(21) 従来、話し手の言語行動における言語表現と非言語表現との関係の研究は、意識調査や実験観察的調査に基づくものが多かったが、発表者はより実証的な調査研究の必要性を感じ、手始めにテレビ放送における言語行動を大量に記録した「マルチメディア・コーパス」を複数作成して、そのパターンを見出すべくいくつかの分析を試みてきた（石井 2009, 石井・孫 2013）。

ただし、発表者が作成したマルチメディア・コーパスは、話し手の発話を文字化したテキストとその発話場面の映像とを同期させ、文字化テキスト上の文字列（単語など）を検索するとその発話時の映像・音声も参照できるように、表計算ソフト“Excel”上に実装した簡易なコーパスであり、いわゆる「マルチモーダル・コーパス」（金子・石崎 1998）のように、非言語行動に関する詳細なアノテーション（タグ付け）を行ったものではない。それは、言語行動を構成する要素・側面やそれらの関係には未解明な部分が多く、分析者が予見した観点＝タグからのみ分析してしまうと、それ以外の部分が見えなくなり、重要な部分を見落としてしまう恐れがあると考えたからである。

こうしたことから、発表者は、マルチメディア・コーパスを使い、注目した言語項目の使用例を検索し絞り込むと、一つ一つの用例の映像・音声を確認して、その言語項目と有意に関係する非言語行動の要素・側面にどのようなものがあるかを検討するという作業を継続してきた。その結果として、これまでに、擬音語・擬態語と身振り、終助詞ネと視線、指示詞と指差し、といった関係を見出し、それらの間にあるパターンを発見してきた（孫 2010a・2010b, 石井・孫 2013）。

しかし、こうした作業は、一つ一つの用例についてその都度映像・音声を参照するキー操作を行わねばならず、せっかく同一の言語項目を絞り込んでいるのに思考が中断され、その言語行動に共通する要素・側面を直観的にとらえることが容易ではなかった。そこで、発表者は、マルチメディア・コーパスの一つである「対談番組のマルチメディア・コーパス」（孫 2010a）において、言語項目＝キーワードを検索して前後の文脈順に配列する、すなわち KWIC 形式に選択・配列した後、任意の用例上でキー操作を行うと、それ以降の用例の映像・音声が自動的に連続して再生される「映像 KWIC」とも呼べる機能を、“Excel”のマクロ機能を使って作成・付加した。

これにより、分析者は、同じキーワードが発話される場面の映像を次々に見ることになり、それらに共通する「何か」を直観的に把握する可能性が高まる。たとえば、後述するように、「～と思います」という発話は多くの場合うなずきを伴うが、その生起には引用節末の肯否の形式も関係しているようだとか、話し手が「やっぱり」と発話して改めて考え直す様子を示す場合には、その前後にポーズが置かれたり、相手に視線を向けなかったりするようだとか、といった「気づき」である。

（孫・石井 2016：165-166, 引用中の文献参照は省略）

言語行動の研究で、言語形式と非言語表現との関係を探索する場合に、こういった映像 KW I C っていうものが使えるのではないかということなんですけれども、これは要するに KW I C っていうのは、連続して並べて見るっていうところが、この直感的な探索、発見ができるいいところで、今までのやり方だと、1 行 1 行ごとに再生して見て、また次の 1 行再生っていう、そういうことで連続して見るってことができない、ちょっと途切れちゃうんですよね。そこで、なんとか連続再生できないかっていうことで、孫さんにお問い合わせしたら作ってくださいって。それは、こういう言語行動の研究にも使えるんだけど、単純にですね、話し言葉に多い言語形式についても、この映像 KW I C で連続再生すると、そこから何か発見があるかもしれないと思って、今日最後に、残り 20 分弱ですけど、連続再生の実例をご覧に入れようと思います。

それで、話し言葉に特有の形式というので、感動詞の「あっ」というものがテレビの対談番組でどういうふうに使われてるかを見てみたいと思います。こ

ここでは、実際に検索したところを画面キャプチャーで録画してきたので、今からそれをお見せします。

(映像KWICによる連続再生の紹介。具体例は省略)

すいません。これずっと長く続くので、ちょっと聞き取りにくいところ、ここで「あっ」って言ってるのかがわからないっていうのもあったと思うんですけど、これをずっと見ていってですね、要するに一番はっきりした特徴は、対談番組で、相手の発話が終わって応答するときに、「そうなんですか」とか「そうですか」とかっていう応答に前接する「あっ」ですね。それと、そうではなくて、自分が1人でこう語ってるんだけど、その中で、自分のこととして過去に何らかのそういう気づきとか発見とかがあったときに、その語りの中で「あっ」で言っている。要するに、一人称の語りの中で現れる「あっ」ですね。この場合、いろんなパターンがあるんですけど、いくつかあったパターンは、「何々したら」とか、「何々すると」とか、「何々したときに」って言って、何かある状況とか、ある刺激を与えられたとき、自分がそこでハタと気づいたことを、この「あっ」に続いて述べるっていうような例が数多くあったと思います。そうじゃないものももちろんありますけどね。

そういう一人称の語りの中で、「あっ」と、自ら気づいたことを自ら表現するのは一体どういうことなんだろう。それは、リアルな、実際の気づきとか驚きっていうのとは、何か違いがあるんだろうか、それとも同じなんだろうかとかね。そういうことも、ここからいろいろ考えてみたら面白いんじゃないか。いや、語りの中のこういう自らの過去の経験を表現するときのこの「あっ」っていうのは一体どんなものだろうかっていうのは、映像KWICっていうものを見ることによって、いろんなことに、それこそ気づくことができたらいんじゃないか。

こういうことが探索的コーパス言語学の方法と言えるかどうかはまだまだわかりませんが、なにかいろんな自覚的、目的指向的な発見とか気づきっていうのはどんなものなんだろうかっていうことをこれからも考えていきたいと思います。結局突き詰めきれずに終わりますけれども、このような形で、これからも、そもそも「探索」とは何かっていうようなことまで少しずつ考えが及ぶよ

うに進めていけばいいなとは自分なりには思っておりますけれども、またいろんなところでお目にかかると思いますので、その際には忌憚のないご意見をいただければと思います。今日は本当にありがとうございました。

## 引用文献

- 石井正彦(2019)『探索的コーパス言語学—データ主導の日本語研究・試論—』大阪大学出版会
- 荻野綱男(2002)「計量言語学の観点から見た語彙研究」『国語学』208, 97-115
- 荻野綱男(2021)「言語統計学入門(1)—統計的な見方と言語研究—」『計量国語学』33(1), 28-39
- 孫 栄爽・石井正彦(2016)「映像KWICによる言語行動の直観的探索—対談番組のマルチメディア・コーパスを例に—」『日本語学会 2016 年度春季大会予稿集』, 165-170
- 宮島達夫(2007)「語彙調査からコーパスへ」『日本語科学』22, 29-46
- 吉田 忠[編](1995)『現代統計学を学ぶ人のために』世界思想社
- 米盛裕二(2007)『アブダクション 仮説と発見の論理』勁草書房
- 渡部洋・鈴木規夫・山田文康・大塚雄作(1985)『探索的データ解析入門—データの構造を探る—』朝倉書店
- Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, 14(3): 275-311.
- McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press. (=2014, 石川慎一郎(訳)『概説コーパス言語学—手法・理論・実践』ひつじ書房)
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.
- Whiteside, K. (2016). *A corpus-driven investigation into the semantic patterning of grammatical keywords in undergraduate History and PIR*

*(Politics and International Relations)*. Unpublished PhD Thesis,  
University of Warwick

## 付記

本稿は、2024年2月14日（水）15時30分～17時に、大阪大学豊中キャンパス文法経講義棟・文41講義室において行われた石井の退職記念講義（最終講義）の講義録である。講義に向けてご尽力くださった基盤日本語学講座の教員各位、学生諸君、関係者の皆様、そして、当日ご参集くださった皆様に感謝申し上げます。

（人文学研究科教授）