様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２８　年　　６　月　１４　日現在

機関番号：　１３９０１
研究種目：　基盤研究(C)（一般）
研究期間：　2013～2015
課題番号：　２５３７０５４９

研究課題名（和文）コーパスデータの信頼性とI言語研究資料としての利用の妥当性に関する考察

研究課題名（英文）On the Reliability and Validity of the Use of Corpus Data as Evidence for I-Language Studies

研究代表者

大名　力（Ohna, Tsutomu）

名古屋大学・国際開発研究科・教授

研究者番号：００２３３２０５

交付決定額（研究期間全体）：（直接経費）　　1,400,000円

研究成果の概要（和文）：特別な知識や技術がなくても利用可能なユーザーフレンドリーな環境の整備・普及は，コーパス研究の促進に寄与すると同時に，様々なレベルでブラックボックス化を進める要因の１つにもなっている。研究代表者の前研究課題では，データ抽出方法，統計処理方法等について検証し，ブラックボックス化の裏で起きている問題を具体的に指摘したが，ブラックボックス化，非明示性の問題は用語・概念，手法，方法論などにも及んでおり，より広い範囲で多角的に検討する必要がある。このような状況を踏まえ，本課題では，研究資料としてのコーパスの利用に関する用語・概念，手法，方法論について，I言語研究の観点から明示的・体系的に検討を行った。

研究成果の概要（英文）：The development and spread of user-friendly environments, which has made it possible for researchers to use corpora with no special knowledge or skills of text processing and statistics, have contributed to the popularity of studies based on corpora. At the same time, however, this has made corpus studies akin to a "black box," as researchers sometimes use corpus data to provide evidence for their theories and hypotheses without examining the reliability and validity of their use. In this project, I conducted research on the reliability and validity of the use of corpus data as evidence for I-language studies, by examining not only corpora and the tools used to process them, but also the terminology, concepts, methods, and methodologies often used in corpus studies.

研究分野：　英語学

キーワード：　コーパス　言語能力　経験科学　科学哲学　妥当性

１．研究開始当初の背景

前課題（平成 22～24 年度科学研究費補助金「言語研究資料としてのコーパスデータの客観性と信頼性に関する考察」基盤研究(C)，研究代表者 大名力）では，言語研究においてコーパスの利用が拡大しているにも関わらず，コーパスデータの客観性と信頼性に関して体系的な調査・考察が行われていない現状を踏まえ，コーパスからのデータ抽出方法，統計処理方法の妥当性を検証し，コーパスデータの客観性・信頼性の確保のために必要な条件の検討を行った。その結果，例えば，コロケーションの研究で共起性の指標としてよく用いられてきた t-score，MI-score のような指標でも，研究者やプログラムの間で使用している計算式に違いがあり，数値を直接比較することができないことがあることが認識されていないなど，研究者間での知識・技術の共有が不十分であるだけでなく，そのことにより問題が生じていること自体がほとんど認識されてないことが，研究発表，講演会等での参加者の反応からも明らかになってきた。

このような状況を生み出している原因の 1 つとして，所謂“ユーザーフレンドリーな”環境の整備・普及が挙げられる。コーパス，文字コード，テキスト処理，統計学などに関して，特別な知識や技術がなくてもコーパスが利用できるユーザーフレンドリーなツールの普及は，コーパス研究の促進に大いに寄与してきたが，しかし同時に，コーパス，処理内容，手法・方法論のブラックボックス化を進める要因にもなっている。ユーザーフレンドリーなツールを使って検索すれば何らかの結果は得られるが，その結果の適否の判断は研究者自身が行わなければならない。適切な判断のためには，入力・処理・出力（処理対象・処理内容・処理結果）の３点をセットとして考える必要があるが，ユーザーフレンドリーなツールでは，入力と処理の部分がユーザーから隠されてしまうため，出力の正しさの検証が難しくなるだけでなく，そもそも，検証の必要性自体が意識されにくくなるという問題がある。

このような状況においては，ユーザーフレンドリーな環境の整備・普及によりコーパス研究がさらに盛んになっても問題の解決には繋がらないため，ホワイトボックス化は無理であったとしても，現状の改善のためには，少なくとも外部から中身が見えるグラスボックス (glass box) にしていく必要があり，そのためには，コーパスの中身，処理内容，手法・方法論の明示化と体系化が不可欠であり，さらには，知識や技術を共有するのみならず，その必要性自体が研究者の間で共有される必要がある。

２．研究の目的

コーパスの利用を謳った研究でも，データとして提示されている数値を実際に該当コーパスでチェックしてみると誤っていることもあり，正しい数値で検証してみると提示されている仮説が支持されないケースもある。もし，誤ったデータで“正しい”仮説が“支持される”とされていたとすると，数値そのものの正しさだけでなく，研究資料としてのコーパスの扱い，また，使用されている“コーパス言語学的手法”も検討する必要が出てくる。

心理測定や教育測定の分野では，「信頼性」(reliability) は一貫性・安定性，「妥当性」(validity) は計りたいものが計れている程度を表すが，コーパスデータに関しても，同様の観点から検討する必要がある。研究において「計りたいもの」は直接観察可能できない理論的構成物 (theoretical construct) であることが多く，測定対象をどう捉えるかは理論に大きく依存するため，理論の重要性は大きい。「理論言語学対コーパス言語学」のように対立的に捉えるのではなく，対象を明確化しモデルを立て，その中でコーパスデータはどのような位置付けを与えられるのか，また，コーパスデータの信頼性・妥当性を高めるためにはどうすればよいかを考えていく必要がある。

「１．研究開始当初の背景」で述べたように，ユーザーフレンドリーな環境の普及とともに様々なレベルでブラックボックス化が進行している現状を踏まえ，本課題では，言語研究（特に人の心の中の言語知識を対象とする I 言語研究）のための資料，証拠としてのコーパスデータ利用の可能性と問題を探るために，データの抽出方法，統計処理方法等だけでなく，コーパスを利用した研究において用いられる用語・概念，手法，方法論について，明示的・体系的検討を行うことを目的とする。

３．研究の方法

平成 25 年度は，前研究課題「言語研究資料としてのコーパスデータの客観性と信頼性に関する考察」の研究成果を基に，コーパスデータの信頼性の問題について整理を行い，本課題で扱うべき用語・概念の検討を行う。具体的には，「演繹法・帰納法」「言語能力・言語運用」「合理論・経験論」「経験的」など，コーパスを用いた研究における方法論について論じられる際によく用いられる用語を取り上げ検討する。また，コーパス研究でよく見られる定量分析の基礎となる「頻度」の意味についても検討を行う。

平成 26 年度は“コロケーション”について検討する。「語と語の慣習的な結び付き」とされる“コロケーション”が，多種多様な共起関係を示す用語として用いられていることを示し，そのような共起を引き起こしている原因を明らかにする。また，“コロケーション”の定義に関わる問題として，概念的定義，操作的定義，証拠の区別・扱いについても検討し整理を進める。

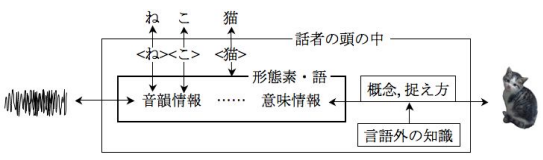
最終年度である平成 27 年度は，I 言語研究のための資料としてのコーパスデータという観点から，「経験論」「合理論」「言語能力」「言語運用」「I 言語」「E 言語」「確率」「確率(論)的」など，コーパス研究に関わる様々な用語・概念の整理を試みる。

４．研究成果

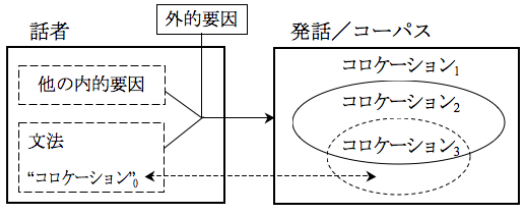
初年度の平成 25 年度は，コーパスデータの信頼性の問題について整理を行い，本課題で扱うべき用語・概念の検討を行った。具体的には，「演繹法・帰納法」「言語能力・言語運用」「合理論・経験論」「経験的」など，コーパスを用いた研究における方法論について論じられる際，よく用いられるが，多義または曖昧で研究者によって異なる意味で用いられていたり，誤った意味で使われていたりすることが少なくない用語を取り上げ検討した。また，コーパス研究でよく見られる定量分析の基礎となる「頻度」の意味についても，I 言語研究という観点から検討を行った。

当初の計画では，２年目の 26 年度に“コロケーション”に焦点を当て問題点を検討する予定であったが，日本英文学会中部支部第 65 回大会で，コロケーションをテーマとしたシンポジウムが企画され，研究代表者がその司会となることとなったため，前倒しでコロケーションに関する検討も進めた。大会では「“コロケーション”研究の諸相」と題してシンポジウムを行い，その司会を務めると共に，「“コロケーション”を解体し，統合する」という題目で発表も行った。研究発表では，「語と語の慣習的な結び付き」とされる「コロケーション」が，多種多様な共起関係を示す用語として用いられていることを，英語および日本語の具体例を基に示し，共起の事実を指摘するのに留まらず，そのような共起を引き起こしている原因を明らかにすることの重要性について論じた。

平成 26 年度は前年度に引き続き“コロケーション”を中心に検討し，このような多種多様な共起関係にある語の組み合わせがまとめて“コロケーション”として扱われる背景を探った。具体的には，語（形態素）はある種のインターフェースとして働くため，語（から構成される句）を介して結び付けられる様々な要素（意味・形式・指示物等）の間の共起関係が，「語と語」の共起関係として捉えられ，それらが区別せずに扱われやすいことを示した。



また，異なる種類の共起関係が心的実在物としての文法(I 言語)の中でどのように位置付けられるものであるかについても検討した。さらに，研究者によって，概念的定義，操作的定義，証拠の区別・扱いが曖昧であったり，混在しているケースがあることも，“コロケーション”の定義を考える際の混乱の原因の１つとなっていることを踏まえ，これらについても検討し整理を進めた。



コロケーション$_1$: 単なる語の組み合わせ，2, 3を含む
コロケーション$_2$: 共起性の高い語の組み合わせ
コロケーション$_3$: 特定の関係にある語の組み合わせ，狭義のコロケーション

最終年度である平成 27 年度は，I 言語研究のための資料としてのコーパスデータという観点から，コーパス研究に関わる様々な用語・概念の整理を試みた。コーパス研究のパラダイムと言語モデルについては，主に，Leech (1992) "Corpora and theories of linguistic performance" の挙げる３つのパラダイムを基に，現在のコーパス研究におけるパラダイムについて検討した。また，コーパス研究において「言語能力」「言語運用」「普遍文法 (UG)」「I 言語」「E 言語」などの基本的な用語が誤った意味で用いられていることが少なくないことから，Chomsky (1965) *Aspects of the Theory of Syntax*, Chomsky (1986) *Knowledge of Language: Its Nature, Origin, and Use* 等での用語の意味を踏まえ，言語能力と言語運用，言語知識の内容，言語能力・言語運用と言語の社会性など，コーパス研究において見られる言説について内容を検討し整理した。コーパス研究で用いられている「確率」「確率(論)的」等の用語は，一見，曖昧性のない用語・概念のように思われるが，実際の論文では異なる意味で使われることがあることを指摘し，生起確率による予測の問題と現象の原因の区別，文法性と生起確率，習得可能性などの観点から整理を行った。また，「経験論・合理論」については，研究対象と研究者，言語習得の問題，言語知識の範囲と内容の問題，言語研究の問題，研究資料の問題に分け検討した。

これらの成果の一部を収録した論文「コーパスと生成文法」，「I 言語研究とコーパスデータ」(仮題)は，それぞれ，英語コーパス研究シリーズ第７巻『コーパスと多様な関連領域』(赤野一郎・堀正広編，ひつじ書房)および『コーパスからわかる言語変化・変異と言語理論』(仮題)（小川芳樹・長野明子・菊地朗編，開拓社，2016 年 11 月刊行予定)に掲載される予定である。(原稿を出版社に提出済みであるが，本報告書作成の時点で初校が出ていないため，下記の「５．主な発表論文等」

には掲載していない。）
　また，2015 年 4 月には，前年度までの検討内容およびそれまでの教育経験を踏まえ，平成 27 年度英語コーパス学会春季シンポジウム「コーパス関連専門科目の授業内容について」において，「コーパス研究の一方法論と情報教育としてのコーパス研究教育」と題して，コーパス研究教育に関する発表を行い，本課題の成果の一部を公開した。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 0 件）

〔学会発表〕（計 2 件）
①大名力 (2013)「“コロケーション”を解体し，統合する」日本英文学会中部支部第 65 回大会 シンポジウム「“コロケーション”研究の諸相」(司会および発表)，2013 年 10 月 6 日，椙山女学園大学星ヶ丘キャンパス

②大名力 (2015)「コーパス研究の一方法論と情報教育としてのコーパス研究教育」平成 27 年度英語コーパス学会春季シンポジウム「コーパス関連専門科目の授業内容について」，2015 年 4 月 25 日，関西大学

〔図書〕（計 0 件）

６．研究組織
(1)研究代表者
大名力　（Tsutomu OHNA）
名古屋大学・大学院国際開発研究科・教授
研究者番号：00233205

(2)研究分担者
なし

(3)連携研究者
なし