

論 文

アカデミックコーパスにおける Coherence 構築のストラテジー – Science の Discussion における Information Order の検証 –

中谷安男・土方裕子・清水眞

Abstract

This article explores how scientific articles develop coherence in Discussion sections. Focus is on how information order is organized in individual paragraphs. The findings are based on quantitative corpus analyses on 100 research papers of *Science*. The results indicate that the writers begin clauses with given information defined as a theme and end them with new information defined as a rheme. The theme of each clause is connected with the topic sentence or the information presented previously in the paragraph. Such information order and chaining structure of themes can maintain the flow of discourse and develop the consistency of the paragraph. When the theme is not located at the beginning of a sentence, metadiscourse tends to be used to get attention of readers and guide them to follow writers' intentions.

1. はじめに

近年、日本の大学においてもアカデミック・ライティング (Academic Writing: AW) 指導の重要性が認識されている。AW は英文のエッセイや修士・博士論文に必要な技能である。また最終目標はインパクト・ファクター (Impact Factor: IF) を持つ世界的学術誌に掲載されるための論文執筆のストラテジー習得とも言える。この具体策の調査結果として Gosden (1993) は、科学分野における論文査読者が英文において最も重要視する以下の2項目を提示した。

- 1) “logical and clear linking of sentences for the reader”
 - 2) “development of the topic from sentence to sentence in a coherent way”
- これは、論理的に明確で分かりやすく、各文に連続性があり、各主題の基に「談話の一貫性 (Coherence)」を持つことである。さらに具体例として、IF を持つ学術雑誌に掲載を認められた論文の評価に以下のようなコメントがあった。

『英語コーパス研究』第18号 (2011), pp.1-16

法副詞 *no doubt* の機能について

—類義表現との比較から—

鈴木大介

Abstract

This study attempts to shed light on the modal adverb *no doubt* through a comparison with two other synonymic expressions: *doubtless* and *undoubtedly*. It is well known that these three adverbials fall into the same semantic category and express a speaker's judgment about the certainty or probability of a proposition. However, the distinctive function of *no doubt* has not been provided in the literature. This paper, therefore, shows the factors that are significant in predicting the use of *no doubt*, by analyzing corpus data. For this purpose, I present three factors to be considered regarding these adverbs: (i) whether or not they co-occur with modal verbs; (ii) in what position they occur; initial, medial, or final; and (iii) whether the subject in the clause where they appear is a pronoun or full NP.

1. はじめに

本研究では、英語の法副詞である *no doubt* について、実際に使用された言葉を対象とし、機能論的な観点からの分析を行う。単独での *no doubt* は、(1a, b) で表されるように、文副詞として用いられる。

- (1) a. It was *no doubt* clever of him to offer his resignation at that point in the proceedings. (Quirk et al. 1985: 622)
b. The thing *no doubt* would have happened differently to another man. (FICT) (Biber et al. 1999: 866)

さらに、「疑いがない」という文字通りの意味ではなく、*certainly*, *probably* といった蓋然性の意味を有するものと同様に、話者のスタンスやモダリティを表す副詞類に分類されている。

しかし、そもそも形式的には *no doubt* という名詞句であるため、Biber et al. (1999: 866) で例外的と述べられているように、他のスタンスやモダリティを

『英語コーパス研究』第18号 (2011), pp.17-31

英文エッセイ評価における指標としての接続語句

北村まゆみ・杉浦正利

Abstract

This article examines the effects of connectives on the quality of essay writing by Japanese EFL university students and the validity of a measure based on the diversity of connectives. Using a decision tree analysis, a total number of 134 essays written on two topics from a corpus written by Japanese university students were analyzed. Correlation analysis was also carried out to investigate whether the measure based on the diversity of connectives was valid. The results showed that connectives were not a significant predictor of writing quality. In addition, no significant correlation was found between the diversity measure of connectives and essay ratings. Through the discussion, we concluded that due to the overuse of connectives by lower level Japanese EFL students, the number of the connectives could not be an influential factor in our decision tree analysis, and our suggested measure based on the diversity of connectives would need further sophistication.

1. はじめに

英文ライティングにおいて文章構造を反映する接続語句の使用が重要であると多くのライティングの教科書で従来から指摘されている。また、接続語句の使用はテキストの結束性を高めるが、過剰使用するとテキストの理解を妨げることがあるという指摘もある (Crewe, 1990)。

Granger & Tyson (1996) でも指摘されているように、Halliday & Hasan (1976) の *Cohesion in English* の出版をきっかけとして、これまでに接続語句を含めた cohesive ties とライティング能力の関係に関する研究が行われてきている (Witte & Faigley, 1981; Homburg, 1984; Ferris, 1994; Liu & Braine, 2005; 杉浦, 2008b; Narita & Sugiura, 2009)。また、1990年代から学習者コーパス研究が盛んになり、英語学習者と英語母語話者のエッセイにおける接続語句を過剰使用や過少使用といった観点から量的に比較分析し、英語学習者の接続語句の使用傾向を調査した研究が行われてきている (Granger & Tyson,

『英語コーパス研究』第18号 (2011), pp.33-48

研究ノート

2つの史的コーパス – その代表性と類似性

塚本 聡

Abstract

This research explores the similarity between two historical corpora and their representativeness: PPCMBE and COHA. They contain samples from nineteenth century texts. Both corpora share the most frequent words, and also show the similar word rankings compared with those in twentieth century corpora. Due to abundant sampling of fiction texts, COHA contains an abundance of personal pronouns. The frequencies of words in PPCMBE fluctuate over shorter periods. This signifies the sampling is not evenly distributed in PPCMBE. A sample observation of the adjective paradigm illustrates that the inflectional form became more common during the nineteenth century. Both corpora have limitations on their representativeness.

1. はじめに

2010 年になり、下記の 2 つの史的コーパスが相次いで公開された。

- ・ Penn Parsed Corpus of Modern British English (以下 PPCMBE)
- ・ Corpus of Historical American English (以下 COHA)

Helsinki Corpus から始まった史的コーパス編纂の流れは、Parsed corpora of historical English プロジェクトとして拡大版が公開されてきている。今回公開された PPCMBE は、1700～1914 年のテキスト約 949,000 語で構成されたコーパスで、前述の Parsed corpora of historical English の一部を成すコーパスである。一方、COHA は、Corpus of Contemporary American English (以下 COCA) をはじめとする Mark Davies による無償オンラインコーパスの一つとして公開されたもので、1810～2009 年にいたる 4 億語を収録したコーパスである。

本研究ノートの目的は、PPCMBE と COHA のうち、重複する 1810～1910 年代のテキストを対象とし、語彙構成、人称代名詞の生起数、形容詞の屈折を観察することにより、両者の 19 世紀コーパスとしての代表性、有用性を検討することである。¹

『英語コーパス研究』第 18 号 (2011), pp.49-59

Web を源とした質情報付き英語科学論文コーパスの構築法

田中 省作・柴田 雅博・富浦 洋一

Abstract

This paper describes a method of building a corpus of scientific papers with information on the quality of their linguistic features from the Web. First, the method collects only electronic papers presumably put on the Web by authors, using a search engine and keywords of scientific categories. Second, it evaluates the quality of the linguistic features, not the content, of each paper automatically, and finally, organizes papers as a corpus of scientific papers with information on the quality. The proposed method helped build a corpus of 280 million words in a period of approximately a week.

1. はじめに

英語科学論文に不慣れな日本人が書いた国際会議等の論文には稚拙な誤りはほとんど無いものの、熟達者の論文との間には何かしらの大きな質的隔りがある。このような明示的ではない日本人英語科学論文に共通の特徴を明らかにすることは、日本人の専門英語教育において重要な課題の一つである。我々はこのような課題に対して、近年様々な言語的特徴の解明に成果を上げているコーパスによる究明を試みている。田中他 (2006) は、情報系の国際会議録に掲載されている科学論文の表現の質（以降、本稿でも「論文の質」は内容の質ではなく、このような表現の質を指すこととする）という観点で、良質な論文 768 編と日本人の論文 363 編を収集し、それらを品詞列で対比し、日本人の論文に特有の文法的特徴を明らかにしている。田中他 (2006) で挙げられた特徴は品詞レベルで規定されるものであったので、比較的小規模なコーパスでも信頼性の高い分析が可能であった。しかし、コロケーションや構文パターンといったより複雑・多様な言語的特徴をとらえようとする、取り扱うべき語・品詞・統語構造といった素性の組み合わせ数が膨大となり、現状のコーパス規模では不十分となる。そこで本研究では、膨大かつ多様なデータが存在する Web から英語科学論文を収集し、それらに質情報を付与することで、より大規模な質情報付き英語科学論文コーパスの効率的な構築を実現する。¹ なお、

『英語コーパス研究』第 18 号 (2011), pp.61-71