【11】

氏名	小林 雄一郎（こばやし ゆういちろう）
博士の専攻分野の名称	博士（言語文化学）
学位記番号	第 25000 号
学位授与年月日	平成24年3月22日
学位授与の要件	学位規則第4条第1項該当 言語文化研究科言語文化専攻
学位論文名	**統計的テキストマイニングによる学習者作文におけるメタ談話標識の研究**
論文審査委員	（主査） 教授 岩根 久 （副査） 教授 沖田 知子　准教授 田畑 智司

論文内容の要旨

学習者が実際に使用した言語の特徴や誤用についての知識を持つことは，英語教師にとって重要なことである。第2言語習得の分野では，誤用は，(1) その学習者が対象言語をどれくらい学んだかについての情報を教師に提供する，(2) どのように対象言語が学習されるかについての情報を研究者に提供する，(3) 学習者が正用と誤用に関する情報を基に対象言語の規則を発見できる，という3つの点で有意義であるとされている (Corder, 1967)。

近年，英語教育の現場では実践的コミュニケーション能力の育成を図ることが求められており，中学校や高校の学習指導要領にも同様の記述が見られる。しかし，円滑で効果的なコミュニケーションをするためには，「何を」伝えるかよりも，「いかに」伝えるかが必要不可欠となる (e.g. Conner & Mbaye, 2002; Young, 2002)。具体的に効果的なコミュニケーションを達成する1つの方法は，対比 (e.g. *while, on the other hand*)，理由 (e.g. *because, so*)，結果 (e.g. *therefore, as a result*)，列挙 (e.g. *firstly, secondly*)，例示 (e.g. *for example, in particular*) といった接続語 (connectives) や談話標識 (discourse markers)，あるいはメタ談話標識 (metadiscourse markers) によって，談話のユニット間の論理関係や意味関係を表すことである (Altenberg & Tapper, 1998)。それらの表現には，談話の結束上の「手がかり」を与えて (Leech & Svartvik, 1994)，読み手や聞き手がユニット間の一貫性を見つけることを手助けし，テキストの意味理解ができるようにする働きがある。しかしながら，学習者にとって，談話標識の「適切な」使用は非常に難しい。談話標識の使用数が増えても結束性の質が上がるわけではなく (McCarthy, 1991, p. 50)，むしろ「人工的で機械的な文章」(Zamel, 1983, p. 27) となってしまうことで，テキストの理解を妨げることもある (e.g. Crewe, 1990)。従って，実践的コミュニケーション能力の育成を図る上で，学習者による（メタ）談話標識の使用傾向を調査し，彼らの談話構造における特徴や誤用を究明することは極めて重要である。しかしながら，これまでの研究では，手作業による談話分析のコストが高いこともあって，限られた数の学習者データしか扱うことができず，そこから得られた結果がどこまで普遍的なものかを検証することが難しかった。さらに，大規模な調査を行う場合は，多くの分析者が必要となり，どうしても結果が個々の分析者による主観に影響されてしまうという欠点があった (e.g. Baker, 2006)。

それに対して，本研究では，日本人中学生，高校生，大学生の英作文を集めた「学習者コーパス」(learner corpus) をテキストマイニングの手法を用いて客観的に解析し，そこから得られた結果を様々な角度から比較検討していく。

テキストマイニングとは，テキストデータをコンピュータで計量的に解析し，有益な情報を抽出するための様々な手法の総称であり，統計学，データマイニング，人工知能，自然言語処理で開発された技術を背景に持っている (e.g. Feldman & Sanger, 2007)。テキストマイニングは，大規模なテキストデータを統一的な視点から少ない労力で客観的に分析することを可能にする。現在のコーパス言語学では，インターネット上の膨大な言語データからコーパスを自動生成する技術が導入されつつあり，個人の研究者であっても数億語から数十億語のコーパスを構築することが可能となり，100億を超えるウェブページに含まれる言語データを蓄積している研究グループも存在する (e.g. 田中, 2011)。このような大規模データを前にしたとき，手作業でデータを解析することは極めて困難となる。しかしながら，分析に利用されるデータが大きければ大きいほど，データ縮約やテキスト分類といったテキストマイニング技術の必要性は高まり，その精度も安定していく。従って，テキストマイニングは，言語研究において，分析データの量および分析結果の質を飛躍的に高めるブレイクスルーをもたらす可能性を秘めている。

テキストマイニングという観点から見たとき，本研究の独創性は，主に以下の2点である。

- コーパスにメタ談話標識を自動付与するプログラムの作成
- 多変量解析やデータマイニングなどの統計的手法の活用

また，分析対象となるメタ談話標識とは，文章の書き手と読み手の間に存在するコミュニケーションや社会的行為に関して，一般的な「談話標識」よりも幅広く，包括的に分析していくための枠組みである。また，談話標識が主に話し言葉の分析に用いられるのに対して，メタ談話標識は主に書き言葉の分析に用いられている。そして，メタ談話標識は，書き手と読み手の双方を2つのレベルで助ける (e.g. Jalilifar & Alipour, 2007)。まず，第一のレベルでは，従来の談話標識の枠組みで扱われてきたような接続表現 (e.g. *and, in addition*) によって，書き手が命題内容を系統立てることを助ける。次に，第二のレベルでは，命題内容に関する書き手の評価や態度を表す表現 (e.g. *perhaps, undoubtedly*) によって，読み手が書き手の立場を明確に知ることを助ける。メタ談話の研究において，最もよく使われる枠組みは，恐らくHyland list (Hyland, 2005a) であろう。このリストは，Vande Kopple (1985) やCrismore *et al.* (1993) による研究をベースとして，10種類のカテゴリーに分類される約400種類の談話表現を網羅的に収録したものである。また，このリストは，コーパスに基づく統計的研究を想定して作成されたものであり，これまでにアカデミックライティングを始め，教科書，学位論文，ビジネスレターなど，様々な言語データの分析で成果を上げている。このリストには，TRA (transitions) やFRM (frame markers) のような多くの研究で談話標識とみなされている接続表現が含まれているだけでなく，HED (hedges) やBOO (boosters) のようなstance markers (Biber *et al.*, 1999) のような書き手の態度や評価を表す表現など，書き手と読み手のコミュニケーションに関わる様々な表現が網羅的に含まれているため，非常に多角的な分析が可能である。本研究においては，Hylandによってリスト化されたメタ談話標識を分析の対象とする。

本研究のRQ (Research Questions) は，以下の6つである。

(a) 日本人英語学習者は，学習段階が進むにつれて，談話表現の頻度や使用傾向はどのように変化するのか
(b) 日本の中高英語教科書における談話表現の提示方法は，日本人学習者の英作文の談話的特徴と何らかの関係があるのか
(c) 日本人英語学習者と英語母語話者の間には，談話表現の頻度や使用傾向にどのような違いがあるのか
(d) 世界の様々な言語を背景とする英語学習者の中で，日本人学習者の英作文はどのような談話的特徴を

持っているのか

(e) 日本の中高英語教科書における談話表現は，母語話者が実際に用いる談話表現に即したものであるのか

(f) 世界の様々な言語を背景とする英語学習者と英語母語話者の間には，談話表現の頻度や使用傾向にどのような違いがあるのか

第1章「はじめに」では，本研究の目的，背景，方法論，使用するデータとソフトウェアについて述べていく。第2章「多変量アプローチ」では，相関分析，対応分析，クラスター分析などを用いて，分析データの全体像を把握し，データの構造を視覚的に提示する。第3章「Interactive resourcesの分析」と第4章「Interactional resourcesの分析」では，メタ談話標識の意味カテゴリー別に詳細な量的分析と質的分析を行う。第5章「NS/NNS分類モデル」では，判別分析，決定木，ランダムフォレストといった教師あり学習法を用いて，母語話者の作文と日本人学習者の英作文を分類し，2つのクラスを識別する特徴を抽出する。第6章「世界の英語学習者の談話的特徴」では，日本語を含む17種類の異なる言語を背景とする書き手の英作文データを統計的に比較する。第7章「中高英語教科書における談話表現」では，日本の中学校・高校の英語検定教科書におけるメタ談話標識の提示のされ方を調査し，日本人学習者によるメタ談話標識の使用傾向との関係を探る。そして，第8章「おわりに」では，本論全体を総括するとともに，今後の展望などについて言及する。

論文審査の結果の要旨

小林雄一郎氏の学位請求論文「統計的テキストマイニングによる学習者作文におけるメタ談話標識の研究」は、日本人英語学習者の作文の特色を統計的に分析することで、その言語使用の実態を明らかにし、英作文の技術の指導に有益な情報を提供することを目的としている。

本論文が対象とする研究は、大規模コーパスを用い、統計という極力恣意性を排した分析手段によって、日本人英語学習者の英作文の特色を抽出しようというものであるが、計量的手段をとる以上、コーパステキストから計量可能な指標をまず選定する必要がある。この選定は、後の分析の質の良否を決定する重要なプロセスである。本研究では、ハイランド(Hyland, *Metadiscourse: Exploring interaction in writing*, New York: Continuum, 2005)が提案するメタ談話標識という言語項目を指標として選定し、分析を成功に導いている。これにより、６つの問題設定すなわち、(1)日本人英語学習者は，学習段階が進むにつれて，談話表現の頻度や使用傾向はどのように変化するのか、(2)日本の中高英語教科書における談話表現の提示方法は日本人学習者の英作文の談話的特徴と何らかの関係があるのか、(3)日本人英語学習者と英語母語話者の間には談話表現の頻度や使用傾向にどのような違いがあるのか、(4)世界の様々な言語を背景とする英語学習者の中で日本人学習者の英作文はどのような談話的特徴を持っているのか、(5)日本の中高英語教科書における談話表現は母語話者が実際に用いる談話表現に即したものであるのか、(6)世界の様々な言語を背景とする英語学習者と英語母語話者の間には談話表現の頻度や使用傾向にどのような違いがあるのか、という問いに明確な解答が与えている。

また、分析にあたっては、相関分析、対応分析、クラスター分析、判別分析、ランダムフォレスト等の統計的手段を駆使した独自の分析手法を編み出し、日本人学習者の談話表現使用と母語話者の談話表現使用の弁別的特性の抽出に成功している。ここで確立された手法は、コーパスなどの要件を満たせば、本論文で扱われた学習者英作文のみならず、文学作品、学術論文、インターネットのブログなどの分析にも適応可能な普遍的な手法であり、その功績は大きい。

なお、本論文の性格上やむを得ないところではあるが、研究成果の実践現場への応用については示唆に留まっており、統計的に得られた結果を具体的にいかに英語教育に生かしていくのかという議論への発展が、今後は期待されるところである。しかしながら、この論文によって得られた重要な知見、およびこの論文において開発されたコーパス分析手法の有用性の意義は特筆に値する。

以上のように、本論文は博士（言語文化学）の学位論文として十分価値のあるものと認める。