

ムーブ分析に基づく英語論文表現データベースの開発 - 京都大学学術論文コーパスを用いて -

金丸 敏幸*¹ マスワナ 紗矢子*¹ 笹尾 洋介*² 田地野 彰*¹

京都大学*¹

(kanamaru@hi.h.kyoto-u.ac.jp, smaswana@gmail.com, akira@tajino.mbox.media.kyoto-u.ac.jp)

Victoria University of Wellington*²

(ysasaojp@gmail.com)

1 はじめに

総合研究大学において何よりも求められるのは、英語による学術論文を通じた研究成果の発表である。全学共通教育と大学院を中心とした専門教育との連携を視野に入れたとき、英語教育の果たすべき役割として、英語論文の執筆指導を含むアカデミック・ライティング (以下、AW) 教育の充実が挙げられる。しかし、現状では指導の根拠となるべき客観的データは存在せず、AW 指導の内容は教員の個人的な経験に依存しているというのが現実である。

そこで、われわれは、各専門分野において代表的な国際学術雑誌に掲載された学術論文の構成や論文内で使用される表現の分析を行うことで、AW 指導の客観的な指標となるデータの収集と、AW 指導に役立つ表現データベースの構築を行うことにした。ただし、専門分野の論文の分析には、専門分野の知識が必要不可欠である。そのため、われわれは、学術雑誌・専門誌への投稿を通じて研究発表を行っている各専門分野の大学院生・オーバードクターの方々に、学術論文の内容分析の協力を依頼した。

本研究では、われわれが、以前、開発した我が国最大規模の英語学術論文コーパス「京都大学学術論文コーパス」(金丸, 笹尾, 田地野 2009) に含まれる学術論文を対象とした。このコーパスは、京都大学の各専門分野の研究者に、学生に推薦する学術雑誌を選定してもらい、それらの雑誌に掲載された学術論文の中から無作為に抽出した論文から成り立っている。したがって、分析の対象とした学術論文は、専門分野の批評に耐えうる十分な質を持ったものであると言える。

本研究の目的としては、これら学術論文に使用される英語表現 (語彙やコロケーション) をムーブと呼ばれる構造ごとに収集することで、AW 教育に供する英語表現データベースを構築することである。また、そのデータベースを利用した AW 指導・学習用教材の開発も視野に入れている。

本研究の成果は、全学共通教育における英語教育は

もちろんのこと、実際に論文執筆に携わる大学院生の教育にも貢献できるものである。全学共通教育から専門教育まで含めた英語教育により、アカデミックディスコースコミュニティにおける、ライティングを通じた実践的なコミュニケーション能力の育成を行うことが可能となる。

以下では、われわれが用いた学術論文コーパスを概説した上で、実際の論文分析作業の詳細とその分析の結果について説明する。

2 京都大学学術論文コーパス

現在、日本の大学英語教育の一つの大きな特徴となっているのが、「特定目的の英語」(ESP: English for Specific Purposes) である。ESP 研究の知見から大学英語教育のカリキュラムを検討すると、主に 1, 2 年生を対象とする一般教育課程における英語教育においては、3 年生以降の専門教育課程における英語教育との連携を視野に入れる必要がある。この両者の関係を整理するための理論的枠組みとしては、田地野ら (田地野, 水光 2005) が参考となる。

田地野らの理論的枠組みに従えば、「英語」はまず、「一般目的の英語」(EGP: English for General Purposes) と「特定目的の英語」(ESP) に大きく分けられる。そのうち、「特定目的の英語」については、「学術目的の英語」(EAP: English for Academic Purposes) と「職業目的の英語」(EOP: English for Occupational Purposes) とに細分化される。両者の区別は、専門家集団 (プロフェッショナル・ディスコース・コミュニティ) とジャンルが、それぞれ異なっていることに起因する。

「学術目的の英語」は、さらに、「一般学術目的の英語」(EGAP: English for General Academic Purposes) と「特定学術目的の英語」(ESAP: English for Specific Academic Purposes) とに分類される。「一般学術目的の英語」は各専門分野に共通する、一般的な学術的言語技能を対象とし、他方、「特定学術目的の英語」は理学や法学といった、ある特定の専門分野において必要となる学術的言語技能を対象とする。したがって、大学

表1 各分野ごとの論文数(本)

| 学部 | 論文数 | 学部 | 論文数 |
|---------|-----|----------------|-------|
| 総合人間学部 | 180 | 理学部 | 176 |
| 文学部 | 150 | 医学部 | 160 |
| 教育学部 | 150 | 薬学部 | 150 |
| 法学部 | 150 | 工学部 | 210 |
| 経済学部 | 150 | 農学部 | 175 |
| エネルギー科学 | 101 | アジア・アフリカ | 100 |
| 生命科学 | 100 | 地球環境学舎 | 100 |
| | | 14 学部 (研究科) 合計 | 2,052 |

の英語教育カリキュラムの観点から考えると、前者のEGAPは、一般教育課程の英語が、後者のESAPは、学部・大学院における専門課程の英語が、それぞれ対象となると考えられる。なお、EGAPとESAPは、はっきりと区別されるものではなく、連続体をなしており、それぞれが有機的に関連づけられることが期待されている。

この理論的枠組みに基づいて、学術目的の英語教育に資する言語資源として構築されたのが「京都大学学術論文コーパス」である。

「京都大学学術論文コーパス」は、京都大学の全10学部と4研究科の教職員・学生の協力を得て、特に学術語彙の学習に資するための言語資源として構築された。学術論文コーパスの構築の手順は、以下の通りである(田地野, 寺内, 笹尾, マスワナ 2007)。

まず、京都大学の総合人間学部、文学部、教育学部、法学部、経済学部、理学部、医学部、薬学部、工学部、農学部の各専門分野の研究者(専任教員)の協力を得て、それぞれの分野の学部生・院生に推薦する英語の学術雑誌を5~15種類、選定してもらった。

次に、選定してもらった雑誌から、主として過去2年間に掲載された学術論文を、1雑誌あたり7~16本ずつ無作為に抽出した。抽出した論文を、人手によりテキスト化して、各分野ごとの論文コーパスを構築し、最後に、これらをまとめて全学共通学術論文コーパスを構築した。

また、これらに加え、アジア・アフリカ地域研究研究科、エネルギー科学研究科、生命科学研究科、地球環境学舎の各独立研究科からも学術雑誌を推薦してもらい、そこから無作為に抽出した論文も加えた全学共通学術論文コーパスを用意した。

各分野のコーパスに含まれる論文数を、表1に示す。

3 学術論文の分析

3.1 ムーブ分析

本研究は、大学の英語教育目的を「学術目的の英語」(EAP)として明確に位置づけた上で、EAPの重要なジャンルである学術論文の執筆指導を目的とした表現データベースを構築する。ここでいうジャンルとは、特定の目的を持ったテキストのことを指す。

ジャンルがどのような性質を持ったテキストであるかを理解するためには、ムーブ分析という手法が使われることが多い。ムーブ分析とは、ある特定のジャンル、たとえば学術論文などのテキストセグメントの機能や目的でテキストを分類、分析することをいう。ムーブ分析の対象となるムーブとは、テキストの構成要素で、テキストの展開において特定のコミュニケーション機能や目的を果たすまとまりを指す。また、ムーブは、通常ステップと呼ばれる下位要素から構成されている。

一般にムーブ分析は、ESP研究において広く用いられている分析手法である。主要なESP研究は、高等教育での英語教員によってなされているが、英語教員が自らが専門とする英語教育以外の専門分野のテキストを言語的特徴のみに基づいて分析することには限界がある。そのため、従来のムーブ分析では対象となる専門分野が限られていたり、論文のセクションもアブストラクトやイントロダクションといったように限定的な範囲のみであったりすることが多い。

本研究では、「京都大学学術論文コーパス」から選ばれた論文を、各専門分野の専門知識を有した大学院生やオーバードクターの方々の協力を得ることにより、分野別および横断的に分析、考察を行うことができた。また、論文のセクションも限定することなく学術論文全体のムーブを分析することができた。

3.2 表現抽出

学術論文においては、イントロダクションや結論など、特定のセクションにおいて観察される特徴的な表現が存在する。われわれは、学術論文の構成をムーブという概念を用いて分析することで、セクションよりもさらに細かいレベルで特徴的な表現を抽出することができると考えた。これにより、学術論文執筆のための包括的、かつ実際の支援につながるものと考えられる。

学術論文からの特徴的な表現の抽出作業も、ムーブ分析を行った協力者によって行われた。抽出の対象となる表現は、学術論文において定型的な表現を中心に、熟語から複数の文にまたがるレベルまでを対象とした。

表 2 各研究科の分析論文数 (本)

| 学部 | 論文数 | 学部 | 論文数 |
|---------|-----|----------------|-----|
| 総合人間学部 | 8 | 理学部 | 58 |
| 文学部 | 47 | 医学部 | 37 |
| 教育学部 | 13 | 薬学部 | 44 |
| 法学部 | 4 | 工学部 | 59 |
| 経済学部 | 8 | 農学部 | 58 |
| エネルギー科学 | 9 | アジア・アフリカ | 23 |
| 生命科学 | 33 | 地球環境学舎 | 14 |
| 情報学 | 8 | 15 学部 (研究科) 合計 | 423 |

表 3 あらかじめ用意したムーブ

| コード | ムーブ |
|-----|---------------------|
| 1 | 研究の背景について述べる . |
| 2 | 関連する先行研究を概観する . |
| 3 | 本研究を紹介する . |
| 4 | データ収集の手順などについて述べる . |
| 5 | 実験手順を描写する . |
| 6 | データ分析の手順について描写する . |
| 7 | 結果を提示する . |
| 8 | 結果について議論する . |
| 9 | 主な結果とその意義について述べる . |
| b | 具体的な結果について説明を行う . |
| c | まとめを述べる . |

3.3 作業協力者と分析論文数

学術論文の分析には、京都大学の 15 研究科から合計 51 名の大学院生、およびオーバードクターの方々に協力を仰いだ。協力者には、「京都大学学術論文コーパス」に含まれる論文タイトルを提示し、自分の専門に近い学術論文を選択してもらった。

分析する論文数は、一人あたりの総量が 100 ページ前後になるようにして調整した。専門分野によって論文の長さが異なるため、一人あたりの論文数は必ずしも一致しない。最終的に 423 本の学術論文が分析の対象となった。各学部、研究科ごとの分析論文数を表 2 に示す。

3.4 ムーブ分析および表現抽出作業の概略

分析者には、われわれがあらかじめ用意したムーブ分類表とムーブの分類コード一覧を参考に、論文テキストのすべてに対して、ムーブとステップの分類を行ってもらった。協力者に配布したムーブの分類表は、Nwogu(Nwogu 1997) を基本にしているが、Salager-Meyer(Salager-Meyer 1990, 1992), Bhatia(Bhatia 1993), Swales(Swales 1990, 2004) なども参考にして分類表を作成した。われわれが用意したムーブの一覧を表 3 に示す。

ステップはムーブごとに定義されており、全部で 33 個用意した。ムーブ 4「データ収集の手順などについて述べる .」を例に挙げると、このムーブには、(1) データのソースを提示する。(2) データのサイズを提示する。(3) データ収集の基準を提示する。(4) データ収集の手順を説明する。(5) 分析対象について詳述する。の五つのステップが含まれる。

分析者には、ムーブとステップを組み合わせたコードを使って、論文テキストをステップごとにタグ付け

してもらった。あるステップの開始となるテキストの前には、<mv:xx>というタグを、終わりには始まりと同じ文字列の前に / のタグをそれぞれ論文テキストに入力する。たとえば、ムーブのコードが 2、ステップのコードが 2 であった場合、<mv:22> </mv:22> のように記入する。

ムーブを分類する際、2 つ以上のムーブやステップに当てはまる場合は、そのテキストが担う主な役割で分類することにした。また、ステップの判断が難しい場合は、ステップは 0 とすることにした。たとえば、ムーブ 5 で、ステップの判断が難しい場合には 50 を記入する。

あらかじめ用意したムーブ分類表のいかなるムーブにも当てはまらないテキストがある場合は、d 以降のアルファベットを順次使用してもらった。また、新たなムーブを追加するごとに、ステップについても追加してもらい、新たなムーブのステップが分かれば 1 から順番に追加していき、不明の場合は既存のムーブと同様に 0 を用いてもらった。新たなムーブを追加した場合、別途、ファイルに分類番号と、そのムーブやステップが担う主な役割を書くように指示した。

ムーブ分析が終わった論文テキストに対しては、学術論文に特徴的な表現の抽出を行った。表現の抽出には、ムーブ分析と同様、タグを付けてもらうことにした。表現の始まりには<[を、終わりには]>を、それぞれ論文テキストに入力する。たとえば、<[The purpose of this study is]>...のように行う。ただし、抽出したい表現の中に省きたい語句がある場合は、<[Findings from a study conducted by <#the Faculty of Medicine (Smith, 2004) #> seem to indicate that]> のように、<# 省きたい語 #>という規則にしたがって、論文テキストに入力してもらった。

表 4 新たに追加されたムーブ

| | ムーブ |
|----|----------------------------|
| 1 | テキスト（原典）を解釈・分析する． |
| 2 | 著者の解釈を展開する． |
| 3 | 評者との論争． |
| 4 | 採用する説明の妥当性を検証する． |
| 5 | 比較法研究の成果を記述する． |
| 6 | 実験のデザインの基礎となる理論モデルを提示する． |
| 7 | 理論モデルを提示する． |
| 8 | 理論モデルの予測を検証するための手順を描写する． |
| 9 | 定義，仮定等を議論し，目的とする結果について述べる． |
| 10 | 補題を提示する． |
| 11 | 命題を提示する． |
| 12 | 定理を提示する． |
| 13 | 理論モデルを評価する． |

4 分析結果

本研究におけるムーブ分析の結果，われわれが用意したムーブに対し，新たに 13 のムーブが追加された．また，既存のムーブに対しても 44 のステップが，新たに追加された 13 のムーブに対しては，75 のステップがそれぞれ追加された．新たに追加されたムーブを表 4 に示す．

各分野ごとのムーブ分析の結果をまとめると，以下のようになる．まず，文学などの文系分野では，論文の内容に古典文献の解釈や抽象概念の分析といったものが含まれ，理系の論文のように，明確に「実験 結果 総論」といったムーブの展開が存在せず，われわれが提示したムーブだけでは，分類が行いにくい論文が多く存在した．そのため，原典の分析や解釈，原典の著者の解釈を展開するといった新たに追加されたムーブがこの分野で特徴的に見られた．

また，文理を問わず，理論系の論文では，数学的な展開が中心となっていた．そこでは，「理論モデルの設定・仮定 問題の定式化 問題を解く 命題・定理などの導出」という，これまで中心的に分析されてきた実験系の論文におけるムーブとは大幅に異なる展開が見られた．

抽出された表現の数は，のべ 34,441 表現となった．抽出された表現には，“for instance” のような一般的な熟語的表現から，“the purpose of this study” のような論文固有の慣用表現まで多様な種類が確認されている．

5 おわりに

本稿では，われわれが行った「京都大学学術論文コーパス」を用いた分野横断的かつ論文全体に渡る学術論文のムーブ分析と，学術論文内で使用される表現の抽出によって構築した表現データベースについて述べた．

特に学術論文のムーブ分析については，これまで，狭い範囲の分析しか行われてこなかったのに対し，われわれは，分野横断的かつ学術論文全体に渡って行った．その結果，これまで明らかになっていたムーブ以外にもさまざまなムーブ，およびステップが存在することが明らかとなった．

現在，われわれは，新たに追加されたムーブ，ステップを含め，ムーブごとに抽出した表現との対応を分析しているところである．この分析によって，抽出された表現が，より細かな観点から利用できるようになる予定である．

参考文献

- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman, London.
- 金丸敏幸, 笹尾洋介, 田地野彰 (2009). “京都大学学術論文コーパスを用いた学術語彙リストの作成.” 言語処理学会第 15 回年次大会 発表論文集, pp. 737-740.
- Nwogu, K. N. (1997). “The medical research paper: Structure and functions.” *English for Specific Purposes*, 16 (2), pp. 119-138.
- Salager-Meyer, F. (1990). “Discoursal flaws in medical English abstracts: A genre analysis per research-and text-type.” *Text*, 10 (4), pp. 365-384.
- Salager-Meyer, F. (1992). “A text-type and move analysis study of verb tense and modality distribution in medical English abstracts.” *English for Specific Purposes*, 11 (2), pp. 93-113.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press, Cambridge.
- Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge University Press, Cambridge.
- 田地野彰, 水光雅則 (2005). “大学英語教育への提言 - カリキュラム開発へのシステムアプローチ -.” 竹蓋幸生, 水光雅則 (編), これからの大学英語教育, pp. 1-46. 岩波書店.
- 田地野彰, 寺内一, 笹尾洋介, マスワナ紗矢子 (2007). “総合研究大学における英語学術語彙リスト開発の意義 - EAP カリキュラムデザインの観点から -.” 京都大学高等教育研究, 13 (1), pp. 121-131.