

英語論文表現データベースを用いた分野横断的ムーブ分析

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 Wellington²

(ysasaojp@gmail.com)

1 はじめに

本研究では、われわれが、以前、開発した「京都大学学術論文コーパス」(金丸, 笹尾, 田地野 2009) から選ばれた論文の構造を各専門分野の研究者が分析した「京都大学英語表現データベース」(金丸, マスワナ, 笹尾, 田地野 2010) を用いて、論文全体のムーブについて分野横断的に分析、考察を行った。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。

これまでの研究では、研究者が対象ジャンルのテキストを恣意的にいくつか選択し分析しており、必然的に研究結果の信頼性および応用性が低いものとなっている (Biber, Connor, and Upton 2007)。また、従来の ESP (English for Specific Purposes: 特定目的の英語) 研究におけるムーブ分析は、主に高等教育での英語教員によってなされていた。そのため、分野的にも範囲的にも限定的なものに留まっていた。われわれの研究においては、ムーブの分析に際して、各専門分野の研究者の協力が得られたため、広範、かつ、専門的知見からの分析結果を得ることができた。論文分析に専門家の協力が必要不可欠な理由として、分析者(通常は英語教員)の専門分野以外の論文テキストを言語的特徴のみに基づいて分析することは非常に難しいことが挙げられる。そもそも、特定ジャンルのテキストは、対象ディスコース・コミュニティのメンバーによって認識されるものであり、特に学術論文はこの性質が強いと言われる。

また、従来の研究では、「Introduction - Method - Results - Discussion」(いわゆる, IMRD) という明確

なセクション分けがされている論文の一セクションのみが主に分析対象となっていた。しかし、実際には明確な IMRD 構造を持たない論文も数多く存在し (Yang and Allison 2004)、一部の論文セクションの分析だけでは論文全体の構造を理解するには不十分であると考えられる。今回、われわれは各専門分野の研究者の参加により、本来難しいとされた論文全体のムーブ分析が可能となった。

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

大学における英語教育の目的を「学術目的の英語」(English for Academic Purposes: EAP) として位置づけた上で、この EAP 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

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

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

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

大学の英語教育の目的を EAP として位置づけた場合, EAP の重要なジャンルである学術論文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必要がある。ここで, ジャンルとは, 特定の目的を持ったテキスト(一連の発話や文の総称)であり, EAP 研究においては “a class of communicative events, the members of which share some set of communicative purposes” (Swales 1990, 58) と定義される。

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て, 分析されたものを用いた。これにより, 分野別横断的にムーブの分析や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 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。

3.3 ムーブ分析作業の概略

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。

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。

協力者のムーブ分析の結果、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。最終的に得られたムーブを表3に示す。

4 ムーブ展開の分析

協力者によるムーブ分析の結果をさらに分析するため、各ムーブやステップがどのように展開しているかを算出した。

まず、ムーブの連続パターンについては、異なりパターンとして236種類、あるムーブからあるムーブへの展開はのべ11,284回であった。特に多く見られたムーブの連続パターンの上位30パターンを表4に示す。表4におけるムーブ遷移で示される数字は、表3に挙げたムーブ番号である。

表3 最終的に得られたムーブ一覧

	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。
d	テキスト(原典)を解釈・分析する。
e	著者の解釈を展開する。
f	評者との論争。
g	採用する説明の妥当性を検証する。
h	比較法研究の成果を記述する。
i	実験のデザインの基礎となる理論モデルを提示する。
j	理論モデルを提示する。
k	理論モデルの予測を検証するための手順を描写する。
l	定義、仮定等を議論し、目的とする結果について述べる。
m	補題を提示する。
n	命題を提示する。
o	定理を提示する。
p	理論モデルを評価する。
a	アブストラクト

ここから分かるように、ムーブ7とムーブ8を相互に遷移しているパターンが多く観察された。ムーブ7は結果の提示であり、ムーブ8は結果についての議論である。いわゆるIMRDに代表されるように、結果と議論については、「Results - Discussion」の順に分けて提示することが望ましいとされてきた。しかし、この結果を見る限り、実際には結果と議論は行き来することが多く、理想的なパターンばかりではないことが分かる。

次に、ステップのレベルで見た場合の連続パターンについては、異なりパターンとして、1,540種類、あるステップからあるステップへの展開はのべ18,227回であった。ムーブの展開パターンに比べ、ステップの展開パターンの方は、異なりパターンの種類で約6.5倍の差がある。しかし、一方、のべパターンの回数差は約1.6倍に留まっている。このことから、同一ムーブ内におけるステップだけの遷移が連続することは少ないこ

表 4 ムーブ展開の上位 30 パターン

順位	ムーブ遷移	出現数	順位	ムーブ遷移	出現数
1	7 8	1625	16	3 4	160
2	8 7	1143	17	5 6	158
3	6 7	508	18	7 5	153
4	1 2	430	19	3 2	151
5	2 3	385	20	b 7	149
6	a 1	340	21	9 b	140
7	5 7	324	22	2 1	138
8	4 7	315	23	4 5	137
9	7 b	305	24	2 7	119
10	b c	302	25	8 6	119
11	7 6	213	26	6 4	108
12	7 4	206	27	7 2	95
13	4 6	206	28	6 5	94
14	8 b	184	29	3 5	93
15	1 3	176	30	d e	89

とが推測される。一例を挙げると、ステップ 74「具体的な分析結果を提示する」からステップ 81「得られた結果の説明を行う」の遷移（74 → 81）が 1,106 回と一番多い一方、逆の遷移（81 → 74）が 428 回と、こちらは 3 番目に多く出現している。

ただし、ステップの遷移パターンのうち、出現回数が多いものを見てみると、同一ムーブ内におけるステップの遷移や繰り返しも目立っていた。こちらの例には、ステップ 73「大まかな結果を提示する」からステップ 74 への遷移（73 → 74）が 441 回あり、逆のパターン（74 → 73）も 179 回出現している。

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の結果について述べた。この分析結果は、アカデミックライティングにおける論文執筆指導の客観的な指標となるデータとして役立つものである。

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。

現在、これらのムーブ、ステップの展開パターンの分析結果と併せて、昨年度の分析時に収集した論文表現データを公開するための準備を進行中である。

謝辞

本研究は、科研費（基盤研究 (B) 22320102）の助成を受けた。

参考文献

- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman, London.
- Biber, D., Connor, U., and Upton, T. (2007). *Discourse on the move. Using corpus analysis to describe discourse structure*. John Benjamins, Amsterdam.
- 金丸敏幸, マスワナ紗矢子, 笹尾洋介, 田地野彰 (2010). “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 -.” 言語処理学会第 16 回年次大会 発表論文集, pp. 522-525.
- 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, 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.
- Yang, R. and Allison, D. (2004). “Research articles in applied linguistics: structures from a functional perspective.” *English for Specific Purposes*, 23 (3), pp. 264-279.