

学術論文の構成要素と構造

宮田洋輔(慶應義塾大学) m@miyay.org
石田栄美(九州大学), 池内淳(筑波大学)
安形輝(亜細亜大学), 上田修一(慶應義塾大学)

抄録

研究の成果を公表するための主要なメディアとして学術論文がある。学術論文は、専門的かつ論理的な記述であるという内容的特性だけでなく、文献の参照、抄録の存在、IMRADのような構造化、といった形式的特性を有している。本研究では、日本で出版された学術論文と海外で出版された学術論文 1,172 件を対象として、学術論文の構成要素と構造に関する調査を行った。その結果、構成要素では雑誌に関するコードの表記に大きな差異があること、構造では、Introduction 以外は明示的には示されていない論文が最も多いことが明らかになった。

1. はじめに

学術論文は研究の成果を公表するための主要なメディアとして用いられている。学術的なコミュニケーションを成立させるために、学術論文は、内容や表現形式が高度に定式化された特性を持っているとされており¹⁾、専門的かつ論理的な記述であるという内容的特性だけでなく、文献の参照、抄録の存在、構造といった形式的特性も有している。しかしながら、後者の形式的特性は、分野や言語、掲載される雑誌等によって様々である。

学術論文の典型的な構造として、IMRAD 形式がある。IMRAD 形式は、序論(Introduction)、方法(Method)、結果(Result)、結論(Discussion)から構成される。デイとガステルは、IMRAD 形式の普及について、学術研究の方法論の重要性が高まる中で、実験の再現性を高めるために、方法に関する記述が独立し、現在の形態が確立されたと指摘している²⁾。

学会が公開しているスタイルガイドや、学術論文の指南書でも、学術論文を構造的に記述することが推奨されている。例えば、アメリカ心理学会(APA)のスタイルマニュアルでは、実証研究の報告に含まれるべき記述内容を包括的に示した学術論文報告基準を設定している。その中で、タイトル、著者の氏名と所属、著者注(著者の所属や謝辞、連絡先など)、アブストラクト、序文、方法、結果、考察、複数の実験、メタアナリシス、リファレンス、脚注、付録およびオンライン補足資料を挙げ、今日求められている学術論文の構造と形式について解説している³⁾。デイとガステルは、科学論文は、はっきりと区別できるいくつかの部分から構成され、それらの関係は統一的なスタイルがある、と

し、IMRAD 形式を推奨している²⁾。

学術論文とはどのようなメディアであるかという問題意識から、その構成要素や構造を把握しようとする試みが行われている。医学分野では、SollaciとPereiraが、1935年から1985年までにBritish Medical Journal, JAMA, Lancet, New England Journal of Medicineに掲載された1,297件の原著論文を対象とした調査を行った⁴⁾。その結果、1985年には医学分野ではIMRAD形式を採用する論文が100%になっており、1955年から75年の20年間で、それまでの4倍にまでIMRAD形式が普及したことを明らかにした。物理学分野では、Bazermanは、1893年から1980年までにPhysical Reviewに掲載された学術論文の、様々な特徴について調査している⁵⁾。その中で、Physical Review掲載論文では、1950年以降、論文が見出しによって構造化されていき、時代が新しくなるにつれ、見出しとして論文固有の名前を用いるのではなく、「Experiment」のようなより抽象化された見出しを持つようになっていったことを指摘している。実験心理学分野では、倉田と坂上は、Journal of Experimental Analysis of Behavior, Journal of Experimental Psychology: Animal Behavior Processes, Journal of Experimental Psychology: Human Learning and Memory, Memory and Cognitionの1975年と1998年の実験を行った全論文を対象とした調査を行った¹⁾。その結果、序論や考察の割合が増え、結果をそのまま提示するものから、議論や解釈の重視へと変化してきていることを示した。2007年に掲載された、工学、応用化学、社会科学、人文学の39分野の433論文を対象とした

Lin と Evans の調査では、多くの実証論文が IMRAD 形式に従っているが、序論-文献レビュー-方法-結果と考察-結論(ILM[RD])という構造を採る論文が最も多かった⁶⁾。

また構成要素については、科学技術情報流通技術基準(SIST)08で、「学術論文の執筆と構成」として、学術論文の構成要素とその記載要領についての基準が示されている⁷⁾。

本研究では、学術論文の表現形式の構造に着目し、分野を問わない複数の学術論文を調査・分析することによって、現代における様々な分野での、学術コミュニケーションの特徴について理解することを目的とする。学術論文における要素や構造を把握することは、学術コミュニケーションにおける学術論文というメディアの役割を明らかにし、さらには学術論文をほかの文書と区別する際の手がかりを与えるだろう。筆者らは、これまで学術論文の構造的特徴に基づいてウェブ上から取得した様々な PDF ファイルに対して、学術論文を自動的に判定し、提供する学術論文検索エンジンを作成してきた⁸⁾。学術論文の構造や形式的特性を把握することは、これら一連の研究の精緻化にも繋がるだろう。

2. 調査方法

学術論文 PDF ファイルを収集し、学術論文の中に現れる構成要素と、見出しに基づく学術論文の構造に関する調査を実施した。

2.1. 調査対象

はじめに調査対象となる学術雑誌の標本を作成した。本調査で対象とする論文は、分野や格付けに依存しない広範なタイトルの集合を得るために、国内で出版された雑誌のタイトルは、「CiNii 本文収録刊行物ディレクトリ」に掲載されたタイトルから(以下、国内集合)、国際的に出版されている雑誌のタイトルは、Web of Science, Journal Citation Reports, Arts & Humanities Citation Index から無作為抽出した(以下、国際集合)。

標本に含まれた雑誌タイトルに対して電子ファイルの提供があるかどうかを確認し、PDF ファイルを入手できる場合は、2010 年第 1 号までの中での最新号から、当該号の中で 2 番目に掲載された論文を調査対象としてダウンロードした。論文ファイルをダウンロードできない場合、その雑誌タイトルは調査対象から除外した。また、PDF ファイル

に含まれる内容が、書誌や詩など論文的な内容でない場合は、調査対象から除外した。その結果、国内集合では、490 ファイルを取得した。国際集合では、682 ファイルを取得した。

2.2. 調査項目

取得した PDF ファイルに対して、構成要素の調査と、構造の調査を実施した。

構成要素の調査では、「論題」、「著者」、「所属」、「抄録」、「キーワード」、「引用文献」の有無、「引用文献の位置」、「掲載雑誌名」、「ISSN」、「DOI」の有無、「書字方向(縦書・横書)」、「見出し」の有無を集計した。各論文には、論文内容に基づいて NDC 分類を付与した。また、対象論文が各章に対して見出しを持っている場合は、見出しの文字列を記録した。表 1 に調査対象論文の、NDC 第 1 区分別の集計を示した。

表 1 調査対象の主題

	国内		国際		総計	
	件数	%	件数	%	件数	%
0類	31	6.3%	10	1.5%	41	3.5%
1類	3	0.6%	51	7.5%	54	4.6%
2類	1	0.2%	18	2.6%	19	1.6%
3類	20	4.1%	90	13.2%	110	9.4%
4類	256	52.2%	384	56.3%	640	54.6%
5類	130	26.5%	51	7.5%	181	15.4%
6類	45	9.2%	17	2.5%	62	5.3%
7類	4	0.8%	25	3.7%	29	2.5%
8類	0	0.0%	10	1.5%	10	0.9%
9類	0	0.0%	26	3.8%	26	2.2%
総計	490	100.0%	682	100.0%	1,172	100.0%

「見出し」を持っている論文は、見出しによって構造化されているとみなし、学術論文構造の調査を実施した。構造調査では、IMRAD を構成する序論(Introduction)、方法(Method)、結果(Result)、結論(Discussion)と、最近の調査で指摘されている、文献レビュー(Literature review)と結語(Conclusion)を追加し、記録された見出しが構造的な意味を持つ場合には、ラベルを付与した。表に示したような語によって見出しの役割が明示的に示されている場合のみ、構造のラベルを付与した。付与されたラベルを結合し連続した箇所をまとめた文字列を、対象論文が持つ構造とした。

例えば、「はじめに」、「対象」、「実験方法」、「結果」、「考察」という 5 つの見出しで構成される論文があった場合、それぞれに IMMRD というラベル

が付与され、連続した M をまとめ、IMRD をその論文が持つ構造とした。

表 2 構造を明示する語の例

	日本語	英語
I	「はじめに」	"Introduction"
	「序論」	"Background"
M	「試料と方法」	"Material and Method"
	「実験」	"Methodology"
R	「実験結果」	"Result"
	「結果」	"Finding"
D	「考察」	"Discussion"
	「議論」	"Implication"
L	「文献レビュー」	"Literature Review"
	「先行研究」	"Related Research"
C	「おわりに」	"Conclusion"
	「結論」	"Summary"

3. 調査結果

3.1. 構成要素

論文の構成要素に関する集計をばらつきが大きかったものに関して対象集合別に示した(表 3)。論題(国内 100%, 国際 100%), 著者(国内 100%, 国際 99.9%), 所属(国内 100%, 国際 97.2%), 掲載誌名(国内 97.8%, 国際 98.5%)という学術論文を識別するための基本的な要素については、国内集合と国際集合との間に大きな違いはなかった。抄録、キーワード、引用文献に関しても調査対象とした集合の間で大きな違いはなかった。また書字方向も全ての論文が横書きであった。学術論文の構造化の指標とした見出しの有無に関しては、国際集合の論文のほうが国内集合の論文に比べて、見出しのない論文が多かった。国際集合で見出しを持たなかった論文 50 件中 44 件(88.0%)が、Arts & Humanities Citation Index から抽出されたものであった。

ISSN と DOI では、集合間で大きな違いがあった。ISSN は国内で出版された学術雑誌ではどれにも含まれていなかったが、国際集合の論文では 4 割以上で論文内に ISSN が含まれていた。DOI に関しても、日本で出版された学術論文にはほとんど含まれていなかった。これは、2012 年 3 月 15 日にジャパンリンクセンターが認定されるまで日本には DOI の登録機関がなく付与が難しかったためと考えられる。

国際集合では、見出しを 1 つも持たない論文(7.3%)が国内集合(0.8%)に比べて多かった。こ

れは国際集合の学術論文が、文学や宗教学のような人文学の論文を多く含んでいたことがその原因として考えられる。

表 3 構成要素の集計

	国内		国際		合計
ISSN					
あり	0	0.0%	293	43.0%	293
なし	490	100.0%	388	56.9%	878
DOI					
あり	1	0.2%	508	74.5%	509
なし	489	99.8%	174	25.5%	663
抄録					
あり	437	89.2%	607	89.0%	1,044
なし	53	10.8%	75	11.0%	128
キーワード					
あり	370	75.5%	431	63.2%	801
なし	120	24.5%	251	36.8%	371
引用文献					
あり	482	98.4%	674	98.8%	1,156
なし	8	1.6%	8	1.2%	16
引用文献の位置					
末尾	482	99.6%	629	93.0%	1,111
脚注	2	0.4%	47	7.0%	49
見出し					
あり	486	99.2%	632	92.7%	1,118
なし	4	0.8%	50	7.3%	54

3.2. 構造調査

次に見出しを持った論文 1,118 件に対して、見出しの役割に関する分析を行った。構造の調査を行う中で、レビュー論文は、調査対象から除外した。その結果、構造調査の対象となったのは 1,086 論文であった。構造調査の結果、35 パターンの構造が見られた。構造パターンの中で総計での上位 10 構造を、表 4 に示した。

表 4 学術論文の上位 10 構造

構造	件数	構造	件数
1 I	380	6 IM	35
2 IMRD	191	7 IMRC	19
3 IMRDC	187	8 IL	16
4 構造なし	136	9 MRDC	14
5 MRD	45	10 IMR	10
その他	54	合計	1,086

その結果、Introduction 以外は、明示的な構造をもたない論文が最も多かった。次に IMRD が続く IMRAD 形式が多かった。IMRAD 形式でも、最後に Conclusion を持つ事例も少なくなかった(全体で 187 件)。また、IMRAD 形式を採っていても Introduction が見出しを持たず、今回の調

査では MRD 型となる事例もあった。このような派生的な形式も含めると、IMRAD 形式が学術論文構造の主流であるといえる。

これらの構造のパターンを、IMRAD 形式との近さを比較するために、各構造のパターンと IMRD に対するレーベンシュタインの編集距離を算出した。表 5 ではパターン間の編集距離を国内と国際で比較した。その結果、国際の方が、いずれの見出しも IMRAD 形式と一致しないものの割合が多かった(編集距離 4: 121 件, 19.6%)。

表 5 IMRAD 形式との編集距離と集合

集合	編集距離					総計
	0	1	2	3	4	
国内	80	142	22	208	17	469
国際	111	126	51	208	121	617
総計	191	268	73	416	138	1,086

次に主題ごとに比較した(表 6)。自然科学分野は第 2 次区分まで展開し、460 から 489 は生物科学にまとめた。各主題区分の中の最も件数が多かったセルを罫線で示した。その結果、生物科学では多くの論文が IMRAD 形式を採っていた。また化学、医学、産業でも、IMRAD に近い形式も多く採られていた。一方で、哲学、歴史、芸術、文学のような人文学分野では、IMRAD 形式とは遠い形式が採られていた。

表 6 IMRAD 形式との編集距離と主題分類

分類	編集距離					総計
	0	1	2	3	4	
総記		6	2	32		40
哲学	3	4	2	8	28	45
歴史		1		1	8	10
社会科学	11	5	7	56	28	107
自然科学						
自然科学				7		7
数学		2	1	22	2	27
物理学	1	10	3	29		43
化学	5	20	2	11	2	40
天文学			2	1		3
地球科学	5	20	6	27	3	61
生物科学	57	27	10	38	6	138
医学	86	110	16	50	25	287
技術	8	37	13	107	8	173
産業	14	24	7	12	5	62
芸術		1	1	9	13	24
言語	1	1	1	4	2	9
文学				2	8	10
総計	191	268	73	416	138	1,086

4. 考察

学術論文の構成要素に関する調査の結果、学術論文の PDF ファイルは、ほとんどの場合、基本的な要素として、論題、著者、所属、抄録、引用文献を持っており、横書きで書かれていた。一方で、雑誌固有のコードに関しては、国内の集合と国際の集合で大きな差があった。

見出しに含まれる語に基づく学術論文の構造調査の結果、Introduction 以外は明示的には示されていない論文が最も多かった。しかし、自然科学系を中心に IMRAD 形式に近い構造が多く採られていた。

引用文献

- 倉田敬子. 学術情報流通とオープン・アクセス. 東京, 勁草書房, 2007, 196p.
- デイ, R・A; ガステル, B. 世界に通じる科学英語論文の書き方: 執筆・投稿・査読・発表. 美宅成樹訳. 丸善. 2010, 321p.
- アメリカ心理学会. APA 論文作成マニュアル. 第 2 版. 前田樹海, 江藤裕之, 田中建彦訳. 医学書院, 2011. 304p.
- Sollaci, L.B, Pereira M.G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. Journal of Medical Library Association. 2004, vol. 92, no.3, p.364-367.
- Bazerman, Charles. Modern Evolution of the Experimental Report in Physics: Spectroscopic Articles in Physical Review, 1893-1980. Social Studies of Science. 1984, vol. 14, no. 2, p. 163-196.
- Lin, Ling; Evans, Stephen. Structural patterns in empirical research articles: A cross-disciplinary study. English for Specific Purposes. (In Press)
- 科学技術情報流通技術基準 (SIST)SIST08: 2010. 学術論文の執筆と構成 http://sist-jst.jp/pdf/SIST08_2010.pdf
- 安形輝, 池内淳, 石田栄美, 野末道子, 久野高志, 上田修一. 日本語学術論文 PDF ファイルの自動判定. Library and Information Science. 2007, no.56, p.43-63.