

# 引用の対象である概念の表現の考察

渡邊 晃一郎<sup>1,3</sup> 影浦 峯<sup>1,2</sup> 関根 聡<sup>3</sup>

<sup>1</sup> 東京大学大学院 教育学研究科

<sup>2</sup> 東京大学 情報学環 <sup>3</sup> 理化学研究所 AIP

kouichirou-watanabe495@g.ecc.u-tokyo.ac.jp

## 1 はじめに

学術領域における引用は、他の文献への参照を示す形で行われ [1]、その参照は“Citation Anchor (CA)”<sup>1</sup>を用いて示される。このような引用が行われたとき、引用文献では被引用文献の内容が記述されるが、ここではその内容を引用の対象とする。このCAの有無の判断は一般に文を単位になされるが、全ての場合において文全体が引用の対象を記述するとは考えにくい。では、引用において対象となるものはどのようなものであるのか。これについて、Small [2] は、引用の対象とはある被引用文献を引用することによって導入される“concept (概念)”であると述べた<sup>2</sup>。以下、本研究では、引用の対象である概念を引用対象概念、それに対応する言語表現を引用対象概念表現とする。これらについて、Small は以下のような例を示している。

Small’s paper (1) presents a completely erroneous interpretation of citation practice.

この例では“(1)”というCAによって示される引用は、“Small’s paper”を引用することにより“completely erroneous interpretation of citation practice”により表現される引用対象概念を導入していると考えられる。

このように引用の対象が概念であることを前提として研究者及び論文の影響力を分析する研究手法である引用概念分析が最近登場したが、この研究手法は名詞(句)のみを考慮しており、節のレベルをそもそも考慮していない [3]。加えて、学術論文要約のタスクでは、被引用文献の内容に関係する内容を記述した言語表現である引用文脈を、そこで引用する文献の要約と考える場合がある (e.g. [4])。しかし、そのような引用に対する考え方は想定に過ぎず、これまで引用の対象に対する分析は行われていなかった。そこで、本研究では引用の対象がどのようなものであるのかを明らかにするために、引用対象概念表現の考察を行う<sup>3</sup>。

<sup>1</sup>“By offering a flexible computational architecture, they allow having more and more industrial modern services (Deng, Huang, & Xu, 2014).”における“(Deng, Huang, & Xu, 2014)”のような言語表現のこと。

<sup>2</sup>Small [2] は引用の対象となる概念について、言語表現のレベルでは語句によって示されるものに限らず、文(節)によって示されるものも含むとしている。

<sup>3</sup>Kang ら [5] は本研究と類似した引用文脈の様相の分析を行っ

## 2 研究方法

大まかには以下の4つの手順を踏む。

1. 引用文の収集
2. 引用対象概念表現の抽出
3. 引用対象概念表現の分類
4. 引用対象概念表現の分類の結果の分析

1つ目の手順では、各論文から引用文(CAの存在する文)を抽出し、2つ目の手順では、1節の具体例で示したような形で、各CAに対しそれと対応する引用対象概念表現を抽出する。3つ目の手順では、引用対象概念表現を2.1と2.2で説明する枠組みに基づき、分類する。このとき、文は節から構成され、節は語句から構成されるため、考えられる単位としては節と語句があるが、それ自体の性質を考慮する語句の分類に対し節の分類は複数の構成要素を考慮することが必要であるため、本研究ではそれぞれ異なる枠組みを適用する。この分類を踏まえて、4つ目の手順では、この結果を分析する。2.1と2.2では3つ目の手順で用いる枠組みについて詳述する。

本研究で行う意味の分類は対象に対する一定の理解が必要であることから、以上の分類を行う対象データとして、筆者の専攻である図書館情報学分野の学術雑誌に掲載されている論文、具体的には *Information Processing and Management, Journal of the Association for Information Science and Technology, Scientometrics* において2018年に出版された論文からランダムサンプリングされた各10件、合計30件の論文を用いる。

### 2.1 語句レベルの引用対象概念表現の分類

語句レベルの概念分類については、これまで多くの研究が行われてきた。ただし、本研究は学術領域という限定された領域を対象とするため、学術領域の語句を対象に概念分類の枠組みを作成しているKoら [6] の枠組みを一部修正し適用する。この分類の枠組みは

ているが、Kangらの分析対象は引用文脈であり、分析が言語表現の単位について行われているという2点で本研究と異なる。

表 1 に結果と併せて示している。本研究では、大まかな分類である 1st level とより詳細な 2nd level での 2 段階の分類を行っている。これに加え、語句が固有名か否かの分類も行う。

## 2.2 節レベルの引用対象概念表現の分類

一般に節を概念を表現するものとして考えることはないため、節が表現する概念を、一旦その節が示す意味として考える。節レベルでの意味の分類はこれまで機能文法分野で行われてきたので、本研究ではその枠組みで節レベルの引用対象概念表現を分類する。しかし、機能文法には主に 2 つの理論が存在するため、そのうちのどれを選択するかを以下で考える。

機能文法には大きく Halliday [7] による理論と Dik [8] による理論があり、Dik の理論は “State of Affair (SoA)” を節の意味の中核と考えるが、この考え方は態が異なると異なる類型に分類される。しかし、本研究では態の差異及びそれによる分類が重要であるとは考えられないため、Halliday の理論を採用する。

Halliday が提唱した機能文法では、経験した事柄を “過程型” に解釈構築することによって人間は節の意味内容を理解するとした [7, p. 106–107]。この過程は過程中心部、参与要素、状況要素の 3 つからなり [7, p. 109]、意味の分類を考えた時、その中心は過程中心部と言える<sup>4</sup>。節において、過程中心部に相当するのは動詞群であり、これは動詞によって示される事象とそれに付随する定性から構成される [7, p. 196]。この定性は時制とモダリティから構成される [7, p. 75]。そのため、節の意味はその中心である過程型とそれに付随する時制、モダリティの 3 つの観点で分類することができる。過程型の類型には物質過程 (外的世界における人や物の動き)、心理過程 (内的世界における経験)、関係過程 (経験の断片を、別の断片に関連づけ、一般化すること)、行動過程 (物質過程と心理過程の境界に位置する、意識と生理的状態を外的に明示すること)、発言過程 (心理過程と関係過程の境界に位置する、意識の中で構築され、言語として表現された記号的な関係を提示すること)、存在過程 (物質過程と関係過程の境界に位置する、事物の存在と生起のこと) の 6 類型がある [7, p. 106–107] ので、本研究では節の意味の中心となる分類として、以上の 6 つの類型への分類を行う。時制は大きく過去、現在、未来の 3 種類からなるが、未来はモダリティとの判別が難しい部分があるため、本研究では過去と現在の 2 つに分類する。それに加えて完了相についても完了形と未完形に分類する。モダリティは “Low”、“Mediam”、“High”、モ

ダリティなし (本研究では No に分類) の 4 類型があり [7, p. 76]<sup>5</sup>、本研究でもその分類を適用する。

## 3 分析結果

対象データから抽出された引用対象概念表現は 1614 個であり、そのうち 39.5% の 638 個が語句レベル、60.4% の 976 個が節レベルであった。

### 3.1 語句レベルの引用対象概念表現の分析

1st level	2nd level	比率	個数	固有名比率	固有名個数
A.Object	A01_Human	0.000	0	0.000	0
	A02_Institution_Organization	0.003	2	1.000	2
	A03_Natural_Object	0.000	0	0.000	0
	A04_Artifacts	0.227	140	0.492	69
B.Action_Function	B01_Action_Activity_Role	0.076	49	0.000	0
	B02_Change	0.009	6	0.000	0
C.Property	C01_Characteristic_Property	0.043	28	0.000	0
	C02_Psychology	0.001	1	0.000	0
	C03_Phenomenon_Issue	0.012	8	0.000	0
D.Theory_Method	D01_Theory_Thought_Ideology_Principle_Rule	0.100	64	0.109	7
	D02_System	0.059	38	0.605	23
	D03_Method	0.202	134	0.485	65
	D04_Technique_Strategy	0.064	41	0.292	12
E.Format_Framework	E01_Form_Type_Style_Genre	0.001	1	0.000	0
	E02_Model_Criteria	0.164	104	0.451	47
	E03_Languages	0.000	0	0.000	0
	E04_Space	0.000	0	0.000	0
X.General_Common	X01_Place_Name	0.000	0	0.000	0
	X02_Period_Time	0.000	0	0.000	0
	X03_Relationship_Interaction	0.014	9	0.000	0
	X04_Result	0.012	8	0.000	0
	X05_Aim	0.000	0	0.000	0
	X06_Cause	0.001	1	0.000	0
	X07_Effect	0.006	4	0.000	0
Total		1.000	638	0.352	225

表 1: 語句の比率; “固有名比率” の個数の行では各項目に占める固有名比率を示し、“固有名個数” では各項目に当てはまる語句において固有名が当てはまる個数を示している。

語句レベルでは、表 1 で示すように、D.Theory\_Method と E.Format\_Framework に該当するものの比率を合計すると約 59% となり、半分以上を占める結果となった。これらの項目は用語が当てはまる項目であり、予想される結果であると言える。その次は A.Object であり、その中でも A04\_Artifacts が多い結果となった。この結果の全体からは、用語を引用の対象にする場合が過半数を占めるものの、用語ではない語句に対する引用が一定数存在することが示されている。固有名比率を見ると、全体ではおよそ 35% が固有名である。固有名割合が最も高いのは A04\_Artifacts であり、個別事例分析で示すように、人名を文献として扱った引用の仕方が影響していると考えられる。

次に、どのような用語ではない語句が引用の対象となっているのかを分析するために、以下で A04\_Artifacts

<sup>4</sup>Halliday [7] にある用語の日本語の訳語については、以下に示す同書の日本語訳を参照している: Michael Alexander Kirkwood Halliday (2001) 『機能文法概説—ハリデー理論への誘い—』 [An Introduction to Functional Grammar, 2nd ed., London, Edward Arnold, 1994] 山口登, 筑壽雄訳, くろしお出版。

<sup>5</sup>モダリティを示す助動詞は以下のように分類される: Low: can, may, could, might, need not, do not need/have to; Mediam: will, would, should, is to, will/would/should not, is not to; High: must, ought/have to, need, must/could/can/might/may not, ought/have not to

に該当する語句について、固有名であるものとそうでないものに分けて事例分析を行う。

A04\_Artifacts に分類されるもので固有名でないものは大きく 2 つに分類することができる。1 つは文献そのものを示すものであり、これが 71 個中 55 個を占める。以下で具体的に例を示しつつ、説明する。以下の例においては、太字の部分が注目する引用対象概念表現である。まず、以下の例では “previous studies” が引用対象概念表現である。

However, the fast growing of predatory publications found in our dataset indicates that there might have some gain achieved through the publishing in this kind of outlet (BL, in our case), which are consistent with **previous studies**(Pyne 2017).

この例においては、被引用文献が示す内容そのものを記述することなく、文献を引用対象概念とすることで被引用文献全体を引用していると考えられる。このように、語句を対象とする引用においては、被引用文献中の具体的な意味内容ではなく、被引用文献そのものを引用する場合があります。この場合においては、被引用文献の内容それ自体は引用文献においては記述されない。

その他の 16 個は用語ではない語句であるものである。以下に具体例を示す。次に示す例では、“the relational charts for the presentation of relative indicators” が引用対象概念表現である。

Inspired by **the relational charts for the presentation of relative indicators** (see Braun et al. 1985; Schubert and Braun 1986), we show the plot of the relative citation rate (RCR, see Schubert and Braun 1986) versus relative usage rate (RUR) in this study.

この例に示されるように、図や質問紙は言語によって表現されるものではなく、このような言語によって表現されるもの以外を引用の対象とする場合が存在する。

A04\_Artifacts に分類されるもので固有名であるものも同様に大きく 2 つに分類することができる。1 つは被引用文献そのものを示すものであり、これが 69 個中 29 個を占める。以下の例では “Li and Sun (2014)” が引用対象概念表現となっている。

As a result, they improved the F-measure by about 0.05% compared to **Li and Sun (2014)**.

この例のように、固有名の場合も文献それ自体を示す表現が引用対象概念表現となる場合が存在する。

その他の 40 個は被引用文献そのものを示さないものである。以下に例を挙げる。

We randomly extracted 2,010 nonduplicate metadata records from two different digital collections: **The UNT Library Catalog** (University of North Texas Libraries, n.d.), and **the Portal to Texas History** (University of North Texas Libraries, 2016).

この例のように、固有名であるものでも、用語ではない語句で引用の対象となっているものがある。

### 3.2 節レベルの引用対象概念表現の分析

項目	値	比率	個数
過程型	物質過程	0.364	356
	心理過程	0.001	1
	関係過程	0.345	337
	行動過程	0.220	215
	存在過程	0.045	44
	発言過程	0.023	23
時制	現在	0.868	848
	過去	0.131	128
完了相	未完了	0.850	830
	完了	0.149	146
モダリティ	No	0.891	870
	Low	0.077	76
	Middle	0.021	21
	High	0.009	9

表 2: 節の比率

表 2 を参照しつつ節レベルの分類の各項目について、報告する。過程型では物質過程と関係過程が多いことがわかる。前者は、実験などの具体的な操作、後者は語句の説明などが影響していると考えられる。時制については、過去、完了形ともに比率は 15% 程度である。このことは、過去の事例を示すような形で引用がなされる節の比率を示唆していると思われる。モダリティについては、多くの文でモダリティは示されないが、これは引用においては過去の事例の引用や他者の見解など、著者の主観が入ることが少ないという引用そのものの性質と、学術論文という対象データの性質の両方に由来すると考えられる。以下、具体的にどのような節が引用の対象であるのかを分析するため、意味の中心である過程型について比率の大きい類型である物質過程、関係過程、行動過程に属する節から各 100 個を無作為抽出して分析する。分析の結果、共通して具体的な事例を記述する節と一般的な事柄を記述する節に分けられると考えられた。

物質過程に該当するもので、具体的な事例を記述する節は 100 個中 54 個であり、一般的な事柄を記述する節は 100 個中 46 個であった。関係過程に該当するもので、具体的な事例を記述する節は 100 個中 19 個であり、一般的な事柄を記述する節は 100 個中 81 個であった。行動過程に該当するもので、具体的な事例を記述する節は 100 個中 60 個であり、一般的な事柄

を記述する節は100個中40個であった。具体的には以下のような節のことである。

**Kuzi et al. (2016) train word embeddings specifically based on the document collection on which queries will be executed.**

この例は個別の研究活動<sup>6</sup>の報告である。

ここで示した具体的な事例を記述する節に関しては、個別の事例について言及する節であるか、複数の研究活動の事例を統合した形で言及する節であるかという、より詳細な分類を行うことができる。複数の研究活動の事例をまとめた、つまり複数の被引用文献をまとめた形で引用するものは以下のようなものである。

**Some reviews categorize the existing research by units of analysis, summarizing works at a micro, meso, and macro level (e.g., Brush 1992; Brush et al. 2009).**

このような引用対象概念表現は、複数の被引用文献を対象にした引用である。

以上で示したように、節のレベルでは、引用対象概念は大きく具体的な事例と一般的な事柄の2類型に分けられ、具体的な事例に関してはさらに複数の研究活動の事例を統合したものと、個別の事例に言及するものがあることがわかる。このことから、引用対象概念表現は以下のような多様な形を有するということがわかる。つまり、(1)一般的な事柄に言及する節、(2)個別の具体的な事例に言及する節、(3)複数の研究活動の事例にそれらをまとめる形で言及し、被引用文献と1対1対応の関係を有さない節である。

このことから、引用対象概念が単純に被引用文献と1対1の対応関係にあるのではなく、被引用文献と引用対象概念、ひいては引用文脈の関係はこれまで考えられてきたような被要約文書と要約文書のような関係で捉えること(e.g. [4])ができないこと、そして引用文脈を研究において用いる際には、それがどのようなものかの考慮の必要性を示唆する。

## 4 おわりに

本研究は、引用対象概念表現の分析を通じ、引用の対象の様相を示すことを目的にした研究である。本研究の結果、語句については、全体として用語が半分以上を占めるものの、用語ではない語句に対する引用が存在することを示し、個別の事例分析では、(1)引用

の対象には言語によって表現されないものも含まれること、(2)固有名か否かに関わらず、文献それ自体が引用対象概念表現となりうること、(3)用語でなくとも引用対象概念表現になりうることの3点を示した。(1)は、言語処理の分野において、一般に図表は処理の対象とならないが、引用においては考慮する必要があることを示すものであり、言語表現に限定した処理の限界と過誤の可能性を示唆するものであるため、特に重要であると思われる。節については機能文法における過程型の枠組みを用いて分析を行い、その中心である過程型については物質過程、関係過程、行動過程が大きな割合を占めることを明らかにした。このことに加え、個別事例分析では節レベルの引用対象概念表現が、(1)一般的な事柄に言及する節、(2)個別の具体的な事例に言及する節、(3)複数の研究活動の事例にそれらをまとめる形で言及し、被引用文献と1対1対応の関係を有さない節のように多様な様相を示すことを指摘した。

## 参考文献

- [1] Joan M. Reits (2004) *Dictionary for Library and Information Science*, United State of America, Libraries Unlimited.
- [2] Henry G. Small (1978) "Cited Documents as Concept Symbols," *Social Studies of Science*, vol. 8, no. 3, p. 327-340.
- [3] Lutz Bornmann, K. Brad Wray, Robin Haunschild (2019) "Citation Concept Analysis (CCA): A New Form of Citation Analysis Revealing the Usefulness of Concepts for Other Researchers Illustrated by Two Exemplary Case Studies Including Classic Books by Thomas S. Kuhn and Karl R. Popper," *Scientometrics*, p. 1-24, available from: <https://arxiv.org/pdf/1905.12410.pdf> (2020-1-10).
- [4] Arman Cohan, Nazli Goharian (2017) "Scientific Document Summarization via Citation Contextualization and Scientific Discourse," *International Journal on Digital Libraries archive*, vol. 19, no. 2-3, p. 287-303.
- [5] In-Su Kang, Byung-Kyu Kim (2012) "Characteristics of Citation Scopes: A Preliminary Study to Detect Citing Sentences," *Computer Applications for Database, Education, and Ubiquitous Computing*, p. 80-85.
- [6] Young Man Ko, Min Sun Song, Seung Jun Lee (2016) "Construction of the Structural Definition-based Terminology Ontology System and Semantic Search Evaluation," *Library Hi Tech*, vol. 34, no. 4, p. 705-732.
- [7] Michael Alexander Kirkwood Halliday (1994) *An Introduction to Functional Grammar*, 2nd ed., London, Edward Arnold.
- [8] Simon C. Dik (1997) *The Theory of Functional Grammar: The Structure of the Clause*, 2nd ed., New York, M. de Gruyter.

<sup>6</sup>“研究活動”について、本研究では研究の対象ではなく、“investigate”や“analysis”のような動詞が中心となる節で表現される、研究を行うことそのものを指す語句として用いる。