

# chapter. 14

## 研究者による情報発信としての 「学術ウェブサイト」の評価の行方

高橋晃一

### 1. 研究者による情報発信はもっと積極的にされてよい

インターネットが普及し、研究環境は大きく様変わりした。かつては、本を探すならカードの図書目録を調べていたが、いまでは大学のオンライン蔵書目録を使う。キーワードを入れるだけで、蔵書に関するさまざまな情報が簡単に手に入る。さらに CiNii Books で検索すれば、全国の図書館の所蔵状況が瞬時にわかる。また、書籍だけでなく論文も CiNii Articles で簡単に見つけることができるようになった。さらに出版形態そのものも変化しはじめている。従来は印刷物として刊行されていた学術雑誌が電子ジャーナルになり、PDF で公開されるようになってきた。これに伴ってオープンアクセスが推奨されるようになり、大学や研究所の機関リポジトリも充実し、いまでは図書館に行かなくとも、論文の PDF が手に入る。また学協会で発行している雑誌論文も、多くは J-Stage で入手できるようになってきた。情報の利用者、受け取り手という立場で見れば、環境は十分すぎるほどに整備されている。しかし一方で、個々の研究者からの情報発信が有効になされているのかどうかはわからない。電子ジャーナルの普及とオープンアクセスの提唱によって、研究成果が広く知られるようになったのは事実だろう。ただし、これは従来、紙に印刷し製本していたものを、PDF で出版するようになっただけで、単なる媒体の変化にすぎない。多くの場合、研究者個人が情報発信のための技術や知識を身に付けているわけではない。しかし、誰でも容易にウェブサイトを開設できる状況を鑑みれ

ば、研究者による情報発信がもっと積極的になされてもよいのではないだろうか。

## 2. 2001年の研究者自身による「オープンアクセス」

そのような、研究者が研究成果を公開するためのウェブサイトを仮に「学術ウェブサイト」と呼ぶことにする。そのような試みはすでになされている。次に紹介するのは、「担保する」という言葉の使い方について、Google 検索したところ、偶然見つけたものである。

ウェブブラウザのアドレスバーに「担保する」と入力すると、Google の検索結果では、最上位に「インターネットを利用して「担保する」を解く」という論文が表示されている。検索結果の表示から、著者は野浪正隆氏、その所属は大阪教育大学、などおよその情報がわかる。クリックすると情報のソースになっているウェブサイトに切り替わる<sup>1)</sup>【図1】。

表示された野浪氏の論文は PDF ではなく HTML で作成されている。操作作用のインターフェースなどはなく、上から順にスクロールして読み進めるようになっていく。論文を読んでみると、「担保する」という表現について、インター

**インターネットを利用して「担保する」を解く**

野浪正隆

**0. はじめに**

**0.1 「担保する」との遺蹟**

近頃「担保する」という言葉をよく聞く。最初は違和感があった。「担保」は「保証」と同じように借金のかたであってモノ概念なのに、スル+つててコト概念になった場合などのような意味になるの不明であったからである。

昨年、住んでいる集合住宅の管理組合の理事をして、住宅保険の説明文書に「担保する」が使われているのを「発見」した。将来損害が生じた場合にそれを補償するのが「保険」だから、「保険スル」でもいいのだから、「〇〇保険は、交通事故を保障します」ではわかりにくくなるので、別の勘弁を使うことにしたのであろうか。保険の広告のキャッチコピーが「担保する」は使われていない。「担保する」という語のかわりに、「保証」が「生涯保証」などという形で使われている。しかし、説明書や契約書には、「担保する」が使われている。語の意味が一意でないといけないので、法律用語としての「担保する」を使うのである。広広に使わず契約書に使うところから判断すると、保険会社は「担保する」が一語に満ちる語でないと考えていることが分かる。

これも、昨年あたりに「発見」したことである。ある教授が教授会での発言中に「担保する」を使っていた。メモを取っていないので不正確だが、「真意の自覚性を内視で担保する」というような内容の発言だったと記憶している。これは、新しい「担保する」である。内視は、将来の損害を補償する機能を持たないからである。「保証」でもないし、「保護」でもないし、「補償」でもないし、損害を前提としていないからである。

「担保する」くらの意味かと思ふ。

このような「発見」から、「担保する」について少し調べてみる必要があると考えようになつた。

**0.2 インターネットページのコーパスとしての有効性**

語彙研究や文法研究や音声研究にコーパス（電子化された言語データ）が利用されている。コーパスに関する二つの言葉を引用しておく。

1. 新しいデータ収集の方法：コーパスの利用 従来、言語研究に利用するデータには、(1) 研究者自らが筆物、録音、新聞などから集めた原文、(2) 辞書、文法書、論文、著書などに引かれている例文、(3) 母語話者が意識によって作った例文、(4) 非母語話者である研究者が作成した例文を母語話者に判断してもらったものなどがあつた。今日では、コンピュータ技術とインターネットの普及によって、新しいデータ収集の方法がもたらされた。コーパスの利用によるデータ収集である。コーパスによるデータ収集は、従来の手法に取って代わるものというより、むしろ相互補完的に使われるべきものであるが、必要なデータを大量に提供し、新たな言語現象を示してくれる豊富な手段であることは疑いない。

コーパスと語彙学 野浪正隆 編著  
『英語教育』(大修館) 平成17年2月号(pp. 43-45)

一方、音声情報処理システムの研究・開発を行うためには、分析・合成・認識の各種の手法を適切に比較・評価することが必要とされる。これを行う方法としては現在のところ、共通の音声データを用いてこれらの処理を行い、その結果を比較するという方法以外は知られていない。このようなことから、共通利用可能な各種・大量の音声データを収録し、保管・公開することは研究・開発過程での利用および評価システムの性能評価の両面から求められる。このような目的に利用される音声データを一般に音声データベースあるいは音声コーパスと呼んでいる。

研究・開発過程での利用および評価システムの性能評価 野浪正隆 編著

研究の評価という点では認定されたコーパスが必要である。しかし、流行語や流行しないかもしれないけれども現時点で使われている語（や構文や文法構成法）については、認定されたコーパス内に含まれていない

図1 野浪正隆氏のウェブサイト

ネットのサーチエンジンをコーパスとして利用し、調査した結果をまとめたものであった。野波氏が調査した時期は、論文中の記述では2001年11月18日となっているが、ウェブ上の論稿には公開年次が明記されていない。論稿末尾の欄外に「この論文は「学大國文」45号（大阪教育大学 国語教育講座・日本アジア言語文化講座）に掲載したものをHTML化したものである」と記されているので、大阪教育大学で発刊している『学大國文』という雑誌が出典だということはわかる。大阪教育大学附属図書館のホームページ (<https://www.lib.osaka-kyoiku.ac.jp/>) にアクセスし、「蔵書検索」で探すと、当該の雑誌は2002年3月に出版されていることがわかった。

実はこの年はオープンアクセスの歴史にとって記念すべき年で、Budapest Open Access Initiative (BOAI) がオープンアクセスに関する宣言を発表したのが2002年2月14日であった。大阪教育大学リポジトリの運営指針は2008年11月5日付になっているので、おそらくリポジトリの開設に先んじて、著者の野浪氏がウェブサイトで自分の論文を全文公開していたのであろう<sup>2</sup>。野波氏がオープンアクセスを意識していたかどうかはわからないが、自分の論文をウェブ上で公開しようとする試みは当時の時流に合致しているだけでなく、研究機関の動向に先んじていたことになる。ちなみに現在では大阪教育大学のリポジトリで、野浪氏の論文のPDFも閲覧できる。これはオープンアクセスという世界的な動向の過渡期に起こった逆転現象であり、研究者個人のアーカイブによるオープンアクセスの努力の跡とも言えるが、いまとなってはリポジトリにあるPDFと個人のウェブサイトのHTMLは、一見すると原本と写しの関係になってしまっている。

ついであるが、Googleをコーパスとして利用する試みに関して言えば、昨今では国立国語研究所でウェブを利用したコーパスの構築が大規模に行われているので、2001年当時とは状況がだいぶ変化してしまっている。野浪氏は結論の中で、「研究のための資料をいかに集めるか整理するかが、研究時間の大部分を占めていた時代から、資料そのものがすでに用意・整理されていて、それをいかに分析するかに時間を使う時代が変わりつつある」と述べているが、資料の用意・整理の仕方自体も組織的な活動が中心になり、状況は大きく変化している。加えて、野波氏の論稿では参照した資料にハイパーリンクが張って

あるが、現在ではほとんどのウェブサイトがリンク切れになってしまっている。これは著者の責任ではないが、ウェブ上の情報の維持管理も非常に大きな問題だということがわかる。

### 3. 個人運営のデータベース Digital Dictionary of Buddhism

しかし、ウェブサイトはその特性を活かせば、有効な研究成果の発信手段になり得る。特にコーパスや電子辞書のように情報を追加していく可能性がある場合、ウェブのように随時更新可能な媒体を用いる方が都合がいい。仏教学の分野では、チャールズ・ミュラー博士の Digital Dictionary of Buddhism がその先駆けで、おそらく個人で運営する学術ウェブサイトとしては最大規模のものの一つであろう<sup>3</sup>。筆者自身も「仏教用語用例集」(通称バウツダコーシャ)というウェブサイトを設計した経験がある。これは仏教の専門用語に特化した資料集である<sup>4</sup>【図2】。

ところで、これらは電子ジャーナルやリポジトリでの論文 PDF と異なり、単にブラウザ上に表示されている見かけの部分だけでなく、データの構築状況やその処理方法が非常に重要な意味を持っている。例えば、ウェブ上で単

**Buddha Koša**  
 仏教用語の用例集 (バウツダコーシャ)および現代基準訳語集

HOME:プロジェクトについて | 仏教用語用例集 | ニュースレター | スタッフ紹介  
 仏教用語の定義的用例集は2014年3月5日に一部更新されました。  
 2015年11月27日に「パーリ文獻の五七五法対照篇」を追加しました。  
 日版:2011-2013年版/2014年版

五位七十五法関連用語の定義的用例集

五位七十五法一覽表 (アビダルマ術語集)

項目をクリックすると、アビダルマ文獻の用例を参照することができます。

色 (11)	視 (0)	caṅkue (觀) śrotra (經) ghrīṅga (鼻) jīvā (衆) kāya (身)
無表 (1)	視 (0)	rūpa (色) ābhā (變) gandhu (香) rasa (味) sroṅṅavya (觸)
心 (1)		avijjāṅgi (無智)
大地 (法 (10))	心 (1)	citta (心) (= vijñāna (識) - manas (意) )
大善 (法 (10))	大地	vedānā (受) saṃjñā (想) cetanā (慧) chanda (欲) aparā (觸) praṅṅā (觸) amṛti (樂) manasikāra (作意) adhimokṅga (勝解) samādhi (三摩地)
大福 (法 (10))	大善	śraddhā (信) vīrya (勤) upeṅṅā (持) hri (樂) apatrāṅṅa (樂) aloha (樂) adveṅṅa (無礙) avihapā (不著) prestrabhi (釋安) apramāda (不放逸)
大福 (法 (10))	大福	āvidyā (無明) pramāda (放逸) kaṅṅidya (懈怠) āradḅhya (不悔)
有為 (法 (70))	心所 (法 (16))	stāyāna (勝況) suḅḅhatya (歸勝)
大不 (法 (10))	大地	āhīkya (無懼) anapatrāṅṅa (無懼)

図2 「仏教用語用例集」ウェブサイト

に文章を表現するだけなら、見出しと段落の構造を整理すれば十分であろう。HTML のタグは <h1> や <p> だけでも十分表現できる。そのほか、リンクを張るための <a> や強調のための <b> などを使えば、見栄えはともかく最低限の内容は表現できるだろう。さらに PDF ならそうしたウェブサイト構築の知識も必要なく、ワードなどで作成したドキュメントを PDF に出力すればよい。この場合は、文章の構造をタグを使って表現する必要もない。

しかし、電子辞書やテキストデータベースの場合、表示結果の見かけだけを表現しても意味がない。例えば電子辞書なら、「登録語彙」「品詞の分類」「意味」「用例」などを区別して整理しなければ、電子データとしては使い勝手が悪い。また研究者が扱うテキストであれば、異読<sup>いどく</sup>の情報が電子データとして整理されていることは非常に重要であり、また、一般名詞と専門用語、固有名の区別などがされてないと、場合によっては膨大な数の検索結果が与えられ、それに目を通すことに無駄な労力を割くことになるだろう。また、研究者がデータを作成する場合、自身の研究成果をデータに反映させる方法があれば、ありがたい。例えば、あるプロジェクトで文献の電子データを作成しているとする。入力している過程で、多くの引用があり、調べた結果、いくつかの引用文の出典が明らかになったとする。従来なら脚注を施すことになるが、電子データであれば、出典の情報などをテキストとは別なレベルで記述することができる。そのように電子データとして整理しておけば、用途に応じて引用であることを強調表示したり、出典の表示・非表示を選択することもできる。さらには引用の情報だけを抽出して整理することもできる。また検索の際に引用文だけに検索をかけたり、逆に引用でない箇所だけに検索をかけることもできる。人名や地名、専門用語などについても同様で、電子データはそれらをあらかじめ区別して記述しておくことができる。一昔前にはカードを取るという作業をしていたことが、電子データの中にメタレベルの情報を記述しておくことで、一つのファイルとして管理することができる。また複数の人との情報共有も容易になる。

#### 4. XML、TEI

こうしたことを可能にするために XML が有効だということは次第に認識さ

れつつある。特に人文学の分野で、XML を用いて情報を整理するための統一規格として TEI のガイドラインが浸透しつつある<sup>5</sup>。先に紹介した「バウダコーシャ」もソースファイルは TEI のガイドライン (P5) に準拠し、XML で作成してある。ところで、テキストデータにタグを付す作業は、単純に表面的な体裁を再現するだけなら、テキストの内容に関する知識がなくてもできるだろう。例えば、章に <chapter>、段落に <p>、引用文に <quote> と付けていけばいい。しかし、実際に文献研究者が求めているものは、単なる体裁の再現ではなく、内容の分析に資するデータであろう。また研究者が電子データの製作者であるなら、テキストの構造を細かく分析し、その結果を電子データに反映させたいと考えるだろう。その結果としてタグ付けは複雑化するかもしれないが、そこには研究者独自の見解が反映されてくることになる。この段階になると、単なる電子テキストというよりは、校訂テキストや翻訳のように、個別の研究成果と言うべきものになってくる。つまり、XML でテキストを構築することは、従来人文学者が試みてきたことと、本質的には何ら変わりがない。よりよく構築されたデータもあれば、役に立たないような構造を持ったデータもあるだろう。つまり、ブラウザ上の表示結果ではなく、XML の構造自体が人文学者の知見を反映しているのであり、それに対して適切な評価がなされなければならない時代になりつつある。

## 5. 電子データをどう提供するか

そうした観点で電子データを評価する試みとして、NINES の peer review の制度がある<sup>6</sup>。そこでは三つの基準があげられている。一つは、電子データの理論的根拠と内容であり、ウェブサイトの内容と質を印刷物と同じ既存の評価基準で判定する。二つ目はインターフェースのデザインと使いやすさで、ほかのデジタル資料との相互運用性、開発者の自発性、斬新さ、独創性を評価対象としている。最後に持続可能性を重視し、独自のフォーマットを極力避けるように勧めている。NINES の基準で示されているように、まず何より研究の質は電子化という行為によって決まるのではなく、その内容によっている。これは電子データであっても、紙の印刷物であっても変わらない。人文学の研究成

果に対する評価は、この点で従来通りの価値基準を保つべきであり、電子化という一点だけで評価されてはならない。それに加えて、電子データの利用方法を的確に理解し、電子データならではの情報提供の仕方が評価される必要がある。直観的に利用できるデザインや、外部の情報との適切なリンクなどは評価されるべきであり、さらに何らかの新しい知見を与える試みがあればそれも評価に値する。最後の基準は電子データを作成する際の約束事を逸脱していないことを求めている。データの構築法が独自のものだと持続的な管理や情報共有に多大な支障をきたす可能性がある。電子テキストの構築は既存の安定した技術に基づいてなされるべきである。そのために NINES は TEI のガイドラインを使用することを推奨している。

日本でも TEI の活動は知られはじめている。しかし、TEI を利用した電子テキストを積極的に評価する動きはまだない。テキストの電子化は今後、ますます進んでいくことは間違いないであろう。そうした成果をどのように評価すべきか考える時期に来ている。

#### 注

- 1 URL は <http://www.osaka-kyoiku.ac.jp/~kokugo/nonami/ronbun/tanporon.html> (2019 年 7 月 30 日閲覧)
- 2 BOAI 宣言文 : <https://www.budapestopenaccessinitiative.org/read> (2019 年 7 月 31 日閲覧)  
大阪教育大学リポジトリ運用指針 : <http://goose.bur.osaka-kyoiku.ac.jp/doc/public/rule/579.html> (2019 年 7 月 31 日閲覧)  
ちなみに東京大学学術機関リポジトリ (UTokyoRepository) は 2006 年 4 月 1 日に公開されている。
- 3 <http://www.buddhism-dict.net/ddb/> (2019 年 8 月 31 日閲覧)
- 4 [http://www.l.u-tokyo.ac.jp/~b\\_kosha/start\\_index.html](http://www.l.u-tokyo.ac.jp/~b_kosha/start_index.html)
- 5 <https://tei-c.org/guidelines/> (2019 年 8 月 31 日閲覧)
- 6 <https://nines.org/about/scholarship/peer-review/> (2019 年 8 月 31 日閲覧)