

『計量国語学』アーカイブ

ID	KK300601
種別	特集・招待論文A
タイトル	Webコーパスの概念と種類, 利用価値 —語史研究の情報源としてのWebコーパス—
Title	The Concept, Types and Utility of Web Corpora: Web Corpora as a Source of Information for Etymological Studies
著者	田野村 忠温
Author	TANOMURA Tadaharu
掲載号	30巻6号
発行日	2016年9月20日
開始ページ	326
終了ページ	343
著作権者	計量国語学会

特集・招待論文 A

Web コーパスの概念と種類, 利用価値

—語史研究の情報源としての Web コーパス—

田野村 忠温 (大阪大学)

要旨

Web コーパスの満たすべき定義的な条件は“インターネット上にある文書を大量に収集、蓄積したもの”ということであろう。Google ブックス、国立国会図書館デジタルコレクション、各社新聞記事アーカイブなどのサイトはその条件を満たさないが、Web コーパスに連続する面を持つ。そうしたサイトを Web コーパスと見立てて利用することにより、近現代の語史の考察のための強力な情報源とすることができる。近年そうしたことが可能になったのは語史研究の方法論に関わる革命的な出来事と言っても過言ではない。

キーワード: Web コーパス, 通時コーパス, 近現代語, 語史, 「立ち上げる」, 「電視」

1. はじめに

コンピュータやインターネットの発達と普及に伴い、言語の研究に利用できるさまざまなものが出現している。そして、一見言語研究には関係が薄そうでも実は大きな利用価値を秘めているものも少なくない。表面的な印象や固定観念にとらわれることなく、個々のものの価値を見抜き、研究に役立てていくことが重要である。

この小論では、インターネット上のさまざまなサイトをコーパスと見立て、そこで得られる情報を総合的に利用することにより、近現代の語史の記述を従来の水準から容易に引き上げることができることを示す。いくつかの研究事例はすでに田野村 (2015a, 2015b, 2016a, 2016b) で述べた。ここでは、それらの考察で用いた方法とその背後にある筆者の考えを明らかにする。

ここでの筆者の関心は、ひとりコーパスでもなければ Web コーパスでもなく、言語研究におけるコンピュータ技術の適用の可能性全般にある。それは、言語研究の一般的、常識的な枠組みや方法論にはとらわれず自由な発想で物事を考えたいということにほかならない。そして、研究を特定の資料や方法の範囲内で完結させなければならないという発想も筆者にはない。研究上有用なものは適宜組み合わせればよいという考えに以下の論述は基づく。

2. Web コーパスの意義

周知の通り、言語研究の世界において、コンピュータ技術の応用として近年広く注目を集めているのはコーパスの使用である。例えば、現代日本語について言えば、国立国語研

究所で5年の期間を費やして作成された『現代日本語書き言葉均衡コーパス (BCCWJ)』(2011年完成・公開)がある。

そのようなものがあるのになぜデータの素性もはっきりしないインターネットに頼る余地、価値があるのか。それには2つの理由がある(田野村(2012, 2014))。

1つは、手作業を介して作成されるコーパスには、作成に要する時間や費用の関係上、自ずと厳しい量的な制約が課せられるということである。現代日本語の計画的に作られたコーパスとして現在最も規模の大きいBCCWJも例外ではない。BCCWJはもしその全体を丸ごと使うとしてもデータ量は小説単行本に換算して1,000冊程度である。これは、文庫本の形にして並べれば書架の2連にも満たない量である。それだけのデータ量で足りる研究テーマもあるが、それでは得られる用例が不足し、満足に研究できないテーマも当然多い。後者の種類のテーマについては、BCCWJの何百、何千、何万倍の規模を持つWebコーパスの利用により解決が得られる可能性がある。

もう1つの理由は、BCCWJは特定の基準に基づいて収集されたテキストの集合体であり、それが日本語の現実のモデル化として唯一的であるわけではないということである。それはBCCWJに固有の問題ではない。いかなるコーパスも言わば“箱庭”であり、その中に日本のあらゆる風景を文句のない形で再現するということはもとより望めない。残存する資料の少ない過去の言語や死語を別として、これさえ調査すれば万全だと言えるようなコーパスが作られることは将来的にもあり得ない。その問題はここでの主題であるWebコーパスにも当てはまるが、特定のコーパスとは異なるコーパスを使うことによって新たな日本語の姿を観察することが可能になる。そして、Webコーパスには、既存のコーパスでは収集の対象外とされたさまざまな種類や時期の日本語のデータが大量に含まれる。使うコーパスごとに異なる結論が得られればその異なりをどう解釈するかという問題が発生するが、それは困った事態でも回避すべき事態でもなく、日本語に関するより深い理解に至るための出発点である。

3. Webコーパスの概念と種類

さて、Webコーパスについて論じるには、まずWebコーパスが何であるかということをはっきりさせる必要がある。しかし、これは一見単純そうで、実のところむずかしい課題である。それは、Webコーパスとそうでないものの区別が昨今曖昧になりつつあるからである。

Webコーパスの基本的な条件は、“インターネット上にある文書を大量に収集、蓄積したもの”ということであろう。そのように考えたとき、Webコーパスにはその存在場所と利用形態に関してまず2種類のものを区別することができる。

3.1 第1種 Webコーパス——ローカルデータとしてのWebコーパス

第1種のWebコーパスは、インターネットから大量の文書を収集し、それをローカルな——すなわち、使用者の手許にあり、インターネット上に公開されていない——外部記憶装置(ハードディスクなど)に蓄積したものである。使用者はそのコーパスを自身の技能の範囲内で望む通りに利用することができる。コーパスの内容も当然すべて自由に見ることができる。

第1種 Web コーパスは上述の Web コーパスの基本条件を忠実に満たす。

3.2 第2種 Web コーパス——サーチエンジンなどの背後にある Web コーパス

第2種の Web コーパスはインターネット上においてサーチエンジンなどのソフトウェア、インターフェースを通じて利用されるものである。この種の Web コーパスも基本的に第1種のそれと同性質であるが、コーパス自体は公開されておらず、その内容を外部の者が直接に知ることはできない。使用者に許されるのはもっぱら、サーチエンジンの表示する検索結果などを見ることである。もっとも、サーチエンジンには検索結果からハイパーリンクを介してインターネット上のものと文書を参照することのできる仕組みが用意されている。したがって、運よくもとの文書が残っており、かつ、内容の改変がなければ、その限りにおいてコーパスが公開されているのと事実上同じであることになる。

第2種 Web コーパスは多くの場合第1種のそれに比べてデータ量ははるかに大きい。しかし、それでもサーチエンジンが検索するのはインターネット自体ではなく、そこから収集した範囲の文書の集積に過ぎないことは注意しておいてよい。

第2種 Web コーパスも概して Web コーパスの基本条件を満たすが、すぐ下で述べる通り、一部に例外的な事実もある。

3.3 第3種 Web コーパス——各種のアーカイブ検索サイト

近年、Web コーパスの基本条件を満たさないが一見第2種 Web コーパスのようにも見えるものはいくつも出現している。それは、インターネット以外のところから収集した文書から語句を検索することのできる各種のサイトである。

最大の規模を有するその例は「Google ブックス」(<https://books.google.co.jp/>)であろう。Google ブックスは米国その他の図書館に所蔵された膨大な数の書籍や雑誌を画像化し、それに光学文字認識 (OCR) によってテキストデータを付与し、Google のサーチエンジンと同様の方法による語句の全文検索を可能にしたサイトである。

Google ブックスのデータはインターネット上の文書ではなく通常の出版物の内容の集積であり、したがって、Web コーパスの条件を満たさない。しかし、Google ブックスと Google のサーチエンジンは、見かけや機能の面で類似していることを別としても、2つの点で連続的である。第1に、今では Google のサーチエンジンで語句を検索したときに Google ブックスの検索結果も同時に検索、表示されることがある。Google のサーチエンジンは実のところもはや純粋な Web コーパスとは言えないのである。3.2 の最後に述べた例外的な事実とはそのことを指している。また、逆に、Google ブックスで検索して該当する書籍、雑誌がない場合は、Google のサーチエンジンの検索結果が表示されるようになっている。このように、Google ブックスと Google のサーチエンジンは相互に重複した形になっている。第2に、今やインターネット上の文書とそうでない文書とのあいだに明確な区別がない。例えば、出版の電子化が普及して以来、同一の文書がインターネットに掲載されると同時に紙媒体で出版されるということがよくある。そのようなとき、サーチエンジンは当該の同一文書のインターネット版、Google ブックスは印刷版を収集して検索の対象とすることになる。以上のような連続性を拠り所として、ここでは Google ブックスを第2種 Web コーパスの延長線上にあるものと見なすことにする。

Google ブックスを Web コーパスの一種と見るとすれば、それに連続するいろいろなものも視野に入ってくる。日本語に関して言えば、「国立国会図書館デジタルコレクション」のサイト (<http://dl.ndl.go.jp/>) では明治・大正期を中心とする多数の書籍、雑誌が画像化されて公開され、全文検索の機能は残念ながら実現していないが、書名や章、節の見出しなどに基づく検索ができるようになっている。また、大新聞のアーカイブのサイト (有償) や国立国会図書館の「国会会議録」のサイト (<http://kokkai.ndl.go.jp/>) も Google ブックスや国立国会図書館デジタルコレクションと同じく、本来インターネット以外のところにあった文書から語句を検索できるようにしたものである。

Google ブックス、国立国会図書館デジタルコレクション、新聞記事アーカイブ、国会会議録などを第 3 種の Web コーパスとした場合、第 1 種、第 2 種の Web コーパスになく多くの第 3 種の Web コーパスにある共通の特徴はデータの歴史性である。第 1 種、第 2 種の Web コーパスに含まれる言語データは基本的に現代、今日のものである。その中に古い言語データは多く含まれず、含まれていても新しいデータの中にただ混じっているだけなので、それを時間の観点から区別して扱うことができない。¹ それに対し、第 3 種 Web コーパスでは過去の言語データを検索することができ、個々の用例の年代も知ることができる。第 3 種 Web コーパスはそれぞれに収録したテキストの種類が異なり、対象とする期間も異なるが、その特性を生かして利用することにより、近現代語の有用な研究手段とすることができる。

以下では、語史の考察に第 3 種 Web コーパスを役立てる可能性を 2 つの事例の簡単な考察を通じて示す。Web コーパスは用例の発見に用いるだけなので、個々の用例をどのコーパスから得たかといったことに関する細かい記述は省く。単に各種アーカイブのサイトで検索してその結果を利用するだけのことである。ただし、Google ブックスは研究上の利用に際して注意を要する問題が多く、それについては後に 1 節を設けて述べる。

4. 「立ち上げる」——辞書記述の改善

1980 年代後半から加速したパーソナルコンピュータの普及とともに「(コンピュータを) 立ち上げる」という複合動詞が社会に広まった。「立ち上げる」には「事業を立ち上げる」のような使い方もある。

ここでは国語辞典における「立ち上げる」の扱いを確かめたくて、その記述の妥当性を Google ブックスを中心とする第 3 種 Web コーパスの利用によって検証し、記述の改善の可能性を考える。

4.1 問題の背景と『日本国語大辞典』の記述

筆者が「立ち上げる」という動詞を聞き知ったのは 1979 年のことである。大学にあった大型計算機 (メインフレームコンピュータ) やミニコンピュータを始動する作業が「立ち上げる」と表現されていた。

1 「Internet Archive」(<https://archive.org/>) のサイトにある Wayback Machine のサービスなどはその例外となる。Wayback Machine では過去のインターネット上のページをその収集日——執筆日ではない——の情報とともに閲覧することができる。しかし、少なくとも現状では日本語研究上の利用価値はきわめて限定的である。

語を知った時期と状況まで記憶に残っているのはその語が新奇で抵抗を感じさせるものであったからにほかならない。その抵抗の一因として考えられるのは、「立ち上げる」という複合動詞の与える“「立ち上がる」から無理に作り出されたもの”という印象である。²しかし、核心的な原因は、そのような理屈ではなく、単にそれまで「立ち上げる」という言い方になじみがなかったということであろう。「(ホコリなどを) 舞い上げる」は「舞い上がる」から同等のプロセスによって作り出されたものと見られるが、筆者自身聞き慣れて抵抗を覚えない。また、「浮かび上げる」という、筆者には従来接した記憶もない表現も Google ブックスで検索してみると予想に反して多数の書籍の著者が用いている。そうした用例に接すると、個人的な言語経験の範囲を超えることは内省では知り得ないという、当然と言えば当然のことを思い知らされる。³

『日本国語大辞典』初版第 13 卷(小学館, 1975 年)には次のような「立ち上げ」という名詞の項目——および同義の「立て上げ」の項目——はあるが、「立ち上げる」という動詞の項目はない。

たち-あげ【立上】〔名〕江戸時代, 和船の筒を挟んで立つ二本の帆柱の受け材。たてあげ。たてあま。*改新造積り書「同五尋弁, 立上げ壺本, 二つわり」

同辞典の第 2 版第 8 卷(小学館, 2001 年)では「立ち上げ」の記述が増補され, 次のような「立ち上げる」の項目が新たに加えられている。

たち-あ・げる【立上】〔他ガ下一〕①機械を稼動させる。特に, コンピュータを起動して, アプリケーションを操作できる状態にする。「パソコンを立ち上げる」②新しい企画, 作業などをはじめる。「新しい辞書の企画を立ち上げる」

これは筆者の内省に頼って読めばおおむね妥当な記述のように感じられる。自分が日ごろ接する 2 通りの用法が記述されているからである。しかし, 例文がいかに辞書項目執筆者の関心と経験の範囲内で作られたものであることが記述の信頼感を少なからず損なっている。内省に頼り, 現実の用法を見ていないという疑いを抱かせるからである。そして,

2 “無理に作り出された”ということについて筆者の考えるところを簡単に述べる。「(人が) 立ち上がる」は“人が立ち、そして、その人が上がる(高い位置に移る)”と解釈することができ、「立ち」と「上がる」とで主体が共通しているという点において意味上一貫しているが、それに対し、「(人がコンピュータを) 立ち上げる」では「立つ」のはコンピュータであり(少なくとも人ではない)、「上げる」の主体は人である。その主体の不一貫が「立ち上げる」に“無理”の要素を感じさせるものと思われる。「立ち」が自動詞で「上がる」が他動詞だからその異種の組み合わせに問題があるということではない。

3 直前の注に書いた理解からすれば首尾一貫していないがそれでも現実に慣用化している複合動詞は多い。例えば、「(人が秘密を) 漏れ聞く」は“秘密が漏れ、人が聞く”ということであるし、「(人が知人から物を) 譲り受ける」は「譲り」も「受ける」も他動詞ではあるが、“知人が譲り、人が受ける”という関係にある。「舞い上がる」から「舞い上げる」が作られた(自動詞→他動詞)のとは逆に、「仕上げる」「焼き上げる」から「仕上がる」「焼き上がる」が作られた(他動詞→自動詞)と推定される事例もある。慣用化しているために気付きにくい「仕上がる」も「焼き上がる」も本来例外的、拡張的である。例えば、「でき上がる」に基づく「でき上げる」、「作り上げる」に基づく「作り上がる」のような類似的表現は一般的ではない。もっとも、本文で触れた「浮かび上げる」と同じく、「でき上げる」や「作り上がる」も Google ブックスで調べてみると現実の書籍中に使用が見出される。

筆者も同じく言語に関わる仕事に携わっているから、筆者の感覚と辞書の記述が合致しても、それが日本語全般の観点から事実には忠実であるとは限らない。

『日本国語大辞典』第2版——以後、『大2』と略記する——における「立ち上げる」の記述は果たして妥当なものになっているだろうか。

4.2 Web コーパスを用いた考察

Google ブックスや国立国会図書館デジタルコレクションなどを使って用例を探してみても分かることは、「立ち上げる」の用法は多様であり、『大2』に記述された2種類の語義の範囲には収まらないということである。

『大2』の記述に何よりも補わなければならないのは、“柱、壁、配管などを上方に向けて設置する”という語義である。これは『大2』の記述する比喩的な語義①、②と異なり、物理的、空間的な動作に関わり、より根源的な語義であると見られる。実際、この語義は語義①、②よりも早くから使用が見られ、かつ、現代に至るまで広く使い続けられている。用例をいくつか示せば次の通りである。

換気設備は、外壁室内側にトタン製の筒を立ち上げ、軒下に開口せしめる予定である。

(島之夫『満洲国民屋地理』, 1940 (昭和 15) 年)

この構築に近い小屋を、私は更に秋田県男鹿市船川町門前、即ち男鹿半島の日本海岸に於いて(中略)発見した。異なるところは堅穴を掘らぬことと、柱を立ち上げて壁をつけたことのみである。

(藤島亥治郎「信濃古建築五論」『地方研究論叢』, 1954 (昭和 29) 年)

「うだち」というのは、ふつう切妻屋根の妻の部分に、せいの低い壁を立ち上げたものを指して呼ばれている。

(大河直躬「日本の民家」『月刊文化財』3月号, 1964 (昭和 39) 年)

修理工事と併行して、防水用水道管を立ち上げ、ドレンチャー装置を施しているのは、きわめて画期的なことといわねばなるまい。

(伊藤延男「東大寺金堂(大仏殿)の昭和大修理」

『文部時報』11月号, 1974 (昭和 49) 年)

土台を据え付け、柱を立ち上げ、梁や胴差し、桁をわたし、構造上主要な骨組みをつくる。

(木村了『わかりやすい造園実務ポケットブック』, 2008 (平成 20) 年)

オープンキッチンで、シンクの手元を隠す小壁を立ち上げただけでも給水の音や皿を洗う音は随分遮へいされます。

(佐川旭・林直樹『最高の住まいをつくる「リフォーム」の教科書』,

2014 (平成 26) 年)

柱や壁は建物の基礎や梁などから見てその上方に——しかも通常建築資材を持ち上げることによって——設置するもので、「上げる」という表現の使用に無理がない。配管の施工は下から上に向けて行われるとは限らないが、それでも配管の起点は——そして施工前の資材も——多くの場合地面や床にあり、柱の設置と同じように捉えられるのも自然なことである。

語義①の初出は調査によって確認できた限りでは 1980 年である。次の第 1 の用例中の「SVP」は「サービスプロセッサ」なるもの、「CPU」は中央処理装置（central processing unit）を指す。

SVP は CPU から独立しており、従来のシステム操作卓機能のほか、システム異常状態監視機能、再試行機能、構成制御機能、保守診断機能、およびシステム立上げ機能などを持ち、システム保全を行うための重要な役割を果たす処理装置である。⁴

（岩根雅彦・佐藤文孝「大形^(マ)計算機における保全性設計」

『電子通信学会論文誌』第 J63-D 巻第 3 号，1980（昭和 55）年）

コンピュータを立ち上げ、システムを管理するための OS（オペレーティングシステム）、ついで管理業務用のプログラムを呼び出し、ファイル準備などの入力を終ると、メインワークステーションのディスプレイ上に、メインメニューが表示される。

（生理遺伝部種子貯蔵管理室「遺伝資源管理のシステム化とコンピュータの利用」

『農業技術研究所年報（昭和 55 年度）』，1981（昭和 56）年）

『大 2』は語義①を“機械を稼動させる”ことと記述しているが、あらゆる種類の機械に関して「立ち上げる」が使われるわけではない。そして、「立ち上げる」は“稼動”自体を表すわけでもない。機械を立ち上げるというのは、少なくとも原初的なイメージとしては、コンピュータのように始動に時間を要し、使用者の指示に基づいて複雑な機能を果たす機械について、それを稼動可能な状態にすることであろう。出版物において「立ち上げる」がよく適用されているコンピュータ以外の機械は原子炉である。また、ソフトウェアを起動することも「立ち上げる」と表現されるので——「Word を立ち上げる」のように——、立ち上げが機械の始動に限られるわけでもない。

語義②に該当すると考え得る用例の確認できた初出は 1979 年である。

はたからは困難のようにも見えただろうとも思いますけれども、当のわたしには、いいことだけが思い出されるのです。（中略）苦勞して一つの会社を立ち上げたときの喜びを、わたしはだれよりも味わっているのではないかと——。

（早川種三「会社再建にかけた人生」

NHK 編『わたしの自叙伝（1）』，1979（昭和 54）年）

この用例での「立ち上げる」の対象は『大 2』の語義②の記述にある“企画，作業”ではなく、組織である。このように組織、あるいは、“企画，作業”よりも規模の大きい事業を表す表現を目的語とする「立ち上げる」の使用はこれ以上実例を挙げるまでもなく一般的である。

以上のように、第 3 種 Web コーパスを使って用例を得て考察することにより、「立ち上げる」の辞書記述の問題点と改善の方向を容易に見出すことができる。また、頻度の低い

4 この例では「立上げ」という表記が使われているので、書き手の意図が「立ち上げ」ではなく「立て上げ」であったということも考え得る。しかし、その可能性はきわめて低いと考え、ここでは「立ち上げ」の用例として扱った。

用例も考慮に入れば、「立ち上げる」の用法は上で見てきたことの範囲には収まらない。「立ち上げる」のより詳細な考察においても Web コーパスがいっそう重要な情報源になることは言うまでもない。⁵

5. 「テレビ」——日中近現代の語史の考察

Google ブックスには中国の書籍、雑誌も多数収められている。また、過去の中国の新聞や雑誌を検索できるサイト（有償）も複数あり、古い時期のものでも全文検索が可能なものもある。日本の新聞記事アーカイブは古いものについては通常全文検索ができず、日本の雑誌に関してはまとまったアーカイブがない。この点で、中国は日本よりも大きく進んだ状況にある。

ここでは近代の日中語彙交流に関わる可能性のある、television を表す「テレビ」という語の初期の使用状況を各種の第3種 Web コーパスの利用によって探ってみる。

5.1 宮島（2008）の記述

20 世紀前半の日本語で「テレビ」の語が使われていたことを宮島（2008）が述べている。『日本国語大辞典』（小学館）は初版、第2版ともこの語を収録していない。現在一般に使われず、古典にも現れない語は辞書編集者の意識に上りにくいのであろう。

宮島は、日本で出版された各種の書籍、雑誌、新聞、辞書、事典および少数の中国の辞書などの調査に基づいて「『テレビ』は 1920 年代から 40 年代の日本語のなかで使われていた」とし、「状況証拠からすると、これが日本製、つまり和製漢語である可能性がたかいたおもわれる」と結論付けている。

しかし、上の記述に見る通り、調査対象が極端に日本側の資料に偏っている。そのような調査に基づく語史の考察が危ういことは自明である。宮島が参照している中国の資料は、辞書以外には、1986 年に出た百科事典と中国語版 Wikipedia だけである。調査方法に問題のあることは宮島も認識しており、「日本側の資料だけで中国側の資料がかけられているので、具体的に日本から中国へ『テレビ』という表現がつたわった経路がわからない。それがつきとめられないかぎり、これが和製漢語だという証明は完成しない。」と述べている。⁶そうした認識を持ちながら、なお“和製漢語”の可能性が高いという判断を結論とした宮島には一定の予断があったものと推測される。すなわち、“近代化の開始に遅れを取った中国が日本より先に television の訳語を必要とし、作り出したとは考えにくい、だから中国の資料を調べたとしても結論は変わらないだろう”という考えである。宮島は中国語版 Wikipedia から“1934 年に南京中央大学で television が開発され、「テレビ」がその中国語名として確定された”という記述を引用し、「日本ではそれ以前にテレビの研究がすすんでいた」ことを説明している。しかし、television の研究開発と television の訳語の考案、使用はそもそも相互に独立した事象である。また、Wikipedia の記述が正しいとしても、それは「テレビ」の名称が 1934 年に中国で初めて使われたことを意味しない。“確定”

5 「立ち上げる」の本格的な考察においては、「立ち上がる」や「立て上げる」、さらには「立ち下がる」「立ち下げる」といった関連表現との関係も検討する価値がある。

6 宮島は、引用に見る通り、「和製漢語」を“日本で作られた漢語”ではなく“日本で作られ、かつ、中国に借用された漢語”の意味に用いている。この用語法における和製漢語は以後“”を添えて表示する。

という表現はむしろ「電視」が以前から他の名称とともに使われていたことを示唆する。宮島は予断にとらわれ、一面的、固定的な思考に陥ってしまっている。⁷

語史を予断や現代の資料に頼って論じることにはできない。「電視」の由来を明らかにするには、当該時期の日中両語におけるその使用状況の確認が不可欠である。

5.2 Web コーパスを用いた考察

各種の第3種 Web コーパスを主たる情報源とする調査によって確認できた限りにおいて、日本の出版物における television への最初の言及は、『東京朝日新聞』の1922（大正11）年8月2日の号に掲載されたXY生「写真電送」という記事に見出される。ただし、そこでの「テレヴィジョン」に関する説明は単に静止画の高速伝送ということにとどまる。

動画の伝送であることを明確にした最初の記述は次の1924（大正13）年の雑誌記事に見られる。記事において television は「無線遠視法」「遠視法」と訳されている。

無線遠視法とは、或場所に於て時々刻々に変る景色を障壁に依て見る事を得ない遠方の地点に於て、之と同期的に^(ママ)かはる像としてあらはし観察し得る方法を云ふ。

（高柳健次郎「無線遠視法（一）」『電気之友』第596号，1924（大正13）年10月1日）

そして、「電視」の訳語は、当時天才発明家として知られ、世界先端の水準でテレビの開発に取り組んでいた青年安藤博の書いた一連の雑誌記事に現れる。初出は1927（昭和2）年4月1日の号の記事である。

放送ラヂオとテレヴィジョンとの相違は、音声の変化を加へる代りに電流が凡ゆる被電視物の明暗に従つて制御せられるの一点である。（中略）テレヴィジョンの先覚者達は前々号にも述べた通り被電視物の像を無数の小部分に分割し、その各小部分を同時に多数の電線を以て送信しようとしたので、実用的に何等の成功を見ることができなかったのであるが（後略）

（安藤博「テレヴィジョンに就て」『アサヒカメラ』第3巻第2号，第4号，1927（昭和2）年2月1日，4月1日）

この変調機は二十メートル乃至四十メートル位の短波長送信機に組合されて居るので、茲に於て初めて短波長電波として、世界の隅々まで電視放送が達成せらるるのである。

7 研究史の観点から正確を期して付言すれば、「電視」を“和製漢語”と見る考えは宮島（2008）よりも早く、「日本国語大辞典第二版オフィシャルサイト 日国.NET」にある「日国友の会」のフォーラム（<http://www.nikkoku.net/tomonokai/>）で述べられている。会員「古書人」氏は2006年10月25日の投稿において、東京朝日新聞学芸部編『学界余談 第一編』（興学会出版部，1928（昭和3）年）における「電視」の用例を報告し——宮島もそれを引用している——、「中国ではテレビ局のことを『電視台』といっていますので、『電視』は是非とも立項すべきだと思います。『電視』という言葉は、元々日本人が考えたということをはっきりさせたい。」というコメントを添えている。宮島の考察は、時期と内容から考えて、その投稿に触発され、かつ、「電視」を“和製漢語”とするコメントをそのまま信用し、用例の補充によってそれを論に仕立て上げるという意識でなされたものであったかも知れない。そう思わせるまでに宮島の論述は一面的で、事実の探求ではなく既定の仮説の証明を目指すものという印象が強い。

なお、『学界余談 第一編』所収の television に関する一文は、後掲（5.2）の『東京朝日新聞』1927（昭和2）年12月17日の記事を再録したものである。

(安藤博「無線^{テレビジョン}電視の新發明」『アサヒカメラ』第4巻第4号, 第6号,
1927(昭和2)年10月1日, 12月1日)

計4件の安藤の記事における「電視」の出現のタイミングと様子から、安藤はおそらく訳語を自ら考案したのではなく、2月号の記事の執筆後に何らかの形で学び知ったのではないかと思われる。しかし、その情報源は不明である。

これらに次ぐ「電視」の使用は『東京朝日新聞』の同年12月17日の記事に見られる。

テレグラフやテレフオンに続いてテレヴキジョンなるものが發明せられ、目下世界科学界注視的となつて居る。邦語でテレグラフを電信といひ、テレフオンを電話と呼んで居る以上「テレヴキジョン」を「電視」といふのが正当であるが、電視ではいまだ意味が分らず、かつギリシヤ語の「テレ」は遠方を意味し、「ヴキジョン」は「視る」であるから、「テレヴキジョン」を「遠視術」と命名したのが適當では無いかと思ふ。

(鎌田弥寿治「テレヴキジョンの話」『東京朝日新聞』1927(昭和2)年12月17日)

記事の著者は訳語の候補としての「電視」に対して否定的であるが、「電視」はおそらくそこに述べられた根拠——すなわち、「電信」「電話」の訳語との形式上の統一——に基づいて作られ、普及した(後述)。

他方、中国の出版物における television への最初の言及は、日本より5年ないし7年も早い1917(民国6)年の雑誌記事に見出される。次はその開始部分である。

電話能達数千英里、而対談者僅營欸相接、不能形貌互睹、引為憾事。然近世科学家、漸發明新法、使用電話者、無論相距幾遠、接談之時、並可互見形貌。

(「電話見形」『科学』第3巻第4期, 1917(民国6)年4月25日)

掲載誌の『科学』は、1915(民国4)年に米国コーネル大学留学生任鴻雋^{しゅん}、趙元任らの設立した學術団体である「科学社」——翌年に「中国科学社」と改称——によって発刊された雑誌である。当時すでに最新の科学に関する情報を米国から直接中国に伝えるルートが確立していたことになる。日本から中国に西洋の知識とともに新しい學術用語が集中的に伝播した19世紀末から20世紀初頭にかけての時期に比べると状況はすでに変わっていたのである。上の記事はテレビ電話の実現に向けた研究を紹介するものであるが、文中に television を表す特定の名称は出て来ない。

「電視」の語は、日本での初出よりも1年以上早く、やはり『科学』に掲載された次の記事に現れる。

電視者、欲藉電力而伝達某種現象於在遠距離人民之目中。此举至今、尚為一種希望、其一二不完全之実験室試験、仍未能与吾儕以能信之証明、及可靠之結果、故未可遽言電視之已成功。

(莊智煥「無線電之前途」『科学』第10巻第7期, 1925(民国14)年10月30日)

これに続く用例は日本での「電視」初出の時期に重なる。

自歐戰發生以迄今日，科學家窮究聲光之奧，無微不至。舉凡入目不見之光，充耳不聞之聲，無不探求其高至低之度以窮其極。關於光浪聲浪之振動次數，超過吾人感覺範圍以外者，均在研究之列。邇來天文學家某氏，已能利用紫烈光線，攝取天空影像。英人復能應用此種光線以造電視器具（Television）。日人又能借助此種光線以製攝影機械，其作用可在黑夜偷攝敵影，未遭覺察。凡此種種，概係光學上之新進步也。

（楊復耀「死聲之發明」『申報』1927（民國16）年5月8日）
各種伝通方法史上成功之最偉大者電視居其一。乃在紐約之賠兒試驗室（Bell Laboratory）中，費幾多人之心血，受埃愛武斯（Herbert E. Ives）博士之指導而抵於成。

（「電視之進步」『科学』第12卷第6期，1927（民國16）年6月20日）
一九二七年四月七日下午，美國霍氏（Herbert Hoover）在華盛頓演說，而聽衆却是在紐約，他們非但可以聽見他的高明演說，同時還可以看見他活動的面貌。這豈不是創造了「順風耳」又有「千里眼」的成功麼？美國電話電報公司裏的貝爾（Bell）實驗室發明這種電視的儀器。今年四月七日第一次公開試驗，竟然成功將華盛頓與紐約間二百多哩的距離完全打破，使人類逐漸接近減少許多因距離而發生的隔膜。

（沈嘉瑞「美國電視試驗之成功」『東方雜誌』第24卷第17号，
1927（民國16）年9月10日）

「電視」の初出が中国，日本の順であり，しかも，両者のあいだに明確な時間差がある以上，「電視」「和製漢語」説はひとまず白紙に戻さざるを得ない。筆者は得られた限りの情報に基づく総合的な判断として，「電視」の訳語は日中両語でそれぞれ独自に作られたと推定する。「電視」が一方の国で作られて他方の国に伝播したことを示す事実が確認できないからである。また，もし television に関する日中の記事の一方が他方の翻訳になっているのであれば訳語の借用の可能性も高まるが，両国の記事は原文と訳文の関係にない。television に関する中国の記事の多くは米国の報道を情報源としており，日本語の影響の印象は薄い。加えて，中国では盲人用読書機 Visagraph の発明の報道に「電視機」という訳語を用いており（『科学新語林 新発明的盲人電視機』『学生雑誌』第15卷第12号，1928（民國17）年），中国における「電視」の使用は少なくとも単なる日本語の模倣ではなかった。確かに，上掲の『申報』記事では「電視器具」の話に日本人による軍事的応用の話が続いて日本語の影響を暗示するようにも見え，また，『申報』1932（民國21）年12月8日の鶴「電視」という記事は「電視」を日本人による訳語と説明しているが，いずれも「電視」の日本から中国への伝播を証明するものではない。日中両語における「電視」の使用がいずれかの方向での借用の結果である可能性も排除できないが，それを肯定するにはしるべき証拠の発見が必要である。

調査によって確かめることのできた「電視」の初期の使用状況を年表の形にまとめると表1のようになる。「電視」以外の名称および television を表さない「電視」の使用などは〔 〕に入れて示す。もっぱら「テレヴィジョン」「テレビジョン」という名称を用いる資料は最初の数年のものに限って示す。項目末の「※」は当該の資料が宮島（2008）で言

及されていることを表す。

表 1:「電視」略年表

年	日本	中国
1917(大正6,民国6)		[科学3-4 テレビ電話の話, 特定の名称は不使用]
1922(大正11,民国11)	[東京朝日新聞8/2 テレビジョン(静止画高速伝送の話)]	
1924(大正13,民国13)	[電気之友596 (無線)遠視法, ラディオ・テレビジョン]※	[東方雑誌21-13 電晤]
1925(大正14,民国14)		科学10-7 電視【初出】 [東方雑誌22-6 無線電影]
1926(昭和1,民国15)	[東京朝日新聞5/16,11/1 ラジオテレビジョン, テレビジョン] [読売新聞11/1 テレビジョン] [早稲田大学新聞11/4 テレビジョン, ラジオ映画]	[東方雑誌23-22,23 電伝影, 電影術, 電景, (無線)電伝影術, 電伝影術] [婦女雑誌12-4 電伝活動影片] [申報10/28,11/6 電線電伝影(劇)]
1927(昭和2,民国16)	アサヒカメラ3-2,4 被電視物【日本初出】[, テレビジョン] アサヒカメラ4-4,6 被電視物, 無線電視, 電視放送 [, テレビジョン] 東京朝日新聞12/17 電視[, テレビジョン, 遠視術] [科学の世界7-1 テレビジョン, 遠視機装置] [工業之大日本24-4,5,6 テレビジョン] [読売新聞4/4,5/23 テレビジョン]	科学12-6 電視, 電視装置, 電視之進歩 東方雑誌24-15,17 電視, 電視(的)儀器 [, 電伝形象, 電伝形象, 電伝照相] 申報1/1,5/8,8/2 電視器具[, 電線電(伝)影]
1928(昭和3,民国17)	東京朝日新聞1/16 無線電視機※ 東京朝日新聞社 学界余談1 前年12/17の新聞記事の再録※ 三省堂英和大辞典 電視, 無線電視[, てれゐ'じょん]※ [日本及日本人161 テレビジョン, 無線遠視] [東京日日新聞11/27 テレビジョン, 無線遠視]	科学13-3 電視[, 電景, (無線電)伝声影機, (電)伝影, 電伝動影] 申報10/21 電視, 電視之実現 [学生雑誌15-12 (盲人)電視機(Visagraphの訳語)]
1929(昭和4,民国18)	小泉武夫・広田栄三 誰れにも解る電送写真の原理と実際 電視[, テレビジョン] 石井重美・原田三夫編 最新科学講座 電視[, テレビジョン, 無線遠視] 電気工学18-1 電視, 電視法, 電視方式, 放送電視劇 [, テレビジョン] 軍事と技術3-11 電視, 電視器 無線と実験10-4 電視, 放送電視[, テレビジョン, テレビジョン, テレビ]※ [海軍及海事要覧 昭和4年版 テレビジョン, 遠視法]	申報1/27,12/25 電視, 電視術, 電視之方式 [, 電望, 電送影片]
1930(昭和5,民国19)	加納八郎他 大衆科学の尖端 テレビジョンの知識 電視, 天然色電視[, テレビジョン] 内外社編 綜合チャールズナリズム講座Ⅱ 電視幕, 電視影像, 電視(受)映機, 電視放送, 電視界[, テレビジョン] 電気工学19-1,5,6,8,9,12 電視, 電視の方式, 電視の試験, 電視劇[, テレビジョン] 電気学会雑誌10月号 電視	東方雑誌27-13,14 電視, 電視(収像)機, 有色電視, 電視伝播台 中学生8,10 (無線)電視, (無線)電視機, 電視電話, 双方電視, 電視(方法) [科学15-1 無線電相] [申報2/6 顕像伝影術]
1931(昭和6,民国20)	加納八郎・平田潤雄 テレビジョン 電視[, テレビジョン] 電信電話学会雑誌105 電視, 電視装置, 電視方式※	申報7/13 電視術, 電視電台 [科学15-3 德律維相]

tele- は遠距離を表し電気を表さないのに両国で独自に翻訳して結果が一致するものかという疑問があり得るが, telegraph, telephone の訳語である「電信」「電話」との関係を考えればそれはさほど不思議なことでもない。次の引用に見る指摘, 記述によれば, 「電視」という翻訳は「電信」「電話」の先例にならったもので——この話は先に見た新聞記事にもあった——, 「電気協会」もその訳語を採用したという。

電信のことを英語でテレグラフといひ, 電話のことをテレフホーンといふのは, 諸君の既に御承知のこと、思ふ。(中略)さて、今度は、Tele と Vision が結合して、テレヴキシヨンとなれば Vision が視るといふ意義であるから、従来の日本の慣例によると、電視と呼ばなければならない筈である。Tele に電気といふ意義は少しも無いのであるが、電信や電話は電流や電波を利用して、これを云ふものであるから、日本では信や話に電の字を冠して、電話、電信と呼んだものと見える、然るに、テレヴキシヨンを電視といふては目下のところ一般には一寸分りにくいかも知れないが、電気協

会でも「電視」といふことに決定して居るから、兎に角電視として御承知願ひ度い。
(小泉武夫・広田栄三『電送写真の原理と実際』, 1929 (昭和 4) 年 11 月 3 日)

中国語について言えば、年表に見る通り、日本語に比べてはるかに多様な **television** の翻訳が試みられているが、「電視」の初出前後から「電語」「電景」「電影術」「電伝形術」などのようにあらゆる訳語に「電」を用いている。そして、調査の限りにおいて、**tele-**を「遠」と訳そうとした形跡すら認められない。**television** だけでなく、文書や写真を遠隔地に伝送する装置を表す *télautographie*, *téléphotographie* という名称も「電書」「電相」と訳されている(黄涓生「電書与電相之新發明」『東方雜誌』第 21 卷第 13 号, 1924 (民国 13) 年)。当時の中国ですでに、**tele-**で始まる電気を利用する発明物の名称の翻訳に「電」を使うことが常識化していたと判断することができる。

ただし、中国語において「電」で始まる多数の訳語候補の中から「電視」が最終的に選ばれた経緯は不明である。『申報』1929 (民国 18) 年 1 月 27 日の号に掲載された「電視之研究」という記事は、“**television** には「電望」「電視」「電送影片」などの訳語があるが、「電視」は「電信」「電話」とともに三つ巴を成す」と述べている。しかし、この説明は不十分で、「電視」の代わりに「電望」——あるいは「電景」など——を選んでも「電信」「電話」と釣り合う 2 字の訳語になる。⁸

「電視」という訳語の発生、選択について宮島は、**television** の直訳である「遠視」は「目の『遠視』と同音衝突をおこす」から避けられたとする考えを述べている。一見もっともらしい話であるが、妥当性は疑わしい。まず、別の語と同音であることが不都合だと言うのであれば、「電視」は日本語においては「電子」と同音衝突を生じる。そして、その 2 語はともに電気工学に関わる用語である。したがって、同音衝突ということ言えば、「電視」は「遠視」以上に都合の悪い訳語である。では、宮島の言う「同音衝突」を「同音同字の衝突」と読み替えればよいのか。すなわち、「電視」は「遠視」に生じる同音同字の衝突を避けるために選ばれた訳語だったと言うことはできるのか。筆者の見るところでは、そうした考慮が要素として働いた可能性はあり得るとしても、それも有効な説明にはならない。なぜならば、**television** を「遠視」と直訳すべきところそれでは医学用語との同音同字衝突が生じて不都合だから「電視」にしたとすれば、その限りでは一見説明できた形になるが、しかし今度は **telegraph** や **telephone** の訳語においても「遠」の代わりに「電」が選ばれた理由、すなわち、「遠」を使うことで生じる不都合を新たに探し出さなければならなくなってしまうからである。結局、「電視」という訳語の成立根拠は、**television** とその訳語だけを見比べて想像を働かせるという学問前の方法によって説明を与え得るものではないと思われる。

「電視」の語史に関して探求すべき課題はほかにもある。例えば、日中の古い「電視」

8 「電話」の語は日本で作られ、中国に伝わったことが確実ようである(荒川(2007)およびそこに引用された研究を参照)。筆者の確認の限りでは、「電話」の初出例は日本では『朝日新聞』1879 (明治 12) 年 7 月 9 日の号の記事、中国では『申報』1883 年 3 月 10 日(光緒 9 年 2 月 2 日)の号に掲載された日本の報道の翻訳記事中に見出される(それぞれ「電話器」「電話機」の形で現れる)。「電話」に先行して成立した「電信」については「電視」と同じく状況が不透明であるが、佐藤(2007)は「日本、中国それぞれ独自に造られたものと考えられる」としている。

の用例を観察して気付くことは、今の「テレビジョン」「テレビ」と異なり、しばしば行為——すなわち、“映像を遠隔地に伝送し、再現する”こと——を表していることである。例えば、「電視せんとする物体」「第二図の如き装置で電視した」「暗闇にある物体を電視し得る」「放送局で電視をやるとしたら」、また、「電視法」「電視方式」「電視の試験」などの表現が多く見られる。安藤博の記事における「電視」の日本初出例も「被電視体」という複合語の形で出現であった。中国語においても、先に挙げた用例にあった「電視器具」や「電視的儀器」のほかにも「電視術」「電視之法」「電視之方式」「電視之実現」などの表現が観察される。考えてみれば、テレビ放送事業が始まり、テレビ受信機が日常的な電気製品として普及する前の時代における television の概念、理解が今日のそれと異なっていたとしても不思議はない。television の日中名称の歴史は、単なる表現形式の選択の問題ではなく、その意味、内実にも関わる問題として考察する必要がある。

また、宮島も触れている関連の問題として、日本語の「テレヴィジョン」「テレビジョン」から「テレビ」という短縮語が作られて普及した過程も未詳である。次は「テレヴィ」「テレビー」「テレビ」という語の早期の用例である。ただし、いずれも見出しの類における使用で、本文では「テレヴィジョン」「テレビジョン」という表現が用いられている。

世界に誇るべかりし野球のテレヴェ放送【節見出し】

(荒川文吾『テレヴェジョンの基礎知識』, 1932 (昭和 7) 年 2 月 24 日)

テレビー愛好者への手引【記事副題】

(伊藤豊「装置の作り方」『東京朝日新聞』1933 (昭和 8) 年 2 月 24 日)

米国のテレビ【段落見出し】

テレビの原理【同上】

(「世界各国のテレビジョンを探る」『読売新聞』1934 (昭和 9) 年 4 月 23 日)

『無線と実験』第 10 巻第 4 号 (1928 (昭和 3) 年) 所載の座談会記録「テレビジョンは何と訳す可きか」では、出席者の 1 人である日本放送協会本部の伊藤豊が座談の冒頭で“訳すかどうかという問題から考える必要がある”との認識を示し、その後「之は訳語ではないが、テレビジョンといふのも長いから、略して『テレビ』としては如何。例へばテレビ送信機、テレビ受信機、将来はこんな使ひ方になるのではないかと思ふ。」と発言している。この予言がその後現実のものとなった。

中国の資料における television の音訳の出現も興味を引く。年表に見る通り、1931 (民国 20) 年の『科学』第 15 巻第 3 期の記事では「徳律維相」という名称が使われている。telephone の音訳「徳律風」は広く知られているが、「徳律維相」はそれに類する television の音訳である。ただし、「維相」の 2 字はおそらく発音に加えて、television の“情景をそのまま遠隔地に伝える”働きを示そうとする音義兼訳であろう。⁹ほかに「徳律維雄」「徳律維生」「徳律維興」という音訳もあった。

「電視」の語史をめぐるさまざまな問題の考察においても語の使用状況の十分な確認が

9 telephone を表す「徳律風」は通常音訳と説明されるが、羅 (1950) はそれを音義兼訳として扱っている。おそらく電話が“消息を伝える”ものであることを「風」で暗示しているという解釈であろう。

必要であり、それには各種の第3種 Web コーパスが大きな力を発揮する。

6. Google ブックスによる用例調査上の注意と工夫

Google ブックスで語句を検索したときに得られる情報の範囲は出版物によって異なり、主なものとして次の3段階がある。

- ①当該の出版物（書籍、雑誌）の全体を画像として閲覧、ダウンロードできるもの
- ②当該の語句を含むテキストの断片（スニペット）が3～4行程度表示されるもの
- ③著者名や書名などの書誌情報だけが表示されるもの

①に該当する場合に利用価値が最も大きいことは言うまでもない。③の場合は、当該の出版物中に問題の語句が実際に現れるという保証もなく、利用価値は小さい。

6.1 用例調査上の注意

事例が最も多い②の場合は注意と警戒を要する。表示される情報にはほぼ確実に誤りが含まれるからである。しかも、はなはだ危険なことに、一見問題のなさそうな場合も例外ではない。したがって、②に該当するすべての場合において当該の出版物現物による確認が不可欠である。

誤りにはさまざまな性質のものがあり、ここではその主なものについて述べる。

まず、書誌情報の誤りがある。表示された刊行年が実際のそれから数年、数十年ずれているという事例はきわめて多い。また、書名も信用することができない。ある出版物における語句の用例として表示されたものが実はまったく別の出版物の一部であることも珍しくない。Google ブックスは膨大な書籍の情報を集積しているので、それなりの頻度で書誌情報に混乱が生じているものと受け止める必要がある。表示されたテキストと書誌情報がまったく正常に見えても、そのまま信用することはできないのである。

歴史的な資料の復刻版については、Google ブックスでは通常その資料の当初の刊行年が表示される。そして、復刻資料にはしばしば解説文が添えられることから、用例年代の混乱が生じることになる。すなわち、ある語句の用例を明治期の資料中に見出せたと思っても、それが実は後年復刻時に書かれた解説文に含まれるものに過ぎないといったことがよくある。

テキスト自体に多くの誤りがあることは Google ブックスを使ってみれば直ちに気付くことである。古い書籍は一般に文字認識の精度が低い。新しい書籍でも字形の似た文字を中心に誤認識は多い。出版物での「電視」が「雷視」として解釈されているなどの誤りである。

容易に気付く文字の誤認識に加えて、レイアウトの誤認識もある。特に、縦書きの書籍において行が左から右に進行するように認識されていることが多い。それによって、実際には存在しない用例が見つかることになる。例えば、縦書きの書籍で、ある行が「上げる」で始まり、次の行が「立ち」で終わる場合、実際には「上げる」と「立ち」のあいだに関係はないわけであるが、行の進行が逆方向に認識されるとそこに「立ち上げる」という語の幻の用例が出現することになる。

また、文字認識に関わる別の問題として、語句の出現箇所を示すページ番号の表示にも誤りが多い。そのような場合、特定の出版物に用例が含まれることが分かっていても、その出現箇所が直ちには分からないことになる。大部の書籍や製本雑誌に目を通して語句の用例を探し出す作業は多大な労力と時間を要する。

6.2 用例調査上の工夫

語句を検索して得られる用例が少ないときや、既知の初出例よりも古い用例を見出したときには、多様なキーワードで検索してみることが重要である。ここで述べる内容は Google ブックス以外での検索にも共通するが、収録データが多様な表記法を含み、誤りも多い Google ブックスでは特に配慮を要する。

例えば、Google ブックスの検索で漢字の字体は区別されるので、複数の字体——新字体、旧字体、繁体字、簡体字など——を持つ漢字を含む語句の用例を多く得たければ可能な表記をすべて検索する必要がある。

漢字と仮名の選択に関わる異表記もあるので、「立ち上げ」に加えて「立ちあげ」「たち上げ」「たちあげ」などの表記も検索する。また、仮名の用法に関わる各種の異表記も多い。まず、送り仮名に関しては、「立上げ」も検索する。ただし、注 4 でも触れた通り、「立上げ」の検索によって得られる用例は「立ち上げ」でなく「立て上げ」の異表記である可能性があることに注意が必要である。また、古い資料で旧仮名遣いが使われていることは言うまでもないが、ほかの面でも現代の慣用と異なる表記は多いので、「立ち上ゲ」「立上ゲ」「立ち上ケ」「立上ケ」などの表記も検索する。

さらに言えば、文字の誤認識を想定して、誤ったキーワードで検索してみるのも有効な方法である。すなわち、「雷視」のような予想し得る誤認識の結果をキーワードに指定して検索するということである。

実際に使ってみれば分かる通り、Google ブックスによる語句の検索結果は精度が低く、表示される検索結果の大半が不要のものであるということも多い。文字認識の誤りにより、あるはずの用例が見つからないという場合もある。また、そもそも Google ブックスにあらゆる出版物が収められているわけではないので、その意味でも不完全な調査しかできない。しかし、それでも Google ブックスを使わなければ得られない情報を豊富に入手できることの意義はきわめて大きい。検索結果から無用のものをふるい分け、有用である可能性のあるものを書籍現物によって確認するという労力を費やせば、Google ブックスは近現代語の研究のための強力な情報源となる。

7. おわりに

大量性と歴史性という 2 つの性質を兼ね備えた第 3 種 Web コーパスは近現代語の研究に絶大な支援を提供する。Web コーパスの利用により、研究者の内省、紙媒体の資料、小規模なコーパスに頼った従来の研究法では考えられないほどの豊富なデータに基づく考察が可能になる。

ただし、Web コーパスのみに頼って近現代語を研究できるということではない。Web コーパスから得られる情報も限定的であり、網羅的と言うにはほど遠い。Web コーパスに期待できるのは用例収集の補助手段としての役割であり、従来の方法による用例収集が

不要になるわけではない。5.2 に挙げた中国の雑誌『科学』に載った television に関する最初の記事（1917 年）も「電視」の初出例（1925 年）も紙媒体の同誌の調査によって見出したものであった。前者は「電視」の語を含まず、また、後者は記事本文に「電視」を含むだけで、いずれも機械的な検索では得られなかった。同誌の 1927 年の号に載った「電視之進歩」という表題の記事が検索によって得られ、それを手がかりにより古い号にさかのぼって調査して見出したということであった。

コンピュータやインターネットの発展は言語研究にも革新をもたらし続ける。ここで述べた研究法は十年前には予想も付かなかったことである。そのことを思えば、十年後には今は予想できないさらに新しい研究法が可能になっていることであろう。

文献

- 荒川清秀（2007）『『電』のつくことば—『電話』を中心に—』内田慶市・沈国威（編）『19 世紀中国語の諸相—周縁資料（欧米・日本・琉球・朝鮮）からのアプローチ—』263-282. 雄松堂出版。
- 佐藤亨（2007）『現代に生きる幕末・明治初期漢語辞典』明治書院。
- 田野村忠温（2012）「BCCWJ に収められた新種の言語資料の特性について—データ重複の諸相とコーパス使用上の注意点—」『待兼山論叢』46（文化動態論篇）：59-82. 大阪大学大学院文学研究科。
- 田野村忠温（2014）「BCCWJ の資料的特性—コーパス理解の重要性—」田野村忠温（編）『講座日本語コーパス 6 コーパスと日本語学』119-151. 朝倉書店。
- 田野村忠温（2015a）「意識地名『牛津』『剣橋』の発生と消長」『大阪大学大学院文学研究科紀要』55：81-137。
- 田野村忠温（2015b）「中国語の外来名表記に関する覚書—地名 Amsterdam の意識をめぐる議論の検討ほか—」『待兼山論叢』49（文化動態論篇）：9-34. 大阪大学大学院文学研究科。
- 田野村忠温（2016a）「『科学』の語史—漸次的・段階的変貌と普及の様相—」『大阪大学大学院文学研究科紀要』56：123-181。
- 田野村忠温（2016b）「真珠湾の日中名称小史」『待兼山論叢』50（文化動態論篇）。大阪大学大学院文学研究科。刊行予定。
- 宮島達夫（2008）「『テレビ』と『電視』—『電視』は和製漢語か—」沈国威（編）『漢字文化圏諸言語の近代語彙の形成—創出と共有—』95-110. 関西大学出版部。
- 羅常培（1950）『語言与文化』国立北京大学。

（2016 年 3 月 13 日受付）

Invited Paper (A) to the Special Issue

The Concept, Types and Utility of Web Corpora: Web Corpora as a Source of Information for Etymological Studies

TANOMURA Tadaharu (Osaka University)

Abstract:

The defining condition of a Web corpus will be that it is a huge amount of text data collected from the Internet. Although Websites such as *Google Books*, *National Diet Library Digital Collections* and newspaper archives do not satisfy the condition, they nevertheless cannot be clearly distinguished from typical Web corpora, and thus it may not be groundless to regard them as a type of Web corpus. This article, drawing upon two case studies, will demonstrate that we can easily enhance the level of the description of the history of Japanese as well as Chinese terms of the modern era with the help of information obtainable from those Websites.

Keywords: Web corpus, diachronic corpus, modern Japanese and Chinese, etymology, *tatiageru* (transitivized form of the verb *tatiagaru*), *densi/dianshi* (Japanese/Chinese term for television)