

漢字文獻 情報處理研究

第 18 号

漢字文獻情報處理研究会 編

好文出版

漢字文献情報処理研究 第18号

目次

論文	4	「権威」の要らない言語学習の可能性 ICTと学習者オートノミー	山崎 直樹
	18	『皇明條法事類纂』電子化割記 文献学・資料学とデジタルアーカイブのはざま	小島 浩之
	29	東京大学デジタルアーカイブズ構築事業 学術資産の利活用に向けた取り組み	中村 覚
	35	【研究ノート】韓国における仏教学とデジタル・ヒューマニティーズ “Digital Humanities and Buddhism – Focusing on Data Mining and Visualization” 参加記	師 茂樹
特集1	入門・中国語形態素解析	41	
	42	形態素解析とは何か	師 茂樹
	46	KH Coder で中国語分析	田邊 鉄
	51	形態素解析の中国語教育への応用	氷野 善寛
	56	古典中国語（漢文）の依存文法解析と直接構成素解析	安岡 孝一
	63	中国語形態素解析エンジン概観	千田 大介
	66	Windows で結巴、THLAC を使う	千田 大介
	72	macOS で HanLP を使う	師 茂樹
	76	コラム：Word で単語区切り	千田 大介
特集2	デジタル時代の中国学リファレンス②	77	
	基礎知識		
	78	解説：CNKI 工具書庫	
	1. 官職をしらべる		
	80	職掌をしらべる	91 解説：官僚登用制度
	82	官品をしらべる	93 清代の官職の官品や職掌をしらべる
	84	官職の別称をしらべる	95 清代の人物の官職をしらべる
	86	官制を視覚的に把握する	97 中華民国時代の人物の官職をしらべる
	88	解説：官制の構造	
	2. 新聞資料をしらべる		
	99	「満洲」の新聞をしらべる	
	3. 仏典をしらべる		
	101	仏典の叢書・テキスト	111 解説：仏典の分類
	106	仏典を検索する	

4. 道教・民間信仰をしらべる			
114	道教経典をしらべる	121	民間信仰の神をしらべる
117	解説：『道蔵』とは	123	解説：民間信仰と民間宗教の資料
5. 古典通俗文学をしらべる			
126	古典通俗小説の概要をしらべる	134	古典戯曲作品の概要をしらべる
128	古典通俗小説の排印本	136	古典戯曲作品の影印本・排印本をさがす
130	古典通俗小説の影印本をさがす	138	古典戯曲作家の情報をしらべる
131	解説：官刻本・家刻本・坊刻本	140	白話語彙の意味をしらべる
132	古典通俗小説の全文検索		
6. 現代中国語をしらべる			
143	中日辞典に載っていない中国語の意味	147	欧米人名の中国語訳
146	日中辞典に載っていない日本語の中国語訳		
特集 3	2018 年夏期中国視察報告	148	佐藤 仁史・千田 大介・師 茂樹
	149	北京創新力博社 北京本社	
	158	北京創新力博社 洛陽入力センター	
	162	CNKI 北京本社	
特集 4	自炊・モバイル時代の中国語 OCR 徹底比較	169	
	170	いまなぜ OCR なのか	千田 大介
	172	PC ソフト 4 種の OCR 機能	千田 大介
	177	PC ソフト OCR 結果比較	千田 大介
	187	Office Lens と Adobe Scan	千田 大介
	191	中国語 OCR スマホアプリ比較	田邊 鉄
レビュー & リソース紹介	195	IIIIF の解説と動向 —— 画像データの国際的な相互利用促進を目指す枠組み / 万方数拠 / Android スマホ用中国語辞書 西岡 千文 / 小島 浩之 / 田邊 鉄	
	206	執筆者紹介	
	208	編集後記	

- ◇ 本誌記事中のソフトウェア名・プログラム名・会社名などは一般に各社の商標または登録商標です。本文中では、™・®等のマークは明記しておりません。
- ◇ 本誌記事の記述に基づいて行われた作業の結果生じたあらゆる損害について、編著者・翻訳者および出版社は一切の責任を負いません。
- ◇ 本誌記事の内容に関するお問い合わせには応じかねます。

「権威」の要らない言語学習の可能性

ICT と学習者オートノミー

山崎 直樹 (やまざき なおき)

回 1. この文章について

◇ 1.1 目的

この文章の目的は、青木 (2018)¹⁾が、学習者オートノミー *autonomy* と ICT を使った言語学習の関係に触れる中で提起した「何が「正しい」情報であるかを判断する権威は誰が持つべきなのか」という問いかけと、山崎 (2015) の提起した「その翻訳はまちがっているかもしれないということを疑っていないかのような態度で、課題に関する解答を提出する」問題の 2 つを出発点に、現在の ICT 環境を利用した学習環境、とくに Web2.0 という呼称で表現されるような ICT 環境が「権威を必要としない言語教育」の道を拓いてくれるのではないか、という可能性を考えることである。

この問題に関心のないかたは、§6 と §7 だけを見ていただくのもよいと思う。2018 年現在のウェブ上の機械翻訳の性能の一端を知ることができるはずである。

◇ 1.2 この文章の概要

まず、§2 で、学習者オートノミーについて概略を説明する。この概念をご存知のかたは、ここをとばしていただいてさしつかえない。

次に、§3 で学習者オートノミーと現在の ICT

環境とがどう関係するのかについて述べる。

§4 では、言語学習における「権威」とは何を指すかについて定義をする。

§5 では、山崎 (2015) が提出した問題 (学習者が自分の産出した言語形式に対して妥当性を検証しようとしな) について簡単な紹介をする。この文章の内容をご存知のかたは、ここをとばしていただいてさしつかえない。

§6 では、ICT による支援の一例として、現在の無料機械翻訳サービスの日→中翻訳の精度を、金子・山崎 (2013) の結果と対比する形で紹介する。

§7 では、現在の機械翻訳の日→中翻訳の出力は、どれほどの *post edit* を必要とするかという実例を見る。

§8 と §9 では、Web2.0 の利点を援用した言語学習では、「権威」の必要性は低くなっているという主張を述べる。

回 2. 学習者オートノミーについて

◇ 2.1 定義

学習者オートノミー (*autonomy*, 自律, 自律性) とは何か。

日本語教育学会 (2005) の「自律学習」の項では、「学習者が自分で自分の学習の理由あるいは目的と内容、方法に関して選択を行い、その選択に基

づいた計画を実行し、結果を評価できる能力である」と規定している（なお、この項の執筆は、本稿でも言及している青木直子氏である）。

※ 2.2 オートノミーが必要な理由

学習者オートノミーは、外国語学習を含む言語の学習にとっても必要であるとの認識が高まりつつある。なぜ、学習者オートノミーが必要なのかを、やや長くなるが、日本語教育学会(2005)の「自律学習」の項から引用して示したい。

第二言語教育は学習者オートノミーを育てることを目標にすべきであると主張する人々の挙げる理由は主に以下の7つである。

- 1) 学習者オートノミーを育てることは、社会の意思決定に責任をもって参加できる市民を育てることにつながる。
- 2) 学習とは学習者が新しい知識を自らの既有知識と結びつけることによって起こるものであるから、学習に成功するためには学習者オートノミーが必要である。
- 3) 第二言語習得は教えられて起こるものではなく、学習者独自のシラバスに基づいているものだから、学習者が自分で学習をコントロールすれば習得の効率上がるはずである。
- 4) 第二言語使用者として成功するためには、自分のことは自分ですること、自信をもつことが欠かせない。
- 5) 自分で選択することが内発的動機づけにつながる。
- 6) 社会の急速な変化に対応するためには生涯にわたって学びつづける必要がある。
- 7) 自分の学習に関する選択を自分で行うことは学習者の権利である。

※ 2.3 何をコントロールするのか

§ 2.2 の 3) で言及された「学習者が自分の学習をコントロールする」とは、何をコントロールす

るのか。Holec (1985:180-182) によれば、それは以下のような事項である（青木 (2018) に日本語による適切な要約があるので、それを以下に引用して示す）。

- a) 目標（明日／来週／来月／来年までに、何をできるようにするか）
- b) 内容（目標達成のために、何を学ぶか）
- c) リソース（どんなリソースを使って学ぶか）
- d) 方法（リソースをどのように使うか）
- e) 評価方法（できるようになったかどうかはどのようにして確認するか）
- f) 順番（何から学び始めるか、次に何をするか、あとでやることにするのは何か）
- g) 場所（どこで学習するか）
- h) 時間（一週間／1日のうち、いつ学習するか、どのくらいの時間学習するか）
- i) 頻度（毎日やるか、週に何回やるか）

☐ 3. ICT とオートノミー

§ 2.3 の「何をコントロールするか」を見ると、現在の ICT 環境は、学習者オートノミーの実現にとって、とてもよい環境にあるように思われる（言語学習に役立つ ICT の具体的な例については、Reinders & White (2011) を参照）。とくに、Web2.0 と呼ばれる「ユーザーの参加・関与」を特徴の1つとする最新の Web 環境は、学習者オートノミーと親和性が高い（*Learner autonomy and Web 2.0*. というタイトルの書籍があるくらいである（Cappellini, Lewis & Rivens Mompean (2017)））。

学習者がインターネット上のリソースをどのように利用して言語学習を進めているかについては、欧麗賢 (2016) に具体的な記述がある（この例は、中国人日本語学習者の例であるが、日本における外国語学習者も利用できるリソースの種類については大差ないと思われる）。

この ICT と学習者オートノミーの親和性との関連において、青木 (2018) は、「これから eラーニングのプログラムを作る時に考えたほうがいいと思われること」と題し、次の問題提起を行っている。

- 教材を作る必要はもうないのではないか
- どのようなリソースがどこにあり、それはどんな目的でどのように使えるかという情報が検索できるようにすることのほうが重要ではないか
- コントロールの分散化をどのように実現するか
- 学習者をサポートするシステムをどう作るか (e.g. アドバイジング)
- 学習者同士が会える場をどのように作るか (バーチャルな場, 物理的な場)
- 市販のアプリや言語学習サイトなど、すでに巨大なポータルが存在する中で、個別の教育機関に利用者を限定したポータルはいらないのではないか
- 何が「正しい」情報であるかを判断する権威は誰が持つべきなのか (太ゴシックによる強調は筆者による)

すべて、極めて妥当な問いかけであるように思われる。本稿では、以下で、とくに最後の「権威」について考えてみたい。

回 4. 言語学習における「権威」

◆ 4.1 何についての権威か

§3 で言及した青木 (2018) の「権威」については、青木自身の詳細な説明はなかったので、「正しい情報」というのが「何についての正しい情報」を指しているのか明らかではないが、本稿では「言語についての正しさ」に的を絞って考えたい。学習者にとっても教授者にとっても「言語の正しさ」は言語学習／言語教育の大きな関心事であるからである¹⁴。

「言語の正しさを判断する権威」について考えるとき、我々は、次の2つの「権威」を別個に考えねばならない。

A) 学習対象としての○○語の「規範性」につ

いての権威

B) 学習者が個別の学習状況において産出した個々の言語形式の「規範性」に対する権威

◆ 4.2 学習対象としての○○語の「規範性」

学校を典型例とする教育機関の教室において (= formal instruction の場において)、権威を持つ1人の教師が、言語 (とくに外国語) を教える場合、一般的にあって、「唯一の正しい標準語」を中心に教えることが求められる。

この理由は、実際にそのような変種 (唯一の正しい標準語) が存在するかどうかは別として、そのような言語の存在を仮定したほうが、教師にとっても、学習者にとっても「てま要らず」であるということであろう。

しかし、これは、学習者の得られるリソースが、教育機関によって選択された教材の中の言語と、教育機関の教師が産出する言語に限られる環境、つまり (formal instruction に限定される) 特殊な外国語学習環境の場合にのみ成立する話である。

現在のように、ウェブ上のあらゆるリソースが言語学習の資源となり、言語学習に特化した SNS までもが存在する環境では、上述した「唯一の正しい標準語」のみを選択的にリソースとすることは不可能である。

たとえば、中国語、英語、フランス語、スペイン語、ドイツ語……のような「複中心地言語 pluricentric language」 (規範を複数もつ言語、規範を定める〈首都〉が複数ある言語) では、ウェブ上のリソースに対し、「唯一の規範」を選択することは困難である (し、また、現実世界の言語使用を考えれば、その必要もない)。

◆ 4.3 学習者が産出した形式の「規範性」

伝統的な外国語学習の場では、とくに formal instruction の場では、学習者が産出した学習目標の言語のさまざまな言語形式に対し、教師が「規範」に照らし合わせ、その形式としての正誤、運用の適不適を判断してきた (このシステムが学習者と教師の間にどのような齟齬を産んだかについては、山崎 (2015) でやや詳しく述べた)。

しかし、このような学習モデルは、formal instruction の外では、もはや成立しない。SNS 等で自由に交流する中で産出された学習者の言語形式は、たとえ、規範に照らして構造的な誤りがあったとしても、あるいは運用上の不適切さがあっても、スルーされる可能性が大いにある。スルーされないまでも、それらを指摘するのは、その言語の規範を代表しない話者の語感である場合もあれば（「規範を代表する人の語感」というのは実は幻想であるのだけれど）、母語話者でないが熟達した学習者の場合もあるし、時として、熟達していない学習者の誤った訂正である場合もある。そして、学習者は、これらの訂正が自分にとって必要なものか判断できないこともあり、また、学習者はその訂正を受け入れるかもしれないし、受け入れないかもしれない（相手は成績を判定する教師ではないので、受け入れなくても、少なくとも短期的には、問題は起らない）。あるいは、誤りのサジェスションに当該の学習者が気づかないこともありうる。

では、訂正は無意味かということ、もちろんそのようなことはなく、学習者は、目標言語の使用者のソーシャル・ネットワークで、あるいは、言語学習を目的としたソーシャル・ネットワークで、さまざまな（そして教室のそれとは圧倒的に量が異なる）リソースとフィードバックに接しながら、意識的に、あるいは、無意識に、自らの内部に構築されたつつある学習目標言語の体系を修正していくはずである。少なくとも、現在のような ICT 環境が提供する言語使用と学習の場では、そのようなモデルで言語習得を考えたほうが実情に近い。

このような状況下では、「誰か 1 人の権威を持った教師が自分の誤りを必ず訂正してくれ、学習者はそれを正しいと受け入れる」ような学習は期待できないし、その必要もない。また、それが効率がよいと考える理由もない。

◆ 4.4 「問いかけ」への回答

ここでは、青木 (2018) の「何が「正しい」情報であるかを判断する権威は誰が持つべきなのか」への一応の回答を、筆者なりに考えてみたい。

上述 (§ 4.1) の A) 「学習対象としての〇〇語の

「規範性」についての権威」について言えば、「そのような権威は誰も持つ必要はない」というのが、筆者の考えかたである。

上述 (§ 4.1) の B) 「学習者が個別の学習状況において産出した個々の言語形式の「規範性」に対する権威」について言えば、「学習者は自らの産出した形式の構造上の誤りや運用上の適切性に対して、常に意識を高めておく必要がある。しかし、自分以外の誰か権威を持つ人に常にその判定を委ねる態度を採るべきではない」というのが、筆者の考えかたである。「意識を高めておく」というのは、その言語を使用するコミュニティの他の成員とのやりとりに常に注意を払い、そこから受け取るフィードバックによって、自分の学習目標言語のシステムを発達させることを自ら心がける、という意味である。

回 5. 出力の妥当性の検証

◆ 5.1 教師と学生の認識のズレ

山崎 (2015) は、外国語を学ぶ授業における学生のある種の課題解決行動（例：翻訳や作文の課題に対し、機械翻訳サービス等を利用して解答を作成するが、その妥当性の検証を行わない）の理由として、教師の願望と学習者の反応の間の認識のズレを指摘した (p.14)。以下のとおりである。

(A) 教師の願望

1) 過程を重視してほしい

- 与えられたインプットがどんな意味か解釈してほしい。
- 文がどんな構造をもっているか分析してほしい。
- 手持ちの材料でアウトプットに必要な構造を再構築してほしい。
- さまざまなストラテジーを駆使して、談話を構成する訓練（ブリコラージュ Bricolage 型の解決行動）をしてほしい。

2) 出力を検証してほしい

- 自分が提出したアウトプットが妥当か

どうか、考えてほしい。

(B) 学習者の反応

1) 教師が重視するのは成果を提出すること

- 評価されるのが成果であるのなら、求められているのは成果を出すことである。
- かけたコストが評価されないのなら、コストはかけない。
- 早く、低コストで、教師の設定したゴールに到達することが、賢い行動である。

2) 結果の適不適を検証するのは教師の仕事

- たとえ妥当な結果でなくても、早く成果物を提出して、教師の判断を仰ぐのがよい。

外国語を教えた経験のあるかたは、自らの学習設計の中に、「学習者が自分の出力の妥当性を検証する学習過程」をシステムティックに組み込んでいたか、それに必要なツール類の存在と利用法をサジェストしたか、そして何よりも、そのような作業の意義を理解させるような学習を試みたかを自省してみるとよいと思う。

このような学習過程がなければ、学習者が「(プロセスのいかに考慮せず) 解答を準備し、教師にそれを渡して、妥当性の検証は教師に任せる」が学習者の役割だと理解しても無理はない。それは「学生は答案に解答を書く。教師はそれに○×をつける」という分業体制と同じである。

◆ 5.2 オートノミーの一部として

これまでに述べてきたことからわかるように、「学習者は自らの産出した形式の構造上の誤りや運用上の適切性に対して、常に意識を高めておく必要がある」「学習者が自分の出力の妥当性を検証する過程を学習に組み込むべきである」は、学習者オートノミーの育成に必要であると考えてよい。

ひとつ注意していただきたいのは、次の2つは別のものであるということである。

- 1) 学習者がいろいろなツールを利用して妥当な出力を自力で生成したいと努力する。
- 2) 学習者が自分の出力が妥当かどうかで自分で検証しようという意識を持つ。

われわれが知る多くの日本の外国語学習者には2)が欠けていること、そしてその理由を、山崎(2015)で指摘した。

以下では、2)を考える手がかりとして、まず1)について考えてみたい。学習者が「いろいろなツールを利用して妥当な出力を自力で生成したい」と考えたとき(ここでの「自力で」の意味は「権威ある教師の力を借りずに」という意味である)、現在のICTはどのような支援をしてくれるのかを、日本語→中国語の機械翻訳を例にして考えてみたい。

□ 6. 機械翻訳はどこまで支援してくれるか

◆ 6.1 近5年の進歩

金子・山崎(2013)は、中国語能力検定試験の準4級～3級の試験問題を使って、ウェブ上の無料翻訳サービスと有料の翻訳ソフトの性能(日→中、中→日の両方向の翻訳)を試した。その際の一応の結論は、「機械翻訳は中検3級に合格しない」であった。

この検証は2013年夏に行われたが、5年後の2018年夏の時点では、翻訳の精度はどのように上がっているのだろうか。果たして実用に耐えるのであろうか。

なお、今回は、各種翻訳エンジンの性能を確認すべく、かなり網羅的に各種サイトのチェックを行ったが、今回は、Google, Bing, 百度の大手3サイトに限って、日→中の方向の翻訳のみをチェックした。

以下は、上述の3サイトの、同一入力による出力結果を、前回と今回で比較した表である。

	2013	2018
Google	我不想去任何地方了,今晚。	我今晚不想去任何地方了。
Bing	今晚再别的地方不想去。	今晚我不想再去任何地方了。
百度	今晚哪儿也不想去。	今晚我再也不想去了。

表1 「今夜はもうどこにも行きたくない」

	2013	2018
Google	因为不知道韩国人你会得到一个她的翻译。	我不懂韩语,所以请她翻译。
Bing	韩国语言不知道,所以有她的翻译。	我不懂韩国语,所以我请她翻译一下。
百度	韩语是因为不明白,请她翻译。	因为不懂韩语,所以请她翻译。

表6 「韓国語はわからないので、彼女に翻訳してもらおう」

	2013	2018
Google	所有刚来到日本时,他可能不会说日语。	我来日本时,他根本不会说日语。
Bing	我刚到日本时,他不能在所有日本语言。	当他来到日本时,他根本不能日语。
百度	他刚来日本的时候完全会日语了。	他刚到日本的时候,完全不会日语。

表2 「彼は日本に来たばかりの頃まったく日本語ができなかった」

	2013	2018
Google	以及驾驶汽车,她还可以修复。	她不仅可以修理汽车的驾驶。
Bing	她不只开一辆车,可以修理。	她不但可以开车,还能修理汽车。
百度	她不仅是开车,也会修理。	她不仅会开车,还可以修理

表7 「彼女は車の運転だけでなく、修理もできる」

	2013	2018
Google	他感兴趣的東西。	他对一切感兴趣。
Bing	他有什么感兴趣。	他对一切都感兴趣。
百度	他对什么感兴趣。	他对什么都很感兴趣。

表3 「彼は何にでも興味を持つ」

	2013	2018
Google	我已经写了很多次,但我不写正确通过各种手段。	我写了很多次,但我写不出来。
Bing	我写了几次,但通过各种手段写。	我已经写了很多遍了,但是我写得不好。
百度	我曾多次写了一下,无论如何都写不好。	我写了几遍,怎么也写不好。

表8 「私は何回も書いてみたが、どうしてもちゃんと書けない」

	2013	2018
Google	它需要3个小时,飞机从这里到东京。	从这里到东京需要三个小时的飞机。
Bing	花3小时从这里到东京乘飞机。	从这里飞到东京需要三小时。
百度	从这里到东京飞机需要三小时。	从这里到东京要花3个小时的飞机。

表4 「ここから東京まで飛行機で3時間かかる」

	2013	2018
Google	周老师太多时间为两年来日本。	我的老师又来到日本两年了。
Bing	再过2年也在老师的周围来到日本。	周先生在日本已经两年多了。
百度	周老师来日本已经2年多。	周老师来日本已经两年多了。

表5 「周先生は日本に来てもう2年あまりになる」

※ 6.2 文レベルでの実用性

どの翻訳サービスも、どの課題においても翻訳精度が確実に上昇していることが見てとれる。とくに**百度**は安定して、よい結果を出しており、事実関係の情報の伝達だけであれば、単文単位では**实用レベル**といってもよいかもしいない(中検は、「情報の伝達可能性」を評価基準にはしていないと考えられるので、合格できるかどうかはわからない)。

※ 6.3 学習者が自己チェックをすることは可能か

上でも述べたとおり、1) 学習者が妥当な出力

を自力で生成したいと努力することと、2) 学習者が出力の妥当性を自分で検証しようという意識を持つこととは別である。

では、学習者が2)の意識を持ったとして、どのように検証すればよいのか。考えられるのは、次のa～cの3パターンである。

- a. 自分の出力を教師に見せる。
- b. 自分の出力を他の学習者に見せる。
- c. ツールを使って自分で検証する。

このような検証を行うためには、教科書にせよ、ウェブ上の教材にせよ、静的なりソースの中を探しても解答が見つかるものでもないの、インタラクティブなりソースを使うしかない。上記のa～cはすべてインタラクティブなりソースである。

aは伝統的な学習方法であるが、教師とコンタクトを取りこのような検証を頼むというのは、学校のような教育機関では、まだハードルが高い。eメール以外でも便利で手軽なコミュニケーションツールが普及しているものの、24時間、何でも気軽に尋ねられるかどうかについては、教師との人間関係にもよるが、まだまだ心理的抵抗が大きいと思われる。

bは、現実社会の学習者仲間に依頼するほか、Lang-8^[3]などの学習者用SNSで質問するのもこれにあたる。

cについては、いろいろな可能性が考えられるが、(機械翻訳を使った)逆翻訳 back translationによる正確さの検証などというのも、よく知られた方法であろう。次の節では、機械翻訳を使った逆翻訳の結果を見てみる。

◆ 6.4 機械翻訳による逆翻訳

6.4.1 逆翻訳の実験

興味深いことに、同一の翻訳サービスの中では、日→中の翻訳の出力結果が不自然な中国語であっても、それを入力にして中→日の翻訳をすると、自然な日本語の出力が得られることが、ときとしてある。

以下に示す資料は、3種類の機械翻訳サービス

を使い、ある1つの機械翻訳による日→中の出力([o]とする)、[o]を入力にした同一のサービス内の中→日の逆翻訳、[o]を入力にした他の2つのサービスの中→日の逆翻訳を並べて示したものである。出発点となる日本語は、「口語的なやや複雑な文構造を持つ日本語」と「文構造を簡略化した日本語」の2種類である。

6.4.2 やや複雑な日本語

入力 (日本語)	程さんはカラオケの応援で盛り上がり過ぎて、ソファから転げ落ちていました。	
Google の出力	先生先生在卡拉OK的支持下太兴奋了, 从沙发上掉了下来。	
逆翻訳	Google	ミスターさんはカラオケのサポートであまりにも興奮し、ソファから落ちました。
	Bing	先生, カラオケのサポートで、彼はあまりにもソファから落ちて興奮していた。
	百度	先生はカラオケでの応援に興奮して、ソファから落ちてしまいました。

表9 Googleの出力を他のサービスで

入力 (日本語)	程さんはカラオケの応援で盛り上がり過ぎて、ソファから転げ落ちていました。	
Bing の出力	摔倒先生太兴奋了, 不能为卡拉ok欢呼, 所以他离开了沙发。	
逆翻訳	Bing	秋さんはカラオケを応援できないほど興奮していたので、ソファを出た。
	Google	倒れた男はカラオケのために応援するほど興奮していたので、彼はソファを出ました。
	百度	転ぶさんは興奮して、カラオケでは歓呼するわけにはいかないので、ソファを離れた。

表10 Bingの出力を他のサービスで

入力（日本語）	程さんはカラオケの応援で盛り上がり過ぎて、ソファから転げ落ちていました。	
百度の出力	程先生因为卡拉 OK 的支持而气氛高涨, 从沙发上摔下来了。	
逆翻訳	百度	程さんはカラオケの応援でテンションが上がって、ソファから落ちました。
	Google	チェン氏の雰囲気はカラオケのサポートによって飛躍し、ソファから落ちた。
	Bing	チェンさんはカラオケのサポートのために上機嫌で、ソファから落ちた。

表 11 百度の出力を他のサービスで

6.4.3 簡略化した日本語

入力（日本語）	程さんはカラオケの応援をしてとても興奮しました。それでソファから落ちました。	
Google の出力	他非常高兴能够支持卡拉 OK。然后它从沙发上掉了下来。	
逆翻訳	Google	彼はカラオケをサポートしてとても幸せです。その後、それはソファから落ちた。
	Bing	彼はカラオケをサポートすることができて非常に喜んでいました。それからソファから落ちた
	百度	彼はカラオケを応援することができて嬉しいです。そしてそれはソファから落ちてきた。

表 12 Google の出力を他のサービスで

入力（日本語）	程さんはカラオケの応援をしてとても興奮しました。それでソファから落ちました。	
Bing の出力	我很高兴能为卡拉 ok 喝彩。所以我从沙发上摔下来了	
逆翻訳	Bing	カラオケで応援してくれて嬉しいです。だからソファから落ちた
	Google	私はカラオケのために応援することが非常にうれしいです。だから私はソファから落ちた。
	百度	カラオケに喝采することができて嬉しいです。だからソファから落ちてきました。

表 13 Bing の出力を他のサービスで

入力（日本語）	程さんはカラオケの応援をしてとても興奮しました。それでソファから落ちました。	
百度の出力	程先生为卡拉 OK 的应援非常兴奋。所以从沙发上掉下来了。	
逆翻訳	百度	チェンさんはカラオケのリクエストにとっても興奮していました。だから私はソファから落ちた。
	Google	チェン氏はカラオケのサポートに非常に興奮しています。それはソファから落ちた。
	Bing	チェン氏はカラオケのサポートに非常に興奮しています。それはソファから落ちた。

表 14 百度の出力を他のサービスで

6.4.4 結果の考察

これだけのデータからは断定的なことは言えないが、少なくとも同一サービス内でおかしな逆翻訳が得られるようでは、その入力となった翻訳に問題がある可能性が高いといったところであろうか。

無料の機械翻訳サービスによる逆翻訳も、学習者を支援するツールとして使えるような感じではある。

7. Post edit の必要性という観点から

7.1 Post edit

では、現実の社会で直面しそうな状況を考えた場合、機械翻訳の出力はどの程度、使えるのか。以下では、人手による post edit（翻訳後の修正）がどれほど必要かという観点で、機械翻訳の出力をチェックする実験を行った。

使用した翻訳サービスは**百度**である。これは §6 での比較の結果、日→中の翻訳では、**Google** や **Bing** に比べ、安定して質の高い翻訳を提供しているという感触を得たためである。なお、今回提示するデータは 2018 年 7 月～8 月に得られたものである。

◆ 7.2 実験の目的

この実験の目的は何かを実証することではなく、現実の社会でありそうな状況の中に置いたとき、機械翻訳の出力に対し、熟達した中国語のユーザーはどの程度の許容をするかという感触を掴むことである。よって、被験者の数は2人にとどめた。

◆ 7.3 実験の概要

使用した課題は全部で4種類である。post editの必要性は、現実の具体的な状況によって要不要の程度が異なるため（誰に対して何の用途に使うのか、時間的な余裕はどれくらいあるのか……）、詳細に場面を想定した。

post editを依頼した被験者は2人で、1人は、高校卒業まで中国で過ごした中国語のネイティブ話者の30代女性（以下ではChと略す）、もう1人は大学入学後に中国語の学習を始めた日本語話者の20代女性、中国での1年以上の留学経験があり、HSKの6級にも合格している上級学習者（といてよいと思う。以下ではJPと略す）。

それぞれ詳細な場面設定をし、最小限のpost editを考えるような要求（以下、PE.）と、もし全面的に書き換えるなら／自分でゼロから書くのなら（以下、FW.）どうするかという2種類の作業をもらった。

以下でその結果を見る。PE.の太ゴシックで示した部分は修正を施した箇所である。詳細な場面設定の説明は、最小限の修正を引き出すためのものである。

◆ 7.4 実験の結果

7.4.1 課題1 待遇度（相手の利益）

あなたは、あるホテルのフロントスタッフです。ある日、マネージャーが「夏の間はロビーで冷たいお茶のサービスをしていたんだけど、温かいコーヒーを欲しがいるお客様もいるから、別に用意したんだ。でも、ポットに入っていると見てもわからないのよね。だからこんな張り紙をしておこうと思って」と言って、【元】に書いてある日

本語の文章を見せてくれました。さらに「中国語の翻訳も付けておこうと思って、機械翻訳の助けを借りて自分で翻訳したんだけど」と言って、「機械翻訳」に示した中国語を見せてくれました。さて、あなたがこの中国語を最小限、修正するとしたらどうしますか（全面的に書き換えてしまったら、店長さんがかわいそうです）。

【元】

こちらは温かいドリップコーヒーです。どうぞご利用くださいませ。

【機械翻訳】

这是温暖的滴漏咖啡。请使用它。

【PE.】(JP)

这是热的滴滤咖啡，请享受它。

【FW.】(JP)

这是热的滴滤式咖啡。请享受一下。

【PE.】(Ch)

这是热的滴漏咖啡。请品尝。

【FW.】(Ch)

ポットの前に「热滴漏咖啡」と書かれたカードを立てておきます。

7.4.2 課題2 待遇度（禁止）

あなたは、ある楽器店の店員です。店長が「こういう〈お願い〉を張り出そうと思ったんだけど」と言って、【元】に書いてある日本語の文章をみせてくれました。さらに「中国語の翻訳も付けておこうと思って、機械翻訳の助けを借りて自分で翻訳したんだけど」と言って、「翻訳」に示した中国語を見せてくれました。さて、あなたがこの中国語を最小限、修正するとしたらどうしますか（全面的に書き換えてしまったら、店長さんがかわいそうです）。

【元】

お客様へのお願い

ハーモニカのご試奏はご遠慮ください。他のお客様にも気持ちよくお買い物をお楽しみいただくためにご協力をお願い

いします。

【機械翻訳】

对客人的请求

请不要进行口琴的试奏。为了让其他的客人心情舒畅地享受购物，请协助我们。

【P.E.】(JP)

对客人的请求

请不要试吹口琴。为了让其他的客人心情舒畅地享受购物，请协助我们。

【F.W.】(JP)

对客人的请求

为了其他客人，请不要试吹口琴。谢谢合作。

【P.E.】(Ch)

对客人的请求

请勿试吹口琴。为了让其他的客人心情舒畅地享受购物，请协助我们。

【F.W.】(Ch)

各位尊敬的顾客

请勿试吹口琴。谢谢您的合作。

7.4.3 課題3 重要事項の伝達

あなたの友人（中国語が少しできます）が、地震の被害に遭った知り合い（中国からの留学生、日本語があまりできない）に、【機械翻訳】に示した中国語のメッセージを大急ぎで送ろうとしていました。どうやら【元】に示した日本語を、機械翻訳で翻訳したのらしいです。さて、あなたは、横から口を挟んでこの中国語を修正しますか。もし最低限ここは修正すべきだと思うところがあったら書いてください。

【元】

そこにいると危険です。いっしょに避難所に行きましょう。避難所は、日本人でも外国人でも誰でも入ることができます。避難所は無料です。避難所には無料の水も食べ物もあります。避難所では毛布や布団を無料で貸してくれて、そこで寝ることもできます。避難所には医者や看護師もいます。困って

いることがあったら、遠慮なく言ってください。

【機械翻訳】

在那里很危险。我们一起去避难所吧。避难所，无论是日本人还是外国人都可以进去。避难所是免费的。避难所里有免费的水和食物。在避难所里可以免费借毛毯和被子，在那里也可以睡觉。避难所里也有医生和护士。如果有困难的话，请不要客气地说。

【P.E.】(JP)

（修正せず）

【F.W.】(JP)

你那里很危险。我们一起去避难所吧。避难所，不管日本人还是外国人，谁都可以使用。那里还有可以免费吃的和喝的。然后借毛毯和被子也不要付钱，还可以睡觉。医生和护士也在。如果你有问题的话，就问我。千万不要客气哦！

【P.E.】(Ch)

（省略）如果有困难的话，请不要客气，尽管说。

【F.W.】(Ch)

在那里很危险，我们这就一起去避难所吧。无论是日本人还是外国人都可以去，而且是免费的。避难所有免费的水和食物。而且在那里还可以借到（免费的）毛毯和被子，又有地方可以睡觉。如果觉得哪里不舒服的话，不用担心，避难所里还有应急的医生和护士。

7.4.4 課題4 手順の伝達

友人の中国語を勉強している学生が、日本に遊びに来る予定の台湾の知り合いに対し、SNS上で【機械翻訳】に示したメッセージを送ろうとしています。どうやら【元】に示した文を、機械翻訳を使って一生懸命中国語に翻訳したようです。相手が無事に目的地に着けるよう、修正が必要なところは修正したいと思います。ただし、友人の学習意欲を削がないよう、修正は最小限にしたいと思います。

【元】

京都駅から JR 京都線の新快速に乗ってね。女性専用車両に乗らないように気をつけてね。高槻で快速に乗り換えて、茨木で降りてね。茨木は新快速は停まらないから気をつけて。茨木から万博公園行きのバスに乗って、終点で下りて。お金は運賃の表示板で確認して、下りるときに払って。

【機械翻訳】

从京都站乘坐 JR 京都线的新快速。注意不要坐女性专用车厢。用高槻换乘快速，在茨木下车。茨木新快速不停，小心点。从茨木乘坐到世博会公园的巴士，在终点站下车。钱在运费的显示板上确认，下车时支付。

【P.E.】(JP)

从京都站乘坐 JR 京都线的新快速。注意不要坐女性专用车厢。在高槻站换乘快速，在茨木站下车。茨木站新快速不停，小心点。从茨木站乘坐往世博会公园的巴士，在终点站下车。钱在运费的显示板上确认，下车时支付。

【F.W.】(JP)

从京都站坐 JR 线的新快速。你一定要注意不要坐女性专用车厢。在高槻站要换乘到快速，然后要在茨木站下车。茨木站新快速不停，小心点。在茨木站下车后，你要坐往世博会公园的巴士，然后在终点站下车。运费可以在巴士里边的电子板上确认，下车时支付就好。

【P.E.】(Ch)

从京都站乘坐 JR 京都线的新快速。注意不要坐女性专用车厢。用高槻换乘快速，在茨木下车。茨木新快速不停，小心点。从茨木乘坐，到世博会公园的巴士，在终点站下车。钱在运费的显示板上确认，下车时支付。

【F.W.】(Ch)

要坐两次电车和一次巴士。首先从京都

站坐 JR 京都线的新快速，在高槻站下车。（小心不要坐女性专用车厢哦。）然后换乘快速电车，在茨木站下车。（新快速在茨木站不停的要小心哦。）在茨木站再坐往世博会公园方向去的巴士，终点下车。巴士里面的显示板上会显示运费的，下车时支付就可以了。

◇ 7.5 結果の考察

被験者間で修正の度合いを比較することは、もっと多くのデータに基づかなければ、意味がない。個人差が大きいと思われるからである。ここでは、被験者各人における P.E. と F.W. の距離に注目したい。これは、その状況におけるその被験者の許容範囲を示していると考えてよい。あくまで、その状況に限定された許容度だが。

ごく大雑把にいて、ニュアンスの厳密な伝達を重視しなければ、また、「自然な」「流暢な」という側面を重視しなければ、**百度**の機械翻訳は実用レベルに近づいていると言ってよい。

◇ 7.6 Pragmatic Transfer, Pragmatic Reformulation

「Pragmatic Transfer 語用論的転移」というのは、以下のような現象である。

日本にふだん居住している日本語話者が、台湾の飲食店で、(フロアに店員の姿が見えないので) 華語を使って従業員を呼びたいと思ったとする。しかし、呼びかたがわからない。この場合、a) 日本語でならどう言うか考え(例: すみませ〜ん!), b) その言語形式を直接、華語に置き換え(例: 對不起!), 発話したとする。これが語用論的転移である。

もし、上記の a) の段階で、台湾の社会では声を出して店員を呼ぶときにはどう言うかを考え、そこで見出した言語形式を、b) の段階で使用するのなら、“對不起!” を使うといったような語用論的転移は起こっていない(声を出して呼ぶということ自体が別の語用論的転移である可能性はある)。

§7 で見た機械翻訳の出力は、結果として、人間がしばしば起こしやすい「語用論的転移」と同じ作業をしている箇所が多く見られる。それを

手っ取り早く知るには、被験者 JP と Ch がそれぞれ、F.W. のところで修正、削除、付加している部分を見ればよい。

例えば、課題 2 の Ch の F.W. では、「他のお客様にも気持ちよくお買い物をお楽しみいただくために」に該当する部分は省略されており、いっぽう「各位尊敬的顧客」（尊敬するお客さまがた）という部分では、原文にはない「尊敬的」という修飾語が加わっている。

また、Ch も JP も、F.W. においては、「谢谢合作」「谢谢您的合作」という結びのことばを補っていることに注意をされたい。これは英語の “Thanks in advance” のように事前に謝意を示す、日本語の習慣では一般的でない社交用語である。

以上に指摘したような、目標言語の使用習慣に依じて配列順を換えたり、削除・追加を行ったりするような出力は、よほど固定した表現でなければ、現時点では機械翻訳に要求することは難しかろう。つまり、語彙レベル・文内部の構造のレベルでは実用の水準に近づいているが、その言語が使用される社会の慣習的ルールに則った談話の生成をするにはまだ至っていないということである。

回 8. 教師という権威は必要か

※ 8.1 人手でなら実現できるか

§ 7.6 で「その言語が使用される社会の慣習的ルールに則った談話の生成をするにはまだ至っていない」と述べた。しかし、これにより、「まだ人手に頼らないといけないのだ」と判断するのは浅慮というものである。

「慣習的ルールに則った談話の生成」を教えることは難しい（固定的なビジネス文書のライティング等を除けば）。非母語話者の教師ではそのルール自体を把握していないかもしれない。母語話者の教師ならだいたいじょうぶかという、母語話者は通常それらを非明示的知識として所有しているだけなので、それを明示的知識として提示するには、専門的な知識とスキルが必要である。

教師が教室において、tips として偶発的にそれ

らを教えることはあるかもしれないが、少なくとも日本の中国語教育においては、それらを系統的に学習する課程や教材の実例を、筆者は知らない。そのような実践を系統的に行なっている例も知らない。

要するに、機械ではまだできないが、人手による教室での教育でそれが実現できているわけではない、ということである。

※ 8.2 誰がそれを教えてくれるか

被験者 JP に、「前もって謝辞を述べる」中国語の習慣は、どこかで教わったのか、自然にそれを習得したのか尋ねたところ、「これは自然に気付きました。注意書きだけでなく、WeChat で中国人からお願い事をされる時も、こちらが先に返事をするよりも早く谢谢と言われるので、断る余地がないと友達と話したこともあったのでよく覚えていました」とのことであった。

この回答の示唆する重要なポイントは 2 つある。1 つは、その場でのやりとりであれ、SNS でのやりとりであれ、誰か他の話者（母語話者であれ学習者であれ）とのインタラクションを通して、このような慣習的ルールを学べるということ（=相手は教師でなくてもかまわない）、もう 1 つは、新しい習慣に会ったとき、それを自分のもつ習慣との対比において異質性を意識化する「気づき awareness」が学習のきっかけになるということである。

総じて、日本の伝統的な外国語学習の環境というのは、教室以外で学習目標言語を使う機会の乏しい環境であった。日本の外国語教育を担う機関としては大学が代表的であるが、かつてそこで採用されていた学習過程のモデルは、「formal instruction の場で構造に関する知識を学び、ネイティブ教師の授業や留学先で運用に関する訓練を行う」というものであった。§ 7.6 で言及した「その言語が使用される社会の慣習的ルールに則った談話の生成」については、「そういうのは留学すれば自然に覚えられるんだけどなあ」としか言及しない教師も珍しくはなかった。しかし、このような学習環境に身をおける学習者はごく一部であ

る（外国語大学／外国語学部という類の教育機関で、当該言語を専攻する学生ぐらいであろう）。

ところが、現在では、上述の Web2.0 と呼ばれるような環境下で、これまでにない形での使用者（この表現は当該言語の母語話者も学習者も含めている）の交流が実現している。ある程度メジャーな言語の場合は、使用者との交流以外にも多様なリソースも手に入れられる（欧麗賢 2016 参照）。もちろん、このような場所では「権威を持った教師」はいないのであるが、では、「正誤を判定する権威としての教師」が必要かといえば、必要ではない、というのが筆者の見解である。

必要ではない理由は、以下の 2 点である。

- 1) 言語構造の正誤を細かくチェックするという役目は、機械翻訳のような支援ツールの精度が上がるにつれ、存在意義が薄れつつある。
- 2) 「その言語が使用される社会の慣習的ルールに則った談話の生成」を習得することは、現状の学校教育を見ていると「権威ある教師がいても難しい」と判断せざるを得ない。かといって、例えば、Web2.0 のような環境において、学習目標言語の使用者のコミュニティで出現する相互交流の中で学ぶ方式が、この方面のスキルを身につけるのに有利であると実証するのは難しい。しかし、不利であると断じる証拠もない。つまり、どちらにせよ、現時点では、この能力を明示的に学ぶのは難しい。

9. 妥当性を検証しようとする意識

上述したように、次の 2 つは別物である。

- 1) 学習者がいろいろなツールを利用して適当な出力を自力で生成したいと努力する。
- 2) 学習者が自分の出力が妥当かどうかを自分で検証しようという意識を持つ。

学習者オートノミーを育てようと思うなら、2)

は欠かせない。では、どのようにしたら、それが育つのだろうか。

山崎 (2015) は、「権威をもった教師が出力をチェックする」というステージを最後に据える伝統的な学習設計の方法では、学習者が自律的に自分の出力の妥当性を検証する意識は育ちにくいのではないかという指摘をした。

これが正しければ、権威をもった教師の存在自体が、学習者のこの方面の意識の向上を妨げていることになりかねない。しかし、問題はそう単純ではない。教師が教室で行う formal instruction の中でも、そのような意識を高める学習活動を導入することは可能であろうし、教師のいない場での自由な交流が必ずそのような意識を高めることにつながると考える根拠もない。

筆者が考えるに、そのような意識を高めるために有効なのは、自分の発したことばが、教室といういわばサンドボックスの中での遊戯として捉えられるのではなく、現実社会での何かしらの課題遂行のためのことばであると捉えられるような言語使用の場である。つまり、自分がまちがえたことを伝えれば、関係者が現実にも不利益を被り、自分が不適切な表現を使えば、相手との人間関係が壊れる、というような言語使用の場である。

日本の教育機関で外国語を学ぶとき、このような場を設けるのは難しい。こう考えてくると、青木 (2018) が「学習者オートノミーと Web 2.0 ～これから e ラーニングのプログラムを作る時に考えたほうが良いと思われること」で行った問題提起は、単なる「新しいテクノロジーを取り入れよう」という提案ではなく、「言語学習の場において〈権威者〉は必要か」という、本質に関わる問題提起であると思われる。

10. まとめ

この文章では、青木 (2018) の問いかけに答える形で、「正誤を判断する権威を持った教師」が言語教育（とくに、日本の外国語教育）の場に必要かということについて、主に 2 つの主張をした。1 つは、文レベルの構造の正誤については、ICT の

支援があれば、そろそろ「構造の正誤をチェックする教師」の存在は不要になるのではないかということ、もう1つは、「学習者が自分の出力の妥当性を検証する意識を高める」ためには、教師の存在は特に必要ではない、ということである。

※ 本稿は、日本学術振興会科学研究費補助金「初級学習者を対象としたコンテンツ駆動型中国語学習基盤の構築」(平成30年度、基盤研究(B)、研究代表者：田邊鉄、課題番号：18H00682)による成果の一部である。

【参考文献】

- Cappellini, M., Lewis, T. & Rivens Mompean, A. (Eds.). (2017). *Learner autonomy and Web 2.0*. Sheffield: Equinox.
- Holec, H. (1985). On autonomy: Some elementary concepts. In P. Riley (Ed.), *Discourse and learning*. New York: Longman, 173-190.
- Reinders, H. & White, C. (Eds.). (2011). Special issue on learner autonomy and new learning environment. *Language Learning and Technology*, 15(3). [http://www.lltjournal.org/collection/col_0125_35919]
- 青木直子 (2018). 「学習者オートノミーを育てる教育実践：これまでの蓄積と新たな展開」第11回大阪大学専門日本語教育研究協議会『学習者オートノミーを育む言語学習とその支援』における講演, 2018.2.16. [講演の際のスライドは, <http://ciec.osaka-u.ac.jp/wp-content/uploads/2018/03/Final-draft.pdf> で入手可能]
- 青木直子・中田賀之 (2011). 『学習者オートノミー：日本語教育と外国語教育の未来のために』. 東京：

ひつじ書房.

- 欧麗賢 (2016). 「日本語学習者がインターネット上のリソースを教室外の学習に利用し始めるメカニズム：修正版グラウンデッド・セオリー・アプローチ (M-GTA) による理論構築」, 『阪大日本語研究』, 28, 77-108.
- 金子真也・山崎直樹 (2013). 「機械翻訳は中検3級に受かるか?」, 『漢字文献情報処理研究』, Vol.14, 161-168.
- 日本語教育学会(編) (2005). 『新版日本語教育事典』. 東京: 大修館書店.
- 山崎直樹 (2015). 「自然言語処理技術の発達が外国語教育にもたらすもの」, 『漢字文献情報処理研究』, Vol.16, 6-16.

注

- [1] 学習者オートノミーの研究に多大な貢献をされた青木直子氏 (大阪大学教授) は、この講演のあと、ほどなくして逝去された。謹んでご冥福をお祈り申し上げます。
- [2] 本稿では「権威」を「言語」に絞って考察したが、これを「文化」に換えても、結論は変わらない。現在の言語教育では「文化をどう扱うか」も重要な課題であるが、その扱う「文化」とは何かを考えると、「学校教科書でよく使われる『わが国の文化』という表現で表されるような『唯一の標準的文化』というものはありうるのだろうか」ということが必ず問題になる。そこから、「文化についての情報を正しいと判断するのは誰か、誰がそのような権威を持つのか」という問いかけが現れる。それへの回答は、「言語の権威」についての回答と同じである。
- [3] <http://lang-8.com>

『皇明條法事類纂』電子化劄記

文献学・資料学とデジタルアーカイブのはざま

小島 浩之 (こじま ひろゆき)

回 はじめに

東京大学総合図書館は、2018年8月15日に皇明條法事類纂^[1] デジタルアーカイブを公開した^[2]。

<https://iif.dl.itc.u-tokyo.ac.jp/repo/s/koumin/page/home>

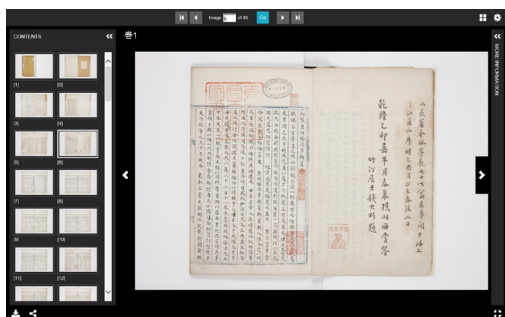


図1 皇明條法事類纂デジタルアーカイブ

筆者はこの構築に関与し得たので、本稿を通じて公開に至るまでの経緯で得た知見、とくに文献学・資料学的研究に基づく成果をまとめ、当該資料を利用する研究者の参考に供したい。

『皇明條法事類纂』の研究は、富山大学教授の徳永洋介氏が主宰する共同研究班で行われ、科学研究費補助金・基盤研究 (C)「宋元時代の刑事政策とその展開」(課題番号: 23520860、代表者: 徳永洋介)の助成を受けて全体のデジタル画像を撮影し、続く基盤研究 (C)「元明時代の法制に関する基礎的研究—『皇明条法事類纂』の分析を中心として—」

(課題番号: 15K02890、代表者: 徳永洋介) において内容の精読と冊子目録『皇明条法事類纂条名目録』(書誌事項は注3を参照)を刊行した。

『皇明條法事類纂』デジタルアーカイブは、これらの成果に基づき、東京大学総合図書館および情報基盤センターの協力を得てデータを整備し公開したものである。

回 1. 法典編纂史上の位置づけ

『皇明條法事類纂』は、東京大学所蔵の鈔本(写本)しか存在せず、明代中期の法制史料として貴重かつ有用なものである。このため、まずは中国の法典編纂史上にこの史料を位置づけておこう^[3]。

滋賀秀三氏は、前近代中国の法典編纂を史的に論ずるにあたり、法令をその構造から①基本法典、②単行指令、③副次法典の三種に分けて分析した。いま滋賀説に従えば、基本法典とは体系的に編纂され、当該時代の法制の中心をなす基本的な法令であり、律や令などが該当する。単行指令は必要に応じて、詔や勅など皇帝の命によって出される臨時的法令のことを指す。副次法典は、単行指令や判決例・裁判例・執務内容・行政処分などの先例から必要なものを抽出し編纂されたもので、基本法典に対する補完・修正・追加の機能を持つ。こういったものを元代では「格例」、明代前半期では「条例」と呼んだのである。

ただし明代の条例は、恣意的な運用を防ぐため、その有効期限を皇帝一代限りとしていた。新帝の即位に際しては、前代条例の一括廃止と、法源の

明律への一元化が、各帝の即位詔の中で繰り返し明言されたのである。しかし、弘治帝の即位詔からは、こういった言及がなくなり、現実の法実務への影響を考慮し、過去の条例をまとめて有用なものを活用する方向に舵がきられる。『皇明條法事類纂』が編纂されたのは、この時期以後のことであり、具体的な事案に関する条例類を的確に検索し得ることを目的とした実務の書として編纂されたものと考えられる。

収録の条例総数は 1,297 で、このうち 24 条は条例名のみで本文を欠く。この 24 条を含む 35 条は繫年不明であるものの、1,262 条は少なくとも年までは確定できる。年ごとの条例数を可視化すると図 2 のようなグラフとなる^[4]。

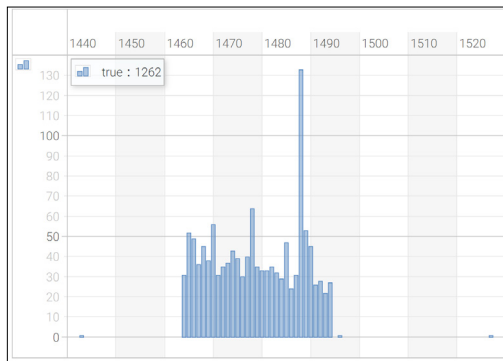


図 2 収録条例数の変化

グラフから、正統 8 年 (1443) と嘉靖 6 年 (1527) の例外的な 2 条を除き、他の繫年の判明する 1,260 条は天順 8 年 (1464) から弘治 9 年 (1496) までの範囲に収まっていることがわかる。弘治の初年を含むものの、基本は弘治帝の父である成化帝在位期間中の天順末年から成化年間の条例集であることが見て取れる。このグラフは、先に述べた明朝における条例取り扱いの変更を雄弁に物語っているとさえ言う。

書名に含まれる「条法」という言葉は法令の箇条書の意味^[5]で、その使用は宋代に遡り、南宋時代の『慶元条法事類』や『吏部条法総類』といった法制史料が今に残っている。また「事類」とは案件や項目ごとに分類整理することを指す。ただし『皇明條法事類纂』は、「事の発端を述べた上

奏文等の引用から始まって、その後の文書のやり取りや各官庁での議論・判断の結果等の記述を経た後、皇帝による裁可の言葉で締めくくられる形を常とし」、官府間の文書の送達過程もが比較的詳細に記録されている^[6]。この点からすれば、『皇明條法事類纂』は書名に条法とあるにもかかわらず、箇条書の法令集でないばかりか、直後に編纂される『問刑条例』などの条例集とも異なる趣を持つ。

一方でその構成を見ると、五刑類・名例類・吏部類・戸部類・礼部類・兵部類・刑部類・工部類という 8 つの大分類からなっている。総則部分の名例を五刑と名例に二分しているものの、総則の後に『周礼』さらには『唐六典』の系譜を引く六部分類を用いる構造は、明律と同じである。

このように『皇明條法事類纂』は、その端々に唐宋時代以来の伝統的な観念を残しつつも、構造的には明律を基本とし、内容的には『問刑条例』など明代の一般的な条例集の成立に至る過渡期の姿を示している。一般的な法令集の枠組を超え、明代中期の行政史料としての価値を有しているとも言えるだろう。

こういった稀少性と重要性に鑑み、1966 年には古典研究会から影印本が、1994 年には楊一凡氏による標点本（『中国珍稀法律典籍集成』乙編 4～6 冊所収、科学出版社）が出版され、これらが現在の通行本となっている。

2. 中国文献学上の諸問題

先述のように『皇明條法事類纂』は、東京大学のみが所蔵する天下の孤本である。基礎的な考察は既に仁井田陸氏によってなされているので^[7]、仁井田説を踏まえつつ、原本の調査から新たに判明した事実を中心に読者の参考に供したい。以下に基本的な書誌事項を掲げるので、これに沿って順にみてゆこう。

皇明條法事類纂五十卷，(明)戴金奉勅編次，鈔本，線装六十四冊（合冊製本十三冊），每半葉十三行，白口，双魚尾，四周双辺，藍格，

本文葉：皮紙，副葉：竹紙，黄表紙，縦 29.1 糎，横 19 糎

印記：「仲魚圖象」（陽刻朱印、3.9 × 2.8 糎）、「簡莊藝文」（陽刻朱印、1.7 × 1.1 糎）、「海寧陳鱸觀」（陽刻朱印、4.3 × 1.0 糎）、「得此書費辛勞後之人其鑒我」（陰刻朱印、3.3 × 2.5 糎）、「香圃收匪書畫」（陽刻朱印、2.0 × 1.4 糎）

◇ (1) 書名表記

各巻首と題簽に「皇明條法事類纂 [巻数]」（ただし、巻首は「皇明條法事類纂目録 [巻数]」となっているものもある）とあるほか、第一巻の序文に「御製皇明條法事類纂序」、続く総目に「大明文淵閣鈔寫永樂大典條法事類纂總目」、総目末尾に「大明文淵閣永樂大典鈔寫 進本條法事類纂總目終」などとある。

ただし、題簽の紙は機械漉の洋紙で、東京大学総合図書館所蔵の他の漢籍にもよく見られる様式である。後述する不分巻との関連からすれば、当初は題簽が無く、東京大学において各巻頭の書名巻数を採って関東大震災後に附したものと考えられよう。

御製とされる序文に関しては、仁井田陸氏により、『万曆会典』と『正徳会典』の序文を併せて字句を修正したものであることが論証されている。このため「御製皇明條法事類纂」という書名は信用し難い。同様に仁井田氏によって『永樂大典』云々もこじつけであることが証明されているので、総目冒頭と末尾の書名もやはり信じることはできない。このため各巻首の「皇明條法事類纂」が今のところ信ずるにたる唯一の書名となる。

◇ (2) 巻数と冊数

『東京大学総合図書館漢籍目録』は五十巻とするのでこれに従った。ただし、実際には 64 冊の線装本からなっており、64 の物理単位で 50 巻分の内容を構成している。64 冊の内訳は、巻首に書名と巻数表示のある線装本が 50 冊、巻首に書名も巻数も記載のないものが 14 冊である。この 14 冊は不分巻、未分巻、または附編などとされ、巻数表記のある線装本の後に排され区別されてきた。

しかし、張建民氏や劉篤才氏が詳細に分析しているように¹⁸⁾、不分巻所載の条文名と各巻の巻頭目録（後述）を対照するに、その多くは巻 3、8、9、11、14、18、21、25、27、28、35、39、45 にあるはずの条文である。したがって、原則的には上記各巻の第二分冊だと判断される（具体的な対応関係は表 1 参照）。ただし、不分巻の 14 冊目だけは、管見の限りどの巻とも対応関係がなく、嘉靖 6 年の条例¹⁹⁾ という所収年代としても例外的であるため、全体での位置付けは不明とせざるを得ない。

不分巻号	本来の巻冊
不分巻-1	巻 25 第 2 冊
不分巻-2	巻 9 第 2 冊
不分巻-3	巻 18 第 2 冊
不分巻-4	巻 21 第 2 冊
不分巻-5	巻 3 第 2 冊
不分巻-6	巻 27 第 2 冊
不分巻-7	巻 8 第 2 冊
不分巻-8	巻 28 第 2 冊
不分巻-9	巻 45 第 2 冊
不分巻-10	巻 39 第 2 冊
不分巻-11	巻 14 第 2 冊
不分巻-12	巻 11 第 2 冊
不分巻-13	巻 35 第 2 冊
不分巻-14	不明

表 1 不分巻と本来の巻冊の対応表

ここから『皇明條法事類纂』全体の本来の順序（原秩序）は、表 2 のようであったと想定される。冊子目録では各条文を原秩序に従って排列し、デジタルアーカイブもこれに準じている。

一方、標点本は『中国珍稀法律典籍集成』乙編の 4 および 5 冊目を正編として、書名と巻数表示を持つ線装本 50 冊分の内容にあて、6 冊目を附編として不分巻にあてている。つまり標点本は現状の冊子排列順序（現秩序）に従って整理されている。

◇ (3) 編纂者

編纂者である監察御史の戴金に関して、仁井田

原秩序	対応巻	類	原秩序	対応巻	類
1	巻 1	五刑	33	不分巻 -1	兵部
2	巻 2	名例	34	巻 26	兵部
3	巻 3	名例	35	巻 27	兵部
4	不分巻 -5	名例	36	不分巻 -6	兵部
5	巻 4	名例	37	巻 28	兵部
6	巻 5	名例	38	不分巻 -8	兵部
7	巻 6	名例	39	巻 29	兵部
8	巻 7	吏部	40	巻 30	兵部
9	巻 8	吏部	41	巻 31	兵部
10	不分巻 -7	吏部	42	巻 32	刑部
11	巻 9	吏部	43	巻 33	刑部
12	不分巻 -2	吏部	44	巻 34	刑部
13	巻 10	吏部	45	巻 35	刑部
14	巻 11	吏部	46	不分巻 -13	刑部
15	不分巻 -12	吏部	47	巻 36	刑部
16	巻 12	戸部	48	巻 37	刑部
17	巻 13	戸部	49	巻 38	刑部
18	巻 14	戸部	50	巻 39	刑部
19	不分巻 -11	戸部	51	不分巻 -10	刑部
20	巻 15	戸部	52	巻 40	刑部
21	巻 16	戸部	53	巻 41	刑部
22	巻 17	戸部	54	巻 42	刑部
23	巻 18	戸部	55	巻 43	刑部
24	不分巻 -3	戸部	56	巻 44	刑部
25	巻 19	戸部	57	巻 45	刑部
26	巻 20	戸部	58	不分巻 -9	刑部
27	巻 21	礼部	59	巻 46	刑部
28	不分巻 -4	礼部	60	巻 47	刑部
29	巻 22	礼部	61	巻 48	刑部
30	巻 23	兵部	62	巻 49	工部
31	巻 24	兵部	63	巻 50	工部
32	巻 25	兵部	64	不分巻 -14	不明

表 2 『皇明條法事類纂』巻冊構成原秩序

氏は最終的に次のように述べる。

戴金がその監察御史であった世宗の嘉靖二年後数年間（あるいはその後）に編纂したとすることだけは、あるいは肯定できるかと思う。しかし戴金を編纂者とできるかどうかさえ、実は確信しかねる。【217 頁】

戴金は実在の人物であるものの、『皇明條法事類纂』の編纂に関しては他に傍証史料がなく、残念ながら今に至るも彼が本当に編者であったのかについてはよく解ってはいない^[10]。

◇(4) 版式・料紙など

匡郭（縦 22.7 糎、横（半葉） 16.3 糎）および界、魚尾は全て藍色で印刷されており、四周双辺で毎半葉 13 行である。版心に葉数その他の情報は記されていない。明鈔本の野線類は万暦以前が黒もしくは無界、以後が藍もしくは朱が主流であるという^[11]。これは、仁井田氏が序文の偽作を万暦 15 年以降、もしくは清代に入ってから作られたものとする点と符合する。

本文の料紙を顕微鏡で観察^[12]すると、長く幅広く薄膜に包まれており、線条痕や結節が観察されるなど楮系統の繊維の特徴を備えている。篋目痕も 1 寸あたり 21 本と粗いことから、楮繊維に

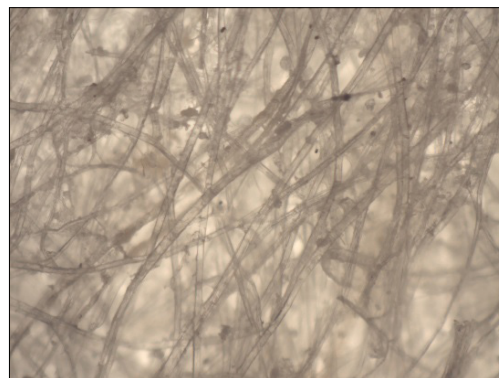


図 3 本文紙の繊維（100 倍で撮影）

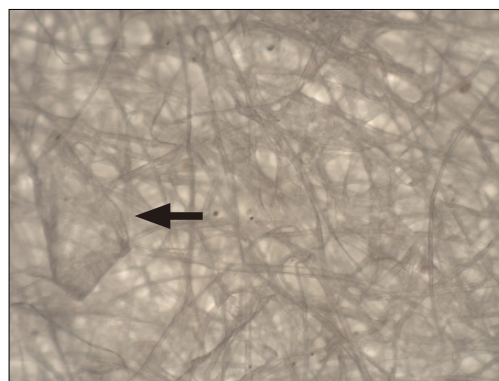


図 4 遊び紙の繊維（100 倍で撮影）

適したものであることがわかる^[13]。繊維は適度に切断されていることから、日本で作られた紙ではなく、棉紙や皮紙などと呼ばれる楮（桑科）の韌皮繊維を主とする中国産樹皮紙と判断される。

これに対し表紙・副葉・遊び紙^[14]は、繊維が細くしなやかで、竹に特徴的な孔紋道管（図4矢印部分）が見られることから竹紙と考えられる。簀目痕は1寸あたり52本と細かく、40本以上を計る中国産竹紙の特徴に一致する。

樹皮紙は竹紙より上質の紙である。これが本文紙に使用され、装幀部分には竹紙が用いられているのは、明代までの理に適った料紙の使い分け方である^[15]。『皇明條法事類纂』は素材や形式の面からみれば素性のよい鈔本であり、年代の上限を明に比定し得る要素を持っている。

◇(5) 装幀

原装は四つ目綴じの線装である。表紙は竹紙に黄色の染料（おそらく黄堊）が塗布されている。

ただし、現状では、線装本5冊ないしは4冊を合本として洋装に仕立て上げ、13冊の物理単位からなっている。1911年の『東京帝國大學附屬圖書館和漢書書名目録』増加第二には「五十卷六四冊」とあるので、この時点では合冊製本されていなかったようである。

1923年の関東大震災で東京大学は壊滅的な被害を受け、図書館も蔵書の大半を焼失してしまう。幸いにも『皇明條法事類纂』は、文学部西洋史学科教授の齋藤清太郎が借り出していたため、奇跡的に難を逃れ、1935年頃に改めて文学部から総合図書館に戻される。図書館の目録や原簿類にはこの頃から13冊と記されるようになることから、震災後の整理の中で合冊製本されたようである^[16]。

ところで、線装本各冊の表紙背側上部には、明治34年（1901）に東京大学が受入・整理した際の請求記号を書いたラベルが貼られている。当時の請求記号は「貳一三八八」^[17]で、さらにこれが一から八までの下位分類で展開されている。つまり東京大学では当初、『皇明條法事類纂』に関して線装本8冊を1組として取り扱っていたこと

線装	合冊	帙/箱	17		34	7		不分1		
1	1	1	18	4	35	5	8	不分2	11	7
2			19		36			不分3		
3			20		37			不分4		
4	2	1	21	5	38	6	9	不分5	12	8
5			22		39			不分6		
6			23		40			不分7		
7	3	2	24	6	41	7	10	不分8	13	8
8			25		42			不分9		
9			26		43			不分10		
10	4	3	27	7	44	8	11	不分11	13	8
11			28		45			不分12		
12			29		46			不分13		
13	4	4	30	5	47	7	10	不分14	13	8
14			31		48					
15			32		49					
16			33		50					

表3 物理単位ごとの対応関係

※ 「線装」は線装本としての物理単位（現状の巻数）、「合冊」は現状の物理単位、「帙/箱」は関東大震災以前の収納単位を指す。

になる。この単位で帙か箱に収納されていたと考えるのが妥当であろう。

以上、『皇明條法事類纂』における複数の物理単位の対応関係を整理すると表3のようになる。これを見ると線装本の各冊が、現状で何冊目に合冊されているのか、また関東大震災前にどのような組み合わせで収納されていたのかが理解できる。

また、1901年に東京大学が受け入れた時点で、表2で示したような原秩序は既に失われていたこともわかる。この状態のまま、不分巻の内容を精査せず、巻首の記述に従って線装本の各冊に題簽を附し、さらに機械的に合冊処理をしてしまったため、この本の全体像が分からなくなってしまったのである。

◇(6) 印記・識語・観款

印記として示した鑑蔵印のうち「香圃收匪書画」以外の4種はいずれも陳鱣（1753-1817）のものである^[18]。

陳鱣は、字を仲魚、号を簡莊といい、浙江海寧の人で、清代後期の儒学者かつ大蔵書家として知られる^[19]。特に「得此書費辛苦後之人其鑒我（この書を得るに、辛苦を費やせり、後の人、其れ我を鑑とせよ）」の印は、陳鱣の書籍収集にかける熱き思いが吐露されているものであり、これらから『皇明條法事類纂』も陳鱣の旧蔵とみて間違いのないだろ



図5「得此書」印



図6「仲魚圖象」印

う。

陳鱣の蔵書は彼の死後に売却されたので、『皇明條法事類纂』もこの時点で陳家の手を離れたのであろう。印記のうち「香圃收匪書画」が次の所有者のものではないかと推察されるが人物を特定できず、後攷を俟ちたい。

さて、仁井田氏の論考以来、本書の中で序文とともにその真偽が疑わしいとされている部分に、巻1冒頭の識語や観款がある。ここでは序・跋・解題的な文章を識語とし、閲覧・観覧の記録のみを記した観款と区別する。これらを表紙に近い方から順に並べると表4のようになっており、みな副葉や遊び紙に書かれている。

巻1の装幀だけは他の巻と副葉や遊び紙の構造が異なっている。表紙に貼り付く副葉は三葉あり、順に竹紙（茶色と白色の紙の貼り合わせ）、竹紙（白色）、楮紙（白色）となっている。現状でこれらは剥がされて、いずれも遊び紙のように見えるが、紙に残るのりしろや剥がれ跡などから、小口部分に糊付けして三葉を貼り合わせてあったことは明白である。遊び紙は他の巻と同じく一葉分であるもの、竹紙ではなく楮紙である。副葉と遊び紙の楮紙はいずれも本文紙と簀目痕が一致するから同じ紙だと考えてよい。

巻頭の冊子には総目もあり、頻繁に参照されて痛みやすく、補強のために最初から装幀を他の冊子と変えてあったものと推察される。

識語1は、剥がされた副葉の二枚目の竹紙のオモテ面に書かれている。内容は宋代の「条法事類」に関してまとめたもので、本書の内容と直接関わるものではない。青みがかった墨（松煙墨カ）で書

識語 1	陳振孫直齋書録解題曰
識語 2	洪氏讀書叢録題一則
識語 3	錢竹汀条法事類纂跋
観款 1	孫星衍観款
観款 2	陳文述観款
観款 3	朱彝尊観款
観款 4	錢大昕観款

表4 識語・観款一覧

かれ、識語・観款の中では最も古い書風にも感じることが、いつ誰が何の目的でこのような場所に書いたのか特定し難い。

識語2は、洪頤煊(1765-1833)『讀書叢録』からの『皇明條法事類纂』に関する記述の抜き書きという建前だが、仁井田氏が指摘するように、『讀書叢録』と同じ内容は見当たらない。偽作としても『讀書叢録』という書名が世に知られてから書かれたものであることは疑いない。『讀書叢録』は道光元年(1821)の序文を持つから、これより後に書かれたものと判断される。陳鱣は1817年に没しているので、この識語は陳家から売却され、東京大学に寄贈される1901年までの間のものと推定される。

この識語には、朱彝尊と錢大昕の著述からの引用があり、仁井田氏はこれらを別の識語とみて分けて考えている。しかし、字配りや筆跡からみて全体で一つの識語であると考えの方が妥当であろう。なお、管見の限り、朱彝尊や錢大昕の著作に該当する内容は見当たらない。

識語3は、錢大昕によるとされる本書の跋文であるが、やはり現存する錢大昕の著作内に該当するものは見当たらず、真偽が定かではない。また、通常、跋文は書物の末尾に書くものであり、冒頭に置くのは例外的な感が強い。

観款は孫星衍(1753-1818)、陳文述(1775-1845)、朱彝尊(1629-1709)、錢大昕(1728-1804)のものとしてされる。錢大昕と孫星衍は陳鱣と学問的交流があったようなので、その意味では彼らの観款が存在してもおかしくはない。

最古の観款は朱彝尊の観款3で、「小長蘆金風亭長七十七翁彝尊、関于海上之紅葉山房、時乙酉

月正立春後二日。」とあり、彼が七十七歳である乙酉年正月つまり康熙44年(1705)と特定できる。次に古いのは銭大昕の観款4「乾隆乙卯嘉平月春皋、携此冊賞鑒。竹汀居士錢大昕題。」で、乾隆乙卯すなわち乾隆60年(1795)のものである。

一方で孫星衍のものとする観款1は、紀年が無い。また、観款2は次のようになっている。

頤道居士文述讀於青柯館
道光八年小春漁沚拜誦

仁井田氏は一行目を陳文述のものとし、二行目を別の観款として扱っている。ただし、筆致から見ると同一人のもののようにも見える。道光8年(1828)であれば陳文述も存命であるから、二行で一つの観款と扱えそうにも考えられる。ただし、「漁沚」という言葉と陳文述との関係性が立証できないため、本稿ではひとまず二行まとめて取り扱い、両論併記しておく。

識語2は、剥がされた副葉のうち楮紙のオモテからウラの半ばにわたり、識語3は続けて遊び紙のオモテの三分の二までを占める。観款1と2は識語3の後、遊び紙のオモテ末尾に、観款3と4は遊び紙のウラに書かれている。

これらから、識語2から観款4までの執筆順序やその信憑性に関して、以下のように考えてみたい。最初、康熙44年(1705)に朱彝尊の観款3が書かれた。この時点では、副葉は剥がされておらず、「竹紙、竹紙、楮紙」の順で表紙に貼り合わせてあったであろう。その後、『皇明條法事類纂』は陳鱣の所蔵するところとなり、遊び紙のウラに各種の鑑蔵印が捺された。そして、乾隆60年(1795)に陳鱣と交流のあった銭大昕がこの書を閲覧して、観款4を書いた。

その後、嘉慶22年(1817)に陳鱣が死去して本書が売却された後、道光元年(1821)以降に、観款3と観款4の存在を前提として識語2が偽作された。この時点で副葉は剥がされ、現状のように遊び紙が4葉存在するような形になったはずである。

識語3も偽作の可能性が高いが、銭大昕と知遇

のあった学究肌の陳鱣の手によるものとは考え難い。また、識語3が書かれているのは副葉のうち楮紙のウラの後半三分の一である。この部分は副葉全体の最末尾にあたり、副葉の剥がされる前から露出している部分であるから、識語2より前に書かれたとすれば、この面の冒頭から執筆されていなければおかしい。前半を識語2が占め、識語3はこれに続けて、違和感のない行間で書かれている。これらから、識語3が識語2より後に書かれたことは確実であろう。

孫星衍の観款1は真筆であるならば、彼の没年から考えて嘉慶23年(1818)以前に書かれていなければならないから、識語2や3より前に書かれたはずである。しかし、他の孫星衍の署名の筆致と比べると、旁である「生」の、一画目の入筆と四画目の処理が通常の「星」字の崩し方と異なっており、全体として「是」字のようになってしまっている。うがってみれば、第三者が書写の際に「星」を「是」に誤ったのではないかという疑問も残る²⁰⁾。

また観款4は、二行目だけは道光8年(1828)に書かれた可能性を残すが、一行目に関して現時点では如何とも評価し難い。

このように、装幀の構造と内容から理詰めで考えると、識語1と識語2は副葉を表紙から剥がした変則的な紙葉に書かれており、識語1は本文との対応関係は不明で、識語2は偽作の疑いが濃厚である。続く識語3も偽作の可能性が高く、観款1と2は真偽の確定には証拠不足の感が否めない。すなわち現段階において、識語1から観款2までの部分は素性の良いものとは言えず、その取り扱いに関しては慎重な判断が必要となろう。一方で朱彝尊と銭大昕の観款3と4は、これをなくしては識語2や3の偽作の意味が無く、書かれている場所や位置にも疑問の余地はないので、他に比べて信頼に足るものと言えようである。

ただし、今回はもっぱら書物の構造と内容からの分析であり、厳密な筆跡調査は行っていないので、より詳細な分析は今後の研究に委ねたい。

回 3. 構成上の諸問題

◇ (1) 条文名の不一致

先述のように『皇明條法事類纂』は、各種の事案を8類に分類している。さらに類の下には具体的な件名により複数の条を排列区分した下位分類がある。伝統的な目録学からすれば、類の下位の小分類なので目と表現できよう。目の下に各条が排列されているので、『皇明條法事類纂』は「類一目一条」という三階層の分類構造になっている。

第1巻の巻頭には総目があり、各巻所収の類名と対応する目名、さらに目ごとの条数が示される。一方、各巻の冒頭には当該巻所収分の詳細な目次（巻頭目録）があり、該当巻に収録される類名、目名、条名が記載されている。

ところが、実際に本文にあたって巻頭目録の条名と、本文見出しの条名を対照してゆくと両者の間で齟齬の多いことに気付く。条文名の文言が異なるもの（単純な誤植から、明らかに条名が異なるものまで様々なレベルがある）、巻頭目録に条名があっても本文が存在しないもの、逆に巻頭目録に記載のない条文が存在するもの等々、非常に多種多彩である。このため、総目に書かれている条数も全くあてにならない。

さらに、既に述べたように不分巻の問題があり、どこにどのような条文があるのかを知ることすら難しい、研究者泣かせの資料なのである。

内容検索のため研究者の便宜を図ったものとして、大澤顯浩『『皇明条法事類纂』目録（I）：五刑・名例・吏部・戸部・礼部』（『大手前女子大学論集』32, 1998年）があるが、巻頭目録や標点本との対応関係の明示はなく、今回、基本情報の整理が最優先だと判断された。

◇ (2) 影印本と標点本

早い時代に影印本は出ていたものの、縮小され印刷も悪いため文字が判読しづらいのが難点である。

残念ながら、影印出版当時の撮影フィルムは東

京大学には保存されていない。年代から考えて劣化のため使用できなくなり廃棄されたものと考えられる。このため、影印本作成時の状況がわからないままであったが、広島大学図書館に当時のものと考えられる紙焼写真（以下、広大本とする）の存在を確認できた。

広島大学図書館に残されている受入関係の書類から判断すると、広島大学文学部東洋史研究室の教員が、文献複写により紙焼を入手し、線装本と同じ単位で64冊（縦14.8糎、横10.3糎）に製本したもので、昭和36（1961）年5月22日付で広島大学図書館に受入登録されている^[21]。

影印本の出版が1966年であるので、5年の間に貴重書指定の大部な図書を複数回撮影するとは、資金的にも管理面からも考え難い。影印本と同じフィルムを元に作られた紙焼写真とみて間違いないであろう。ただし、広大本はネガからの直接の焼き付けのため、濃淡もはっきりし細かな文字も確実に視認できる。昭和30年代の『皇明條法事類纂』の状況を知る上で大変貴重であろう。

広大本は、見開き撮影、つまり一枚が前葉のウラと後葉のオモテの組み合わせの写真からなっているのに対して、影印本は一枚の写真が一葉分の姿で掲載されている。原本には糸を切って再綴した形跡はないから、影印本はフィルムから紙焼したものを切り貼りして一葉単位の写真とし、改めてそれを撮影したフィルムを元に出版されたとしか考えようがない。そういった視点から影印本を子細に見ると、版心部分で左右がずれているものが散見される。

フィルムは撮影ネガが最も解像度が高く、複製されればされるほど解像力は落ちてゆく。上記の工程であると、影印本は「原本→撮影ネガ→紙焼→再撮影ネガ→印刷物（影印本）」というように、手法は全て同じでないものの原本からみれば第5世代の複製に当たるもので、解像度の低下は如何ともしがたい。かつ、縮小した上、グラデーションで表現されているモノクロ写真を、白黒のみの二値の印刷で表現せざるを得なかった。これでは手間暇かけて読みにくくしているようなもので、せめて撮影ネガをそのまま使えば、「印本の

文字が読めないという問題」も、そこまで深刻なものとはならなかったであろう。

標点本は、明示していないが影印本を底本にしていると考えられ、校訂に難渋した跡がみられる。一方で、影印本からの字句の読み取り間違いや、逆に根拠の明示のない字句の訂正が散見され、研究利用のためには原本との丁寧な照合が欠かせない。

もちろん影印本も標点本も、作成された時代における可能な限りの努力によって作られたものであることは重々承知している。しかし、テキストクリティークという観点からみればいずれも問題多しとせざるを得ない。特に標点本は、影印本の問題点を軽減しようとして、結果的に新たな問題を増やしてしまっている。これらに目をつぶったまま『皇明條法事類纂』を用いた歴史研究を行うことは不可能である。そこで今回のデジタルアーカイブ化においては、誰もが利用できるいわば第三の通行本を提供したいというのがねらいの一つでもあった。

◆ (3) 冊子目録とデジタルアーカイブ

これまで述べた『皇明條法事類纂』に関する全ての問題点を勘案した結果、まず冊子目録を以下の方針で作成した。

- (1) 標点本を底本として巻頭目録と本文の見出し標題の一覧を作成し、それに対応する原本の各標題を入力し、両者を一つの表にまとめる。
- (2) 対応する標点本と影印本の掲載箇所、さらには撮影した原本画像のコマ数を(1)の表に付け加える。
- (3) 条文排列は原秩序に基づくものとし、それに合わせて50巻の中に不分巻を組み入れて表現する。

つまり、冊子目録では原秩序による全体像の把握と、通行本二種の条文名のそのままの対比、および掲載箇所の一覧化が何よりも優先された。

一方のデジタルアーカイブでは、冊子目録の

データをメタデータとしつつも、以下の点について別途考慮した。

- (1) デジタルアーカイブは対応する原本画像を公開するものであるため、メタデータについては画像データを基準としたものに改め、原本画像と標点本が対比されるようにした。
- (2) 紙媒体と異なり、種々の観点からデータのソートができるようにした。
- (3) 検索はメタデータ単位つまり、条例単位とした。ただしこれだけでは、序文・総目・巻頭目録・識語・観款などが閲覧できない。また、各条例と原本冊子との対応もわかりにくい。これらを解決するために別途各巻を通覧できるようにもした。

□ おわりに

以上、人文科学系研究者が歴史資料のデジタルアーカイブの作成に関与するにあたり、その公開のために情報を収集し内容を検討してきた過程の一端を示してみた。資料をモノの側面と内容の両面から検討し、どういった構成が利用する側にとって望ましいかを常に念頭においたものであることをご理解いただければありがたいと思う。

前近代の資料にはボーンデジタルのデータが存在しない以上、作成されるデジタルアーカイブは、基となる資料についてモノと内容の両面から本質を捉えたものでなければ、研究に有用なものとはなり得ない。そうでなければ、作成されたデータを基礎として、たとえば図2のような可視化による分析を進めても意味のないものになり、いかにデジタル・ヒューマニティーズの方法論を駆使しても、絵に描いた餅になってしまう。

そこで思い出すのは、大学図書館の目録の電子化が始まった頃、量が質に転化するというような妄想にとらわれ、過去のカード目録データを機械的に処理してデータ化し、目録システムの構築のために利用を試みたことである(Reconデータ)。しかしデータの精度が悪くて使いものにならず、大学図書館は現物に基づいて目録データを

作成することが、質的量的いずれにとっても最善であることを学んだのである。つまり大学図書館はモノに基づかないデジタル化がいかに空しいものか、十二分に理解しているはずである。一方で、人文社会系の研究者は、『皇明條法事類纂』における影印本から標点本への流れのような悪循環に泣かされてきた。両者が手を結ぶならば、大学図書館も研究者も同じ轍を踏んではならないし、研究者は今後さらに積極的に、デジタルアーカイブの構築に寄与すべきである^[22]。

本稿は、そのために人文科学系研究者は何ができるのかを実践してみた記録ではあるものの、推論・仮説だらけになってしまった。仁井田論文で提示された疑問点を解決するどころか、余計に混乱を深めた感すらあるが、ひとまずの記録・試論として提示しておきたい。

【謝辞】本稿の執筆にあたり、広島大学図書館の西村朋子氏、東京大学総合図書館の石川一樹氏、谷口瑞枝氏、東京大学情報基盤センターの中村覚氏には貴重な情報を提供いただき、幾重にも配慮を賜った。また東京大学総合図書館および情報基盤センターの教職員の方々には、資料の閲覧・調査・撮影・公開まで格別の配慮をいただき、辛抱強く研究の進展を見守っていただいた。心より御礼申し上げます。

なお、冊子目録とデジタルアーカイブのメタデータは、徳永洋介（富山大学）、中村正人（金沢大学）、矢野正隆（東京大学）、小林晃（熊本大学）、高橋亨（東北大学）各氏との共同での成果によるものである。

また、本稿は、「はじめに」で言及した各科研費のほか、JSPS 科研費 15K12938, 16H03351 およびサントリー文化財団研究助成の研究成果の一部でもある。

注

- [1] 本稿の地の文では「条法」を用いるが、それ以外は引用元、参照元の原表記に従って、「條法」と「条法」を使い分けるものとする。
- [2] 公開のポリシーやシステムの詳細については、中村覚「東京大学デジタルアーカイブズ構築事業：学術資産の利活用に向けた取り組み」（本誌所収）を参照。
- [3] 以下、法制史料としての位置づけに関する部分の記述は、王毓銓「《皇明条法事類纂》読後」『明史研究論叢』1, 1982年、佐藤邦憲「明律・明令と大詔および問刑条例」滋賀秀三編『中国法制史：基本資料の研究』東京大学出版会, 1993年、滋賀秀三「法典編纂の歴史」『中国法制史論集：法典と刑罰』創文社, 2003年、徳永洋介「解題」『皇明条法事類纂条名目録』[富山大学人文学部], 2018年によった。
- [4] 条例数に関しては研究者によって見解の相違がある。本稿では、今回のデジタルアーカイブのメタデータ数に基づいている。可視化のグラフは東京大学情報基盤センターの中村覚氏によるものである。
- [5] 梅原郁編『慶元條法事類語彙集覽』京都大学人文科学研究所, 1990年の「はじめに」参照。
- [6] 中村正人「『皇明条法事類纂』講読雑感」『東洋法制史研究会通信』29, 2016年 < http://www.terada.law.kyoto-u.ac.jp/tohoken/29_nkmr.htm > (2018年9月17日確認)
- [7] 仁井田陸「舊鈔本『皇明條法事類纂』私見」『補訂中国法制史研究：法と慣習・法と道徳』東京大学出版会, 1980年（初出は1940年）。以下、本稿で仁井田氏の諸説に言及する場合はこの論考を指す。
- [8] 条文レベルでの詳細な検討は、張建民「影鈔本《皇明条法事類纂》未分卷文書初考：《皇明条法事類纂》考校之一」『人文論叢』2001や劉篤才「破解《皇明条法事類纂》之謎」『中外法史研究』2017-5などがある。張氏は影印本に、劉氏は標点本に基づいて分析している点異なるが、各巻の巻頭目録の条文名と本文見出しの条文名の対照が主眼となっている点は同じである。
- [9] 具体的には不分巻14冊目の「戸部爲祛夙弊清鹽法以備國儲以足民用事」。標点本は一冊全体を一条として戸部類にあったものと考えている。冊子目録では附録として巻50工部類の後に排している。なお、前掲注3王論文や注8張論文は、途中で分割し二条と判断する。
- [10] 戴金の詳細な伝記は、李劍雄「戴金事迹小述」『史林』1987-3があり、これに基づき呉啓琳「《皇明条法事類纂》成書年代及作者考」『歴史档案』2010-3では、戴金の官歴から法律や官文書に精通していると考え、戴金が編纂者である蓋然性が高いことを論じている。
- [11] 魏隱儒・王金雨著；波多野太郎・矢嶋美都子訳『漢籍版本のてびき』東方書店, 1987年, 180頁。

- [12] 顕微鏡による繊維観察や繊維鑑別方法については、ひとまず小島浩之・矢野正隆「漢字・字喃經典料紙調査概要：東南アジア地域文献の史料論的研究序説」『東京大学経済学部資料室年報』8, 2018 年を参照。
- [13] 多くの手漉き紙には簀と簀を編んだ糸の痕跡が残り、前者を簀目、後者を糸目という。これらは紙の種類や地域、時代により特徴があり、紙を鑑別するための一指標となる。楮繊維は太く長いため簀も太く粗いものでよく、竹繊維は細いため細く密な簀が好まれるのが一般的な傾向である。
- [14] 副葉は、表紙と本文紙の間に置かれている紙葉で、表紙に貼り付いているものを指す。『皇明條法事類纂』では、表紙と本文紙の間に二葉あり、一葉分は表紙に貼られ、もう一葉分は独立した白紙葉となるものが多い。後者に関して中国文献学に適当な言葉が見当たらないため、ひとまず「遊び紙」としておきたい。これは、洋装本において表紙と本文紙の間にある紙で、表紙と一体となっている方を見返し紙、独立しているものを遊び紙とする言い方に倣ったものである。なお、副葉は実際には茶色に染めた竹紙と白紙を半葉ずつ繋ぎ一葉としている。
- [15] 中国における紙の品質と使い分けに関しては、小島浩之「中国における記録媒体の変遷再考：文書料紙を中心として」『東アジア古文学の構築：現状と課題』東京大学経済学部資料室, 2018 年を参照。
- [16] 東京大学総合図書館石川一樹氏のご教示による
- [17] 石川一樹氏によれば、「貳一三八八」のうち、「貳」の部分は「門(部)」、「一三八八」の部分は「号」と呼ばれ、書庫内書架図書分類を表すもので、関東大震災前の図書館案内によれば、第貳門は法律・政治・経済・社会・統計を指すとのことである。「号」の部分の数字は、図書館における分類付与の常識から考えて、貳門の中での受入順の番号なのであろう。
- [18] このほかに東京帝国大学の蔵書印(朱印)と受入印(墨印)が捺されている。
- [19] 陳鱣については、陳鴻森「清儒陳鱣年譜」『中央研究院歴史語言研究所集刊』62(1), 1993 年、同「陳鱣事蹟辨正」『伝統中国研究集刊』1, 2006 年のほか、「清代蔵書家陳鱣資源庫」<<http://chenzhan.hnlib.com/Default.aspx>> (2018 年 9 月 20 日確認)を参照した。
- [20] 孫星衍の署名は、台湾故宮博物院所蔵「清孫星衍行書七言聯」<http://painting.npm.gov.tw/Painting_Page.aspx?dep=P&PaintingId=31187>、北京市東城区図書館所蔵「清孫星衍篆書橫幅」<http://www.bjdclib.com/dclibpic/examJC/KJZL/201110/t20111024_117991.html> (いずれも 2018 年 9 月 23 日確認)などと比較した。なお、仁井田陞氏の録文【220 頁】も「是」とし、横に(星?)と注記がある。
- [21] 広島大学図書館西村朋子氏のご教示による。
- [22] 図書館のデジタルアーカイブへの研究者の関わりについては IFLA (国際図書館連盟) が、貴重書や特殊コレクションのデジタル化のためのガイドラインで、計画策定段階から研究者が関与することへの配慮を謳っているという。安形麻理「デジタルアーカイブの動向と研究者の関わり」『漢字文献情報処理研究』17, 2017 を参照。

東京大学デジタルアーカイブズ 構築事業

学術資産の利活用に向けた取り組み

中村 覚 (なかむら さとる)

回 1. はじめに

2015年に公表された「東京大学ビジョン2020」の一つに「学術の多様性を支える基盤の強化」が掲げられ、「東京大学が保持する学術資産のアーカイブを構築し、その公開と活用を促進することで、学術の多様性を支える基盤を強化する。」という方針が示された。これを受け、2016年9月に「東京大学学術資産等アーカイブズ委員会」が設置された。本委員会は東京大学附属図書館長、東京大学総合研究博物館長、東京大学文書館長、東京大学情報基盤センター長を中心として構成され、東京大学内のMLA（博物館 Museum・図書館 Library・文書館 Archives）機関が参画している点に特徴がある。委員会では、東京大学の保持する多様な学術資産等のデジタルアーカイブ化を行い、かつ国内外に向けて広く公開し、その活用を促進するための「東京大学デジタルアーカイブズ構築事業」（以下、「本事業」）を2017年度から実施している。

本稿では、本事業における学術資産の公開と利活用支援に向けた取り組みについて述べる。2章では、本事業における学術資産の公開の側面を中心に述べる。3章では、公開された学術資産の利活用を支援するための取り組みについて述べ、4章にて具体的な活用事例について紹介する。これらの取り組みを踏まえた結論を5章で述べる。

回 2. 学術資産のデジタル化と公開支援

本事業の取り組みに関する全体像を図1に示す。東京大学内の学術資産のデジタル化と公開支援に向け、具体的には以下の3点に取り組んでいる。

- 学術資産のデジタル化の促進
- 画像等の公開支援
- ポータルシステムの構築

※ 2.1 学術資産のデジタル化の促進

東京大学内に所蔵されている学術資産のデジタル化を促進するための取り組みとして、学内公募により選定した事業を対象として、デジタル化の予算配分を行なっている。対象資産の学術的価値・社会的価値、研究教育活動での利用可能性、唯一性・希少性を考慮し、事業選定を行っている。

2017年度は計8部局の学術資産のデジタル化を実施した。これらの中には、2018年8月に公開した、明治時代の博物学者である田中芳男（1838-1916）が収集・所蔵した貴重な資料群である「田中芳男・博物学コレクション」¹⁾などが含まれる。その他の資料についても、現在公開に向けた準備を進めているところである。

◆ 2.2 画像等の公開支援

組織の人的リソース等の制約により、2.1 でデジタル化した学術資産を公開するシステムの構築・運用が困難な部局も存在する。本事業では、このような組織に対して、デジタル化した学術資産の公開を支援するシステムとして、「東京大学学術資産等アーカイブズ共用サーバ（以下、共用サーバ）」を提供している。支援が必要な部局から画像やメタデータを受け取り、公開サイトの構築までをサポートする。

本システムは2018年1月に公開を開始し、2018年8月末時点で14コレクションを公開している。これらのコレクションには、明代中期の条例（皇帝の判断に基づく法令や先例）を事案ごとに分類してまとめた貴重な資料群である「皇明條法事類纂デジタルアーカイブ」²⁾が含まれる。

◆ 2.3 ポータルシステムの構築

本事業における3点目の取り組みは、学内の各部局が公開するデジタルアーカイブや2.2で述べ

た共用サーバからメタデータを収集し、部局や各アーカイブを横断した検索を可能とするシステム「東京大学学術資産等アーカイブズポータル（以下、ポータル）」の構築である。2018年度内のシステム公開を目指している。本システムを構築することにより、東京大学内の学術資産の発見可能性を向上させるとともに、国立国会図書館や国立情報学研究所等の外部機関とのデータ連携におけるハブとして機能することを目的としている。

2017年度には、オープンソースソフトウェアの次世代OPAC「VuFind」を用いたプロトタイプシステムを構築し、閲覧者を限定した学内公開を行なっている。本プロトタイプシステムの機能検証を目的として、ダミーデータ（推進室が学内ウェブサイトから独自に抽出したデータ）を登録している。2018年2月7日時点では、16部局116,192件のアイテムが登録されている。本システムを用いることにより、例えば「情報学環・学際情報学府」「法学政治学研究科・法学部」「総合研究博物館」が各々所蔵する「郵便報知新聞」をまとめて取得することができ、対象資料の網羅的な検索が可能となっ

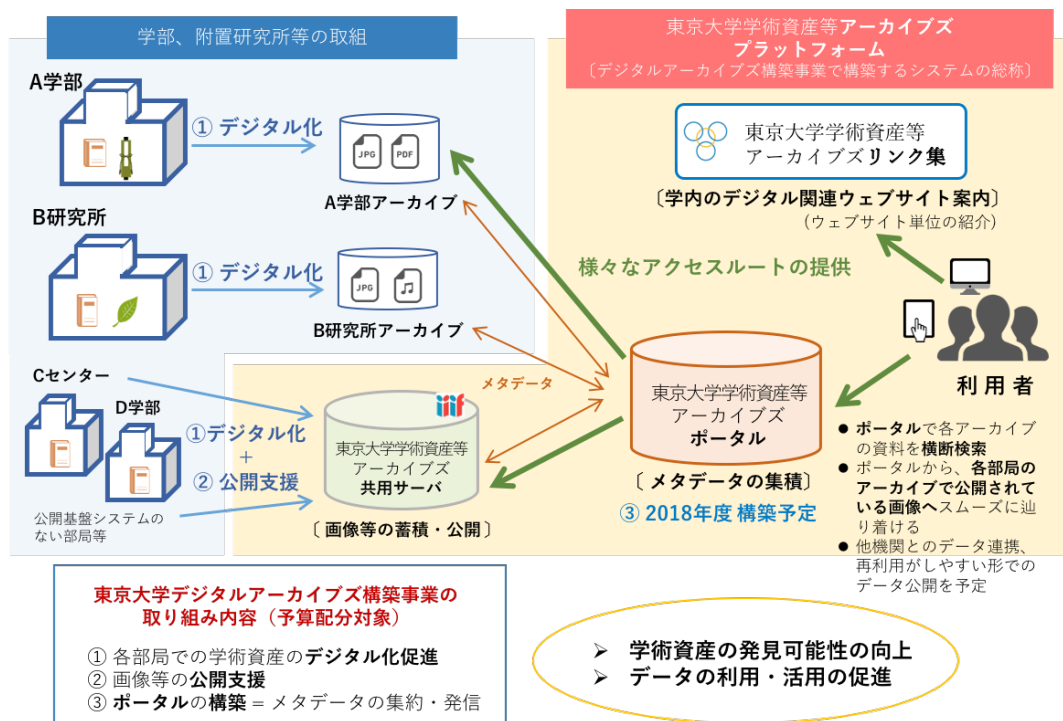


図1 東京大学デジタルアーカイブズ構築事業の全体像

た。このような事例から、異なる部局が所蔵する学術資産に対する横断検索を可能とするポータルシステムの有用性を確認した。

また、ポータルシステムの構築に先立ち、2017年12月に「東京大学学術資産等アーカイブズリンク集（以下、リンク集）」を公開した。リンク集は、既に学術資産のデジタル公開を行っている学内ウェブサイトを一覧形式で紹介するものであり、本学で所有する学術資産を俯瞰的に閲覧可能にすることを目的としたシステムである。2018年7月末時点で、41部局・186コレクションの学内ウェブサイトが登録されている。日本語ページだけでなく、英語ページも提供することで、学術資産の国内外に向けた発信を行っている。

3. 学術資産の利活用支援に向けた取り組み

2章では、本事業における学術資産のデジタル化と公開支援に向けた取り組みについて述べた。本章では、公開した学術資産の利活用支援に向けた取り組みについて述べる。具体的には、以下の4点に取り組んでいる。

- API の提供
- データのオープン化
- LOD の提供
- 国際的な標準規格の導入

これらの要素は、2.2の共用サーバに実装し提供している。図2に共用サーバの画面例を示す。以下では、これらの各取り組みについて説明する。

3.1 API の提供

1点目の取り組みは、API（Application Programming Interface）の提供である。APIとは、ソフトウェアが持つ機能の一部を外部のプログラムなどから簡単に利用できるようにする、機能の呼び出し手順などを定めた仕様のことを指す。APIを利用することにより、プログラムを効率的に開発することが可能となり、多様な利活用を促すことができる。



図2 共用サーバの画面および提供機能例

共用サーバのベースシステムとして、オープンソースのWebパブリッシングのプラットフォームであるOmeka S³⁾を採用している。Omeka SはWeb APIの一種であるREST APIを標準で提供しており、データの取得（GET）や作成（POST）を外部のプログラムから実行することができる。例えばAPIを利用することで、システム上で公開されているデータを一括取得するなど、UI（ユーザーインターフェース）上で提供されている以上の機能を利用することができる。これにより、データ提供者の意図を超えたデータの利活用を支援する。

3.2 データのオープン化

様々なデータセットを組み合わせる際などには、データの相互運用性が重要となる。この相互運用性の確保を目的として、公開データのオープン化（オープンデータ化）を積極的に進めている。オープンデータは、二次利用や商用利用が認められた形でのデータ提供を指し、第三者によるデータの利活用を実現するための最低条件であると言える^[4]。

東京大学総合図書館では、2018年6月にデジタル画像等の利用条件を変更した。具体的には、東京大学総合図書館が公開する、著作権の保護対象ではない総合図書館所蔵資料の画像データ、それに関連するメタデータ等については、クリエイティブ・コモンズ・ライセンスの「CC BY」

(クリエイティブ・コモンズ表示 4.0 国際ライセンス) 相当の条件で利用可能とした。これにより、様々な目的に応じたデータ利用を促すことができると考える。

◆ 3.3 LOD の提供

本事業ではデータのオープン化にとどまらず、LOD (Linked Open Data) の提供を行っている。LOD は Linked Data の技術に基づくオープンデータである。Linked Data はセマンティック Web 技術の応用として、データの意味論を RDF (Resource Description Framework) などを用いて記述する^[4]。データ形式が機械可読な形で標準化されることにより、様々なシステムやアプリケーションでのデータ利用が容易となる。

Omeka S は LOD を標準でサポートしている。そのため、共用サーバ上で公開されている各資料のデータを、JSON-LD や RDF/XML といった形式で取得することができる。さらに、本稿執筆時点においては、筆者個人による試験的な取り組みであるが、RDF クエリ言語である SPARQL を用いたデータ検索が可能な SPARQL Endpoint を提供している。これを利用することにより、利用者は目的に応じた任意のデータを検索・取得し、アプリケーション開発などにデータを活用することができる。

なお、この SPARQL Endpoint は、クラウド上の RDF トリプルストアのサービス Dydra^[5] を用いて提供している。先述した REST API を利用し、共用サーバ上で公開されている RDF データを一括取得し、当該 RDF トリプルストアに登録している。

◆ 3.4 国際的な標準規格の導入

Linked Data に加え、国際的な標準規格、デファクト標準を積極的に採用している。国際的な標準規格を採用することにより、国内外における意味共有、データ利用を支援する。

画像については、画像の効果的・効率的な共有を行うための規格である IIF (International Image Interoperability Framework) を用いた画像公開を行っている。IIF の利点として、画像へのアクセスが標

準化され、多様な利活用を促す点が挙げられる。例えば、北本^[6]は、IIF のキュレーションツールである IIF Curation Viewer を開発している。本ソフトウェアを使用することにより、利用者は各々の関心に合わせて IIF 画像を収集し、切り抜き等を行うことができる。この例に見るように、利用者は IIF が定める API に準拠したアプリケーションやソフトウェアを使用することにより、IIF 画像の提供機関を問わず、共通の仕組みで画像を取り扱うことが可能となる。

また、全文テキストデータについては、人文学資料のマークアップのためのルールを定めているガイドライン TEI (Text Encoding Initiative) に準拠したテキストデータを公開している。この公開については、TEI データの保存・公開・共有のためのリポジトリサービスである TAPAS プロジェクト^[7]を利用している。さらに、メタデータの記述については、Dublin Core や bibo (Bibliographic Ontology) などの既存語彙の利用を基本として、データ単体での情報流通を見据えたデータ作成を行っている。

◆ 3.5 まとめ

以上、本事業における学術資産の利活用支援に向けた取り組みについて述べた。ここで述べた要素はいずれも新しいものではないが、これらの要素を満たしたシステム、デジタルアーカイブは、特に日本国内においては、比較的希少であると言える。これらの基本的な要素に対応することが、データ利用者の多様なニーズに応えるために、データ提供者が取り得る一つの方法であると考えられる。

回 4. 学術資産の活用事例

本章では、筆者個人による取り組みとして、2章で紹介した学術資産の活用例について紹介する。これにより、3章で提示した要素の有用性について検証することを目的とする。

❖ 4.1 IIIF を用いた複数機関の所蔵資料の比較

4.1.1 課題設定

人文学研究においては、各地に所蔵される資料画像を拡大縮小しながら対比・閲覧できる環境が求められる。例えば、山田^[8]は「百鬼夜行図」を対象として、巻物によって異なる場合がある妖怪の配列の分析を通じて書写の系譜を明らかにする研究を行なっている。ここでは、IIIF を活用し、このような複数機関の所蔵資料の比較を支援する環境構築を目的とする。

4.1.2 アプローチ

具体的には、IIIF 画像を公開する各機関から、各資料を表示するための情報が格納された IIIF Manifest を収集する。これらを、IIIF 対応ビューアの一つである Mirador に読み込ませることで、同一ビューア上に複数機関が公開する画像を表示することができる。

例えば「百鬼夜行図」の IIIF 画像は、国立国会図書館^{[9][10]}、国文学研究資料館^[11] および東京大学総合図書館^[12] から公開されている。これらを同一ビューア上で表示した例を図3に示す。異なる機関で公開されている各絵巻に描かれた妖怪の細部を比較することができる。

4.1.3 まとめ

先述したように、IIIF が定める API に準拠したアプリケーションやソフトウェアを使用することにより、IIIF 画像の提供機関を問わず、多様な利活用が可能となる。本事例においても、IIIF 対応ビューアの一機能を用いることで、画像の対比と細部の閲覧が可能となる環境を構築することができた。これは、データの利活用支援に向け、データ提供者が IIIF のような国際的な標準規格に準拠したデータを提供する利点を示す一例である。

❖ 4.2 LOD を用いたアプリケーションの構築

4.2.1 課題設定

本事業では、2018年8月末時点において、共用サーバ上で公開する複数コレクションのコンテンツを横断して検索可能なシステムを提供していない。これは、2章で述べたポータルシステムが

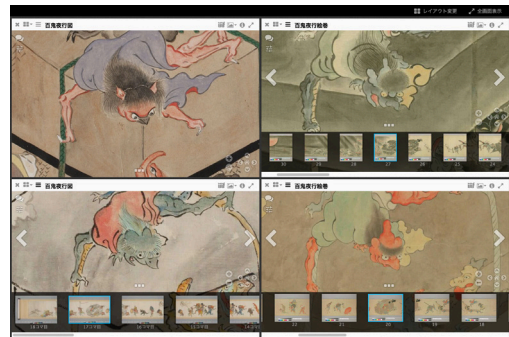


図3 IIIF を用いた「百鬼夜行図」の比較例

担うべき機能として位置付けているためである。しかしこの結果として、本事業が公開する IIIF 対応コンテンツの一覧を把握することができず、その発見の困難性が課題の一つである。

4.2.2 アプローチ

この課題に対して、共用サーバ上で公開されているデータを任意に検索し取得可能な SPARQL Endpoint を利用し、IIIF 画像の一覧を表示するアプリケーションを構築する。本事例では、SPARQL Endpoint に対して、各資料のタイトルと IIIF Manifest URI の一覧を取得する SPARQL クエリを発行する。その結果を、IIIF で個別資料に対応する Manifest をグループ化する IIIF Collection の形式で再構成する。この IIIF Collection を、例えば Mirador で読み込むことにより、図4に示すように、共用サーバで公開されている IIIF 対応コンテンツの一覧を閲覧可能なアプリケーションを構築することができる。

4.2.3 まとめ

本事例では、利用者が目的に応じたアプリケー

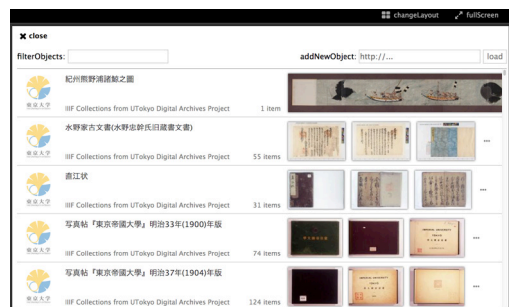


図4 IIIF 対応コンテンツの一覧表示例

ションを構築する例について述べた。本事例で対象とした課題のように、データ提供者が提供できる機能には限界があり、利用者の多様なニーズに応えるサービスを提供することは容易ではない。したがって、データ提供者は第三者が利活用しやすいフォーマットでデータ提供を行い、さらにそれらを利用しやすい環境を提供することが求められると考える。これらを実現するための一手段として、本事業では、前者については IIIF や TEI などの国際規格に準拠したデータ提供を行い、後者については LOD や SPARQL Endpoint、API の提供を行っている。

回 5. 結論

本稿では、東京大学デジタルアーカイブズ構築事業における学術資産の公開と活用に向けた取り組みについて述べた。

学術資産の公開に関する取り組みとしては、「学術資産のデジタル化の促進」「画像等の公開支援」「ポータルシステムの構築」の3点について紹介した。

また単なる公開にとどまらず、公開した学術資産、データの利活用を支援する取り組みとして、東京大学学術資産等アーカイブズ共用サーバに実装した要素を中心に述べた。具体的には、「API の提供」「データのオープン化」「LOD の提供」「国際的な標準規格の導入」を要素として挙げ、それぞれの目的や仕組みについて説明した。

さらに、これらが学術資産の利活用に資するかを検証することを目的として、二つの活用事例について述べた。IIIF 規格を活用した画像の対比・閲覧環境の構築や、LOD を用いたアプリケーション開発を通じ、各取り組みが学術資産の利活用支

援に機能しうることを確認した。

本稿で述べた取り組みが、学術情報、デジタルアーカイブの発信者と受信者の対話における材料の一つになれば幸いである。

注

- [1] 田中芳男・博物学コレクション, <https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/tanaka/>, (2018.08.31).
- [2] 皇明條法事類纂デジタルアーカイブ, <https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/koumin/>, (2018.08.31).
- [3] Omeka S, <https://omeka.org/s/>, (2018.08.31).
- [4] 大向一輝, オープンデータと Linked Open Data, 情報処理, Vol.54, No.12, pp. 1204-1210, 2013.
- [5] Dydra, <https://dydra.com/>, (2018.08.31).
- [6] 北本朝展, IIIF Curation Viewer の開発と利用, IIIF Japan シンポジウム〜デジタルアーカイブにおける画像公開の新潮流〜, 2017. <http://agora.ex.nii.ac.jp/~kitamoto/research/publications/iiif17b-ppt.pdf>, (2018.08.31).
- [7] Julia Flanders, Scott Hamlin, TAPAS: Building a TEI Publishing and Repository Service, Journal of the Text Encoding Initiative, Issue 5, 2013. DOI: 10.4000/jtei.788
- [8] 山田奨治, 「百鬼夜行絵巻」編集の系譜: 情報学からの解明, 日本研究, 国際日本文化研究センター, pp. 103-128, 2009.
- [9] 国立国会図書館, 百鬼夜行絵巻, <http://dl.ndl.go.jp/info:ndljp/pid/2540972>, (2018.08.31).
- [10] 国立国会図書館, 百鬼夜行絵巻, <http://dl.ndl.go.jp/info:ndljp/pid/2541003>, (2018.08.31).
- [11] 国文学研究資料館, 百鬼夜行図, <http://doi.org/10.20730/200016403>, (2018.08.31).
- [12] 東京大学総合図書館, 百鬼夜行図, <https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/hyakki/document/fbd0479b-dbb4-4caa-95b8-f27e1c423e4b>, (2018.08.31).

【研究ノート】韓国における仏教学 とデジタル・ヒューマニティーズ

“Digital Humanities and Buddhism – Focusing on Data Mining and Visualization” 参加記

師 茂樹 (もろ しげき)

回 はじめに

日本の仏教学は、他の人文学諸領域と比べて、比較的早くからコンピュータ利用が盛んであったが^①、近年、SAT 大正新脩大蔵経テキストデータベース^②を中心にデジタル・ヒューマニティーズへの取り組みが活発化している。デジタル・ヒューマニティーズの国際学会連合 Alliance of Digital Humanities Organizations (ADHO) に加盟する日本デジタル・ヒューマニティーズ学会の下田正弘会長が、SAT を開発する大蔵経テキストデータベース委員会の代表委員であるとともに、日本印度学仏教学会の理事長であることは (本稿執筆中の 2018 年 9 月時点)、日本の仏教学界とデジタル・ヒューマニティーズとの強い結びつきを示している。

韓国の仏教学界においても、『高麗大蔵経』や『韓国仏教全書』のデジタル化など、仏典のデジタル化は盛んに行われていた。しかし、上記のような国際的なデジタル・ヒューマニティーズの動向のなかにはまだ加わっていない。

そのような状況のなか、2018 年 6 月 2 日、韓国・ソウルの東国大学校において、仏教学とデジタル・ヒューマニティーズに関する国際会議「Digital

Humanities and Buddhism – Focusing on Data Mining and Visualization デジタル・ヒューマニティーズと仏教：データマイニングと視覚化を中心に」が開催された。登壇者は以下の通りである。

- 下田 正弘 (東京大学) 「The Significance of Buddhist Studies in the Diversity of Digital Humanities. (デジタル・ヒューマニティーズの多様化における仏教学の重要性)」(基調講演)
- ドナルド・スタージョン (ハーバード大学) 「Optical Character Recognition of Chinese Buddhist texts. (中国仏教文献の OCR)」
- マークス・ビンゲンハイマー (テンブル大学) 「Who was “Central” in the History of Chinese Buddhism? - A Social Network Approach. (誰が中国仏教史における「中心」人物なのか? ソーシャルネットワークによるアプローチ)」
- 朴普藍^{パクボラム} (東国大学校) 「Authorship Attribution in Huayan Texts by Machine Learning using N-gram and SVM. (N グラムとサポートベクターマシンを用いた機械学習による華嚴文献の著者推定)」
- 師茂樹 (花園大学) 「Remarks on the 3DCG Visualization of Lost Buddhist Temples. (仏教寺院の 3 次元 CG 復元について)」

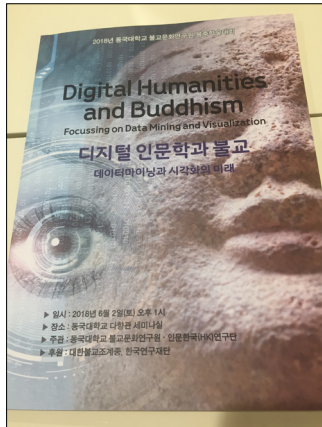


図1 「デジタル・ヒューマニティーズと仏教」予稿集

このプログラムは、下田正弘氏の基調講演をはじめ、上記のデジタル・ヒューマニティーズの国際的な動向に対する韓国側の関心を表しているものと思われる。以下、各発表の概要について、簡単に報告をしたい。

回 デジタル・ヒューマニティーズの多様性に向けて

近年、様々な分野で「多様性」あるいは「ダイバーシティ」という言葉を耳にするようになった。哲学研究の世界でも、哲学＝西洋（欧米）哲学という状況に対して、中国哲学や仏教、イスラームなど、非西洋の哲学的伝統を含めて「哲学」と呼ぶべきだ、という議論がなされている¹⁴。

下田正弘氏の基調講演「デジタル・ヒューマニティーズの多様化における仏教学の重要性」は、そのタイトルの通り、人文学におけるデジタル・ヒューマニティーズの重要性と、デジタル・ヒューマニティーズが多様化するために仏教学がデジタル・ヒューマニティーズに取り組むことの重要性について論じたものである。

講演では、まず下田氏がこれまで繰り返し論じてきた¹⁴ 人文学におけるコンピュータ利用の方法論的重要性について再確認された。それとともに、欧米が中心のデジタル・ヒューマニティーズの世界で、SAT を中心とした日本の仏教学の取り



図2 下田正弘氏の基調講演

組みが存在感を増していることを踏まえ、東洋学・アジア学等の非欧米の人文学においてデジタル・ヒューマニティーズを推進する意義についても論じられた。人文学やデジタル・ヒューマニティーズの多様性（ダイバーシティ）のためにも、仏教学におけるコンピュータ利用の推進が必要だ、というのである。

下田氏の主張は首肯し得る点が多いものの、多様性という観点では議論すべき点が残されているように思う。ADHO はデジタル・ヒューマニティーズの盛んなヨーロッパ、アメリカ、カナダ、南アフリカ、台湾、そして日本の団体で構成されており、TEI や IIF をはじめとする文献デジタル化の標準化やオープンサイエンスなどの動きと連動している面がある。一方で、それとは異なる価値観のもと、独自の方法で大規模な文献デジタル化とその商用化を進める中国は、このような「国際」的な流れとは一線を画しているように見える。「国際」的に推進されているオープン化の流れに、中国も参加するよう促す、というのも一つの考え方もかもしれないが、このような文献デジタル化の「文化」の違いを、(デジタル) 人文学の多様性（ダイバーシティ）をめぐる議論のなかでどのように考えるかについては、今後さらなる議論が必要なのではないだろうか。

回 漢字文献デジタル化の現場

ドナルド・スタージョン氏は、Chinese Text



図3 ドナルド・スタージョン氏(奥)と通訳の朴普藍氏(手前)

Project (<http://ctext.org>) を開発、公開し、デジタル中国学 (Digital Sinology) を提唱するなど、中国学領域でのデジタル・ヒューマニティーズを先導する研究者の一人として知られている。

講演の内容は、Chinese Text Project の概要と、版本の OCR 処理における技術的な問題点が主であった。また、この発表は「中国仏教文献の OCR」と題されているが、Chinese Text Project には幅広い前近代中国文献 (一部近代以降の文献も含む) が収録されており、実際には仏教文献に限った話ではない。

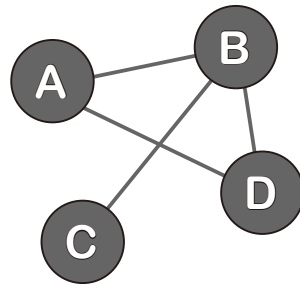
国際会議は、登壇して発表することだけがすべてではない。デジタル・ヒューマニティーズの世界における漢文文献デジタル化のキーパーソンを招聘し交流したことは、今後の韓国におけるデジタル・ヒューマニティーズの進展において、少なからぬ意義があったと思われる。

㊦ 歴史上の人物のソーシャルネットワーク分析

マークス・ビンゲンハイマー氏の「誰が中国仏教史における「中心」人物なのか？ ソーシャルネットワークによるアプローチ」は、中国の高僧伝 (高僧の伝記集) を TEI によってマークアップすることで人物間のネットワークをデジタル化し、ネットワーク分析をしたものである¹⁵⁾。

ネットワーク分析とは、下図のように、何かと何かの関係をノード (点) とエッジ (辺、リンク)

からなるネットワークとして表現することで、グラフ理論などの数学的手法によって分析できるようにしたものである。人物間ネットワークの場合には、ノードが一人一人の人物となり、エッジがある場合にはつながったノード (人物) どうしに何らかの関係 (友人である、同じ職場にいる、家族である etc.) があることを示す。



ネットワークのなかでどれが重要なノードなのか (人物間ネットワークであれば、誰が重要人物なのか) を示すのが中心性 (centrality) である。中心性にはいくつか種類があるが、ビンゲンハイマー氏はそのなかの次数中心性 (degree centrality) と媒介中心性 (betweenness centrality) の分析結果について報告した。

次数中心性とは、もっともエッジ (リンク) がつながっているノードのことである (上の図では B)。中国の高僧伝に登場する人物のなかで、次数中心性の上位 5 人は以下の通りである (発表では 20 位まで示された)。

1. 楊堅 (隋文帝)
2. 玄奘
3. 楊広 (隋煬帝)
4. 蕭子良
5. 勝荘

皇帝や玄奘のような有名人が上位を占めるなか、興味深いのは勝荘の存在である。勝荘は 7 世紀に活躍した新羅の学僧であるが、「たしかな伝記がない」¹⁶⁾ と評されるなど、研究史的にはマイナーな存在であった。その勝荘が玄奘などの有名人と並んでこのように上位に来たことは、東アジア仏教研究者の「常識」からすれば意外なことであり、

コンピュータを使った新たな知識発見の可能性を示すものとして興味深い結果であった（加えて、勝荘が新羅僧ということも、韓国の研究者には強いインパクトを与えたようである）。

次の媒介中心性とは、二つのノード間の最短経路で、よく通過されるノードほど中心性が高いとする考え方であり、中心性のなかでよく用いられる方法である。ビンゲンハイマー氏によれば、中国の高僧伝に登場する人物のなかで、媒介中心性の上位5人は以下の通りである。

1. 馬祖道一
2. 法眼文益
3. 楊広（隋煬帝）
4. 南嶽懷讓
5. 慧安

ここでは、楊広（隋煬帝）以外はすべて禅僧であり、次数中心性の結果とは大きく変わっている。この結果の違いから、ビンゲンハイマー氏は中国仏教史における変化——皇帝などをパトロンとする教団のあり方から、師資相承を重視する教団のあり方へ——を読み取ろうとしている。これについては様々な批判があるろうが、中国仏教史への新たな視点、あるいは問題提起をしているという点では有益であると思われる。

回 N グラム分析の先へ

本誌前号で筆者は、2017年に韓国で開催された元暎に関する国際会議において、N グラムを用いた文献分析が韓国の研究者に大きなインパクトを与えたことを紹介した⁷⁾。朴普藍氏の「N グラムとサポートベクターマシンを用いた機械学習による華厳文献の著者推定」は、それを受けての発表である。N グラム分析については、クラスター分析など、これまでいくつか分析手法が提案されているが⁸⁾、このような新たな分析手法の開発によって、より分析が多角化するなど、研究上の意義は大きいと思われる。

サポートベクターマシン (SVM) とは機械学習



図4 マークス・ビンゲンハイマー氏（奥）と通訳の朴普藍氏（手前）

において用いられるパターン認識手法の一つで、認識性能の高さから広く用いられているものである。サポートベクターマシンは、何かの集合を二つのグループに分ける場合に用いられる。朴氏のテーマでいえば、ある仏教文献の集合があったとして、各文献の特徴を何らかの方法で数値化したうえで、著者Aの文献のグループとそれ以外の文献に分ける、といった場合にサポートベクターマシンが使える。著者不明の文献が、著者A・B・C…の文献と同じグループに入るか、入らないかを分析することによって、著者推定をすることが可能になるのではないか、という見込みである。

今回は、サポートベクターマシンを用いた著者推定の試作段階での発表であったことから、今後の進展が期待される。現在、人工知能ブームということもあって、機械学習のツールが数多く公開され、様々な手法を試すことが以前より容易にな



図5 朴普藍氏の発表

りつつある。今後、このような研究がますます発展していくのではないと思われる⁹⁾。

回 文化遺産 CG 復元の問題点

最後に、筆者の発表「仏教寺院の3次元CG復元について」である。この発表だけ、文献研究が中心の（ADHO的な）デジタル・ヒューマニティーズとは異なり、コンピュータ・グラフィックス（CG）を用いた文化遺産の復元がテーマである。発表では、筆者がこれまで行ってきた文化遺産の3次元CG復元について紹介したあと¹⁰⁾、3次元CG復元の1) 技術的な問題と2) 方法論的な問題について検討した。

1) 技術的な問題としては、3次元CG技術の限界（マシンパワーなど）や、3次元CGが技術・デバイスの発展とともに陳腐化してしまうことの問題などを指摘した。現在さかんにVRコンテンツが作られているが、その多くは近い将来、デバイスの世代交代などによって閲覧不可能になることが予想される。一過性のコンテンツとして消費するあり方もあってもよいだろうが、人文学の研究の成果として3次元復元CGが作られた場合、あるいは教育での利用などを考えた場合には、長期的な保存・活用も必要であろう。

2) 方法論的な問題としては、3次元復元CGを

制作すること自体が、歴史叙述と同様の実践になっていることを指摘した。昨今の歴史修正主義の問題を見てもわかるように、どのように／どのような歴史を書くか、ということには、ナショナリズムなどのイデオロギー的な問題がしばしばからんでくる。また、世界文化遺産がしばしばナショナル・アイデンティティーや観光政策などと結びつけられて論じられるように、文化遺産をめぐる言説には政治的、経済的な議論が伴うことが多い。そして、ナショナリズム的な立場から歴史を書きたい人々、あるいは文化遺産で地域を活性化したい人々が思い描く「歴史」と、歴史学者によって学術的に見出された「歴史」は、しばしば対立することがある¹¹⁾。3次元CG復元もまた過去を画像・映像で表現しようという試みである以上、これらと同様の問題が起り得ると考えられる。

また、3次元復元CGは、歴史学的な知識を視覚化（visualization）する手法の一つであるが、仏典の『観無量寿経』が *The Sutra of Visualization of the Buddha of Infinite Life*（無限の寿命をもつブッダ〔阿弥陀仏〕の視覚化に関する経典）などと英訳されることからわかるように、visualizationは仏教における修行実践（観仏など）と深い関係がある。実際、宗教界にはVRをはじめとする3次元CG技術を用いている人々もおり、その意味でも寺院の3次元CG復元には宗教的な実践性があると言えることを指摘した。

回 おわりに

以上、雑駁ではあるが、韓国で開催されたデジタル・ヒューマニティーズの国際会議について報告した。

下田正弘氏が述べているように、デジタル・ヒューマニティーズは人文学に対して方法論的な議論を促すものであり、今回の各発表はまさに、これまでの仏教研究では見出せない視点を提供する（あるいは提供するための基盤を準備する）もの、新たな方法論的議論を喚起するものであったと思われる。韓国を含めた様々な地域の研究組織が、デジタル・ヒューマニティーズの実践と、それをめ



図6 筆者（師）の発表（香山氏撮影）

ぐる議論に参加することを期待したい。

注

- [1] 日本の仏教学におけるコンピュータ利用についての論文・記事は多数あるが、ここでは拙稿「仏典のデジタル化の現状」(『ASCL』22(11)、1998年11月)、「仏教研究におけるコンピュータ利用の現状—INBUDSを中心に」(『人文学と情報処理』24、1999年9月)、「コンピュータは電子仏典の夢を見るか」(『三色旗』761、2011年8月)、「仏教文献とデジタル人文学：日本の事例を中心に」(『한글불교 문헌의 정본화와 확장성 (韓国仏教文献の定本化と拡張性)』、2017年10月)、本誌の学術リソースレビュー等をあげておく。
- [2] <http://21dzk.l.u-tokyo.ac.jp/SAT/>
- [3] Jay L. Garfield and Bryan W. Van Norden, "If Philosophy Won't Diversify, Let's Call It What It Really Is." *New York Times*, May 11, 2016. タイトルは「もし哲学が多様性を持たないのであれば、「西洋哲学」という本当の名前で呼ぶことにしよう」といった意味。著者のガーフィールドはチベット仏教の研究者、ノーデンは中国哲学の研究者。
- [4] 下田正弘「比較思想と人文情報学」(『比較思想研究』44、2018)など。
- [5] Marcus Bingenheimer, Jenjou Hung and Simon Wiles, "Social Network Visualization from TEI Data." *Literary and Linguistic Computing* 26(3), 2011, 271-278. DOI: 10.1093/lilc/fqr020
- [6] 東国大学校仏教文化研究所編『韓国仏書解題辞典』(国書刊行会、1982)、49ページ。
- [7] 拙稿「【研究ノート】2017年仏教学国際学会参加記」(『漢字文献情報処理研究』17、2018)。
- [8] 拙稿「N グラムモデルとクラスター分析を用いた漢文古典テキストの比較研究 — 『般若心経』の異訳の比較

を例に」(『京都大学大型計算機センター第69回研究セミナー 東洋学へのコンピュータ利用』、2002)、「大規模仏教文献群に対する確率統計的分析の試み」(『漢字文化研究年報』1、2006)など。

- [9] こういった研究が進展するためには、人文系の「数学忌避」とでも言える傾向を是正していく必要があると考えられる。グラフ理論を知らずにピングエンハイマー氏の研究を批判したり、確率・統計的な知識がないのに朴普藍氏の研究を批判したりすることはあり得ないと思うが、実際には「コンピュータには文献は読めない」というような根拠に乏しい「批判」がまだ少なくないと思われる。
- [10] 筆者が取り組んできた3次元復元CGは、①平安京右京二條三坊八町にあった平安初期の貴族邸宅(2008年)、②下京梅忠町にあった江戸末期の京町家(2010年、<http://www.nicovideo.jp/watch/sm10311336>などで公開)、③安土城天主(2012年、スマートフォンアプリ「VR安土城」などで公開)、④平安京羅城門・西寺(2017年、「AR西寺・羅城門」<http://www.ar-rajomon.jp>として公開)。
- ②については、松田隆行ほか「下京・梅忠町の家屋敷景観の復元および家督相續の實証的研究」(『花園大学文学部紀要』43、2011)、師茂樹・明珍健二「京町家の3次元CG復元をめぐる問題」(『第16回公開シンポジウム「人文科学とデータベース」論文集』、2010)を参照されたい。
- [11] 厳密に言えば、歴史修正主義者による歴史叙述も、学術的な歴史叙述も、歴史を書くという実践という意味では同じである。歴史修正主義的な歴史叙述がマイノリティーなどへの暴力や抑圧となることはよく知られているが、学術的な歴史叙述であっても、時に過去の人々などに対する暴力として働く可能性は否定できない(北條勝貴「〈書く〉ことと倫理—自然の対象化/自然との一体化をめぐる—」『GYRATIVA』3、2004など)。

入門・ 中国語形態素解析

自然言語処理は、私たちにとって身近な技術の一つである。たとえば、この文章は日本語入力システムで書いているが、ローマ字入力された仮名の文字列を適切に分節し、文脈に応じた適切な変換候補を一瞬で提示することができるのは、長年の自然言語処理技術の発展の賜である。ほかにも、検索エンジンや機械翻訳サービスなど、自然言語処理技術がなければ成り立たないものは多い。

自然言語処理は、昔から人工知能研究と深い関係がある。近年、ディープラーニングなどの新技術の発展によって、人工知能の研究・開発が活気づいているが、それにともなって自然言語処理もまた発展してきている。特に、IT や情報科学の研究・開発の成長が著しい中国においては、近年数多くの論文やツールが発表されている。

本特集では、自然言語処理のベースとなる形態素解析に焦点を当てて、無料で使える中国語自然言語処理ツールの紹介とレビューを行う。また、中国語教育での応用と、古典中国語（漢文）形態素解析ツールの開発についての報告もあわせて収録している。これらの記事を通じて、現在の中国語自然言語処理が、教育や研究において十分に実用的であることがわかるのではないかと思う。

本誌はこれまで、第 2 号で N グラム分析（統計的自然言語処理の一種）の特集を組むなど、自然言語処理には関心を持ち続けてきた。統計的自然言語処理は形態素解析でも重要な技術である。本特集を機に、中国語自然言語処理への関心が少しでも高まれば幸いである。

CONTENTS

形態素解析とは何か	師 茂樹…	42
KH Coder で中国語分析	田邊 鉄…	46
形態素解析の中国語教育への応用	氷野 善寛…	51
古典中国語（漢文）の依存文法解析と直接構成素解析	安岡 孝一…	56
中国語形態素解析エンジン概観	千田 大介…	63
Windows で結巴、THLAC を使う	千田 大介…	66
macOS で HanLP を使う	師 茂樹…	72
コラム：Word で単語区切り	千田 大介…	76

形態素解析とは何か

師 茂樹 (もろ しげき)

✳ 自然言語処理の基礎技術

形態素解析とは、与えられた文を形態素単位で分割し、それぞれに品詞などの情報を付与する処理のことである。機械翻訳などの高度な自然言語処理の多くがその前段階で形態素解析を行っており、自然言語処理の基礎技術の一つとなっている。そのため、日本語や中国語を含め、数多くのソフトウェアが開発されている。

形態素解析がどのような処理なのかは、実例を見たほうがわかりやすいだろう。形態素解析ソフトウェアの MeCab^[1] を用いて「裏庭には二羽、庭には二羽鶏がいる」を形態素解析すると、以下のように出力される（見やすくするために出力結果の一部を省略している）。

裏庭	名詞, 一般, *, *
に	助詞, 格助詞, 一般, *
は	助詞, 係助詞, *, *
二	名詞, 数, *, *
羽	名詞, 接尾, 助数詞, *
,	記号, 読点, *, *
庭	名詞, 一般, *, *
に	助詞, 格助詞, 一般, *
は	助詞, 係助詞, *, *
二	名詞, 数, *, *
羽	名詞, 接尾, 助数詞, *
鶏	名詞, 一般, *, *
が	助詞, 格助詞, 一般, *, *
いる	動詞, 自立, *, *, 一段, 基本形, いる

つまり、「裏庭」「に」「は」「二」「羽」…が形態素であり、それぞれに名詞 / 一般、助詞 / 格助詞…といった品詞情報が付されている。これが形態素解析である。

最初から分かち書きをする英語などと違い、日本語ではどの文字からどの文字までが形態素なのか、という判定が単純ではなく、処理のなかで大きな比重となっている。その意味で形態素解析は、日本語の自然言語処理において特に重視されてきた技術である。

「形態素」という言葉は morpheme の訳語であり、「意味を持つ最小の言語単位」などと定義される。もともと欧米語の言語学から来ている術語でもあるため、日本語をはじめとする欧米以外の言語における morpheme が何を指すのかは一定しない。日本語形態素解析の分野では、形態素は「辞書にある項目(文字列)とそれに付与されているタグ(品詞など)の組み」^[2] とされる。

英文等を扱う自然言語処理の入門書では、「形態素」のかわりに「トークン」という言葉が用いられ、文をトークン単位に分割するトークン化と、トークンに品詞情報等を付加するタグ付けが別個の処理として説明されることがある^[3]。中国語は日本語同様、分かち書きをしない言語であるが、中国語自然言語処理でも「分詞」(単語分割)と「詞性標注」(タグ付け)に処理を分けることが多く^[4]、形態素解析(morphological analysis)として一括することは一般的ではないようである^[5]。

本特集では、日本において一般的な「形態素解析」という術語を用いる。

✦ 辞書とコーパス

文を形態素に区切る処理をする場合に用いられるのが辞書である。辞書を整備することが、適切な処理につながる。

たとえば新語の「エモい」⁹⁶という言葉を使った「あの人の発言はエモかった」という文を MeCab で解析してみると、次のようになる(執筆時点の結果)。

```
あの 連体詞 ,*,*,*
人     名詞 ,一般 ,*,*
の     助詞 ,連体化 ,*,*
発言  名詞 ,サ変接続 ,*,*
は     助詞 ,係助詞 ,*,*
エモ  名詞 ,一般 ,*,*
かつ  動詞 ,自立 ,*,*, 五段・ワ行促音便 ,
連用タ接続 ,かう
た     助動詞 ,*,*,*, 特殊・タ ,基本形 ,た
```

ここでは「エモかった」が、「エモ」「かつ」「た」に分割されているが、実際には「エモい」の活用形であるから、以下のような結果(「エモかつ」「た」)になるべきであろう。

(前半省略)

```
エモかつ 形容詞 ,自立 ,*,*, 形容詞・ア
ウオ段 ,連用タ接続 , エモい
た         助動詞 ,*,*,*, 特殊・タ ,基本
形 ,た
```

このようにならなかったのは、MeCab の辞書(ここでは標準で用いられる IPA 辞書)に新語の「エモい」という単語が登録されていないためである。辞書にない単語が処理できないわけではないが、精度が落ちてしまうので、形態素解析には辞書の充実が必要なのである。

また、辞書に語彙が登録されていたとしても、確実に分割できるわけではない。たとえば「美国人」という中国語は、「美国/人」(アメリカ/人)

のほかに、「美/国/人」(美しい国の人)のようにも分割できる。このように同じ文字列で複数の分割のしかたができる例は無数にある。このような場合、どちらの分割のほうにより妥当かを判断するために用いられるのがコーパス⁹⁷である。コーパスを用いれば、「美国/人」と「美/国/人」のどちらの頻度が高いかが統計的にわかる。また形態素情報付きのコーパスを用いれば、名詞-名詞というつながり方と、形容詞-名詞というつながり方のどちらが確率が高いのか、といった品詞間のつながりやすさも統計的に判断できる⁹⁸。辞書とともにコーパスも重要となってくるのである。

✦ 形態素解析の利用

形態素解析の結果は様々な処理に利用される。たとえば、Google などの検索エンジンで「京都の明日の天気」などと入力すると、「京都」「明日」「天気」などに分割されて AND 検索がなされるのは、形態素解析の応用の一例である。また、日本語入力の際のかな漢字変換でも、形態素解析が行われている。

文献研究にも利用可能である。文章の特徴を統計的に分析する計量文献学⁹⁹や、テキストデータ群から有用なデータを見つけ出すテキストマイニング¹⁰⁰などでは、しばしば前処理として形態素解析が行われる。

本誌では何度か N グラムを用いたテキスト分析を特集してきた¹⁰¹。ここでは、下のような文字単位での N グラムを用いていたが、

```
吾輩は猫
輩は猫で
は猫であ
猫である
である。
```

形態素解析を用いれば、下のような形態素単位での N グラム分析も可能である。

```
吾輩 は 猫
```

は 猫 で
猫 である
である。

残念ながら、形態素解析エンジンは現代語向けのものばかりで、中国古典語（漢文）を処理できるものは存在しない。古典文献の分析のためにも、安岡孝一氏のグループによる漢文形態素解析エンジンの開発には期待がかかる（本特集の安岡論文参照）。漢文形態素解析エンジンが実用化すれば、白文を読むときの手助けにすることができるなど、様々な応用が考えられる。

語学教育への応用は、本特集の氷野論文を参照されたい。

✳ 形態素解析の先へ

もっとも形態素解析だけでは、文の意味を理解したり、文を別の言語に翻訳したり、といった高度な処理には不十分である。そのためには、形態素解析の結果を踏まえ、構文解析などのさらなる処理が必要となる。

先に見たような形態素解析の結果だけでは、ある形容詞がどの名詞を修飾しているのか、というような形態素間の関係はわからない。このような関係を解析するのが構文解析である。構文解析は、中国語では「句法分析」と呼ばれる。漢文の場合、たとえばコンピュータに自動的に返り点を打たせる、といった作業をさせる場合には、構文解析が必須となる。

本特集の安岡論文で紹介されている「依存文法解析」は、構文解析の一種である。依存文法とは、二つの単語（形態素）について、依存する／依存される、という関係（係り受け構造）を記述するものである。たとえば「あの / 人」という二形態素の組み合わせであれば、「あの」は「人」に依存しているので、「あの」←「人」という具合に矢印等を使って依存関係を記述する。

構文解析の以外にも意味解析、文脈解析…と続くがここでは省略する¹¹²⁾。

注

- [1] <http://taku910.github.io/mecab/>
- [2] 山下達雄「日本語形態素解析入門 Version 0.9.1」(<http://nais.to/-yto/doc/tech/jma/jma19990514.pdf>, 1999)
- [3] たとえば、Python を用いた自然言語処理について概説した Steven Bird, Ewan Klein, Edward Loper 『入門 自然言語処理』(オライリー・ジャパン、2010) は、英文中心であるため、原書には「形態素解析」としてまとめて説明する部分がなく、日本語訳者の一人、萩原正人氏によって書き下ろされた第 12 章「Python による日本語自然言語処理」のなかで形態素解析が詳しく説明されている（この章はウェブで公開されている。<http://www.nltk.org/book-jp/ch12.html>）。
- [4] たとえば、「FudanNLP 开源中文自然语言处理工具箱」(<https://code.google.com/archive/p/fudannlp/>) など。
- [5] 中国語では morpheme の訳語として「語素」「詞素」「形素」などがあり、それぞれ意味するところが異なるようである。宗世海「“語素” 说、“词素” 说理由评析——兼论汉语语素的分类」(『暨南学报(哲学社会科学)』1997 年 04 期)、袁玲玲「现代汉语中的语素和词素不是同一概念」(『山西大同大学学报(社会科学版)』2007 年 01 期) など参照。
- [6] Wikipedia によれば「エモいは、英語の「emotional」を由来とした、「感情が動かされた状態」、「感情が高まって強く訴えかける心の動き」などを意味する日本語の形容詞」とのこと (<https://ja.wikipedia.org/wiki/エモい>、2018 年 9 月 23 日最終確認)。
- [7] 現代日本語の形態素情報付きコーパスとしては、国立国語研究所の「現代日本語書き言葉均衡コーパス」(http://pj.ninjal.ac.jp/corpus_center/bccwj/) などが、中国語では、台湾中央研究院の Academia Sinica Balanced Corpus of Modern Chinese (<http://ckip.iis.sinica.edu.tw/CKIP/engversion/20corpus.htm>) などがある。
- [8] 本特集の安岡孝一氏の論考では、McCab を用いた形態素解析において、複数の形態素の候補から最適なものを選ぶ方法が説明されている。
- [9] 計量文献学については、村上征勝『真贋の科学 計量文献学入門』(朝倉書店、1994)、村上征勝・金明哲・土山玄・上阪彩香『計量文献学の射程』(勉誠出版、2016) 等参照。
- [10] KH-Coder のサイト (<http://khcoder.net/>) で様々な事例が

紹介されている。本特集の田邊鉄氏の記事も参照。

- [11] 本誌第2号・特集「N-gram が開く世界 確率・統計的手法による新しいテキスト分析」(2001年)、第5号・特集「N-gram による漢字文献の分析」など。

- [12] 自然言語処理全般を見渡す入門書として、黒橋禎夫『自然言語処理』(放送大学教育振興会、2015)、奥野陽ほか『自然言語処理の基本と技術』(翔泳社、2016)をあげておく。

KH Coder で中国語分析

田邊 鉄 (たなべ てつ)

✿ 1. KH Coder とは

KH Coder は樋口耕一氏が開発、公開している、計量テキスト分析のためのフリーソフトウェアである。アンケートの自由記述欄や、インタビュー記録などを形態素解析にかけ、その結果を様々な量的データとして取り出すことができる。

もともと形態素解析エンジンとして ChaSen や MeCab を用い、日本語と英語、およびイタリア語・フランス語・オランダ語等、一部ヨーロッパ言語のテキストを対象としていた。Version3 ではそれらの言語で、より正確に語を抽出できるように改良が施されるとともに、対応言語にカタロニア語・スロベニア語・中国語(簡体字)・朝鮮語・ロシア語が追加された。簡体字のみではあるが、中国語の形態素解析・テキスト分析に対応、ほとんどワンタッチでデータを得ることができる。

KH Coder は Windows・Mac OS・Linux に対応しており、Windows の場合はダウンロードした zip ファイルを、書き込み権限のある場所に展開するだけで、すぐ使い始めることができる。ただし、Windows10 の S モードのように、デスクトップアプリが実行できない環境では利用できない。

Mac OS、Linux は Perl のソースコードから起動して使うことはできるが、補助ソフトを自分でインストールする必要がある、ややハードルが高い。

本稿は、Windows 版を中心にインストールから使用方法までを解説する。

✿ 2. KH Coder を使う準備

KH Coder は公式サイト^①で配布している。2018 年 9 月現在の最新版は Version 3.Alpha.14 である。Version3 系は α バージョンであるが、機能・仕様についてはほぼ固まっており、筆者も使用中にトラブルに見舞われたことはない。Version3 最大のウリの一つは多言語対応なので、特に中国語の分析に使用する場合は Version2 系の「安定版」を選ぶ理由はない。

配布パッケージをダウンロードして開き、ダイアログで解凍先を指定して Unzip をクリックする。解凍先は、書き込み可能なフォルダ(たとえばドキュメント)とする。解凍が終了すると、デスクトップに「KH Coder 3 Folder」というショートカットが作成される。これは文字通り、インストール先フォルダを開くだけのショートカットで、1 万を超えるファイル・フォルダが含まれているが、ユーザが直接操作するのはプログラム本体である kh_coder.exe のほかは、マニュアル・チュートリアルおよびサンプルくらいである。

筆者は勤務先で複数の PC にインストールしてみたが、まれに exe 化された zip ファイルを解凍する際に「このプログラムは実行できません」と出て解凍がスタートしないことがあった。この場合、いったん zip ファイルを捨て、PC を再起動後にあらためてダウンロードしたところ、問題なくインストールできた。インストール作業は単に zip をほどこいて、フォルダのショートカットをデスクトップに作っているだけだから、他の PC で解凍して、できたフォルダを丸ごと、USB メモ

リ等を使ってコピーしてもよい。

KH Coder は必須ソフトウェアとして、内部的に ChaSen・MySQL・Perl・R を使用している。Windows 版ではこれらのソフトウェアもパッケージに含まれており、設定済みなので、kh_coder.exe を起動するだけで使い始めることができる。ただし、日本語・英語以外の言語を解析する場合は、Java が必須である。インストールしていない場合は Java のサイト^[2]からインストールしておく。

Linux や MacOS 上で利用したい場合は、GitHub リポジトリ^[3]から Perl のソースをダウンロードして起動するが、必須ソフトウェアや Perl の各種モジュール、R のライブラリはあらかじめ自分でインストールし、ターミナルを使って複雑な設定作業を行う必要がある。MacOS の場合、これらの設定をほぼ自動で行ってくれる「自動設定ソフトウェア」が 3980 円で提供されている。KH Coder サポートストア^[4]で購入することができるので、設定が面倒、または、自信がない、という向きは、こちらを利用するといいたいだろう。

必須ソフトには含まれていないが、分析結果の閲覧には、Microsoft Excel など、Excel 形式・CSV 形式が扱えるソフトが必要である（もっとも、計量テキスト分析を行おうかという人のパソコンに、Excel などが一切インストールされていない、ということはちょっと考えにくいと思うが）。

得られたデータは、KH Coder 内だけでも分析できるが、共起ネットワーク分析などの多変量解析を行う際には、内部で用いている R の制限から、データが大きい場合など、処理に時間がかかりすぎることもある。SPSS など統計ソフトがあるなら、そちらにデータをエクスポートして分析を進める方がいいだろう。

いずれの OS 上でも、インストールが完了しさえすれば、中国語の分析もすぐに行える。いきなり中国語のテキストを読み込ませてもいいが、KH Coder の利用が初めて、または計量テキスト分析に慣れていないなら、付属のチュートリアル^[5]の「2. プロジェクト作成と前処理」以降を読み、それに従って操作してみることをおすすめする。夏目漱石『こころ』のテキストを材料に、形

態素解析から簡単なテキスト分析や結果の可視化までひととおり体験できる。

✳ 3. プロジェクトの作成

KH Coder は分析対象を「プロジェクト」として管理している。タグ等を使った複雑な分析をするのでなければ、一般には「プロジェクト＝対象となるテキストファイル」と考えて差し支えない。

インストールしたフォルダから kh_coder.exe を起動する。メニューの「プロジェクト」「開く」から、処理対象となるテキストファイルを開く。中国語は簡体字・utf-8 のテキストだけしか対応していないので、必要に応じて変換しておく。

KH Coder は複数ファイルの分析には対応していない。したがって、例えば「20 個あるテキストファイルの中から、共通して特徴的な語を取り出す」ような作業をする場合、ひとつずつ読み込んで分析し、結果を保存して、と、効率の悪い作業になってしまう。これを避けるため、ひとつのファイルの中に、区切り文字で区切った、複数のテキストを含めて、複数ファイルのように扱う機能がある。区切り文字には通常、<h1>～</h1> など、HTML の見出しタグで囲まれた文字列が用いられる。

「そういうテキストファイルを用意する方が面

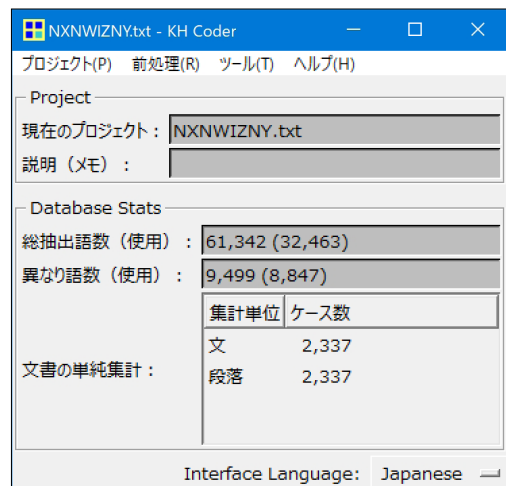


図 1 起動直後の画面

倒じゃないか」と思われるかもしれないが、KH Coder には、複数のテキストファイルを区切り文字を挟みながら結合する機能が備わっている。メニューの「ツール」「プラグイン」「データ準備」「テキストファイルの結合」を実行し、複数のテキストファイルを指定するだけで、一本につなぎ合わせたファイルを作ることができる。

テキストは何の加工もしていない生のテキストでもいいし、行単位で分けて、コードを付した CSV ファイルでもよい。CSV の場合は、表のどのカラムをデータとして読み取るか指定してやると、それ以外のカラムのデータは外部変数として扱うことができる。生のテキストだけを読み込ませておき、外部変数を後から別ファイルで与えることもできるが、別ファイルを作るのが面倒なうえ、データと外部変数との対応がわかりにくくなるので、公式サイトではあらかじめ Excel 等で整理しておくことを推奨している。

✳ 4. 前処理

KH Coder の前処理とは、形態素解析を行い、形態素を抽出することである。プロジェクトを開いた状態で、メニューの「前処理」「前処理を実行」を選ぶと、処理が開始される。

#	抽出語	品詞/活用	頻度
1	不	Adv	601
2	沈佳仪	ProperNoun	578
3	就	Adv	399
4	都	Adv	360
5	很	Adv	343
6	会	Verb	312
7	世	Adv	287
8	人	Noun	241
9	要	Verb	219
10	说	Verb	214
11	喜欢	Verb	204
12	李中华	ProperNoun	161
13	但	Adv	146
14	想	Verb	144
15	到	Verb	143
16	看	Verb	143
17	还	Adv	141
18	可以	Verb	127

図 2 抽出語リスト

中国語の分析には、形態素解析エンジンとして Stanford POS Tagger^[5] が用いられる。前処理の設定として、対象とする品詞、強制的に一語として抽出する語、抽出の対象としない語を指定できる。

6 万語あまりの小説の解析は、おおむね 45 秒前後で終わった。メニューの「ツール」「抽出語」「抽出語リスト」で、結果が確認できる。また、抽出語リストを Excel ファイルとして保存することもできる。標準で出現頻度が 100 位までの語が抽出される。

形態素解析エンジンの Stanford POS Tagger の中国語解析能力を検証するため、本特集の千田氏の比較記事で用いられている例文 (p.69 参照) を解析してみた。

- (1) 结婚 Verb/ 的 / 和尚 Noun/ 未 Adv/ 结婚 Verb/ 的 / 都 Adv/ 成了 Verb/ 和尚 Noun.
- (2) 他 / 和 / 老板 Noun/ 正在 Adv/ 量 Verb/ 和服 Noun/ 尺寸 Noun.
- (3) 最 Adv/ 巧 Adj/ 的 / 是 / 两 / 人 Noun/ 的 / 名字 Noun/ 相似 Verb/, / 骨傲天 Noun/ 本名 Noun/ 铃木 Noun/ 悟 Verb/, / 史莱姆 ProperNoun/ 本名 Noun/ 三 / 上 Noun/ 悟 Verb/, / 且 / 两 / 人 Noun/ 生前 Adv/ 都 Adv/ 是 / 一般 JJ/ 社员 Noun/.

(1) のように複数の区切りが考えられる場合でも、「和尚」が拾われているので、他の区切り方は検討されていないようだ。(3) は固有名詞をどう扱っているか見たが、モンスターの「スライム」は固有名詞として、『オーバーロード』の通称でもある「骨傲天」は普通名詞として、それぞれ拾ってはいるが、スライムを「オタク用語」として採ったのか、氏名としての形をしているから採ったのかはわからない。日本人の名前は、「鈴木」のようなよく使われる姓以外は全く拾っていない。分析対象のテキストや分析の目的によっては、使いにくいかもしれない。

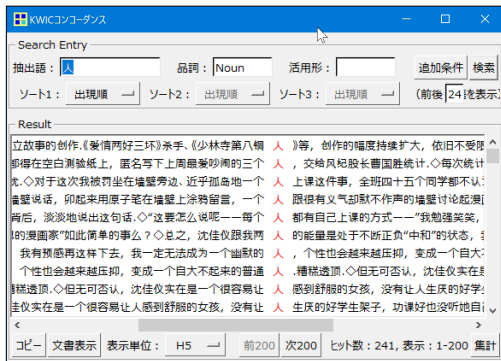


図 3 KWIC コンコーダンス

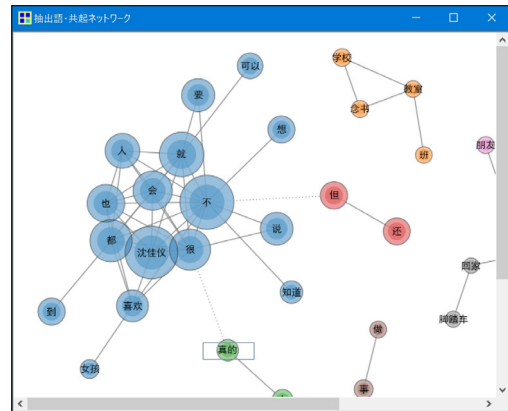


図 4 共起ネットワーク分析

✿ 5. KWIC コンコーダンス

抽出語リストの語をクリックすると、KWIC コンコーダンスを見ることができる。用例の抜き出しは、注目語から前後 24 語を抽出しているが、必要に応じて、もっと長くも短くもできる。

コンコーダンスを表示するダイアログの右下にある「集計」ボタンをクリックすると、コロケーションを集計して示してくれる。

試みに日本で映画がリメイクされた台湾の小説『那些年，我們一起追的女孩』^[6]を分析してみると、「人」という語が頻出している、「三個人」（悪友三人組）の使用が多い、などが見て取れる。

✿ 6. 語の共起を探索する

メニューから「ツール」「抽出語」「共起ネットワーク」を実行すると、よく一緒に現れる（共起する）語を、共起ネットワーク図として示してくれる。円の大きさが語の使用頻度、線の長さが共起関係の「距離」を示す。

このほか、自己組織化マップ・多次元尺度法による 3D 図・クラスター分析の結果などを、それぞれ作図することができる。

これらの分析と作図には、内部で R を使っている。R は中国語の分析に対応しているものの、システムのロケールと異なる言語を含む分析では問題が生じることがある。

✿ 7. 特徴語の抽出と対応分析

複数の文書を対象に分析している場合、各文書の特徴付ける語を、出現頻度に加えて逆文書頻度を用いて抽出することができる。

メニューの「ツール」「特徴語一覧」を実行し、文書単位の検索を指定すると、各文書の特徴語を Excel ファイルとして得ることができる。

たとえば、『那些年，我們一起追的女孩』はこの小説を書いた経緯を含んだ、メタフィクション構造をしている。では、作中に描かれる「作家」は著者と同じと言ってよいか、という目的を設定して分析してみる。特徴語抽出では、「小説」「創作」「題材」「故事」「写作」等の語が「前書き」と第 25 章に特徴的であり、「前書き」を著者によるこの小説の執筆動機の解説と見なすなら、作中の柯景騰（及びその成長した姿として描かれる作家九把刀 A）＝著者九把刀 B とみなして構わないだろう。

KH Coder には他にもタグ付けによる詳細な分析を行う機能があり、形態素解析からテキストマイニングまでほぼこれ一本でまかなうことができる。中国語の計量テキスト分析入門には、非常にありがたいソフトウェアである。ただ、中国語の形態素解析エンジンは、十分優秀であるとは言えず、繁体字に対応していないこともあって、利用範囲は限られるだろう。

※ 本稿は、日本学術振興会科学研究費補助金「初級学習者を対象としたコンテンツ駆動型中国語学習基盤の構築」(平成 30 年度、基盤研究 (B)、研究代表者: 田邊鉄、課題番号: 18H00682) による成果の一部である。

注

- [1] <http://kncoder.net>
- [2] <https://www.java.com>
- [3] <https://github.com/ko-ichi-h/kncoder>
- [4] <https://kncoder.stores.jp>
- [5] <https://nlp.stanford.edu/software/tagger.html>
- [6] 九把刀『那些年, 我們一起追的女孩』春天出版, 2011 をもとに、簡体字で入力して utf-8 で保存した。

形態素解析の中国語教育への応用

氷野 善寛 (ひの よしひろ)

＊はじめに

数年前にも似たようなことを書いたが、今一度、形態素解析の中国語教育への応用について考えてみたい。

形態素解析は、文法的な情報の記述のないテキストデータを解析し、形態素単位に分析することである。結果として、区切りのない文を単語単位で分割し提示することができる。中国語には英語のように単語を区切るスペースや日本語の助詞のような単語の区切りを示すマーカーになる単語はない。そのため、形態素解析は容易ではないが、Mecabなどの形態素解析エンジンに中国語の辞書を組み込むことで形態素解析を行うことができる。

この単語に区切った文をどう利用するかという点については場面により異なる。中国語教育の場面において、特に学習者に近いところでは、「分かち書き」した文を提示して、単語の切れ目を理解させるといった形での使用するというのが思い至るのではないだろうか。特にピンインを使う場面では、単語の切れ目で分けて提示することが必須で、こういった処理は形態素解析を利用することで実現することができる。その他では、品詞情報を付加することで、文の構造を正確にする手助けになるかもしれない。

＊中国語の分かち書き、ピンインの入力

中国語の教材を作る際に、中国語そのものを入力するだけではなく、ピンインを入力する場面に

ぶち当たる。初級の教材プリントを作る際に、ピンイン入力は必須で、ChineseWriterなど市販のIMEを利用していなければ、直接入力することが困難で、苦勞された経験がある方も多いのではないだろうか。Windowsでは、上記の市販ソフト以外は、ピンインを直接入力できるものもなく、Macはそもそもピンインを入力するソフトすら無く、声調符号付のアルファベットを入力しようと思えば工夫が必要となり、非常に面倒で手間のかかる作業となる。そのため、ピンイン変換サイトを利用し、簡体字中国語を入力した上で、ピンインに変換し、そのピンインをコピーして利用することが考えられるが、単語単位で出力されるウェブサイトはほとんどなく、多くのウェブサイトでは、単漢字単位でしか変換できなかつたり、多読字に対応していない。この種のピンイン変換サービスに形態素解析を組み込むことでこれらの問題は解決する。前後の文脈から単語を区切り、多読字のその場における正しい読み方を提示することができるのである。

そこで今回は形態素解析を組み込んだ「変換」サイトとして利用できる「Chinese Text Analyzer (以下CTA)」と「Google 翻訳」という2つウェブサイトの使用方法について、具体的な例を挙げながら紹介していきたい。

● CTA を利用したテキストの解析

中国語初級学習者を対象とした授業では、分かち書きやピンインの付加が必要となる場面がある。中国語の初級教科書では、本文は必ずと言っていいほど、単語単位で分かち書きされており、ピンインが上下のいずれかに付されている。

そこで、具体的に教科書の1課の本文を使ったプリント教材を作るということを想定して、以下のテキストを分かち書きしてみたい。

高桥，你家在哪儿？
我家在大阪。
你家有几口人？都有什么人？
我家有五口人，爸爸、妈妈、哥哥、妹妹和我。
你妹妹多大了？
我妹妹今年十三岁了。

手作業で分かち書きすると以下のようになり、こういった結果が形態素解析を通じて得られることが理想である。

高 桥 ， 你 家 在 哪 儿 ？ 我 家
我 家 在 大 阪 。
你 家 有 几 口 人 ？ 都 有
什 么 人 ？
我 家 有 五 口 人 ， 爸 爸 、
妈 妈 、 哥 哥 、 妹 妹 和 我 。
你 妹 妹 多 大 了 ？
我 妹 妹 今 年 十 三 岁 了 。

この作業を常に手作業するのは非常に面倒なので、CTA^[1]の解析にかけてみる。

- ① CTAのウェブページにアクセスし、区切りなしのテキストを入力欄に貼り付ける^[2]。CTAでは繁体字処理用や少し古い中国語の解析など用途別に複数の辞書を搭載しており、それを選択することで使い分ける



図 1：テキストデータの入力

設定になっている。ここではごくごく一般的な簡体字中国語で書かれたテキストを分析する際に利用する「標準辞書」を選択して解析を行う。

- ② テキストを貼り付けたら、「分割」ボタンをクリックする。分割結果が下に表示され、同時に「のべ語数」と「異なり語数」の集計も行われる。「異なり語数」をクリックすると解析したテキストの中で使われている単語の一覧と出現数を統計したデータが表示される。



図 2：分割結果の表示

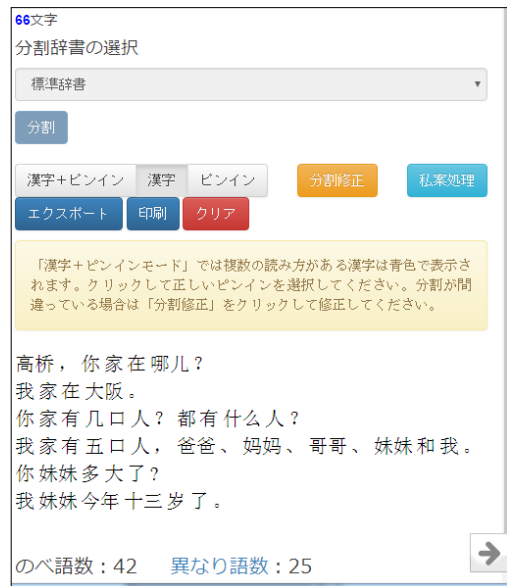


図 3：分割結果の切り替え（単語）

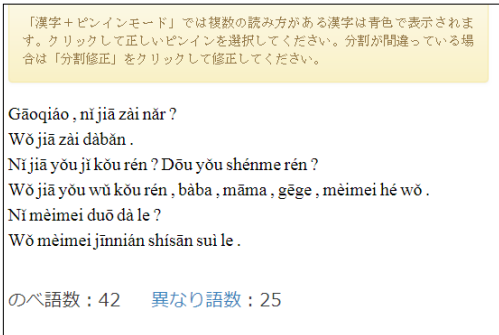


図 4：分割結果の切り替え（ピンイン）

③分割結果は「漢字とピンイン」「漢字のみ」「ピンインのみ」の3つの形式に切り替えて表示することができる。

④CTAに搭載している辞書は現在構築中のため、固有名詞を中心に解析精度が低く、間違えた分析をすることがある。そのため分割結果に間違いがある場合や微調整をしたい場合は、「分割修正」機能を利用する。「表示切替」ボタンのとりにある「分割修正」ボタンをクリックする。

開いた「分割修正」ウィンドウでは、単語の区切りの「/」をクリックすることで単語をつなげ、分割したい手前の文字をクリックすることで分割の区切りマーカーである「/」を入れることができる。分割修正をしたら「Apply」ボタンをクリックすることで結果画面に反映することができる。

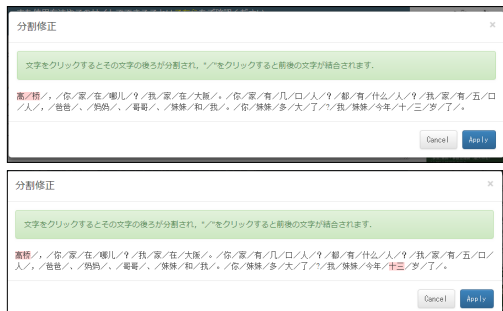


図 5：分割結果の修正

⑤多読字についても、解析しきれなかった箇

所については手作業で調整することができる。修正は、「漢字+ピンイン」モードで実施するが、システム的には、多読字の対照リストを持っており、複数の読み方がある漢字については、結果を表示する際にこの辞書からデータを照合し、該当するピンインが青色で表示される。この青字で表示されたピンインをクリックすることで、登録されている別のピンインに切り替えることができる。例えば下の例では「都」という漢字の下にあるピンインをクリックすることで、dōu と dū の2つを切り替えることができる。

都	有	什么	人	？
Dū	yóu	shénme	rén	？

都	有	什么	人	？
Dōu	yóu	shénme	rén	？

図 6：多読字の修正

⑥得られたピンインは「エクスポート」ボタンから好きな形で出力することができる。たとえば句読点で改行して、中国語・ピンインの順に表示したい場合には、「漢字×ピンイン」「文章記号で改行（。?!を指定）」としてやることで次のような結果を得ることができる。

高桥，你家在哪儿？

Gāoqiáo , nǐ jiā zài nǎr ?

我家在大阪。

Wǒ jiā zài dàbǎn .

你家有几口人？

Nǐ jiā yǒu jǐ kǒu rén ?

都有什么人？

Dū yǒu shénme rén ?

我家有五口人，爸爸、妈妈、哥哥、妹妹和我。

Wǒ jiā yǒu wǔ kǒu rén , bàba , māma ,

gēge , mèimei hé wǒ .

你妹妹多大了？

Nǐ mèimei duō dà le ?

我妹妹今年十三岁了。

Wǒ mèimei jīnnián shísān suì le .

その他に、指定した文字数や、単語単位での出力にも対応しているので、作成したい教材用途に応じて使い分けができる。

⑦分割した単語は、HSKの単語リストや中国語教育学会の「中国語初級段階学習指導ガイドライン」に収録されている単語リストなど、既存の公開されている単語リストと照合し、使用されている単語の有無や、レベルを視覚化することができる。白黒のため分かりにくい、図7では前段のテキストとHSKの6段階に分けられた単語リストと照合している。一致する単語は色が付けられ、暖色系の色ほどレベルが低い単語で、寒色系の色をほどレベルが高い単語であることを示している。このテキストの場合は、初級の会話ということもあり、LV1、LV2を中心とした赤やオレンジの暖色系の色が占めており、比較的簡易な単語だけテキストが構成されていることが一目で見て取れる。



図7：単語リストとの照合

試しに難易度が高いテキストということで、『人民日報電子版』の9月7日の記事の一部で解析を実施し、単語リストと照合をしてみた。色味とし

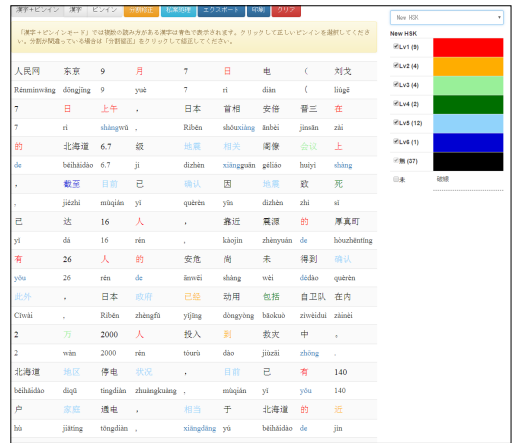


図8：『人民日報電子版』9月7日より抜粋

てはLV5、6で使われている濃い青色や薄い青色、あるいは未知語を示す黒色が多くなることが分かった。このため、このテキストの解析結果からは、記事内で使用されている単語が初級レベルの学習者にはやや難しい、あるいは未習であることが一目で分かり、授業前にそういった色分けをされた単語を中心にまずは学習者に理解をさせる必要があるということが分かる。

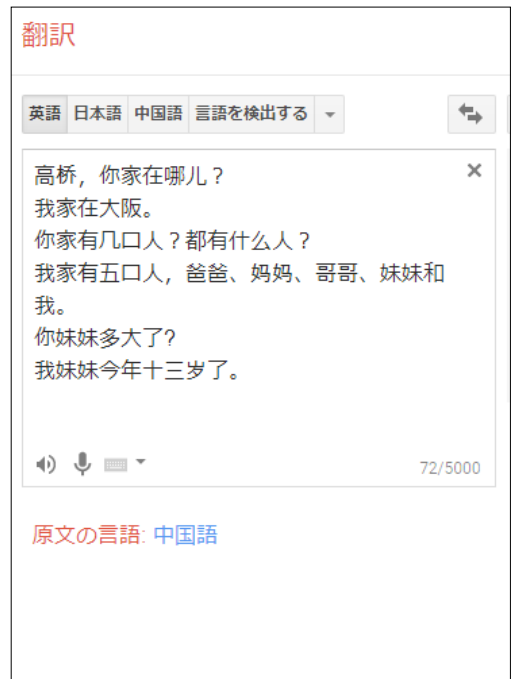


図9：Google 翻訳への貼り付け

以上 CTA の機能について簡単にふれたが、いずれの機能もテキストを形態素解析した上で、得られる結果である。

● Google 翻訳を利用した形態素解析によるピンインの獲得

さて、次に、Google 翻訳の機能を見たい。Google 翻訳では特定の言語を別の言語に翻訳するために、翻訳対象となるテキストを形態素解析している。この機能を利用することで、中国語の形態素解析を行い、その結果としてのピンインを得ることができる。

- ① 「Google 翻訳」^[3] にアクセスし、翻訳言語を「中国語(簡体)」とし、変換先の言語も「中国語(簡体)」を選択し、ピンイン変換したいテキストデータを入力する。
- ② 翻訳すると同時に、中国語が解析され、貼り付けたテキストの下に分かち書きされたピンインが表示される。

解析結果を見てみると、いくつか気になる点がある。

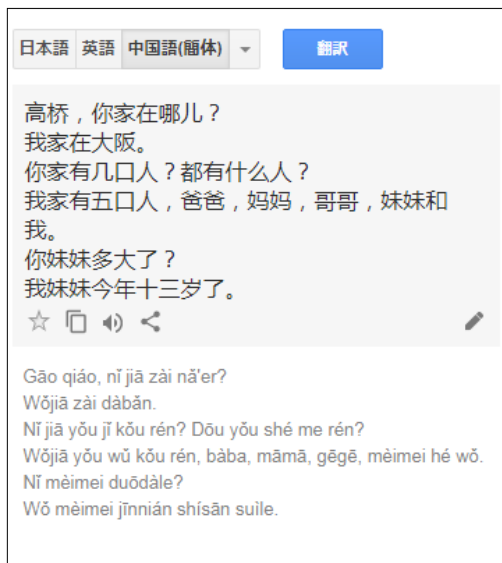


図 10：解析結果

Gāo qiáo, nǐ jiā zài nǎ'ér?

Wǒjiā zài dàbǎn.

Nǐ jiā yǒu jǐ kǒu rén? Dōu yǒu shé me rén?

Wǒjiā yǒu wǔ kǒu rén, bàba, māmā, gēgē, mèimei hé wǒ.

Nǐ mèimei duōdàle?

Wǒ mèimei jīnnián shísān suìle.

nǎ'ér の箇所は本来 nǎr となるべきところであるが、どうやらアール化に対応していないようで正確なピンイン表記となっていない。その他、分かち書きのミスとしては、Wǒjiā、duōdàle などが分割して欲しいところで、逆に Gāo qiáo はくっついて欲しいところである。このように分割の基準がピンインの分割ルールと異なるところがまだ散見される。いずれにせよ、簡単に単語の分割からピンインという処理をしてくれる点は、有効なツールであると言える。

付記：本稿は、日本学術振興会科学研究費補助金「中国語における形態素解析システムの構築と中国語基本語彙研究に関する試み」(平成 30 年度、若手研究 (B)、研究代表者：氷野善寛、課題番号：16K16839) と「小中高大連携を見据えた外国語教育と ICT の接点を探る研究ならびにアーカイブの開発」(平成 30 年度、基盤研究 (C)、研究代表者：神谷 健一、課題番号：17K02956) による成果の一部である。

注

[1] <http://www.7384uc.sakura.ne.jp/chmecab/>

[2] 1 回で解析できる最大文字数はサーバーの処理能力の関係で 1000 文字に制限している。

[3] <https://translate.google.com/>

古典中国語（漢文）の 依存文法解析と直接構成素解析

安岡 孝一（やすおか こういち）

✧ はじめに

言語処理という側面から見ると、古典中国語（漢文）の白文というのは、かなりやっかいなシロモノである。単語と単語の間に区切りがない。文と文の間にも区切りがない。漢字がズラズラと、切れ目なく並ぶだけである。こんなもの、どうやって読めばいいのか。

筆者が班長を務める京都大学人文科学研究所共同研究班「東アジア古典文献コーパスの実証研究」では、現在、古典中国語における文法解析の自動化に全力で取り組んでいる。漢文の白文に対し、形態素解析・依存文法解析・直接構成素解析を順におこなうことで、白文の統語構造が解析可能となる、というのが、われわれの見通しである。形態素解析^[1]によって、単語切りをおこなうと同時に、各単語の品詞を得る。依存文法解析^[2]によって、単語と単語の間の係り受け関係を解析すると同時に、文の切れ目を得る。直接構成素解析^[3]によって、各文の統語構造を解析木の形で得る。例として「孟子見梁惠王王曰叟不遠千里而來」という白文に対し、形態素解析・依存文法解析・直接構成素解析を順におこなう際の流れ（イメージ図）を、図1に示す。

✧ 漢文の形態素解析

古典中国語の形態素解析^[4]において、われわ

れは、MeCab という汎用の形態素解析ソフトウェアを用いることにした。MeCab は、もともとは日本語の形態素解析用だった^[1]が、言語、辞書、コーパスに依存しない汎用的な設計がなされており、辞書とコーパスを準備すればいかなる言語にも対応できる、というのが売りである。

MeCab の辞書には「品詞」（複数の階層が可能）が必要なことから、われわれは、日本語と漢文をつなぐ「構造」の一種である訓読に着目し、返り点を「品詞」に反映させることを考えた。すなわち、訓読における返り点が、漢文の動賓構造を表しているとき、動詞類に「v」という「品詞」を、賓語に「n」という「品詞」を、その他の語に「p」という「品詞」を、それぞれ、MeCab 漢文辞書の「第1階層の品詞」として定めた。次に「第2階層の品詞」を「名詞」「代名詞」「数詞」「動詞」「前置詞」「副詞」「助動詞」「助詞」「感嘆詞」の9種類とした。従来の漢文文法などで見られた「形容詞」を廃止して、「動詞」と統合している^[5]のが特徴である。さらに「第3階層の品詞」として44種類の意味素性を、「第4階層の品詞」として88種類の小素性を定義し、形態素解析の結果として得られる各単語を、意味の面からも捉えやすいよう工夫した。

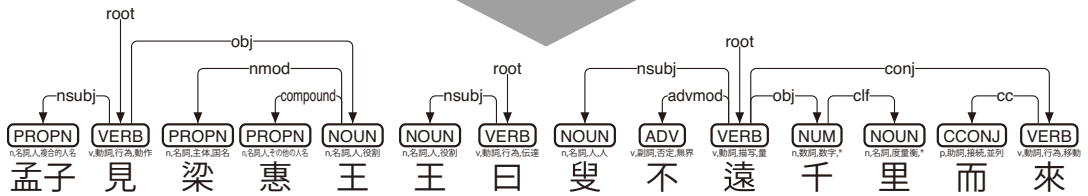
MeCab を用いた形態素解析において、その中心となるアイデアは、CRF（Conditional Random Fields）である。われわれの漢文形態素解析に即して言えば、解析したい白文を MeCab 漢文辞書に基いて、可能性のある全ての単語（4階層の品詞を含む）の組み合わせの列に変換する。例として「孟

孟子見梁惠王王曰叟不遠千里而來

形態素解析



依存文法解析



直接構成素解析

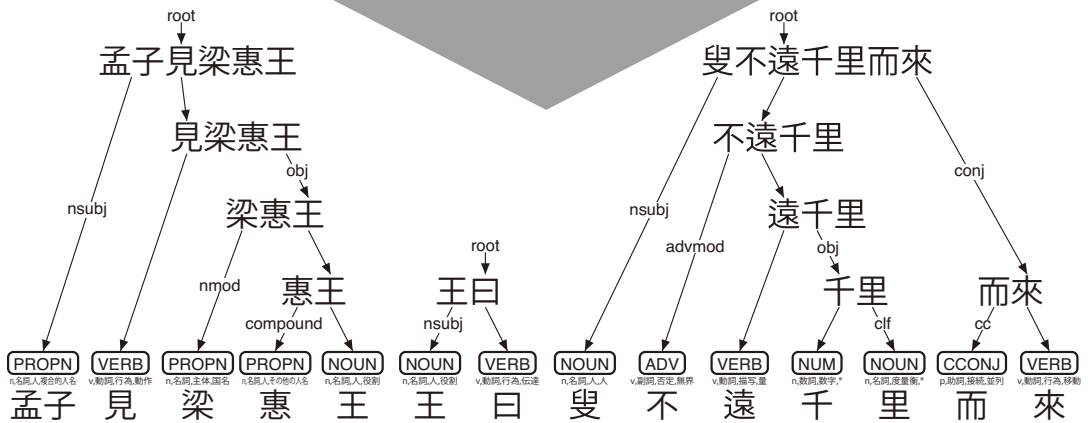


図 1: 漢文の形態素解析・依存文法解析・直接構成素解析

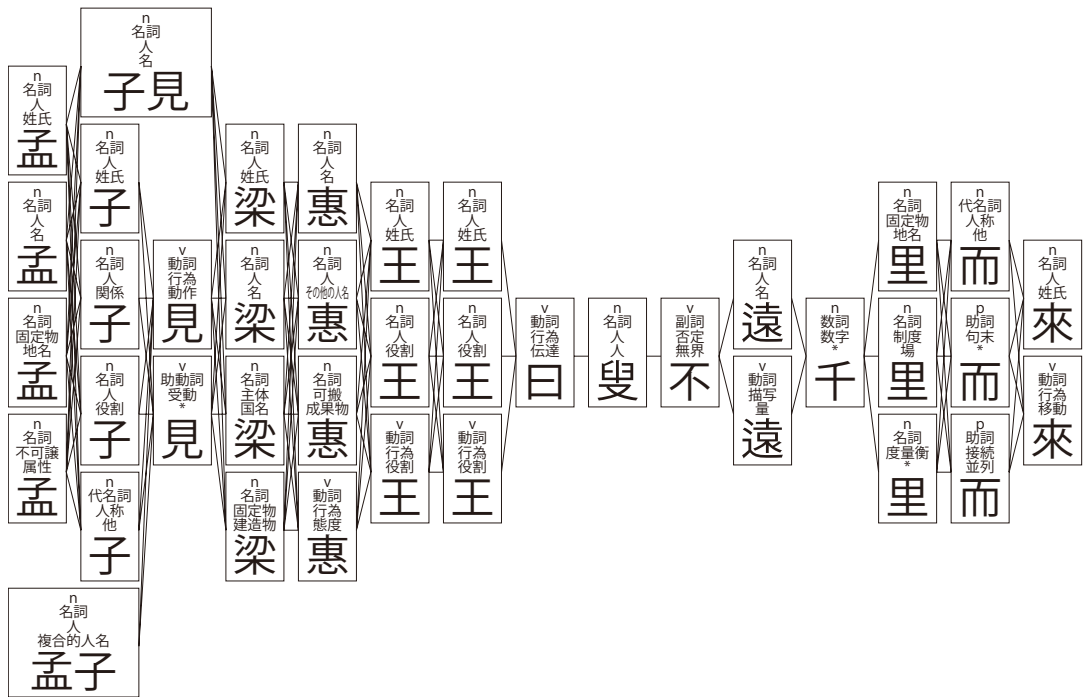


図 2: CRF を用いた漢文の形態素解析

子見梁惠王王曰叟不遠千里而來」という白文に対する CRF (の一部) を、図 2 に示す。「孟子」には「n, 名詞, 人, 複合的人名」という「品詞」が与えられており、その一方で、「孟」には 4 種類の「品詞」が、「子」には 4 種類の「品詞」が与えられうることから、これらの組み合わせが全て列挙されている。さらに「子見」に「n, 名詞, 人, 名」という「品詞」が与えられていて、「見」は「v, 動詞, 行為, 動作」と「v, 助動詞, 受動, *」の可能性もある。これらの組み合わせに対し、MeCab は、各単語の出現確率と、隣り合う単語どうしの共起確率から、全ての組み合わせの中で最も確率が高くなるような単語列を抽出する。そのようにして、図 1 の形態素解析結果が得られるわけである。

✳ 漢文の依存文法解析

古典中国語の依存文法解析^[6]において、われわれは、Universal Dependencies^[7] (以下「UD」) という、言語横断的な依存構造記述を用いることにした。UD は、品詞・形態素属性・依存構造情報を、

言語に依存せず記述する手法である。句構造を考慮せずに係り受け関係を記述できるよう、全ての構文構造を「単語」間の依存関係で記述するのが特徴である。

UD には 17 種類の品詞が準備されているが、これらのうち、われわれの古典中国語 UD では PROPN・NOUN・PRON・NUM・VERB・ADP・ADV・AUX・PART・SCONJ・CCONJ・INTJ の 12 種類を使用している^[8]。残りの 5 種類 (ADJ・DET・PUNCT・SYM・X) は、使用していない。12 種類の UD 品詞へは、漢文形態素解析の結果として得られる 4 階層の品詞を、自動変換している。

「単語」間の係り受け関係に対しては、UD 依存構造の「単語」間リンクを用いて表現し、各リンクに表 1 の UD 依存構造タグ 34 種類を付与している^[8]。タグのうち 30 種類は、もともと UD で規定されているものであり、4 種類 (nsubj:pass・csubj:pass・discourse:sp・flat:vw) は、その派生形である。root はリンク元を持たないが、他のタグによるリンクは、リンク元の「単語」とリンク先の「単語」を 1 つずつ有する。たとえば、漢文の動賓構

	Nominals	Clauses	Modifier Words	Function Words
Core arguments	nsubj 主語 → nsubj:pass [受動文] obj 目的語 iobj 間接目的語	csubj 節主語 → csubj:pass [受動文] ccomp 節目的語 xcomp 節補語		
Non-core arguments	obl 斜格補語 vocative 呼称語 expl 形式語 dislocated 外置語	advcl 連用修飾節	advmod 連用修飾語 discourse 談話要素 → discourse:sp [文助詞]	aux 動詞補助成分 cop 繫辞 (copula) mark 標識(marker)
Nominal dependents	nmod 体言による連体修飾語 nummod 数量による修飾語	acl 連体修飾節	amod 用言による連体修飾語	det 決定詞 clf 類別詞 case 格表示
Coordination	MWE	Loose	Special	Other
conj 接続 cc 接続詞	compound 複合 (endocentric) flat 並列 (exocentric) → flat:vv [動詞類]	list 細目 parataxis 隣接表現		root 親

表 1: 古典中国語に対する UD 依存構造タグ

造は、動詞をリンク元、賓語をリンク先、とする obj というリンクで表現する。リンクの本数は「単語」の個数に等しく、各リンクのリンク先は、全て互いに異なっている。すなわち、図 1 の依存文法解析結果にも示すとおり、各「単語」から出るリンクは複数ありうるが、各「単語」に入るリンクは 1 つだけである。また、リンクはループしない。さらに、われわれの古典中国語 UD では、リンクどうしが交差しない、root をまたぐリンクも存在しない、という制限も設けている。

依存文法解析のための手法は、これまでに数多く提案されているが、われわれの古典中国語 UD のように、複数の root を持ち (dependency forest)、UD 依存構造のリンクどうしが交差せず (planar)、root をまたぐリンクがない (projective)、という条件においては、arc-planar^[9] という (非決定性) アルゴリズムが、有効だと考えられる。arc-planar は、単語列の先頭から末尾に向かって「垣根」(stack-buffer boundary) を移動していく、というイメージで処理をおこなう。「垣根」がおこなう遷移は、Shift・Reduce・Left-Arc・Right-Arc の 4 種類である。

- Shift 「垣根」を右に 1 単語分、移動する。
- Reduce 「垣根」のすぐ左の単語を除去して、解析結果へ移す。
- Left-Arc 「垣根」のすぐ右の単語から、すぐ左の単語へリンクを繋ぐ。

- Right-Arc 「垣根」のすぐ左の単語から、すぐ右の単語へリンクを繋ぐ。

単語が全て Reduce されて、「垣根」がポツンと取り残された時点で、arc-planar は終了である。arc-planar による「孟子見梁惠王王曰叟不遠千里而來」の依存文法解析の様子を、図 3 に示す。あとは、リンクが入っていない「見」「曰」「遠」に root を刺すことで、図 1 の依存文法解析結果が得られるわけである。

ただし、arc-planar における「垣根」の遷移は、実際には非決定的^[10]である。図 3 では解析過程を一本道で示したが、現実には、各局面において複数の可能性が、枝分かれとして存在する。これら複数の可能性については、それぞれの遷移を選択した場合を、確率的に並行して解析することになる。

✳ 漢文の直接構成素解析

依存文法解析によって得られる UD 依存構造に対し、われわれは、階層化 UD 依存構造^[11]を導入した。階層化 UD 依存構造は UD 依存構造の変形であり、リンク先として辿っていくことができる単語を全て、リンク元に集約する形で階層化をおこなう (図 4)。ただし、UD 依存構造のリンクどうしが交差しない、root をまたぐリンクがな

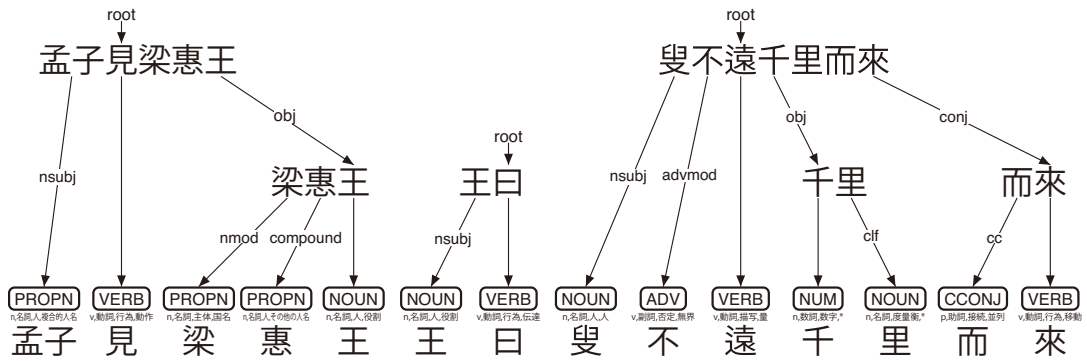


図 4: 階層化 UD 依存構造

い、というのが、階層化 UD 依存構造への変形可能条件である。

階層化 UD 依存構造は、直接構成素解析の前段階とみなすことができる。図 4 の階層化 UD 依存構造であれば、これに「見梁惠王」「惠王」「不遠千里」「遠千里」の 4 つの中間ノードを加えることで、直接構成素解析の結果解析木 (図 1) を導出できる。では、どういう形で中間ノードを加えるのが適切なのか。少し議論してみよう。

図 4 の「孟子見梁惠王」というノードからは、nsubj と obj のリンクが両方^[12]出ている。一方、図 1 解析木の「孟子見梁惠王」からは、nsubj は出ているが、obj は出していない。これを nsubj < obj という半順序で表すことにする。同様に、図 4 の「梁惠王」というノードからは、nmod と compound のリンクが両方^[12]出ている。一方、図 1 解析木の「梁惠王」からは、nmod は出ているが、compound は出していない。これを nmod < compound という半順序で表すことにする。

これらの半順序を階層化 UD 依存構造に適用し、半順序を満たすような中間ノードを加えることで、直接構成素解析が可能となる。たとえば、図 4 の「叟不遠千里而來」からは、nsubj と advmod と obj と conj のリンクが出ている^[12]が、ここに nsubj < obj および nsubj < advmod という条件があれば、obj と advmod のリンクを nsubj より後に出すために、「不遠千里」という中間ノードを加えるのが適切、ということになる。さらに、advmod < obj が追加されて、nsubj < advmod

< obj という条件ならば、「不遠千里」と「遠千里」という 2 つの中間ノードを加えるのが適切、ということになる。

ただし、これらの半順序は、多数の例文の間では時に矛盾を孕む。したがって、これらの半順序は確率として扱うべきであり、直接構成素解析における中間ノードの追加は、それぞれの可能性を、確率的に並行して解析することになるだろう。

❁ おわりに

古典中国語の形態素解析・依存文法解析・直接構成素解析に関して、ざっと、その概要を述べた。これらの解析を、できる限り正確におこなうためには、人手（もしくは半自動）で作成した漢文コーパスが必要となる。形態素解析における共起確率や、依存文法解析における遷移確率や、直接構成素解析における半順序確率は、大量の漢文コーパスから導出するしか、現実的な方法がないからだ。われわれが構築中の漢文コーパスは

<https://corpus.kanji.zinbun.kyoto-u.ac.jp/gitlab/Kanbun>

において、逐次公開中である。ぜひ一度、触れてみてほしい。

注

- [1] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (July 2004) , pp.230-237.
- [2] Igor A. Mel'čuk: Dependency Syntax: Theory and Practice, New York: State University of New York Press (1988) .
- [3] Rulon S. Wells: Immediate Constituents, Language, Vol.23, No.2 (April-June 1947) , pp.81-117.
- [4] 安岡孝一, ウィッテルン クリスティアン, 守岡知彦, 池田巧, 山崎直樹, 二階堂善弘, 鈴木慎吾, 師茂樹: 古典中国語 (漢文) の形態素解析とその応用, 情報処理学会論文誌, Vol.59, No.2 (2018年2月) , pp.323-331.
- [5] 山崎直樹, 守岡知彦, 安岡孝一: 古典中国語形態素解析のための品詞体系再構築, 人文科学とコンピュータシンポジウム「じんもんこん 2012」 論文集 (2012年11月) , pp.39-46.
- [6] 安岡孝一, ウィッテルン クリスティアン, 守岡知彦, 池田巧, 山崎直樹, 二階堂善弘, 鈴木慎吾, 師茂樹: 古典中国語 Universal Dependencies への挑戦, 情報処理学会研究報告, Vol.2018-CH-116 (2018年1月) , No.20, pp.1-8.
- [7] Joakim Nivre: Towards a Universal Grammar for Natural Language Processing, CICLing 2015: 16th International Conference on Intelligent Text Processing and Computational Linguistics (April 2015) , pp.3-16.
- [8] 安岡孝一: Universal Dependencies にもとづく古典中国語 (漢文) の依存文法解析, センター研究年報 2018 (2018年10月) .
- [9] Carlos Gómez-Rodríguez, Joakim Nivre: A Transition-Based Parser for 2-Planar Dependency Structures, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (July 2010) , pp.1492-1501.
- [10] 「垣根」が単語列の先頭にある局面では Shift を、末尾にある局面では Reduce をおこなうしかなく、これらの場合だけは決定的である。また、Left-Arc の直後を Reduce に、Right-Arc の直後を Shift に、それぞれ決め打ちすることは (局面に応じて) 可能である。
- [11] 守岡知彦: 古典中国語 UD コーパスの IPFS を用いた表現の試み, 情報処理学会研究報告, Vol.2018-CH-118 (2018年8月) , No.6, pp.1-7.
- [12] head を表すリンク^[11] (図4ではタグなしリンクで表現) も出ているが、このリンクは中間ノードの付加に用いることとし、半順序の比較には用いない。

中国語形態素解析エンジン概観

千田 大介 (ちだ だいすけ)

＊はじめに

近年、中国では形態素解析を含む自然言語処理の研究・開発が盛んである。CNKIで「电子技术及信息科学」カテゴリから「分词」を検索すると、執筆時点(2018年9月)で8,000以上の文献がヒットするし、検索範囲を優秀修士・博士論文データベースに絞っても、なお4,700件以上になる。

こうした状況が、近年の中国におけるネット社会の急速な進展を背景とした、ビッグデータ解析、各種インターネットサービス開発やマーケティングといったニーズを反映していることは、自然言語処理サービスを提供する博森社(<https://bosonnlp.com/>)トップページの画像に書かれた「从海量文本获得商业价值」(大量のテキストから商業的価値を獲得する)が端的に示している。

それゆえに中国語の自然言語処理・形態素解析エンジンには、大手デジタル企業が提供する商用のものも多いが、一方で、中国国外も含めて、研究機関、理工系大学や個人研究者が開発・配布しているフリーのものも複数存在している。

以下ではそうした形態素解析エンジンについて、ごく簡単に紹介したい。

＊形態素解析エンジン紹介

中国のネットを検索すると、数多くの形態素解析エンジンについて紹介している記事が幾つか見つかると、それらにもとづき、日本から無償で使える中国語形態素解析サービスを、以下にピックアップする。

● フリーツール

ネット上で配布されており、ユーザーが自由にダウンロードして利用できるツールには、以下のようなものがある。

◦ NLPPIR-ICTCLAS

<http://ictclas.nlpir.org/>

中国科学院^{アカデミー}が提供する、フリー中国語形態素解析エンジンの草分け。Java・C・C#・Pythonなどのプログラミング言語で使うことができ、Javaによる使用方法について、日本語で詳細に解説しているブログエントリーもある^[1]。

また、Python版のPyNLPPIRもあり、以下にインストールと使い方の説明がある。

<https://pypi.org/project/PyNLPPIR/>

Pythonが入っていればpipでインストールして使うことができる。

◦ ハルビン工科大学 LTP

<http://ltp.ai/>

ハルビン工科大学はMSピンインIMEの開発に携わるなど、自然言語処理の分野では有力な大学の1つである。LTPはフリーで提供される自然言語処理ツールで、形態素解析にとどまらず、多彩な機能を備えている。

しかし、新しいバージョンではインストールが上手くいかないケースが多く、特にWindowsでは使いづらくなっている。旧バージョンをMAC・Linux環境で使うことになろう。

また、上記サイト内の「在线演示」(オンラインデモンストレーション)では、ボックスにテキストを入れることで、LTPの多彩な機能を試用することができる。



ハルビン工科大学 LTP のデモンストレーション

▫ 清華大学 THULAC

<http://thulac.thunlp.org/>

清華大学自然言語処理与社会人文計算実験室が開発した形態素解析エンジン。C++、Java、Python 版がある。5,800 万字のコーパスでトレーニングした正確性を売りにしている。

後で、Windows での Python 版の使い方を詳述する。

▫ FudanNLP

<http://nlp.fudan.edu.cn/>

上海復旦大学自然言語処理実験室が LGPL3.0 で提供するオープンツール。インストール方法や使い方に関する解説が充実しているとはいいがたく、デモンストレーションページも執筆時点ではアクセスできなくなっている。一般のパソコンユーザーには敷居が高いといえよう。

▫ Hanlp

<http://hanlp.com/>

初期バージョンは上海林原信息科技有限公司の開発で、現在は大快搜索が主導する、オープンソースの自然言語ツール。形態素解析のほか、依存文法解析など多彩な機能も備える。上記サイトで、テキストボックスに任意の文を入れて、試用することができる。

後で、macOS + Python 環境における使い方を詳述する。

▫ 結巴中文分詞

<https://github.com/fxsjy/jieba>

MIT ライセンスで提供される、オープンソースの中国語形態素解析エンジン。上記 URL は

Python 版で、導入・利用が容易であることから、広く使われている。このほか、Java、C++、R、Ruby、PHP、iOS など、さまざまな言語版がある。

後で、Windows + Python 環境における使い方を詳述する。

▫ SnowNLP

<https://github.com/isnowfy/snownlp>

Python の形態素解析エンジン。pip でインストールして手軽に利用することができる。感情分析にも対応する。

▫ xmnlp

<https://github.com/JoeWoo/xmnlp>

Java の中国語分析パック。JDK1.6 対応。ApacheLicence 2.0 で提供される。さまざまな単語分割方法に対応し、ピンイン変換など多彩な機能を持つ。

▫ SCWS 中文分詞

<http://www.xunsearch.com/scws/index.php>

PHP に対応した形態素解析エンジンで、Unix 系環境での使用が推奨されている。BSD ライセンスによるオープンソース。上記サイトでデモ版を試用できる。

▫ 中文断詞系統

<http://ckipsvr.iis.sinica.edu.tw/>

台湾の中央研究院が提供する形態素解析ツール。学術目的の試用は無償だが、ユーザー登録が必要。Python 版もあるが (https://github.com/ldkrsi/ckip_python)、やはりユーザー登録してパスワードを受け取る必要がある。

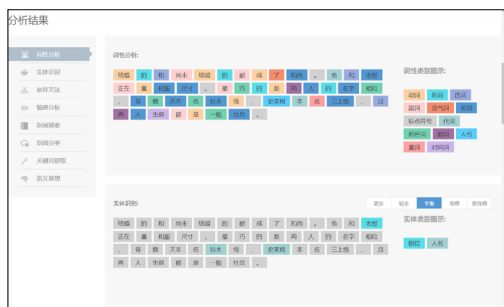
▫ Stanford Word Segmenter

<https://nlp.stanford.edu/software/segmenter.shtml>

米国スタンフォード大学による単語分割ツール。アラビア語と中国語に対応する。KH Coder の中国語エンジンに使われているので、詳細については、本誌の田邊鉄氏の論考「KH Coder で中国語分析」を参照されたい。

● Web ベースのサービス

商用サービスのデモ版サービスというような形態で、Web 上で形態素解析など、自然言語処理ツ-



BosonNLP のデモンストレーション

を提供しているサイトもある。大量のテキストを処理して統計分析するといった使い方は難しいが、それでも数百字程度のテキストを解析したり加工したりする際には重宝しよう。

▫ BosonNLP

<https://bosonnlp.com/>

正式名称は「玻森中文語義開放平台」(ボソン中国語語義オープンプラットフォーム)。API の利用は有償だが、ユーザー登録すると、多彩な機能を持つデモ版を使うことができる。

▫ 搜狗実験室クラウドサービス

<http://www.sogou.com/labs/websevice/>

搜狗系の検索ポータル、搜狗が提供する自然言語処理のデモ。テキストボックスに文を貼りつけると、形態素解析してくれる。テキストファイルのアップロードによる処理は、筆者が日本から試したところでは、上手くいかなかった。

▫ 騰訊文智

<http://nlp.qq.com/>

テンセントが提供する、資源言語処理プラットフォーム。各種 API の利用は有償だが、「体験」をクリックしてデモ版を試用することができる。

● 比較と評価

それぞれの形態素解析エンジンは、アルゴリズムや開発言語、動作環境が異なるため、性能、あるいは癖に違いがある。単語句切りについても、例えば「日本豆腐」を 1 語と取るか「日本」「豆腐」

の 2 語ととるか、各エンジンによって対応が分かれる。また、ビッグデータを扱う際には動作速度が問題になるので、そうした観点から形態素解析エンジン機能・性能を比較した記事も見られる。

形態素解析エンジンの性能比較に、1 つのモデルを提供しているのが、2005 年に実施された ICWB2 (第二届中国汉语分词测评) である。これは ACL (計算言語学会) の中国語処理分科会、SIGHAN が 2005 年に主催したコンテストだが、その際に使用されたトレーニング用、評価用のテキストファイルがネット上にアーカイブされており (<http://sighan.cs.uchicago.edu/bakeoff2005/>)、中国語形態素解析エンジンの性能比較の際に、たびたび利用されている。

その内容や評価については、自然言語処理の門外漢である執筆者の能力を超えているため、以下に 2 つの記事へのリンクを掲げるだけに留めたい。

▫ 11 款开放中文分词引擎大比拼

<https://ruby-china.org/topics/28000>

BosonNLP による比較レポート。当然のことながら、BosonNLP がトップ評価になっている。

▫ python | 六款中文分词模块尝试

https://blog.csdn.net/sinat_26917383/article/details/77067515

Python 対応エンジン 6 種の比較。機能比較のみならず、計算方法なども解説している。

※ 本稿は、日本学術振興会科学研究費補助金「初級学習者を対象としたコンテンツ駆動型中国語学習基盤の構築」(平成 30 年度、基盤研究 (B)、研究代表者: 田邊鉄、課題番号: 18H00682) による成果の一部である。

注

- [1] <https://language-and-engineering.hatenablog.jp/entry/20150207/ChineseTextMorphologicalAnalysisByNlpirle-tclas>

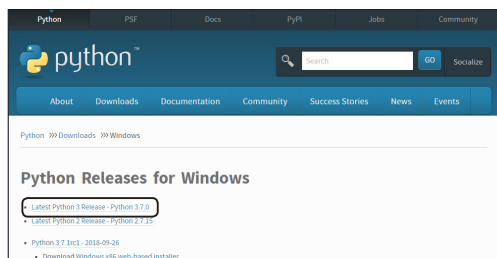
Windows で結巴、THLAC を使う

千田 大介 (ちだ だいすけ)

で右クリックし「システム」を選択、「デバイスの仕様」の「システムの種類」を見る。確認したらウィンドウを閉じる。



- Python 公式サイト (https://www.python.org/downloads/windows/) の「Latest Python 3 Release - Python 3.x.x」をクリックする。執筆時点の最新バージョンは 3.7.0。



- Python の選択したバージョンのダウンロードページが開く。ページをスクロールし、「Files」以下で「Windows x86-64 web-based installer」をクリック。

Version	Operating System	Description	MD5 Sum	File Size	GPB
gdbped source tarball		Source release	4186c956b4147a1e0517a7d9a580271	22745726	SGP
XZ compressed source tarball		Source release	4b8c2a6d1447050813c02714af46813	18922100	SGP
macOS 64-bit (32-bit installer)	Mac OS X	for Mac OS X 10.6 and later	ca3e8409d00f6d0e42e28374338e	34274481	SGP
macOS 64-bit installer	Mac OS X	for OS X 10.9 and later	ae0717a02efca30eb34aad680d498	27631276	SGP
Windows help file			46562a856c2049d0cc7680348180da	8547689	SGP
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	cb084f0d979a3025873ed541def10a5	6946082	SGP
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64	5313c18211e0a4107b6d6c6a129b3e	2628280	SGP
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	3c5d44c0d30a75aac122b40504d2	1327160	SGP
Windows x86 embeddable zip file	Windows		49941c0281699f532309c20723706f	6395982	SGP
Windows x86 executable installer	Windows		48b644c284c1447e902d87381af0	2556632	SGP
Windows x86 web-based installer	Windows		779c4085464eb3ee5b1a49ff0ebac4	1298280	SGP

＊はじめに

中国語形態素解析エンジンは、1つのエンジンが C++、Java などさまざまな言語のバージョンで提供されていることが多いが、以下ではプログラミングに慣れない人にとっても比較的取り扱いが容易であると思われる Python 版の結巴と THULAC について、その使い方を解説する。

Python は汎用プログラミング言語の一種だが、プログラミングコードがシンプルかつわかりやすくなるように設計されている。また Python で書かれたパッケージソフトを管理する pip というシステムによって、多くのプログラムが容易にインストールできる。

以下ではまず、Windows への Python の導入方法と使い方をごく簡単に解説した上で、中国語形態素解析エンジンを使って、入力した文字列、およびテキストファイルを、品詞情報付きで単語句切りする方法を解説する。

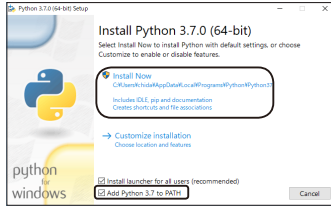
＊ Python の導入

Python の導入方法や使い方の詳細については、Python.jp (https://www.python.jp/) をはじめとする多くの Web サイトや書籍でわかりやすく解説されているのでそちらを参照していただくとして、ごく大まかな流れだけを以下に紹介する。

● インストール

- 使っている Windows が 32 ビット版か 64 ビット版かを確認する。スタートボタン上

- ④ダウンロードしたインストーラをダブルクリックして実行する。インストーラが起動したら、「Add Python 3.x to Path」をチェックし、「Install Now」をクリック。

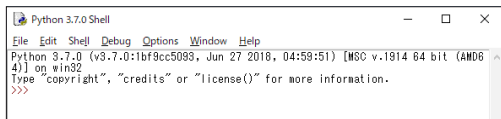
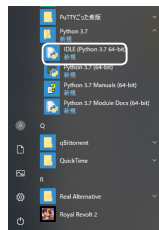


- ⑤ユーザーアカウント制御の警告が出たら、「はい」をクリック。
⑥インストールの完了画面が表示されたら、「Close」をクリック。

以上で、Python のインストールは完了である。

● IDLE をつかう

Python のプログラミングやプログラムの実行には、IDLE を使うのが便利である。IDLE は、スタートメニューの「Python 3.x」をクリックして展開し、「IDLE (Python 3.x nn-bit)」をクリックして起動する。



「>>>」の後にプログラムコードを書いたり、コピー＆ペーストしたりして実行する。また、プログラムを保存して実行することもできる。

✦ 結巴の使い方と形態素解析結果

結巴 Python 版のセットアップ方法や使い方は、以下のページに詳しい。

<https://github.com/foxsjy/jieba>

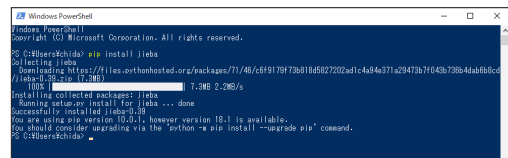
その内容をふまえつつ、インストールおよび形

態素解析の基本的手順を解説する。

● 結巴のインストール

結巴は pip を使ってインストールする。

- ①スタートボタンを右クリックし、「Windows PowerShell」を起動する。
- ②「pip install jieba」と入力しエンターキーを押す。インストールが始まる。
- ③インストールが完了したら、Windows PowerShell を閉じる。



● IDLE で結巴を使う

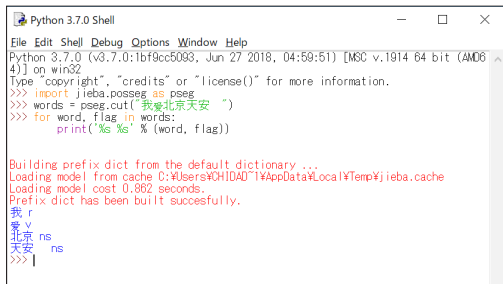
準備が整ったら、実際に結巴を使ってみよう。結巴 Python 版のページ (<https://github.com/foxsjy/jieba>) を開き、「4. 词性标注」以下の品詞情報付き形態素解析のサンプルコードを試してみる (行頭の数字は行番号)。

```
1 import jieba.posseg as pseg
2 words = pseg.cut("我爱北京天安门")
3 for word, flag in words:
4     print('%s %s' % (word, flag))
```

1 行ずつ、Web ページからコピーして、IDLE 上で右クリック、「Paste」して、エンターキーを押す。

2 行目は、IDLE 上で簡体字が欠けて表示されないかもしれない。これは表示用フォントに簡体字対応フォントが指定されていないために欠けたもので、データはちゃんと Python に渡せているので心配ない。

3 行目を入れると、4 行目の行頭は自動でインデントが入った状態になる。4 行目をペーストし



たら、エンターを2回押すと、「我爱北京天安门」の解析結果が表示される。

簡体字が欠けていても、IDLE からエディタ・ワープロなどにコピー & ペーストすると、全ての文字が見えるようになる。

```
我 r
爱 v
北京 ns
天安门 ns
```

なお IDLE で簡体字を表示したい場合は、「Options」→「Configure IDLE」を選択して、フォントを「SimHei」などの簡体字フォントに切り替えればよい。

結巴の品詞略号は以下ようになる。

ag	形语素	a	形容词	ad	副形词
an	名形词	b	区别词	c	连词
dg	副语素	d	副词	e	叹词
f	方位词	g	语素	h	前接成分
i	成语	j	简称略语	k	后接成分
l	习用语	m	数词	ng	名语素
n	名词	nr	人名	ns	地名
nt	机构团体	nz	其他专名	o	拟声词
p	介词	q	量词	r	代词
s	处所词	tg	时语素	t	时间词
u	助词	vg	动语素	v	动词
vd	副动词	vn	名动词	w	标点符号
x	非语素字	y	语气词	z	状态词
un	未知词				

● テキストファイル进行处理する

結巴は、以下のコマンドでテキストファイルを直接処理することができる。

```
python -m jieba news.txt > cut_
result.txt
```

しかし Windows では PowerShell、コマンドプロンプト、いずれもコマンドを打ち込んで直接テキストファイルを読み込もうとすると、ShiftJIS コードとして扱ってしまうためエラーになる (macOS や Linux などではこの問題が発生しない)。

このため、文字コードを指定してファイルを読みこみ、形態素解析して別ファイルに保存する、といった処理は、スクリプトを保存して実行するのが便利である。

以下、スクリプトのサンプルを掲出する。解析するテキストファイルを D:\test\ に「input.txt」の名称で保存していることとする (以下のスクリプトでは「¥」は「\」と表示されている)。またテキストファイルは、文字コードを BOM 無し UTF-8 で保存する。メモ帳は BOM 付き UTF-8 でしか保存できないので、EmEditor などのエディタを使う。

- ① IDLE を起動し、左上の「File」をクリック、「New File」を選択。
- ② スクリプトの新規作成画面が開くので、以下の内容を入力¹⁾。

```
1 # -*- coding: utf-8 -*-
2 import jieba
3 import jieba.posseg as pseg
4 f=open(r"d:\test\input.
  txt","r",encoding='utf-8')
5 string=f.read()
6 words = pseg.cut(string)
7 result=""
8 for w in words:
9     result+= str(w.
  word)+" /"+str(w.flag)+" "
```

```

10 f=open(r"d:\test\output.
    txt","w",encoding='utf-8')
11 f.write(result)
12 f.close()

```

9 行目の行頭では tab を一度押して字下げにする。同じ行で、「"/」の / は単語と品詞の区切り文字、「" "」の半角スペースは単語間の区切り文字になるので、適宜書き換えてかまわない。

- ③ 「File」→「Save」でフォルダを選択し、名前を付けて保存。ここでは、「d:\test¥script」に「jieba1」の名前で保存。
- ④ F5 を押してプログラムを実行。

これで、D:\test フォルダの「output.txt」に解析結果が保存される。スクリプトを保存しておけば、IDLE の「File」→「Open」で保存したスクリプトを開き、必要に応じてフォルダ名・ファイル名などを書き換え、繰り返し使うことができる。

● 形態素解析結果

実際の解析結果を、以下の 3 つの例文で確認してみる (例文は田邊氏・師氏と統一している)。

- ① 结婚的和尚未结婚的都成了和尚。
- ② 他和老板正在量和服尺寸。
- ③ 最巧的是两人的名字相似，骨傲天本名铃木悟，史莱姆本名三上悟，且两人生前都是一般社员。

①と②は中国語形態素解析エンジンのテストでしばしば使われる例文である。①は「和尚」が 2 箇所出現するが、前者は「和 / 尚未」と切るのが正しい。②は「和服」を「和 / 服」と切らずに一語の名詞と認識できるかがポイントとなる。③は執筆時点で最新の日本アニメ関係の記事からの引用である。作品名『オーバーロード』のネットスラング「骨傲天」を語彙として認識できるか、「史莱姆」はここでは作品名『転生したらスライムだった件』の略称として使われているが、アニメ・ゲー

ムに出てくるモンスター「スライム」の訳語として認識できるか、「鈴木悟」「三上悟」という日本人名を認識できるか、などを確認する。

- ① 结婚 /v 的 /uj 和 /c 尚未 /d 结婚 /v 的 /uj 都 /d 成 /v 了 /ul 和尚 /nr 。 /x
- ② 他 /r 和 /c 老板 /n 正在 /t 量 /n 和服 /nz 尺寸 /n 。 /x
- ③ 最巧 /a 的 /uj 是 /v 两 /m 人 /n 的 /uj 名字 /n 相似 /v ， /x 骨 /ng 傲 /a 天 /q 本名 /n 铃木 /nr 悟 /v ， /x 史莱姆 /nr 本名 /n 三 /m 上悟 /v ， /x 且 /c 两 /m 人 /n 生前 /t 都 /d 是 /v 一般 /a 社员 /n 。 /x

①②に関しては問題ない。③については「骨傲天」を語彙として認識せず、「鈴木悟」「三上悟」は、それぞれ「悟」「上悟」を動詞として認識した。「史莱姆」は史姓の人名として解釈したようだ。

● 結巴のその他の機能

結巴には以上の他にもさまざまな機能がある。単語句切りについても、品詞情報を付加しない場合は、「jieba.cut」および「jieba.cut_for_search」コマンドを使うことになる。「jieba.cut」コマンドには、「cut_all」(デフォルトは False)、「HMM」(デフォルトは True) 2 つのオプションがあり、「jieba.cut_for_search」は「HMM」を指定でき、たとえば「jieba.cut(" 我爱您。", cut_all=True)」のように記述する。③の例文の一部がどうなるかを比較してみよう。

◦ jieba.cut

● デフォルト

骨傲 / 天 / 本名 / 铃木 / 悟 / ， / 史莱姆 / 本名 / 三上 / 悟 / 。

● cut_all=True

骨 / 傲 / 天 / 本名 / 铃木 / 悟 /// 史莱姆 / 莱姆 / 本名 / 三 / 上 / 悟 //

● HMM=False

骨 / 傲 / 天 / 本名 / 铃木 / 悟 / ， / 史莱姆 / 本名 / 三 / 上 / 悟 / 。

● cut_all=True, HMM=False

骨 / 傲 / 天 / 本名 / 鈴木 / 悟 /// 史萊姆
/ 萊姆 / 本名 / 三 / 上 / 悟 //

▫ jieba.cut_for_search

● デフォルト

骨 傲 / 天 / 本名 / 鈴木 / 悟 /, / 萊姆 /
史萊姆 / 本名 / 三上 / 悟 /。

● HMM=False

骨 / 傲 / 天 / 本名 / 鈴木 / 悟 /, / 萊姆 /
史萊姆 / 本名 / 三 / 上 / 悟 /。

jieba.cut_for_search は、検索用の逆引きインデックス作成のためのコマンドになる。cut_all オプションは、テキスト中の単語である可能性のある語を、重複ありでリストアップする。HMM は未知の語彙に対応する機能で、デフォルトの HMM=True 状態では「三上」を語として認識していることがわかる。前の品詞情報付きの分割では、この機能が働いていないために精度が落ちていようだ。

jieba.cut でテキストファイル进行处理するのであれば、例えば以下のようなスクリプトになる。

```
# -*- coding: utf-8 -*-
import jieba
f = open(r"d:\test\input.txt", "r", encoding='utf-8')
string = f.read()
words = jieba.cut(string, cut_all=True, HMM=False)
result = ' / '.join(words)
f = open(r"d:\test\output.txt", "w", encoding='utf-8')
f.write(result)
f.close()
```

結巴は、また「jieba.analyse」コマンドによって、キーワードを抽出することができる。公式サイトにサンプルスクリプトへのリンクが張られているので、興味をお持ちの方はそれらを使って試して頂きたい。このほか、ユーザー辞書や辞書のトレーニングといった、辞書関係の機能も整っている。

✳ THULAC を使う

清華大学の THULAC も、Python で容易に取り扱うことができる。

まずインストールだが、pip に対応しているので、「Windows PowerShell」を起動して、「pip install thulac」と入力してエンターを押せばよい。

● 単語の分割とオプション項目

まずは、THULAC のサイト (<http://thulac.thunlp.org/>) の説明に従って、IDLE を使って形態素解析してみよう。

サイトの「使用方式 > 1. 分詞和词性标注程序 > 1.1. 命令格式」の以下のコマンドを、IDLE に 1 行ずつコピー & ペーストしてエンターを押していく。

```
import thulac
thu1 = thulac.thulac()
text = thu1.cut(" 我爱北京天安门 ",
text=True)
print(text)
```

解析結果は以下ようになる。

我 _r 爱 _v 北京 _ns 天安门 _ns

THULAC の各種オプションは、「thulac()」の () 内に記述する。オプション項目は以下の通り。

- user_dict：ユーザー辞書の設定。
- T2S：繁体字から簡体字に変換。
- seg_only：単語区切りのみを行い、品詞を注記しない。
- filt：意味の無い語句をフィルタリング。
- model_path：モデルファイルのパス。

例えば、上の例で単語区切りのみを行う場合は、2 行目を以下のように書き換える。

```
thu1 = thulac.thulac(seg_
only=True)
```

3行目の「.cut("我爱北京天安门", text=True)」の「text=True」であるが、「text=False」とすると、出力結果が以下ようになる。

```
[['我', 'r'], ['爱', 'v'], ['北京', 'ns'], ['天安门',
'ns']]
```

THULAC についても、テキストファイル进行处理するコマンドがあるのだが、Windows ではうまく動作しないので、やはりスクリプトを保存して実行するのが楽で良い。以下にサンプルを掲げる。フォルダ、テキストファイルなどについては、前の結巴と同様である。

```
# -*- coding: utf-8 -*-
import thulac
seg = thulac.thulac()
f = open(r"d:\test\input.
txt", "r", encoding='utf-8')
string=f.read()
words = seg.cut(string,
text=True)
f = open(r"d:\test\output.
txt", "w", encoding='utf-8')
f.write(words)
f.close()
```

● 例文の解析結果

- ①结婚_v 的_u 和_c 尚未_d 结婚_v 的_u 都_d 成_v 了_u 和尚_n。_w
- ②他_r 和_c 老板_n 正在_d 量_v 和_c 服_v 尺寸_n。_w
- ③最_d 巧_a 的_u 是_v 两_m 人_n 的_u 名字_n 相似_a, _w 骨傲天本名铃木_id 悟_g, _w 史莱姆_np 本_r 名三_n 上_f 悟_g, _w 且_c 两_m 人_n 生前_t 都_d 是_v 一_m 般_q 社员_n。_w

①②については特に問題ない。③では、うまく解析できずに「骨傲天本名鈴木」となっており、また日本人名も認識できなかった。

なお、THULAC の品詞略号は以下ようになる。

n	名词	np	人名	ns	地名
ni	机构名	nz	其它专名	m	数词
q	量词	mq	数量词	t	时间词
f	方位词	s	处所词	v	动词
a	形容词	d	副词	h	前接成分
k	后接成分	i	习语	j	简称
r	代词	c	连词	p	介词
u	助词	y	语气助词	e	叹词
o	拟声词	g	语素	w	标点
x	其它				

✿ おわりに

以上、ごく簡単に結巴、THULAC の使い方について解説した。筆者は計量言語学については門外漢であるし、Python も初心者にはほかならないので、プロパーの方々からすれば、常識以前のことを書き連ねていることは重々承知している。

しかるに中国語形態素解析は、文学研究、語学教育、文化研究などに幅広い応用が可能であると思われるが、そもそもやり方そのものが余り知られていないのが実情である。本稿がそうした状況を変える一つのきっかけになれば、幸いである。

※ 本稿は、日本学術振興会科学研究費補助金「初級学習者を対象としたコンテンツ駆動型中国語学習基盤の構築」(平成30年度、基盤研究(B)、研究代表者: 田邊鉄、課題番号: 18H00682) による成果の一部である。

注

- [1] スクリプトは以下のサイトに掲載されているもののもとに、若干の修正を施したものである。<https://blog.csdn.net/caicai1617/article/details/21556073>

macOS で HanLP を使う

師 茂樹 (もろ しげき)

✳️ はじめに

HanLP は、初期バージョンを上海林原信息科技有限公司が開発し、現在は大快搜索が開発を主導する中国語自然言語処理環境である。形態素解析 (中文分詞・词性标注) だけでなく、依存文法解析 (依存句法分析) やキーワード抽出 (关键词提取)、自動要約 (自动摘要) など、様々な機能を有している点に特徴がある。

HanLP は Java で開発されているが、Python のインターフェースも公開されている。ここではインストール等が比較的容易な Python 版について簡単に紹介する。

なお、形態素解析などの自然言語処理は、それだけで何かを行うというより、他の処理 (テキストマイニングなど) の一部として用いられるものである。したがって、GUI などのインターフェースがあるものは少なく、Java や Python といったプログラミング言語の知識を前提とするものが多い。以下の説明では、プログラミング言語についてある程度知識があることを前提とし、ターミナルを用いた CUI の操作についても解説しない。

✳️ macOS へのインストール方法

以下に述べる方法は一例であり、OS のバージョンなど、ユーザの環境に応じてこの方法ではインストールできないこともある。その場合は各自でマニュアル等を参照されたい。

● Java の導入

HanLP は Java で開発されているので、Java のサイト (<https://www.java.com/ja/>) から Java のインストーラをダウンロード、インストールする。

● Python の導入

macOS には最初からバージョンは 2.x 系統の Python がインストールされている。HanLP は 2.x、3.x のどちらでも動作するが、Unicode についてはバージョン 3.x 系統のほうが扱いやすいので、そちらをインストールしたほうがよいだろう。3.x 系統をインストールする場合には、Python のサイト (<https://www.python.org/downloads/mac-osx/>) から最新バージョン (ただし、バージョンのあとに「~rc1」などと付いているのは開発中なので避けたほうがよい) をダウンロードして実行、インストールするか、MacPorts (<https://www.macports.org>) や Homebrew (<https://brew.sh>) などのパッケージ管理システムを使ってインストールをする。

● pip の導入

Python 版の HanLP は Python のパッケージ管理システムである pip で提供されているので、pip をインストールする。OS 付属の Python の場合、ターミナルで、

```
$ sudo easy_install pip
```

とすればインストールできる (sudo を使う場合には、管理者のパスワードが必要)。MacPorts の場合は、


```
$ sudo port install py-pip
```

とすれば、インストールされている Python に応じた pip がインストールされる。その後、

```
$ sudo port select --set pip pip37
```

とすれば (最後の 37 は、Python のバージョンが 3.7 の場合)、Python 3.7 用の pip をデフォルトに設定することができる。Homebrew の場合は、Python と同時に pip が同時にインストールされる。

● Command Line Tools for Xcode の導入

pip はコマンドラインで動作する C++ コンパイラが必要なので、App Store で Xcode をインストールした後 (図 1)、ターミナルで、

```
$ sudo xcode-select --install
```

を実行し、Command Line Tools for Xcode をインストールする。

● pyhanlp の導入

以上のインストールが終わったら、ターミナルで以下のコマンドを実行することで Python 版の HanLP がインストールされる。

```
$ sudo pip install pyhanlp
```

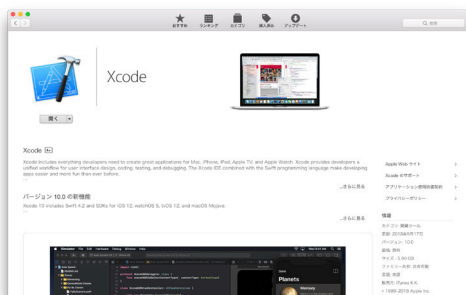


図 1 App Store で Xcode をインストール

次に、ターミナル上で

```
$ python
```

と実行して対話型プログラミングモードにし、

```
>>> from pyhanlp import *
```

を実行すると、HanLP の本体を自動でダウンロード、インストールを開始する (ダウンロードに時間がかかるが、随時進捗が表示される)。

✿ Python 版 HanLP の使用方法

ここでは、Python の対話型プログラミングモードを使った対話的な操作方法について説明する。先のインストール時と同様、ターミナル上で、

```
$ python
```

と実行して対話型プログラミングモードにし、

```
>>> from pyhanlp import *
```

を実行すると、HanLP を使う準備が整う (インストールが完了していない場合はインストールが始まる)。

では、形態素解析を行ってみよう。以下のように、HanLP.segment() という関数で形態素解析を行うことができる。

```
>>> print(HanLP.segment('我爱北京天安门'))
```

```
[我 /rr, 爱 /v, 北京 /ns, 天安门 /ns]
```

結果にある「rr」や「v」などは品詞の略号である。

rr	人称代名詞	v	動詞
ns	地名		

HanLP の品詞は 148 種類もあるため、ここで

は略号一覧を載せることはしない。「HanLP 词性标注集」(<http://www.hankcs.com/nlp/part-of-speech-tagging.html#h2-8>)を参照されたい。

✿ 形態素解析結果

この記事と同様、以下の3つの例文を HanLP で形態素解析してみたい(例文のポイントは、千田大介氏の記事を参照されたい)。

1. 结婚的和尚未结婚的都成了和尚。
2. 他和老板正在量和服尺寸。
3. 最巧的是两人的名字相似，骨傲天本名铃木悟，史莱姆本名三上悟，且两人生前都是一般社员。

結果は以下の通り。

1. 结婚 /vi, 的 /ude1, 和 /cc, 尚未 /d, 结婚 /vi, 的 /ude1, 都 /d, 成 /v, 了 /ule, 和尚 /nnt, 。 /w
2. 他 /rr, 和 /cc, 老板 /nnt, 正在 /d, 量 /n, 和 /cc, 服 /v, 尺寸 /n, 。 /w
3. 最 /d, 巧 /a, 的 /ude1, 是 /vshi, 两 /m, 人 /n, 的 /ude1, 名字 /n, 相似 /a, , /w, 骨 /ng, 傲 /a, 天 /qt, 本名 /n, 铃木 /nz, 悟 /vg, , /w, 史莱姆 /nr, 本名 /n, 三 /m, 上 /f, 悟 /vg, , /w, 且 /c, 两 /m, 人 /n, 生前 /t, 都 /d, 是 /vshi, 一般 /ad, 社员 /nnt, 。 /w

上の結果に出てくる品詞略号は以下の通り。

a	形容词	ad	副形词
c	连词	cc	并列连词
d	副词	f	方位词
m	数词	n	名词
ng	名词性语素	nnt	职务职称
nr	人名	nz	其他专名
qt	时量词	rr	人称代词
t	时间词	ude1	助词(的、底)
ule	连词(了、喽)	v	动词

vg	动词性语素	vi	不及物动词(内动词)
vshi	动词“是”	w	标点符号

1については問題ないようである。2については「和服」を2語として解釈してしまっている。3については、ネットスラング「骨傲天」、日本人名「鈴木悟」「三上悟」を認識できていない^[4]。「史莱姆」は人名として認識してしまっている。

● 依存文法解析

ある形態素がどの形態素にかかっているかを分析する依存文法解析をするためには、以下のようになる。

```
>>> print(HanLP.parseDependency('结婚的和尚未结婚的都成了和尚。'))
```

結果は以下の通り。

- 1 结婚 结婚 v v_8 主谓关系 __
- 2 的 的 u u_1 右附加关系 __
- 3 和 和 c c_5 左附加关系 __
- 4 尚未 尚未 d d_5 状中结构 __
- 5 结婚 结婚 v v_1 并列关系 __
- 6 的 的 u u_5 右附加关系 __
- 7 都 都 d d_8 状中结构 __
- 8 成 成 v v_0 核心关系 __
- 9 了 了 u u_8 右附加关系 __
- 10 和尚 和尚 n n_8 动宾关系 __
- 11 。 。 wp w_8 标点符号 __

左から7番目の数字は、その形態素がどこにかかっているかを示している。たとえば、副詞である7の「都」は、動詞である8の「成」にかかっている。この例文の分析は、うまくいっているようである。

✿ 終わりに

以上、HanLP のインストール方法と使い方に

ついでごく簡単に紹介した。筆者は、多機能な HanLP のごく一部をテストしただけであり、十分な紹介ができたとは言えない。ぜひとも以下のサイトにある詳細情報 (中国語) を参照されたい。

<https://github.com/hankcs/HanLP>

注

- [1] HanLP には「日本人名識別」機能があり、以下の手順で実行できるはずであるが、実行結果は先にあげたも

のと変わらなかった。

```
>>> a = HanLP
>>> a.enableJapaneseNameRecognize
= True
>>> print(a.segment('最巧的是两人的名字相似，骨傲天本名铃木悟，史莱姆本名三上悟，且两人生前都是一般社员。'))
```

HanLP には他にも「中国人名識別」、「音译人名識別」、「地名識別」、「实体机构名識別」などの機能がある。

コラム：Word で単語区切り

千田 大介 (ちだ だいすけ)

✳ Words オブジェクトの応用

単純に文を単語に区切るだけならば、Microsoft Word でも対応できる。

Word の VBA マクロを使うと、ドキュメント中の単語を取得できる。たとえば、

```
ActiveDocument.Words(12)
```

は、現在の Word 文書の 12 番目の単語、という意味になる。

MS Office では多言語校正ツール集である Proofing Tools を導入すると、日本語以外の言語の単語も認識できるようになるので、VBA を使った中国語の単語句切りも可能になる。

筆者は、この機能を応用し、Word ドキュメントの言語を指定して単語句切りする「ワードブレイカー」マクロを公開している^[1]。

それによって、前の記事中使ったサンプル文書を分割すると、以下のようになる。

- ① 结婚 / 的 / 和 / 尚未 / 结婚 / 的 / 都 / 成 / 了 / 和尚 / 。 /
- ② 他 / 和 / 老板 / 正在 / 量 / 和服 / 尺寸 / 。 /
- ③ 最 / 巧 / 的 / 是 / 两 / 人 / 的 / 名字 / 相似 / ， / 骨 / 傲 / 天 / 本名 / 铃木 / 悟 / ， / 史莱 / 姆 / 本名 / 三 / 上 / 悟 / ， / 且 / 两 / 人 / 生前 / 都是 / 一般 / 社员 / 。 /

①②に関しては「和尚」・「和」などを正確に区切れている。③では、「スライム」を「史莱 / 姆」

と区切り、日本人名のうち「鈴木 / 悟」は問題ないが「三 / 上 / 悟」はうまく認識できていない。ただ、これらは多くの中国語形態素解析エンジンでもうまく区切れなかった語である。

Word は、単に単語を区切るだけならば、あまたの形態素解析エンジンと遜色ないレベルにあるといえよう。

✳ おわりに

ネットを検索すると、Word ドキュメント中の単語一覧を作成したり、一覧を Excel に出力したりするマクロが幾つか公開されており^[2]、それらは、Proofing Tools が入っていれば、中国語ドキュメントでも利用できる。

人文系研究者にとってプログラミングの敷居は高いかも知れないが、プログラミングができなくても、VBA マクロの扱い方を知っておきさえすれば、ネットで入手したプログラムを利用できるようになり、応用の可能性が広がるので、是非、これらの VBA マクロを試してみたい。

注

[1] <https://wagang.econ.hc.keio.ac.jp/index.php?download>

[2] たとえば、以下のようなものがある。

- 【コード】文書に含まれる単語を調べるマクロ_並べ替え付き (Collection オブジェクト) (<https://www.wordvbalab.com/code/3043/>)
- 【VBA】Word 文書中の単語一覧を Excel に出力する (<http://jazzmaster.hatenablog.com/entry/2012/05/01/013352>)

特集 2

デジタル時代の 中国学リファレンス ②

前号に引きつづき、情報化時代をふまえた中国学の調査方法について解説する。今号では、官職、仏教、道教、通俗文学、現代中国語などの分野の、基本的な調査方法や参考図書・叢書などの扱い方について、取り上げている。これからそうした専門課程を学ぼうという学生・大学院生だけでなく、広く周辺領域に目配りしたいと考えている方々にも、ご活用頂けたら幸いである。

なお、本特集では著作権法など、法的な問題の可能性のあるコンテンツについては原則として取り上げていない。また、参照ポイントには、前号掲載分への参照も含まれるので、ご注意頂きたい。

CONTENTS

基礎知識

解説：CNKI 工具書庫.....78

1. 官職をしらべる

職掌をしらべる.....80
官品をしらべる.....82
官職の別称をしらべる.....84
官制を視覚的に把握する.....86
解説：官制の構造.....88

解説：官僚登用制度.....91
清代の官職の官品や職掌をしらべる.....93
清代の人物の官職をしらべる.....95
中華民国時期の人物の官職をしらべる.....97

2. 新聞資料をしらべる

「満洲」の新聞をしらべる.....99

3. 仏典をしらべる

仏典の叢書・テキスト.....101
仏典を検索する.....106

解説：仏典の分類.....111

4. 道教・民間信仰をしらべる

道教経典をしらべる.....114
解説：『道蔵』とは.....117

民間信仰の神をしらべる.....121
解説：民間信仰と民間宗教の資料.....123

5. 古典通俗文学をしらべる

古典通俗小説の概要をしらべる.....126
古典通俗小説の排印本.....128
古典通俗小説の影印本をさがす.....130
解説：官刻本・家刻本・坊刻本.....131
古典通俗小説の全文検索.....132

古典戯曲作品の概要をしらべる.....134
古典戯曲作品の影印本・排印本をさがす.....136
古典戯曲作家の情報をしらべる.....138
白話語彙の意味をしらべる.....140

6. 現代中国語をしらべる

中日辞典に載っていない中国語の意味.....143
日中辞典に載っていない日本語の中国語訳.....146

欧米人名の中国語訳.....147

解説：CNKI 工具書庫

❖ 新デザイン移行

前号で紹介した CNKI 工具書館ですが、今年 2018 年に新デザインのサイトが登場しました。トップページのロゴでは名称も「工具書庫」に改められています（トップページの HTML では「工具書総庫」になっているのですが、読者のわかりやすさを考えて、以下は「工具書庫」に統一します）。執筆時点では、海外向けサーバ（oversea.cnki.net）は旧版のままですが、近い将来、新デザインに移行することでしょう。

新デザインでは、機能は旧版とさして変わりませんが、インターフェイスが大幅に変更されています。このため、ここで新インターフェイスの使い方について、解説しておきます。

❖ 検索と閲覧

◆ 基本的な検索・閲覧手順

トップページの検索ボックス①にキーワードを入力し「検索」（検索）②をクリックします。このとき、検索ボックス右端のプルダウン③で、検索範囲を以下の 4 つから選択できます。

- 条头（精确）…見出し語（完全一致）
- 条头（模糊）…見出し語（あいまい）
- 条目…項目の全文
- 图书…収録書籍のタイトル

デフォルトでは「条头（精确）」になっていて、たとえば「居易」で検索すると、「居易」という項目だけを探します。検索結果画面の検索ボックス右端④で「条头（模糊）」に切り替えて検索すると、項目タイトルに「居易」が入っている「白居易」「郭居易」などもヒットします。「条目」は、項目の解説文も含めて検索します。デフォルトの設定でうまく見つからないときは、「条头（模糊）」や「条目」に切り替えて試してみましょう。

また検索ボックス下の「分館」「工具书类型」右の各項目⑤をクリックすることで、検索範囲を絞り込むことができます。複数ジャンルも選択できます。クリックしたジャンル名は検索ボックスの下に表示されます⑥。選択を解除するときは、検索ボックス下のジャンル名にカーソルを合わせて、右側に表示される「×」⑦をクリックします。

検索結果一覧画面で、項目見出し⑧をクリックすると、個人アカウントの場合は課金の確認が出たあと、工具書の当該項目が全文表示されます。

❖ その他の機能

トップページの検索ボックス上のタブ⑨で、検索対象を画像・表などに限定することができます。

ジャンルごとの工具書とその一覧を閲覧するには、工具書庫トップ画面上のメニュー⑩をクリックします。その右の「工具书旧版」⑪から、旧バージョンのトップページにジャンプできます。分類・



ジャンルの一覧は、トップページの下の方にも掲載されていますので、スクロールしてそこからたどることもできます。

検索条件を細かく設定するには、トップページ「検索」ボタン右の「高級検索」(高級検索)をクリックします。通常の検索ボックスでも「?」(任意の一字)「*」(任意の一字以上の文字)を使ったワイルドカード検索ができます。

❖ ユーザーアカウントに関する変更

CNKI は 2018 年より、中国国内でチャージされたアカウントによる海外からの使用に大幅な制

限が掛けられ、有償コンテンツへのアクセスは年に数回だけしかできなくなりました。海外の代理店の要請を受けた措置だそうです。

このため、日本国内から、所属機関が契約していない CNKI の有償コンテンツにアクセスするには、代理店である東方書店から日本国内向けの CNKI カードを購入することがが必須になりました。

カードの購入や、アカウントへのチャージ方法については、東方書店にお問い合わせ下さい。

(千田 大介)

官職をしらべる

❖ 職掌をしらべる

Q: 三国志で曹操の腹心であった^{じゅんい}荀彧が就いた「侍中尚書令」^{じちゅうしょうしょれい}とはどのような官職でしょうか。

後漢の建安元年（196）、献帝は曹操の庇護を受け曹操の根拠地であった許に遷都します。これにより、後漢の朝廷は曹操の意のままとなり、彼の傀儡といってよい状態に陥ります。この時、功績のあった荀彧が就任したのが侍中尚書令です。

荀彧は曹操配下の腹心として有名ですが、元々は27歳で孝廉科（☞解説：官僚登用制度）に推挙され、若くして県の長官になった後漢のエリート官僚でした。董卓の乱で官を棄てた彼が、めぐりめぐって侍中尚書令となるのですから、乱世とは何が起きるかわからないものです。

❖ 辞書や事典を調べる

『大漢和辞典』や『漢語大詞典』などの大型辞書は、官職など制度上の用語もある程度網羅しています。しかし、漢語辞書による官職名の解説は時代別の細かな違いがわかりません。歴史事典や官職に特化した専門辞典類を参考にしましょう。

①『アジア歴史事典』（平凡社、1961初版）

アジア史に関する唯一の大事典です。刊行後60年近く経っていますが、現在でも十分に活用できます。検索すると「侍中」と「尚書令」の項目がありました。ここでわかるように侍中と尚書令という二つの官職であったのです。

中国語の辞典では官制や官職に関する専門辞典が充実しています。なかでも全時代を対象としているものには次のようなものがあります。

②『中国官制大辞典』上・下（俞鹿年編著、黒竜江人民出版社、1992）

③『中国官制大辞典』[重訂本]（徐連達編著、上海大学出版社、2010）

④『中国歴代官制大辞典』修訂版（呂宗力主編、商務印書館、2015）

どれも1990年前後に刊行されたものですが、③と④は近年改訂版が出されて、最新の研究成果も盛り込まれたものとなり、入手しやすくなりました。なお④の初版は北京出版社が1994年に同じタイトルで、③の初版は安徽教育出版社が1991年に『中国歴代官制詞典』として刊行したものです。②および④の初版はCNKI 工具書庫（☞解説：CNKI 工具書庫）でも検索することができます。

どの辞典にも侍中と尚書令の項目がありました。解説の分量や精粗は辞書により差がありますが、大要はほぼ同じです。中国の官職名は同じ名称でも時代により別の職掌であることがままあります。このため、調べたい官職がいつの時代のものであるかを確認した上で、辞書の記述を読み解かねばなりません。この場合は、後漢に該当する記述をさがしましょう。

これらによれば、後漢の侍中は皇帝の近くにあつて庶務を管理し、政務に関する顧問となる職掌であったことがわかります。一方の尚書令は皇帝の秘書官として政務全体を総攬する立場にあつたことがわかります。

ところで『後漢書』や『三国志』の荀彧伝をみると、正確には「侍中・守尚書令」となっています。「守」が何を意味するのかについても、官職辞典類で守を検索してみればわかります。後漢時代には一年間は試用期間として官職名の上に「守」がつき、試用期間が終わって任に適う人物だと判断されればこれが外されることになっています。

❖ 史料や先行研究にあたる

こういった事典・辞典類を使えば、中国前近代の官職についてはかなり詳しい情報を得ることができます。概要を知りたいだけであればこれで十分ですが、レポートや卒業論文に使う場合や、研究として深く考察したい場合は、実際の史料や先行研究にあたる必要が出てきます。ここでは官職に関する史料の調べ方を中心にみてゆきます。

各王朝の正史（公式な歴史書）には、志と呼ばれる部門史の記述があります。志の一つに職官志や百官志など官職に関する部分があり（『漢書』のみ表）、各官職の沿革・職掌・官品が記されています。このため、官職に関する史料を確認するならば、まずは正史職官志の類を見るとよいでしょう。

范曄の『後漢書』は志がなかったのですが、後に司馬彪『統漢書』の志を補って今の形なっています。また『隋書』の志は別に作られた『五代史志』を取り込んだもので、梁・陳・北齊・北周・隋の五つの王朝の記録が含まれます。

王朝によっては正史に職官志の類がない場合があります。こういった場合は、上古から唐前半期までの官制の沿革を注にまとめている『唐六典』や、典章制度の沿革を記録した政書と呼ばれる『通典』（上古から唐後半期）、『文獻通考』（唐後半から南宋中期）、『続文獻通考』（南宋中期から明末）などの官職に関する部分で補うことができます。

『漢書』百官公卿表と『後漢書』百官志には、日本語訳注もあります。また『宋史』職官志には宮崎市定による詳細な解説があり参照すべきです。

⑤ 『『漢書』百官公卿表訳注』（大庭脩監修、朋友書店、2014）

『漢書』	百官公卿表（巻19）
『後漢書』	百官志（『統漢書』志24～28）
『晋書』	職官志（巻24）
『宋書』	百官志（巻39～40）
『南齊書』	百官志（巻16）
『魏書』	官氏志（巻113）
『隋書』	百官志（巻26～28）
『旧唐書』	職官志（巻42～44）
『新唐書』	百官志（巻46～49）
『旧五代史』	職官志（巻149）
『宋史』	職官志（巻161～172）
『遼史』	百官志（巻45～48）
『金史』	百官志（巻55～58）
『元史』	百官志（巻85～92）
『明史』	職官志（巻72～76）

表1 正史職官志類一覧

⑥ 『全譯後漢書』第9冊「百官志」（渡邊義浩主編、汲古書院、2013）

⑦ 宮崎市定「宋代官制序説：宋史職官志を如何に読むべきか」（『宮崎市定全集』10、岩波書店、1992（初出1963））

さて、『後漢書』（『統漢書』）の百官志によれば、尚書令は人事部門と行政文書全般を統括しているかなり重要な役職です。ただし、歴史の浅い官職のため、役職が重い割に官秩は千石とあまり高くありません。また、侍中については百官志に目新しい記述はありませんが、藤井律之「侍中領衛考」（『魏晋南朝の遷官制度』京都大学学術出版会、2013）によれば、皇帝の詔勅は皇帝と尚書令の印によって二重に封印した上で発信されることになっており、皇帝の印璽は侍中が所持していたので、侍中と尚書令の兼務は国家の情報発信を一手に掌握し得るものだったということがわかります。

なお、官僚にとって官秩や官品の等級を示す役割をもつ官職を本官といいます。ここでの荀彧の本官は尚書令であり、侍中は兼任です。ただし侍中の方が尚書令より上位（比二千石）の官のため、侍中・守尚書令の順になっていると考えられます。

（小島 浩之）

❖ 官品をしらべる

Q: 北宋の蘇軾^{そしよく}が左遷された後、政界に復帰する時の官職ほどの程度のランクのものだったのでしょうか？

蘇軾（1036-1101）は唐宋八大家の一人で、日本でも詩人・蘇東坡としてよく知られている人物です。一方で彼は、22歳で科挙に合格した優秀な官僚でもありました。ただし、蘇軾が官僚として政治の第一線にいた時代は、宋代で最も政治闘争が激しい時代であり、彼は人生で二度の左遷を経験し、二度目の左遷の際には中央政界に復帰することなく病により世を去ります。

蘇軾が一度目の左遷の時に作ったのが、後漢末の赤壁の戦いを題材とした「赤壁の賦」です。作賦の約二年後、皇帝が代替わりしたこともあり、彼は中央政界に復帰します。

この時期の蘇軾の官職は、元豊8年（1085）8月に朝奉郎・知登州となり、中央に召されて礼部郎中、さらに起居舎人にうつり、翌元祐元年（1086）に中書舎人となっています。そののち翰林学士にうつり、元祐三年（1088）には侍読を兼ねたとされています。

❖ 官職辞典類を検索する

まずはこれらの官職に関する職掌や官品などについて官職辞典類を使って調べましょう。前掲の通時代的な官職辞典（☞官職をしらべる）だけでなく、宋の場合は宋代官制を専門に扱った辞典もあり便利です。

- 『宋代官制辞典』増補本（龔延明編著、中華書局、2017）

この辞典は1997年の初版ですが、つい最近になって増補版が出されました。初版はCNKI工具書庫（☞CNKI工具書庫）でも検索できます。

宋代の官制は元豊年間（1078-1085）の官制改革の前後で大きく変わります（☞解説：官制の構造）。

蘇軾の最初の左遷は元豊二年ですから、復帰した時は元豊の官制改革後です。このため辞典類を確認する際には、北宋の元豊官制改革以後の記述を参照しなければなりません。たとえば朝奉郎は、元豊官制改革以前は散官で以後は寄禄官ですから、『宋代官制辞典』では両者で見出し語を分けて記述しています。今回は正七品の寄禄官について説明してある方を見なければなりません。

次に知登州は、このまま辞典の項目を探しても該当するものはありません。これは登州という州の長官を表しています。州の長官は総称して知州と呼ばれますが、「知+州名」で個別の州長官を表します。ですから、辞典の見出し語としては「知州」や「知某州」、「知某州事」などとしてあります。『宋代官制辞典』では「知某州軍州事」の見出し語に集約され、元祐の官品令では上州知州は正六品、中・下州の知州は従六品であると記載されています。蘇軾が知登州となったのは元祐改元の直前ですが、令はよほどのことが無い限り大改正されないで、ほぼ同じと考えてよいでしょう。そうであれば、登州は上州の等級なので知登州は正六品官ということになります。つまり、蘇軾は俸禄の等級としては正七品であるものの、実職は正六品官相当の官職に就いていたことになります。

その後、すぐに中央に呼ばれて礼部郎中となります。元豊の官制改革前は寄禄官名ですが、改革以後は職事官（差遣）名となり、国家の礼典を担当する責任者を指します。官品は従六品です。

蘇軾は短い期間で起居舎人さらには中書舎人と異動します。起居舎人は皇帝の言動を記録する担当者で従六品、中書舎人は詔勅起草の担当官（草制の官）で正四品です。『宋代官制辞典』では中書舎人も寄禄官と職事官の二種類の見出し語があります。元豊官制改革以後のことなので、職事官の

方が確認すべき項目になります。

翰林学士は中書舎人同様に草制の官ですが、中書舎人が宰相府からの命令で文書を起草するのにに対し、翰林学士は皇帝からの直接の指示で文書を起草します。つまり同じ草制の官でも翰林学士の方がはるかに格上で、宰相へと向かうエリートポストでした。

翰林学士に兼職として与えられた侍読とは、経筵官と呼ばれるもので、皇帝の学術的顧問というべき立場です。このように、蘇軾は復帰から3年ほどの間で、瞬く間に官階を駆け上がり、皇帝と直接に相対することができ、宰相のポストも射程に入るほどの位置を占めていたのです。

❖ 官職辞典以外からの情報

さて、これまでの説明は、実は官職辞典だけでは不可能です。なぜならば官職辞典の記述から、個々の官名に対応する職掌や官品への理解はできませんが、これらを官僚の履歴という時間の流れの中で考える場合、辞書の記述から相互の関連性を見出すことは難しいからです。この欠を補うために、この解説の執筆にあたっては次のようなことをしています。

まず、蘇軾の履歴を追うために、宋代の人物の伝記を簡便に知ることができる『宋人伝記資料索引』（解説：伝記資料）を活用しました。また、竺沙雅章『蘇東坡』（人物往来社、1967）をはじめとする各種の伝記類、官制の全体像を理解するために宮崎市定「宋代官制序説」（官職をしらべる）、梅原郁『宋代官僚制度研究』（同朋舎出版、1985）、熊本崇「宋神宗官制改革試論：その職事官をめぐる」（『東北大学東洋史論集』10、2005）などを参照しています。元豊官制以後の地方行政区画の等級は宋代の地理書である『元豊九域志』の記述を確認しました。

官職じたいは官品による階層構造の中に位置付けられ、法令で役割を規定されたものです。しかし、人間の肩書きとして官制の中でどのように作用しているかということになると、時代や昇進順序、さらには誰がその官職に就いたかなどで意味

が異なってきます。したがって、少しでも後者の立場から分析するようになると、官職辞典の記述ではもの足りない部分が出てきます。この段階になると、関係史料と先行研究をじっくり読み込むことが必要となってきます。

❖ 官品と官職の対応関係

今回の質問のように官職のランク（等級）を考える場合、絶対的な等級と相対的な等級があります。絶対的な等級は官品です。官品と官職の対応は正史職官志や政書類（官職をしらべる）を見れば書いてありますが、簡便なツールとしては兪鹿年編著『中国官制大辞典』下巻の「歴代官制表析」や、呂宗力主編『中国歴代官制大辞典』修訂版の「歴代職官品位表」があります（官制を視覚的に把握する）。

宋代に関しては前掲『宋代官制辞典』末尾に「官制表格」として、宋代の官制に関する詳しい図表が掲げられています。また『新編東洋史辞典』（清代の官職の官品や職掌をしらべる）の附録に散官・寄祿官と官品の対応表があります。

官品の上下で官職を並べた場合に、同品の官職が横並びとなります。これらは同じ官品であっても時代や社会の要請に応じて自然と重要度が異なってきます。その結果、同品の官職の中で、その重要度によって相対的な上下関係が生じてきます。これが先に触れた相対的な等級です。官品を縦軸とすれば、横軸方向の等級となるでしょうか。

官僚の人事はこういった二つの等級を指標として動いており非常に複雑です。なかでも相対的な等級は、官品のような絶対的な指標がなく、時期によって随時変化するので、これを理解するためには、前述のような、人間の肩書きとして官職がどのように作用しているかという観点から考えなければなりません。

こういった方向からの官制研究や官僚制研究はまだ歴史が浅く、これからの研究成果が十分に期待される分野です。

（小島 浩之）

❖ 官職の別称をしらべる

Q: 清代の学者、王昶の『柳南隨筆』を読んでいたら、「徐汝讓は大司空徐栻の従孫（兄弟姉妹の孫）だ」という記述がありました。が、「大司空」を官職辞典で調べても明清時代の官職ではないようです。これは何でしょうか？

❖ 倣古別称

『中国歴代官制大辞典』修訂版（呂宗力主編、商務印書館、2015）で調べてみると、古代の官職名であることがわかりますが、明清時代に関する具体的な記述はありません。

ただし、大司空の項目の5つ目の説明として工部尚書の別称とあります。実際に徐栻の伝記史料にあたってみると、明代の工部尚書（『明史』巻220）であったことがわかります。このように古代の官名を用いて官職を表現したものを倣古別称といいます。

なぜこのような面倒なことをするのかと言えば、伝統あるいにしへの制度に倣って表現する方が雅であると考えからです。直接的な表現は避け、知っている人にはわかるという婉曲的な表現を前近代中国の知識人たちは好んだのです。文章における修辞の一つと考えればよいでしょう。『柳南隨筆』によれば、明清時代の知識人は、特に司馬遷『史記』の記述に合わせて官名を表現することが多かったようです。また逆に、倣古別称が記録を混乱させる悪弊ととらえている知識人も多数いたようです。

こういった別称の事例はわが国にもあります。たとえば、徳川光圀のことを水戸黄門と呼ぶのは、光圀が権中納言であり中納言の別称が唐の官名である黄門侍郎だからです。

❖ その他の別称

一方で、古代ではなく同時代の正式の官職名や官名と名称がかぶってしまう別称もあります。

唐代は、宰相のことを「同中書門下三品」、「同中書門下平章事」と呼ぶことがあります。これは正式の宰相である三品官の中書省と門下省の長官と同列という意味であり、宰相の別称ということもできます。また、宰相府の正式名称は政事堂でしたが、徐々に中書門下という言葉が宰相府の別称として使われるようになります。このため史料上で「中書門下」とある場合、「中書省&門下省」を指す場合と宰相府を指す場合があります。さらに、門下省の政治力が徐々に後退してゆくと、中書門下を単に中書と簡略化して書くようになります（简称）。このため、史料上に中書門下や中書と書いてある場合、本来の官署を示すものなのか、宰相府の別称や简称なのか、前後の文脈で判断しなければなりません。

また、その官職に関する故事から、もしくは特徴を端的に表現したりデフォルメしたりして付けられた、通称や愛称とも言うべき別称があります。これらは場合によってはシニカルに、はたまたその逆で遊び心満載のしゃれっ気たっぷりに付けられています。

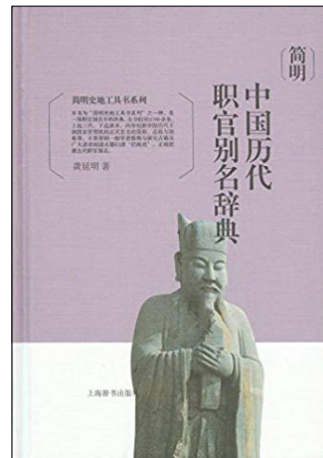
南宋・洪邁の『容齋四筆』巻15には、こういった別称の例が挙げられています。たとえば、監察官の侍御史は、端公・南牀・横榻・雜端・脆梨など、正八品上の文散官である給事郎は夕郎・夕拜、詔勅起草官の知制誥は三字、監査を行う役所である比部は昆脚皆頭など、その真の意味はわからなくとも、漢字を見ただけで苦笑いしてしまうようなものまであります。こういった例は漢代から清代まで数多くみられます。

❖ 別称の調べ方

最初に見た『中国歴代官制大辞典』のように、大概の官制辞典では別称についても情報を載せています。ただし、官職名の別称に関しては以下のような専門辞典もあります。

- ①『中国歴代職官別名大辞典』（龔延明著、上海辞書出版社、2006）
- ②『簡明中国歴代職官別名辞典』（龔延明著、上海辞書出版社、2016）

①は9,400条余りを収録する大辞典ですが、現在入手が困難です。②は①から2,700条あまりを精選して、50万字程度に収めたものです。一般的な別称は簡明版でも十分に対応できます。単に意味を説明するだけでなく、典拠も明示してあります。一般的な官制辞典では別称にまで典拠も含めた細かな説明のスペースは割かれていないので、こういった専門辞典は大変重宝します。



残念ながら、これらはCNKI 工具書庫 (CNKI 工具書庫) には入っていません。ただし、CNKI 工具書庫がコンテンツに含む官制辞典の解説の中で別称にまで触れていれば、情報を得ることができるかもしれません。

(小島 浩之)

❖ 官制を視覚的に把握する

Q: 官制や職官制度に関する記述を読んでも、官僚組織の具体的なイメージが湧きません。図や表で視覚的にわかるものはないでしょうか。

官制や官職といった職官制度に関する記述を読んだだけでは、官僚組織の具体的な枠組までイメージするのは難しいと思います。職官制度や官僚組織、さらには官制の全体像を理解するためには、図や表による可視化が最も有効な手段です。

❖ 組織図

官制を俯瞰しやすく可視化したものに組織図があります。組織図について、平凡社『世界大百科事典』第二版では次のように述べています。

経営組織の構造を図表化したもの。職位明細書または職務権限明細書と並んで組織便覧の中心をなすものであり、職能関係の構造および職位の権限関係の全貌を鳥瞰的に示すものである。したがって当然それには、明りょう性、正確性が求められる。組織図はその内容により機構図、指揮系統図、機能図に分けられるが、注意を要するのは、〈組織〉とこの組織図との関係である。両者を同一視することは避けなければならない。（下線は筆者）

前近代中国の官制の組織図は、同時代人ではなく現代の研究者が史料を読み解き可視化したものです。ある研究者の見地から可視化された組織図であって、実際の組織の全てが表現されているわけではありません。『世界大百科事典』からの引用文の下線部分にも関わりますが、組織図を利用する際には、この点、十分に意識しておきましょう。

●『中国歴代官制簡表』（衛文選著、山西人民出版社、1987）

●『中国歴代官制大辞典』 修訂版（呂宗力主編、商務印書館、2015）

『中国歴代官制簡表』は、殷から清に至る歴代王朝の詳細な官制の組織図（機構図）を樹形図で載せています。出版から30年以上を過ぎましたが、全ての王朝の官僚機構を図として概観できる大変便利なツールです。

一方で、『中国歴代官制大辞典』の附録二「歴代中央機構簡表」では、秦から清までの中央官制の機構図を表の形で整理しています。樹形図に比べて定員や職掌などを同一組上で表現できていますが、上下関係などの構造的な部分は樹形図の方がわかりやすいです。また中央官制に限定さ

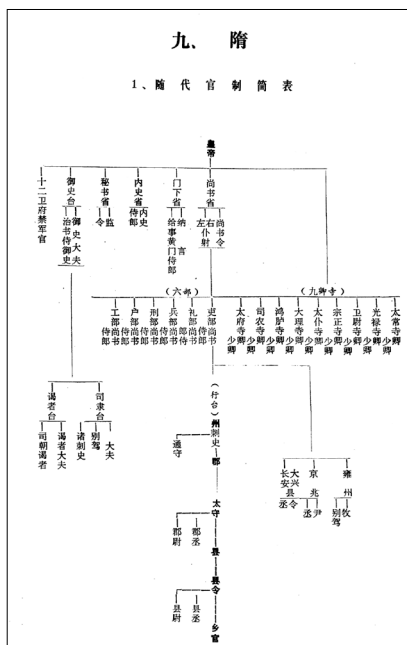


図1 『中国歴代官制簡表』の一部

れているのが難点です。日本語のものでは、『新編東洋史辞典』（京大東洋史辞典編纂会編、東京創元社、1980）の巻末に「中国歴代官職表」があります。

❖ その他の表形式のツール

官職名称の時間的変遷を追うものとして『歴代職官表』があります。これは清の乾隆年間に勅命により編纂されたものです。ただしこの本が72巻と膨大であったため、道光年間に黃本驥が簡略版を作成しました。勅命によるものを『欽定歴代職官表』と呼び、単に『歴代職官表』といえば簡略版を指します。『歴代職官表』は1965年に中華書局から出された点校本があり、2005年に上海古籍から復刻されています。

『歴代職官表』は清朝の官名を基本として、それらが上古から明の各王朝でどのような名称であるかを一覧にしたものです。このため清朝の考える官制の系譜に基づき、その変遷を可視化したものですから全てを鵜呑みにできません。

官制に関する一覧表としては、官職と官品の対応表や官僚の俸禄表なども考えられます。官職・官品・職掌の三者の対応を完備した表を掲載する

ものには兪鹿年編著の『中国官制大辞典』があります（編者の異なる同名辞典がありますから注意してください）。

- 『中国官制大辞典』上・下（兪鹿年編著、黒竜江人民出版社、1992）

この辞典は上下二巻からなっていますが、下巻の三分の一（329頁分）が「歴代官制表析」として、上古から民国時代までの各官職・品秩（官品もしくは官秩）・職掌等の対応を時代ごとに表形式でまとめてあり、職官制度の変遷を通覧するのに大変便利です。残念ながら1992年に出版されてから再版されておらず、国内の大学図書館も66館しか所蔵していません（2018年10月9日現在）。CNKI工具書庫（☞CNKI工具書庫）にも収録されていますが、本文と異なり図表は課金しないと全く見ることができません。

また、前掲『中国歴代官制大辞典』の附録一には「歴代職官品位表」があります。こちらは実際には野線のある表ではなく、各時代の官職を品秩ごとに羅列したもので、官品との対応関係をみるだけなら便利です。

官僚の俸禄に関しては、前掲『中国歴代官制簡表』の附録に、漢・北齊・北周・隋・唐・宋・元・明・清の官秩・俸禄が表形式で載せられています。

ここでは、通史的に図表を載せている参考図表を紹介してきましたが、同様のものは各時代を取り扱った歴史の概説書や、官僚制に関する研究書にも必要に応じて載っています。

こういった図表は、ざっと全体を把握するには大変便利なツールです。ただし、繰り返しになりますが、どれも作成者の主観が大きく反映されていますので、論文等で使用する場合には、最新の研究成果や典拠史料と照合し、その正否について自分自身で判断するようにしてください。

なお、図表により官制を可視化したオンラインツールで、学術の見地から公開されているものは管見の限りありません。

士 學 讀 侍	士 學 院 掌 院 林 翰	1
上外周 士史	天中内周 大史	三代
		秦
		漢
中夫中文 耶大學		後漢
		三國
史左矣 耶著鎮 國 作	今東矣 酒觀鼎韓 今東蜀 觀 祭文 觀	晉
耶著 作	作大 著	梁宋
學直 學士 學瑣 耶著 士省 士林 士史 作	作大 祭東 著	陳齊
耶著 作	作西 著	北魏
士館文 耶著 學林 作	非林例 作大 館文 著	北齊
士殿麟 天下外 學証 大史	天中内 大史	後周
耶著 作		隋
耶著 學侍集 作 士職院	學翰 承學翰 士林 旨士林	唐
	學翰 承學翰 承學翰 士林 旨士林 旨士林	五季
耶著 學侍翰 作 士職林	學翰 承學翰 士林 旨士林	宋
牙左 林	牙部翰 學翰 承學翰 承學翰 林林 士林 旨士林 旨牙面 林面	遼
學侍翰 士職林	學翰 承學翰 士林 旨士林	金
士職院國翰 耶侍史林	士職院國翰 百院國翰 學史林 承史林	元
士職院翰 學侍林	士職院翰 學林	明

図2 『歴代職官表』の一部(1965年中華書局本)

(小島 浩之)

解説：官制の構造

❖ 官と吏

中国において理想の国家とされる周の時代の支配者層には、王や諸侯の臣下の身分に「卿（公卿）一大夫一士」の三階層がありました。この臣下の三階層がのちの官の範囲に相当します。官の中での具体的な職掌が職で、あわせて官職とも言われます。官の機構を体系化したものが官制、官職を持つ者が官人（官僚）ということになります。

一方、非支配者層は庶人といい、庶人の中でも官の下請け的な業務を生業とする人々は、官と庶人の間にあって、のちに吏や胥吏と呼ばれるようになってゆきます。ただし日本語でも官吏という言葉が役人を指しているように、本来、官と吏は同義語であり、史料上では吏が官の意味を指すこともあります。

◆ 官僚の組織化

中国の歴代王朝は、国家運営を円滑に進めるために、階層的な身分秩序に基づいた官僚の組織化に腐心してきました。

漢代の官僚は俸禄の等級（官秩）により序列が決まっており、たとえば地方長官（郡太守）は秩二千石であることから、二千石がその別称にもなりました。

後漢末に九品官人法が施行されると、人物評価に基づいた官品が官僚の身分秩序となります。官品は、一品から九品までを正・従・上・下の語などで展開して、20 から 40 等級に階層化したものです。等級の最小単位を階といい、品と階の組み合わせからなる官制序列の指標を品階といいます。本来、階は人事昇進の基準となる単位で、官品とは異なる序列であったものが、北朝において九品官制の中に組み込まれていったものと考えられています。

九品官制において、三品以上は公卿、四・五品は大夫、六品以下は士に相当します。このため、

任用方法から待遇まで、三品以上、五品以上、六品以下で厳密な区別がありました。

王朝によっては、級・班・命などの語が品の代わりに用いられることもありましたが、いずれも考え方は九品官制と変わりません。官僚はこういった官品による階層的な身分秩序によって、清末まで組織化されていたのです。

図1は官と庶および九品官制を表わした模式図です。官の部分は、上位の官品になるほど資格対象者は限定され、一品を頂点としたピラミッド型の秩序が構築されていることがわかります。

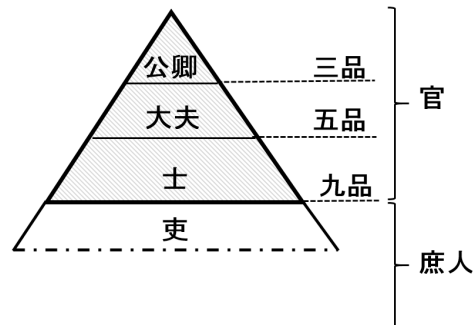


図1 九品官制の構造

官の世界では、給与や肩書はもちろんのこと、服装や席順などの礼制に関わる部分までもが官品によって左右されました。言い換えれば、官品による階層化された身分秩序と礼制による秩序維持が複雑に絡み合い、官僚の生活全般をコントロールしていたのです。

❖ 品階の表現

品階は、原則として数字による直接的な表現を避け、古代の官職名を使って表わします。これらは散官、散階、階官（以下、散官に代表させます）などといい、原則として品階を表すために機能し、文官・武官の二系統がありました。時代によって

変遷はあるものの、文散官の五品以上は「大夫」、六品以下は「郎」、武散官の五品以上は「將軍」、六品以下は「校尉」の呼称を用いることを原則としています。具体的には大夫を例にすると、光祿大夫、中散大夫などのように、品階により上記の呼称と組み合わせる名称が変化します。

❖ 職と散官の変遷

唐になると実職を有する官のことを職事官しよくじかんと呼ぶようになります。職事官は、職務の重要性や官としての歴史・伝統などに基づいて品階が決められていました。官僚は科挙などの資格試験に合格することでまず散官を得たのち、実務試験に合格してはじめて、先に得た散官の品階と同程度の職事官に任命される原則でした。

しかし、唐代後半期になり律令体制が崩壊してゆく過程で、時代や社会の変化に対応するため、官品とは直接に対応関係のない臨時の職が登場します。このように、律令の枠外に必要なに応じて置かれた職を令外官りようげのかんといいます。特に財政関係や軍事関係は重要性が増したため、これに関わる多くの令外官が登場します。代表的なものに節度使、塩鉄使などの使職や、皇帝のブレーンである翰林学士などがありました。また、使職の下には、幕職官ぼくしよくかんと総称される現地採用の令外官が多数おかれていました。

同じ頃、職事官の体系の中でも官と職との乖離りが進行してゆきます。たとえば、戸部尚書判吏部事こふぶしょうしよはんりふぶじなどというように、本来は財政長官である戸部尚書の肩書を持ちながら、実際には人事関係の責任者であるというようなケースが頻繁に見られるようになります。この場合の戸部尚書は官僚の等級を示す指標にすぎません。つまり職事官の階官化が進んで実際の職は別に与えられるものとなってゆくのです。ちなみに、日本語にもなっている知事という言葉は、「判吏部事」と同様で、本来「知県事」といって「県事(県の政治)をつかさどる」というところから出てきた名称です。

北宋になると、ついに職事官は給与の等級を示すだけの寄祿官きよろくかんとなり、実際の職務は差遣さけんと呼ばれるようになります。先の例で言えば、戸部尚書

が寄祿官、判吏部事や知県事が差遣です。官僚の肩書は、品階をあらわす散官、給与の等級を表す寄祿官、実職を表す差遣という組みあわせが基本となるのです。ただし、北宋の元豊年間に官制改革(元豊の官制改革)が行われると、散官が品階と俸祿の等級の指標となり、これを新たに寄祿官と呼ぶようになります。このため、宋代の官僚の肩書は、元豊以前か以後かで同じ官名でも全く意味が異なるので注意が必要です。

以後、中国では、元豊官制が基軸となり、各王朝において随時の修正が加えられてゆきます。品階を示す(時代によっては俸祿の等級の意味も含む)散官と、実際の職務内容を示す職の二つが存在する構造は原則として変わりません。ただし、徐々に散官は意味を失ってゆき、職の官品が官僚の等級を表すものとなってゆきます。そして散官は、封贈という名誉的に与えられる称号のような位置付けになってゆきます。生前に与えられる場合が封、死後に与えられる場合が贈で、両者をまとめ封贈といえます。

❖ 爵と勲

官僚が帯びる主要な肩書には、散官と職のほか爵と勲があります。爵(爵号)は官品の登場以前から存在する礼に基づいた階層的な身分秩序です。官僚が与えられる爵に限って言えば、大夫以上で封建された者(封爵)や功績ある者に与えられます。このため、九品官制が成立すると五品(大夫)以上の官品と関連付けられることとなります。伝統的な「公・侯・伯・子・男」の五等級が知られていますが、これを基本として王朝によって様々な等級付がなされていました。爵位は世襲できますが、子孫の爵位は父祖のものより下位の等級となり、世代を重ねるたびに等級が下がります。

勲官は、西魏に淵源をもち唐代に成立したものです。軍功によって与えられる官でしたが、濫発によって庶人にも多く与えられたため、一介の兵卒じようちゆうこくが上柱国(正二品)や柱国(従二品)を持つことも珍しくありません。このため価値が低下し、元豊の官制改革で一度は廃止されますが、そ

の後復活して明までは存在します。

❖ 胥吏

中国古代において、官は中央政府の資格審査による採用、吏は各官庁の判断による現場採用という違いがあったとされています。

九品官人法により人物評価が官への採用の基礎となるしくみができ、官品の体系が整備されるようになると官と吏の身分差は大きく開いてゆき、官すなわち九品官制の枠内を流内、九品官制の枠外の吏を流外と呼ぶようになります。

唐代になると、吏は二品から九品まで8段階の流外品で階層化されて九品官制の下に位置付けられ、各官署で書記などの役割に従事しました。このほか、流外品をも有しない雑任や職掌人と呼ばれる存在や、吏と官の中間的な存在として特定技術により採用された技術官などもありました。流外品の吏は、成績評価と試験によって官への採用も可能となっていました。九品官となってもその昇進は著しく制限されており、科挙などの資格試験による採用者と比べて差別されていました。

しかし、官僚が人事異動で頻繁に交替するのに対して、吏は長期で当該官庁での業務にあたることから、徐々に実務面で吏の力が大きくなってゆきます。宋代になると吏は胥吏と呼ばれ、官から俸給を与えられず、一般民衆が役所での手続に必要な書類を代行して手数料を徴収し生活の糧とします。この結果、ほしいままに法外な手数料を要求したり、手数料じたいが賄賂化したりするなどし、「胥吏」は民衆を苦しめる代名詞のようになりました。こういった行政における胥吏の暗躍と、政府の取り締まりのイタチごっこが、清朝まで続くのです。

❖ 参考文献

- ①『支那官制発達史：特に中央集権と地方分権との消長を中心として』影印版（和田清編、汲古書院、1973）
- ②谷井俊仁「官制は如何に叙述されるか：『周礼』から『会典』へ」（『人文論叢：三重大学人文学部文化科学研究紀要』23、2006）
- ③『九品官人法の研究：科挙前史』（宮崎市定著、中央公論社、1997、初出は1956）
- ④『中国歴代官制』増補本（鹿謂慧 [ほか] 主編、齊魯書社、2013）

中国の官僚制度の全般について日本語で扱っているものはあまり多くありません。①は1942年初版の影印で、殷から元までの官僚制度の通史です。明・清については別に刊行される予定でしたが未刊に終わっています。古い研究成果ではあるものの、現在もその価値を失っていない部分があります。②は歴史書の中で官制がどのように記述されてきたかを、その思想的背景も含めて論じています。学術論文なので初学者には難しい部分もありますが、官制叙述の通史として他に類をみないものです。③は九品官制の成立に関する基本文献です。元々胥吏の研究から発展したものであることから、胥吏や技術官などへの目配りも十二分になされています。中国語による官制の概説書は多くありますが、新しいものでかつ記述が詳細ということでひとまず④を挙げておきます。

このほか図表で官制を概観できる参考図書もあります。これらについては「官制を視覚的に把握する」に詳しく紹介していますので、そちらを参照してください。

（小島 浩之）

解説：官僚登用制度

❖ 任子（恩蔭）・選挙・辟召

官僚の登用制度を知る上で理解しておかなければならないのが、任子・選挙・辟召の言葉です。

任子とは、父祖の官職によって子孫が官僚に登用される制度のことで、唐からは蔭や恩蔭などと呼ばれます。一定以上の官品（漢では官秩）を有する官職に就けば、子孫は自動的に父祖より下位の官職に任じられるよう制度設計されていました。つまり任子とは、血縁関係を基礎とした世襲制度ともいえるでしょう。

選挙とは官僚登用制度のことだと説明されますが、狭義には任子のような世襲を除き、推薦や試験による選抜制のことだと考えて差しつかえありません。

辟召は、各官庁の長官の裁量範囲内で下僚を採用できる制度です。任子や選挙が高級官僚を中心とした登用制度であるのに対して、辟召は下級官僚の登用制度です。

❖ 漢から唐へ

漢では任子制度に加えて、有用な人材を中央の高官や地方長官が推薦することにより官僚に登用する制度が整備されました。儒教の考え方に則って、該当者の郷里における評判を重要視する建前であったことから、この方式を郷举里選きょうきょりせんと呼びます。郷举里選には推薦理由の別により、秀才・孝廉など複数種類あり、これが後に科挙の試験科目の種別へとつながってゆきます。

後漢末に九品官人法が施行されると、人物評価は各地に置かれた中正官（中正）の職務となります。中正は管轄下の地域において、人物を評価した上で一品から九品に格付けしました。この中正により付される官品きょうひんを郷品と呼びます。郷品を得た人物は、中央政府により郷品から四級下の官品の官職に任じられるしくみでした。たとえば郷品が二

品であれば、六品の官職に任官されたのです。

南北朝時代になると、南朝は九品官人法に基づきますが、北朝では官僚の身分指標として官品を用いつつも、漢代以来の秀才・孝廉などを復活させます。

隋の文帝は、589年に陳を滅ぼして中国を統一すると、中央集権化のための大規模な制度改革に乗り出します。この改革の一環として、九品官人法が廃止されて科挙がはじまり、また下僚の辟召も廃止となり、人事権が中央政府に一元化されたのです。辟召は、唐後半期に幕職官の採用のために復活しますが、宋以降、再び見られなくなります。

❖ 科挙

科挙とは科目による選挙という意味で高等文官資格試験のことです。なお、武官には別に武科挙と呼ばれる制度がありました。

科挙の制度は隋・文帝の598年（諸説あり）にはじまり、宋代に大きく発展を遂げました。元の初めに30年ほど実施されない期間はあったものの、清朝末の1905年まで約1,300年間続いたのです。科挙には複数の試験科目が用意されており、当初は儒教の經典に関して出題される明経科が重要視されました。しかし、徐々に論文試験である進士科の重要性が高まり、北宋・王安石の科挙改革によって進士科以外が全廃されると、「科挙合格者＝進士」という図式が一般的なものとなります。なお、進士のなかでも一番の成績を収めた者にはじょうげん状元の称号が与えられました。

科挙は地方試験（唐は郷貢、宋は解試、明清は郷試）に合格したのち、中央での試験（唐は貢舉、宋は省試、明清は会試）を受けることとなります。唐代では貢舉に合格しても、次に人事担当官庁である吏部が実施する選考試験（吏部試）が待ち受けており、吏部試に合格しなければ実職に就くことはできませんでした。科挙が採用試験ではなく、官僚とな

るための資格試験だと説明されるのは、本来このような制度であったためです。

宋になると省試の後に皇帝直々の試験である殿試が加わります。これにより、最終選考を皇帝みずからが行った上で合格者に官職を授けるようになり、資格試験と採用試験との境界が徐々に曖昧になってゆきます。また、北宋中期からは科挙の実施が三年に一度となりました。こういった「地方—中央—皇帝」という三段階からなる試験制度や三年ごとの実施は、科挙制の基本的な枠組みとして定着し、明清時代へと受け継がれてゆくのです。

試験制度と教育を組み合わせ、優秀な官僚を養成するために設けられたのが学校です。ただし、唐代は高位高官の子弟しか入学できず、魏晋南北朝以来の家柄や血統を重視する気風を、科挙制度・学校制度ともに脱却できませんでした。宋になると、地方には官学のほか私学（書院）も多く設立され、学校は地方の優秀な人材を中央へと送り込むための重要なルートとなってゆきます。ちなみに、宋代における木版印刷の隆盛が、こういった学校制度や受験制度の影響であることも忘れてはなりません。

明清時代は地方学校の生徒であることが、科挙受験の条件でした。このため明清の学校の入学試験や定期試験は、科挙の試験の一段階として組み込まれてゆきました。府・州・県など地方学校の生徒のことを生員せいえいんといい、彼らが郷試に合格すると挙人きよじんと呼ばれました。清になると、こういった学校試に加えて、郷試以後の各段階に小試験や予備試験が設けられ制度が複雑化してゆきます。受験生たちは最終的な合格に至るまでに、過酷とも言える競争を強いられました。何度も受験に失敗し、進士に合格した時には既に老人であったという逸話はよくあります。途中で挫折した者はその時点に見合ったポストが用意されており、これを得て中・下級官僚として生涯を終える者も多数

いました。ただし、ひとたび進士になれば高位高官への出世が約束されているわけですから、立身出世を望む教養人の多くが、進士を目指して日々研鑽を積んだのです。

❖ 参考文献案内

日本語で書かれた中国の官僚登用制度の研究書や一般書で、まず参照すべきものを挙げるならば以下の5つでしょう。

- ①『九品官人法の研究：科挙前史』（宮崎市定著、中央公論社、1997、初出は1956）
- ②『科挙史』（宮崎市定著、平凡社、1987）
- ③『科挙と官僚制』（平田茂樹著、山川出版社、1997）
- ④『科挙：中国の試験地獄』（宮崎市定著、中央公論社、1963）
- ⑤『科挙と中国近世社会：立身出世の階梯』（何炳棣著；寺田隆信・千種真一訳、平凡社、1993）

①は九品官人法の成立からその展開を実証的に論じた学術論文ですが、前後の時代との関連を重視しているので漢から唐への官僚登用制度の流れを概観するのに便利です。②は元々戦前に執筆されたものを平凡社東洋文庫の一冊として復刻したものです。清代の科挙についての研究ですが、その前提として唐から明までの科挙についても詳しく論じています。③は宋代の科挙と官僚制度についてまとめたブックレット、④は清代の科挙についての一般書で、中公新書の一冊として刊行された後に中公文庫にも収められました。⑤は明清時代の科挙と官僚制、さらには社会構造に至るまで幅広い観点から論じたものです。社会史的視点からの史料分析によるデータを駆使した力作です。

なお、上記のうち①は『宮崎市定全集』（岩波書店）第6巻に、②と④は第15巻に収録されています。

（小島 浩之）

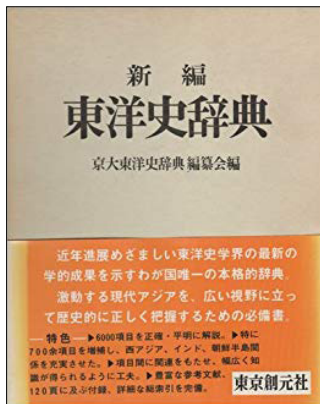
❖ 清代の官職の官品や職掌をしらべる

Q: 清代の内閣中書はどのような官職でしょうか？また、他の官職との関係はどうなっているのでしょうか？

❖ 概説的な工具書にあたる

中国の歴代の官職と対応する官品についての概況を知るためには、手頃な工具書にある記載を最初にあたってみるとよいと思います。

- 『新編東洋史辞典』（京大東洋史辞典編纂会編、東京創元社、1980）



本書の巻末には、歴代の主要官職沿革表が掲載されており、漢代、唐代、宋代、明代、清代の状況を掴むことができ便利です。また、時代ごとの官職の変遷を理解する上でもっとも基本的なことを把握できます。しかしながら、宋代が職員と官品との対照がわかるのに対して、清代の部分については対照する官品が記述されていないため、中央官職と地方官職の概要と、職位間の上下関係が判明するにすぎません。

清代について手っ取り早く知るには、他の工具書に当たる必要があります。以下で紹介するものも、東洋史学（特に中国史）専攻の院生は必携のものです。

- 『中国歴史公文書読解辞典』（山腰敏寛編、汲古書院、2004）

本書は清末民国期の中国における公文書を読解するために編まれた工具書であり、独特の用法と体裁をもつ当時の公文書を正しく読むために大変便利です。非常に行き届いた工具書で、付録部分には、読解のための手引きや各種対照表も収録されています。2～3頁に「主要品級一覧」があり、先に見た『新編東洋史辞典』収録の「清代の官職表」よりも詳しく、職位と官品との対照表がつけられているのが便利です。ここで内閣中書とは内閣に設けられた従七品官であったことがわかります。簡単に確認する程度であれば、これらの情報で大づかみにしておけばよいでしょう。

品級	正一品	从一品	正二品	从二品	正三品	从三品	正四品	从四品	正五品	从五品	正六品	从六品	正七品	从七品	正八品	从八品	正九品	从九品
宰相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相	丞相
内閣中書																		
翰林院																		
御史																		
知府																		
知縣																		

❖ さらに詳しくしる

さらに詳しい情報を得るためには、上述の工具書が参照した史料や書籍などに当たる必要があります。『中国歴史公文書読解辞典』が引用する「清代の官制」『官場現形記・上』（中国古典文学全集第

よる検索（「人名査訊」、官職名による検索（「官職査訊」）に分かれています。



『清季職官表附人物録』は清代職官資料庫に包括されているので、データベースは一つで事足りるようにも見えますが、清季職官表查詢系統では着任・離任時期、離任の理由などもわかる点が清代職官資料庫にない特徴です。

❖ Web 検索してみる

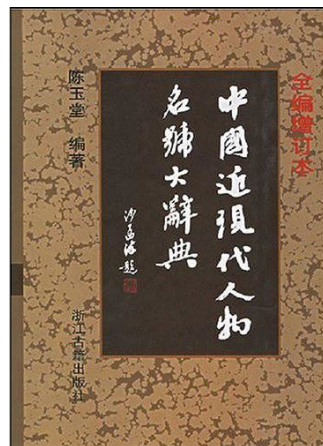
しかしながら、このデータベースを使ってすぐに気がつくのは、収録範囲が上級の職位にあったものに限定されているので、例えば、彼が最初に任官したのはどの職位なのか、どれくらい任官したのかなどの情報は含まれていないことです。さしあたって、「丁日昌」をキーワードとして Google 検索を行ってみると、Wikipedia の英語版、中国語版、日本語版に項目がたっており、前 2 者の記述が比較的詳しくなっています。この記載によれば、彼は 20 才で生員になったのち、郷試には及第しなかったものの幕僚として政界に入り、太平天国軍撃退の軍功によって、瓊州府学訓導に任命され、次いで江西省廬陵の知県となります。したがって、彼の官界での職歴は瓊州府学訓導からはじまったこととなりますが、「清季職官表查詢系統」ではその点はわかりません。また、曾国藩の幕僚を経て、同治 3 年に蘇淞太兵備道に、翌 4 年には兩淮塩運使に就任したこともこのデータベースからは判明しません。さらに注意しなく

てはならないのは、複数の職を兼任した場合、それがデータベースに反映されていない可能性があることです。彼の例でいえば、福建巡撫在任時に、福建船政大臣と台湾学政を兼任している点にそれにあたります。

❖ 人名辞典にもあたる

したがって、丁日昌の官歴の全体像を把握するには、人物事典をあたってみるのが不可欠になります。近現代の人物について調べるのに最も便利なのが以下の工具書です。

- 『中国近現代人物名号大辞典（全編修訂本）』（陳玉堂編、浙江古籍出版社、2005）



この工具書は書名にもある通り、号や筆名からも検索をすることができます。なお、CNKI 工具書庫 (<http://gongjushu.cnki.net/rbook/>) には本書を含めた類書が収録されているので、こちらも有用です。

上に紹介したデータベースや工具書からでは十分に明らかにならない経歴は、彼の出身地の新編地方志に人物伝がある可能性もありますが、ない場合には清代職官資料庫が土台とした各種一次史料を丁寧にあたって追跡することが必要となります。

（佐藤 仁史）

❖ 中華民國時期の人物の官職をしらべる

**Q: 張作霖政権下で民政を担った王永江の官職歴は、どのよう
にしらべたらよいのでしょうか？**

❖ オンラインデータベースにあたる

オンラインで中華民國期の官職について調べるには、国立政治大学が作成した「中華民國政府官職資料庫」が便利です。

- 中華民國政府官職資料庫 (国立政治大学)
(<http://gpost.ssic.nccu.edu.tw/index.php>)

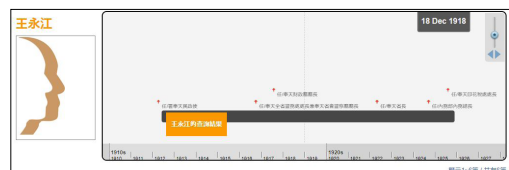
紹介文によれば、本データベースには中華民國体制下において各時期の総統府によってなされ、政府公報に掲載された人事情報が集められています。ここでいう人事情報には文官と武官ともに含まれています。収録されている政府公報と発行時期とは次の通りです。

- ①『臨時公報』：辛亥年 12 月 26 日（民国元年 2 月 13 日）～民国元年 4 月 26 日
- ②臨時政府公報：民国元年 1 月 29 日～ 4 月 5 日臨時政府結束
- ③政府公報：民国元年 5 月～民国 17 年 6 月
- ④軍政府公報：民国 6 年 9 月 17 日～民国 7 年 5 月 14 日
- ⑤陸海軍大元帥大本營公報：民国 11 年 1 月

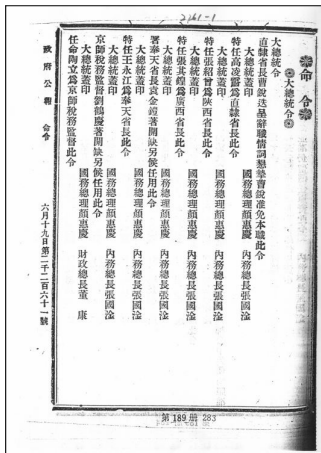
30 日～民国 14 年 5 月 20 日

- ⑥国民政府公報：民国 14 年 7 月 1 日～民国 37 年 5 月 19 日
- ⑦總統府公報：民国 37 年 5 月 20 日以降（現在に到る）

それでは、さっそく「王永江」で検索をしてみると、彼が、民国元年から民国 14 年にかけて、署奉天民政使、奉天全省警務處處長兼奉天省會警察庁庁長、奉天財政庁庁長、奉天省長、内政部内務総長、奉天印花稅處處長に任命されたという結果が表示されます。このデータベースの特徴は、各官職の着任時期をタイムテーブルの上に表示することで、直感的な理解ができることです。



また、情報の末尾には、「源自」という欄があり、根拠とされた公報の画像ファイルにリンクが張られていて便利です。ところで、奉天系軍閥と言えば、張作霖に反旗を翻した郭松齡がよく知られています。そこで、彼がどのような官職にあったのか検索してみたところ、ヒットしませんでした。つまり、歴史上の人物としては有名ではあったものの、総統府によって任命された官職にはなかったということが判明したのです。このような人物の場合、他の工具書に当たる必要があります。特に、北洋政府期の軍閥政権の官職を理解するには、一次史料に丁寧にあたるしかありません。



❖ Web 検索してみる

上述のデータベースは単に各時期の總統府によって任命された官職の任命時期が判明するだけであるので、ある人物の官職に関連する活動についてはやはり人名辞典を調べる必要があります。概要を大雑把に掴むだけであれば、先に検索した二人とも人名をキーワードとした Google 検索によって、Wikipedia に項目が立っていることはすぐに確認できます。しかし、工具書としての信用度の問題から、専門家によって編集された人物事典や関連する一次史料に当たる必要があることはいうまでもありません。

❖ 人名辞典にあたる

近現代の人物については清代の欄で挙げた『中国近現代人物名号大辞典（全編修訂本）』（陳玉堂編、浙江古籍出版社、2005）が便利です。それ以外にも、民国の人物については、類書がいくつかあります。例えば、中国社会科学院近代史研究所『民国人物伝』1-12 卷（中華書局、1978-2005）や徐友春主編『民国人物大辞典 増訂版』（河北人民出版社、2007）、劉国銘主編『中国国民党百年人物全書』（團結出版社、2005）などがそれにあたります。

ところで、人物を手がかりにではなく、ある特

定の官職に誰がついたのかについて調べる場合、先に紹介した・中華民国政府官職資料庫から「任意詞」の欄に検索したい官職名を入力して検索する方法もありますが、紙版の工具書について言えば次のものがよく知られています。

- 『民国職官年表』（劉寿林ほか編、中華書局、1995）

この本は、1912 年 1 月の南京臨時政府成立から 1949 年 10 月の南京国民政府終了までを範囲として、中央政府については、総統、各部の総長と次長、司局長、参事、主席、五院の院長、委員、各部の部長、司長、および歴代の議員について、地方政府については、都督、省主席、将軍、督軍、省府委員、行署主任などの歴任者が明らかにされていて便利です。巻末には人名録と字索引が付されていて、それぞれ人名や字号から検索することができます。

9. 江蘇省軍政民政司法職官年表

職名	年次	民國七年(1918年)	民國八年(1919年)
都督	李純	李純	
副都督	李純	李純	
江蘇鎮守使	曹錕	曹錕	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	
常州鎮守使	朱瑞	朱瑞	
徐州鎮守使	朱瑞	朱瑞	
鎮江鎮守使	朱瑞	朱瑞	
南通鎮守使	朱瑞	朱瑞	
蘇州鎮守使	朱瑞	朱瑞	

新聞資料をしらべる

❖ 「満洲」の新聞をしらべる

Q: 日本敗戦前後の満洲、特に長春の日本人の状況についてしらべていますが、どのような新聞・雑誌資料がありますか？

この問題は、①「満洲国期」（以下、括弧を省略）における日本語新聞の刊行状況、②敗戦前後における日本語及び中国語新聞の刊行状況、③中華民国に復帰後の中国語新聞に分けて考える必要があります。

❖ 大規模データベースに含まれない日本語新聞

近年の中国近現代史研究を取り巻く学術環境の変化として、大規模商用データベースの登場が挙げられます。そのうち枢要の位置を占めているものに、新聞雑誌データベースがあります。

- 全国報刊索引 (<http://www.cnbkisy.com/>)
- 大成 老 旧 期 刊 全 文 数 拠 庫 (<http://laokan.dachengdata.com/tuijian/showTuijianList.action>)

前者には「晚清期刊全文数拠庫(1833-1911)」と「民国期刊全文数拠庫 (1911-1949)」などに、上海図書館所蔵の新聞、雑誌が全文で収録されています。後者は清末民国期の雑誌データベースです。また、中国近代史研究によって不可欠の基礎史料といっても過言ではない『申報』も愛如生の『中国近代報刊庫』の一部として公開され、現在では広く使用されています。しかしながら、これらの商用データベースは租界や植民地・占領地で刊行されていた日本語新聞は含まれていません。

もちろんこれらのデータベースからも関連情報を得ることができます。例えば、「日僑」「遣送」

などのキーワードなどから検索をすると、満洲国崩壊後の日本人引き揚げに関連する記事が少なからずヒットします。

❖ 満洲国期における日本語新聞の刊行状況

この点については、国立国会図書館のウェブページに詳細な紹介があるので、屋上屋を架すのは避けます。

- 国立国会図書館リサーチナビ (https://rnavi.ndl.go.jp/research_guide/entry/theme-asia-49.php)

当該記事から明らかになるのは、満洲国における出版の中心地は大連や奉天であり、長春を拠点とした新聞はあっても刊行期間が極めて短かった



という事実です。また、国立国会図書館に所蔵されているもので判断すると1944年初頭までしか刊行されておらず、これらからはアジア・太平洋戦争下の満洲国の状況を知ることはできても、敗戦前後の満洲国において日本人がどのような状況に置かれていたのかを知ることができません。したがって、ほかの情報源にあたる必要があります。

❖ 敗戦前後における日本語及び中国語新聞の刊行状況

敗戦後より引揚げに至るまでの時期の日本人の状況については、ソ連軍侵攻後に東北地方の大部分を接収した南京国民政府が設立した東北日僑管理処による新聞があります。

- 『東北導報』（長春版）58号～151号（1946）^[1]



ここからは、日本人居留民の引揚げへの動き、国民政府による各種政策、外地から流れ込んできた避難民の救済問題などを明らかにできます。

ただ、これらは刊行機関も短く残存している号も多くないので、同時に後で取り上げる中国語新聞やその他の編纂物と対照させることが不可欠になります。真っ先にあたるべきは、『満蒙終戦史』と『満洲国史』です。これらは1950年代に関連する種々の史料（日記や手記を含む）や関係者への聞き取りをもとに編纂されたもので、新聞記事のより深い読み込みや新聞記事に登場しない事項を

知るのに便利です。

❖ 中華民国復帰後の中国語新聞

日本敗戦後、中華民国に復帰した東北三省の状況を知るてがかりとして中国語新聞があります。東北地方全体の動向を知るのに便利なのが『東北日報』（出版地は瀋陽）です。ただ、注意しなくてはならないのは、新聞の政治的な立場です。中国国民党系なのか、中国共産党系なのか、独立系なのかによって、報道の仕方が相当異なるからです。例えば、中国共産党が進めていた土地改革についても政治的な立場によって、場合によっては正反対の評価が報道されたこともあり得ます。なお、『東北日報』は中国共産党によって発行された機関誌としての性質が濃厚な新聞です。

東北地方の具体的な地域に即した状況を知りたい場合は、市や県レベルの名前を冠した地方紙を調べる必要があります。東北地方で刊行されていたこれらの地方紙は、現段階では大型商用データベースに収録されていないので（現在入力中で将来収録される可能性は排除できません）、閲覧をするには、中国国家図書館の縮微閲覧室に赴く必要があります。



注

- [1] 長春版は、加藤聖文 監修・編『海外引揚関係史料集成・補遺篇』補遺第3巻（ゆまに書房、2002）に収録されています。なお、奉天版は、1号～457号（1946～1947、欠号あり）が国立国会図書館に所蔵されています。

（佐藤 仁史）

仏典をしらべる

❖ 仏典の叢書・テキスト

Q: 仏教經典について調べようと思って図書館に行ったら、『〇〇大蔵經』『△△全書』のような本が何百冊も並んでいて、途方に暮れています。

❖ 「八万四千の法門」?!

たしかに仏教には、「八万四千の法門」とよばれるぐらい、たくさんの經典・文献があります。またそれらを取録した叢書(☞解説:叢書)も数多くあります。図書館にずらりと並んだ仏典を前にして、仏教研究者以外の方が、どれを使えばいいのか途方に暮れてしてしまうのは当然のことでしょう。

ここでは中国仏教研究でよく使われる仏教叢書(大蔵經・一切經)について、簡単に紹介したいと思います。

❖ 代表的な叢書とその特徴

『〇〇^{だいぞうきょう}大蔵經』『△△全書』と題された仏教叢書はたくさんありますが、中国仏教に関する文献を調べたいのであれば、まずは『大正新脩大蔵經』と『大日本統蔵經』から始めるべきでしょう。この二つは、データベース化も進んでおり検索しやすいので、文献を探したり全文検索をしたりすることも容易です(☞仏典の検索方法)。

◆『大正新脩大蔵經』

『大正新脩大蔵經』(1923-1934、『大正蔵』『正蔵』などと略す)は、近代的な文献学によって編纂された初めての大蔵經であり、世界中の仏教研究者が使用している、現在もっとも普及している大蔵經です。

和装本と洋装本がありますが、和装本はめった

に目にすることはありません。洋装本にはハードカバー版とソフトカバー版(普及版)がありますが、後者は前者にあった誤字等の修正がなされています(ただし、網羅的な修正ではありません)。そのため、全体から見ればわずかですが、内容が異なります。ですので、どちらかと言えばソフトカバー版(普及版)を用いたほうがいいでしょう。

全体で100巻ありますが、そのうち前半85巻が漢文仏典で、インド・中国(朝鮮半島)・日本の仏典が収録されています(☞解説:仏典の分類)。これに加え、図像部12巻と、仏典目録である昭和法宝総目録3巻で、合計100巻となります。現在、図像部も含めてデジタル化されており、オンラインでの閲覧や検索が可能です(☞仏典を検索する)。

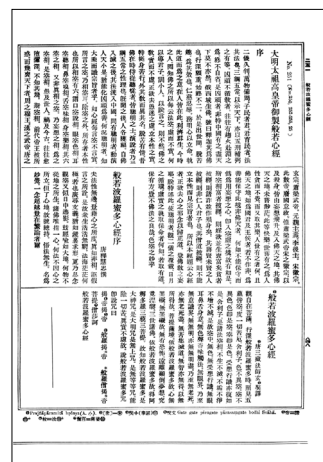


図1 『大正新脩大蔵經』

『大正新脩大藏經』は、主に『高麗大藏經』を底本とし、他の版本・写本、そして一部の文献についてはサンスクリット本などの校訂情報を明示しているなど、他の大藏經と比べれば学術的な信頼性は高いです。ただし、校正が不十分な文献があったり、返り点・句読点などの付け方が間違っている文献もあったりと、厳密な研究に用いる際には底本である『高麗大藏經』などをチェックする必要があります(後述)。今から百年近く前、大正時代から昭和初期にかけて編纂されたものだという事は忘れないでおきましょう。

論文などに引用をする際の出典の書き方は、日本印度学仏教学会が指定する執筆要領^[1]に準拠するのがよいでしょう。仏教学の論文で「大正」や「T」という略号があれば、『大正新脩大藏經』のことを指すのは常識となっています。

日本や中国では従来、「大正四五・二七一上」のように、

“大正” + 卷数 + “・” + ページ番号 + “上”“中”“下”段

という形式で書くことが多いです。図1のように、『大正新脩大藏經』は1ページ三段組になっていますから、巻数・ページ・段を書くのが一般的でした。日本印度学仏教学会の執筆要領でも縦書き論文の場合はこの形式で書くよう指定しています。

一方欧文の論文では、「T1861, 45, 271a8」のように、

“T” + 文献番号 + “;” + 巻数 + “;” + ページ番号 + “a”“b”“c”段 + 行番号

と行まで書くことが多く、日本印度学仏教学会の執筆要領では横書きの場合この形式で書くよう指定しています。ちなみに複数行に渡る場合には、「T1861, 45, 271a8-c1」のように、

“T” + 文献番号 + “;” + 巻数 + “;” + 開始ページ番号 + 開始“a”“b”“c”段 + 開始行番号 ((+ 終了ページ番号) + 終了“a”“b”“c”段) + 終了

行番号

となります(終了ページ番号・段は、開始ページ番号・段と重複する場合には省略)。近年、英語論文が増えていますので、後者の形式が徐々にスタンダードになっていくかもしれません。

◆『大日本統藏經』

中国仏教研究に欠かせないのが、『大日本統藏經』(『統藏經』『統藏』などと略す)です。一部重複していますが、『大正新脩大藏經』に収録されていない中国撰述仏典を多数含んでいることから、中国仏教研究で広く用いられています。ただし、『大正新脩大藏經』よりも前に刊行されたこともあって、底本や校訂の情報が不明確であるなど、学術的には問題点も多くあります。

『大日本統藏經』を使う上で注意しなければならないのは、三つのバージョンがあることです。

- 1.『大日本統藏經』(全150套、藏經書院、1905-1912)
- 2.『卍統藏經』(全150冊、新文豊出版、1976)
- 3.『新纂大日本統藏經』(全90巻、国書刊行会、1975-1989)

1が初版で、和装本です。2は1を影印し、いくつか新しい文献を追加して出版した洋装の海賊版です。本来なら2ではなく1を使用すべきでしょ

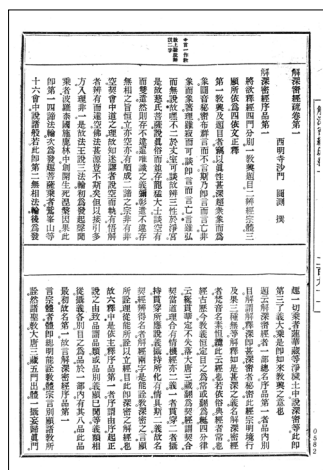


図2 『卍統藏經』

うが、1を所蔵している図書館に限られていることもあり、個人でも入手しやすい2が世界中で普及しています。3は文献の配列や組版を大幅に変えてあり（1・2は二段組だが、3は三段組）、校訂等も1・2と比べればしっかりとしていますが、2と比べると普及はしていません。

引用する場合にも、それぞれ出典の示し方が異なります。1の場合は、「第一編・第四十二套・第四冊」のような書き方になります。2・3は『大正新脩大蔵経』と同様、巻・ページ・段の表記となります。

『大日本統蔵経』もCBETAでデジタル化されていますので、全文検索が可能です（☞仏典を検索する）。

◆ 影印刊行されている大蔵経

『大正新脩大蔵経』や『大日本統蔵経』は近代の日本で刊行された大蔵経ですが、中国にも多数の大蔵経が伝わっており、近年出版が相次いでいます。大きな図書館では『大正新脩大蔵経』などと並んで配架されているところも増えてきていますので、簡単に説明しておきましょう。

中国の大蔵経は大きく分けて開宝蔵系、契丹蔵系、江南諸蔵系の三系統があると言われ、それぞれに以下のような大蔵経が知られています（『高麗大蔵経』は高麗のもです）^[2]。

- 開宝蔵系…『開宝蔵*』（一部）『高麗大蔵経*』『趙城蔵（金蔵）*』など。
- 契丹蔵系…『房山石経*』（一部）など。
- 江南諸蔵系…『崇寧蔵』『毘盧蔵』『磧砂蔵*』『普寧蔵』『元官蔵』『洪武南蔵（初刻南蔵）*』『永樂南蔵』『永樂北蔵*』『乾隆大蔵経（龍蔵、清蔵）*』『嘉興蔵』など。

これらの大蔵経のなか、*をつけたものは影印出版されています。また、『趙城蔵（金蔵）』をベースに、『房山石経』『磧砂蔵』『高麗大蔵経』などと対校して編纂された『中華大蔵経』も出版されています^[3]。

これらの大蔵経の一部が日本に伝わり、日本で刊行された大蔵経の元となっていますので、仏典の本文批判を行うためには重要な資料です。また、

『大正新脩大蔵経』などに収録されていない文献も含まれています。

ここにあげた中国の大蔵経の一部は、テキストデータや画像データとしてデジタル化されており、オンラインでの閲覧や検索が可能です（☞仏典を検索する）。

◆ その他

このほか、朝鮮半島で作られた仏典を集めた『韓国仏教全書』や、日本の仏典を集めた『日本大蔵経』『大日本仏教全書』、また日本仏教については『天台宗全書』『真言宗全書』等々、宗派ごとに編まれた叢書などもあります。

❖ 敦煌文献と日本古写経

『大正新脩大蔵経』や『大日本統蔵経』は、全文検索もできるということもありますので、世界中で用いられています。しかし、先に述べたように、どちらも一世紀近く前に作られたものですので、いろいろな点で問題があります。問題点の一つは、近年発見された新文献が収録されていないこと、もう一つは校訂が不十分であることです。これらの問題をクリアするには、先に紹介した『高麗大蔵経』『中華大蔵経』などに加え、敦煌文献や日本古写経などを用いる必要があります。

敦煌文献は、その名の通り、二十世紀初めに中国の敦煌で発見された文献の総称で、そのなかに大量の仏典が含まれています。そこに含まれている文献のなかには、それまで知られていなかったもの（「敦煌新出文献」などと言ったりします）も多く、中国仏教研究に大きなインパクトを与えました。敦煌文献の一部は『大正新脩大蔵経』第85巻に収録されていますが（☞解説：仏典の分類）、大部分はまだ叢書等に収録されていません。

敦煌文献の網羅的なデータベースとしては、大英図書館を中心に世界各国の大学・研究機関が参加するInternational Dunhuang Projectがあります。

- International Dunhuang Project (<http://idp.bl.uk>)

このデータベースでは、一部の文献が画像で公開されていますが、全体のデジタル化はまだ終



図3 International Dunhuang Project (龍谷大学のサイト)

わっていません。デジタル化されていないものについては、『敦煌宝藏』などで確認する必要があります。

- 『敦煌宝藏』全141巻(黄永武主編、新文豊出版社)
- 『国家図書館蔵敦煌遺書』全146巻(任統憲主編、北京図書館出版社)
- 『英国国家図書館蔵敦煌遺書』既存50冊(方広鋸・Frances Wood(呉芳思)主編、広西師範大学出版社)

『敦煌宝藏』は写真がやや不鮮明です。

また、日本には古い仏典の写経が数多く残されており(『日本古写経』と総称されます)、現在も調査・研究が続いています。特に近年、国際仏教学大学院大学が中心となって研究が進んでいる七寺(名古屋)・金剛寺(大阪)・興聖寺(京都)などに残された日本古写経は、唐代写本を写したものであること注目を集めています。先に紹介した中国の各種大蔵経が宋代以降のもので、日本古写経はそれより古い姿を残していると考えられ、実際に本文を比較すると相違点が多く見つかります。

日本古写経について調べたい場合には、国際仏教学大学院大学・日本古写経研究所の「日本古写経データベース」が便利です。



図4 日本古写経データベース

- 日本古写経データベース (<https://koshakyo-database.icabs.ac.jp>)

このデータベースでは、写本の冒頭しか画像を公開していません。全体を閲覧したい場合には、国際仏教学大学院大学の図書館で閲覧申請をするか、以下のような出版物、データベースを利用しましょう。

- 『宮内庁正倉院事務所蔵 聖語蔵経巻 カラーデジタル版』(宮内庁正倉院事務所編、CD-RまたはDVD-R、丸善、2007)
- 『七寺古逸經典研究叢書』全6巻(七寺古逸經典研究会編、大東出版社、1994-2000)

敦煌文献、日本古写経ともに、これ以外にもたくさん資料があり、ここで紹介したのはそのなかのごく一部です。また、これらの写本を翻刻(筆文字を活字にすること)・校訂して出版した書籍や論文も数多くあります。CNKIやCiNiiなどの研究文献データベースを、「敦煌写本」「新出文献」などのキーワードで検索してみましょう。

❖ 仏典の内容を知りたい

何千もある仏典にどのようなことが書いてあるのか知りたい場合、まずはジャパンナレッジやCNKI 工具書庫（☞解説:CNKI 工具書庫）などを使って、各種辞書、百科事典等を引いてみるのがよいでしょう。

しかし、一部のメジャーな文献を除くと、辞書に載ってない仏典のほうが多いでしょう。そのような文献を含む、大蔵経全体にわたる解題辞典としては、以下のようなものがあります。

- 『仏書解説大辞典』15冊（小野玄妙編纂、大東出版社、1975-1981）
- 『大蔵経全解説大事典』（鎌田茂雄ほか編、雄山閣出版、1998）
- 『一切経解題辞典』（鎌田茂雄総監修、浅井円道ほか監修、大東出版社、2002）

一部の主要な仏典に関する解題は、以下のようなものがあります。

- 『仏典解題事典』（水野弘元ほか編、春秋社、1977）
- 『大乘経典解説事典』（勝崎裕彦ほか編著、北辰堂、1997）

- 『中国仏教史籍概論』（陳垣著、西脇常記・村田みお訳、知泉書館、2014）
- 『韓国仏書解題辞典』（東国大学校仏教文化研究所編、国書刊行会、1982）

また、漢文仏典に関しては、「国訳」とよばれる書き下しが網羅的に行われており、そのなかの『国訳一切経』（大東出版社、1930）や『新国訳大蔵経』（大蔵出版、1993-、刊行中）には解題があり便利です（もっとも、解題の質にはバラツキがあります）。岩波書店をはじめ、出版各社から現代語訳も出ていますので、そこに付された解題も内容を知るのによいでしょう。

注


- [1] <http://www.jaibs.jp/article>
- [2] 中国の大蔵経については、佛教史学会編『仏教史研究ハンドブック』（法蔵館、2017）106～109ページにコンパクトにまとまっています。
- [3] 『中華大蔵経』は、漢文仏典だけでなく、チベット語仏典も収録されています。また、『趙城藏（金藏）』だけに収録されている珍しい文献については、『宋蔵遺珍』としても出版されています。

（師 茂樹）

❖ 仏典を検索する

Q: 晩唐の詩人・李商隱の「北青蘿」詩に「世界微塵裏，吾寧愛与憎」という句がありました。注釈を見ると「世界微塵裏」は仏典からの引用のようです。どの仏典から引用されているか、どのように調べればよいでしょうか？

❖ SAT・CBETA を使うのが基本

漢文仏典には様々な種類がありますが（ 仏典の叢書・テキスト）、テキスト検索をする場合には下の二つのサイトを使いましょう（以下の説明は、2018年9月段階のものです）。

- SAT サット 大正新脩大蔵経テキストデータベース
(<http://21dzk.l.u-tokyo.ac.jp/SAT/>)
- CBETA シーベータ 中華電子仏典協会 (<http://cbeta.org/>)

SAT と CBETA は、どちらも『大正新脩大蔵経』（大正蔵）を中心とした大規模な仏典データベースであり、それぞれに特徴がありますので、どちらか一方を使えば事足りるというわけではありません。それぞれの特徴を把握して両方使うことが肝要です。

両データベースが収録している文献の違いは、右表のとおりです。中国仏教文献で言えば、CBETA のほうがより多くの文献を検索することができます。しかし、散逸してしまっいつぶんた中国仏教文献の逸文（元の文献は失われたが、他の書物に引用されたことで部分的に残ったもの）が、日本仏教文献に引用されている例も少なからずありますので、こちらもチェックしたい場合には SAT を使いましょう。

また SAT と CBETA では、大蔵経のテキストデータベース以外に、仏教辞書など、様々なデータベースと連動しています。

SAT の場合、

	SAT	CBETA
大正蔵 インド・中国	○	○
大正蔵 日本	○	×
大正蔵 その他	○	○
大正蔵 図像部	○	×
大日本統蔵経	×	○
嘉興蔵	画像	テキスト(一部)
金蔵・高麗大蔵経・宋蔵遺珍・房山石経・中華大蔵経・洪武南蔵・永樂北蔵・乾隆大蔵経・仏教大蔵経・中正蔵・国家図書館善本仏典・漢訳南伝大蔵経・正史仏教資料類編・北朝佛教石刻拓片百品・蔵外仏教文献・大蔵経補編選録・中国仏寺史志彙刊選録	×	一部
浄土宗全書	○	×

- チャールズ・ミュラー編「Digital Dictionary of Buddhism」(<http://buddhism-dict.net/ddb/>)
- 石田瑞麿『例文仏教語大辞典』（小学館、1997）¹⁾
- BDK パラレルコーパス（仏教伝道協会の英訳大蔵経との並行箇所を表示）
- 現代日本語訳パラレルコーパス（本稿執筆時点では『勝鬘経』『馬鳴菩薩伝』『龍樹菩薩伝』『比丘尼伝』の四点）
- インド学仏教学論文データベース INBUDS (<https://www.inbuds.net>)
- CiNii Articles (<https://ci.nii.ac.jp>)

と連動しており、検索結果から辞書や論文情報を

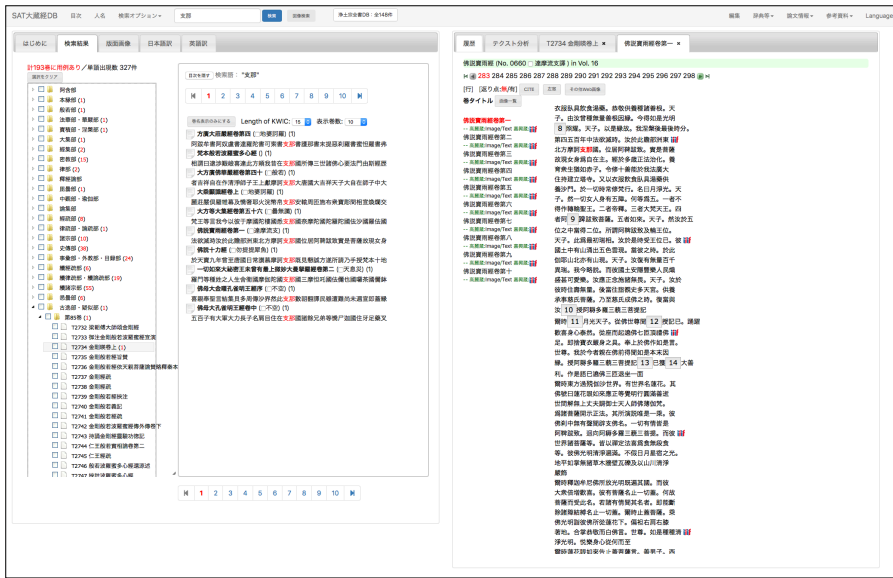


図1 SAT 大正新脩大蔵経テキストデータベース 2018

検索することもできます。一方のCBETAも、

- 丁福保『仏学大辞典』(1922)
- 辛嶋静志『A Glossary of Dharmaraksas Translation of the Lotus Sutra 正法華経詞典』(創価大学国際仏教学高等研究所、1998)
- 辛嶋静志『A Glossary of Kumarajiva's Translation of the Lotus Sutra 妙法蓮華経詞典』(創価大学国際仏教学高等研究所、2001)
- 辛嶋静志『A Glossary of Lokaksema's

Translation of the Astasahasrika Prajnaparamita 道行般若経詞典』(創価大学国際仏教学高等研究所、2010)

- 辛嶋静志『Die Abhisamacarika Dharmak Verhaltensregeln fuer buddhistische Moenche der Mahasamghika-Lokottaravadins. Grammatik, Glossar und Nachtraege (比丘威儀法詞典)』(創価大学国際仏教学高等研究所、2012)
- 釈智論『南山律学辞典』(西蓮浄苑、1996-2000)

などの仏教辞典が用意されており、文献検索をしながら辞書を引くことができます。

❖ SAT・CBETAの検索の仕方

◆ ウェブ上で検索する

両サイトの使い方は難しくありません。SATの場合、ブラウザで「SAT 大正新脩大蔵経テキストデータベース 2018」のページを開き、左上のテキストボックス(入力欄)に検索したい文字列を入力して検索ボタンを押せば検索できます。異体字を自動的に同一視してくれるほか、AND・OR検索や正規表現(一部記号のみ、8文字まで)などが使えます。他にも様々な機能がありますので、ウェ



図2 CBETA 漢文大蔵経

プ上のヘルプなどを確認してください^[2]。

CBETA では、ウェブ上での検索のほか、CBReader という専用ソフトを使う方法があります（後述）。ウェブ上での検索は、トップページの右上のテキストボックスに検索したい文字列を入力し、大蔵経搜尋ボタンを押せば検索できます。この検索システムは Google の検索機能を使っていますので、Google で使えるオプションはそのまま使えます。

■ CBReader を使う

CBETA にはウェブ版のほかに、Windows・macOS にインストール可能なダウンロード版がありますので、ネット環境がない状況でも検索することができます。CBETA トップページの上のメニューにある「下載」→「CBReader 下載」から行くことができるダウンロードページでは、以下のようなファイルがダウンロード可能となっています。

- CBETA DVD 30 光碟映像檔…以下のソフトウェア、データ全体と、EPUB ファイルなどを追加収録した ISO ファイル。DVD-R に焼いて配布することが可能（直接マウントして閲覧することも可能）。ファイルサイズが大きいので注意。
- CBETA CBReader 2X…下の「経文資料檔」を閲覧・検索するためのソフトウェア。Windows 版と macOS 版がある。
- CBETA CBReader 2X 経文資料檔…仏典のデータ。
- GoldenDict 跨平台辭典…各種仏教辭典。Windows 版のみ。
- 悉曇字型及蘭札字型…悉曇等のフォントを収録。

インストール方法は OS ごとに異なります。ここでは「CBETA DVD 30 光碟映像檔」を使ったインストール方法を簡単に紹介します。

- Windows…DVD のなかにある「CBReader」

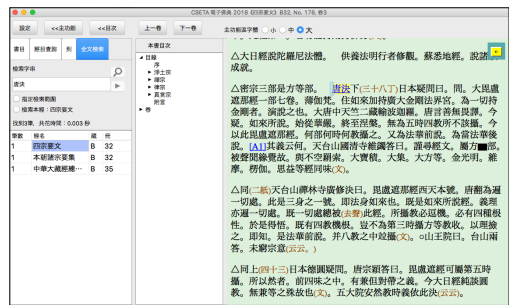


図 3 Mac 版 CBReader

フォルダを任意の場所にコピーする（3ギガバイト程度あるので注意）。

● macOS

- Finder で自分のホームディレクトリ (/User/(ユーザー名)/) を開く。そのなかにある「ライブラリ」フォルダ (/User/(ユーザー名)/Library/) に「CBETA」というフォルダを作る (Finder のメニューの [ファイル] → [新規フォルダ])。
- DVD のなかにある「CBReader」を開き、そのなかにある「Bookcase」を、上で作った「CBETA」フォルダにコピー（ドラッグ・アンド・ドロップ）する（3ギガバイト程度あるので注意）。
- DVD の「Mac」フォルダ中にある「CBReader」を「アプリケーション」にコピー（ドラッグ・アンド・ドロップ）する。

CBReader を起動すれば、各種検索が可能になります。全文検索をしたい場合には、左上の「全文検索」から可能です（図 3）。

■ 検索の実際

さて、前置きが長くなってしまいましたが、質問にあった李商隱の詩「北青蘿」の「世界微塵裏」について調べてみましょう。

まず、SAT と CBETA (CBReader) で「世界微塵裏」を検索すると、残念ながらどちらも結果は 0 件でした。逆に「世界 微塵」（「世界」と「微塵」の AND 検索）では、SAT だと 1,832 巻^[3]、CBETA で

は 2,036 件の用例がヒットしてしまい、これをいちいちチェックするのは大変そうです。

検索しても結果が 0 の場合、いくつかの理由が考えられます。

1. 引用元の文献が現存しない
2. 引用元の文献がデータベース化されていない
3. データベースに誤植がある
4. 引用が原文通りではない
5. 引用者による記憶違いや間違い、捏造

1 の場合は、逸文として残っている可能性がありますので、道教文献（⇒道教経典をしらべる）をはじめとする他の文献や、日本仏教文献などにも調査の範囲を広げる必要があるでしょう。2 の場合は、図書館などで原文にあたる必要がありません（けっこうな手間ですが、いろいろ勉強になります）。3、4 の場合、検索語の表現をいろいろ変えて検索してみると、見つかる場合があります。5 は、現在伝わっているバージョンには存在しなくても、かつて存在したバージョンには存在したかもしれませんので、徹底して調査しても見つからない場合でも「現時点では見つけることができなかった」という暫定的な結論にするほかないでしょう。

ここでは 4 のケースを考えてみます。「世界微塵裏」の「裏」は「～の中」という意味でしょうから、「世界微塵中」で検索してみたらどうでしょう。そうしたらいくつかヒットしました。

SAT :

大方広仏華嚴經 (0279) 0218c27-0220c18: 自処起一微塵中入無数世界微塵中起無数世界微塵中入一 [show] (2 matches)

華嚴經關脈義記 (1879) 0659b12-0660b15: 界等十方法界各尽同類世界微塵中刹一一重重滿十及尽主 [show] (1 match)

宗鏡録 (2016) 0864b12-0866b13: 自処起一微塵中入無数世界微塵中起無数世界微塵中入一 [show] (2 matches)

CBETA (CBReader、下の表は一部のみ) :

筆数	経名	蔵	経	巻
2	大方広仏華嚴經	T	0279	42
1	華嚴經關脈義記	T	1879b	1
2	宗鏡録	T	2016	81
2	華嚴經合論	X	0223	69
2	華嚴綱要	X	0240	42
2	大方広仏華嚴經疏鈔会本	L	1557	42
2	華嚴經疏論纂要	B	0002	75

検索結果を見ると、禅文献の『宗鏡録』^{すぎょうろく}以外は、『大方広仏華嚴經』（『華嚴經』）とその注釈書のようです。実は『華嚴經關脈義記』を除くと、すべて一番上の『華嚴經』巻四十二と同じ箇所を注釈あるいは引用しているものなので、実質的な用例は上の二つになります。いずれも「世界の中に存在する極小の粒中に」「世界を構成している原子のような極小の粒の中に」というような意味になるでしょうか。

では次に、「世界微塵裏」に対する注釈を見ましましょう。注釈では次のようになっています。

【程注】楞嚴經：「由是引起塵勞煩惱起為世界。」法華經：「譬如有経巻，書写三千大千世界事，全在微塵中，時有智人破彼微塵，出此経巻。」【馮注】金剛經：「若以三千大千世界碎為微塵。」此種語極多。^[4]

このなかの「程注」「馮注」は、凡例を見ればそれぞれ程夢星『重訂李義山詩集箋注』と馮浩『玉谿生詩集箋注』という清代の注釈書だとわかります。これらの注釈書は、「楞嚴經」「法華經」「金剛經」という経典を典拠としてあげています（「此種語極多」とありますから、他にもたくさん用例があるということでしょう）。

注釈書であげられている経典の引用文を検索してみると、『楞嚴經』（『大仏頂如来密因修証了義諸菩薩萬行首楞嚴經』）は同じ文が見つかりました（T945, 19, 120a12-13）^[5]。「金剛經」は『金剛般若経』という経典に、「若善男子善女人以三千大千世界碎為

微塵」という下線部だけが異なる用例がありました (T235, 8, 752b6-7)。

問題は程夢星が「法華經」としてあげている經文です。これと類似する表現は『法華經』には見つからず、仏跋陀羅撰『華嚴經』卷三十五に以下のような文が見つかります (下線部が一致する部分)。

仏子。譬如有一經卷、如一千大千世界、大千世界一切所有無不記錄。(中略) 彼三千大千世界等經卷、在一微塵内。一切微塵亦復如是。時有一人、出興於世、智慧聰達、具足成就清淨天眼、見此經卷在微塵内、作如是念。云何如此廣大經卷在微塵内、而不饒益衆生耶。我当動作方便破彼微塵、出此經卷饒益衆生。(T278, 9, 623c27-624a11)

どうも程夢星は『華嚴經』と『法華經』を混同していたようです。この一文は、三千大千世界のことをすべて記録している「經卷」があり、それが極小の粒(微塵)のなかにあることを知った智慧のある人が、微塵を破壊して經卷を取り出した、という話になります。

文学作品に対してはいろいろな解釈が可能なのですが、仮に「世界微塵裏」という一句が「世界に存在する微塵のなかに」のような意味だとすれば、最初の『華嚴經』卷四十二の用例が一番近いだろうと思います⁶⁾。

❖ おわりに

以上、駆け足で仏典検索の方法について紹介してきました。ほかにもいろいろなキーワードで検

索することで、新たな発見があるかもしれません。

現在は検索可能なデータベースが充実してきていますが、それで何でも解決するわけではありません。ちょっとウェブで検索したぐらいで「見つかりませんでした」とあきらめるのは、大学院生以上ではやってはいけないことです。いろいろな角度から粘り強く調査をすることで、検索テクニック、文献調査技術を磨きましょう。

注

- [1] ジャパンナレッジ (<https://japanknowledge.com>) に収録されているものなので原則として有料。
- [2] 少し古いですが、永崎研宣「SAT 大藏經テキストデータベース 人文学におけるオープンデータの活用に向けて」(『情報管理』58(6)、2015、DOI: 10.1241/johokanri.58.422) が参考になります。
- [3] 正規表現を使った「世界・微塵」では 434 件。
- [4] 劉学鐸・余恕誠『李商隱詩集解 増訂重排本』(中華書局、2004)、2086 ページ。
- [5] 『大正新脩大藏經』の引用典拠の示し方については、「仏典の叢書・テキスト」参照。
- [6] 細かいことを言えば、一つ目の『楞嚴經』にある「塵勞」は煩惱を表す語であり、極小の粒のことを指す『金剛般若經』『華嚴經』の「微塵」とは意味が異なりますので、適切な用例とは言えないかもしれません。また、『金剛般若經』は三千大千世界を粉々に破壊した結果できる微塵のことを言っているのに対して、『華嚴經』の例は三千大千世界の情報が詰まった微塵を破壊して情報を取り出す、というようなことを言っていますので、「世界」と「微塵」との関係が異なります。

(師 茂樹)

解説：仏典の分類

❖ はじめに

『大正新脩大蔵経』などの仏典の叢書（大蔵経）には、各巻に「阿含部」「諸宗部」などのサブタイトルがついているものがあります（図1）。これは仏典の分類を表していますが、それぞれがどのような分類なのかは専門外の人にとってはわかりづらいと思います。

現在はデータベースで全文検索をすることもできますので（☞仏典を検索する）、そもそもこのような分類を意識することも少なくなってきましたが、それでも仏典データベースの多くは伝統的な分類を念頭に置いて作られていますので、大まかに知っておいたほうがよいでしょう。仏教の叢書はたくさんの種類があり（☞仏典の叢書・テキスト）、それぞれに独自の分類をしています。ここでは、漢文仏典のなかで現在もっともよく使われている『大正新脩大蔵経』の分類を解説したいと思います。

『大正新脩大蔵経』の構成を表1にまとめておきましたので、これを見ながら以下の解説を読んでいただければと思います。この表にもあるように、漢文仏典の分類を大きく規定しているのは、「三蔵」と「三国」です。この二つは、『大正新脩大蔵経』の書誌情報にはどこにも書いてありません



図1 『大正新脩大蔵経』の背表紙

が、『大正新脩大蔵経』の構成を大きく決定しています。

❖ 仏典分類の基本は「三蔵」

インドにおける仏典の分類の基本は「三蔵」であり、『大正新脩大蔵経』をはじめとする多くの仏教叢書がこの分類を基本にしています。『西遊記』でおなじみの「三蔵法師」は、この「三蔵」から来ています。「三蔵」は仏典の分類から転じて、仏典全体を指す言葉にもなっているため、三蔵法師は「仏典全体に通じたお坊さん」といった意味になります。

さて、三蔵とは経蔵・律蔵・論蔵の三つを指します。

- 経蔵…釈迦をはじめとするブツダが説いた教え（「経」と言った場合、狭義にはこれを指す）。
- 律蔵…釈迦をはじめとするブツダが制定した仏教教団や僧侶のためのルール（戒律）。
- 論蔵…ブツダの教えを解説した（インドの）仏弟子の著作。

『大正新脩大蔵経』では、経蔵・律蔵・論蔵は第1巻～第32巻に収録されています。これらはすべてインドで編纂されたり、書かれたりしたものですので、『大正新脩大蔵経』に収録されているのは漢訳（中国語訳）ということになります（一部、中国で撰述された偽経なども含みます）。

経蔵のなかには「阿含部」「般若部」などの下位分類がありますが、これは伝統的な經典の分類方法で、概ね内容で分類されています。阿含部は、釈迦在世時に説かれたとされる初期仏教經典（「小乗仏教」という言い方をすることもあります、差別的な用語なのでできるだけ使わないほうがよいでしょう）を集めたもの、「本縁部」は釈迦の前世譚（本生譚）を集めたもの、「般若部」は般若經典と総称される

		地域による分類			その他
		インド	中国・朝鮮半島	日本	
三蔵	經	阿含部 (1・2)	經疏部 (33～39)	統經疏部 (56～61)	古逸部 (85) 疑似部 (85)
		本縁部 (3・4)			
		般若部 (5～8)			
法華部 (8)					
華嚴部 (8・9)					
宝積部 (11・12)					
涅槃部 (12)					
大集部 (13)					
經集部 (14～17)					
密教部 (19～21)					
	律	律部 (22～24)	律疏部 (40)	統律疏部 (62)	
	論	釈經論部 (25～26)	論疏部 (40～44)	統論疏部 (63～70)	
		毘曇部 (26～29)			
		中觀部 (30)			
		瑜伽部 (30～31)			
		論集部 (32)			
その他		(史伝部)	諸宗部 (44～48)	統諸宗部 (70～84)	
		(外教部)	史伝部 (49～52) 事彙部 (53～54) 外教部 (54) 目録部 (55)		
				悉曇部 (84)	

表 1 『大正新脩大蔵經』の分類（カッコ内の数字は巻数）

「空」を主張する大乘經典を集めたもの……といった具合です。

各部の順序は、概ね歴史的な成立順です。『大正新脩大蔵經』以前の蔵經では、「華嚴部」を最初に置くものもあります。『華嚴經』という經典は、釈迦が悟りを開いた直後に説かれた、という設定になっていましたので、『華嚴經』を含む「華嚴部」を最初に置いたというわけです。『大正新脩大蔵經』はそういった伝統的な分類を採用せず、近代以降の学術的な研究に基づく実際の成立年代に基づいていますが、完全に学術的な根拠に基づいているわけではなく、実際には伝統的な分類・順序と、歴史的な分類・順序の折衷になっています。

❖ 地域別分類

『大正新脩大蔵經』で、三蔵とならびもう一つの分類の軸となっているのが、「三国」です。昔、「三国一の花嫁さん」のような言い方がありましたが、その「三国」のことです。具体的には、仏教が伝来してきたインド・中国・日本を指します（実際には中国に朝鮮半島が含まれる）。それぞれの地域で作られた文献を、インド撰述部、中国撰述部、日本撰述部などと総称することもあります。


インドの三蔵、すなわち經・律・論に対して、中国・朝鮮半島で作られた注釈書（しよ）がそれぞれ經疏部・律疏部・論疏部としてまとめられ、日本で作られた注釈書が統經疏部・統律疏部・統論疏部として分類されています。

インドの三蔵には「阿含部」などの下位分類がりましたが、中国・朝鮮半島・日本の場合にはそのような分類はありません。しかし、続経疏部のなかの文献の順序は、インドにおける順序を踏襲しています。

❖ 三蔵以外の文献

『大正新脩大蔵経』には、これまで述べてきた三蔵の枠組みに収まらないもの、言い換えれば中国・朝鮮半島・日本で独自に作られた文献も、収録されています。

- 諸宗部・続諸宗部…天台宗、三論宗、法相宗、禪宗、浄土宗など、中国・朝鮮半島（諸宗部）や日本（続諸宗部）で成立した学派・宗派内で作られた文献のなかから、代表的なものを宗派ごとに収録。
- 史伝部…仏教史や高僧伝など、仏教者によって作られた歴史書や地誌を収録。一部インドで作られたものの漢訳も含む。
- 事彙部…辞典・事典的な文献。
- 外教部…『老子化胡経』や『大秦景教流行

中国碑』（石刻資料をしらべる）など、仏教以外の宗教の文献を収録。一部インドで作られたものの漢訳も含む。

- 目録部…前近代に作られた仏典目録。
- 悉曇部…悉曇しつたんとは梵字（東アジアに伝来したインド系文字の一種）のこと。ここでは、日本で作られた悉曇に関する文献のほか、仏教音楽の楽譜なども収録されている。
- 古逸部・疑似部…『大正新脩大蔵経』が編纂されている時期と同じ頃に発見・公開された敦煌文献を収録。三蔵全体、インド・中国両方にまたがる。十分に時間がなかったためか、残念ながら翻刻・校訂の精度は他の巻と比べてよくない。

これらの分類は『大正新脩大蔵経』独自のものですが、他の仏教叢書の分類に使われることもあります。

（師 茂樹）

道教・民間信仰をしらべる

❖ 道教経典をらしべる

Q:『封神演義』でよく出てくる『黄庭経』とはどんな経典でしょうか？有名な経典らしく、ネットでも訳文がありますが、いまひとつ内容が理解できません。

❖ 有名でも難解

『黄庭経』は、王羲之が書跡を残したとされ、そのために日本でもよく知られています。後世に与えた影響も大きいです。『封神演義』に名前が出てくるのも、この経典が非常に著名であるからだと考えます。

しかし、いざこの経典を読んでも、内容は非常に難解で、理解しにくい面があります。単に直訳した文章のみを見るだけでは、わかりにくいでしょう。

そもそも「黄庭」が、人体のどこを指すのかについても、議論がいろいろわかれます。筆者も、この経典の内容を的確に説明しろといわれても、お手上げの状態です。

この経典では、道教の「内観法」という修行法が語られています。人間の身体の中には数多くの神々がいて、各部位を守護していますが、その神々が身体内から離脱していくと、人間に死が訪れます。そのため、不老不死、あるいは長命を願うなら、自分の体内の神を知り、自身の精と気によって神を養わなければなりません。その方法を説いているのだということです。またこの経典は、読誦するだけで効用があるとされます。

ただ、いま『黄庭経』とされるのは、『黄庭外景経』と『黄庭内景経』の二種があります。このほかに、『黄庭経』と題する経典はたくさんあります。

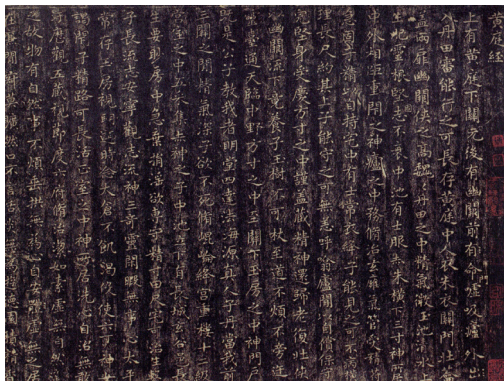
ただ翻訳したものを読むだけでは、その背景の思想を含めて、簡単には理解できないと思います。ですので、ここではいくつかの解説書を挙げて紹介します。

❖ 解説書いくつか

『黄庭経』は、むろん専門の論文を参照してもよいのですが、その論文自体が理解しにくい面があります。

まずは、道教経典を一般に解説したいいくつかの書籍を見て、そこから他経典との関連も考えるとよいのではないかと考えます。

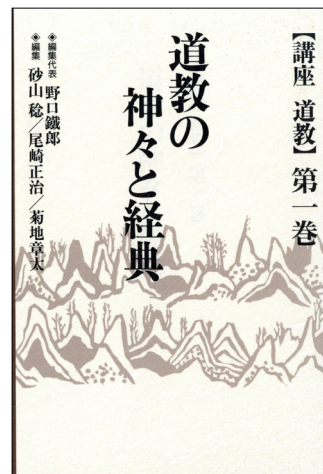
大修館書店の「あじあブックス」シリーズのひとつに、増尾伸一郎・丸山宏編『道教の経典を読む』（大修館書店、2001）があります。この本は、『太平経』『黄庭経』『度人経』『洞淵神呪経』『清静経』



伝王羲之書『黄庭経』墨拓（『道教の美術展図録』読売新聞社、2009）50ページ



『道教的經典を読む』（大修館書店、2001）



『道教的の神々と經典』（雄山閣、1999）

『陰符経』『悟真篇』『道法会元』『上清靈宝大法』『玉皇経』などの知名度の高い經典を、それぞれの經典に詳しい研究者が解説するというもので、一般向けの本ではありますが、内容は充実しています。經典の本文の一部を訳出しているのも大きいでしょう。まずは、この書籍を手がかりにしていたきたいと思います。

『黄庭経』については、加藤千恵氏が詳しい解説が書かれています。こちらをまずは参照することが大事だと思います。

また雄山閣『講座 道教』の第一巻に当たる編集代表野口鐵郎・編集砂山稔・尾崎正治・菊池章太『道教的の神々と經典』（雄山閣、1999）も大いに参考になる書籍です。

この本では、『太平経』『抱朴子』『真誥』『黄庭経』『雲笈七籤』などが取りあげられており、それぞれ解説が加えられています。ただ中身はやや専門的なものも含まれます。

『黄庭経』については、前田繁樹氏が第Ⅱ部2章の「黄庭経とその時代」という文章で解説されています。ここでは、『抱朴子』や『老子中経』などの様々な經典を参照しながら解説しており、単に『黄庭経』だけを見ただけではわからない面について丁寧な説明がなされています。

あるいは『黄庭経』をひとつだけ取りあげるだけでなく、『抱朴子』や『悟真篇』『周易参同契』『雲

笈七籤』など多くの道教經典を互いに参照しながら理解したほうがいいかとも思います。

こういった書籍をふまえて、さらに野口鐵郎・坂出祥伸・福井文雅・山田利明編『道教事典』（平河出版社、1994）の関連各項目を参考にさせていただけるとよいかとも思います。

あとは進んで、道教經典の叢書である『道蔵』に収録する經典を参照することになるとと思います（☞解説：『道蔵』とは）。

❖ その他の情報

日本語のウィキペディア (<https://ja.wikipedia.org/wiki/>) だと、『黄庭経』の項目は立てられていない形です。むしろ「王羲之」の項目に解説が含まれています。

中国語版のウィキペディア (<https://zh.wikipedia.org/wiki/>) には、かなり詳しい解説があります。ただ、見てわかるとおり、「黄庭」の語の解釈だけでも、目であるとか、脾臓であるとか、下丹田であるとか、様々な説が挙げられています。

中国語であれば、「CNKI 工具書庫」(<http://gongjushu.cnki.net/rbook/>) を参照するのもよいかもしれません（☞解説：CNKI 工具書庫）。こちらには、「黄庭経」で検索すると 38 箇条の情報が出てきます。王羲之の書跡とされるためか、書道に関連する情報も多く出てきます。

原文については、ウィキペディアの姉妹サイト「維基文庫（ウィキソース）」（<https://zh.wikisource.org/wiki/>）に『黄庭外景経』と『黄庭内景経』の全文が収録されています。ただ、この経典については、注釈を参照しながらでないと思ひにくく思ひま

す。劉処玄の『黄庭内景玉経註』も収録されていますので、こちらも参照するとよいかもしれません。

（二階堂 善弘）

解説：『道蔵』とは

❖ 『道蔵』とは

われわれがいま一般的に使う『道蔵』は、明の正統年間（1436-1449）に編纂された『正統道蔵』のことを指します。さらに万暦年間（1573-1620）に『統道蔵』が編纂され、一般的にいま『道蔵』といった場合は、この両者を合わせたものを意味します。全部で5,485巻にもなる、膨大な叢書です。

こういった道教経典の編纂の作業は、だいたい歴代の王朝によって行われてきました。唐の玄宗の時、宋の真宗の時、そして金・元の時代にも、それぞれ『道蔵』が編纂されました。ただ、いずれも一部を除いて残っていません。

とはいえ、宋の時代の『道蔵』である『大宋天宮宝蔵』は、その内容をダイジェストしたものが『雲笈七籤』となって残っています。これ以前の道教経典の概略を知るには、この『雲笈七籤』を見るのが一番です。

清の時代には王朝によって『道蔵』を編纂することは行われませんでした。しかし民間では数次にわたって道教経典が編纂され、光緒年間には四川の二仙庵において『道蔵輯要』が刊行されました。これは貴重な書物を多く含むものです。

現代においては、明の『道蔵』のあとを継ぐ形

の『中華道蔵』が編集され、2004年に華夏出版社から刊行されました。

これとは別に、1990年代には、巴蜀書社から『海外道書』が出版されました。これは明の『道蔵』から漏れた経典を集めたものです。

現代の読者からすると、たぶん『中華道蔵』のほうが使いやすいと思うのですが、相変わらず『道蔵』といえば明の『正統道蔵』を指すので、そこは注意が必要です。ただ、最近は論文でも、『中華道蔵』を典拠にする研究者も増えてきました。

なお、『道蔵』の影印本については、上海涵芬楼本が芸文印書館・上海書店及び文物出版社・中文書店などから出版されていますので、大きな図書館、あるいは大学図書館に行けば見られると思います。大きな図書館であれば、『道蔵輯要』なども所蔵していることがあります。

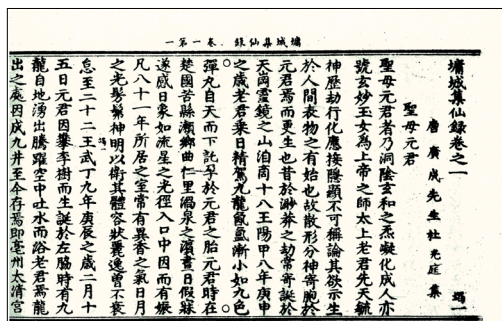
❖ 様々な内容

『道蔵』には様々な種類の経典が含まれていて、時にほとんど統一性が感じられないほどに感ずるのではないかと思います。

まず『老子』や『莊子』の注釈書がたくさん収められているのは理解できると思います。さらに『韓非子』の注釈が入っているのも、法家と道家の密接な関係を考えればわかるでしょう。

しかし、『道蔵』には墨家思想の書である『墨子』も含まれますし、『孫子』のような兵家の書も含まれます。要するに、諸子百家の儒家以外のものは、どんどん取り込んでいくという形でしょうか。『黄帝内経』などの医学書もかなり多く含まれます。道教の養生思想から、病気を治したり、薬草を用いることがさかに行われたためです。歴代の道士は、同時に医者である人も多かったです。

2015年にノーベル生理学・医学賞を受賞した屠呦呦の業績は、葛洪の医学書『肘後備急方』に



『道蔵』の一部（『壩城集仙録』）

触発されてのことでしたが、この書も『道蔵』のなかに入っています。

道家を含む諸子百家、医学書や地理書、それに仙人や道士の伝記など、様々な内容を含むのが『道蔵』なのです。

それから、道教の経典は似たような名前が多く存在するので、注意が必要です。

「太上」「洞玄」「靈宝」「上清」などという名称が組み合わされるものが多く、はじめの何文字かを見ては区別できないものが大半です。たとえば、『玉皇経』という経典は、正式なタイトルは『太上洞玄靈宝高上玉皇本行集経』となります。それを省略して『玉皇経』と呼んでいるわけです。

❖ 『道蔵』の分類「三洞四輔」

『道蔵』の分類法は特異なもので、「三洞」「四輔」というものです。これがまた厄介で、一般の人にはなかなか理解しにくい分類となっています。

まず「三洞」ですが、これは「洞真」「洞神」「洞玄」の3つを指します。これは道教で考える天界に応じたものです。

六朝時代、陸修静という有名な道士が、この三洞によって当時増えてきた道教経典を分類しました。すなわち、「上清経」のグループを洞真部に、「三皇経」のグループを洞神部に、「靈宝経」のグループを洞玄部に所属させました。

その後、さらに「四輔」という分類が加えられました。「道德経」とその注釈を中心とした太玄部、「太平経」を中心とした太平部、外丹の法術を説いた経典を集めた太清部、そして天師道系の経典を集めた正一部の4つです。これら三洞(「洞真」「洞神」「洞玄」と四輔(「太玄」「太平」「太清」「正一」)を合わせた7部が、道教経典の基本的な分類として成立したのです。現在の『道蔵』も、これによって分類されています。

ただ、あまりにも古くにその分類が決まったため、後世、道教経典が増えていきますと、どうしてもその枠組みに合わないところが出てきます。

たとえば、金の時代から発展した全真教の経典は、7分類のいずれに属するか、いまひとつ判然としません。おそらく、どこかで全真部を作って

おくべきだったのかもしれませんが、そうはなりません。現代の『中華道蔵』ですと、そういった点は一部改善されています。

そのほかにも、どうしてこの経典がこの部に入っているのか、疑問に思うところがたくさんあります。ただ、この点はそういうものだとご理解いただくしかないかもしれません。

また、三洞四輔のさらに下の区分として「十二類」があります。

「本文類」「神符類」「玉訣類」「靈図類」「譜録類」「戒律類」「威儀類」「方法類」「衆術類」「記伝類」「讚頌類」「表奏類」という形で、これは内容に即した分類です。こちらも実際の内容との齟齬は結構あります。

道教経典は、「〇〇部〇〇類」でその収録箇所を示すことが多いです。たとえば、『陰符経』であれば、「洞真部玉訣類」の収録となります。

❖ 参考書あれこれ

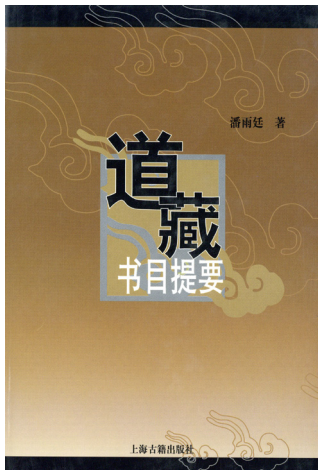
先にも述べました通り、『道蔵』は膨大な道教経典が収録されているので、扱いが厄介です。

ここでは、『道蔵』を扱うためのいくつかの参考書を紹介します。

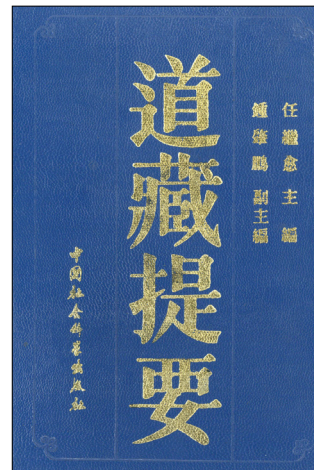
道教経典が論文に引用されている時、先に示したように「〇〇部〇〇類」という書き方が一般的です。しかし、また一方で「H.Y.167」とか、「S.N.108」というナンバーが付されている場合も多いです。

これは何かと申しますと、「H.Y.」のほうは「ハーバード・イェンチン・ナンバー」と呼ばれておりまして、ハーバード大学の『道蔵子目引得』(『哈佛燕京学社引得』25、1942)において『道蔵』に収録されている経典に通しのナンバーが振られたものです。そして「S.N.」のほうは、「シペール・ナンバー」と呼ばれて、クリストファー・シペール(Kristofer Schipper)『道蔵通検』(フランス極東学院、1975)において、通しナンバーが振られたものです。シペール・ナンバーのほうは、数が少し多くなっていますが、これはハーバード・ナンバーの数え間違いを正してあるからだとされます。

この両者を比較し、整理したものが、野口鐵郎・



『道蔵書目提要』（潘雨廷、上海古籍出版社、2003）



『道蔵提要』（任繼愈主編、中国社会科学出版社、1991）

坂出祥伸・福井文雅・山田利明編『道教事典』（平河出版社、1994）の付録に付いています。専門の研究者でなければ、こちらを参照したほうがわかりやすいと思います。

道教経典は、成立年代不明のものが多いです。経典のなかを見て「元始天尊によって下された」などと書いてあって、年代が全然当てになりません。経典によっては、各研究者で意見が分かれたりします。

道教経典の中身や性格については、いくつかの「提要」があって、そこで論じられています。有名なものには、潘雨廷『道蔵書目提要』（上海古籍出版社、2003）があります。ただ、ここで論じられている経典の数はそれほど多くはありません。

もっと多く、収録されたすべての経典についての解説が欲しいという場合は、任繼愈主編『道蔵提要』（中国社会科学出版社、1991）が参考になります。『道蔵』に含まれる多くの経典について、その性格や内容について紹介します。時に成立年代についても考証しますが、すべてではありません。

また英文の書籍になりますが、『道蔵』のほぼすべての経典について論じたものに、シペールとヴェルエレン両氏の Kristofer Schipper and Franciscus Verellen "The Taoist Canon" (The University of Chicago Press、2005) があります。各経典について、その成立年代についても深く論じています。

『道蔵』については、数多くの研究書がありますが、よく参考にされるのは、陳国符『道蔵源流考』修訂版（中華書局、2014）です。初版は1949年で古い本ですが、いまでも参考にされています。

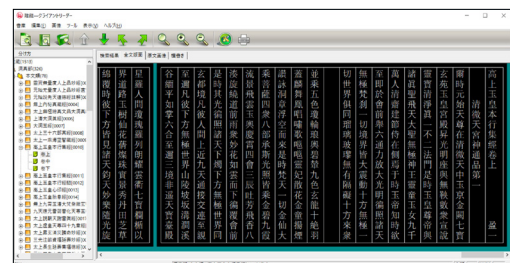
❖ 『道蔵』のデータ

『道蔵』については、愛如生や凱希メディアサービスによって全文検索が可能なデータベースが作成されています。ただ、これらは研究機関向けのもので、かなり高価なものとなっています。

インターネットには、『道蔵』に含まれるかなりの数の経典のデータがアップされています。

まず、台湾の中央研究院「漢籍電子文庫」(<http://hanji.sinica.edu.tw/>)にはデータがアップされていますが、それほど多くはありません。

ウィキペディアの姉妹サイト「維基文庫」(<https://zh.wikisource.org/wiki/>)のなかには、かなりの数の道



凱希メディアサービス『道蔵』

教経典が収録されています。ただ、やはり『道蔵』全部ではありません。

「中国哲学書電子化計画」(<https://ctext.org/zh>)においても、『道蔵』に含まれる経典が続々とアップされつつあります。しかもここは、画像までアップしてくれています。とはいえ、テキストのほうのデータは、校正がまだまだのようで、残念ながら使えるデータにはなっていません。

『中華道蔵』のデータも含めて、いま数多くのサイトでデータがアップされている状況ですが、そ

れぞれのサイトで重視している部分が異なっており、品質にもバラつきがある状態です。

道教経典でいえば、「漢リポ Kanseki Repository」(<http://www.kanripo.org/>)のデータが充実しています。こちらは『道蔵輯要』のデータがすべて公開されていますし、使い勝手も非常によいです。とりあえずは、まずこちらでデータを検索することをお勧めします。

(二階堂 善弘)

❖ 民間信仰の神をしらべる

Q: 台湾に行ったら、あちこちに天后廟というものがありました。そこに祀られているのがどういう神さまか、どうやって調べたらよいのでしょうか？

❖ 天后媽祖とは

台湾や、中国大陸の福建・広東地方に行くと、それはもう、あちこちに天后廟があります。台湾では廟と廟の間を多くの信者が巡礼する大規模な活動があり、それによっても知られています。

天后とは、「媽祖」と呼ばれる神さまで、名を「林默娘」という女性が神になったものです。宋代に実在した女性だとされています。

また天后廟でなくても、媽祖を祀るところは数多くあります。台北の龍山寺は、主として祭祀するのは観音菩薩ですが、奥の殿には多くの神々とともに媽祖を祀ります。

媽祖は、海の神で、航海における守護神とされます。そのため、中国では船のひとつひとつにこの媽祖の像が置かれていました。陸にあがると、その媽祖の像は廟に安置し、出航するときまた船に乗せます。

日本の長崎にきた清の商人たちの船も、上陸し

たときに媽祖の像を崇福寺・興福寺などの媽祖廟に安置しました。これを「ボサあげ（菩薩揚げ）」と称しました。長崎の崇福寺・興福寺には、いまでも媽祖堂があって媽祖が祀られています。

マレーシアやベトナム、タイなど、東南アジアの各地に行っても、至るところに媽祖廟があります。中華系の移民の行く港には、必ずといってよいほど媽祖廟が存在しています。

その信仰の本山となっているのは、福建の莆田湄洲島という島で、もう島全体が聖地となっている形です。

部下の神として、千里眼と順風耳がおります。媽祖廟に行くと、だいたいこのふたつの鬼の像が、前面を守護しています。

❖ 媽祖の資料

媽祖に関する資料は、中文資料の場合は、かえって多すぎて混乱するかもしれません。ここでは基



莆田湄洲島の媽祖像



呂宗力・欒保群著『中国民間諸神』（河北教育出版社、2001）



朱天順著『媽祖と中国の民間信仰』（平河出版社、1996）

本的なものから紹介していきます。

まず日本語のウィキペディア (<https://ja.wikipedia.org/wiki/>)「媽祖」の項目があります。日本語ウィキペディアは、民間信仰などについては手薄な印象がありますが、こちらの情報はそれなりに揃っています。

さらに、中国語ウィキペディア (<https://zh.wikipedia.org/wiki/>)「媽祖」の項目は、かなり豊富な内容で、非常に充実しています。

ただ、こういった情報は媽祖信仰がかなり流行してからのものですので、信仰の発展の経緯については、不明な点も多いです。

中国の民間信仰の神々については、いまはいろいろな書籍があって、どれも参考になります。た

だ、もっとも基本的かつ網羅的な書籍は、呂宗力・樂保群著『中国民間諸神』（河北教育出版社、2001）になります。天后媽祖についても、非常に詳しい記録があり、信仰の発展についての足跡を追うことができます。

もう少し専門的な知識が必要であれば、朱天順著『媽祖と中国の民間信仰』（平河出版社、1996）が、一般向けながら、豊富な内容を持っています。

中文の専門書については枚挙にいとまがありません。いくつか書名を挙げれば、以下の通りです。

- 『媽祖信仰史研究』（徐曉望著、海風出版社、2007）
- 『閩台媽祖文化傳播研究』（吉峰著、廈門大学出版社、2017）
- 『媽祖伝説研究』（羅春榮著、天津古籍出版社、2009）
- 『海神媽祖』（呉玉賢主編、外文出版社、2001）
- 『台湾的王爺与媽祖』（蔡相輝著、台原出版社、1989）

このほかにも数多くの研究書があります。

なお、これらの研究の淵源となっているのは、日本語の書籍で李献章『媽祖信仰の研究』（泰山文物社、1979）です。古い本ですが、網羅的で非常に質の高い研究となっています。ただ、現在は非常に入手しにくい書籍となっています。図書館で閲覧するのがよいでしょう。

（二階堂 善弘）

解説：民間信仰と民間宗教の資料

❖ 民間信仰と民間宗教

少し細かすぎて区別が難しいと思いますが、中国の宗教文化では、「民間信仰」と「民間宗教」を分けて考えることが多いです。もっとも、これらは部分的に重なる面もあります。道教や仏教とも、それぞれ重なるところがあります。

中国や台湾・香港の廟に行くと、たくさんの人々が関帝廟や娘娘廟などにお参りしています。

そして神さまの誕生日になりますと、廟会というお祭りを行ったりします。こちらでは人々は特定の教団に属すわけではなく、その街や村においてお祈りを行い、お祭りに参加するだけです。

こういった地域に根ざした庶民の宗教文化は、「民間信仰」と呼ぶことにしています。

一方で、地域に根ざしてはいるものの、一定の教団組織があり、ある程度の信者を抱えているものを、「民間宗教」と呼びます。

昔から、羅教・八卦教・紅陽教などの教派があり、教団を組織しています。これらは白蓮教の流れをくむところが多いです。またこれらの教派の活動は、時に官憲からは警戒されました。こちらは「宝巻」という一種の経典を有していることが多いです。羅教の教団での『五部六冊』が有名です。

これらの民間信仰や民間宗教は、仏教や道教と交流し、その神さまや仏さまを信仰しています。ですので、関帝や観音菩薩は、民間信仰や民間宗教でも重要な神さまとなります。

関帝の場合などは、どこの宗教でもすべて祀られるので厄介です。関帝は仏教のお寺では伽藍守護神として祀られます。さらに、道教の道観では関聖帝君として上位の神として機能しています。

そして、民間宗教系の善堂などでは、関帝、または関恩主として祭祀されます。民間信仰では、普通に廟のなかに関帝が祀られ、関帝廟があちこ

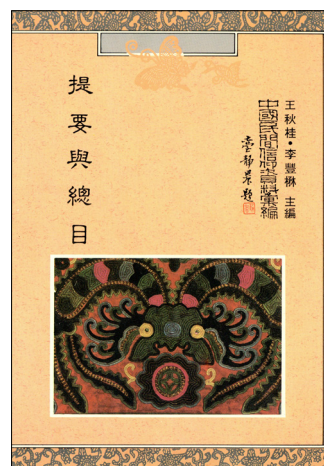
ちにあります。同じ関帝という神が、いくつかの系統が異なる宗教集団でそれぞれ祀られることになっているのです。単に関帝廟といわれても、それだけでは、どの系統なのかはわかりません。

❖ 民間信仰の資料

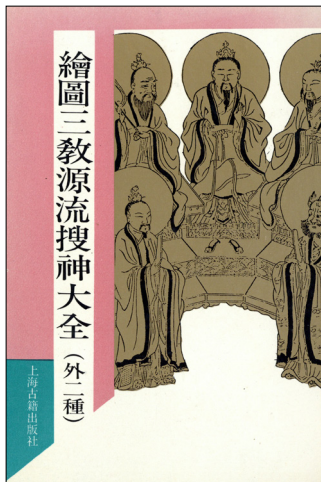
実は民間信仰については、『大蔵経』や『道蔵』などの経典類はほとんど存在しません。なぜなら、近代以前の民衆の多くは文字を読めませんでした。そのため、民衆は講談やお芝居などの演芸や、廟にある神さまの像や絵画などを見て、そこから神々の物語を体得していきました。

こういった信仰に関しては、信仰している本人たちはほとんど記録を残せません。その信仰については、文字を書ける読書人たちが、外部の目から記録を残すことになります。これが民間信仰に関する資料になります。

たとえば、『集説詮真』という書籍があります。これは、もともと清末の時代に、キリスト教を奉ずる人たちが民間信仰を「迷信」と決めつけ、徹



王秋桂・李豐楙編『中國民間信仰資料彙編』
(第1輯・台湾學生書局、1989)



『三教搜神大全』(上海古籍出版社、1990)

底批判するために書いた本です。ただ、批判対象として、非常に豊富な民間信仰の資料を収めているので、それで「逆に」民間信仰の資料として使えるようになっているのです。このように、逆の立場から残された民間信仰の資料も多々あります。

この『集説詮真』のほか、『神仙通鑑』などの民間信仰の資料を網羅的に集めたのが、王秋桂・李豊楙編『中国民間信仰資料彙編』(第1輯・台湾学生書局、1989)となります。清代以前の資料は、ほぼこちらに集められていますので、この本を参照することが必要になります。

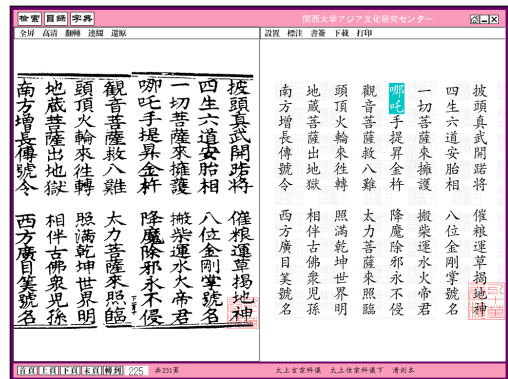
上記の『中国民間信仰資料彙編』のなかにも含まれますが、元から明の時代の神々の伝記を集めた『三教搜神大全』も重要な資料です。

❖ 民間宗教の資料

民間宗教においては、一定の文字が書ける層が属している場合があり、彼らは教団のなかに経典を持っていることがあります。ただ、一般によく知られているのは宝巻となります。

宝巻は、書物ですが、文字の読めない民衆のために読み上げられることが多いものです。演芸の台本に似ている役割を持ちます。宝巻を説くことを、「宣巻」と称します。

その教団の教理が説かれることが主ではありませんが、完全に娯楽としての物語を説くものもあり



愛如生『宝巻新集』

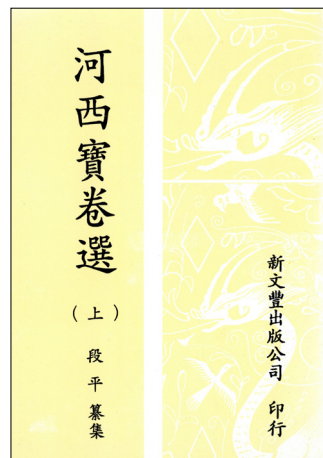
ます。演芸と宗教活動の両面を併せ持つものです。

先にも書いた通り、羅教の『五部六冊』はよく知られた宝巻です。このほかにも、宝巻を經典として扱う教団は多くあります。

宝巻に関する研究としては、基礎となるものに澤田瑞穂『増補 宝巻の研究』(国書刊行会、1975)があります。なお、澤田氏の所蔵していた宝巻の多くについては、いま早稲田大学の「古典籍総合データベース」の「宝巻」の項 (<http://www.wul.waseda.ac.jp/kotenseki/furyobunko/hokan.html>) で見ることができます。

近年は多くの宝巻を集めた『宝巻初集』(山西人民出版社、1994)などの叢書も数多く出版されています。

そういった宝巻類を集めてデータベースとした愛如生の『宝巻新集』も、いくつかの研究機関で



使用されています。

ただ叢書類にしろ、データベースにしろ、高価でかつ原文がそのまま出されているため、実際には読みにくい場合が多いです。

その場合は、活字として整理され、句読点を施

した本のほうをしてみるべきでしょう。

段平編『河西宝巻選』（新文豊出版公司、1992）などでは、いくつかの宝巻が整理され、読みやすくなっています。

（二階堂 善弘）

古典通俗文学をしらべる

❖ 古典通俗小説の概要をしらべる

Q: 『三下南唐』という小説について、作者やあらすじなどはどうやって調べたらよいのでしょうか？

❖ 日本語の参考図書

明・清の時代には通俗文学が花開き、『三国志演義』・『水滸伝』・『西遊記』・『金瓶梅』・『紅樓夢』といった長編小説、あるいは三言二拍に代表される短編小説が数多く世に出ました。それらには口語に近い文体である白話の文体が使われているので、白話小説とも呼ばれます。

我が国でも、多くの古典通俗小説が邦訳されていますが、評価の高い作品、あるいは有名な作品に限られているのが実情です。

さて、質問の『三下南唐』ですが、まずはGoogleで検索してみましょう。そうすると、小説のテキストデータを集めたアーカイブが幾つか見つかり、『趙太祖三下南唐』あるいは『宋太祖三下南唐』というタイトルになっていることがわかります。これを念頭に、調査を進めましょう。

中国の通俗小説について掲載されている日本語の参考図書としては、以下の2点が挙げられます。

- 『中国文化史大事典』（尾崎雄二郎・竺沙雅章・戸川芳郎編、大修館書店、2013）
- 『中国学芸大事典』（近藤春雄著、大修館書店、1978）

両者ともに通俗小説の作品について項目を立てていますが、邦訳があるような有名な作品に限られますので、『三下南唐』は、検索で見つかった異称も含めて、みあたりません。

❖ CNKI 工具書庫でしらべる

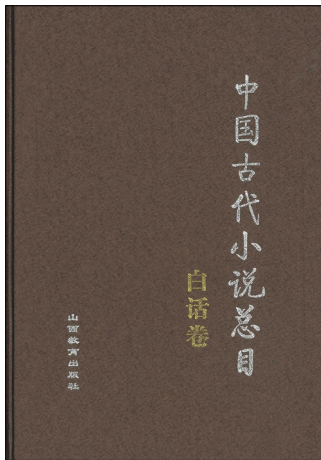
つぎに小説『三下南唐』について、中国語でしらべてみましょう。CNKI 工具書庫（※解説：CNKI 工具書庫）で「三下南唐」を検索します（「条头（模糊）」に設定）。すると、『五百種明清小説博覧』（張兵主編、上海辞書出版社、2005）の「宋太祖三下南唐」の項がヒットします。

CNKI 工具書庫が使える環境であるならば、クリックして全文を閲覧してみましょう。『宋太祖三下南唐』が清代の「白話長編神怪史話小説」で、全称『趙太祖三下南唐被困寿城』、別名が、全53回で「好古主人撰」と題しているがどのような人物であるのかわからない、清の咸豊年間の成立である、といったことがわかります。物語は、宋の太祖の南唐征討を扱っており、寿州城で南唐軍に包囲された太祖が、女將軍・劉金錠らの活躍で救出され、南唐を平定する、といったもので、仙人が登場し道術が飛び交う、『封神演義』のような趣の作品であるようです。

❖ 参考図書でしらべる

あらすじや概要をしらべるだけでしたら、CNKIの『五百種明清小説博覧』でも十分だといえますが、しかし、この本は項目数が500に絞られていて、古典通俗小説が網羅的に収集されているわけではありませんので、小説によっては見つからないこともあるでしょう。

古典通俗小説の目録・概説の定番といえるのは、



以下の書籍になります。

- 『中国古代小説総目』（石昌渝編、山西教育出版社、2004）

「文言巻」・「白話巻」・「索引巻」の3巻からなり、明清の通俗小説は白話巻に収められています。

通俗小説の目録としては、『中国通俗小説書目』（孫楷第著、国立北平図書館、1933）があり、1980年代に『日本東京所見小説書目』（孫楷第著、人民文学出版社、1981）、『倫敦所見中国小説書目』（柳存仁編輯、書目文献出版社、1982）、『増補中国通俗小説書目』（大塚秀高著、汲古書院、1989）といった目録が編まれましたが、本書はそれらを集大成する形で作られています。

特に、小説の版本とその所蔵についての解説が充実しており、例えば『三国志演義』の項目では、1990年代に進展した版本研究の成果に基づいて系統ごとに解説されています。ちなみに『五百種

明清小説博覧』では、主要な版本を列挙しているだけです。一方、あらすじの解説は『中国古代小説総目提要』の方が簡略です。

さて、『宋太祖三下南唐』に関しては、版本の種類も少なく、CNKIの『五百種明清小説博覧』と大差ありませんが、版本の特徴や所蔵についてはやはりこちらの方が詳しく書かれています。

❖ 古典通俗小説調査の注意点

古典通俗小説調査の起点となる『中国古代小説総目』ですが、利用にあたっていくつかの注意点もあります。

まず、索引の問題です。項目として立てられている正式名称でしか検索できないため、質問にある『三下南唐』というタイトルだけでは『宋太祖三下南唐』にたどり着くことができません。『俠義奇女伝』・『第一俠義奇女伝』といった異称も載っていません。このため、索引を引いてもみつからないときには、Google検索やCNKI工具書庫の全文検索などを駆使して、その小説の異称がないか、しらべる必要があります。小説などの通俗的な出版物では、往々にしてタイトルの他に略称・別名などがありますので、注意しましょう。

また、『中国古代小説総目』はピンイン順で排列されているため、たとえば『宋太祖三下南唐』と近い時代を扱った小説にどのようなものがあるか、といったことは、うまく調査できません。

ある小説を研究テーマに取り上げるのであれば、内容に即して分類されている『中国通俗小説書目』や『増補中国通俗小説書目』に目を通して、周辺に目配りしておくべきでしょう。

（千田 大介）

❖ 古典通俗小説の排印本

Q: 小説『三下南唐』を読んでみたいのですが、お勧めの活字本はありますか？

❖ 古典通俗小説の排印本

中国では文化大革命の時期には古典通俗小説の出版が禁止されていましたが、1980年代半ば以降、所謂「四大名著」(『三国志演義』・『水滸伝』・『西遊記』・『紅樓夢』)から順次解禁されていき、1990年代になるとマイナーな作品も含めて多種多様な古典通俗小説の排印本が、書店に並ぶようになりました(エロティックなものを除いて)。

質問の『宋太祖三下南唐』ですが、試みに中国国家図書館 OPAC で検索してみると、執筆時点以下のものがヒットします。

- 春風文芸出版社 (1994、2017)
- 山西人民出版社 (2009)
- 安徽人民出版社 (2013)
- 遠方出版社・内蒙古大学出版社 (2001)
- 岳麓書社 (1998)
- 中国文聯出版公司 (1998)
- 華夏出版社 (1995)
- 江西人民出版社 (1992)

これらのうちどれがお勧めかというと、特にお勧めはない、逆にいえば、どれを読んでもかまわない、ということになります。

❖ 研究者向けか一般向けか

中国の出版社から刊行されている古典通俗小説の排印本(活字本)は、研究者向けのものと、一般読者向けの娯楽読み物とに分かれます。両者を区別するのは、端的に言えば校注の有無です。

版本や鈔本(☞解説:刊本・鈔本・影印本・排印本)を活字に起こす際には、原本の欠落や誤りを修正しなくてはなりません。こと土人から「低俗」と

のそしりを受け続けた通俗小説は、その大半が坊刻本(☞解説:官刻本・家刻本・坊刻本)で、民間の商業主義的な出版物でしたので、厳密さが追求されないばかりか、安かろう悪かろうなものも多く、俗字が多用され誤字もしばしば見られます。それをどのように解釈し、確定したのかを校注に留めて評価を待つのが、学術的な態度です。

また、中国にあって繁体字で刊行される書籍は、それだけで繁体字の知識を持つ研究者を対象としていることになりしますので、学術的なニーズを意識した、品質の高いものである蓋然性が高くなります。

もしも研究のために小説を読むのであれば、当然のことながら、研究者向けの排印本を選ぶべきです。

一方、研究者ではない中国の一般の読者の立場からすると、簡体字で校注がない方がスムーズに読み進められるのでよい、ということになるでしょう。活字に起こす作業も、厳密に校注をつけないでざっくりやる方が、遥かに楽です。

中国では、中華書局や上海古籍出版社といった権威ある出版社からも、一般向けの古典小説の簡体字・排印本のシリーズが出版されていますが、それらの翻刻の質は、他の出版社に比べて決して優れているわけではありません。句読点の切り方や、異体字の同定を比べてみると、ものによっては他の出版社の方が良いことさえもあります。つまり、一般向けの排印本の良し悪しは、出版社などによって一概に決められるものではないのです。『宋太祖三下南唐』については、前に挙げた排印本は、いずれも校注のない簡体字横組みのもので、一般向けの読み物としての性格が強いものだといえます。その意味では、いずれも大きな差があるとは思えませんので、どれを読んでもか

まわらない、ということになるのです。書店で見つかった本、あるいは読みやすい装丁の本を選べばよいでしょう。

❖ 研究に使うときには

小説を読むのであれば、読みやすい一般向けの排印本を読んでもいっこうにかまいませんが、論文やレポートに引用したい、となると、事情は少々変わってきます。

小説を翻刻する過程で、ストーリーレベルの大きな誤りが発生することはまずありえないでしょう。しかし、細かい字句レベル、また句読の区切りなどには、多々、誤りがあるものです。単純なミスや、俗字・難読字の読み違いもありえますし、底本で欠けている字を適当に補う、といったことも行われているものです。

このため、研究で小説を引用する際には、最善の版本や鈔本（の影印本）（㊦古典通俗小説の影印本をさがす）にあたって、少なくとも引用箇所については、しっかりと対校するべきです。

❖ オンラインテキスト

Googleで「宋太祖三下南唐」を検索すると、全文テキストデータを収録するアーカイブがいくつもヒットします。それらのテキストデータは、適当な排印本をもとに入力したものが大半ですが、底本が明記されず、タイポがあったり、難字が書き換えられていることがあったりと、一般向けの排印本よりも、さらに品質が落ちることは避けられません。

とはいえ、オンラインテキストは無償で簡単に手にはいりますので、こと著作権が切れている古典通俗小説については、品質に問題がある可能性をふまえた上で、読んだり、全文を検索するためにデータをダウンロードしたりと、積極的に活用するべきだと思います。引用する際に、版本や鈔本にあたって確認しなくてはならないのは勿論ですが。

（千田 大介）

❖ 古典通俗小説の影印本をさがす

Q: 小説『三下南唐』の版本は、どのように確認したらよいでしょうか？

❖ 古典通俗小説の影印本

論文・レポートなどに古典通俗小説を引用する際には、版本や鈔本の確認が欠かせません。どのような版本・鈔本があるのかは、『中国古代小説総目』(☞古典通俗小説の概要をしらべる)などでしらべることができます。

版本や鈔本をしらべるとき、文献学的な調査では、紙の質や実際のサイズ計測が欠かせないので現物を見る必要がありますが、小説の内容や表現を確かめたいのであれば、影印本(☞解説:刊本・鈔本・影印本・排印本)を閲覧すればよいでしょう。

古典通俗小説の大規模な影印叢書には、以下の3種があります。

- ①『明清善本小説叢刊』初編・続編(天一出版社、1985・1990)
- ②『古本小説叢刊』(中華書局、1987-1991)
- ③『古本小説集成』(上海古籍出版社、1990-1994)

①は台湾で刊行されたもので、計343種の小説を収録しており、この叢書にしか収録されていないテキストもあります。通俗小説の大半は坊刻本(☞解説:官刻本・家刻本・坊刻本)で、摩耗した版本で刷られたような質の悪いテキストも多いのですが、それを差し引いても、紙焼きコピーを元に影印したとおぼしき、字がつぶれて読みづらいものが多く含まれているのが玉に瑕です。

②と③は北京と上海をそれぞれ代表する出版社が競い合うように刊行した、小説の大規模影印叢書です。②は明・清の小説および元の平話や明の説唱詞話など169種を、③は明・清の小説428種を、それぞれ収録していますが、両叢書で重複し

ているテキストも多々みられます。この3種の影印叢書で、古典通俗小説の大多数の作品の主要なテキストは、ほぼ網羅されています。

閲覧したい小説が影印叢書に収録されているかどうかは、中国国家図書館 OPAC(☞解説:中国国家図書館 OPAC)や NBInet(☞解説:NBInet 聯合目錄)、全国漢籍データベース(☞解説:全国漢籍データベース)、および各図書館の OPAC などで検索してしらべることになります。

『三下南唐』については、『古本小説叢刊』と『古本小説集成』に収録されていますが、前者は大英博物館所蔵英文堂刊本、後者は北京大学図書館所蔵紫貴堂刊本と、テキストが異なります。

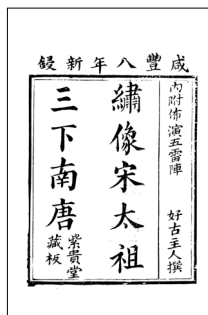
『中国古代小説総目』によると、最も古い版本は後者であるとのことですので、『古本小説集成』所収本(下図)を見るべきでしょう。

❖ その他の叢書・単行本

主要な小説のテキストは、上記3種の叢書でほぼ網羅されていますが、図書館のコレクションや研究上貴重なテキストなどについては、ほかにもさまざまな影印叢書や影印本が刊行されています。

それらは、やはり中国国家図書館 OPAC、NBInet などで探すことになります。

(千田 大介)



解説：官刻本・家刻本・坊刻本

❖ 官刻・家刻・坊刻

前近代中国の木版印刷物を出版の主体によって区分すると、官刻・家刻（私刻）・坊刻の三つに分けられます。

官刻とは、政府によって印刷刊行されたものを指し、中央政府によるものと地方官庁によるものがありました。

官に対する民の出版物には家刻と坊刻がありました。家刻は個人による出版物を、坊刻は民間による商業出版物を指します。民間の出版物のうち営利目的ならば坊刻、非営利ならば家刻と言い換えることもできます。

印刷は唐代にはじまると言われていますが、明確な時期は定かではありません。現存する紀年のある最古の印刷物は敦煌出土の「金剛波羅蜜經」（大英博物館所蔵）で、咸通9年（868）のもので、唐後半になると、こういった仏教経典や暦などが坊刻により多数出版されていることがわかっています。ただし、唐までは原則として鈔本（写本）中心の時代です。印刷物としての書物の流布が一般的となるのは宋以後のことです。

官刻は、五代・後唐の時代に宰相・馮道の命で国子監において儒教の経典を印刷刊行したことが始まりとされています。国子監とは国立の最高学府のことです。これ以後、国子監には印刷設備が整えられ、儒教経典をはじめとする官僚養成に必要な標準テキストが刊行されました。これら国子監の刊行物を監本かんぼんと呼び、特に明の監本は、テキスト校訂の行き届いた素性のよい本が多いことで知られています。清朝では中央の武英殿でんぼんや地方の書局で書物が印刷され、前者の印刷物は殿本、後者の印刷物は局本きょくぼんと呼ばれています。武英殿では

書物の分量により印刷方法（版木か活字か）を使い分けており、活字本は武英殿聚珍版しゅうちんとして知られています。殿本は精緻で美しい仕上がりですが、清朝に都合の悪い部分や異民族に関する記述などが書き換えられているので注意が必要です。

家刻では、時代が下るにつれ官僚・学者・蔵書家らがこぞって出版を試みました。たとえば、大蔵書家として知られる明の毛晋は、蔵書中の貴重書（善本）を復刻出版したり、校訂本を出版したりしています。これらは毛晋の蔵書楼（書庫）の名前から汲古閣本と呼ばれています。このように家刻本には学術的価値の高いものも多く、汲古閣本のように、学者や蔵書家たちの書齋名や蔵書楼名をもって通称とするのが一般的です。

❖ 坊刻と書肆

坊刻の版元かつ販売元を書肆しよし・書坊・書林などといいます。一般には坊刻の「坊」は書坊の「坊」だと説明されます。官刻や家刻に対して坊刻は、大量印刷・大量出版を旨とするものですから、内容・印刷・素材の全てにおいて玉石混交です。

たとえば、宋代に坊刻が盛んであった麻沙鎮（福建）の出版物は麻沙本と呼ばれて、後世に質の良くない本の代名詞のようになりました。

ただし、既に失われてしまった官刻のテキストが、坊刻により残っていることもあります。また、坊刻本は民間の需要を反映した出版物であるため、通俗文学のように官刻や家刻では決して出版されないようなジャンルの書物が数多く出版されています。このため文学研究ではもちろんのこと、歴史や思想の分野においても坊刻本は貴重な情報源となり得ます。

（小島 浩之）

❖ 古典通俗小説の全文検索

Q: 小説『三下南唐』で仙人「黄石公」が出てくる場面を探したいのですが、どうすればよいでしょうか？

❖ 大規模古典文献データベース

ある人物や言葉が出現する箇所を探すには、全文検索をかけることとなります。そうすると、文献データベースの出番です。学術的な大規模データベースの構築は、儒教経典や歴史書、詩文などが優先されてきたため、通俗的な小説・戯曲の類を収録するものは、以下の2つに限られます。

- 中国基本古籍庫
- 中国俗文庫

いずれも中国の愛如生社の製品で、日本国内では東方書店が代理店になっています。

このうち「中国基本古籍庫」は、明清の通俗小説を110種ほど収録していますが、比較的著名なものに限られるため、『三下南唐』は収録されていません。

中国俗文庫は、小説・戯曲・説唱など、通俗文学全般を収録しており、『三下南唐』も含まれますが、高価なデータベースであるため、残念ながら、国内ではまだ導入実績がないようです。中国に渡航したときに、現地の図書館などで使うしかないでしょう。

❖ ネットで検索する

大規模データベースが使えないとなると、ネットでデータを探すことになります。古典通俗小説は大半がテキストデータ化されており、そうしたデータを集めたオンラインテキストアーカイブも多数存在します。

そうしたオンラインデータの大半は、学術的なものではなく、きちんとした校訂を経ていませんので、質が高いとはいえませんが、それでも文献

の全文検索に利用することができます。

Googleで『三下南唐』を検索し、ヒットしたテキストアーカイブを見ていくと、いくつかのサイトでは、目次の第一回に出てくる人名「余鴻」が「餘鴻」になっています。姓の場合は繁体字でも「余」が正しいのですが、簡体字から機械的に繁体字に変換して「餘」になってしまったのでしょうか。こうした誤変換の含まれる繁体字のサイトを避けて、「余鴻」あるいは簡体字で「余鸿」になっているサイトを探します。そうした中から、2つのサイトを選んで、検索方法を解説します。

- 維基文庫 (<https://zh.wikisource.org/>)

フリーなテキストデータのアーカイブであるウィキソースの中国語版です。経書・史書から通俗文学まで多くのテキストデータが登録されています。『三下南唐』も繁体字で全文が収録されていますが、タイトルが『趙太祖三下南唐』になっています。

ウィキソースでは、長編のテキストは巻や章、古典小説であれば回ごとにページを分けて収録していますが、検索対象のページの指定ができません。



ので、「黄石公」(繁体字)で検索するとウィキソース全体から膨大なページがヒットしてしまいます。このため、検索するときに「黄石公 趙太祖三下南唐」のように、書名も合わせて指定しましょう。これで、検索結果の上位に『三下南唐』が並びます。

◆ オンラインテキストアーカイブ

『三下南唐』が収録されたオンラインテキストアーカイブは複数ありますが、ここでは簡体字データを収録する以下のサイトで検索してみます。

- 夢遠書城 (<http://www.my285.com>)

こうしたアーカイブを検索するときには、Googleの「site:」コマンドを使います。このサイトで『三下南唐』の目次のURLは「<http://www.my285.com/gdwx/xs/xy/ztz/index.htm>」、第一回は「<http://www.my285.com/gdwx/xs/xy/ztz/01.htm>」となっており、「<http://www.my285.com/gdwx/xs/xy/ztz/>」以下にデータが置かれていることがわかります。

そうしたらGoogleで、以下のように検索語を入力します。



site:www.my285.com/gdwx/xs/xy/ztz/ 黄石公

「site:」以下に <http://> を除いたディレクトリのURLを記述(コピー&ペースト)することで、検索範囲を指定できます。そのあとに、半角スペースを挟んで簡体字で「黄石公」と入力します。これで、『三下南唐』だけを対象に「黄石公」の登場する箇所を探すことができます。

この「site:」コマンドによる検索は、いろいろと使えますので、覚えておくとよいでしょう。

(千田 大介)

❖ 古典戯曲作品の概要をしらべる

Q: 『永团圆』という伝奇がどのような作品なのか、どうやってしらべればよいのでしょうか？

❖ 戯曲・雑劇・伝奇

「戯曲」という言葉は、日本語で演劇の台本のことをいいますが、中国語では「戯」は演劇・芝居、「曲」は歌曲を指し、伝統的な歌劇スタイルの演劇のことをいいます。

中国では、金・元の時代に北方で「雑劇」と呼ばれるスタイルの演劇が形成され、元から明にかけて、全国的に行われました。南方では南宋から元にかけて南曲が形成され、明代になってから発展して全国で盛んに演じられるようになり、その台本は「伝奇」と呼ばれるようになりました。これらはいずれも歌劇の形式をとっており、その流れを汲む京劇を初めとする現在の伝統演劇も、中国語では「戯曲」ということになります。

このため、伝統的スタイルの演劇の台本も「戯曲」と呼ばれます。逆に、現代中国の話劇（現代劇）の台本は、歌曲形式ではないので、中国語で「戯曲」と言うことはできず、「劇本」になります。

❖ 参考図書でしらべる

さて、中国古典戯曲について項目を立てている日本語の参考図書は多くありません。『中国文化史大事典』（尾崎雄二郎・竺沙雅章・戸川芳郎編、大修館書店、2013）と『中国学芸大事典』（近藤春雄著、大修館書店、1978）は、戯曲作品の項目も立てていますが、一部の有名作品に限られ、『永团圆』はみあたりません。

となると、中国語の参考図書をしらべるしかありません。古典戯曲の辞書・目録として、代表的なものは以下の3種になります。

①『中国古典戯曲存目彙考』（莊一弘編著、上海古籍出版社、1982）

②『古本戯曲劇目提要』（李修生主編、文化艺术出版社、1997）

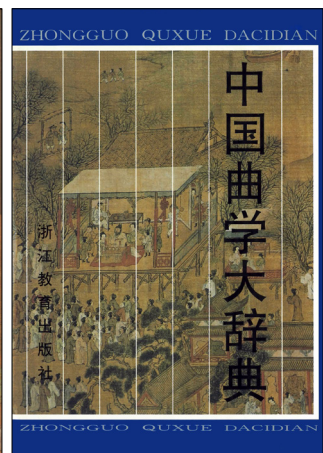
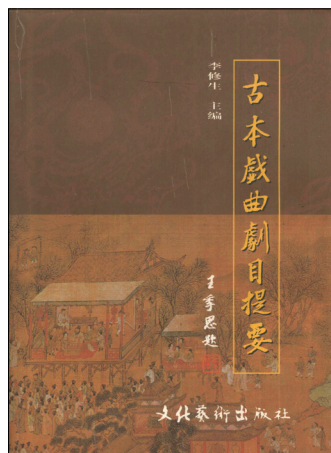
③『中国曲学大辞典』（齊森華・陳多・葉長海主編、浙江教育出版社、1997）

質問の場合は伝奇ですので、以下の書籍も参考になります。

④『明清伝奇綜録』（郭英徳著、河北教育出版社、2005）

収録する作品数が最も多いのは①で、現存しないものも含めて網羅的に収集し、作者の略伝や版本の情報、ごく簡単なあらすじなどを付しています。ただ、1982年の刊行であるため、内容の古さは否めません。

②は、現存の戯曲のみを収録しています。タイトルに「提要」とあるように、作品のかなり詳し





いあらすじが書かれています。作者の情報や、批評、版本などの情報も掲載しています。

③は、古典戯曲の総合的な事典になります。戯曲作品については、現存しないものも含めて項目を立てていますが、項目数は①の2/3ほどにとどまります。当時の最新の研究成果をふまえて、作者、版本、あらすじなどの情報がまとめられています。あらすじは①よりは詳しいですが、②に比べるとはるかに簡略です。

④は現存する南戯・伝奇の総合的な目録で、作者、版本、あらすじなどの情報を掲載しています。実際に版本や鈔本を調査した上で書かれており、記述には信頼が置けます。

これらのうち、最も便利なのは間違いなく③なのですが、解説に問題のある項目も若干見られるので、①は他で見つからないときの最終手段として、②④もあわせてしらべるようにしましょう。なお、②③はCNKI 工具書庫(☞解説:CNKI 工具書庫)に収録されていますので、そちらで閲覧することもできます。

❖ 古典戯曲調査の実際

さて、実際に『永团圆』をしらべてみましょう。まず作者ですが、①②は李玉の作で馮夢竜の改定本の存在に言及しています。③は李玉作、馮夢竜改定とし、④は李玉作と馮夢竜改定本を別々に収録しています。

①②④の記述からわかるように、『永团圆』は

李玉の作で、李玉自らの定本が刊行されていますが、明末の著名文人である馮夢竜の改定本もあります。ところが③は、馮夢竜の改定本しかないかのような書き方になっています。現存テキストについては④が詳細で、明崇禎年間刊本の影印が『古本戯曲叢刊三集』に収録されていること、馮夢竜改定本は明末墨憨齋刊本、清乾隆59年宝研齋刊本があり、後者が『墨憨齋定本伝奇』として影印出版されていることがわかります。②はやや簡略で、①は前者にしか言及せず、③は明刊本があるとは書いていません。

物語については、②④が詳細で、金陵(南京)の才子・蔡文英が、許嫁の江蘭芳、その妹の蕙芳の2人と結ばれるまでの紆余曲折を描いていることがわかります。

上述の①～④でしばしば取り扱いが食い違っているのが、南宋～明初の南曲の戯文(台本)のタイトルだけが記録されており、それと同名の明代中期以降の伝奇が残っている場合の扱いです。②と④は、現存の伝奇に集約して記述しています。①は戯文と伝奇を分けたり集約したりまちまちで、③はできるだけ分けて項目を立てています。

例として『還帯記』を見てみると、①②④は現存する伝奇と戯文とを同じものとして解説しています。③は戯文『裴度香山還帯記』と伝奇『還帯記』、それぞれの項目を立てていますが、どちらも明金陵世徳堂刊本が残るとしており、実質的には同じものを指しています。

特に成立が古い南曲系の作品をしらべる際には、複数の項目に分かれている可能性を考えなくてはなりませんし、その時には『裴度香山還帯記』と『還帯記』のように、違う項目名になっている可能性も考慮する必要があります。

このように、古典戯曲の作品の調査は、1つの参考図書だけで完結できない点に、難しさがあるといえるでしょう。

なお、古典戯曲の調査効率を高めるため、筆者が中心となって構築している中国古典戯曲データベース(<http://ccddb.econ.hc.keio.ac.jp/wiki/>)では、複数工具書の総合索引を公開する準備を進めています。

(千田 大介)

❖ 古典戯曲作品の影印本・排印本をさがす

Q: 『永团圆』伝奇の原文を確認したいのですが、どのように探せばよいのでしょうか？

❖ 影印本を探す

戯曲作品について、目録類で調査すると、どのような版本があるのか、またそれが影印出版されているのかがわかります。

『永团圆』については、前項「古典戯曲作品の概要をしらべる」で見たように、『古本戯曲叢刊三集』と『墨憨齋定本伝奇』の2種類の影印本に収録されていることが分かっています。

このうち『古本戯曲叢刊』は、中国の大規模な古典戯曲の影印叢書で、これまでに1～6集と9集が刊行されています。1～3集には明初から明末にかけての戯文・伝奇が、概ね年代順に収録されています。4集は元・明の雜劇、5集は明末清初の伝奇、6集は清初から乾隆年間までの伝奇、9集は清の宮廷で演じられた長編伝奇になります。

『古本戯曲叢刊』は古風な線装本(糸綴じ本)スタイルでとても高価なうえに、とりわけ初期に刊行された1～3集は国内でも収蔵している機関が稀でしたが、6集の刊行とともに既刊各集が洋装本として再版されたお陰で、だいぶアクセスしやすくなりました。

あとは、一般の中国書籍を探す要領で、CiNii Books (☞解説: CiNii Books)、全国漢籍データベース (☞解説: 全国漢籍データベース) および各図書館のOPACなどを駆使して、最寄りの所蔵機関をさがしてください。なお、国内のOPACやデータベースでは、戯曲叢書の収録作品名(子目)にしばしば入力ミスがありますので、検索しても見つからない場合は中国・台湾のサービスでも試すようにしてください。

このほかにも近年、戯曲の大規模影印叢書の刊行が相次いでいます。北京大学図書館の馬廉旧蔵戯曲を収めた『不登大雅文庫珍本戯曲叢刊』(北

京学苑出版社、2003)、著名な戯曲研究者であった傅惜華のコレクションを影印した『傅惜華蔵古典戯曲珍本叢刊』(学苑出版社、2010)などがそれです。これら新しい叢書は、収録作品がOPACに登録されていないことしばしばです。作品名と叢書名を指定してGoogle検索し、ネット上の目次情報にヒットするのを期待するしかないでしょう。

❖ 排印本探しには手間がかかる

古典戯曲の排印本も多数刊行されていますが、ともすると影印本よりも探すのに手間がかかります。

古典戯曲の排印本には、いくつかの刊行形態があります。まず、戯曲作品の単行本があります。中華書局の「明清伝奇叢刊」、上海古籍出版社の「古代戯曲叢書」といったシリーズが知られており、タイトルをOPACで検索すればヒットします。

残念ながら『永团圆』は、単行本としては刊行されていません。

また、戯曲の選集・総集に収録されている可能性があります。古典戯曲の排印本になっている代表的選集には、以下のものがあります。

● 雜劇

- 『元曲選』(中華書局、1958)
- 『孤本元明雜劇』(商務印書館、1939。1958年に中国戯劇出版社がリプリント)

● 戯文・伝奇

- 『六十種曲』(中華書局、1958)

『元曲選』・『六十種曲』はいずれも明末に編まれたもので、古典戯曲の定本というべきものです。『孤本元明雜劇』は中華民国時代に刊行された元明雜劇の全集に近い選集です。いずれも、目

録などでしらべれば収録情報が載っていますし、また中国国家図書館 OPAC (☞解説：中国国家図書館 OPAC)、NBINet (☞解説：NBINet 聯合目録)、全国漢籍データベースなどで、収録作品を検索することもできます。

影印本と同様、排印本でも近年、いくつかの新たな戯曲選集が編まれています。やはり収録作が OPAC に登録されておらず、検索できないケースが目立ちます。

このほか、別集（個人の作品集）に戯曲が収録さ

れていることもあります。李玉・馮夢竜で検索してみると、『李玉戯曲集』（上海古籍出版社、2004）、『馮夢竜全集』（江蘇古籍出版社、1993）がヒットします。あとは Google 検索などで目次を探してみれば、『永团圆』が収録されていることが確認できるでしょう。

こうした戯曲選集・別集の収録作品の検索についても、今後、中国古典戯曲データベース (<http://ccddb.econ.hc.keio.ac.jp/wiki/>) で対応していく予定です。

（千田 大介）

❖ 古典戯曲作家の情報をしらべる

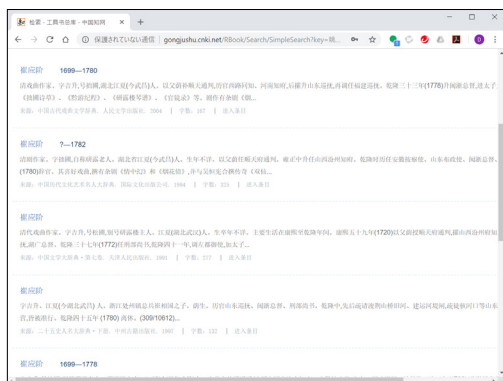
Q: 清代の崔应階という戯曲作家について、どのようにしらべたらよいのでしょうか？

❖ CNKI 工具書庫で検索する

中国の歴史的人物の調査については、「歴史的な人物の伝記情報をしらべる」で解説しましたが、こと戯曲作家となると、戯曲専門の辞書で調べた方が良いでしょう。元の関漢卿、明の湯顯祖、清の江昇といった著名作家でしたら一般の辞書にも載っていますが、大半の戯曲作家は科挙で進士に合格できなかった地方のローカルエリートであるため(☞解説：官僚登用制度)、一般の人名辞典には収録されないこともしばしばです。また、戯曲専門の辞書の方が、戯曲研究の成果が反映されやすい傾向にあります。

さて、CNKI 工具書庫で簡体字「崔应阶」を検索すると、執筆時点で6つの項目がヒットします。それらを見比べると、平生の事績や主要作品などの説明は、だいたい似かよっていますが、生没年が食い違っていることがわかります。以下にまとめてみましょう。

- 1699—1780：『中国古代戯曲文学辞典』（人民文学出版社、2004）



CNKI 工具書庫「崔应阶」検索結果

- ?—1782：『中国歴代文化芸術名人大辞典』（国際文化出版公司、1994）
- 1699—1778：『中国曲学大辞典』（浙江教育出版社、1997）
- 生卒年不詳：『中国文学大辞典』第七卷（天津人民出版社、1991）
- 明記せず：『二十五史人名大辞典』下冊（中州古籍出版社、1997）、『南陽歴代郡守知府』（三秦出版社、2006）

いったいどれが正しいのか、ちょっと踏み込んで調査する必要があるようです。

❖ 古典戯曲データベースで検索する

筆者が中心となって運営する中国古典戯曲データベース (<http://ccddb.econ.keio.ac.jp/wiki/>) では、さまざまな古典戯曲・伝統演劇関係の参考図書類の人名に関する項目の総合索引機能を提供しています。現在、以下の書籍の人名関係項目のデータを登録済みです。

- 『中国戯曲曲芸詞典』・『中国曲学大辞典』・『方志著録元明清曲家伝略』・『明清戯曲家考略統編』・『中国古典戯曲存目彙考』・『明清伝奇綜録』・『中国古代戯曲家評伝』・『明清戯曲家考略』・『明清戯曲家考略統編』・『明清戯曲家考略三編』・『明清戯曲家考略四編』・『中国戯曲志』・『中国崑劇大辞典』・『京劇知識詞典』・『中国豫劇大辞典』

検索すると項目の掲載ページがまとめて表示されるので、戯曲関連人名を調査するときに、まずここで検索しておくとういでしょう。『中国戯曲志』全巻の目次データなど、現代の伝統劇関連の



中国古典戯曲データベース「崔應階」検索結果

人名もあわせて、合計1万1千以上のデータを登録していますので、古典戯曲ばかりでなく伝統劇関連の調査にも活用できるでしょう。

さて、同データベースで繁体字「崔應階」を検索すると、『明清戯曲家考略』p.93に掲載されていることがわかります。『明清戯曲家考略』は、戯曲史研究家の鄧長風による、戯曲作家の伝記的記録の考証を集めた書籍です。上海古籍出版社から1994年に刊行されたのち、『続編』(1997)、『三編』

(1999)が刊行され、2009年には既刊三編と1999年に没した鄧氏の遺稿をまとめた『四編』を合わせた『明清戯曲家考略全編』が出版されています。

さて、崔應階の生没年ですが、『明清戯曲家考略』では、1699年生、1780年没であると考証しています。CNKI 工具書庫の検索結果の中では、『中国古代戯曲文学辞典』が正しいことになります。

もちろん、『明清戯曲家考略』の考証が正しくない可能性もありますので、CNKIで関連論文を探せば確実ですが、崔應階に関しては資料・考証に特にあやふやな点は見あたりませんので、問題ないでしょう。

ローカルエリートの伝記情報については、資料的な不確かさがつきまとい、また調査・研究する人が多くないために、一人のちょっとしたミスが広く流通するような事態も発生しやすくなります。そのような戯曲作家を調査する際には、複数の参考図書を対比して、場合によっては研究書や論文をも参照する必要がある、どうしても生じてしまうのです。

(千田 大介)

❖ 白話語彙の意味をしらべる

Q: 清の李漁の小説『無声戯』第九回に「毎日也不過替人上兩双鞋子，打幾個兪頭，有什么大進益」とありますが、「兪頭」が辞書にみあたらず、どういう意味なのかわかりません。

❖ 文言と白話

元・明・清の通俗文学には、白話の文体や語彙が使われています。「白話」とは、歴史的な書き言葉（ほぼ日本でいう漢文に相当）の「文言」に対して、話し言葉のことを指します。古くは六朝の小説に使われており、以後、唐宋の伝奇小説、敦煌変文、詞などにも見えますが、代表的な白話文学のジャンルといえば、元・明・清の通俗文学、戯曲・小説などということになります。

質問の「兪頭」は、そうした白話の語彙になります。日本の中日辞典・漢和辞典では、『中日大辞典』（第三版、愛知大学中日大辞典編纂所編、大修館書店、2010）が清代後期の語彙をある程度収録しているくらいで、諸橋『大漢和辞典』などには白話語彙があまり収録されていません。このため、中国の辞書を使うことになります。

❖ 白話語彙の辞書

『漢語大詞典』（漢語大詞典出版社、1986-1993）は、

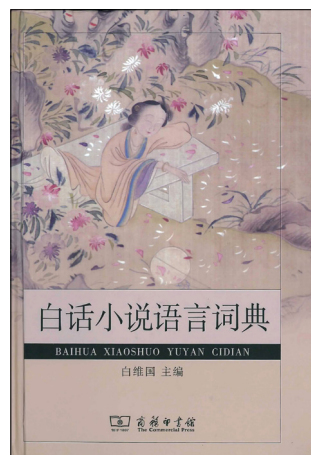
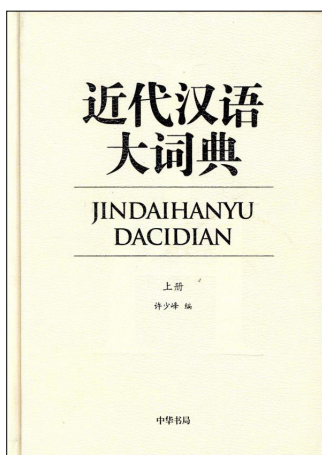
白話語彙をかなりの程度収録しています。ただ、編纂時期が若干古いこともあって完全とはいえませんので、白話語彙の辞書を使うべきです。

辞書調査といえば CNKI 工具書庫ですが、こと白話語彙の辞書に関しては、全くといっていいほど収録されていませんので、紙媒体に頼ることになります。

比較的新しく、かつ収録語彙数の多い白話語彙の辞書は以下の2つです。

- 『近代漢語大詞典』（許少峰編、中華書局、2008）
- 『白話小説語言詞典』（白維国主編、商務印書館、2011）

収録語彙数は前者が約 50,000 語、後者が約 49,000 語になります。前者は六朝や唐宋の小説、元代以降の小説・戯曲や説唱芸能などに見える語彙を網羅するのに対して、後者は白話小説のみを対象としています。



これらの辞書で「兪頭」を引くと、『近代漢語大詞典』は「zhǎng tóu」と読み「釘在鞋跟上的垫块」（靴のかかどに打ち付ける補強材）とし、『白話小説語言詞典』は「zhǎng tou」と読み「为保护鞋底而另加的前底与后跟」（靴底を保護するために付け加える、前側の底とかかど）としています。いずれにせよ靴の底に取り付ける補強材のことを指していることがわかります。

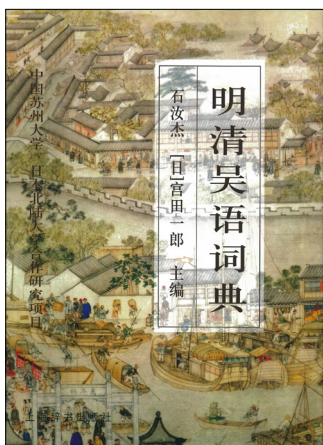
これをふまえて質問の部分を読すと、「毎日、頼まれては二足ばかりの靴を作り、靴底を幾つか補強する、大儲けなんてありやしない」といったところでしょ。

❖ 方言辞書・専門辞書・外来語辞書

明清の白話文学には、当時、文化・経済の先進地域であった江南地方、とりわけその中心都市であった蘇州をバックボーンとするものが数多くあり、このため作中に蘇州方言、呉語が使われていることもしばしばです。明清時代の呉語については、以下の辞書があります。

- 『明清呉語詞典』（石汝傑・宮田一郎主編、上海辞書出版社、2005）

特に江南の地域性が考えられる作品については、白話語彙の辞書とあわせて、この辞書も引いておくべきです。その他の地域については、明清時代の方言辞書が刊行されていることは稀ですが、現代



の方言辞書が参考になることもあります。

一部の有名な白話小説については、専門の辞書が刊行されていることもあります。以下に3冊だけ、例を挙げます。

- 『水滸語詞典』（李法白・劉鏡英編著、上海辞書出版社、1989）
- 『金瓶梅詞典』（白維国編、中華書局、1991）
- 『紅樓夢語言詞典』（周定一主編、商務印書館、1995）

いずれも20世紀の刊行で、前掲のより新しい2冊があれば用は足りるのですが、それぞれの作品を原文で読み込む際には重宝します。

元雜劇には、モンゴル語などに由来する外来語がしばしば使われています。「必赤赤」・「米罕」などがそれで、明・清代の戯曲でも踏襲されています。このような外来語については、以下の辞書で調べます。

- 『古典戯曲外来語考釈詞典』（方齡貴、世紀出版集団・漢語大詞典出版社・雲南大学出版社、2001）

この辞書を引くと、「必赤赤」がモンゴル語「bičičie」の音訳語で「秘書」であることが、「米罕」が「miqan」・「miqa」の音訳語で「肉」を意味することがわかります。収録語彙数は約430語とさほど多くはありませんが、外来語としての由来・意味を詳細に知ることができますし、『近代漢語大詞典』未収録の語彙も多く項目が立っています。

❖ 簡体字と異体字の問題

以上の辞書はいずれも簡体字中国語で書かれています。簡体字では、「髮」「發」が「发」になったり、「鬪」が「斗」、「雲」が「云」にまとめられたりするなど、複数の繁体字が1つの簡体字に統合されている例が多々ありますので、辞書を使う際にはもとの繁体字との対応に注意する必要があります。

これは意外と厄介な問題です。例えば『近代漢

語大詞典』では p.863 と p.868 の 2 箇所「家伙」が立っています。用例の原典を確認すると、後者が「傢伙」の簡体字であることがわかりますが、繁体字が注記されていないため、同じ語が重複掲載されているように見えます。このほかにも「干結」・「凶器」など十数語、同様の例が見られます。

一方、『白話小説語言詞典』は「家伙」を 1 つしか立てていませんが、それはそれで繁体字の「家伙」・「傢伙」の区別が消えています。

もちろん、通俗文学のテキストは、大半が商業出版の坊刻本であり（☞解説：官刻本・家刻本・坊刻本）、一般に正しい漢字表記を厳密に突き詰めているような性質のものではありませんし、そもそも白話の語彙は発音が前にあり、それにどの漢字をあてるのか、正書法が確立しているとは言いがたく、「家伙」と「傢伙」が意識的に書き分けられていたのかも疑問ではあるのですが、いざ「傢伙」という語を調べようとしたときに、簡体字で「家伙」に書き換えられていることが分かっていないと、うまく見付けられないこととなります。各辞書の凡例に目を通しておき、繁体字と簡体字との関連についても、中国語辞書で逐一確認するようにしましょう。

❖ 拡張簡体字に注意

もう 1 つ厄介なのが、いわゆる「拡張簡体字」の問題です。通俗文学のテキストには、多くの異体字が出てきます。その大半は誤字で、翻刻の際に書き換えられてしまうようなものですが、白話

語彙の辞書では、そうした異体字についてもある程度、項目を立てています。そうした異体字を簡体字の辞書に収録する際には、中国の法令に従い、簡単字の一覧「簡化字総表」に収録されない漢字についても、当該漢字の構成部品に簡体字がある場合には、簡体字化しなくてはなりません。このため、例えば「攔」（刀を撫でつける。『水滸伝』などに見える）は、「拦」という簡体字に書き換えられています。こうした、通常使われることのない「簡化字総表」未収録の漢字を無理矢理簡体字化したものを、俗に拡張簡体字と呼びます。

白話語彙の辞書で、テキストに実際に見える異体字を検索できるように、他の代表的字形に置き換えることなく、簡体字化して掲載すると、結果的に多数の拡張簡体字が出現することになります。

例えば、『西遊記』第五十九回、「鎖鞵猿馬」の「鎖鞵」は、辞書には「锁鞵」という拡張簡体字で収録されています。さらに面倒なのは、通俗文学作品の排印本やオンラインテキストでは、こうした難読字を別の字に書き換えているケースが見られることです。質問の「允頭」も、「鞍頭」・「掌頭」に書き換えられているテキストがあります。これは「古典通俗小説の排印本」でも説明したように、一般向けの排印本やオンラインデータの品質が低いことを物語っていますが、いずれにせよ、白話語彙の辞書を引く前には、定評のある版本にあたって漢字表記を確認するとともに、繁体字と簡体字との対応に注意する必要があります。

（千田 大介）

現代中国語をしらべる

❖ 中日辞典に載っていない中国語の意味

Q: 中国語の雑誌を読んでいたら、大学関係の記事で「蹭課」という語が出てきましたが、中日辞典に載っていません。どうやって調べればよいのでしょうか？

❖ 日本の辞書の限界

中国の書籍・雑誌・新聞などを読んでいると、日中・中日辞典に載っていない語にしばしば出くわします。

よくあるのが新語です。従来なかった新造語のほか、本来とは全く異なる意味で使われるようになった同形新義語もあります。例えば「粉丝」は、本来「はるさめ」の意味ですが、ゼロ年代はじめから「ファン」の音訳語としても使われるようになってきました。紙版の辞書の改訂には長い時間と労力が必要ですので、こうした新語の収録は思うに任せません。

また辞書には分量の制限があります。日本の中日辞典は、大きくても十数万字の中辞典クラスに過ぎませんので、使用頻度の落ちる語彙は割愛せ

ざるを得ません。角川書店の『中国語大辞典』もありますが、中国の爆発的経済発展と社会の大変革が始まる以前の1994年の刊行ですので、内容の古さは否めません。

❖ オンライン辞書をしらべる

紙版の辞書で見つからなかったら、オンライン辞書で検索してみましょう。中国語のオンライン辞書としては、以下のサービスがよく使われます。

- Weblio 日中中日辞典 (<https://cjc.weblio.jp/>)
- 北辞郎 (<http://www.ctrans.org/>)

Weblio 日中中日辞典は、白水社『中国語辞典』のほか、以下のオンライン辞書・辞典を一括検索できます。



weblio 日中中日辞典



北辞郎

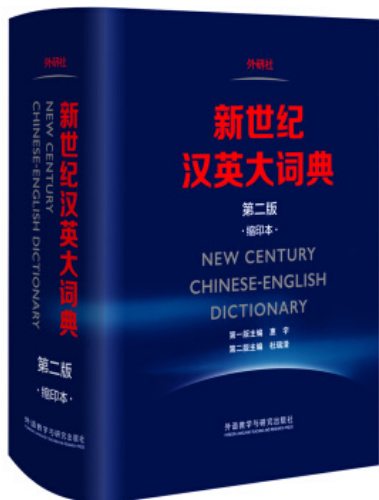
Weblio 中国語翻訳辞書、EDR 日中対訳辞書、日中中日専門用語辞典、中英英中専門用語辞典、Weblio 中日対訳辞書、Wiktionary 日本語版・中国語版

ウェブサイトの説明によれば「60 万語以上の日本語、100 万語以上の中国語」を検索可能であるようですが、各辞書・辞典で重複登録されている語が相当に多いので、実数はずっと少なくなると思われる。

また Weblio 収録辞書のうち、白水社『中国語辞典』は定評のある紙版の辞典ですが、それ以外は質がいささか落ちますので、注意が必要です。

北辞郎は、翻訳家などが共同で構築しているオンライン辞書で、新語や技術用語なども豊富に登録されていて、重宝します。名前が「英辞郎」のもじりであることからわかるように、翻訳家が翻訳時に参照する単語リスト的な性格のもので、用例や語句の解説は基本的に付されません。学習者ではなく、中国語プロパー的な辞書であるといえるでしょう。

さて、質問の「蹭课」という語ですが、上記 2 つのオンライン辞書で検索しても、ヒットしませんでした。



❖ 中英辞典をしらべる

中日辞典やオンライン辞書で見つからなかったら、今度は中英辞典を引いてみましょう。なぜ英語の辞書なのかといえば、中国国内のニーズが高く、それだけに出版社間の競争も激しく、よい辞書が出版されているからです。

中国では中英辞典が多数刊行されていますが、とりわけ現代の新しい語彙に強いのが、以下の辞書です。

- 『新世纪汉英大词典』第二版（外语教学与研究出版社、2016）

この辞書を引いてみると、「蹭课」が立っており、英語の語釈は「sit in on a class (for free); sneak into a class; audit a class (for free)」となっています。日本語に訳せば「授業を（ただで）受ける、授業に忍び込む、（ただで）聴講する」、要するに「授業にもぐる」という意味だとわかります。次の項目「蹭网」が「WiFi タダ乗り」というのも、いかにも現代らしいといえるでしょう。

蹭 cèng (動) ① rub; scrape; scratch: ~破手上的皮 graze one's hand || 混凝土地面把我的膝盖~疼了。I scraped my knee painfully on the concrete floor. || 在马鞍下放个软垫以免~破马背。Put a soft pad under the saddle to avoid chafing the horse's back. || 这双新鞋太紧了,老~我的脚后跟。The new shoes are so tight that they keep rubbing on my heels. ② rub against sth and get stained: ~一脸灰 smear one's face with dirt || ~一身泥 stain one's clothes with mud all over || 浅色的地毯太容易~脏了。Pale carpets stain too easily. || 油漆未干,小心别~着。Mind the wet paint. ③ (口) cadge; mooch; scrounge: ~饭 have a free meal; scrounge/sponge a meal; bum a meal off sb || 他想在我这儿~吃~喝。He tried to mooch meals and drinks off me. || 下班后我想~汤姆的车回家。I want to cadge a lift home from Tom after work. || 这家伙总想从父母那儿~点钱。This guy always intends to scrounge some money off his parents. ④ move slowly; drag one's feet; trail along; shuffle; 他一步一步慢慢往前~。He inched his way forward. / He dragged along. ⑤ delay; dawdle; 别~时间,赶快干。Don't dillydally/dawdle; just get on with it. ⇨磨~

蹭蹭 cèngcèng (动) (书) meet with setbacks; be down on one's luck; be frustrated in one's career

蹭课 cèngkè (动) sit in on a class (for free); sneak into a class; audit a class (for free); ~文化已经席卷了各大高校。The trend of auditing university classes has swept across universities.

蹭网 cèngwǎng (动) siphon Wi-Fi; siphon off an Internet connection; piggyback onto a Wi-Fi connection; 设置一个高强度密码防止别人~ set a strong password to prevent people from siphoning off your Internet connection || 免费~ siphon a free Wi-Fi || ~者 Wi-Fi intruder

このように『新世紀漢英大詞典』は「読者第一、実用性重視」をモットーに、現実生活に密着した新しい語彙を、さまざまな分野から広く収集しています。収録語彙は15万項目前後とそこまで多くはありませんが、収録語彙の鮮度が高く、他の辞書が収録していない語がこの辞書で見つかることもしばしばです（この辞書の邦訳中日辞典版を、是非とも出版してもらいたいものです）。

なお、老舗の商務印書館からも『新時代漢英大詞典』第2版（2014）という似たタイトルの辞書が出ていますが、こちらは比較的オーソドックスな作りで、「蹭課」も立っていません。

❖ ネットで「用英语怎么说」検索

以上のいずれの方法でも見つからなかった語は、ネット検索を駆使して意味を調べるしかありません。これにも、いくつかのコツがあります。

まずは、「蹭課を日本語で」とGoogleで検索してみます。執筆時点では、Google翻訳は「ラブレスン」となり、以下の検索結果でも、妥当な結果は得られませんでした。

そうしたら次に、Googleで「蹭課用英语怎么说」（蹭課は英語で何というのか）と中国語で入力して検索してみましょう。すると、百度知道（日本の「はてな」と同種のサービス）や、中国語・英語対訳のページなどがヒットしますので、そこから英訳語を確定し、さらに日本語訳を調べることで、意味を知ることができます。

日本でも中国でも、最も学習者が多く社会的ニーズが高い外国語は、当然のことながら英語です。辞書やオンラインコンテンツも日中・中日に比べて遥かに充実しています。このため、英語を経由した方が意味が確定できる語も多くなります。意味の分からない中国語を調べる最終手段として、この「用英语怎么说」検索を覚えておくと、何かと役に立つことでしょう。

❖ 急激に変化する中国語

20世紀末以来、中国語は急激に変化しています。経済発展や社会の変化につれて現れた、新し



「蹭課」の「用英语怎么说」検索結果

い概念や事象を指す言葉が次々と生まれているほか、同じ言葉であっても数年のうちに意味やニュアンス、使われ方が一変してしまうこともしばしばです。

例えば「小姐」という語は、台湾・香港では以前から「Miss」の訳語として定着していましたが、大陸では1990年代前半に、若い女性やウエイトレスなど女性接客員に呼びかける際に、「お姉さん」くらいの感覚で使われるようになりました。ところが2000年前後に個室カラオケやバーの「小姐」（ホステス）が大量に出現すると、この語が水商売的なニュアンスを帯びるようになり、一般の女性への呼びかけに使われなくなりました。

また「宅男」・「宅女」は、ゼロ年代初め、日本語の「おたく」の訳語の一つとして現れましたが、「宅」という字面に引きずられて、おたくコミュニティはいざ知らず、一般では「ニート」・「ヒッキー」に近い意味で使われるようになっており、『現代漢語詞典』第7版もそのような意味の語として収録しています。

人は一度学び覚えたことを、ともすると永久不変であるかのようにとらえがちですが、こと「分進秒歩」ともいわれるほどに変化の激しい中国においては、言葉もダイナミックに変化し続けているのであり、中国語のプロパーはそうした変化を常に見つめ続けていなければならない宿命を負っていることを、肝に銘じておく必要があるのです。

（千田 大介）

❖ 日中辞典に載っていない日本語の中国語訳

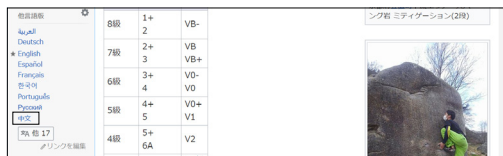
Q: 東京五輪の正式種目になった「ボルダリング」は中国語で何というのでしょうか？日中辞典には載っていません。

❖ Wikipedia を活用する

名詞の中国語訳を調べる際に重宝するのが Wikipedia です。まず、日本語 Wikipedia で「ボルダリング」のページを開き、次に左下の「多言語版」から「中文」をクリックします。すると、中国語版の「ボルダリング」のページに切り替わり、中国語で「抱石」ということがわかります。

中国語ページを開いたときには、記事の冒頭の概説部分をよく読むようにしましょう。同義語が書かれていることがありますし、場合によっては、日本語とは意味のずれる語になっていることもあります。

もう1つ注意が必要なのが、中国・台湾など地域による言い方の違いです。ページ上部の「不转换」をクリックすることで、中国・香港・台湾などのページに切り替えて確認できます。ただし、簡体字ないしは繁体字に自動で変換しているだけ、というケースもありますし、また Wikipedia に間違いは付きものですので、見つかった中国語を検索して実際の使われ方を確認する必要もあるでしょう。



❖ オンライン辞書を使う

「中日辞典に載っていない中国語の意味」で紹介したオンライン辞書は日本語語彙の検索にも使えます。

北辞郎 (<http://www.ctrans.org/>) で「ボルダリング」を検索すると「抱石」という訳語が出てきます。一方、Weblio 日中中日辞典 (<https://cjjc.weblio.jp/>) では、EDR 日中対訳辞書で「徒手攀登巨岩」が、Weblio 中日対訳辞書で「抱石」がヒットしますが、収録辞書間のゆれがあり、これだけで訳語を確定することができません。

❖ 英語経由でしらべる

Wikipedia は百科事典という性質上、学術タームや名詞を多く収録していますが、動詞・形容詞はあまり掲載されていません。その他のオンライン辞書も、新語への対応には限界があります。

例えば近頃ネットでよく目にする「バズる」という言葉は、執筆時点では、Wikipedia やオンライン辞書に項目が立っていませんし、Google で「バズる 中国語で」と検索しても、「パズル」の訳語が並んでしまっとうまく調べられません。

こうしたケースでは、まず「バズる」の英語訳を調べます。「バズる 英語で」で検索すると、いくつもの英語関連サイトで「go viral」であるとされていることがわかります。次に「go viral 中文」あるいは「go viral 用中文怎么说」で検索してみましょう。すると、決まった訳語はないものの「迅速走红」「变得火爆」「疯传」「火了」「爆红」などと訳されていることがわかります。これらの語を改めて検索してみて、使われかたを確認し、訳語を決定するとよいでしょう。

(千田 大介)

❖ 欧米人名の中国語訳

Q: ドイツのボルダリング選手、アレックス・メゴスの中国語のニュース記事を探しています。中国語でどのような表記になるのでしょうか？

❖ Wikipedia でしらべる

現代の人名ですので、まずは Wikipedia 日本語版にページが立っていないかを確認しましょう。残念ながらありませんでした。

次に、Google で「アレックス・メゴス」を検索し、ドイツ語の原文表記を伴っている記事を探します。そうすると、「Alex Megos」であることがわかります。

Wikipedia の言語をドイツ語「Deutsch」に切り替えて、「Alex Megos」を検索します。ドイツ語版にはページが立っており、アレックスは愛称で、正式には「Alexander」であることがわかります。中国語ページは残念ながら作られていません。

さて、人の名前、よほどのキラキラネームでもない限り、同じ姓、同じ名の人がいるものです。そこで、Wikipedia（日本語で見つからない場合は英語版・ドイツ語版などを使いましょう）で姓・名を別々に検索し、同姓あるいは同名のドイツ人を探します。見つかったページが中国語ページに切り替えられれば、どのように訳されているのかが分かります。できるだけ、複数の人物について訳語を調べた方が良いでしょう。

残念ながら、「Megos」姓の人物で Wikipedia 中国語ページが立っている人はいないようです。

❖ 工具書でしらべる

次に、オーソドックスに工具書でしらべてみましょう。中国でドイツ人名の訳名を調べる際には、以下の工具書を使います。

- 『徳語姓名訳名手冊(修訂本)』(新華社訳名室編、商務印書館、1999)

中国国営の通信社である新華社は、海外人物の中国語訳名を決める役割を果たしており、このハンドブックが海外人名の中国語訳を決定する際の規範になります。ドイツ語のほか、英語・フランス語・スペイン語・ポルトガル語・ロシア語・イタリア語についても、姓名訳名ハンドブックを出版しています。

ハンドブックで「Alex」が「亚历克斯」、 「Alexander」が「亚历山大」になることがわかりますが、「Megos」は掲載されていませんでした。

❖ 最終的にはネット検索

これまで得られた情報をもとに、「Alex Megos 亚历克斯」、「Alexander Megos 亚历山大」で Google 検索してみます。すると、「亚历克斯・梅格斯 (Alex Megos)」という例がみつかりました。従って、「亚历克斯・梅格斯」あるいは「亚历山大・梅格斯」が中国語訳名ということになります。

ただ、Google・百度などで検索してみてもほとんどヒットしないので、中国国内ではまだ報道もされず知名度も低いことが窺えます。ともなると、将来的に中国で知名度が上がると、新華社によって別の訳名が与えられるかもしれません。

ところで、現代の人名については、中国と台湾とで中国語訳名が食い違っていることがしばしばあります。例えば、アメリカ第 45 代大統領のドナルド・トランプは、中国では「唐纳德・特朗普」、台湾では「唐納・川普」と訳されます。調査にあたっては、こうした可能性も考慮して、万全を期して下さい。

(千田 大介)

特集3

2018年夏期 中国視察報告

佐藤仁史・千田大介・師茂樹

❖ 視察の主旨

漢字文献情報処理研究会および日本学術振興会科学研究費補助金「情報時代における中国学研究・教育オープンプラットフォームの構築」（平成28年度～30年度、基盤研究（B）、研究代表者：二階堂善弘、課題番号：16H03351）では、研究の一環として、中国の学術情報系企業、創新力博社とCNKI社の視察を行った。参加者は、佐藤仁史・千田大介・師茂樹の3名である。

創新力博社について、千田が本誌第6号「中国の人文情報処理企業の最新動向——中易中標と創新力博」で紹介しているが、10年以上が経過し、洛陽入力センターが開設され、同社が手がけた『中華經典古籍庫』が発売されるなど状況が変わっているため、改めて訪問してお話を伺うことにした。

CNKIは今や日本の中国学研究者で知らぬものはない、中国を代表するオンラインジャーナルである。本誌でも創刊号のレビューで紹介しているし、第15号では論文検索について、第17号では工具書館について、それぞれ詳細に解説している。同社の概略的情報はWebなどを通じてある程度わかるが、同社の生い立ちや論文やデータの扱いなどなど、さまざまな疑問点もあったため、今回視察のはこびとなった。視察にあたって仲介の労をとって頂いた東方書店に感謝申し上げる。

❖ 日程と概要

インタビューは、事前に質問項目を送っておき、当日あらためて質問する、という方式で行った。

◇ 北京創新力博社 北京本社 (p.149)

8月20日午前に訪問し、王曉波社長にインタビューした。インタビューには、同社の朱江氏も同席した。

◇ 北京創新力博社 洛陽入力センター (p.158)

8月21日午前に訪問し、同社の劉智氏の案内で古典文献入力の工程を見学した。また、同氏に前日の補足的なインタビューを行った。

◇ CNKI 北京本社 (p.162)

8月23日午前に訪問した。CNKI側では劉蓉・馮齊・李昕蔚・劉燕清の4氏にご案内いただいた。うち劉蓉・劉燕清両氏は日本語が話せる。

当日はまず、同社内の文献デジタル化作業の現場を見学し、その後、会議室にて同社の概要についてプレゼンテーションを受け、質疑を行った。プレゼンテーション部分については、QA形式を採らずに概要をまとめて掲載している。

なお、以下の視察・インタビュー報告は、テープ起こした中国語（CNKIは一部日本語）を千田が翻訳・整理・編集し、それを佐藤・師が確認・修正したものである。

北京創新力博社 北京本社

2018.8.20 AM9:30 ~ 於北京市昌平区竜徳・紫金

◇ 王曉波氏のプロフィール

—どちらのご出身ですか？

王曉波（以下「王」）：河南の洛陽です。

—あなたのプロフィールを教えてください。

王：大学は合肥の中国科技大学です。もとは北京にありましたが、合肥に移転したんです。専攻したのは数学でした。数学はやはり抽象的でしたね、実際の応用という点で。内容が理論的な方面に偏っていて、将来、どのように応用できるか見通せませんでした。今ではAIやディープラーニングがありますが、すぐく勉強ができる同級生がいて、賞を取ったりしていました。彼は理論を研究していましたが、私たちは実践するしかなかったんです。その後、中国科学院ソフトウェア研究所の大学院生になりました。当時、ある会社で実習していたのですが、その会社が『四庫全書』を開発していて、私はその責任者になり、卒業後はそこで仕事することになりました。書同文社は皆さんご存じですよ。

書同文で実習から仕事まで、だいたい6年働きました。その後、2003年に書同文を離れて、新しい会社を作りました。というのはその頃、社会が大きな変革期にさしかかっており、インターネットが盛んになっていろいろな構想も出てきていたんですね。でも、私はエンジニアとして97年から5～6年文献デジタル化に従事して、やはりこの分野でやっていきたかったので、書同文を離れてから2003年に今の会社を立ち上げました。いろいろな会社がありましたが、考え方が合わなかったの

で、自分でやることにしたのです。基本的には、文献のデジタル化の技術とサービスを重視しています。会社を立ち上げてから、自分でこのデジタル化システムの開発を始めて、2004年には最初の実用的なバージョンができました。愛如生社は知っていますよね、中国基本古籍庫を作っている会社ですが、彼らも我々のシステムを使ってデータを作っています。

—愛如生は書同文のシステムを買ったと聞いていますが？

王：彼らは2000年に書同文のシステムの使用権を買いましたが、3年後に我々のシステムを使うようになりました。文献をデジタル化したいのは専門の会社や一部の大学でしょうが、このシステムを使うところはあまり多くありません。私たちは主に顧客向けのデジタル化代行サービスを中心にやっています。じつのところ、文献デジタル化システムはWordなどとは違って、使いこなすには専門知識や一定の時間が必要で、しかも人員を揃える必要





があります。多くの企業・機関にしてみれば、すぐに使えない上に手間がかかりますね。だから我々はデジタル化代行サービスに特化して、たとえば書籍を預かって、我々がスキャニングを行って最終的に完成したデータを編集して登録して、データベースシステムを完成させる、というふうにやっています。

—数学科のご出身でその後ソフトウェア研究所の大学院生になられたとのことですが、同級生であなたのように創業した人はいますか、それとも社員が多いですか？

王：大学院では3年のうち、一緒に授業を受けたのは1年だけで、その後、みな指導教員に従ってプロジェクトに従事しましたから、大部分の同級生とは連絡がありません。一人、年上の友人は、やはりソフトウェアをやっていたが、『四庫全書』に従事しているときに、彼は全文検索エンジンを作っていました。指導教員が同じだったので仲が良かったのですが、彼は確か2000年に企業を立ち上げて、チップの設計をやって、私より成功していました。

—同窓会は開かれないのですか？

王：学部はありますが、大半はアメリカに渡ってしまいました。アメリカから帰ってきた同級生は、大学で教授になったのが多いですね。アメリカ在住の同級生も、研究者が多いです。会社を立ち上げた人もいますが、多くないです。アメリカで会社を経営している人もいる

ようですが、逆に国内で創業した人は少ないですね。

◇ 創新力博社設立の経緯

—あなたが書同文社を離れてわざわざ会社を立ち上げたのは、なぜですか？

王：書同文は実際のところ、『四庫全書』電子版の完成だけを目標に作られた会社でした。だから、彼らのソフトには汎用性がありませんでした。『四庫全書』電子版の閲覧ソフトを使ったことがあると思いますが、枠組みが固定されていて、新たに本を登録できませんよね。その後の『四部叢刊』も同様で、閉じられた具体的な目標のためのシステムになっています。我々がやったのは、この作業をオープンなものに作り替えよう、ということです。我々のシステムは、必要な書籍をデジタル化して、完成後にはさまざまな書籍とあわせて、大きなデータベースを構築できます。

デジタル化を始めたころ、2000年よりも前の頃はまだ模索の段階で、どうしたらいいのか誰にも分かりませんでした。その頃、唯一の基盤はユニコードでしたが、まだ3.0かそれより前のバージョンで（訳者注：Unicode 3.0は1999年、3.1は2001年。）、でも作り方も分からなければ、ツールもありませんでした。我々の発想は、工業の経験を模倣して、流れ作業のシステムを構築しようというものでした。文献のデジタル化は多くのタスクを内包しています。そもそも内容は非常に多く、1ページの中にも複雑な形式が含まれていて、たとえば本文と注釈があり、表や標題があるかもしれないし、千田さんに頼まれた戯曲はとりわけ複雑で、曲牌・セリフ・ト書き・注などが含まれていますね。スタイルが複雑なのはどの古籍も同じで、そもそも使っている文字からして現在（の簡体字）と全然違います。それに、高品質である必要があります。我々が納品する際の錯誤率は、いずれも一万分の一以下です。

従業員の問題も避けられません。中国では一人っ子政策で少子化が進んでいて、今の若者はみな家では宝物なんです。父方、母方の祖父母、それと両親、6人が1人の子どものご機嫌を取っています。そうした人が成人した結果、中国の現代の若者からは、勤勉とかいったことが、かなりの部分、失われてしまいました。私たちの会社でも、若者を雇用すると、知識などは親世代よりもしっかりと学んでいるかもしれませんが、ちょっとでも叱ろうものなら、翌日といわず、その日のうちに辞めてしまうんですよ。だから、知識や教養だけでなく、素質を見ないと、デジタル化という業務をやっていけません。もしも、本の内容がわかって、繁体字も分かって、入力もできれば、マークアップもできる、そうした完全な人材を養成するには長い時間が必要で、熟練するのに2～3年はかかります。そのあとで、パッと辞められてしまったら、人材養成が無駄になってしまいます。

だから我々の業務は、人の技能を分散して、流れ作業の生産ラインを通じて行う必要があります。そうすれば、仕事に必要な技能が単純化され、2～3日の訓練でできるようになります。いくつかの技術はちょっと複雑ですが、全体の工程の比率でいえば、100人のうち、中核となるしっかりした人が5～6人ほど、各段階の仕事の責任者になってくれれば大丈夫です。会社もうまくコントロールできます。100人のチームを組めば、かなり大量の仕事をごこなせます。

このため、我々はデジタル化を、いくつかの工程に分割しました。大きく、レイアウトの分割、文字の校正、入力、マークアップに分かれますが、この4つの工程にも、制作・点検などが含まれるので、典型的な作業工程は17～18段階に分かれます。こうすれば、才能や経験が一切ない人が来ても、2～3日練習すれば仕事できるようになりますし、何か事情があって辞める人がいても、簡単に交替できます。

流れ作業のほかに重視しているのは、コンピュータを人が校正するサポートに活用する、という点です。レイアウトの分割を例に取れば、まず自動で認識させますが、100%正確ではあり得ません。でも自動で認識されていれば、人は修正するだけでよくなります。要するに、まず機械に作業させ、人は誤りの修正に専念することで、作業量を減らすという思想です。また、作業を細分化し、一工程で一つの作業だけをするようにして、それに対応したツールを提供します。これは、効率の向上と品質の向上、2つの面で効果があります。品質の向上については、ある種の間違ひが見つかったら、そうした箇所を強調表示するなど、さまざまな対策を講じています。例えば文字を入力するとき、一行一行見ながら入力していると、どうしても行や字を飛ばしてしまうものです。我々は、一文字一文字を切り出して、独立したものとして扱い、その文字に一つのエリアを与えます。そして、レイアウトを分析したら、レイアウトのエリアを各文字のエリアにかぶせて対応させます。文字が漏れていたなら、カバーされていないので、画面上に表示されます。このようにタスクを細分化した上で、それぞれのタスクについて検討し、改善しました。全体の改良が難



しくても、具体的な個々の作業に落とし込めば可能になるからです。2004年に初期バージョンを開発してから現在に至るまで、15年近くになりますが、絶えず改良を続けています。

またさまざまなコンテンツへの対応も進めています。データは処理する対象と不可分ですので、新たなコンテンツを処理しようとする、新たな作業が生じます。初めのころ、我々のシステムは表組みを処理できませんでした。後に対応しました。以前、固有名詞や書名の傍線は処理していませんでしたが、『中華経典古籍庫』の書籍には傍線があります。中華書局から大量の委託を受けたときに、そうした機能を開発しました。一般のOCRソフトでは、普通の書籍はうまく処理できず、処理できないコンテンツに出くわすと、滅茶苦茶になってしまいます。処理するコンテンツが増えると、その工程に対応したツールを開発するという方針を採っています。このように我々のシステムは、顧客のニーズに基づき発展してきました。

中華書局は現在、国内最大の古籍出版社ですが、古籍については全てデジタル化を完了しています。現在、彼らが提携した多くの出版社が提供した古籍の入力を委託されていて、完成後、『中華経典古籍庫』に登録して発売していきます。それは、古籍のデジタル化において、最大規模ではありませんが、データの品質は最も高いといえます。

そのほかの大きな進展としては、2011年に、当時の新聞出版総署が、3つの大規模文化プロジェクトを、第十次五ヶ年計画の文化発展綱要に盛り込みました。その一つが中華字庫プロジェクトで、これは文化界にとっても極めて重要なものでしょう。全体的な構想は、中国の文化を起源、つまり最初期の甲骨文から、金文、秦簡・漢簡などの簡帛文字、篆書、さらには宋・元以前の全ての文献、明・清の大部分の文献、それから現代の、中華民国と現代の代表的な文献を、全てデジタル化する

というものです。我々の通常の本のデジタル化では、文献の内容、つまり文献に何が書いてあるのかを重視します。いくつかの稀にしか使われない漢字は、文献の内容にはさほど影響しません。だから技術的には、『四庫全書』や『四部叢刊』は大量の外字で処理していますし、そのほかの書籍でも大量の外字を使ったり、どの文字とどの文字を組み合わせる、というように表示しているものもあります。

中華字庫プロジェクトでは、そうした文字コードから欠落した文字、さらには文字の差異に注目しています。文字というのは、細かく見ると大きな差異があります。簡単な例を挙げると、「峰」の字は、現在通用しているのは左右構造ですが、上下構造の「峯」が使われることもあり、漢字の形式は複雑です。そこで、あらゆる字の形式について、どの程度の差異ならば同じ字と認定できるのか、厳密な規範を作ろうというのです。その規範に基づいて、あらゆる古籍を整理すれば、ある文字の源流を、甲骨文字から金文、簡帛、小篆、木版本そして現在に至るまで、整理することができます。

以前、読んだのですが、現在ユニコードを管理しているのはユニコードコンソーシアムですが、実際には各国のコンピュータで使われている文字コードを総合したものです。そうした作業を行っているのは、大部分がコンピュータの専門家、彼らはそれが何という文字なのか、どう読んで、意味はどこに起源するのかといったことには、あまり興味がないそうです。人が通常読めるのは数千字程度ですが、2万字、3万字ともなると、大部分の漢字はわかりませんので、コンピュータの専門家が、これは一つの漢字だと、我々に提供していることになります。ある漢字は、出典がどこなのかが誰にも分からなくなっており、後に、日本のJISコードで、国勢調査に基づいていたことが分かったそうです。

書き間違いから生まれた漢字も多くあります

が、そうした字は由来がわからなかったり、間違っていたりします。あらゆる文献を整理して、ある字が文献の流伝の過程で、どこに出現したのか、どのような状況でなぜ書き間違えられたのか、またその字が歴史的にかつて意味を持っていたといえるのか、このプロジェクトが完成すれば解明できるでしょう。ユニコードのどの漢字に使い道がないのかを確定することもできます。文字と文字との間の具体的な関係や、確かに存在したのに文字コードに入っていない文字などもわかります。このような文字を補充して比較的完全な枠組みが形成されれば、将来、出版や情報交換の際に、作字する必要もなくなるでしょう。現在、少しマイナーな本では、一冊ごとに作字しなくてはなりませんし、作字した漢字には汎用性がありません。このように、中華字庫プロジェクトの意義は非常に大きいと思います。完成の暁には、漢字文化圏全体にかなりの影響を及ぼすのではないのでしょうか。

このプロジェクトで、我々は入力プラットフォームを担当し、他の落札した機関が入力作業するためのツールを提供しました。また、2つのサブプロジェクトの具体的な入力を担当しました。これは私たちのここ数年の主要な仕事の一つです。

—中華フォントプロジェクトは、何年に終わるのですか？

王：元の計画では2016年に完成するはずでしたが、プロジェクトの途中で問題が発生したため、来年の中頃に完成するはずでした。

◇ 社名の由来と会社の概要

—御社の創新力博という名称には、どのようないわれがあるのですか？

王：2003年ころ、アメリカの経営管理学者のドラッカーの本をよく読んでいましたが、企業の生命は「^{イノベーション}創新」にかかっていると感じました。でも実際にはイノベーションを起こすのはとても難しく、せいぜい小さなことしか

できません。「力博」は、当時、データベース、漢字文化のデータベースを作りたいと思っていたからです。でも結局、この方面のツールとサービスを提供していて、自分でデータベースを作れてはいません。

—英語の社名「LIBO」は、どういう意味なのでしょう？

王：「L」は、2003年当時、インターネットが流行しており、「e」と同じように「I」が流行していました。「力博」はデータベースを作りたいということで「Library」から取りました。反省を込めていうと、データベースを作ることができなかったのは、エンジニアの限界ですね。当時は技術が非常に重要だと思っていたのですが、しかし実際には、企業の発展の観点から技術について考えないといけません、エンジニアの悪いくせで、技術に力を入れすぎるんですね。私はエンジニア出身なので、今でも相変わらず、いつもコードを書いています。当時は、30を過ぎててもコードが書けるのか心配していたものですが、今、40を超えてもまだ書き続けています。

—御社には何人のエンジニアがいますか？

王：5～6人です。最も重要なコードは私が書いていて、補助的な作業や試験などを担当しています。

—ソフト開発に使っているプログラミング言語は何ですか？

王：主に使っているのはC++です。C++ならばユーザーがコントロールする余地が大きく、またある部分の性能が比較的低くても、探し出して自分で交換できます。一部のハイエンドのプログラミング言語では、手出しできる部分が比較的少ないので、いまのところあいかわらずC++を主に使っていますが、人工知能ではPythonを使ってみるかもしれませんですね。

—入力作業を行う従業員は何人ですか？

王：80～90人くらいです。

—みな洛陽支社の方ですか？

王：主に洛陽です。北京ではソフトのテストと、

顧客から書籍を受け取って初歩的な整理を行ったり、最終的にデータを整理したり、初めと終わりの作業をやっています。主なデータ加工業務は洛陽の方です。

—従業員は女性が多いですか、それとも男性が多いですか？

王：女性が多いです。実のところ、我々は従業員の性別について特に要求していないのですが、現実問題として、どうしても社内ですっと座って仕事しなくてはならないので、比較的落ち着いた性格が必要になります。男性は、いろいろと考えたり束縛をきらったりで、数日で我慢できなくなって辞めてしまうことが多く、女性の方が世間から自分自身のポジションを獲得するように求められており、でかいことを成し遂げなくてはというようなことは少ないですね。このため、基本的には9割以上が女性です。

—従業員は高卒ですか、それとも大卒ですか？

王：今ではほとんどが大卒です。というのは、中国で大学の定員拡大が始まってから、都市部では7～8割の人が大学に行くようになり、大卒でない人の方が少なくなりました。

—ここ数年、1年でデジタル化する文献は何文字くらいになりますか？

王：20～30億字ですね。

—1年でですか？

王：1年ではなくて合計ですね。1年では数億字になります。現在、比較的大きな取引先は中華書局ですが、毎年、データベースに3億字程度を増やしています。そのほかにも、文書処理のスピードアップのサポートや、編集作業などもありますが、相当程度を占めるのは図書館からの委託ですね。それと中華字庫プロジェクトや、その他の機関からの委託があり、毎年およそ数億字になります。現在、中国政府は文化関係に相当な資金を投下していて、多くの大学や出版社は十分な資金をもっているんですよ。

—洛陽支社は、何年に開設したのですか？

王：2013年のはずです。そのころ、中華書局が

経典古籍庫の構築を始めました。2005年頃から中華書局の仕事は何度か受けていましたが、当初、彼らはデジタル化について、紙版の販売への影響などを心配していました。多くのデータベースを作りましたが、いずれも内部システム向けのもので、おそらく書籍の再版に使っていたのだと思います。2013年からシステムを発売する準備を始め、我々が落札しましたが、北京では大量の委託をこなすことができなかったので、洛陽に生産センターを作ったのです。

—洛陽では毎年、何文字くらいの生産能力がありますか？

王：たぶん6～7億字は問題ありません。いまのところ、仕事が立て込んでいるわけではないのですが、それでも4～5億字になります。洛陽のセンターを拡充するのは比較的簡単で、既にあちらで5年間やっていますので、従業員の離職率が高いですが、それでもある程度熟練した従業員が留まっています。留まっているのはみな、座っているのが好きで歩き回りがたがらないような人ですね。そうした人が、中核的な従業員になっています。比較的流動性が高いので人員の拡充も簡単で、プロジェクトが多ければ人を多めに募集し、少ないときには募集をかけなければ従業員が自然と減っていきます。解雇しなくても、人を補充しなければ次第に減っていくのです。

—洛陽は北京に比べてコストはどのくらい低くなりますか？

王：まず、不動産が安いですね。北京ではこの狭い部屋でも毎月1万元近くかかります。従業員にとっても、地元ならば住むところを確保しなくて済みます。我々のデジタル化経費では、従業員の人件費がコストの大半を占めています。地方であれば、北京の7割程度で済みます。利益率がさほど高くなく、北京では赤字になってしまうプロジェクトでも、地方ならば10%の利潤が得られます。

—洛陽でここと同じくらいの広さの物件は、月幾らくらいになりますか？

王：ここはマンションですが、洛陽で借りているのはオフィスビルで、条件がずっと良いです。価格は1/3か1/4ですね。行ってみればわかりますが、1,000平米くらいでとても広いです。ここは100平米もありません。だから、ここ数年、労働集約型の企業は、大量に北京・上海・広州・深圳から離れています。ここでは生き残れないんですよ。一部の不動産業者などは家賃を高騰させて、下層の人を北京から追い出そう、そんな考えをもっていました。エグゼクティブであっても、月給2～3万元では、北京で暮らしていくのは相当に困難です。月給3万だと手取りは2万前後、それで北京で部屋を借りて家族を養うと、かつかつですね。だから、下層民を追い出して、エグゼクティブも引き留められなくなってしまいます。ただ、エグゼクティブにとって北京はチャンスが多いのですが。

◆ 主要な取引先

—中華書局の他に、どのような企業や機関と取引していますか？

王：南京大学図書館出版社、人民文学出版社、それと語文出版社ですね。出版社が多いです。フォントについては、北京師範大学と提携しています。

—出版社は自前の文献データベースを開発しようとしているのですか？

王：現在はそう考えていますね。というのは中華書局の『経典古籍庫』が非常に好評を博していて、とりわけ国家指導者がこれは非常に意義があると発言したインパクトは凄かったです（訳注：2015年4月23日に劉雲山・劉延東が中華書局を訪問し、『中華経典古籍庫』を称賛した。）。それに倣って、多くの出版社がデジタル化を始めています。『中華経典古籍庫』のWeChat版は、比較的便利で使いやすいですが、多くの大学や研究機関はやはりミラーサイト版を購入していますね。ミラーサイト版のソフトも我々が開発したもので、機能が豊富でイン

ターフェイスも綺麗です。簡単に使えるのはWeChat版ですね。スマホで使えますし、価格も手ごろですから。

—我々以外に、日本の大学や研究機関からの委託はありますか？

王：ほとんどありませんね。千田さんの委託くらいです。日本の場合は基本的に教員個人の委託で、量も少ないです。逆に中国国内では、教員個人の名義はほとんどなく、みな機関です。日本ではここ数年、この方面の経費も少なく、あまり重視されていないようですね。現在、中国の大学は、経費が多すぎて使い切れないんですよ。大学の教員のみなさんは、経費をどう使えばいいんだ、会計検査が厳しくて、うまい使い道がみつからない、と。

◆ データの精度・文字コード

—文献を入力する際の精度は、何によって決まるのでしょうか？

王：それについては顧客のニーズによります。例えば、顧客が錯誤率一万分の一の精度を必要とするのか、一万分の三でよいのか、それによって仕事の工程を調整します。中華書局の作業では、彼らの要求は一万分の一でしたが、実際にはそれよりも遥かに低く、彼らが数千万字分チェックした結果、数文字しか間違っていないでして。版本だと恐らくもう少し悪くなるでしょう。実際には、この基準を決めてから、顧客の求めるデータ形式と錯誤率に基づき、工程を調整しています。例えば、顧客の錯誤率の要求が高かったならば、まず一度対校し、その後、改めて対校した、A・B二つの結果を合わせます。さらに要求が高ければ、三度の結果を合わせることもあります。そうすると、一人が見落としていても別の人が修正した箇所がわかります。見つかった1～2%の誤字については、後でまた間違えているかもしれないので、さらにチェックすることもあります。また、システム上の誤りもあります。たとえば、「人」と「入」は

似ていますが、他の漢字の部品になっていることもあり、「陝」の繁体字は「陝」で、「陝」は別の字ですが字形は似ています。従業員に違いが分からず、3人とも見分けられなかったら、それがシステム上の誤りになります。そうした字については、訓練で対応しますし、また数はさほど多くないので、それだけを抜き出して解決しています。このようにして、最終的なデータは、顧客の要求以上のものになります。一万分の一を求められた場合は、十万分の一、百万分の一になるかもしれません。前半の90%の仕事は簡単で、最後の1%の仕事が最も難しいものです。ポイントは、字の入力ではなく、誤字を探し出すことにあるのです。

データベースを公開したばかりのあるプロジェクトでは、正確率は99.9%でよく、ユーザーが使っていく過程で誤りを見つけて修正していくという方針をとることで、コストを削減しています。一方、我々が渡したデータで直接本を刷ることもあります。その場合は価格が多少高くとも高い品質を求められます。これまでに出版社の仕事を受けてきましたが、彼らのやり方は伝統的な3回の校正を行うスタイルによって、一万分の一を実現するとしていましたが、全く基準に達してないことが分かりました。編集者は、読むときに意味を読んでしまうんですね。そうすると、たくさん字について細かい字形までチェックしなくなってしまう。人の目は細かいところを見逃しても、概ね読めてしまうみたいで、誤字を見つけれないのです。ある実験によると、文の文字の順番を入れ替えても、人は読めてしまうそうです。だから大量の校正をするときに、編集者が一字一字真面目に見ていかないと、校正漏れが生じてしまいます。我々の校正では、作業員は読めない字であっても、他の字との違いはわかりますので、逆に微細な違いを見つけれられるのです。面白いのは、中華字庫プロジェクトの過程では、一般の字形と異なる字を選び出すと、そ

れが新たな漢字であることでした。あるとき、作業員が私に誤字を示しました。しばらく見ても一般の字形との違いが分かりませんでしたが、横棒が左払いになっていると言うのです。彼女はその漢字を読めませんでした。ルールに則って字と字の違いを判断することはできます。

我々が見たところ、校正が行き届いた原稿でも、それでも一万分の二から三の誤字があります。このため、我々と出版社との提携は上手くいっています。30年前、有名大学の卒業生が出版社の編集者になるのは、相当良い仕事でした。仕事の傍ら、1冊の本を校訂すると、原稿料は十数元にしかなかったかも知れませんが、それでも少なくない額でしたし、意義もありました。今では、編集者の校訂の原稿料は安くて、真面目な良い編集者が少なくなりました。これは大きな問題です。——中国で御社のデジタル化システムを購入した大学や機関はどのくらいありますか？

王：十幾つかです。大学では、南京の金陵科技学院、広西大学、長江学院などです。

——彼らはシステムを買って、何を作っているのですか？

王：大学の場合は、研究のために買っています。本をデジタル化しているところも無くはないですが、大半は研究室でデジタル化を研究することが目的です。実際に入力に使っているのは愛如生のような会社ですね。中華字庫プロジェクトでは、北京師範大学・漢王・方正なども業務を請け負ったために、我々のソフトを使っています。

——独自のデータベース製品を開発する予定はありますか？

王：いまのところありません。というのは、我々がデータベースを開発したら、顧客に自分たちの製品と競合するのではないかという懸念を与える恐れがあります。また、データベースを開発しようにも、現在はテーマの選択が難しいです。『四庫全書』『四部叢刊』や、あるいは地方志、その他、比較的重要なテー

マについては、既にやられています。しかも現在、全面的なものとして愛如生の『基本古籍庫』があり、その後、『中華經典古籍庫』のような出版社のものがあり、さらに中華字庫で作られた大量の文献がこれからネットで公開されるかもしれません。また底本の入手もネックになります。図書館は底本を提供しながらないのです。今のところ、すぐにデータベースを作るつもりはありません。

—あなた方のシステムは、現在、ユニコードのどのバージョンをサポートしていますか？

王：最近、10.0まで対応しました。最新は12.0ですが、まだあまり使われていませんね。文字コードの枠組みには対応済みで、どの区域にどの文字を増やすかということに過ぎませんので、12.0のサポートに問題はありませぬ。新たに増えた字について、対応するフォ

ントと入力方法さえあれば大丈夫です。ただ、10.0で顧客のニーズはほぼ満たせていると思います。

—ルビに対応することはできませんか？

王：入力の際に、ということですよ？受託したプロジェクトで必要があり、ソフトウェアを開発すべきだと判断したらシステムで対応しますが、データが完成し、なんらかのファイル形式で表現しても、今度は検索・閲覧システムがサポートしなくてはなりませんので、大変なコストがかかります。相当大規模でない限り、人に頼んで入力して貰った方がよいでしょう。また日本語のルビの場合、ひらがなとカタカナの字形が似ていて、日本語の十分な知識がない我々の従業員では錯誤率が高くなってしまいます。

北京創新力博社 洛陽入力センター

2018.8.21 AM9:00 於河南省洛陽市洛竜区

◇ 洛陽入力センターの概要

—洛陽の入力センターの責任者はどなたですか？

劉智（以下、劉）：こちらの王曉麗支社長です。洛陽ではデータ加工を行っています。

—どのくらいの職員がいますか？

劉：70人余りがいます。

—勤務時間は何時から何時ですか？

劉：8時半出勤です。通常、冬は5時半、夏は6時退勤で、休憩時間もあります。

—幾つのフロアを借りているのですか？

劉：2フロアです。この階には4部屋あり、上の階には大きな部屋があります。ここは将来、さらに規模を拡大することになっても、空き部屋が多いので対応しやすいでしょう。

—人員はどのように配置されているのですか？

劉：我々の作業プロセスは細かく分かれており、2つのフロアのいくつかの部屋に分かれて作業しています。簡単に説明すると、一冊の本を入力するとき、初めにスキャンして、紙の本を画像に取り込み、それから画像の係員がレタッチ処理します。その上で、デジタル化

システムに取り込みます。まずコンピュータが自動的にレイアウトを分析しますが、若干の間違いがあるので、それを人が修正します。レイアウト分析も、我々は細かく段階を分け、多くの人が関与するようにして、ミスを潰し品質を高めています。その次に、本の中の文字を一文字一文字小さな画像に分割して、校正を行います。校正もいくつかの段階に細かく分かれていて、錯誤率が1万分の1以下になるようにしています。

プロセスごとにそれぞれ別の人員が配置されており、比率からいうと、画像が少なめ、レイアウトが多め、校正は正確性を保障する重要なプロセスですので最も多くなります。

◇ デジタル化システムについて

—あなた方のシステムは、以前の書同文のものとどのような違いがあるのですか？

劉：似ていますが、我々の方が先進的です。我々のシステムは、ユニコードの大規模文字集合の全ての漢字を認識できます。また、表組や



挿絵なども処理できますし、XMLにも対応しています。基本的な理論には似ているところがありますが、我々は長年、システムの改善を重ねてきています。作業工程も細かく分かれており、全部で16～17あります。

—中国語・日本語・英語以外に、どのような言語に対応できますか？

劉：我々が主にやっているのは中国語で繁体字を得意としています。日本語・英語はごくわずかで、さほど専門的とはいえません。日本語は識別できることはできますが、ソフトウェア開発の重点は中国語に置いています。英語は、実のところ認識が簡単ですし、英語混じりの文章は認識できます。もしも英語や日本語を専門にやるとしたら、ソフトを特化させる必要があるでしょうし、フォントも必要になります。

—チベット語などはどうですか？

劉：難しいです。以前、満州語の依頼はあったのですが、識別する方法がみあたらなかったですし、そもそも読める人もいませんでした。当時探したところでは、東北のいくつかの大学で専門に研究・教育しているとのことでしたが。

甲骨文字や少数民族言語は、文字認識技術上、単独のシステムが必要ですし、また文献も多くないので、力を入れていません。甲骨や金文などは中華字庫プロジェクトで扱っていますが、彼らはそれらを標準漢字ではなく非標準漢字に分類しています。そしてまず、その文字を囲うようにフレームを描きます。それから筆画に従ってなぞって書けば、認識できるようになります。線が切れていても、自動で接続して認識できます。それから字形を描いて、フォントを作ります。ただ、コンピュータによる識別の効率は、今のところあまり良くありません。

コンピュータは標準的であればあるほどうまく識別できるので、手書き文字は木版本ほどうまく読めません。というのは、字を書いている人が違くと、書風も違ってきますから。

以前、清史関連の仕事を請け負ったとき、起居注のようなものがあった、多くの大臣の手書きなんですね。標準的とはいええないもので、認識に手間がかかりました。結局、一部分を入力した後で辞書を作って、作業をして学習させて、なんとかしました。

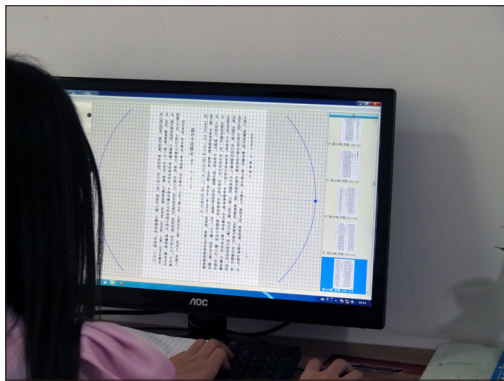
—版本ごとに書体や書風が異なりますが、この問題をどうやって解決しているのですか？

劉：我々のシステムには識別用の辞書機能があり、再学習できるようになっています。システムにかけると、繁体字の大部分は認識できます。手書きの本であれば、まず10～50ページほどやってみます。完成した文字を修正した後で、小型の辞書を作って学習させ、残りの部分を作業します。こうすることで、レイアウトや文字フレームの認識率、そして認識の正確性を高めることができます。実のところ、このデジタル化技術は枯れた技術で、この業界ではどこでも大差ありません。結局、フォントや字書の品質が、自動化レベルの高さ、ひいては認識率の高さとコストとの低さを左右するのです。

◇ 受託業務の内容

—近頃、台湾や香港の企業の仕事は受けていますか？

劉：以前はありました。最近は、国内の仕事が増えています。というのは、政府が古籍を重視していて、毎年、数億字もの発注があるからです。国内のいくつかの大学の委託を受けて、法律古籍・民国古籍・民国期刊・民国時代の広告などといった、いろいろな専門のデータベースを作っています。民国の広告には日本語の広告がとても多いですね。これらは、以前よりも細かくなった印象があります。以前は大規模な古典籍のデータベースを手がけていましたが、今は大学や機関が自分で整理した、単独の小さなデータベースが多くなっており、次第に研究に特化した方向に発展しているように感じます。それと『中華經典古籍



レイアウト分割作業

庫』ですね、近頃の大きな仕事は。また国家図書館をはじめ、図書館の仕事もあります。図書館は毎年の経費が固定なので、量は多くないですが、毎年あります。そのようなニーズに応えるために、我々の従業員も増えています。

—研究者や研究機関はどのようなデータベースを作っているのですか？

劉：地方文献が多いですね。

—研究者の興味によって作られるケースもありますか？

劉：研究者単独は少ないですね。一人ではそんなにたくさんの経費を申請できませんから。古籍に投入する経費が潤沢な、学校なり図書館なり出版社なりが大半です。デジタル化して、刊行したいんですね。民国以前の、刷りが悪いものや木版本の数が多ければ、それだけ予算を投下します。みなさん、事前に学校なり市なり国なりに経費を申請して、予算を獲得

してから、やるんですね。以前、江蘇の先生の仕事を受けたことがあります。個人の興味で、1～2冊だけでしたし、デジタル化した後は、本の内容を分析して論文を書き、科研の成果としていました。今では、機関が作るものは数千万字だと小規模で、基本的にみな一億字以上の規模になっています。

—そうしたデータベースにはどのような特徴がありますか？

劉：量が多いですから、基本的に細かく選び出したりせず、年代やジャンルを大ざっぱに区切って、その範囲をまるごとやるのが多いです。例えば、図書館蔵書の、南京周辺のものとか、民国のものとかいう具合です。図書館のやり方は研究者とは違いますね。以前、山東大学の仕事を受けたのですが、彼らは『易経』を重点的に研究していて、『周易』関係の本を専門的に細かく選んでいました。異なる版本と版本を比較して、時代による変遷や発展を研究するんですね。それに対して機関が作るものは、大ざっぱに言うと、規模の大きさを追求します。図書館も書籍のデータベースを購入するのに、5億字とか10億字とか言われると、蔵書がすごく増えたように感じるでしょう。

それと、近頃、一定量を占めているのが、出版社に委託された作業で、いうなれば、出版社が編集者の仕事を我々にアウトソーシングしているんですね。手で入力するよりも効率がよいので、出版社が古籍を出版するときに、写真で再版するのではなく、我々が本文と構造をデータ化し、それを彼らが組版するのです。出版社も、正規の編集者がとても少ないようで、それで我々に委託しているのです。

◇ デジタル化作業見学

劉：こちらは研修室で、新規採用の従業員はまず一ヶ月、ここで研修を受けて、それから正式に配属になります。

この部屋で画像をスキャンします。本が届け

られたら、まずここでスキャンします。現在はスキャナを2台設置していますが、量が多い場合はもっと増やします。

—どのようなスキャナをつかっていますか？

劉：この種の（エッジの細いフラットベッド）スキャナです。一般のフラットベッドスキャナでは、本が壊れやすいので、普通はこれを使っています。スキャンは、通常、2人が作業します。次に3人が画像をレタッチ処理し、処理が終わった画像はネットであちらの部屋に送られます。

<隣室に移動>

次にこちらの部屋でレイアウト分割の作業をしています。

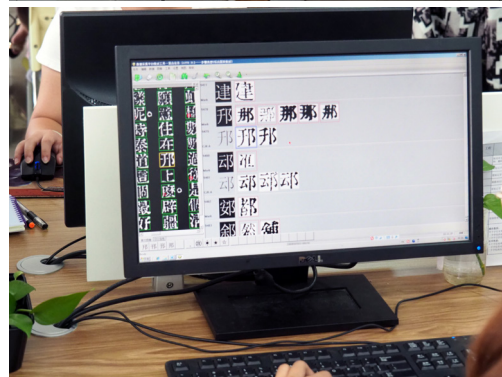
こちらはレイアウト分割の一つのプロセスで、一文字一文字の画像に分割しています。一つのフレームが一つの字を表します。

次のプロセスは、校正ですが、上の階で人員も多いです。

<上の階に移動>

ここでは同じ文字に認識された箇所を抽出して、まとめて校正しています。こうすると、誤字を見つけやすくなります。修正が完了したら、完成データを出力します。

大まかな流れはこのようになりますが、それぞれのプロセスがさらにいくつかの段階に分かれます。



校正作業

CNKI 北京本社

2018.8.23 AM9:30 於北京市海澱區東昇科技園

◇ CNKI の概要と沿革

CNKI 社の正式名称は、同方知網デジタル出版グループで、清華大学傘下の持ち株会社である清華ホールディングスに属する。同社はまた、中国学術雑誌（CD版）電子雑誌社、同方知網デジタル出版技術有限公司、同方知網（中国）技術有限公司の三つの子会社を持つ。

CNKI（中国知識基礎施設工程）の構築は1996年に始まる。同年、まず中国学術雑誌（CD版）が創刊され、1998年には中国学術雑誌（CD版）電子雑誌社が設立された。そして翌1999年に、中国知識基礎施設工程すなわちCNKIが正式にスタートしている。中国の国家重点情報化建設プロジェクトとして、国内外において高く評価されている。

現在、社員の総数はおよそ3,000人、そのうち北京本社に数百人がある。また、山西省太原にデジタル化基地を作っており、中国各地の31の支社がある。



CNKI 本社

◇ 主要なデータベース製品

CNKIの核心的プロジェクトは、中国知識資源総庫であり、雑誌論文、学位論文のみならず、新聞、会議論文、年鑑、工具書、百科事典、特許、標準、科学技術成果、法律・法規などを収録している。各種データベース製品は20種を超えている。以下に主要なものを挙げる。

- 中国雑誌全文データベース（CJFD）
中国大陸の雑誌の99%を網羅。
- 中国博士／修士学位論文全文データベース（CDMD）
中国全国の416の博士学位授与機関の優秀な学位論文、約331万本を収録。学位授与後45日以内にデジタル出版。
- 中国学術コレクション電子ブックデータベース（CNKI e-Books）
民国時期から現代までの書籍、4,100冊を収録。閲覧および全文検索に対応。



CNKI 本社ロビーにて

- 中国工具書ネット出版総庫 (CRWO)
本誌でも取り上げている CNKI 工具書庫。
- 国際会議論文全文データベース (IPFD)
27,000 以上の中国国内の学会・国際学会発表論文、約 230 万本を収録。学会終了後 1 ヶ月以内に登録。
- 中国重要新聞全文データベース (CCND)
2000 年以降の中国国内発行 633 紙を収録。
- 中国年鑑ネット出版総庫 (CYFD)
1949 年以降に出版された中央・地方・業界・企業の年鑑をおよそ 4,000 種、34,000 冊以上収録する、全文データベース。
- 法律知識資源総庫 (CLKD)
法律・法規・条例・判例・論文など、法学関連の総合データベース。
- 中国引用文献データベース (Chinese Citation Database)
引用文献の検索・分析、文献の出力、また論文の著者・出版社・掲載誌などの分析。

◇ ソフトウェア製品

ソフトウェア製品として、「学術不正文献検索システム」、「CNKI デジタル学習プラットフォーム」、世界各国のデジタル学術リソースを一括検索できる「CNKI 学術検索」、出版社のオンライン出版ソリューションである「CNKI 電子ブックネット出版プラットフォーム」その他、情報管理、デジタル出版関係ソフトがある。

このほか最近、個人向けの研究情報管理ツールである「CNKI 研学平^{プラットフォーム}台」(<http://x.cnki.net/>) をリリースしており、ネット上で研究テーマ関連論文の収集・管理などを行うことのできる、いかなればジャーナルの閲覧と引用管理ツールが融合したような機能を持っている。

◇ 近年の動向

—CNKI では近年、新しい動きがありますか？

答：これまで、CNKI は主に文献処理、すなわち中国のあらゆる学術資料、雑誌や学論文のデ



CNKI 研学プラットフォーム

ジタル化を行ってきましたが、ここ 2 年ほどいささか方針を変えています。

例えば、ソフトウェアについては、デジタル加工サービスを行うようになっています。現在、中国では多くの図書館が、我々 CNKI のサービスを使っています。例えば、データベース構築ソフト、デジタル加工などで、現代中国語であれ、古い文字や歴史的文献の処理であれ、大学図書館・公共図書館を問わず、非常に多く使われています。

以前は主に中国で展開しており、距離の問題から、海外でのテスト拠点は少なかったのですが、この二年ほど、香港・マカオ・台湾などでの売り込みを始め、昨年、『マカオ日報』のプロジェクトを完成させました。『マカオ日報』の大量の古い新聞を加工して、データベース化したものです。

よりよいサービスを追求して、我々は二台のスキナーを購入してマカオ日報の作業場に持ち込み、会社の専門スタッフを送り込みました。現場泊まり込みでデジタル化を行い、しっかりと前宣伝を行い、プロジェクトを完成させました。

◇ 論文引用数の統計について

—CNKI では論文の引用数の統計を取っていますが、これは学者や研究機関の評価に使われていますか？

答：CNKI では中国語の大部分の学術リソースを入力しており、データに対して統計を取って



CNKI スタッフのプレゼンテーション

います。CNKIでは専門の文献計量センターを作って、学術評価と文献の計量を行っています。また、この方面に関する学術研究も行っており、中国の著名な学術雑誌に研究成果を発表しています。我々はビッグデータの統計に基づいて、例えば、どの学科、どの大学の研究の成果率が高いか、あるいはどの教授の評価が指標を超えているか、といったことを研究しています。

—そうしたデータは、誰が利用するのですか？
例えば、大学が教授の研究成果を評価するときや、国が大学を評価するときにも使われるのですか？

答：はい、いろいろなところで応用できます。例えば多くの大学に、科学研究処があって、自分の学校の教員の学術業績について注目しています。だから、我々の研究が完成したならば、各校の教員の研究業績がどうであるのか知りたいというニーズに応じて、製品あるいはレポートを提供できるでしょう。例えば、成果の発表状況と被引用数などのデータをレポートしたり、ここ数年の趨勢を示すこともできます。この引用ですが、我々の中国の雑誌における引用のほか、アメリカの『サイエンス』のような国際的な雑誌での引用もあります。海外における引用全体についても計量しており、我々が開発した指標もあります。たとえば、CI、クラウド・インデックスという指標は、ある研究者の被引用数とインパク

トファクターを同時に考慮して、影響力の角度から計測しています。

大学のほか、病院でも同様の指標があります。病院も、病院間の比較のほか、病院内の比較もあって、報告書を出さなくてはならないし、彼らの職務評定にも影響があります。どの業界でもそれぞれこのようなパフォーマンス評価をしています。

そのほか、学術雑誌や雑誌社も、自らの国際的な影響力に注目しています。CNKIでは毎年11月・12月頃に非常に大きな規模の「期刊未来大会」を開催しています。毎年、中国の90%以上の学術雑誌が会議に参加します。会議の規模も非常に大きく、2,000～3,000人もの規模で、全員が一緒に中国の学術雑誌の発展について討論します。会議で我々は毎年、年間報告を発表して、学術雑誌の国際的な影響力について話します。中国の雑誌は毎年、学術的評価を経たレポートを発表しますが、我々のデータを分析したものです。会議では毎年、学術雑誌の年間トップ5とトップ10を選出していますが、このような学術雑誌の表彰は、彼らにとってとても重要な認証になっています。また現在、中国ではたくさんの書籍が出版されていますが、それらの書籍、特に学術的な書籍の影響力がどうであるかについても、我々は統計を取っています。統計の方法は、主に中国のあらゆる学術文献を分析し、どの書籍を引用しているのかを見て、評価の指標に基づいて統計を取っており、中国の学術書籍全般についても評価の体系を作っています。これは今年ようやく研究が完成した成果です。

—今お話にあったインパクトファクターは、どのように算出していますか？

答：インパクトファクターは、ICIの公式に基づいて計算していますが、CNKIでは中国のコンテンツと国際的なアメリカ『サイエンス』の一部のコンテンツをまとめて扱っていますので、ICIと完全に同じではありません。

—被引用数をカウントする場合、プラスの評価

のときもあれば、マイナスの批判的な評価のときもありますが、それをどのようにカウントしていますか？

答：詳しいことは評価センターの所属ではないので分かりませんが、プラス評価、マイナス評価や、自分の論文ばかりを引用するといった点など、非常に細かくやっているはずですよ。この種のことを識別する方法を彼らは持っていますが、私はおおまかな状況しか知りません。

◇ 論文収録と契約

—あなた方は論文の収録にあたって研究者一人ひとりと契約しているのですか？

答：雑誌の場合、我々は雑誌社と契約しています。それと、タイプによって提携先の性質も変わるんですね。例えば、書籍の場合は出版社とお話するかもしれませんが、雑誌であれば雑誌社などとお話します。そして契約を交わしますが、契約も状況に応じて、1年だったり3年・5年だったり、さまざまです。そして、論文の場合は、おそらく著者が大学と契約を結ぶはずで、それでCNKIに収録することに同意したことになります。その後、規格に従って入力します。

—論文を収録するとき、金銭的な保障はしていますか？

答：雑誌論文を収録する場合は、雑誌社などと相談します。雑誌社によっては、著作権使用料を求めてきます。著作権使用料は、年々値上がりしており、CNKIのコストも次第に高くなっています。だから、著作権使用料を支払うほか、イベントなどの形式も採っています。書籍の場合は、我々が売ってしまうと、収入を分配しなくてはならなくなるでしょう。CNKIの収録文献は、著作権についてはクリアで、弊社では遵守しなくてはならない原則になっています。

◇ システムとサーバ

—同じ作者が同じ雑誌の複数の号に論文を発表しているのに、検索できるものとできないものがあり、また、CNKIのスマホアプリでは検索できた、といったことがありました。これはなぜでしょうか？

答：確認してみますが、入力ミスではなさそうなので、サーバの問題かもしれません。我々は2つのサイトを持っていますが、センターサイトは北京のサーバにあり、それと我々が通常使っている海外向けサービス専用サーバは青島にあります。両方のサーバの内容は同じはずですが、検索でちょっとした問題が出ているのかも知れません。特にデータを更新するときなどに問題が出るがありますが、データ自体はアップされていますので、簡単に修正できます。今年の年末頃に、青島サーバをリニューアルする予定です。リニューアル後は、北京センターと同じになり、機能が改善し、リソース安定性も向上するでしょう。

—oversea.cnki.net というドメイン名称が、青島サーバになるのですか？

答：そうです。海外のユーザーは同じURLで最新版に移行する予定です。また、スマホアプリ「手機知網」ですが、現在、新しいエクスプレス版、グローバル学術モバイル版（全球学术快报）が出ており、いち早く研究成果を閲覧できるようになっています。みなさん、スマホを通じて情報を収集することが多くなっているので、中国だけでなく海外のユーザーも、モバイル版を通じて情報が得られるようにしています。まだ、海外ではあまり宣伝していないのですが。

—それは、理系だけですか？

答：全分野です。

—論文によってはCAJ形式でしか読めないものもありますが、PDFに統一しないのですか？

答：中国国内では、多彩な機能を持つCAJビュー

ワの方が良い、という人が多いのです。また新しい論文はHTML形式で、オンラインで閲覧できます。

◆ 論文の収録基準と不正行為対応

—CNKIでは論文の収録に基準を設けているのですか？

答：文献の収録基準は、製品によって少々異なります。雑誌の場合、ISSNコードを持っていれば、中国の正式出版された雑誌、海外版のものなどをできる限り収録しています。ただ、一部機関の内部刊行物については一部分を収録していますが、全部収録することはできません。また、博士・修士論文の場合、全ては収録しません。というのは、大学によってレベルの差が大きく、酷い論文は収録しないからです。ですから一般に、優秀な論文、比較的優秀な学校や学位授与機構、あるいは学科のものだけを収録します。だから中国の多くの学生にとって、CNKIに収録されることは一種の榮譽になっています。その他のデータなどについては、直接、データを公表している政府機関と提携しており、データの信頼性を保証しています。

—論文が優秀であるかどうかは、どうやって決めているのですか？

答：一般に、我々は論文を個々に判断するのではなく、学校全体の影響力や学科のランキング、

それと学科の中国における信頼性などを考慮します。また、CNKIでは学術不正について、非常に強力な検査を行っています。あらゆる大学が卒業論文にCNKIのチェックソフトを使って、論文が剽窃をしていないかを確認して、少なくともこの点については学術成果の真実性を保障しています。今では、30%以上の内容が重複していたら、剽窃だとして不合格になります。チェックソフトでは、剽窃した内容の比率、あるいは引用文献、引用箇所などを色分け表示します。そして、その論文が何パーセント他と共通しているのかを、教えてくれます。

—十数年前、CNKIで「～省の××産業の展望」というタイトルの「～省」の部分を変えて検索すると、データが若干違うだけで内容がほぼ同じ論文がたくさん見つかったと、中国のネット上で話題になっていました。

答：今では厳しくなっています。あらゆる学校が、CNKIで検査することを義務づけています。

—もしも剽窃が見つかった場合は、どのように対応するのですか？

答：さきほども申し上げましたように、一般にこうした問題はあまり発生しません。雑誌論文については、契約の際にチェックソフトを各雑誌社に配布しています。かれらは自分自身の名譽のため、学術的に酷い論文を掲載したくありませんので、原稿を受け取る際に注意しています。それでも、たまにチェック漏れ



文献のスキャン



裁断、スキャン済み雑誌の封筒（管理用のバーコード付き）

の問題があり、告発を受け取ったり、話を聞かされたりします。不正行為だと指弾されたら、我々は公開を取りやめるでしょう。でもやはり基本的に検証を経てからということになります。誹謗中傷の可能性もありますので。——政治的な原因で論文を修正したり公開を取りやめたりすることはありますか？

答：こうした政治的問題は、中国ではあり得ます。とりわけ、社会科学ではこうした問題が存在し得ます。ですが、CNKI が論文を出版したわけではなく、一般に二次的出版になりますので、政治的問題は一次出版の段階で各出版社がクリアしており、基本的にはそのような問題が存在しないはずで。しかし、我々にも確かに政治部門はありますし、やはりもう一度見てみて、比較的重大な問題があったら、雑誌社と連絡を取って、確認するかもしれません。一般的にはよほどのことがない限り、ほとんど関与しません。

◇ データの入力と漢字処理

——論文を収録する際には、自分でスキャンして加工しているのですか？

答：始めたばかりのころは、全てスキャンしてOCRで認識させていましたが、今では大半の雑誌社が Word ファイルあるいはデジタル原稿を持っています。我々は彼らに出版ツールを提供しており、そのプラットフォームを使うと、彼らの方で原稿が確定した後、送信すれば、我々の方で自動的に出版できる、という便利な状況になっています。デジタル版を提供できないようなケースでは、やはりスキャンしています。提携先が提供する内容に応じて作業しています。

我々が入力するものについては、論文管理センターが論文ひとつひとつにバーコードをつけて、どのプロセスにあるか一括管理しています。

——あなた方の OCR ソフトは、独自に開発したものですか？



文献のデジタル化作業風景

答：OCR ソフトは、中国や世界で比較的流行している、3種類のソフトを使っています。3種を同時に使って、それぞれの結果を付き合わせて、認識率を確保しています。

——あなた方のシステムやOCRはユニコードのどの範囲まで処理できますか？

答：ユニコードのバージョンは更新され続けていて、現在は 11 でしたか。バージョンアップにはゆっくりと対応しています。公表されたからすぐ使えるようになるわけではありませんから。現在、11 まで対応できるのかどうかは確認していません。

——異体字への対応ですが、例えば工具書庫で、簡体字の「说文」と繁体字の「説文」の検索結果は同じですが、「説文」は別扱いになっている、というような例があります。こうした所謂コードセパレート文字に対応できませ



んか？

答：これらは専門の箇所漢字のテーブルを作っていますが、この問題には注意していませんでした。これは確かに面倒な問題ですが、将来的に改善できるでしょう。実のところ英語の検索でも似たような問題が発生しており、英語の検索結果と中国語の検索結果が完全に一致しないのです。この問題の解決は確かに簡単ではなく、中国語と英語を関連付けるしかなさそうです。

——私は以前から、中国でユーザー登録して日本で使っていたのですが、今年になってから工具書庫にログインできなくなりました。これは、東方書店が工具書庫のサービスを始めたことを関係しているのですか？

答：昨年中国国内でチャージした場合、海外では使えなくなりました。海外で販売しているカードと価格差があって、海外の代理店から販売に差し障りがあるとクレームを受けて、停止したのです。



自炊・モバイル時代の 中国語 OCR 徹底比較

紙のドキュメントから電子テキストを作成する簡易かつポピュラーな方法が、OCR (光学文字認識ソフト) の利用である。

中国語 OCR は 1990 年代後半、台湾系の丹青 OCR・蒙恬 OCR などの登場によって、中国語文献電子化の現実的方法として我々の視野に入ってきた。当時、そうした OCR ソフトは高額であったし、スキャナも安くはなく、なにより、動かすためには台湾版や中国版の Windows を導入する必要があった。

OS の多言語処理機能の発展につれて、OCR ソフトも多言語化していった。かつては単独パッケージで販売されていたものが、やがて劇的に安くなったスキャナや複合機に同梱されるようになり、さらにさまざまな PC ソフトやスマホアプリに OCR モジュールが組み込まれたことで、ユーザーが意識的に利用しているかどうかはさておき、どこにでも存在するありふれた技術に変わっている。ペーパーレス化が進む情報化社会を支える、基盤の一つになっている、ともいえよう。

こうした OCR を取り巻く状況の変化は、デジタル文書のあり方や活用方法にも影響を及ぼすものであるし、それはまた中国学の研究でもさまざまに活用されるべきものであると考える。

かかる観点から、本特集では OCR モジュールが組み込まれた PC ソフト・スマホアプリを幾つか取り上げ、使い方を解説するとともに、中国語認識機能を中心に OCR としての性能を検証してみたい。

CONTENTS

いまなぜ OCR なのか	千田 大介… 170
PC ソフト 4 種の OCR 機能	千田 大介… 172
PC ソフト OCR 結果比較	千田 大介… 177
Office Lens と Adobe Scan	千田 大介… 187
中国語 OCR スマホアプリ比較	田邊 鉄… 191

いまなぜ OCR なのか

千田 大介 (ちだ だいすけ)

❖ OCR ニーズの変化

OCR ソフトは、テキストを校訂するツールと一体になっており、認識結果を逐一チェックしながら、誤認識文字を修正していくようにできている。1990 年代当時、テキストデータの蓄積がまだ多くなかったこともあり、中国学において OCR は、主に分析の対象となる文献を丹念にコツコツと電子テキスト化するツールとして使われていたと思われる。

そうした状況は、近年、大きく変化した。中華圏において数多くの文献データベースが開発され、またオンラインで提供されるフリーテキストデータも充実したことで、研究対象となる文献の電子テキスト化を個人が行う必要性が乏しくなったし、また中国の人文系情報企業に電子テキストの作成を委託することも可能になった。機械が誤認識する漢字は人間も誤認識しやすいために、OCR の認識結果の校正にかなりの手間暇がかかることを考慮すれば、もはや研究者が OCR ソフトで電子テキストをコツコツ構築する時代ではなくなった、といえよう。

一方、この 10 年ほどで ScanSnap に代表されるドキュメントスキャナの普及に伴い、書籍の全冊スキャン、所謂「自炊」が一般化した。その背景には、PC スペックの向上、オンライン・オフラインのストレージ容量の増加、インターネット回線の強化などによって、自炊された画像ファイルが以前より遥かに軽く容易に取り扱えるようになったことがある。またスマートフォンでも専用アプリによって、書籍やドキュメントを手軽にスキャンできるようになっている。

そうした状況下において OCR は、高精度の電

子テキストを作成するツールとしてよりも、むしろ自炊・撮影した画像ベースの PDF を検索・編集可能にするためのツールとして利用されるシーンが増えていると思われる。その場合、認識精度が多少落ちてでも、必要なことが書かれている箇所を見つける手助けになればよい、と、割切って使うことになろう。

また、スマホのスキャンアプリでは、撮影・整形そして検索可能な PDF への加工までをこなせるものも増えており、書籍の引用したい箇所を手軽にテキストデータ化することができる。

❖ 取り上げるソフト・アプリ

かかる現状をふまえて、本特集では検索可能 PDF を作成するツール、また画像 PDF を編集可能なドキュメントに変換するツールについて検証するが、OCR 専用ソフトは取り上げず、OCR モジュールを備えた代表的な PC ソフト・スマホアプリを取り上げる。

まず多言語 OCR 機能を中心に使い方を解説する。その上で、サンプル文書の認識結果をもとに、それぞれの OCR 機能の精度や特色を比較・検討する。たとえある程度割切った使い方をするにせよ、複数の選択肢があるのであれば、やはり効率の高い方法で検索可能な PDF や電子テキストを生成するに越したことはないのは自明であろう。

テストした PC ソフト・スマートフォン(タブレット) アプリは、以下の通りである。

◇ PC ソフト

- Adobe Acrobat Pro DC (アカデミックライセンス版)
- ScanSnap Manager

- Microsoft Word 2016
- Google ドキュメント

◇ スマートフォンアプリ

- Office Lens
- Adobe Scan
- Google 翻訳アプリ
- OCR
- CamScanner

◇ テスト環境

PC ソフトについては、Windows 10 Education 64bit 環境でテストした。なお、Google ドキュメントは厳密に言えば PC ソフトではないが、PC 環境からテストしたため、便宜上、ここに分類する。

スマホアプリのうち Office Lens と Adobe Scanha については Xperia XZ2 (Android 8.0.0) を、Google 翻訳アプリ・OCR・CamScanner の 3 種については、Moto G5 Plus (Android7.0) を、それぞれ使用した。

◇ 試験に用いたドキュメント

簡体字雑誌記事 ①

『FOOD & WINE』2012 年 5 月号の p.60・61 を用いた。p.60 は白抜き文字が認識できるか否か、p.61 は雑誌記事の複雑なレイアウトを認識できるかを、主に確認する。

簡体字横書き書籍 ②

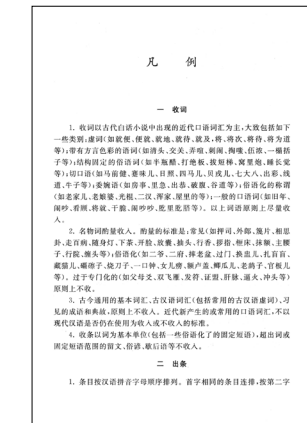
『白話小説語言詞典』(白維国主編、商務印書館、2010) の一部を用いた。二段組みのレイアウトやピンインを認識できるかをチェックしたほか、Word・Google ドキュメントがどの程度の大きさの PDF にまで対応できるかどうかの試験にも用いた。また、簡体字の認識効率の比較に、一段組みの凡例ページを用いた。

繁体字縦書き文献 ③

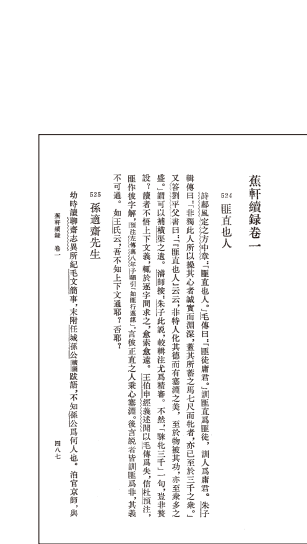
『蕉軒續録』(方溶師撰、中華書局、1995) の p.487 を用いた。繁体字の認識率のほか、縦書きレイアウトの認識、および傍線の施された箇所を抜いたことを確認した。



①



②



③

PC ソフト 4 種の OCR 機能

千田 大介 (ちだ だいすけ)

❖ Adobe Acrobat DC

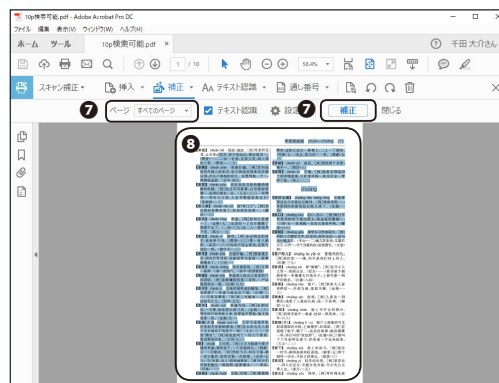
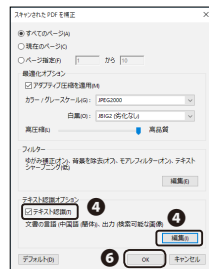
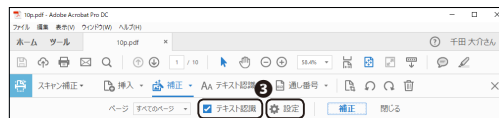
ネイティブの PDF ソフトである Adobe Acrobat は、当然のことながら OCR 機能による検索可能 PDF の作成や、OCR で認識したテキストの書き出しなど、極めて充実した機能を持っている。

PDF の作成・編集にフル対応するのは、有償版の Adobe Acrobat DC になる。無償版の Adobe Acrobat Reader DC では、検索可能 PDF からテキストをコピーすることはできるが、OCR 機能が使えないため、検索可能 PDF を作成したり、画像ベースの PDF からテキストベースの Word・Excel ファイルを書き出したりといったことはできない。

◇ 検索可能 PDF の作成

Acrobat DC では以下の手順で、検索可能 PDF を作成できる。

- ①加工したい PDF を開き、「ツール」→「スキャン補正」をクリック。
- ②「補正」→「スキャンした文書」をクリック。
- ③「テキスト認識」がチェックされていることを確認し、「設定」をクリック。
- ④「テキスト認識オプション」で「テキスト認識」がチェックされていることを確認し、「編集」をクリック。
- ⑤「文書の言語」を選択する。「出力」は「検索可能な画像」とする。
- ⑥「OK」をクリックして、全てのダイアログを閉じる。
- ⑦「ページ」が「全てのページ」になっていることを確認して「補正」をクリック。



⑧元の画像の傾きや汚れが処理され、検索可能な PDF が作成される。

②のスキャン補正ツールで「テキスト認識」を選ぶと、画像の補正なしに OCR 処理だけを行うことができる。

⑦の「ページ」で対象ページを指定すると、一部分のみに OCR をかけることができる。ページ数の多い PDF では OCR 処理に時間がかかるので、引用の必要のある箇所のテキストが欲しいような場合には、ページを指定した方がよからう。

Pro 版では、⑤の「出力」で「編集可能なテキストと画像」を選択できる。こちらを選ぶと、文字ごとに細分された画像と透明テキストとが一体化したり、あるいは元の画像に似たフォントで表示されるように変換されたりして、文書の編集が可能になる。とはいえ、OCR の誤認識の可能性や、誤って字句を削除してしまう危険性を考慮すると、少なくとも研究に利用する文献・ドキュメントでは、画像がそのまま保持される「検索可能な画像」を選択した方が安全であろう。

認識されたテキストは、「ツール」→「スキャン補正」→「テキスト認識」→「認識されたテキストを修正」で、校正することができる。これは、読み取り結果に自信のない箇所を順に表示・確定する機能だが、使い勝手は今ひとつである。


なお、以上のスキャン補正を行うと、ドキュメントの画像が、背景と文字とに分割され、「PDF を編集」で背景画像や地色を消去することができるようになる。

◇ PDF からテキストを書き出す

Acrobat では PDF を Word ファイルなどの形式に書き出すこともできる。このとき、検索可能に加工していない画像の PDF に、直接 OCR をかけて書き出すこともできる。Word ファイルに書き出す場合は、以下の手順になる。

①書き出したい PDF を開き、「ツール」→「PDF を書き出し」をクリック。

②「Microsoft Word」、「Word 文書」が選択さ

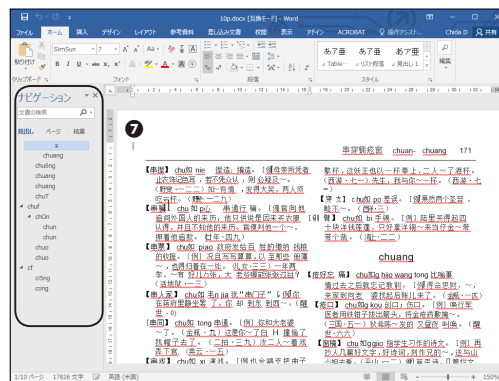
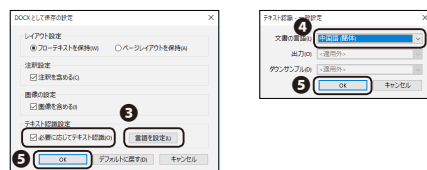
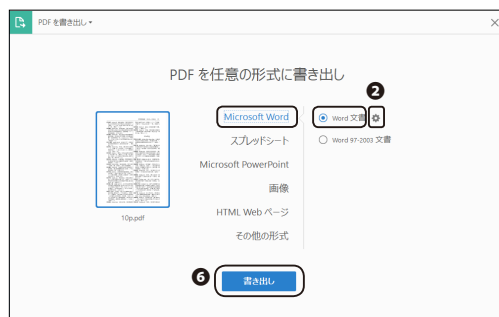
れていることを確認し、「Word 文書」右の「」をクリック。

③「テキスト認識設定」で「必要に応じてテキスト認識」にチェックを入れる。「言語を設定」をクリック。

④「文書の言語」を選択する。

⑤「OK」をクリックして、ダイアログを全て閉じる。

⑥「書き出し」をクリック、保存先フォルダ



とファイル名を指定する。

⑦ Word ファイルが書き出される。

書き出された Word ファイルでは、段組などのレイアウトもある程度再現されているほか、ナビゲーションウィンドウを見ると、見出しスタイルも自動設定されていることがわかる（正確さはさておく）。

また、Word のほか、Excel・PowerPoint・HTML・画像・RTF・テキストなどの形式で書き出すこともできる。検索可能 PDF からプレーンテキストで書き出すと、既存の透明テキストがそのまま書き出される。

◇ Adobe Acrobat Reader からの書き出し

無償の Acrobat Reader であっても、検索可能 PDF を開いて、テキストデータをコピー＆ペーストすることは可能である。

Acrobat Reader では、「ファイル」→「その他の形式で保存」→「テキスト」で、テキストファイルに書き出すこともできるのだが、筆者の環境では OCR 時の言語設定が日本語であるファイルしか、テキスト保存できなかった。また、Word・Excel ファイルへの書き出しは、有償で機能追加する必要がある。

❖ ScanSnap

ScanSnap は、富士通という有名メーカーから発売された、比較的安価かつ高性能なドキュメントスキャナとして、我が国において同ジャンルのデファクトスタンダードの製品になっている。「破壊自炊」（裁断して背を切り離れた書籍の全文スキャン）は、ScanSnap とともに一般化したと言ってもよからう。

ScanSnap では Acrobat を使うことなく、スキャンしたドキュメントから検索可能な PDF を作成することができる。

◇ スキャンして検索可能 PDF を作成する

ScanSnap で検索可能 PDF を作成する方法は、



主に 2 つある。1 つはドキュメントを検索可能な PDF としてスキャン・保存する方法である。

- ① スタートメニューから「ScanSnap Manager」→「ScanSnap Manager の設定」を開く。
- ② 「ファイル形式」タブをクリック。
- ③ 「ファイル形式の選択」で「PDF」を選択。
- ④ 「検索可能な PDF にします」をチェック。
- ⑤ 「対象言語」を変更する。はじめて言語を変更した際に、多言語 OCR パック等の追加ダウンロード画面が開くので、必要な言語・ツールをチェックする。
- ⑥ 「対象ページ」を「全ページ」に設定。
- ⑦ 「OK」をクリックして閉じる。

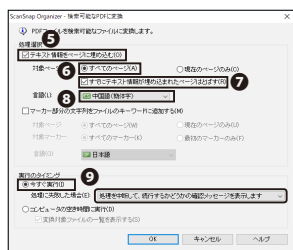
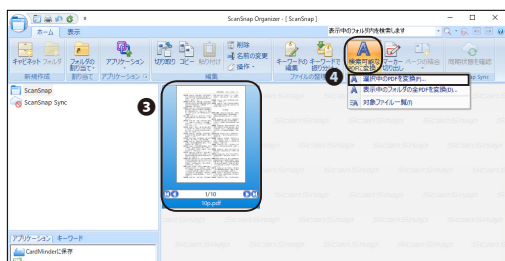
自炊時には、「読み取りモード」タブの「オプション」で「裏写りを軽減します」、「白黒読み取り時の濃度を調整します」、「文字をくっきりします」、「文字列の傾きを自動的に補正します」を、いずれも選択しておいた方がよい。画像ファイルが読みやすくなるばかりでなく、ScanSnap の OCR が裏写りや汚れを、文字に誤認してしまう危険性を軽減することができる。

これで、ScanSnap のスキャンボタンを押してスキャンすると、検索可能な PDF が作成される。ただし、ページ数が多いと OCR に時間がかかることがあるので、スキャン時間を節約したい場合には、画像 PDF としてスキャンしておき、後から ScanSnap Organizer によって検索可能 PDF に変換する方がよからう。

◇ 既存の PDF を検索可能 PDF に変換する

ScanSnap に同梱されるスキャンドキュメント管理ソフトである ScanSnap Organizer では、ScanSnap Manager で作成した画像 PDF ファイルに OCR をかけて、検索可能 PDF に変換することができる。

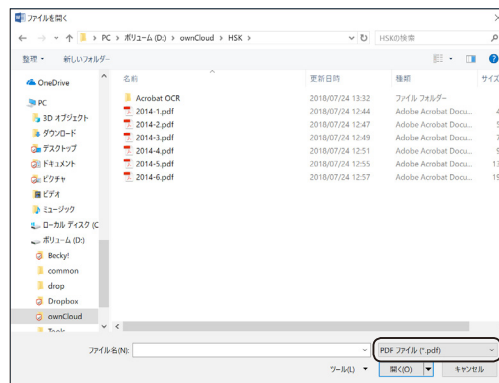
- ① ScanSnap フォルダ (通常、ドキュメントの「ScanSnap」フォルダ) に、変換したい PDF ファイルを保存する。
- ② スタートメニューから「ScanSnap Organizer」→「ScanSnap Organizer」を開く。
- ③ 変換したい PDF をクリックして選択する。
- ④ 「検索可能な PDF に変換」→「選択中の PDF を変換」をクリック。
- ⑤ 「テキスト情報をページに埋め込む」がチェックされていることを確認する。
- ⑥ 「対象ページ」で「すべてのページ」を選択。
- ⑦ 「すでにテキスト情報が埋め込まれたページはとばす」のチェックを外す。
- ⑧ 「言語」の種類を選択する。
- ⑨ 実行のタイミングを選択する。ここでは「今すぐ実行」をチェック、「処理に失敗した場合」は「エラーを無視して処理を続行します」とする。



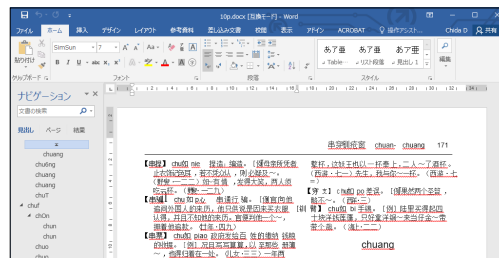
なお、検索可能 PDF に変換できるのは、ScanSnap でスキャンして作成した PDF ファイルに限られるので、注意されたい。他のソフトでレタッチするなどした画像ファイルでも、ScanSnap で作成した PDF を ScanSnap Organizer ビューアで開き、「ページの挿入」で画像ファイルを指定する、という方法で、対応できることはできる。

❖ Microsoft Word 2016

Word は PDF ファイルの読み込み・編集に対応している。PDF ファイルを開くには、Word を立ち上げて「他の文書を開く」でフォルダを指定し、ファイルの種類を「PDF」に設定し、目的のファイルを指定すればよい。PDF ファイルを右クリックし、「プログラムから開く」→「Word 2016」でもよい。



このとき、対象の PDF が文書画像であった場合、自動で文字を認識し Word ドキュメントに変換してくれる。この時点ではまだ PDF ファイル扱いなので、Word ファイルとして保存する。

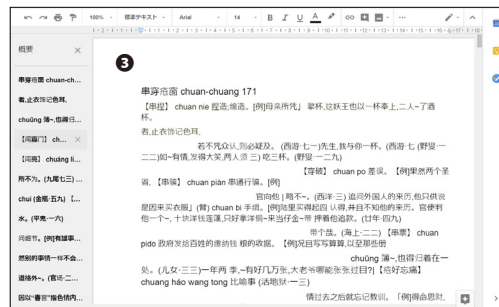


PDFを開くとき、言語の設定などのダイアログは表示されず、Wordが自動で判別する。ProofingToolsを導入しているか否かに関係なく、簡体字・繁体字中国語、さらには複数言語が混在していても認識する。

認識できるファイルサイズの上限は512Mだが、サンプル②で実験したところでは、300dpiでスキャンしたA5判300ページ(336MB)は読み込めたが、400ページ(447MB)は読み込みに失敗した。ページ数が増えると、それだけ認識にも時間がかかる。

なお、検索可能PDFを読み込ませても、Wordは透明テキストを無視して、改めて画像データから文字を認識しなおす。

認識結果は、図のように、レイアウトがある程度再現され、また自動で見出しスタイルも設定される。



❖ Google ドキュメント

Google ドキュメントは、Googleがオンラインで提供する、Wordライクなワープロである。Word同様、画像PDFをテキストに変換して読み込む機能を、以下の手順で使うことができる。

- ❶ Googleドライブに変換したいPDFファイルをアップロードする。
- ❷ アップロードしたファイルを右クリックし、「アプリで開く」→「Googleドキュメント」をクリック。
- ❸ 変換された文書が、Googleドキュメントで表示される。

Googleドキュメントで変換できる画像PDFは、比較的小さなものに限られる。サンプル文書②で実験したところ、20ページ約22Mは読み込めたが、25ページ約27Mでは変換に失敗した。短めの論文一本、本の一節程度の分量までと考えたほうが良さそうだ。

また、スマホアプリ版のドキュメントでは、PDFファイルを開くことができない。変換機能が使えるのはPCのブラウザ版だけであるようだ。

このように機能・性能は明らかにWordに劣るもの、無償で提供されるオンラインアプリであることを考えれば、高望みはできない。

ScanSnap Manager

右段→左段の順にレイアウトを認識し、かつ最後の一行を読まなかった。

Microsoft Word 2016

段組の再現が上手くいかずに文字列の幅が一部狭くなっているものの、文字そのものはほぼ問題なく読んでいる。

Google ドキュメント

文字は全て読み取ったものの、段組を横書き一段として認識した。また、元の白抜き文字を再現しようとしたのであろう、文字色が薄黄色に設定された。

◇ 雑誌の複雑なレイアウト

原文は、見出しの下に、段組のブロックが複数

存在する、かなり複雑なレイアウトになっている。

なお、写真下の枠は人物紹介で独立している。正しい順番は仮に③としたが、①②でも問題ない。

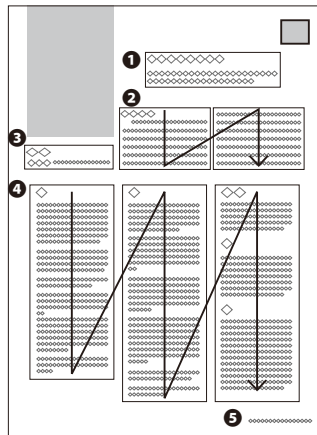
ソフト	結果
Adobe Acrobat	○
ScanSnap Manager	○
Microsoft Word 2016	○
Google ドキュメント	△

各ソフトがサンプルのレイアウトブロックをどの順番で認識したのかを、以下に図示する。

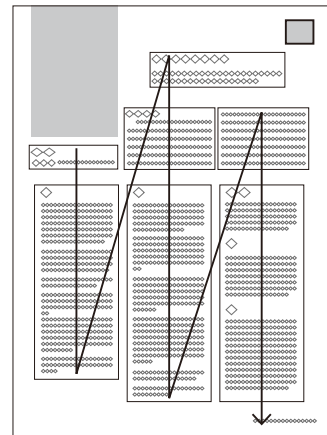
Adobe Acrobat

全体を3段組として認識している。

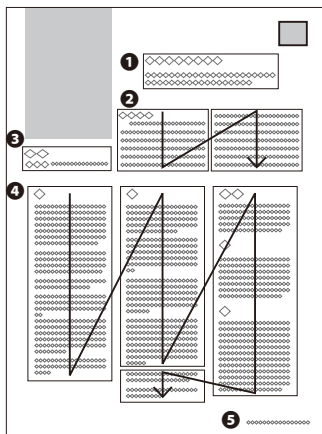
ScanSnap Manager



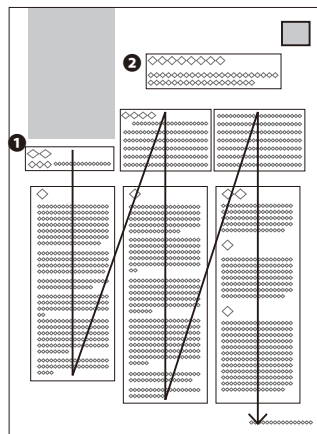
正しいテキスト枠の順番



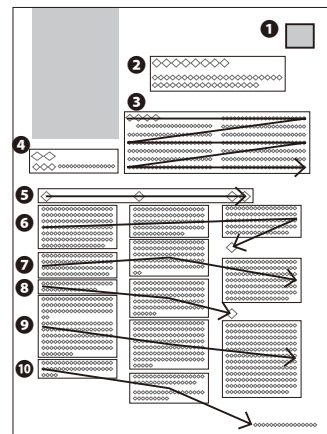
Adobe Acrobat



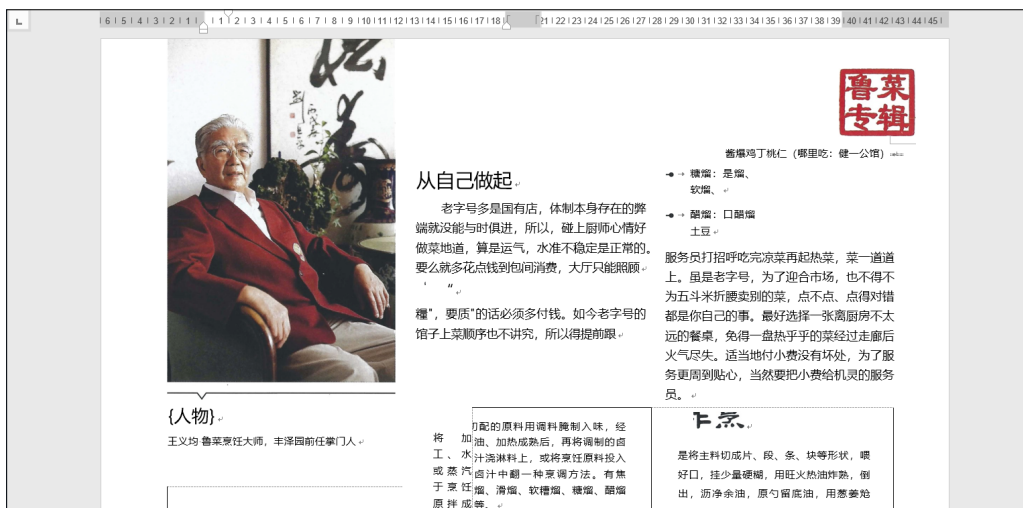
ScanSnap Manager



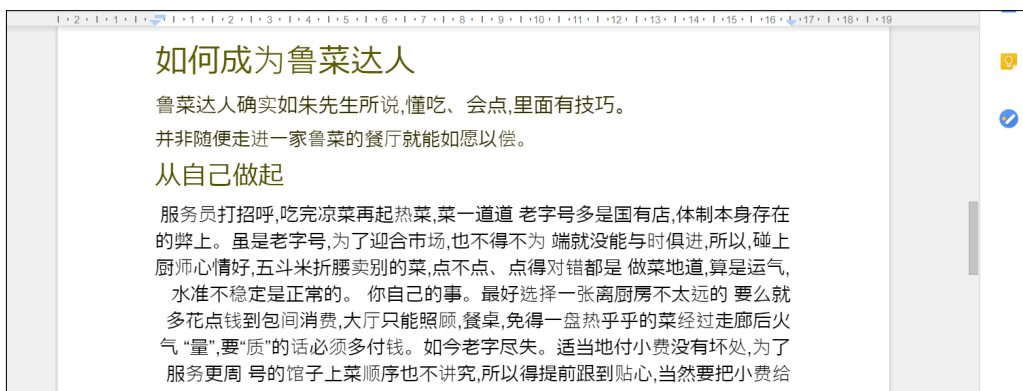
Microsoft Word 2016



Google ドキュメント



Microsoft Word 2016



Google ドキュメント

今回試した中では、原文ともっとも近い順序で認識した。

Microsoft Word 2016

Acrobat 同様に全体を3段組のドキュメントとして読んだが、見出し部分の枠を本文と別枠として認識できている。また、変換後のWordファイルでは、画像の配置、本文の段組や罫線など、かなりの程度、元の誌面を再現してる。段組に見られる乱れは、用紙の高さが不足してテキストが段から溢れたことに起因するようだ。

Google ドキュメント

見出し下の2段組のブロックを1段で認識した。その下の3段組については、段落は認識しているものの、横方向に繋げようという指向が強い。原文のレイアウトは再現されない。

総じて、各ソフトともに段落のブロックは認識できており、それぞれブロックをどのように繋がっていると解釈するかのアロゴリズムの相違が、認識結果の良し悪しに影響したものと思われる。

❖ サンプル文書②

◇ 2段組レイアウトとピンイン

②の辞書本文の横書き2段組のレイアウトは、細かい問題の起きるソフトが多かった。

Adobe Acrobat

検索可能テキストの場合は段を左から右に、正しく認識している。ただし、奇数ページのヘッダは、右段の一部として認識された。

Adobe Acrobat Pro DC

凡例

一收词

1. 收词以古代白话小说中出现的近代口语词汇为主, 大致包括如下一些类别: 虚词(如就便、便就、就地、就待、就及、将、将次、将待、将为道等); 带有方言色彩的语词(如清头、交关、弄喧、刺闹、掏嗤、伍浓、一榻括子等); 结构固定的俗语词(如半瓶醋、打绝板、拔短梯、窝里炮、睡长觉等); 切口语(如马前健、**囊**味儿、日照、四马儿、贝戎儿、七大八、出彩、线道、牛子等); 委婉语(如房事、里急、出恭、破腹、谷道等); 俗语化的称谓(如老家儿、老娘婆、光棍、二汉、浑家、屋里的等); 一般的口语词(如旧年、闹吵、看顾、将就、干脆、闹吵吵、**脱里胞脰**等)。以上词语原则上尽量收入。

2. 名物词酌**倡**收入。酌量的标准是: 常见(如押司、外郎、**蔑**片、相思卦、走百病、随身灯、下茶、开脸、**放囊**、抽头、行香、**才多指**、**木匣**床、抹额、主腰子、行院、缠头等); 俗语化(如二爷、二府、摔老盆、过门、换盅儿、扎盲盲、藏猫儿、**缠碗**子、烧刀子、一口钟、女儿**捞**、额卢盖、**卿**瓜儿、老**鸽子**、官板儿等)。过于专门化的(如父母**交**、双飞雁、发符、证盟、肝脉、逼火、冲头等)原则上不收。

3. 古今通用的基本词汇、古汉语词汇(包括常用的古汉语虚词)、习见的成语和典故, 原则上不收入。近代新产生的或常用的口语词汇, 不以现代汉语是否仍在使用为收入或不收入的标准。

4. 收条以词为基本单位(包括一些俗语化了的固定短语), 超出词或固定短语范围的留文、俗谚、歇后语等不收入。

二出条

1. 条目按汉语拼音字母顺序排列。首字相同的条目连排, 按第二字

あることを、それぞれ示す。スペースについては、本文中に挟み込まれ検索に影響すると思われるもののみを「」に置き換えた。改行や句読点・符号の全角・半角についてはチェックしていない。

Adobe Acrobat

誤字が延べ15字(元の漢字を2字に分割して誤認識したものは1字とした)、句読点の抜けが1字であった。どのような素性の漢字を誤認識したのかをま

ScanSnap Manager

凡例

一收词

1. 收词以古代白话小说中出现的近代口语词汇为主, 大致包括如下

一些类别: 虚词(如就便、便就、就地、就待、就及、将、将次、将待、将为道

等); 带有方言色彩的语词(如清头、交关、弄喧、刺闹、掏**嚏**、伍浓、**一棍**括

子等); 结构固定的俗语词(如半瓶醋、打绝板、拔短梯、窝里炮、睡长觉

等); 切口语(如马前健、**壹**味儿、日照、四马儿、贝戎儿、七大八、出彩、线

道、牛子等); 委婉语(如房事、里急、出恭、破腹、谷道等); 俗语化的称谓

(如老家儿、老娘婆、光棍、二汉、浑家、屋里的等); 一般的口语词(如**日**年、

闹吵、看顾、将就、干脆、闹吵吵、**屹里屹脰**等)。
以上词语原则上尽量收

人。

2. 名物词酌量收**人**。酌量的标准是: 常见(如押司、外郎、**蔑**片、相思

圭卜、走百病、随身灯、下茶、开脸、**放囊**、抽头、行香、**掺**指、**框**床、抹额、主腰

子、行院、缠头等); 俗语化(如二爷、二府、摔老盆、过门、换盅儿、扎盲盲、

藏猫儿、**僵**膘子、烧刀子、一口钟、女儿**痹**、额卢盖、**卿**瓜儿、老**鸽子**、官板儿

等)。
过于专门化的(如父母**交**、双飞雁、发符、证盟、肝脉、逼火、冲头等)

原则上不收。

3. 古今通用的基本词汇、古汉语词汇(包括常用的古汉语虚词)、习

见的成语和典故, 原则上不收**人**。近代新产生的或常用的口语词汇, 不以

现代汉语是否仍在使用为收**人**或不收**人**的标准。

4. 收条以词为基本单位(包括一些俗语化了的固定短语), 超出词或

固定短语范围的留文、俗谚、歇后语等不**收人**。

二出条

1. 条目按汉语拼音字母顺序排列。
首字相同的条目连排, 按第二字

とめたのが、前ページの表である。[] で括ったのは、2文字に分割して誤認識された文字を表す。

GB2312コードに収録される文字が8字、収録

Microsoft Word 2016

凡例

▼收词

1・收词以古代白话小说中出现的近代口语词汇为主,大致包括如下一些类别:虚词(如就便、便就、就地、就待、就及、将、将次、将待、将为道等);带有方言色彩的语词(如清头、交关、弄喧、刺闹、掏嗤、伍浓、一榻括子等);结构固定的俗语词(如半瓶醋、打绝板、拔短梯、窝里炮、睡长觉等);切口语(如马前健、蹇味儿、日照、四马儿、贝戎儿、七大八、出彩、线道、牛子等);委婉语(如房事、里急、出恭、破腹、谷道等);俗语化的称谓(如老家儿、老娘婆、光棍、二汉、浑家、屋里的等);一般的口语词(如旧年、闹吵、看顾、将就、干脆、闹吵吵、**胞里胳膊**等)。以上词语原则上尽量收**入**▼

2・名词酌量收**入**。酌量的标准是:常见(如押司、外郎、篋片、相思卦、走百病、随身灯、下茶、开脸、放囊、抽头、行香、**柜**床、抹额、主腰子、行院、缠头等);俗语化(如二爷、二府、摔老盆、过门、换盅儿、扎盲盲、藏猫儿、**僵**子、烧刀子、一口钟、女儿痲、额卢盖、鲫鱼儿、老鸽子、官板儿等)。过于专门化的(如父母交、双飞雁、发符、证盟、肝脉、逼火、冲头等)原则上不收。

3・古今通用的基本词汇、古汉语词汇(包括常用的古汉语虚词)、习见的成语和典故,原则上不收**入**。近代新产生的或常用的口语词汇,不以现代汉语是否仍在**使用**为收**入**或不收**入**的标准。

4・收条以词为基本单位(包括一些俗语化了的固定短语),超出词或固定短语范围的留文、俗谚、歇后语等不收**入**。

▼出条

1・条目按汉语拼音字母顺序排列。首字相同的条目连排,按**第二**字

されない文字が6字(延べ7字)であった。ここから、Acrobatの簡体字OCRが、GB2312収録文字対応であることが窺える。

実用的な精度が得られているといえよう。

ScanSnap Manager

誤字が延べ36字、うち句読点の間違いが延べ5字であった。

このうち漢字は延べ31字、19文字を誤認識し

Google ドキュメント

凡

例

-收词

[1. 收词以古代白话小说中出现的近代口语词汇为主,大致包括如下一些类别:虚词(如就便、便就、就地、就待、就及、将、将次、将待、将为道等);带有方言色彩的语词(如清头、交关、弄喧、刺▼、▼**嘴**、伍浓、一榻括子等);结构固定的俗语词(如半瓶醋、打绝板、拔短梯、窝里炮、睡长觉

等);切口语(如马前健、▼**味**儿、日照、四马儿、贝▼**儿**、七大八、出彩、线道、牛子等);委婉语(如房事、里急、出恭、破腹、谷道等);俗语化的称谓(如老家儿、老娘婆、光棍、二汉、浑家、屋里的等);一般的口语词(如旧年、闹吵、看顾、将就、干脆、闹吵吵、**胎**里▼**脑**等)。以上词语原则上尽量收**入**。

2. 名词酌量收**入**。酌量的标准是:

常见(如押司、外郎、**篋**片、相思卦、走百病、随身灯、下茶、开脸、放囊、抽头、行香、**拇**指、**柜**床、抹额、主腰子、行院、缠头等);俗语化(如二爷、二府、摔老盆、过门、换**蓝**儿、扎盲盲、藏猫儿、**僵**子、烧刀子、一口钟、女儿**疹**、额卢盖、鲫鱼儿、老**鹤**子、官板儿等)。过于专门化的(如父母交、双飞雁、发符、证盟、肝脉、逼火、冲头等)原则上不收。

3. 古今通用的基本词汇、古汉语词汇(包括常用的古汉语虚词)、习见的成语和典故,原则上不收**入**。近代新产生的或常用的口语词汇,不以现代汉语是否仍在**使用**为收**入**或不收**入**的标准。

4. 收条以词为基本单位(包括一些俗语化了的固定短语),超出词或固定短语范围的留文、俗谚、歇后语等不收**入**。

二出条

1. 条目按汉语拼音字母顺序排列。首字相同的条目连排,按**第二**字

た。GB2312外字を軒並み誤認識したほか、「一」を「一」「_」に読み違えている箇所が目立ち、出現6字中、正確に読めたのは一箇所にとどまった。また「入」を「人」に誤る、「旧」を「|日」とするなど、基本的な文字の誤認識が目立った。

認識結果は全体としてAdobe Acrobatに劣り、

「一」などの基本的文字の誤認識の多さが気になる。ただ、誤字の発生率自体はそこまで高くはないので、検索可能な PDF を作成するツールとしては、検索時に若干の注意を要するものの、実用的なレベルには達しているといえよう。

Microsoft Word 2016

誤認識した漢字が延べ 10 字、漢字の脱落が 2 字、句読点の脱落が 2 字であった。脱落は見出しの漢数字 2 箇所、変換された Word ファイルでは画像として処理されていた。

漢字の認識では「人」を 6 箇所全てで「亼」に誤った。そのほかの誤認識はいずれも GB23312 未収録文字であるが、「箴」・「爻」は認識できていた。

今回試した中で、認識率が最も高かった。

Google ドキュメント

Google ドキュメントでは、数ヶ所で本文中にスペースが入り込んだが、いずれも原文の行末の箇所であった。

誤認識された漢字が延べ 13 字、脱落した漢字が 5 字、計 18 字であった。Acrobat などと同様に、GB2312 未収録文字を誤認識・脱落したものが目立つほか、「腦」・「簽」は誤認識した結果が繁体字になっていた。

まとめ

以上の各ソフトのテスト結果をまとめると、以下ようになる。

ソフト	合計 (全 630 字)		漢字計 (全 492 字)	
	誤・脱字数	錯誤率	誤・脱字数	錯誤率
Acrobat	16	2.54%	15	3.05%
ScanSnap	36	5.71%	31	6.30%
Word	14	2.22%	12	2.44%
Google	18	2.86%	18	3.66%

GB2312 未収録文字が出現する、少々意地悪なテキストで試みたため、錯誤率が高めになっているが、それでも Acrobat・Word・Google ドキュメントの認識率は 95% を超えている。

最も認識効率の高かったのが Word である。Word は前述のように、読み込み可能ページ数に限界があり、また文字を認識した Word ファイル

として出力されるので、検索可能 PDF 作成用途には使えないが、さほど長くないドキュメントを電子テキスト化して加工・引用したいような場合の、ファーストチョイスとなろう。

検索可能な画像 PDF を作成するのであれば、Adobe Acrobat がベストである。ScanSnap Organizer は、他に比べて結果が一等劣っている。ScanSnap はしばらく前から Adobe Acrobat の同梱を取りやめているが、そうであるならば、Acrobat と遜色ない機能を独自ソフトに持たせてもらいたいものである。

Google ドキュメントは、文字認識のスコア自体は悪くないが、認識できなかった文字が脱落するのは困ったものである。ページ読み込み数の限界も Word 以上にシビアである。Word が使えない環境で、その代替として使うものと位置づけられよう。

❖ サンプル文書③

◇ 縦書きレイアウトの認識

いずれも、縦書きレイアウトの認識は、横書きに比べていささか問題が多いようである。

Adobe Acrobat

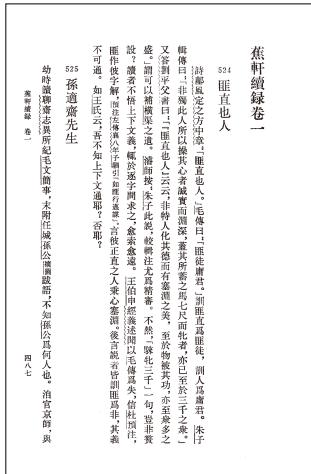
行の前半部分と後半部分の順番が、2 箇所に入れ替わった。入れ替わった箇所は、いずれも傍線部の前後であるので、これはあるいは縦書きレイアウトの認識ではなく、傍線処理の問題に起因するのかも知れない。2 行目の縦中横の数字も、後の項目タイトルと順番が入れ替わっている。コロン、二重鉤括弧なども上手く認識できなかった。

ScanSnap Manager

行の順序は正しく認識しているが、3 行目、本文の冒頭を 1 行飛ばしている。鉤括弧は縦書き用の 90 度回転したものを使っている。二重鉤括弧は認識できず、コロンも 2 つのピリオドになっている。ページ末尾は、汚れを文字として拾ったようで、意味不明な記号の羅列になっている。

Microsoft Word 2016

縦書きレイアウトはほぼ再現できているが、本



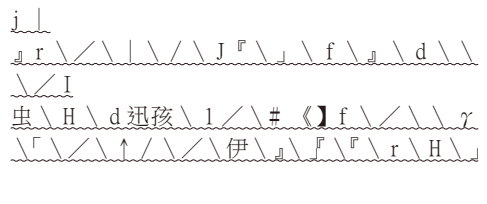
ScanSnap Manager

蕉軒續錄卷一
“匪直也人

輯傳曰..一非獨此人所以操其心者誠實而淵深
蓋其所蓄之馬七尺而牝者’亦已至於三千之眾。
一又答劉平父書曰..一「匪直也人」云云’非
特人化其德而有塞淵之美’至於物被其功’亦
至眾多之
盛。一謂可以補橫渠之遺。濬師按..朱子此說’
較輯注尤為精審。不然’一驪牝三千▼三句’豈
非贅設？讀者不悟上下文義’輒於逐字間求之’
愈索愈遠。王伯申經義述聞以毛傳為失’信杜
預注’匪作彼字解’（韻注）麟麟麟引一
如匪行邁謀一。言彼正直之人秉心塞淵。後言
說者皆訓匪為非，其義不可通。如王氏云’吾
不知上下文通耶？否耶？

錙孫適齋先生

幼時讀聊齋志異所紀毛文簡事’末附任城孫公
隨園跋語’不知孫公為何人也。洎官京師’與
蕉軒續錄卷一四八七



Adobe Acrobat Pro DC

蕉軒續錄卷一

匪直也人

524

「匪徒庸君。」訓匪直為匪徒，訓人為庸君。隸
仔

寫定之寫神窟「匪直也人。屬傳曰

..

..

輯傳曰▼「非獨此人所以操其心者誠實而淵深，
蓋其所蓄之馬七尺而牝者，亦已至於三千之
眾。」

「『匪直也人』云云，非特人化其德而有塞淵之
美，至於物被其功，亦至眾多之

又濬劉渾回父書曰

..

lii

11 按

盛。」謂可以補橫渠之遺。▼隸子此說，較輯注
尤為精審。不然，「驪牝三千」一句，登非贅

..

設？讀者不悟上下文義，輒於逐字間求之，愈
索愈遠。王伯申經義述聞以毛傳為失，信杜預
注，匪作彼字解，（韻注）麟麟麟引「如
匪行邁謀」。言彼正直之人秉心塞淵。後言說者
皆訓匪為非，其義

不可通。如王氏云，吾不知上下文通耶？否耶？
525

孫適齋先生幼時讀聊齋志異所紀毛文簡事，末
附任城孫公隨園跋語，不知孫公為何人也。洎
官京師，與
蕉軒續錄卷一
四八七

文最後のフレーズが、ページ番号の後に来てしまっている。縦中横数字は画像として処理されており、鉤括弧・二重鉤括弧・コンマ・コロンなども上手く認識できていない。縦書き時の見かけを重視しているかのように思える。

Google ドキュメント

縦書きレイアウトは認識できているが、コロン・二重鉤括弧などは、やはり上手く認識できていない。

◇ 傍線と割り注

傍線と割り注の処理はいささかハードルが高かったようで、処理が上手くいっていないソフトが目立った。認識結果では、傍線箇所の誤字が白抜き、脱字が「▽」になる。

Adobe Acrobat

前述の認識順問題が発生したほか、傍線を文字の一部として認識してしまい、「詩廊風定之方中

Microsoft Word 2016

蕉軒續録卷一
 匪直也人

詩鄘風定之方中章・・→匪直也人。▼尾傳曰・
 ・→匪徒庸君。...訓匪直為匪徒'訓人為庸君。
 ▼厚輯傳曰・→非獨此人所以操其心者誠實而淵
 深'蓋其所蓄之馬七尺而牝者'亦已至於三千之
 衆▼又▼劉平父書曰・▼匪直也人~云云'非
 特人化其德而有塞淵之美'至於物被其功'亦至
 衆多之

盛。▼謂可以補▼之遺。▼按・▼子此說'
 較輯注尤為精審。不然'→駮牝三千▼二句'豈
 非贅設? 讀者不悟上下文義'輒於▼字問求之'
 '愈索愈遠。王伯申經義述聞以毛傳為失'信杜
 預注'

匪作彼字解'▼注▼襄八年子▼引→如匪行邁
 謀。言彼正直之人秉心塞淵。後言說▼皆訓匪
 為非'其義

不可通。如▼氏云'吾不知上下文通耶? 否耶?
 ▼孫適齋先生
 幼時讀聊齋志異所紀毛文簡事'未附▼▼▼▼▼
 圖跋語'不知▼公為何人也。
 蕉軒續録 卷一 四八七

泊官京師'與

章」が「溝澶澤定之消冲塵」に、「毛」が「尾」に、「朱子」が「洙子」に、それぞれ誤認識された。割り注でかつ傍線が施された「預注左傳襄八年」は「博注埤溝壤扒准」になってしまった。

ScanSnap Manager

傍線箇所認識がかなり上手くできており、「答劉平父書」・「濬師」・「朱子」などを正確に認識した。流石に割り注+傍線は辛かったようで、「預注左傳襄八年子駟引」が「頤注焮喃籐呖咩庠陔引」になってしまった。

Microsoft Word 2016

本文冒頭の「詩鄘風定之方中章」を正確に認識し、傍線を線として文字と区別して表示した。その一方で、「毛」を「尾」に誤るなど、傍線の切

Google ドキュメント

蕉軒續録卷一
 w 匪直也人

trairwhetter

詩▼定之方中章:「匪直也人。▼毛傳曰:「匪徒庸君。」訓匪直為匪徒,訓人為庸君。朱子▼輯傳曰“▼非獨此人所以操其心者誠實而淵深,蓋其所蓄之馬七尺而牝者,亦已至於三千之衆。」▼又答劉平父書曰“TT 匪直也人▼云云,非特人化其德而有塞淵之美,至於物被其功,亦至眾多之盛。謂可以補橫渠之遺。濬師按“朱子此說,較輯注尤為精審。不然,「驟▼三千▼二句,豈非贊設? 讀者不悟上下文義,輒於逐字間求之,愈索愈遠。王伯申經義述聞以毛傳為失,信杜預注,▼匪作彼字解,預注左傳襄八年子駟引「如匪行邁謀」▼言被正直之人秉心塞淵。後言說者皆訓匪為非,其義不可通。如王氏云,吾不知上下文通耶? 否耶?

11

孫適齋先生

幼時讀聊齋志異所記毛文簡事,未勝任城孫公擴圖跋語,不知孫公為何人也。泊官京師,▼蕉軒續録 卷一 四八七

り分けが上手くいっているのは一部分だけに止まる。傍線付き文字が認識されずにスペースになっている箇所も目立つ。傍線付き割り注は、画像で処理されていた。

Google ドキュメント

傍線部の認識は最も正確であった。傍線を完全に無視して、正確に読んでいる。傍線と割り注が組み合わさった「任城孫公擴圖」を正確に読んだのは特筆に値する。

◇ 繁体字の認識

コードセパレート文字の扱い

繁体字の認識では、所謂コードセパレートの扱いが問題になる。サンプル文書では「衆」(U+8846)がそれにあたり、「眾」(U+773E)の字形が台湾 Big5 コードでは採用されている。また、「為」(U+70BA)は一見、旧字体=繁体字が「爲」(U+7232)だと思えるが、Big5 の収録字形は「為」になる。

サンプル文書は「爲」になっている。

各ソフトの認識結果は以下のようになった。

Adobe Acrobat	衆	爲
ScanSnap Manager	眾	× (烏)
Microsoft Word 2016	衆	爲
Google ドキュメント	眾	為

Acrobat と Word は、原文のまま「衆」・「爲」となったが、Google ドキュメントは繁体字環境の一般的字形である「眾」・「為」に変換している。ScanSnap は「衆」を「眾」に変換した一方、「爲」を全て「烏」に誤認識した。Acrobat と Word がある程度 Unicode の異体字に対応しているのに対して、ScanSnap と Google ドキュメントは繁体字中国語を Big5 コードの文字コレクションの枠内で処理しているようだ。

誤認識文字

右表は、各ソフトの誤認識文字の一覧である。

Acrobat は、繁体字の文字認識効率自体は悪くないことがわかる。ScanSnap も Acrobat に次いで良い成績である。

Word は他のソフトがいずれも誤った「駭」を読むなど、優れた点もみられるが、「曰」・「讀」・「否」をうまく認識できないなど、凡ミスが目立った。

Google ドキュメントは、「贊」・「胜」が簡体字に誤認識されているのが目に付く。サンプル文書を繁体字・簡体字混在文であると認識したのであろうか。

まとめ

縦書き繁体字文書の認識については、各ソフト一長一短であり、決め手に欠けている。

今回サンプル文書に使ったのは中華書局の繁体字排印本であったが、台湾などで刊行されているもの、あるいは大陸でも傍線を付していないもの

正	誤	Acrobat	ScanSnap	Word	Google
曰	日		✓	✓	
	𠂔			✓	
實	責		✓		
己	己		✓	✓	
衆	乘			✓	
牝	牝				✓
	×				✓
駭	驟	✓	✓		✓
一	三		✓		
	二			✓	
為	烏		✓		
豈	登	✓			
贅	贊				
讀	譚			✓	
逐	×			✓	
輒	輻				✓
信	佶	✓			
者	×			✓	
否	杏			✓	
末	未				✓
附	勝				✓
泊	泊				✓
軒	阡	✓			

をスキャンしたならば、もう少し成績が良かったのかも知れない。

しかし、古典文献の繁体字排印本の多くが大陸から刊行されており、傍線が付されているものには相対的に良い排印本が多いことを考えれば、やはりサンプル文書のような文献をある程度高い認識率で読んでくれないと、中国学研究において十分に使える検索可能 PDF にはなるまい。

Office Lens と Adobe Scan

千田 大介 (ちだ だいすけ)

❖ はじめに

ここでは、スマホのカメラによる文献スキャンに対応し、それに OCR をかけることのできるアプリとして、Office Lens と、Adobe Scan を取り上げる。

PC の利用を前提とするのであれば、これらのアプリを文献のスキャンのみを使い、PC の Acrobat なり Word なりで OCR をかけたほうが速度・精度ともに高くなるが、出先などでスマホ・タブレット上で執筆する必要がある場合などでは、こうしたアプリが役に立つ。

なお、今回のテストでは裁断済みの雑誌・書籍のページを撮影したため、実際に書籍を撮影するのに比べて、認識率が良くなっている可能性がある。また、両アプリともに iOS 版・Android 版等を提供しているが、今回試用したのは Android 版であり、他の OS 版との機能比較を行っていない



点、ご承知置きいただきたい。

❖ アプリの概要

◇ Office Lens

Office Lens は Microsoft のスマホ用スキャンアプリである。本誌第 16 号「特集 3 文献撮影のためのスマホアプリ & ツール」でも取り上げたが、無償ながら比較的優秀な成績を残している。

Office Lens では、ドキュメントを撮影した後、「Word (OCR 文書)」で保存すると (左下図)、文字認識した Word ファイルに変換される。撮影した画像は、Word ファイル中に配置される。

利用にあたっては、マイクロソフトアカウントが必要になる。アカウント取得は無料なので、持っていない人はサインアップする。

Office Lens では、撮影時に右上の「:」→「解像度」をタップして、写真の解像度を設定できる。筆者が実験した限りでは、解像度が高いほうが OCR 結果も良くなるようだ。以下の認識実験では、筆者のスマホの最高画質である 19.2M を選択した。

認識結果の Word ファイルは、OneDrive に保存される。他のクラウドなどに保存する場合は、Office Lens の撮影画面で右上の「:」をタップ、「自分のファイル」でファイル一覧を開き、ドキュメント右下の「:」をタップ、「共有」で「Dropbox に追加」「ドライブへの保存」などを選択する。

◇ Adobe Scan

Adobe 社によるスキャンアプリで、カメラ撮影したドキュメントを自動レタッチして、画像や PDF として保存することができる。2017 年 6 月



に公開された、比較的新しいアプリである。「自動キャプチャ」機能を備えており、オンしておく、シャッターを押さなくても、文書であると認識すれば自動で撮影してくれる。

OCR 機能を使うには、アプリを起動して左上の「×」をタップし、ファイル一覧画面の左上の「≡」→「環境設定」をタップ、「保存した PDF に対してテキスト認識を実行」オプションをオンにして、「テキスト認識言語」を設定する（上左図）。違う言語のドキュメントを認識するたびに、設定を変える必要があり、何の設定も必要なく OCR してくれる Office Lens よりも手間がかかる。

利用にあたっては、Adobe アカウント（無償）が必要となる。認識結果の PDF は、Adobe 社のクラウドサービスである Document Cloud（2G まで無料）に保存される。また、ドキュメントの下に表示される「共有」をタップして、任意のクラウドサービスに保存することもできる。

❖ 認識結果

◇ サンプル文書①

白抜き

Office Lens は二段組みとして認識できた。画像解像度が低いと、日本語中国語混じりの意味不明なテキストになったが、高解像度でテキスト部分

にフォーカスを合わせて撮影すると、簡体字中国語として認識した。冒頭部分の認識結果は以下の通り。

请客人吃饭，朱先生一家也有自己的家训。一般会点几道菜先吃着，因为一次点完所有的菜需要占用很长一段时间，加上不短道客人喜欢吃什么、有何禁忌，又不好直问，间接通过前面点的菜，观察客人吃得量多量尘继续同一两个问题，这时大家经过聊天，谈话的氛围比较轻松了再做调整比较合适。

一方の Adobe Scan は白抜き文字は認識できなかった。

複雑なレイアウト

Office Lens は段組であることは認識しているが、レイアウトの再現は必ずしも上手くいっていない。上右図は Word の閲覧モードで認識結果を表示したものである。

Adobe Scan はページ全体を三段組として認識した。

サンプル文書②

両者ともに、簡体字横書き一段のレイアウトは問題なく認識する。以下は、Office Lens の認識結果の一部である。

凡例

收词

1. 收词以古代白话小说中出现的近代口语词汇为主, 大致包括如下一些类别: 虚词(如就便、便就、就地、就待、就及、将、将次、将待、将为道等); 带有方言色彩的词语(如清头、交关、弄喧、刺闹、掏嗤、伍浓、一榻括子等);

見出しの「一」は画像として処理された。

Adobe Scan の結果は以下ようになる。各行ごとに改行が入っていたが、比較の便宜上、本文については削除してある(以下同)。

凡例

一收词

1. 收词以古代白话小说中出现的近代口语词汇为主, 大致包括如下一些类别: 虚词(如就便、便就、就地、就待、就及、将、将次、将待、将为道等); 带有方言色彩的词语(如清头、交关、弄喧、刺闹、掏嗤、伍浓、一榻括子等)

両アプリともに、認識結果は良好である。

横書き二段組みの部分であるが、両者ともにレイアウトは認識できた。サンプルとして、初めの二項目を引用する。

〔串捏〕chumniö 捏以适, [例] 母亲所凭者, 正衣饰记色耳諾不凭众认必疑及一 “(野叟·二二) 如、有情, 发得大笑, 两人须吃三杯, (野叟·一二) 〇

〔串〕chudjipidn 串通行骗, [例] 官向他追问外国人的来历, 他只供说是因来买衣服认得, 并且不知他的来历。官便判他一个。“ 押着他追款 〇 廿年, 四九)

数ヶ所、文字認識できなかった箇所が画像で貼りつけられていた。例えば、2行目「必疑及一」は、本来「必疑及、叟·一二二」である。括弧類や句読点の誤認識も目に付く。また、二項目目は本来「串偏」であるが、「偏」が落ちている。

Adobe Scan も、二段組みは正常に認識したが、文字の認識では、漢字をアルファベットや記号に

誤認した箇所が目立つ。

(I 一役) Chu 扣 1116 捏 {aI 编浩. 「例 Jij J, 东所凭甘. J.I: 衣饰记色 Jf, 若个凭众认, 则必疑及~. (野叟·心一. 二) 如~ 打情, 发 1f} 大笑, 两人须吃三杯. (野叟·一二九) [串偏] chun pion 串通行骗. [例] 官向他川间外国人的来历, 他只供说是因来买衣服认 1 小并且不知他的来历. 官便判他一个~ 〇 押着他追款 (廿年·1/Y J L)

両者ともにいささか問題があるが、Office Lens の方が良好であるといえよう。

サンプル文書③

縦書きの認識には対応している。繁体字も認識できているが、傍線の付された箇所は、やはり読めていない。

冒頭部分は、Office Lens は以下ようになった。

蕉軒續録卷一匪直也人

詩鄘風 ' 匪直也人 傳曰· 匪徒庸 訓 匪直爲匪徒 ' 訓人爲 君 '

傍線付きの文字はごく一部しか認識できておらず、字が飛ばされたり誤字になったりしている箇所が目立った。縦書きカギ括弧の認識が上手いかわからないのは、Word 2016 と同様であるが、文字の認識率は Word よりもかなり落ちる。

一方の Adobe Scan は以下のものであった。

蕉軒續録卷一

墳置豈認廩「匪直也人。元傳曰 ' r 匪徒庸 君。J 訓匪直爲匪徒・訓人爲席君

一中略一

524 匪直也人

小見出しは頭の縦中横数字のためか、左欄外の署名の後になっていた。縦書きレイアウトの認識には、いささか問題があるようだ。

❖ おわりに

以上の結果から、単純なレイアウトの文書については、両アプリともに中国語の認識率そのものはそれなりに高いものの、二段組み・ピンイン混じり、あるいは縦書き・傍線付きといった、やや複雑なドキュメントになると、問題が多いことがわかる。

どちらかということになれば、認識率が比較的良く、OCRの設定が不要で撮影画質の設定ができる、Office Lensの方がベターといえよう。Wordファイルとして保存されるので、すぐに編集・再利用ができるのも嬉しい。

Adobe Scan は白抜き文字に対応しておらず、またスキャンアプリとして見た場合も、画像解像度が選択できないなど、全体に今一歩な感がある。新しいアプリなので、今後の機能向上に期待したい。

また、今回の試験を通じて痛感したのは OCR 結果がスキャニング (=撮影) 状態の良し悪しに、かなりの程度、左右されるという点である。つまり OCR の精度は、ある程度、スマホの写真機能、および撮影者のスキルに規定されることになる。その意味で、以上の比較結果は一つの参考にしかかなりえない。読者諸賢には、無償で利用できる両アプリを積極的に活用して、上手な使い方をあみ出して頂きたい。

中国語 OCR スマホアプリ比較

田邊 鉄 (たなべ てつ)

❖ 1. はじめに

ここでは Android スマートフォン用の、中国語に対応した OCR アプリの中から、比較的人気の高いものを Google Play から 3 つ選んで比較した。使用したアプリは次の通り。

- ① Google 翻訳アプリ
- ② OCR (shinozaki gen 氏)
- ③ CamScanner (INTSIG Information)

Google ドライブには OCR 後に Google ドキュメントに出力する機能があり、本来ならこちらを使うべきなのだろうが、スマートフォン用アプリからは使えず、PC 用 Web サイトにアクセスして使うことになるので、ここでは扱わず、代わりに Google 翻訳アプリを取り上げる。

OCR 機能の検証には、本特集の PC ソフトの比較記事 (千田大介氏) で用いている 5 つのサンプルをスマートフォン内に jpg 画像として保存して用いる。

- a 現代文 (飲食についての説明)
- b 現代文 (飲食についての説明)
- c 白話の辞書
- d 白話辞書の凡例
- e 繁体字

これらについて、変換結果を示す。

❖ 2. Google 翻訳アプリ

Google 翻訳アプリは、対応言語間の翻訳機能を提供するアプリだが、その中に、「カメラで撮影したり、端末に保存された画像を読み込んだりして、OCR にかけてテキスト化してから翻訳する」機能を含んでいる。

起動するとテキスト入力画面になるが (図 1)、ここでテキストボックス左下にあるカメラのアイコンをタップすると、カメラモードになる。紙文書を撮影すると、ただちに OCR がかけられる。保存済みの画像を使う場合は、シャッターボタンの左にある「画像へのアクセス」アイコンをタップし、OCR・翻訳の対象としたい画像を選択する。

画像が読み込まれると、自動的に OCR がかけられる。OCR をかけた結果に対して形態素解析¹⁾も行われており、形態素ごとに四角で囲まれて表示される。ユニークなのは、この画面を指でな



図 1

図 2

ぞると、なぞった部分だけがハイライトし、そこだけの OCR の結果と翻訳が表示されることだ(図 1)。もちろん「すべて選択」をタップすれば、そのページ全体をまとめて選択することもできる。

一部または全部を選択してから、翻訳欄の青い右矢印のアイコンをタップすると、見慣れた Google 翻訳の画面で原文と翻訳が表示される。ここから、原文や翻訳をコピーし、他のアプリで利用したり、「共有」メニューからメール送信したりすることができる(図 2)。

読み取りの結果は以下の通りである。

a 現代文 (簡体字)

请客人吃饭, 矢先生一家 (段組) 也有人品
尝到北万的酱香, 就跟厨布 11 自己的夙
il, . . .

広い版面から、文字のある部分を適切に判定できているようだが、段組が無視されているのは致命的である。また、右下隅のキャプションは読み込まれていない。レイアウト認識はちょっと甘すぎる。

また、ゴシック体は明朝体等の縦横で筆画の太さが異なる書体に比べると、読み違いは少なくなることが多いと考えられるが、このアプリでは縦画を落としてしまつての誤認識が特に多い。ネット等で「日本語などに含まれるアルファベットに強い」という評価を目にすることがあるが、それは英数字への誤認識が多いということでもある、と知った。

b 現代文 (簡体字)

如何成为鲁菜达人
莱汰人确实如朱先生无诚,
憍吃, 会原・零而有拉巧。并非随便走进一
家鲁菜的發厂渐能如愿以偿

a とは違い、囲み線があるためか、レイアウトの誤認識はしなかったが、誤字は a 以上に多い。特に画数の多い文字が全滅に近い。

ちなみに同じ部分を Google ドライブの OCR 機能で変換してみると：

如何成为鲁菜达人

鲁菜达人确实如朱先生所说, 懂吃、会点, 里面有技巧。并非随便走进一家鲁菜的餐厅就能如愿以偿。

となり、逗号と顿号の違いや改行位置も含め、完全に再現できている。これがスマホアプリからも簡単に利用できればいいのだが。

c 白話 (辞書)

【申捏】擎杯, 这妖工也以一杯奉上, 三人 --T 猶杯 (西游・七一, 先生, 我与你~一杯。(西游”七三) chuanni ē 吃三杯。(野叟, 一二儿) chuan pan 【穿破】chu à np ò 差误。[例] 果然两个圣笏, 【案骗】串通行骗。

やはり段組が苦手で、中央は線で区切られているのに、レイアウトは認識されていない。【申捏】の項目にあるのは、この語についての説明ではなく、隣の段の説明がここに挿入されている。レイアウトの取得がこれほど奔放だと、ちょっと使いづらいだろう。

d 白話 (凡例)

凡例收词 1, 收词以古代白话小说中出现的近代口语词汇为主, 大致包括如下一些类别: 虚词 (如就便, 便就, 就地, 就待, 就及, 将, 将次, 将待, 将为道等); 带有方言色彩的语词 (期清头, 交关, 弄喧, 刺闹, 掏唾, 伍浓, 一榻括子等);

見出し行の「一」がとれていないが、単純な 1 段のレイアウトで、程度もよいため、誤字もほとんどない。約物の判別が今ひとつ。

e 縦書き繁体字

52 蕉軒續錄卷 14 匪直也人蒔鄙風愷之方中律 “[匪直也人。”毛傳曰 TH 徒磨。1 船直扁匪徒, 訓人爲膚考幹傳曰: “非獨此人所以操其心者誠實而淵深, 蓋其所蓄之馬七尺而牝者, 亦已至於三千之衆。”又答劉平父喜 1 日.. 『『匪直也人』云云, 非特人化其德而有塞淵之美, 至於物被其功 m 至衆多之盛。]

巻一の「一」がとれなかったこと(!でとっている)

を除けば、まずまずの認識率と言えようか。524 という番号が 52 と 4 に分かれて認識されたのは、52 を割注とみたのだろうか。

❖ 3. OCR

shinozaki gen 氏が公開しているシェアウェア。

iOS や Android 用の OCR ソフトとしては、日本語認識の精度の高さで注目されているが、中国語はどうだろうか。

a 現代文 (簡体字)

请客人吃饭,朱先生一家也有(段組)人品
尝到北方的酱香,就跟厨师协
自己的家训。

残念ながら段組はこのアプリでも認識されず、一行につながってしまっている。文字の間違いはほとんどなし。

b 現代文 (簡体字)

如何成为鲁菜达人
鲁菜达人确实如朱先生所说,"懂吃、会点,
里面有技巧并非随便走进一家鲁菜的餐厅就
能如愿以偿。

コラム部分含め、レイアウトの間違ひもなく、ほぼ完璧に読み取れていたが、サンプルにない“記号が余分に入っている。「说」につられて、



図3 OCR

ということだろうか。

c 白話 (辞書)

【串捏】

chuan nie

捏造; 编造。

[例] 母亲所凭!

擎杯,这妖王也以一杯奉上,二人~了酒杯。
者,止衣饰记色耳,若不凭众认,则必疑及~!
(野叟·一二二)如~有情,发得大笑,两
人须|

吃三杯。(野叟-一九)

(西游,七一)先生,我与你~一杯。

やはり、段組は認識されていない。中央の縦線は、!や|として認識している。ただ中央に「何らかの区切りがあること」はわかっている、そこで改行が入っているのも、Googleのように不規則に行が繋がっているよりは、「復元作業」は多少楽になる。

d 白話 (凡例)

凡例

收词

收词以古代白话小说中出现的近代口语词汇为主,大致包括如下

一些类别:虚词(如就便、便就、就地、就待、就及、将、将次、将待、将为道等);带有方言色彩的语词(如清头、交关、弄喧、刺闹、掏嗤、伍浓、一榻括子等);结构固定的俗语词(如半瓶醋、打绝板、拔短梯、窝里炮、睡长觉等);

見出し行の「一」がとれていないが、それ以外は誤字・脱字なく認識している。

e 縦書き繁体字

蕉軒續録卷!

sed 韭直也人

詩廊風定之方中承-T 噩宜也人° 1 毛傳曰
-r 熙徒賡君 01 訓匪胤爲匪徒,訓 4% 膺册
4° 朱子

韓籒曰“非獨此人所以操其心者誠實而淵深,蓋淇所蓄之馬七尺術牝者,亦已至於

三千之衆。」

縦書きの約物に引っかかっているところを見ると、「(繁体字) 中国語」の縦書きが特に苦手なようだ。日本語の縦書きは、句読点などもごく正確に認識している。誤字も簡体字とは比べものにならないほど多い。

OCR アプリの中でトップクラスを誇るだけあって、シンプルな操作画面や横書きの普通の印刷物に対する対応の確かさは非常に優れている。今後の展開も期待したい。

❖ 4. CamScanner

スマホによる「スキャン」という発想を早い時期に具現化した、この分野のパイオニア的存在。

a 現代文 (簡体字)

请客人吃饭, 朱先生一家也有 (段組) 人品
尝到北方的酱香, 就跟厨师协自己的家训。

やはり、二段組は認識されなかった。スマホアプリでは段組レイアウトは想定していないように見える。

b 現代文 (繁体字)

如何成为鲁菜达人 鲁菜达人确实如朱先生
所说, 「董吃、会点, 里面有技巧。并非随

便走进一家鲁菜的餐厅就能如愿以偿。

上述の OCR 同様、「说」に引きずられたのか、カギ括弧が入っている。スマホの OCR アプリが絵と文字を単純に対応づけているだけではない、ということを感じさせる。認識率自体は高く、コラム部分も含め、ほぼ間違いなく読み取っている。
c 白話 (辞書)

【申捏】 chuonn 匠捏造编造。【例】母亲所凭 (段組) 擎杯, 这妖王也以一杯奉上, 二人~了酒杯。(段組) 者, 止衣饰记色耳, 若不凭众认, 则必疑及~。(西游·七一) 先生, 我与你~一杯。(西游·七(野叟·一二二) 如~有情, 发得大笑, 两人须三)

やはり段組は認識されておらず、行が途中で切れてしまっている。文頭が「者,」で始まっていることに疑問を持つ AI は、少なくともスマホにはまだない。

d 白話 (凡例)

凡例 一收词 1. 收词以古代白话小说中出现的近代口语词汇为主, 大致包括如下一些类别虚词 (如就便、便就、就地、就待、就及、将、将次、将待、将为道等);

項目番号や句読点なども含めて、ほぼ完全に再現されている。

e 縦書き繁体字

(文字を認識できず)

残念ながら、このサンプルでは全く文字を認識できず、数文字の記号 (~~~) を吐き出しただけだった。簡体字中国語と繁体字中国語の選択は可能で、簡体字を OFF にして、繁体字だけの読み取りに設定しても、結局文字は取得できなかった。

注

[1] 本号『入門・中国語形態素解析』も参照されたい。



図 4 CamScanner

IIIF の解説と動向 —— 画像データの国際的な相互利用促進を目指す枠組み

西岡 千文 (にしおか ちふみ)

❖ 1. はじめに

貴重資料等を公開するデジタルアーカイブにおける画像公開の国際的な規格として、近年 IIIF (International Image Interoperability Framework) ^[1] の採用が世界で広がっている。本稿は、それらのデジタルアーカイブで資料を扱う研究者と IIIF を活用した画像データの発信を検討している研究者・開発者を対象に、IIIF に関する基礎的な知識ならびに最近の動向を概説することを目的とする。第 2 章では、IIIF の歩みについて、IIIF の背景・目的とともに述べる。第 3 章では、IIIF が提供する API について、IIIF Image API と IIIF Presentation API を中心に紹介する。第 4 章では、IIIF に関連するソフトウェアとして、代表的な画像ビューワを挙げ、第 5 章をむすびとする。

❖ 2. IIIF の歩み

過去 20 年以上にわたり、図書館・美術館・博物館等で所蔵する資料のデジタル化・公開が進展した。ユーザは Web 上でそれらの画像を自由に閲覧することが可能となった。しかし、下記が依然として課題として指摘されてきた ^[2]。

- ユーザは、異なる機関から画像を使用しようとするたびに、それぞれのデジタルアーカイブの操作方法を習得する必要がある。結果、複数機関の資料を扱う研究者の学術的ワークフローは非効率的であった。
- コンテンツ提供機関は、ハードウェア、ソフトウェア、システム管理等に多大なコストを費やしてきた。類似した機能を有するシステムが、個別の機関で実装されており、非効率的であった。

IIIF は、画像やそれらに付随するデータの配信に関

する規格として API を定める。データの配信方法を API によって統一することで、機関のデータはサイロから解放される (図 2)。公開した画像をあらゆるソフトウェアで利用することが可能となり、様々な機関の画像を同一ソフトウェアで利用することが可能となる。API に対応したソフトウェアを IIIF コミュニティで開発・共有していくことで、開発に要するコストが削減される。また、画像の相互利用、様々な分野での利活用が促進される。より詳細な IIIF の背景、設立の経緯等については、文献 ^[3] を参照されたい。

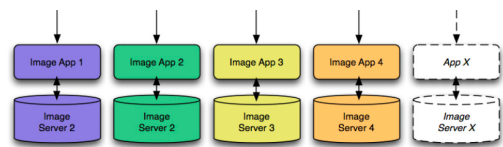


図 1 IIIF 誕生以前のデジタルアーカイブの様子である ^[4]。各機関が、個別に画像サーバ、ビューワ等のソフトウェアを実装していたため、画像データはサイロに閉じ込められ、画像の利用は特定のソフトウェアのみに限られていた。また、類似した機能をもつソフトウェアの実装が重複しており、非効率的であった。

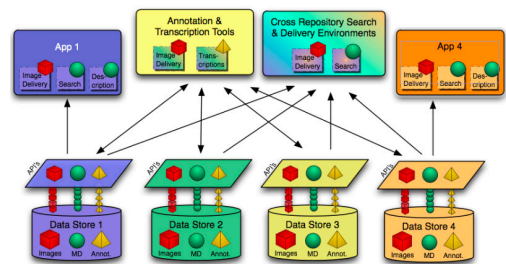


図 2 IIIF が定める共通の API によって、様々な機関のソフトウェアから画像データを利用することが可能となる ^[5]。よって、画像の相互利用、利活用が促進される。

IIIF の持続可能な発展・運営のために、2015 年 6 月に IIIF コンソーシアムが設立された。コンソーシアムメンバーの機関は、年会費 10,000 米ドルを支払い、IIIF の発展・運営に貢献する。2018 年 8 月現在、欧米の図書館・美術館・博物館を中心に 51 機関がコンソーシアムメンバーとなっている ^[6]。日本からは東京大

学が2016年にコンソーシアムメンバとなり、その後、京都大学図書館機構、関西大学、国立情報学研究所が加盟している。コンソーシアムメンバではなくとも、IIIFが提供するAPIは自由に利用でき、技術面に関するディスカッションへ参画することは可能である。

IIIFの導入状況についてであるが、欧米の図書館・美術館・博物館を中心に、全世界に広がっている。2017年5月に公表された調査結果^[7]では、IIIFに準拠して公開されている画像は3億枚以上であることが報告されている。日本の機関での導入も急速に広がっている。2018年4月から2018年8月現在まで、九大コレクション(九州大学附属図書館)、小城藩日記データベース(佐賀大学)、近畿大学貴重資料デジタルアーカイブ、国立国会図書館デジタルコレクションがIIIFを採用した^[8]。日本におけるIIIFの導入状況については、サイト^[9]にまとめられている。2018年5月に開催されたIIIF Conference 2018にて、IIIFサイトへの国別アクセス数において、日本は、米国、英国について3位であると報告された^[10]。このことから、日本におけるIIIFへの関心の高さを伺うことができ、今後の採用の広がりが期待される。

❖ 3. IIIF が提供する API

IIIFは、画像をサイロから解放し、相互利用を促進するための規格としてAPIを定めている。2018年8月現在、IIIFからは下記のAPIが提供されている。

- IIIF Image API^[11]

画像の配信方法を定める。

- IIIF Presentation API^[12]

資料の構成、メタデータなど、ビューワに資料を表示するために必要な情報の記述方法を定める。

- IIIF Search API^[13]

資料に付与された翻刻等のアノテーションの検索方法について定める。検索対象は、基本的に個別の資料であり、全資料のアノテーションを対象とした検索には利用できない。

- IIIF Authentication API^[14]

画像の利用を特定のユーザに制限したい場合に、認証プロセスの導入方法を定める。

本章では、上記のAPIのうち、多くのIIIF準拠デジタルアーカイブで利用されているIIIF Image APIとIIIF Presentation APIについて紹介する。

■ 3.1. IIIF Image API

IIIF Image APIは下記のURI構文によって画像を呼び出す^[15]。

```
{scheme}://{server}/{prefix}/
{identifier}/{region}/{size}/
{rotation}/{quality}.{format}
```

URI構文では、呼び出す画像の領域(region)、サイズ(size)、角度(rotation)等を指定する。図3に、IIIF Image APIによる画像の呼び出し例を示す。図3上段では、呼び出す画像の領域(region)を指定している。図3下段の左から2番目の画像は、呼び出す画像のサイズとして、横幅を90ピクセルに指定している。図3下段の左から3番目と4番目の画像についてであるが、角度(rotation)が「!345」と指定されている。これは、時計回り方向に345度回転させ、反転させた(角度の数値の前に「!」を記述)画像を呼び出す。よって、左から3番目の画像では画像が反転され、4番目の画像では345度時計回り方向に画像が回転している。最後に、左から5番目の画像では、画像がグレースケールになっている。これは、画像のqualityをグレースケール(gray)に指定したためである。他に、フルカラー(color)、二値(白黒)(bitonal)、デフォルト(default)を値として指定できる。多くの画像サーバで、デフォルトはフルカラーに設定されている。

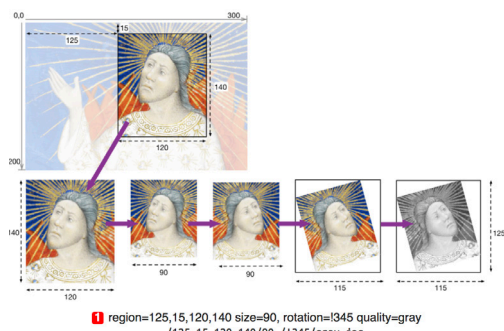


図3 IIIF Image APIによる画像の操作例^[16]

IIIF対応ビューワ(後述)では、ユーザは画像に対して拡大、縮小、回転など様々な操作をする。ユーザがビューワで画像を操作するたびに、ビューワはユーザが要する画像の領域・角度等の値を取得して、IIIF

Image API で画像を呼び出している。

IIIF Image APIによる画像の配信へ対応するためには、IIIF へ対応した画像サーバに画像を保存する必要がある。多くの画像サーバがフリーソフトウェアとして公開されている。IIIF では高解像度の画像を効率的に配信するために、サイズが大きい画像については分割(タイル画像化)して必要なタイルのみを配信するという仕組みを採用している。タイルサイズの設定等、画像配信における留意点については、文献^[17]を参照されたい。

画像のサイズ、タイル等の情報は、下記の URI 構文によって取得する^[18]。

```
{scheme}://{server}/{prefix}/
{identifier}/info.json
```

画像を効率よく呼び出すために、タイル等の情報をあらかじめ取得して、サイズ等を検討することが可能である。

■ 3.2. IIIF Presentation API

IIIF Presentation API は、資料の構成などの情報の記述形式を定義する。IIIF Presentation API では、一つの資料がマニフェストという JSON オブジェクトに相当する。各マニフェストは、マニフェストファイルとして JSON 形式のファイルに保存される。資料のマニフェストには、それらをビューワで表示するために必要となる以下の情報^[19]が含まれる。

- ファイルで表現されている JSON オブジェクトがマニフェストあることを示す情報
- 資料全体に関する情報 (資料名、権利情報等)
- 資料の構成に関する情報

資料全体に関する情報^[20]であるが、通常は label というフィールドに資料タイトルが記述される。また、資料の帰属を示すフィールドである attribution、ライセンスを示すフィールドである license が用意されている。その他のメタデータについては、metadata というリストで、任意のラベルと値のペアを使用することで表現する。IIIF Presentation API はその名称が示す通り、「表示」のための規格である。よって、メタデータのスキーマについては、各機関に任されている。機械可読であるメタデータは別途ファイルに保存し、seeAlso

というフィールドにその URI を記述することで、提供される。seeAlso で公開されているメタデータは、統合検索サイトなどからハーベストされ、インデキシングなどに利用されることが期待される。しかし、現在のところ、seeAlso で情報を提供する機関は一部に限られており、形式も統一されていない。統合検索サイトなどでの本格的な活用のためには、コミュニティでの合意の形成が必要だろう。他に、マニフェストではページ方向を指定できる。東洋の資料は、西欧のものとは異なり、ページ方向が右から左へのものが多い。ビューワで右から左へのページめくりを実現するために、viewingDirection というフィールド^[21]が用意されており、right-to-left という値をフィールドに与えることで設定できる。Universal Viewer (後述) などの主要なビューワは、viewingDirection に対応している。

続いて、資料の構成に関する情報について述べる。マニフェストにおいて資料の各頁はカンバスという要素に対応する。各カンバスは、カンバスに描画する画像・注釈(翻刻・翻訳等)についての情報を含む。マニフェストは、カンバスを順番に並べたシーケンスを含むことで、資料の頁の順序を表現している。IIIF Presentation API のデータ表現によって、同頁に属していたが散逸してしまった資料を、デジタル空間上で再統合することが可能である^[22]。また、カンバスに画像を描画しないことにより、欠損した頁を表現することが可能である。

その他、IIIF Presentation API についての詳細は、記事 [23,24,25,26] を参照されたい。

ビューワは、マニフェストファイルを読み込むことで、資料を表示する。マニフェストファイルを公開することで、あらゆる IIIF 対応ビューワで資料を開覧することが可能となる。多くの IIIF 対応デジタルアーカイブでは、IIIF マニフェストアイコン (図4) によって、マニフェストファイルへのリンクが提供されている。Mirador (後述) 等のビューワでは、このアイコンをドラッグアンドドロップすることで、資料を読み込む。

以上は、現行の IIIF Presentation API 2.1.1 をもとにした解説である。以下、最新の動向として、公開予定である IIIF Presentation API 3.0 について述べる。IIIF Presentation API 3.0 については、現在 α 版が公開されている^[27]。なお、IIIF Presentation API 3.0 にあわせて、IIIF Image API 3.0 も公開される。IIIF Presentation API

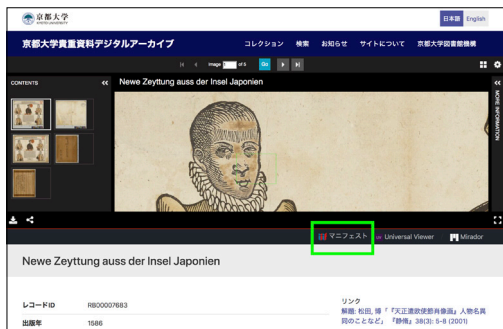


図 4 IIF マニフェストアイコンの表示例（枠内）(天正徳政使節肖像画・京都大学所蔵)

3.0 は、動画・音声といった時間軸をもつデータの公開へ対応する。また、W3C の Web Annotation Model や JSON-LD 1.1 の公開に応じて、更新がなされている。大きな変更点として、例えば AnnotationPage の導入が挙げられる。IIF Presentation API 2.1.1 では、キャンバスに画像をアノテーションとして直接貼り付けることで、当該頁に表示する画像データを表現していた。対して、IIF Presentation API 3.0 では、キャンバスに AnnotationPage というオブジェクトをアノテーションとして貼り付け、AnnotationPage に画像を他の注釈とともに貼り付けることが必要となる。このような変更点によって、現行の IIF Presentation API 2.2.1 との互換性が失われる^[28]。画像の相互利用のための互換性が、対応している API のバージョンによって失われないように、ビューワ等のソフトウェアは、現行のバージョンと後続のバージョンどちらにも対応する必要がある。ソフトウェアの開発状況や、他機関での対応の状況などを注視し、対応を検討する必要がある。

❖ 4. IIF に関連するソフトウェア

IIF に関連するソフトウェアとして、画像ビューワを中心に述べる。なお、IIF に関連するソフトウェアのリストが、Awesome International Image Interoperability Framework (IIF)^[29] として提供されている。

■ 4.1. 画像ビューワ

IIF コミュニティにおいて代表的な画像ビューワである、Universal Viewer と Mirador について短く紹介する。どちらのビューワも右から左へのページめくりに対応している。さらに、IIF の機能を活用した特徴的

なビューワとして、IIF Curation Viewer を取り上げる。

(1) Universal Viewer^[30]

Universal Viewer の特長として、3D 画像、音声、動画、PDF など、様々なメディアデータの表示に対応している点が挙げられる。また、画像のダウンロード機能も実装されている。

(2) Mirador^[31]

Mirador は、スタンフォード大学、ハーバード大学を中心に開発が進められている。Universal Viewer と比較すると、Mirador は研究者を対象とした機能が充実している。代表的な機能として、画像の比較、注釈の付与が挙げられる。画像の比較方法について述べる。Mirador 右上に位置するアイコンを操作することで、スクリーンを分割する。分割して生成されたスクリーンに比較対象の資料の IIF マニフェストアイコン（前述）をドラッグアンドドロップすることで、その資料が読み込まれ、資料の比較が可能となる。また、Mirador は注釈機能を提供しており、画像の任意の箇所に翻刻等の注釈を与えることが可能である。デフォルトの状態では、注釈はブラウザに保存される。ユーザが与えた注釈を保存したい場合は、別途アノテーションサーバの導入といった作業が必要となる。代表的なアノテーションサーバとして、Simple Annotation Server^[32] が挙げられる。

(3) IIF Curation Viewer^[33]

人文学オープンデータ共同利用センター (CODH) によって提供されている IIF Curation Viewer は、複数の資料から任意の頁や領域を抽出して、任意の順序で閲覧する機能の特徴として有している。Mirador と同様に、IIF マニフェストアイコンのドラッグアンドドロップにも対応している。IIF Curation Viewer を使用し、資料のキュレーションを行うことで、資料へ付加価値が与えられる。

❖ 5. おわりに

本稿では、国際的な画像の相互利用促進の枠組みである IIF について紹介した。IIF は欧米の図書館・美術館・博物館を中心とした世界中の機関で採用が広がっている。IIF が定める規格として、IIF Image API と IIF Presentation API を中心に述べた。これらの API が多くの機関で受け入れられている理由として、IIF が、W3C が定める Web Annotation 等の標準に沿って

発展していることが挙げられる。しかし、IIIF のコミュニティは、欧米の機関が中心であり、欧米以外の機関の関与は少ない。文化的多様性を考慮した IIIF の今後の発展のためにも、国内機関からの積極的な関与が期待される。

注

[1] <http://iiif.io/> (最終確認日: 2018 年 8 月 31 日)

[2] Snyderman, Stuart, Robert Sanderson, and Tom Cramer. "The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images." Archiving Conference. Vol. 2015. No. 1. Society for Imaging Science and Technology, 2015.

[3] 永崎研宣. "デジタル文化資料の国際化に向けて: IIIF と TEL." 情報の科学と技術. 67 巻 2 号, pp. 61-66, 2017.

[4] 画像は注 2 の文献より引用した。

[5] 画像は注 2 の文献より引用した。

[6] <https://iiif.io/community/consortium/> (最終確認日: 2018 年 8 月 31 日)

[7] <https://iiif.io/news/2017/05/25/newsletter> (最終確認日: 2018 年 8 月 31 日)

[8] カレント・アウェアネスポータルにて「IIIF」と検索し、検索結果から今年度 IIIF の導入がなされた機関・サイトを集計した。 http://current.ndl.go.jp/search/apachesolr_search/iiif

[9] historyanddigital. "日本の図書館等における IIIF 対応 デジタルアーカイブ一覧." NAVER まとめ. <https://matome.naver.jp/odai/2152584366126558001> (最終確認日: 2018 年 8 月 31 日)

[10] IIIF. "Welcome and state of the IIIF Universe." IIIF Conference 2018. <https://docs.google.com/presentation/d/1xaZt8DApy5RCGuxByAU7d29fwZpX9NEtmlZzmnwuu9M/edit?usp=sharing> (最終確認日: 2018 年 8 月 31 日)

[11] <https://iiif.io/api/image/> (最終確認日: 2018 年 8 月 31 日)

[12] <https://iiif.io/api/presentation/> (最終確認日: 2018 年 8 月 31 日)

[13] <https://iiif.io/api/search/> (最終確認日: 2018 年 8 月 31 日)

[14] <https://iiif.io/api/auth/> (最終確認日: 2018 年 8 月 31 日)

[15] <https://iiif.io/api/image/2.1/#image-request-parameters> (最終確認日: 2018 年 8 月 31 日)

[16] 画像は下記より引用した。 <https://iiif.io/api/>

[image/2.1/#order-of-implementation](https://iiif.io/api/image/2.1/#order-of-implementation) (最終確認日: 2018 年 8 月 31 日)

[17] 永崎研宣. "国際的な画像共有の枠組み IIIF の課題と展望." デジタルアーカイブ学会誌. 2 巻 2 号, pp. 111-114, 2018.

[18] <https://iiif.io/api/image/2.1/#image-information> (最終確認日: 2018 年 8 月 31 日)

[19] 神崎正英. "画像共有の新しい標準 IIIF." LOD Diary. <http://www.infocom.co.jp/das/loddiary/2017/01/20170127001583.html> (最終確認日: 2018 年 8 月 31 日)

[20] <http://iiif.io/api/presentation/2.0/#presentation-resource-properties> (最終確認日: 2018 年 8 月 31 日)

[21] <https://iiif.io/api/presentation/2.1/#viewingdirection> (最終確認日: 2018 年 8 月 31 日)

[22] 永崎研宣. "バラバラになった各地の西洋中世写本断片をまとめて表示: IIIF の活用例." digitalnagasaki のブログ. <http://digitalnagasaki.hatenablog.com/entry/2017/07/03/030708> (最終確認日: 2018 年 8 月 31 日)

[23] 神崎正英. "リンクの機能を柔軟に生かすデータのウェブ." 情報の科学と技術. 67 巻 12 号, pp. 622-627, 2017.

[24] 注 19 と同

[25] 神崎正英. "IIIF マニフェストと注釈." LOD Diary. <http://www.infocom.co.jp/das/loddiary/2017/04/20170410001609.html> (最終確認日: 2018 年 8 月 31 日)

[26] 神崎正英. "一歩進んだ IIIF マニフェストの利用." LOD Diary. <http://www.infocom.co.jp/das/loddiary/2017/03/20170314001601.html> (最終確認日: 2018 年 8 月 31 日)

[27] <https://iiif.io/api/presentation/3.0/> (最終確認日: 2018 年 8 月 31 日)

[28] <https://iiif.io/api/presentation/3.0/change-log/#1-breaking-changes> (最終確認日: 2018 年 8 月 31 日)

[29] <https://github.com/IIIF/awesome-iiif> (最終確認日: 2018 年 8 月 31 日)

[30] <https://universalviewer.io/> (最終確認日: 2018 年 8 月 31 日)

[31] <http://projectmirador.org/> (最終確認日: 2018 年 8 月 31 日)

[32] <https://github.com/glenrobson/SimpleAnnotationServer> (最終確認日: 2018 年 8 月 31 日)

[33] <http://codh.rois.ac.jp/software/iiif-curation-viewer/> (最終確認日: 2018 年 8 月 31 日)

万方数拠

小島 浩之 (こじま ひろゆき)

❖ 万方数拠とは

万方数拠 (Wangfang Data) とは、万方数拠社が運営するオンラインデータベースで、中国の学術雑誌の電子版 (オンラインジャーナル) のほか、国家標準 (国家規格)・中医薬 (漢方)・中国企業の製品情報など様々な情報を有料で提供するデータベースの総称である。



● 万方数拠 <http://g.wanfangdata.com.hk/>

万方数拠については、既に本誌 15 号の特集「東洋学論文指南」において、千田大介氏のレビュー (「中国の論文を探す」) がある。この時点では国内代理店が存在しなかった。現在、文生書院が国内代理店となって提供しているので、国内での利用も容易になった。日本などアジア地域への提供サーバーは香港に置かれているという。

ただし、JJ15 当時と比べて収録コンテンツやインターフェイスはほとんど変わっていないので、本稿では改めて言及はしない。具体的な使用方法などは千田氏のレビューを参照されたい。

なお、文生書院 Web サイトトップページ (<https://www.bunsei.co.jp/>) 下部の「取扱商品」メニューにある「データベース」をクリックすると、文生書院の取り扱っているデータベース一覧が表示される。ここで「Wanfang Data」を選択すると商品情報を確認することができる。

❖ 利用料金

万方数拠は先行する CNKI 同様に、大学や研究機関とのライセンス契約と、個人用の課金システムの両方が完備されている。

個人利用の場合、800 ポイントから購入でき、ポイントは 2 年間有効である。800 ポイントは \$1,600 に相当し、為替レートに従って円建てで文生書院に支払うことになる。サイトトップページの課金欄 (充值卡) からは課金できず、文生書院に直接申し込む必要があるので注意されたい。

現在のところ初年度に限っては 400 ポイント (\$800 分) から購入できるキャンペーンを実施している。また、文生書院では 2018 年末までの期間限定でダウンロードサービスのモニターを募集しており、登録すると無料で 100 ポイント (23,000 円相当、有効期限 2018 年 12 月 31 日) が進呈される。ただし、モニターは最終的に 400 字程度の感想を文生書院に提出する義務を負う。

文生書院によれば、学術雑誌と学会報告は 1 頁あたり約 73.3 円、学位論文は 1 頁あたり約 22 円だという。CNKI の学術雑誌が 1 頁あたり 140.4 円、学位論文が 1 頁あたり約 42.1 円であるから、万方数拠は CNKI の約半額の価格設定になっているといえる。

ただし、通常の最低ポイント購入価格が \$1,600 であるから、一回の購入金額は消費税を入れると 20 万円近くなる。また円安となれば価格も上昇する。つまり個人契約可能といっても、事実上、大学の研究費を使える研究者しか対象としていないということなのであろう。

一方の CNKI は 20 ポイント 7,020 円からプリペイドカードが購入できるから、単価は倍といえ個人顧客を広く対象とする戦略であることがわかる。

❖ 学術論文のコンテンツによる CNKI との比較

学術論文の収録数を見てみると 2018 年 10 月 10 日現在で、万方数拠は学術論文という名称のコンテン

ツに 83,847,613 篇が収録されている。これに対して CNKI は期刊というコンテンツに 104,108,679 篇が収録されており、CNKI の方が約 2 割増である。

この差は検索結果にも表れており、同じ条件でキーワード検索すると CNKI の方が、万方数拠に比べてヒット数が 2 割から 3 割多い。ただし、歴史系でざっと検索する限り、近年の主要なものは万方数拠でもほぼ網羅されているので、数量の差はさほど気にならない。ただ、CNKI の PDF が最近の論文は全てフォント埋め込みデータとなっているものが多いのに対し、万方数拠の PDF は画像データに OCR を掛けたものが多かった。この点は、雑誌版元と両社の契約の関係によるのであろうか。また古い論文に関しては、1993 年以前のバックナンバーをコンテンツとしても

つ CNKI と比べると、どうしても見劣りしてしまう。

一方で、検索結果が出るまでの時間や、PDF のダウンロードにかかる時間は、圧倒的に万方数拠が早くスムーズである。これは CNKI が北京と青島のサーバーであるのに対して、万方数拠が香港のサーバーを使っているという差かもしれない。

いずれにせよ、これまで CNKI しか選択肢のなかった日本において、第二の選択肢ができたことは歓迎すべきことであろう。寡占状態は利用者にとって必ずしも好ましいこととはいえないので、複数選択肢を維持できるよう、代理店の文生書院にはぜひ頑張っていたきたいものである。

Android スマホ用中国語辞書

田邊 鉄 (たなべ てつ)

❖ 0. はじめに

本稿は、Android スマホ・タブレット等で利用できる、中国語辞書のレビューである。中国語のフルコンテンツを搭載したIC電子辞書から、メーカーが次々と撤退する中、あらためて注目されているのが、スマートフォン用の辞書アプリとオンライン辞書である。

iOSでは中国語アプリが比較的充実していたが、Androidでは従来、フルコンテンツの中日辞典は、選択肢がほとんど小学館『中日日中辞典』ビッググローブ版だけしかなく、特に中級以上の学習者には満足できない状況にあった。「〇〇詞典全テキスト」のような海賊版を除けば、今も紙の辞書を電子版で出す、という趣旨のアプリには、魅力的な製品が見当たらない。

ここでは、Androidで使える中日・日中辞典として、スタンドアロン・アプリのもと、オンライン辞書(Web辞書)のうちAndroid向けに専用検索アプリを提供しているものを取り上げ、使い勝手やコンテンツの良否を検証する。いずれもPlayストアからダウンロード、インストールすることができる。記事の内容は2018年9月現在のものである。

❖ 1. 小学館『中日日中辞典』ビッググローブ版

まず、スタンドアロン・アプリから。

残念ながら「第2版」である。したがって、新語・流行語の類には弱い。

「スライム」や「サーバル」はいない。iOSではビッググローブ版とは別に「物書堂」が第3版を電子化して出しているの、Androidの方でもバージョンアップを期待したいところだが、まだ音沙汰ない。

インターフェースは非常にシンプルで、テキストボックスが1個あるきりで、ここに日本語・中国語・ピンインを入れて検索する。中日日中辞典という名前だが、中日辞典と日中辞典は使い分けの必要がなく、シームレスに検索される。

ピンインを入力する際には、一声～四声のボタンか

ら声調をつけることもできる。また、アルファベットの略号(NHK、ASEANなど)は、収録されているものについては、そのまま小文字で入力しても、ピンインと区別され、ちゃんと中国語訳が表示される。その際、JASのように、ピンインにはない綴りは「JA」と打ったところでJASだけが候補になるが、「JIS」のようにピンイン綴りでもあり得る場合「JIS」と打った上で、候補の中から探す必要があり、スクロールしなければならない場合もあるので、多少注意は必要だ。

いずれにせよ、入力する際に文字種や言語を気にしなくてもよく、事前に中日・日中すら選ばなくてもいいという辞書引き体験は快感である。辞書データを全てスマホ内に保持しているので、検索スピードは速く、安定している。ただし、日本語で検索した場合は、見出し語ズバリでないヒットしない。例文中の語はその日本語が含まれていてもヒットしないことがある。

ごく一部の見出し語には、音声がついている。TTSではなく、女声の録音である。ただ、ごく基本的な語である「学」「学生」「学習」にはいずれも音声がついておらず、もう少し増やして欲しいとおもう。

中日日中辞典はPlayストアで3900円。紙の辞書だと中日・日中の2冊で16,000円以上(第3版)する。中日辞典1冊でも、アプリの倍である。大学の初修科



図1 小学館『中日日中辞典』

目で使うだけなら、このアプリ1本で十分だろう（授業中にスマホで辞書をひけるかどうかは教員の判断によるが）。

2. all 中国語辞書 中国語⇄日本語

続いてオンライン（Web）辞書を紹介する。オンラインでは Weblio が比較的良好に使われている。翻訳と辞書の両方が使えること、白水社の『中国語辞典』がオンラインで使えること、などがメリットと感じられているようだ。スマホ用には専用アプリが用意され、同じサイトにある翻訳なども同時に利用できるようになっている。ただし辞書の中身としては、例文がほとんどないなど、アプリ版や紙版には及ばないが、多くのサイトで「複数のサイトからの結果をまとめて表示する」機能を搭載しはじめてから、じわじわと利用が増えている。

ここで紹介するオンライン辞書のラッパーアプリである、all 中国語辞書について特筆すべきことは、対応辞書の多さだ。

- Weblio
- NAVER
- Yahoo
- 北辞郎
- 沪江
- 百科
- 愛詞霸
- Excite
- Goo
- コトバンク
- 有道
- 词典网
- 汉字字典
- BitEx

- Eduu
- Google 画像検索
- Google Books
- ウィキペディア
- Google Web 検索
- Google ニュース検索
- ウィクショナリー

Google の各サービス、およびウィクショナリー・ウィキペディアは中国語版が開く。

「中国語」辞書ではないものも含まれているが、中国語版の画像検索など、授業準備などで便利に使えそうだ。

繁体字への対応は、利用している外部サイトの仕様による。ざっと試してみたところでは繁体字対応は一部のとどまっているようだ。

アプリを起動すると真っ黒な画面が表示される。筆者のスマホはモトローラの Moto G5 で設定によって Android 標準の「▽ ○ □」を表示させず、画面下の物理ボタンを左右にスワイプすることで機能を代替できるようになっている。だが、本アプリでは、物理ボタンの機能は生きているが、「▽ ○ □」は消えない。

画面の一番上には二つのテキストボックスがあり、左が中日、右が日中になっている。日本語を入れたときと中国語を入れたときで、使用する辞書が違うので、はっきり分けているのだろう。

どちらかのボックスに語を入力し、仮想キーボードに表示される虫眼鏡のキーをタップすると全ての辞書で検索をかける。画面の下部にあるタブで辞書を選ぶ

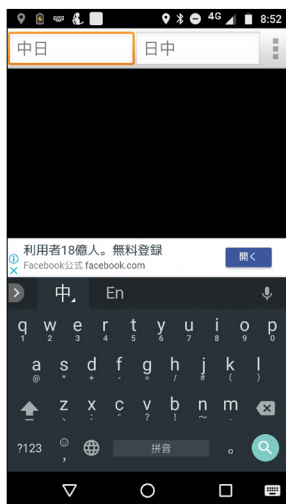


図2 all 中国語辞書



図3 検索表示画面

と、その辞書で検索した結果が表示される。

使用する辞書を追加・削除することはできない。画面右上の設定アイコンをタップし、「言語」を選ぶと中日・日中それぞれに使われている辞書の一覧が表示される。辞書名の横にある↑をタップすることで、辞書の順番は変えることができる。中日、日中それぞれ、よく使う辞書が上にくるようにしておくといいたいだろう。

有料オプションはなく、全ての機能を無償で使えるのはいいが、広告のキャンセルができない。できれば広告が表示されない有料版を作ってほしい。

❖ 3. LINE Dictionary 中国語 - 英語

「あの」LINE である。ただし、LINE と連携するとか、LINE アカウントがないと使えない、ということではない。Collins 中英辞典の 300 万語、15 万例文を収めたフルコンテンツの辞書で、単なる辞書だけではなく、中国語学習者向けのコンテンツを多数収録していることが特徴である。

辞書は一つのテキストボックスに中国語または英語を入れると、訳の候補が表示される。テキストボックスの下に表示される中国語・英語の表示の間にある緑の矢印をクリックすると、中文英訳と英文中訳を切り替えられる。

オンライン辞書は基本的にオフラインでは使用できないが、LINE Dictionary ではオフラインでも訳語だけは表示される。

繁体字にはほとんど対応していないようだ。

たとえば「台湾」と「臺灣」はどちらも収録されているが、別の語として、異なる内容が表示される。また繁体字の「學」では検索されないことがあるなど、収録されていても簡体字同様に使えるわけではない。

辞書の検索履歴は、「単語カード」に記録することができ、後から見返したり、SNS でシェアしたりすることができる。

例文・語の一部は音声に対応している。男声・女声がバラバラで、複数の声の主がいることから、音声は複数のリソースから採集しているのではないかと思われる。

また、漢字の一部は書き順のアニメも提供している。「学生」のような二文字の語も、アニメの下に出てくる語の表示で、それぞれの字をタップすると書き順を個別に参照できる。「今日の中国語」は、1日1フレーズを学習できる会話教材である。1フレーズだけでなく、そのフレーズが含まれた4行~6行の簡単な対話文も音声付きで用意されている。たとえば「このツアーは全部で何人ですか」という疑問文が、この日のキー・フレーズならば：

A 这次旅游团一共多少人？

B 出了地陪，一共十五人。

といった対話が表示される。

「文章解析器」は、「もしや形態素解析ができるのか」



図4 LINE Dictionary

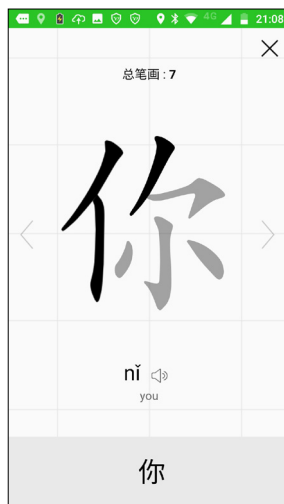


図5 書き順アニメ



図6 文章解析器

と思い、試しに短文を解析にかけてみた。だが、結果は形態素解析というよりは、大学等での初級授業で用いる教科書の分かち書き程度である。チャンク、と言えなくもないだろう。

❖ 4. おわりに——これからの辞書

一時期、クラスの半分近くが所有していた IC 辞書だが、今年ほどのクラスも 1 人か 2 人しか見かけなくなっている。

1979 年発売のシャープの電訳機や、90 年代の 8 セ

ンチ CD を使った“電子ブック”に始まり、その後 IC 電子辞書へと進化した、電子辞書専用端末の命脈は、ほとんど尽きようとしているのかもしれない。百科事典や青空文庫の小説を入れたり、音声ファイルのようなマルチメディア情報を入れたりして延命をはかっているが、スマートフォン用の電子辞書がこれだけ普及してしまっただけでは、IC 辞書＋スマホ、のような持ち方はとても非効率に見えるのだろう。

もちろんまだオンライン辞書は発展途上のコンテンツなので、今後の展開に期待したい。

執筆者紹介

小島 浩之 (こじま ひろゆき)

1971年岐阜県生まれ。東京大学大学院経済学研究科講師・経済学部資料室室長代理。専門は東洋史学および歴史資料の保存と活用に関する研究。最近是中国古文学書の理論体系の構築に精力を傾けている。

中村 覚 (なかむら さとる)

2017年東京大学大学院新領域創成科学研究科博士後期課程修了。博士(環境学)。同年東京大学情報基盤センター学術情報研究部門助教。現在に至る。Linked Dataをはじめとする情報技術を用いたデジタルアーカイブの構築と活用に関する研究や、オントロジーを用いた知識処理に関する研究を行っている。情報処理学会、人工知能学会、Alliance of Digital Humanities、日本船舶海洋工学会などの会員。

佐藤 仁史 (さとう よしふみ)

1971年生まれ。一橋大学大学院社会学研究科教授。専門は中国近現代史。著書に、『近代中国の郷土意識——清末民初江南の在地指導層と地域社会』研文出版、2013年、『嘉定県事——14至20世紀初江南地域社会史研究』(共著)広東人民出版社、2014年、『垂虹問俗——田野中的近現代江南社会与文化』(共著)広東人民出版社、2018年、がある。

二階堂 善弘 (にかいどう よしひろ)

1962年東京の下町生まれ。東洋大学文学部卒業、早稲田大学大学院文学研究科博士課程退学。現職は関西大学文学部教授。博士(文学)・博士(文化交渉学)。専門は中国の民間信仰研究。最新の著作は『元帥神研究』(齊魯書社)。サイトは、<http://www2.itc.kansai-u.ac.jp/~nikaido/>

田邊 鉄 (たなべ てつ)

1963年京都府生まれ。大阪外国語大学大学院外国語学研究科修士課程修了。北海道大学情報基盤センター准教授。CALL授業を自動学習から「能動学習」へと変えるためにジタバタ中。目下のテーマはパンダモチーフ。

西岡 千文 (にしおか ちふみ)

京都大学附属図書館研究開発室助教。慶應義塾大学理工学部卒業、慶應義塾大学大学院理工学研究科博士前期課程修了、キール大学大学院工学研究科博士後期課程修了を経て現職(博士後期課程時の研究拠点は、ドイツ国立経済学図書館)。博士(工学)。専門は情報工学。学術情報のマイニング、セマンティックウェブに関する研究に携わる。

千田 大介 (ちだ だいすけ)

1968年東京都生まれ。早稲田大学大学院文学研究科中国文学専攻博士課程中退。慶應義塾大学経済学部教授。専門は、ここ1,000年くらいの中国のサブカル。成人病の百貨店になりつつある。Webサイト: 電脳瓦崗寨 <http://wagang.econ.hc.keio.ac.jp/>

氷野 善寛 (ひの よしひろ)

1980年大阪生まれ。目白大学外国語学部中国語学科専任講師。N高等学校中国語講座の教材制作およびカリキュラム設計を担当。e-Learning教材として『NetAcademy2 中国語コース』(アルク教育社)、共著として『キクタン中国語』シリーズ(アルク)、『官話指南の書誌的研究』(好文出版)。

師 茂樹 (もろ しげき)

1972年生まれ。花園大学教授。博士(文化交渉学)。『論理と歴史—東アジア仏教論理学の形成と展開』(ナカニシヤ出版、2015年)、『『大乘五蘊論』を読む』(春秋社、2015年)など。Twitter ID: @moroshigeki

安岡 孝一 (やすおか こういち)

1965年大阪府生まれ。京都大学大型計算機センター助手、京都大学人文科学研究所附属漢字情報研究センター助教授などを経て、現在、京都大学人文科学研究所附属東アジア人文情報学研究センター教授。京都大学博士(工学)。専門は人文情報学。著書に『文字符号の歴史 欧米と日本編』(共立出版)、『キーボード配列 QWERTY の謎』(NTT 出版)、『新しい常用漢字と人名用漢字 漢字制限の歴史』(三省堂)などがある。

山崎 直樹 (やまざき なおき)

1962年生まれ。関西大学外国語学部/外国語教育学研究科教授。専門は中国語教育(学習課程の設計、授業設計)/中国語学(談話の構造)。目下関心があるのは言語的少数者のための情報保障、言語教育のユニバーサル・デザイン化、言語教育におけるインクルージョン。<http://www2.itc.kansai-u.ac.jp/~ymzknk/>
Twitter ID: @ymzknk

編集後記

『漢字文献情報処理研究』第18号をお届けする。前号が本年1月刊だったので、1年以内に2冊という、ハイペースの刊行になってしまった。このためレビュー記事数は少なくなっているが、その分、論考・特集は、何時にも増して充実していると思う。

2018年夏、『大漢和デジタル版』発売のニュースが流れたが、本文が画像であるということで、ネットにおける反応は概して芳しくなかった。具体的な機能については発売後に検証することとして、ここでは10万円という販売価格について考えてみたい。

語彙索引・補巻込みの15冊セットの定価は24万円、古書は「日本の古本屋」で検索すると、語彙索引・補巻を除いた13冊セットで15,000～30,000円程度であるが、『大漢和』の市場価格はもっと安い。ヤフオクに3,000円で出品された『大漢和』が流れた、と話題になったのは2016年の暮れだったと思うが、現在でも、落札価格は2,000～3,000円ほど、4,000円を超える出品の大半が流れる状況が続いている。

つまり、ドキュメントスキャナさえあれば、3,000円ほどで『大漢和』を自炊してPDF版を作れることになる。ネットで配布されている親書・熟語一覧を使って掲載ページを検索し、PDFで閲覧することもできる。このような3,000円自炊版に対して、10万円の『大漢和デジタル版』は、果たして価格差に見合ったベネフィットを提供できるだろうか？

本号の編集は、あいもかわらず、突貫工事の泥縄作業であったため、遺漏も多いものと思われるが、ご寛恕いただけたら幸いである。末筆ではあるが、本誌の刊行に有形無形のご援助を賜った漢情研会員諸賢、校正作業にご協力頂いた小山萌さん・倉脇麻友さん、そしてなにより、引き続き本誌の出版をお引き受けくださった好文出版の尾方社長に、衷心より御礼申し上げます。(△)

漢字文献情報処理研究 第18号

発行日	2018年10月29日	
定価	本体2,000円＋税	
編集	©漢字文献情報処理研究会 http://jaet.sakura.ne.jp/	
編集委員	○千田 大介 小島 浩之 田邊 鉄 師 茂樹	上地 宏一 佐藤 仁史 二階堂善弘
デザイン & DTP	電脳瓦崗寨： http://wagang.econ.hc.keio.ac.jp/	
発行人	尾方敏裕	
発行所	株式会社好文出版 〒162-0041 東京都新宿区早稲田鶴巻町540 林ビル3F TEL:03-5273-2739 FAX:03-5273-2740 URL: http://www.kohbun.co.jp/	

◎本誌に関する訂正・補足情報は、漢字文献情報処理研究会サイト (<http://jaet.sakura.ne.jp/>) に掲載します。

※本誌および本誌掲載論文・記事は、日本学術振興会科学研究費補助金「情報時代における中国学研究・教育オープンプラットフォームの構築」(平成28年度～30年度、基盤研究(B)、研究代表者:二階堂善弘、課題番号:16H03351)による成果の一部である。

ISBN978-4-87220-216-8

C3004 ¥2000E



9784872202168



1923004020005