

民事判決原本データベース構築の歩み

新田 一郎・高久 俊子

はじめに

明治初年以来昭和18年末までの民事判決原本は、全国10の国立大学における一時保管を経て、現在、国立公文書館への移管が進行中である。この間に、一時保管先の大学関係者を中心とする「民事判決原本データベース作成協議会」において立案されたデータベース構築作業が、1997年度から2006年度まで10年間にわたり、文部省科学研究費補助金「研究成果公開促進費」（開始時の研究代表者は石井紫郎国際日本文化研究センター教授、のち石井教授の退官に伴い合庭惇同センター教授に交代）の交付を受けて、国際日本文化研究センター（以下、「日文研」）で行われた。

データベース作成作業の基本的方向は以下の通りである。

明治23年末まで（民事訴訟法施行前）の分を、10年計画でDB化する。

検索（書誌）DBは、当事者名簿等、当面必要と思われる諸項目について入力するものとし、将来における項目増設の余地を作っておく。

原本は、スキャナにより原色画像で入力する。ただし、必要があればプライバシー保護のための字句のマスキングを行なう。

検索DBとテキスト画像DBを連動させ、オンラインで必要なデータを引き出せるように構築する。

利用資格は、プライバシー保護の観点から原則として研究者に限る。

以上の方針のもとでデータベース作成の対象となった簿冊の数量と年度ごとの進行状況の概略については、別表1「数字からみた民事判決原本DB作成作業」を参照されたい。以下本稿では、第一章に最終的に確定された手順の概略を示し、第二章に作業の中で生じた問題点とその克服の過程を記す。

新田一郎（にった・いちろう）：民事判決原本データベース作成協議会委員・東京大学大学院法学政治学研究科教授

高久俊子（たかく・としこ）：民事判決原本データベース作成プロジェクト員・国際日本文化研究センター情報課

第一章 データベース作成作業の手順 (図1 「作業手順」 参照)

1. 原本借り受け

日文研から原本保管大学に出向き、対象となる原本¹を調査票とともに借り受ける。1冊ずつ簿冊の状態をチェックし、冊数を確認する。原本は特別トラック便で輸送する。

2. 受け入れ作業

借り受けた原本は、日文研で用意した大型文書箱(中性紙箱)に移し替える。その際に簿冊毎に中性紙ティッシュで包み、簿冊番号札をつける。次に、簿冊の解綴前写真を撮影する。撮影箇所は表・裏・背・小口・天・地・斜めの計7箇所。その後、簿冊毎に保管票を作成する。表題、綴じ方、絵図の有無、虫損・破損等の有無、解体・養生の要不要等、受け入れ時の原本の状態を詳細に記録する。

3. 解綴

原本の多くは和紙に毛筆で書かれているが、綴じ方は各裁判所、年代によって、四つ目綴じから厚表紙の製本まで様々である。各簿冊の解綴にあたっては、「解綴記録表」に簿冊表紙の特徴、綴じ方、綴じ穴の位置を記入し、紐をはずしながら、1丁ずつ解綴し、すべての丁に鉛筆で丁数番号を記入する。丁数番号の記載位置は、復元したときに見えないようにするために原則として裏向けの右端、綴じ穴と綴じ穴の間とする。同時に件分け(第二章参照)をしながら、以後の作業の基本となる「画像入力記録表」を作成する。なお、この作業の過程で、修復すべき丁や大型絵図を選びだし、さらに、後に必要と思われる種々の情報を記録として残す。差し込み・はさみ込み²等も、元通りに復元する際に必要な情報であるためなるべく詳しく記録する。絵図等の複雑な折りたたみ方は、別紙に展開図を作成することもある。

4. 解体・養生・修復

頑丈な製本、複雑に作られた表紙、酸性劣化のためポロポロと崩れる背表紙等、解綴が困難で作業員の手に残ると判断された簿冊は専門業者に解体・養生を依頼する。簿冊にはしばしば虫損が見られた他、水に濡れてフケ状になったものや、焼けこげているものなどもあり、その状態は様々であった。このうち、虫損、破損等の程度が著しく、スキャナにかけることができない原本、あるいは復元に支障をきたす原本については専門業者に修復を依頼する。修復は、漉嵌(すきばめ=欠損部分に和紙の繊維を補填する方法)修復を基本にしたが、使用されている紙や、対象原本の状態によって、裏打ち、繕い等の修復方法も採用した。ただし、虫損による固

¹ 明治初年から明治23年までの原本が対象だが、同簿冊内に対象年度外の事件が含まれていることもある。その場合は全てデータベース化した。

² 袋とじになっている丁の中に差し込まれた本紙及び紙片を「差し込み」、原本と原本の間にはさみ込まれた本紙及び紙片を「はさみ込み」と名付けた。

着・劣化のひどい簿冊の一部については、費用等の関係で修復作業ができずに解綴・データ採録を断念したものもある（別表1の「特記」欄を参照）。

5. 撮影委託する絵図

原本には、多くの絵図が付随していた。争論地の境界線を示す絵図がその主なもので、美しく彩色された絵図も多く存在した。大きさも様々で、半紙サイズのものから、縦1メートル、横2メートルを超える大型絵図もあった。

スキャナのA3サイズをはるかに超える大型絵図は、業者に撮影を委託する。業者は、絵図中の最小の文字サイズから適切な方法を選択し、デジタルカメラあるいは4×5インチカラーポジフィルムで撮影する。各々の絵図には簿冊番号、丁数番号、スケール、カラーチャートを写し込んで撮影する。

6. 画像入力

原本は、貼り紙、付箋から紙片に至るまですべてスキャンし、カラー画像として取り込む。オリジナルの高精細画像（約2200×1560pixels、200～400KB）と簡略画像（約770×550pixels、約100KB）を作成し、ともにweb上で見るようにする。

7. 文字入力

画像に合わせて、件ごとに主要な事項をテキスト入力し、検索できるようにする。この作業は後に詳述するように10年間を通して大きな変遷をたどった。入力するデータ項目は、表1「入力項目の推移」のように推移してきた。

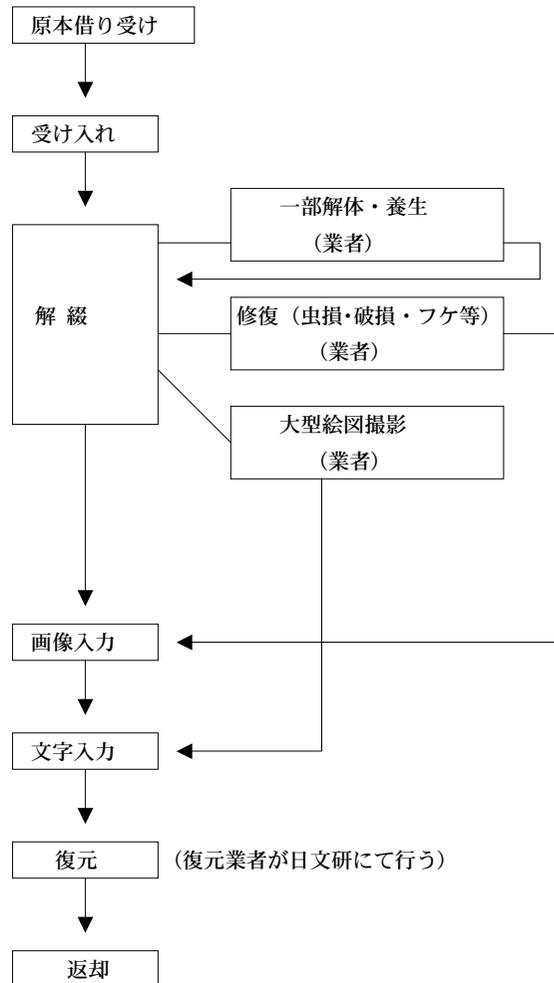


図1 作業手順

表1 入力項目の推移

[1997年]	[1998 - 99年]	[2000年以降]
1 簿冊番号	1 簿冊番号	1 簿冊番号
2 簿冊内番号	2 簿冊内番号	2 簿冊内番号
3 事件番号	3 表題	3 関連文書代表番号
4 事件名	4 事件番号	4 表題
5 原告 1/2/3/4	5 事件名	5 事件番号
6 被告 1/2/3/4	6 原告 1/2/3/4	6 事件名
7 裁判年月日	7 原告代理人	7 原告
8 裁判所名	8 原告側訴訟関係人	8 原告代理人
<補正入力した項目>	9 被告 1/2/3/4	9 被告
・ 原告代理人	10 被告代理人	10 被告代理人
・ 被告代理人	11 被告側訴訟関係人	11 訴訟関係人
・ 訴訟関係人	12 裁判年月日	12 裁判官
・ 裁判官	13 裁判所名	13 裁判年月日
	14 備考	14 判決裁判所
	15 裁判官	15 備考

8. 復元

簿冊を復元する作業は専門の業者に委託する。原本の性格上、持ち出しを禁じ、日文研所内で行う。業者は画像入力記録表に記載された情報をもとに復元していく。元の綴じ穴を用い、元の紐と同じように綴じ、製本し、受け入れ時の状態にするのが基本である。綴じ紐については、原則として使用可能なものはそのまま使用する。しかし、原本は幾度かの再編綴を経て多数の綴じ穴を持つものが多く、紐を一度に引き抜いてしまうと肝心の本綴じの穴や、差し込み・はさみ込みの情報がわからなくなるため、大部分の紐は解綴の際切ってしまう。そのため、復元の際には、染色や漂白、蠟引き等の加工が施されていない新しい麻紐を使用する。

9. 返却

すべての作業を終えた原本は、一簿冊ごとに中性紙ティッシュで包み、借り受け時の箱に入れる。返却するものは簿冊、調査票であるが、簿冊に使用されていた紐・紙縫・中に入っていた洋紙・ポストイット付箋紙などは、中性紙封筒に入れて返却する。原本は借り受けの時と同様特別トラッ



原本

ク便で輸送する。

第二章 作業上の問題点とその克服

次に、作業の中で生じた問題点とその克服の過程について記す。

1996年、データベース作成協議会の発足に先立ち、1名の非常勤職員による事前調査が行われた。大阪大学から4冊の原本を借り受け、画像入力、文字データ入力に関する具体的な方針を決定した。1997年8月に、科研費の正式決定を受け、パソコン10台、スキャナ8台、作業員16名によってプロジェクトが開始された。初年度の作業は解綴と画像入力为中心で、文字データ入力は業者に委託した。

1. 件分けの問題

事前調査時に扱った原本は、明確に判決文とわかる文書（これを「判断文書」と呼ぶことにした）、あるいは判決文に口供などが附属しているものが多かった。このため、当初の方針は、判決文ごとに一件のデータとし、同じ事件でも控訴審は別の一件として扱う、口供などの関係文書は「附属文書」として判決文に従属させ、併せて一件とする、とされた。当時の裁判制度についての専門的知識を持たない大勢の作業員が質の揃った作業を行うため、基本的に、文書表題から一件のデータの区切りを判断するという具体的・即物的な作業方法を採用することになった。

ところが、いざ初年度の東京地方裁判所旧蔵簿冊を開けてみると、言渡書（判決文）とともに、命令書、訴状、願書、口供、証、請書、和解調書、絵図、勧解不調、送達状、鑑定命令書など、多様な関係文書類がひとまとまりに括られているケースが多く見出された。事前調査時の予想を超える多様な文書群の中から「判断文書」と「附属文書」とを弁別するために、具体的な基準を次々に整備する必要があり、この件分け作業に大きな労力と時間が費やされることになった。

ところで、作業現場で答えの出せない作業上の疑問や、不明な点の解決は、データベース作成協議会の委員との質問・回答の往復によりなされた。このような「質問 回答」のシステムがあったからこそ、最終的にはかなり充実した内容のデータベースを作り上げることができたのではないかと思われる。次々出現する新しい表題に悪戦苦闘しながらも、「質問」を繰り返しながら一步一步前に進んだ。

2. 文字入力、業者委託から日文研作業員で入力へ

初年度の文字入力は業者に委託し、入力項目も「簿冊番号」、「簿冊内番号」、「事件番号」、「事件名」、「原告1/2/3/4」、「被告1/2/3/4」、「裁判年月日」、「裁判所名」とわずか8項目だった。指示内容も簡単なものだったため、第二水準を超える文字も入力され、その結果web上では文字化けする障害も出現した。また、こちらの意図がしばしば正確に伝わらず、データ修正に二重の時間を割かざるを得なかった。こうした難点を回避するため、翌1998年からは日文研作業員が入力

することになった。

これを機に、「表題」、「原告代理人」、「被告代理人」、「訴訟関係人」の項目も追加することになった。また、1999年度からは「裁判官」も入力することにした。これで、原本に登場する主要な人物は、ほぼ不足なく採ることができるようになった。

3. データ採取の困難

日文研作業員で文字入力を始めてみると、データ採取の種々の困難が明らかになった。初期の判決文の書式は一定ではなく、原告記載の範囲の中に原告情報、代理人情報がそれぞれの住所、寄留住所・身分・職業も含めて区切りもなく続けて書かれていることも多い。この中から原告の情報、代理人の情報を間違いなく採取しなければならない。また、事件名は、おおむね本文冒頭にある「貸金催促ノ訴訟」などの記載のほか、本文中に「身代限」とか「糶売(せりうり)」、「競売」などの形で命令内容を記載した部分から抽出しなければならない場合がある。当時の執行裁判は、必ずしも現在のように本案判決と判然と区別され独自の表題が付されているわけではなく、混然と記載されていることが多いために起こる難しさであった。しかも、それらが多様なくずし字で書かれているため難しさはさらに増した。

4. 件分けの根本的な改変

こうして作業が進むに連れて「判断文書に附属文書を附属させて一件のデータとしていく方法」には、限界があることが明らかになった。

例えば、「判断文書(原審)」と「判断文書(控訴審)」の後に、「附属文書(原審の口供)」が綴じられていることがある。この場合、事件の流れをあまり細かく区切りたくないという理由から、原審と控訴審は別であることを承知で、「判断文書(控訴審)」+「附属文書(原審の口供)」で一件とせざるを得なかった。また、「裏書案」や「願書(袖判決)」³など、「判断文書」とも「附属文書」とも簡単には割り切れない中間的な性格を持つ文書も多く出てきた。

このような実情を解決するために提案されたことは「件分けを根本的に変えること」だった。具体的には、「附属文書」は「雑文書」と呼び方を変えた上で、「判断文書」と並ぶ独立した一件として数えることとし、すべての文書は「判断文書」「雑文書」「不全文書(判断文書とも雑文書とも判断できない文書)」のいずれかに分類することにした。この結果、それぞれの文書はバラバラに扱われることになるが、検索画面で「関連文書」を相互に表示することによって、事件の流れに沿った文書のまとまりを維持することとした。この根本的な改変は、2000年3月に日文

³ 「裏書案」とは、当事者から提出された文書に裁判所の判断を示す文言を裏書きする案文を指し、いい、「願書(袖判決)」とは「願書」の袖に裁判所の判断を示す一筆が書き加えられているものを指している。現在は、「袖判決」を「裁判袖書」と名付けている。

研で開催されたデータベース作成協議会のシンポジウムで提案され承認された。

5. 共同体訴訟、寄留、身分の情報を追加

2000年度以降も、入力内容の修正・拡充は続いた。まず、地域に深くかかわって「集団の共通の権利を主張する訴訟」は、共同体訴訟として住所も入力することにした。当事者として個人の記載のみあるが、事件名を見ると、入会訴訟、漁場争い、水利権争い等、地域が争論の内容に深く係わる訴訟の代表であることも多かった。この場合には、個人名の前に記載されている住所も必要な範囲で採録することにした。また、本来の戸籍地を離れて居住している「寄留」に関する情報についても、必要との声があり、採録することにした。さらに、身分の情報も、従来の「華族」に加え「士族」も入力することにした。

以上のように、このデータベースは、10年間に文字入力マニュアルの大きな改変・修正を行いながら、内容を充実させてきた。これに対応して、項目の追加に関してはほぼすべてを遡って補正入力した。しかし、共同体訴訟情報・寄留情報・身分情報等は、前半3年間分について遡及入力されないままになっている。

6. ふたつの検索システム

民事判決原本データベースは、1999年までは、独自に開発された検索システムを用いていた。しかし、前項のような状況下、入力マニュアルやデータ項目の見直しに即時に対応することが必要だったにもかかわらず、この独自の検索システムではプログラム修正は容易ではなかった。そこで、2000年度入力分から、日文研の共通検索システム（博物館収蔵品管理システム<Musetheque>）を利用することにした。これにより、項目の変更や追加が容易にできるようになった。

しかしこの結果、1997-99年作成の古い独自の検索システム「区分1」と、2000年度以降の共通検索システム「区分2」とが並存することになり、「区分1」のデータを「区分2」へ移行し、データベースを統合することが、次の課題になった。

7. 「区分1」と「区分2」の統合

統合作業の前段階として、文字化けや検索結果不表示などの不備がある「区分1」のデータを整備する必要があった。具体的には、第二水準を超える文字を直し⁴、裁判年月日を8桁に統一した。

次に、「区分1」のデータを「区分2」へ移行させ統合した。この移行の際に足りない項目を補正入力したが、それに伴い、画像の壊れ、撮り漏れも多数発見された。すでに原本は返却されていたが、各保管大学に撮影を依頼するなどしてデータを整えることができた。移行作業は2003年6月から始めて、2005年秋までかかった。

⁴ 第二水準を超える漢字は、web上で文字化けするため、直せるものは常用・第一水準・第二水準のいずれかの漢字に直し、できないものは で表示した。

おわりに

このようにして、10年の長きにわたったプロジェクトは、最終的に54万件（約130万ファイル）を超えるデータを持つ壮大なデータベースとして結実した。「区分2」に統合されたデータベースは、2006年3月に新しいサーバに移され、以前に較べて格段に検索のスピードが速くなり、検索の画面も一新されている。

なお、関係者の努力で、課題となっていた初期3年間入力分データの補正作業が、2007年6月からスタートすることとなった。計画通りに整備が進めば、民事判決原本データベースは、研究者にとってさらに重要な史料となるに違いない。

*本稿は、データベース構築作業の一応の完結を機に高久がまとめた記録を基とし、新田が補訂を施したうえで、最終的には両者の協議によって成ったものである。内容についての責任は両者が均しく負担する。

別表1 数字からみた民事判決原本DB作成作業

年度	大学	借受時冊数	作業確定冊数(表紙のみ含む)	劣化により解綴せず返却簿冊(表紙のみデータ採取)	復元簿冊数	修復冊数	修復丁数	委託絵図丁数	特記
1997	東京大学	253	253	3	250	40	不明		内3冊は虫損劣化のため、解綴せず返却。表紙のみデータ採取
1998	北海道大学	58	56	0	56	2	123		2冊は明治24年以降のため着手せず返却
	東北大学	472	472	0	472	9	226		コピー簿冊(9冊)は日文研で再編綴
1999	東北大学	281	281	0	281	0	0		
	熊本大学	194	194	0	194	26	1610		
2000	東京大学	467	466	0	466	20	317	2714	1冊は明治24年以降のため着手せず返却
2001	東京大学	463	463	2	461	32	925	54	2冊は虫損劣化のため、解綴せず返却。表紙のみデータ採取
2002	東京大学	126	89	0	89	29	384	7	2冊は明治24年以降のため着手せず返却。3冊は修復のみ。他の作業は次年度へ。虫損固着簿冊32冊は2004年度以降へ
	香川大学	264	264	0	264	11	281	96	
2003	大阪大学	535	157	0	157	26	408	22	リスト漏れ簿冊(但2002年借受け) 東大固着簿冊修復は2004年度分(11冊分)
	岡山大学	121	121	0	121	1	1		
	東北大学	91	91	0	91	0	0		
	東京大学		3	0	3	11	571		
2004	大阪大学		378	9	369	28	970	69	9冊は虫損劣化のため、解綴せず返却。表紙のみデータ採取(2003年借受)
	東京大学		11	0	11	0	0		(2002年借受)
2005	九州大学	280	161	0	161	25	554	144	但2004年借受 (2002年借受)
	名古屋大学	186	186	0	186	4	128		
	広島大学	291	24	0	24	0	0		
	東京大学		5	0	5	5	498		
2006	広島大学		267	0	267	0	0	53	(2005年借受)
	九州大学		119	0	119	0	0		(2005年借受)
	東京大学	13	13	0	13	0	0		リスト漏れ簿冊
	東京大学		16	7	9	9	306+		内7冊は虫損劣化のため、解綴せず返却。表紙のみデータ採取(2002年借受)
	大阪大学	21	21	0	21	0	0		リスト漏れ簿冊
計		4116	4111	21	4090	278		3159	