

CA1899 レファレンス協同データベースの登録事例から垣間見る日本のレファレンスサービス

さとう しょう
佐藤 翔*
よしだ みつお
吉田光男†

はじめに

レファレンス協同データベース（レファ協）は国立国会図書館（NDL）が全国の図書館等と協同で構築する、調べ物のためのデータベースである。参加する各図書館等におけるレファレンス事例等の調査内容を登録・蓄積し、インターネットを通じて提供することで、図書館等におけるレファレンスサービスはもちろん、一般利用者の調査研究活動をも支援することをその目的としている⁽¹⁾。

2005年4月の本格運用開始から10年以上を経て、レファ協は参加機関が700館以上、登録事例が18万件以上、一般公開事例に絞っても10万件を超える、大規模なデータベースとなっている。レファ協に関する研究も複数行われており⁽²⁾⁽³⁾⁽⁴⁾、中でも谷本は、登録データはもちろんのこと、参加館に対する質問紙調査等も実施し、事例登録をめぐる問題点や障害に関する丹念な報告を行っている⁽⁵⁾。一方で、レファ協を通じて公開されているレファレンス事例の中身についての詳細な検討は、必ずしも行われてはいない。谷本の指摘のとおり、多くの図書館は未だ事例を登録していないとは言え、公開されている事例だけでも10万件にのぼるレファ協のデータは、日本の図書館のレファレンスの現況を知る材料としても貴重なものになっていると考えられる。そこで本稿では、レファ協で公開されているデータの内容を分析することで、日本のレファレンスサービスに寄せられる質問とそれに対する回答の傾向を垣間見ることが試みたい。

レファ協には参加機関で行われた質問回答サービスの記録であるレファレンス事例に加え、調べ方マニュアル（特定のテーマ等に関する情報源の調べ方）、特別コレクション（参加機関が所蔵する個人文庫などの特殊コレクションの情報）、参加館プロフィール（参加機関の情報）の計4種類の情報が登録されている。このうち本稿で分析の対象とするのはレファレンス事例の情報である。分析対象データはレファ協のAPI 2.0⁽⁶⁾を用いて取得した、一般に公開されている事例とする。データ取得は2017年2月中旬に実施した。取得時点での参加館のみまたは自館のみが参照できる事例を除く

一般公開事例9万6,504件を分析した⁽⁷⁾。分析に使用したデータ及びAPIから最新のデータを生成するプログラムについてはZenodoにて公開している⁽⁸⁾。

1. 事例登録年、登録図書館

図1は登録年ごとの事例数を示したものである。館種別等の詳細は前述の谷本による調査⁽⁹⁾が詳しいのでここでは割愛するが、基本的に登録データは2010年以降、年間8,000から9,000件程度で推移している。2011、2012年の登録数が突出して多いのは、滋賀県立図書館等、この時期にまとめて過去分までデータ登録を行った図書館があるためである。

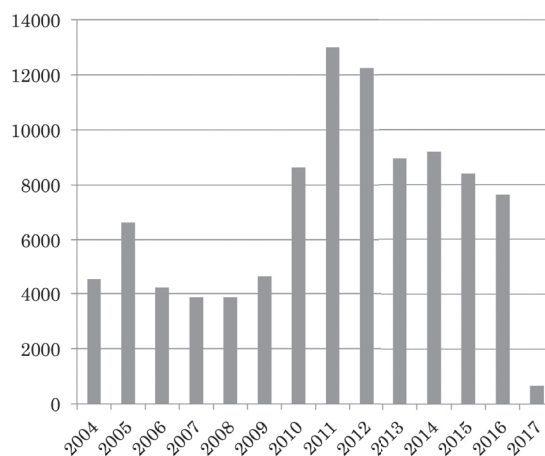


図1 登録年ごとの事例数 (縦軸は件数、横軸は登録年。2017年は2月中旬までの値)

表1は公開されている登録事例数が多い図書館を上位10位まで示したものである。NDLが最も多いのは当然として、2位の埼玉県立久喜図書館を筆頭に、都府県立図書館が10館中7館を占める。大学図書館では近畿大学、市町村立図書館では豊中市（大阪府）の公開事例数が最多である。また、表には含まれないが、専門図書館では日本貿易振興機構アジア経済研究所図書館（1,016件登録）が最多である。

表1 登録事例数上位10機関

| 図書館名 | 登録事例数 (件) | 公開事例中の割合 (%) |
|-----------|-----------|--------------|
| 国立国会図書館 | 14,145 | 14.7 |
| 埼玉県立久喜図書館 | 7,994 | 8.3 |
| 近畿大学中央図書館 | 6,954 | 7.2 |
| 滋賀県立図書館 | 4,569 | 4.7 |
| 香川県立図書館 | 3,736 | 3.9 |
| 岡山県立図書館 | 2,565 | 2.7 |
| 大阪府立中央図書館 | 1,675 | 1.7 |
| 岐阜県立図書館 | 1,626 | 1.7 |
| 豊中市立図書館 | 1,576 | 1.6 |
| 東京都立中央図書館 | 1,428 | 1.5 |

*同志社大学免許資格課程センター
†豊橋技術科学大学情報・知能工学系

2. 分類・キーワード

レファ協では主題による検索を実現するため、レファレンス事例にその事例に該当する日本十進分類法(NDC)の分類記号を任意で付与することができる(最大3つまで)。本稿で対象とする9万6,504件のうち、なんらかの分類記号が付与されていたのは6万7,936件(約70%)であった。

表2は特に付与件数の多い分類記号上位10位を、表3は0類(総記)から9類(文学)までそれぞれの登録件数と割合を示したものである。レファ協においては付与するNDCは第三次区分(要目)までとすることとされており、4桁以上の記号を付与している例はなかった。表2に示した第三次区分の水準だと、最も多い「210 日本史」を筆頭に、日本の歴史・地理に関わる分類や、日本の文学に関わる分類が付与されている事例が多いことがわかる。一方で、表3に示した通り、第一次区分(科目)の水準で見ると、最も多いのが2類(歴史)に関する事例であることは第三次区分の上位と同様であるが、次に多いのは3類(社会科学)であり、2類(歴史)にひけをとらない事例数が存在する。3類(社会科学)については特定の要目に集中せず、社会に関する様々な質問が寄せられているものと考えられる。逆に事例数が少ないのは0類(総記)、1類(哲学)、8類(言語)などで、特に言語に関する事例は全事例数の3%未満にとどまっている。

表2 登録事例数上位の分類記号(10位まで)

| 分類記号 | 項目名 | 登録事例数(件) |
|------|--------------------------|----------|
| 210 | 日本史 | 6,423 |
| 289 | 個人伝記 | 2,706 |
| 291 | 日本(地理・地誌・紀行) | 2,592 |
| 911 | 詩歌(日本文学) | 2,103 |
| 913 | 小説・物語(日本文学) | 1,273 |
| 910 | 日本文学 | 1,137 |
| 090 | 貴重書・郷土資料・ その他特別コレクション | 1,129 |
| 213 | 関東地方 | 1,078 |
| 767 | 声楽 | 1,035 |
| 386 | 年中行事・祭礼 | 935 |

表3 NDC 第一次区分(科目)ごとの事例登録数

| 科目 | 登録事例数(件) | 公開事例中の割合(%) |
|---------|----------|-------------|
| 0類 総記 | 5,671 | 5.9 |
| 1類 哲学 | 4,276 | 4.4 |
| 2類 歴史 | 20,051 | 20.8 |
| 3類 社会科学 | 18,096 | 18.8 |
| 4類 自然科学 | 6,800 | 7.0 |
| 5類 技術 | 7,601 | 7.9 |
| 6類 産業 | 6,223 | 6.4 |
| 7類 芸術 | 9,176 | 9.5 |
| 8類 言語 | 2,766 | 2.9 |
| 9類 文学 | 8,051 | 8.3 |

また、レファ協では分類記号と同じく主題による検索を実現するために、「レファレンス事例の中心的な内容や主要な概念を表現している語」を「キーワード」として付与することができる(付与するか否かは任意で、自由語を複数登録できる)。なんらかのキーワードが付与されていた事例は6万6,486件(公開事例の約69%)で、分類記号と同程度の付与状況であった。

表4は付与件数の多いキーワード上位5位を示したものである。「城跡」、「香川県」、「郷土資料」、「歴史」など、地域や歴史に関するキーワードが多いことがわかるが、最も付与数が多いキーワードでも「城跡」の874件にとどまった。付与されたキーワードは全体でのべ10万4,522件に及び、登録されている事例の内容は多様であることがうかがえる。

表4 登録事例数上位のキーワード(5位まで)

| キーワード | 登録事例数(件) |
|-------|----------|
| 城跡 | 874 |
| 香川県 | 681 |
| 郷土資料 | 626 |
| 歴史 | 569 |
| 統計 | 503 |

3. 解決状況

レファ協にはレファレンス質問が解決したのか、未解決のままなのかを任意で付与できる、「解決/未解決」の項目がある。解決状況が登録されていた事例は7万8,229件(公開事例の約81%)で、そのうち解決済みのものは7万2,815件(解決状況登録事例の約93%)、未解決のものは5,414件(同じく約7%)であった。レファ協登録事例のほとんどは解決済みの事例であることがわかる。

若干ではあるが、分類によって解決状況には異なる傾向が見られる。表5は付与されている分類記号(第一次区分)ごとに、解決事例数、未解決事例数、未解決事例の割合を示したものである。なお、分類記号は一事例に複数付与できるため、表5の事例数の合計は全事例数に一致しない。3類(社会科学)の事例で未解決の割合が5.9%と低い一方で、0類(総記)、8類(言語)などでは未解決の割合が約10%に至っている。類ごとに統計的有意差を見ると、3類(社会科学)に属する事例は他の事例より統計的に有意に未解決割合が低く(カイ二乗検定、片側検定、 $p < 0.01$)、0類(総記)、6類(産業)、7類(芸術)、8類(言語)、9類(文学)は有意に未解決割合が高かった(カイ二乗検定、片側検定、 $p < 0.01$)。レファ協登録事例に限って言えば、3類(社会科学)に属する質問はより解決済みのものが多い一方、0類(総記)、6類(産業)、7類(芸術)、8類(言語)、9類(文学)は未解決のものがより多いと言える。

表5 NDC第一次区分(科目)ごとの解決状況

| 科目 | 解決数(件) | 未解決数(件) | 未解決数の割合(%) |
|---------|--------|---------|------------|
| 0類 総記 | 4,271 | 489 | 10.3 |
| 1類 哲学 | 3,143 | 248 | 7.3 |
| 2類 歴史 | 14,576 | 1,076 | 6.9 |
| 3類 社会科学 | 12,566 | 785 | 5.9 |
| 4類 自然科学 | 4,513 | 313 | 6.5 |
| 5類 技術 | 5,364 | 392 | 6.8 |
| 6類 産業 | 4,328 | 409 | 8.6 |
| 7類 芸術 | 6,426 | 576 | 8.2 |
| 8類 言語 | 1,850 | 205 | 10.0 |
| 9類 文学 | 5,848 | 559 | 8.7 |

4. 質問・回答文の長さ

10万件近くにも及ぶ(データ取得時点)レファレンス事例中の質問・回答について、詳細な内容分析を行うには時間がかかるが、単純に質問文と回答文の長さを分析するだけでも、一定の知見を得ることができる。

表6は質問文、回答文それぞれの単純な統計量を見たものである⁽¹⁰⁾。質問文は平均値約66文字、中央値40文字に対し、回答文は平均値約402文字、中央値220文字で、当然ながら質問文よりも回答文の方が長く、5倍程度になっている。また、これも当然と言えば当然ではあるが、質問の長さとは回答の長さには有意な正の相関関係があり、質問が長いほど回答も長い傾向がある。ただし、両者の相関は必ずしも強くはない(スピアマンの順位相関係数。p<0.01、 $\rho = 0.222$)。なお、質問文・回答文ともデータの偏りが大きく(一部の極端に長い文が平均値に影響している)、正規分布していないことから、以下の分析では代表値として中央値を、統計的検定にはノンパラメトリック検定(データが正規分布していることを前提としない検定)を用いる。

表6 質問文・回答文の統計量

| | 質問文(文字) | 回答文(文字) |
|------|---------|---------|
| 平均値 | 65.7 | 401.9 |
| 中央値 | 40 | 220 |
| 標準偏差 | 84.77 | 560.43 |

分類によって、質問文・回答文の長さには異なる傾向がある。表7は付与されている分類記号(第一次区分)ごとに、質問文・回答文の長さの中央値をまとめたものである。質問文については1類(哲学)・2類(歴史)に属する事例では短く、0類(総記)・9類(文学)に属する事例で長い傾向がある。一方で回答文については5類(技術)が最も長く、その他6類(産業)・1類(哲学)・4類(自然科学)で長く、8類(言語)や9類(文学)では短い。この分類による傾向の違いは、8類(言語)における質問文の長さを除き、全て統計的に有意である(マン・ホイットニーのU検定、p<0.05)。詳細につ

いては文の内容の分析に踏み込む必要があるが、主題によって質問や回答の長さには異なる傾向があるようである。

また、レファレンス事例の解決状況と質問文・回答文の長さにも有意な関係が認められる。表8は解決状況と質問文・回答文の長さの中央値をまとめたものであるが、表のとおり、回答文についてはわずかな差であるが、質問文については未解決事例の方が顕著に長い。この差はいずれも統計的に有意である(マン・ホイットニーのU検定、p<0.05)。より長い質問の方が、解決しない事例が多いことがここからうかがえる。

表7 NDC第一次区分と質問・回答文の長さ(中央値)

| 科目 | 質問文(文字) | 回答文(文字) |
|---------|---------|---------|
| 0類 総記 | 45 | 192 |
| 1類 哲学 | 31 | 265 |
| 2類 歴史 | 31 | 188 |
| 3類 社会科学 | 37 | 237 |
| 4類 自然科学 | 35 | 252 |
| 5類 技術 | 35 | 279 |
| 6類 産業 | 36 | 265 |
| 7類 芸術 | 36 | 207 |
| 8類 言語 | 37 | 178 |
| 9類 文学 | 43 | 174 |

表8 解決状況と質問・回答文の長さ(中央値)

| 解決状況 | 質問文(文字) | 回答文(文字) |
|------|---------|---------|
| 解決 | 42 | 222 |
| 未解決 | 62 | 227 |

5. 言及URL

レファ協には回答を作成するにあたって参考にした資料を登録する「参考資料」欄が設けられているが、同欄は主に人間が閲覧することを前提にした、自然言語による情報の登録がなされており、機械的にその内容を分析することは(例えば登録されている資料が図書なのか論文なのか等を特定し、書誌事項を同定することは)必ずしも容易ではない。しかしその中でウェブページのURLについては、機械的に同定し、よく言及されるウェブサイト等を特定することが比較的容易である。そこでなんらかのURLを含むレファレンス事例について、そのURLを抽出し、分析した(なお、ここではあえて参考資料欄に限ることはせず、質問文・回答文中などのURLも分析に加えている)。なんらかのURLに言及していたレファレンス事例は3万3,204件(公開事例の約34%)で、言及されていたURLのべ数は11万2,742件であった。

表9はレファレンス事例からの言及数が多いFQDN(完全修飾ドメイン名。Fully Qualified Domain Name)上位10位までを示したものである。最も多いのは国立

国会図書館サーチのURLで、2万5,000件以上と群を抜いている。次いでCiNii (ArticlesとBooks、Dissertationsのすべてを含む)、国立国会図書館デジタルコレクション、レファ協自身と続く。そのほかりサーチ・ナビやWebcat Plus、NDL-OPACなど、NDLをはじめ国立情報学研究所 (NII) など、図書館関係のリソースへの言及が多いのは予想の範囲内である。一方で、Yahoo! 百科事典、コトバンク、Wikipedia日本語版など、オンライン事典類への言及も一定程度以上、存在する。ここで特徴的なのは一般により広く普及していると考えられるWikipedia日本語版よりも、Yahoo! 百科事典(『日本大百科全書』をベースとしたオンライン事典。2013年にサービス終了) やコトバンク(朝日新聞社が主体となって運営するオンライン事典) への言及の方が多くことである。レファ協に事例を登録する図書館において、オンライン事典に言及する場合であっても、編集・運営者の素性が重視されていることがここからうかがえる。

表9 言及数上位のFQDN (10位まで)

| FQDN | ウェブサイト名 | 言及数 |
|----------------------|-------------------|--------|
| iss.ndl.go.jp | 国立国会図書館サーチ | 25,745 |
| ci.nii.ac.jp | CiNii | 4,082 |
| dl.ndl.go.jp | 国立国会図書館デジタルコレクション | 3,050 |
| crd.ndl.go.jp | レファレンス協同データベース | 2,574 |
| 100.yahoo.co.jp | Yahoo! 百科事典 | 1,927 |
| kotobank.jp | コトバンク | 1,918 |
| rnavi.ndl.go.jp | リサーチ・ナビ | 1,722 |
| webcatplus.nii.ac.jp | Webcat Plus | 1,104 |
| opac.ndl.go.jp | NDL-OPAC | 1,074 |
| ja.wikipedia.org | Wikipedia 日本語版 | 1,039 |

おわりに

本稿ではレファ協に登録・公開されているレファレンス事例9万6,504件の分析から、日本のレファレンスサービスにおける質問・回答の概況を示すことを試みた。あくまでレファ協参加機関が、公開しても構わないと考えた事例のみに基づく結果ではあるものの、一定の傾向は垣間見えたと言えよう。分析結果からわかった主な傾向は以下のとおりである。

- ・分類としては2類(歴史)、3類(社会科学)に関する質問が多い。2類(歴史)の中では日本の歴史・地域の歴史や郷土に関する質問が多い。0類(総記)、1類(哲学)、8類(言語)に関する質問は少ない。

- ・登録された質問の大多数は解決済みである。中でも3類(社会科学)は解決済みの質問が多いが、0類(総記)、8類(言語)については未解決の質問も1割程度、

登録されている。

- ・質問・回答の長さには分類によって異なる傾向がある。また、未解決の事例は質問文が長い傾向がある。

- ・なんらかのURLを記載する場合、国立国会図書館サーチやCiNiiをはじめ、図書館関係のリソースか、オンライン事典類がよく参照されている。オンライン事典の中ではWikipediaよりもコトバンク等、編集・運営者が確かであるものへの言及が多い。

本稿では扱わなかったが、事例登録館を区別しての分析(例えばNDLと都道府県立図書館、市町村立図書館、大学図書館、専門図書館など)や、質問文・回答文の内容の分析にまで踏み込めば、より有益な知見を見出すこともできるであろう。レファ協のデータはAPIを通じて誰でも利用できる状況にあり⁽¹¹⁾、日本のレファレンスサービス研究の共通のデータ基盤としても、レファ協の存在には大きな意義がある。

- (1) 国立国会図書館. “レファレンス協同データベース事業実施要項”. レファレンス協同データベース. http://crd.ndl.go.jp/jp/library/documents/collabo_ref_guide.pdf, (参照 2017-03-17).
- (2) 谷本達哉, 兼松芳之. レファレンス事例データベースの協同構築事業におけるデータ登録の現状と問題点: 国立国会図書館「レファレンス協同データベース」を対象として. 図書館情報メディア研究. 2013, 11 (1), p.11-21. <http://hdl.handle.net/2241/120075>, (参照 2017-03-17).
- (3) 谷本達哉, 兼松芳之. 図書館の情報サービスが持つ可能性: 国立国会図書館レファレンス協同データベース事業, その軌跡と展開. 図書館界. 2012, 64 (2), p.142-153. http://doi.org/10.20628/toshokankai642_142, (参照 2017-05-24).
- (4) 谷本達哉. 国立国会図書館レファレンス協同データベース事業に関する研究. 筑波大学, 2016. 博士論文. <http://hdl.handle.net/2241/00145213>, (参照 2017-03-17).
- (5) 前掲.
- (6) 国立国会図書館. “外部提供インタフェース (API1.0,API2.0)”. レファレンス協同データベース. <http://crd.ndl.go.jp/jp/help/crds/api.html>, (参照 2017-03-17).
- (7) “レファレンス協同データベースのインターネット公開件数が10万件を突破しました!”. 国立国会図書館. http://www.ndl.go.jp/jp/news/fy2016/_icsFiles/afieldfile/2017/03/01/pr170303.pdf, (参照 2017-03-17).
- (8) Collaborative Reference Database as of 2017-02-17. <https://doi.org/10.5281/zenodo.573265>. (accessed 2017-06-01).
- (9) 谷本. 前掲.
- (10) 質問文・回答文の文字数については全角・半角を問わずに集計している。そのため、質問文・回答文中にURLを含む場合などには文字数を見かけの長さよりも多めに計算している場合がありうる。また、回答に至るまでのプロセスについては回答そのものとは別に、「回答プロセス」(ans-proc)項目に記述されている場合があるが、回答プロセスについては回答文の長さの集計には加えていない。
- (11) レファ協 API の使用にあたっては以下のマニュアルが用意されている。国立国会図書館. “レファレンス協同データベース・システム操作マニュアル (一般利用者用)”. レファレンス協同データベース. <http://crd.ndl.go.jp/jp/help/general/api.html>, (参照 2017-03-17).

[受理: 2017-05-15]

Sato Sho
Yoshida Mitsuo

Current Status of Reference Services in Japanese
Libraries Analyzed with the Collaborative Reference
Database