

# テキスト解析による紋章情報の分類<sup>¶</sup>

柴田みゆき\*, 生田敦司\*, 横澤大典\*, 杉山正治\*, 平塚聡<sup>§</sup>

大谷大学文学部人文情報学科\*, 四條畷学園短期大学<sup>§</sup>

## 1. はじめに

西洋の紋章は、親子関係・婚姻関係・血縁関係に加え、個人の社会的地位や異動を表すコンパクトな図像情報である。しかし、その内容はきわめて厳密かつ膨大な情報を含む上、例外も多い。このため、紋章学が登場し、紋章の収集・抽出・分類が行われてきた。

一方、ICTの発展により、それまでの紋章学を元にインターネットを利用した紋章の幅広い収集とデータベース化、およびその検索に関する研究が行われるようになった[1]。しかし、紋章の検索には紋章学に基づいた分類法が用いられるため、紋章学用語を知らないユーザは検索が困難である。この問題を解決するため、複数の図像から選択肢を提示する研究も存在する[2]。しかし、扱えるデータ数が極めて少ない。この他、画像認識を利用した研究もある[3]。しかし、現状では簡単な要素抽出に留まり、研究は進んでいない。

このような状況に対し、我々はこれまでに、紋章の図像化規則が比較的解明されている中世イギリスを中心にデータベース化のための要件整理やデータフィールドの研究を行ってきた[4]。この研究を通して、図像としての紋章のみでは必要な情報が得られない場合があることが明らかになった。例えば、図1の黒い楕円で示した紋章の上部には文字情報が欠落し、図像だけが示されている[5]。これは紋章のデザインが新規に制定される際に、先祖に関する情報が極めて判読困難な形で組み込まれたケースである。すなわち、紋章のデータベース化には図像解析だけでは不十分であり、様々なテキストを用いて紋章情報を補強する必要がある。

そこで、本研究では紋章が内包する情報を補強可能なテキストの資料として、図2に例示するような「紋章の認定証」に着目した[6]。紋章の認定証は紋章を必要とする人に対し、国家機関である紋章院が発行する紋章情報の確認とそれを利用する人とを同定する公文書である。この中で、我々の研究蓄積[7]のある人物として、紋章とその認定証が公開されている中世イングランドのシェイクスピアを取り上げ、彼がデザインした紋章の認定証に記載されたテキストを利用し、解析した結果を示す。

## 2. テキストの選定

シェイクスピア家には紋章が無かったが、彼の父が紋章継承権を持つ女性と結婚した。紋章継承権を持つ女性が婚姻した場合、相手の紋章が内包する家系や個



図1: “The Dering Roll” にみられる情報の欠落 [5]



図2: イギリスの紋章院発行の認定証の例 [6]



(a) 認証された紋章 [9]



(b) 申請のためのドラフト [8]

図3: シェイクスピア家の紋章資料

人情報を反映した変更を、自らの持つ紋章に施す必要がある。そこで、父のためにシェイクスピアは2度紋章申請を行い、いずれも認められている。したがって、紋章院からの認可証は2通発行されている[8]。図3(a)は2度目の申請に基づくシェイクスピア家の紋章であり、図3(b)はその申請に利用したドラフトである。今回は、シェイクスピアが最終的に利用した2度目の紋章の認可証を利用することとする。

本資料は1599年に発行されたものであり、初期近代英語 (early modern English) により記述されている。正書法は現在利用される現代英語 (contemporary English) と類似点が多いが、綴りが異なるか不安定である場合が多い。これが一次データである紋章の認定証の正式な表記なのか、書籍化された際のタイプミスなのかは不明である。

<sup>¶</sup>An approach for analyzing coat-of-arms from expository text  
\*Miyuki Shibata, Atsushi Ikuta, Daisuke Yokozawa, and Seiji Sugiyama: Human Informatics, Otani University

<sup>§</sup>Satoshi Hiratsuka: Shijyonawate Gakuen Junior College

Words	頻度	割合
1 and	45	7.49%
2 of	39	6.49%
3 the	32	5.32%
4 arms	12	2.00%
5 in	10	1.66%
6 for	9	1.50%
6 his	9	1.50%
6 to	9	1.50%
9 or	8	1.33%
9 that	8	1.33%

(a) 高頻度 10 位

3-grams	頻度	割合
1 of arms and	3	0.56%
1 said john shakespeare	3	0.56%
1 the said john	3	0.56%
4 arden of wellingcote	2	0.37%
4 cote of arms	2	0.37%
4 king of arms	2	0.37%
4 laws of arms	2	0.37%
4 of our offices	2	0.37%
4 shalbe lawfull for	2	0.37%
4 shieldes of arms	2	0.37%
4 the same with	2	0.37%

(b) 3-gram

図 4: 単語集計の結果

テキストは Lambert 編によるシェイクスピアの資料集の PDF 版から入手した [8]. 但し, 同書籍には 2 通ともテキスト情報として掲載されており, 一次資料の出典に関する言及が無い. 一方, 森は現存しないと主張している [9] ため, 信憑性については注意が必要である. ソフトウェアを利用したテキスト変換では誤認識が多かったため, 手入力での抽出した. また, 従来より作成してきた紋章学用語ファイルを参照ファイルとした.

### 3. 解析手法

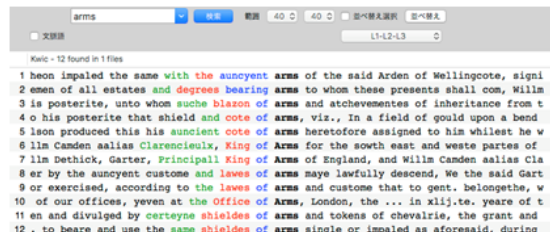
通常, テキストマイニングでは重要語句の頻度を指標とするが, 「紋章の認可証」は一定の形式に則って記述される公文書であるため, 公文書の特徴抽出となってしまう可能性がある. そこで, 自然言語処理において多用されるコンコーダンス解析を行う. これは, 紋章用語が含まれるテキストや前後の文脈, 共起解析などを行うものである. この結果から紋章を見るだけでは得られない情報が得られれば, 紋章の認定証をテキスト解析する意義があると判断する.

この目的にかなうコンコーダンスとして, Casual Conc バージョン 2.0.3 [10] を利用した. 理由は, (1) 英語を主対象としたコンコーダンスであること, (2) 異綴り処理が行えること, (3) 後日の作業拡張を見据え, 統計処理ソフトである R や形態素解析ソフトである TreeTagger との連携が容易なこと, (4) これまで紋章の画像処理を行ってきた macOS 上でも動作すること, の 4 点である.

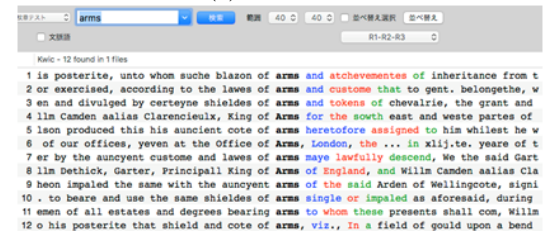
### 4. 解析結果

単語集計の結果を, 単語で高頻度上位 10 位まで示したものが図 4(a) である. 紋章用語 「arms」 と 「or」 がランクインした. そこで, 値を 3 として n-gram 解析をした結果が図 4(b) である. ここから得られた 「arms」という紋章用語で KWIC 文脈表示させた. 図 5(a) が左 3 語, 図 5(b) が右 3 語での表示結果である. さらに共起関係を調べたものが図 6 となる.

この結果から, Arden of Wellingcote 家とシェイクスピア家の関係の言及を抽出できる. これは新規に制作された紋章の画像情報からだけでは得られない情報



(a) 左 3 語抽出



(b) 右 3 語抽出

図 5: 単語 「arms」 による KWIC 文脈表示

単語	共起数	1	2	3	4	5	6	7	8	9	10	11	12
1 a field of	1	0	1	0	0	0	0	0	0	0	0	0	0
1 aforesaid during his	1	0	1	0	0	0	0	0	0	0	0	0	1
1 and atchevementes of	1	0	1	0	1	0	0	0	0	0	0	0	0
1 and custome that	1	0	1	0	1	0	0	0	0	0	0	0	0
1 and tokens of	1	0	1	0	1	0	0	0	0	0	0	0	0
1 and weste partes	1	0	1	0	0	0	0	0	0	0	0	0	1
1 and willm camden	1	0	1	0	0	0	0	0	1	0	0	0	0
1 arden of wellingcote	1	0	1	0	0	0	0	0	0	0	1	0	0
1 as aforesaid during	1	0	1	0	0	0	0	0	0	0	0	1	0
1 assigned to him	1	0	1	0	0	0	1	0	0	0	0	0	0
1 atchevementes of inheritance	1	0	1	0	0	0	1	0	0	0	0	0	0

図 6: 単語 「arms」 での 3-gram 共起関係

である. 一方, 同家の名称は 2 系統存在する [9]. 本認可証がどちらの系統を指すものか, 今後詳細な調査が必要である.

### 5. おわりに

以上, 紋章の認可証をテキスト解析し, 紋章の画像解析では得られない家系情報の抽出に成功した. 今後はこの手法を利用して, 紋章情報に関するテキストデータを収集し, 紋章データベース構築に有用な要素の選定を進める予定である.

謝辞: 本研究は, 科学研究費・基盤研究 (C) [課題番号: 26503015] の研究成果の一部である.

### 参考文献

- [1] European Foundation, “Europeana Collection”, <http://www.europeana.eu/portal/en>
- [2] University of Toronto Libraries, “British Armorial Bindings”, <https://armorial.library.utoronto.ca/>
- [3] Robert McKeag, “Heraldic applications of computational linguistics, computational geometry and image processing”, 17th Joint International Conference of the Association for Computers and the Humanities and the Association for Literary & Linguistic Computing, 2005.
- [4] 柴田, 生田, 横澤, 杉山, 平塚, 三浦, “紋章資料が内包する情報整理の検討”, 情報処理学会第 78 回全国大会, 7F-02, pp. 4-555~4-556, 2016.
- [5] British Library, “Add Roll 77720”, [http://www.bl.uk/manuscripts/FullDisplay.aspx?ref=Add\\_Roll\\_77720](http://www.bl.uk/manuscripts/FullDisplay.aspx?ref=Add_Roll_77720)
- [6] College of Arms, “History”, <http://www.college-of-arms.gov.uk/about-us/history>
- [7] 三浦, 柴田, “紋章における画像化規則の複雑性の検討”, 情報処理学会第 77 回全国大会, 2H-04, pp. 4-557~4-558, 2015.
- [8] Collated and chronologically arranged by D.H. Lambert, “Cartae Shakespearianae: Shakespeare documents, a chro-nological catalogue of extant evidence relating to the life and works of William Shakespeare”, G. Bell, London, 1904. <https://hdl.handle.net/2027/mdp.3901500052458>
- [9] 森護, “シェイクスピアの紋章学”, 大修館書店, 1987.
- [10] “CasualConc” バージョン 2.0.3, <https://sites.google.com/site/casualconcj/Home> (2016).