

民間伝承の計量分析： 「怪異・妖怪伝承データベース」の俯瞰的分析

佐藤 浩輔（株式会社バンダイナムコ研究所）

中分 遥（高知工科大学, University of Oxford）

概要：近年、人文社会科学においてデータベースのデジタル化が進展している。それに伴い、民話や神話などに対して計量分析を行う計算民話学・計量民俗学と呼ばれる取り組みがなされている。本研究では、国内の民間伝承に関する主要なデータベースである「怪異・妖怪伝承データベース」を計量的に分析する端緒として、書誌情報を用いて全容を把握するための基礎的な分析を行った。収録されている事例数について都道府県ごとの分布を可視化したところ、地域による資料数の多寡に差があることが明らかになった。主要な「怪異・妖怪」に関するクラスター分析の結果から、怪異・妖怪の分布の地域的特性が示唆された。データベースを用いた計量分析の展望について議論する。

キーワード：デジタル人文学, 計算民俗学, 怪異・妖怪, データベース

Quantitative Analysis of a Japanese Folkloristic Database: A Statistical Overview

Kosuke Sato (BANDAI NAMCO Research Inc.)

Yo Nakawake (Kochi University of Technology, University of Oxford)

Abstract: With the rise of digital databases in the humanities and social sciences, new computational approaches have appeared in folkloristics, known as ‘computational folkloristic’ or ‘quantitative folkloristic’. In the present study, we quantitatively analysed ‘Folktales of Strange Phenomena and Yokai Database’, one of the major databases on folklore in Japan. We analyzed the bibliographic information of the database. We visualized the distribution of the materials and the major items by prefecture. Further, cluster analysis showed regional characteristics of the distribution of ‘strange phenomena and yokais’. The prospects for future quantitative analysis of the database were discussed.

Keywords: Digital Humanities, Computational Folkloristics, Strange Phenomena and Yokai, Database

1. 序論

1.1 はじめに

近年、民話(folktale)や伝説(legend)を含む民間伝承のデジタル・アーカイブ化が進んでいる。こうしたアーカイブ化された民俗資料に対して、自然言語処理や統計解析など計量的な研究を行う計算民俗学(computational folkloristics)[1]・計量民俗学[2]と呼ばれるアプローチが盛んになり、これまでの研究手法と共に用いられるようになってきた。本稿では、本邦のデジタル・アーカイブである「怪異・妖怪伝承データベース」[3]に含まれる書誌情報を用い、計量分析を行った結果について報告する。

1.2 物語・民間伝承の計量的研究の動向

近年、民俗学・人類学において民話・神話を計量的に研究する手法として隆興しているのが、文化系統学と呼ばれる進化生物学における系統学の手法を文化に適用する手法である。このアプロ

ーチでは、物語の系統関係や起源の推定などを数理的に検討している。民話・神話への適用例としては、魔法物語[4], [5]や「赤頭巾」[6]・「ポリュペーモス」[7]のヴァリエーションを用いて物語間の系統関係、物語の分布と地域・言語との関係を検討した研究がある。これらは物語の地理的な分布や文章に出現するモチーフなどが分析対象であったが、「百鬼夜行研究絵巻」の系統関係を扱った山田(2009)[8]のように図像について同様のアプローチをとっている研究もある。

この分野における研究では、分析対象である民話・神話に出現する登場人物やモチーフ等を研究者が手動でコード化することが多かった。しかし近年、自然言語処理技術を利用し、物語のテキストそのものを計算機で扱う手法が隆興している。Reagan et al. (2016)[9]は、1000以上の物語について時系列的に感情スコアを算出し、感情が推移するパターンを検証した。また、著者らも国際昔話話型カタログに含まれる動物昔話を分析した[10]。その他、情報科学的な技術を使って物語の自動生成やモチーフのタグ付け・分類といった研究もなされている[11], [12]。

1.3 物語の認知科学

物語は伝統的には人文学分野で扱われてきた。近年、社会科学・認知科学においても関心が高まっており、特に認知科学では宗教の起源や「物語る存在である人間 (homo fabulans)」の本性にアプローチする手段として注目されている。

なぜ、妖精や幽霊などといったものが（実在しないにも関わらず）世界のあらゆる地域で語り継がれているのだろうかという問いに対して、認知科学では「ヒトが乳児期から持つ直観に『最小限に違反する』 (minimally counterintuitive) 対象に人々が注目しやすい・記憶しやすい」という説明がある (Minimally counterintuitiveness モデル[13], [14])。実際にグリム童話で人気な物語ほどキャラクターが最小反直観的であることが計量分析によって示されてきた[15]。

また、想像界に人々が惹きつけられる理由として未知のものを人々が追求する探索傾向と関連を指摘する研究もある[16]。こうした想像界に関する文化資料の分析は、物語や伝承を生み出し維持する人間自身の性質の理解へとつながる。

1.3. 本研究の目的

人文学の枠を超えた多様な領域において物語や想像界への関心が高まっている中、市井の人々の間で語り継がれてきた怪異や妖怪伝承についてもその関心が高まっている[16]–[18]。

本邦においてこれら対象に特化したデータベースが「怪異・妖怪伝承データベース」である。

物語や伝承といった文化資料データを認知科学・社会科学的な問いを検討する題材として扱おうとした場合、たとえば資料数の地域的な偏りといった全体の記述統計的な概容を把握する必要がある。

当該データベースを利用した計量研究の試みとして[2], [19]など特定の主題に関する分析を行う研究はある一方で、データベースに対してその全体像を与えるような分析を行っている研究は今のところないようである。

そこで本研究では、「怪異・妖怪伝承データベース」の計量分析に向け、その足掛かりとして、全体像を把握することを目的とする。

2. 方法

2.1 データベース

国際日本文化研究センター（以下、日文研）が Web 上に公開している「怪異・妖怪伝承データベース」¹のデータを用いた。

怪異・妖怪伝承データベースは民俗調査などで報告された怪異・妖怪の事例を網羅的に収集したデータベースであり、国内の民俗学雑誌などを中

心に採取された事例 35,414 件 (2021 年 8 月現在) が収録されている。Web 上で利用できる情報としては、事例番号、事例の呼称、出典、話者 (文献)、地域、要約がある。著作権上の配慮から、事例の詳細を記したカードについては日文研の内部からのみアクセスできるようになっている。

当該データベースの Web ページでは、地域・呼称などの項目に基づいた検索のほか、全文検索、代表的な怪異・妖怪のリストからも検索できるようになっている。

入力ミスの修正や記述の修正等で定期的にバージョンが更新されており、直近では 2021 年 2 月および 7 月に更新がなされている。

2.2 データセット

2021 年 3 月時点で Web 上でアクセスできる事例の書誌情報 (2.3.1 版) を用いて分析用のデータセットを構築し、加えて以下の処理を行った。**地域**: 各事例の項目「地域 (都道府県名)」をもとに、事例番号と都道府県の組を抽出した。ひとつのレコードに複数の都道府県が対応する一対多の関係を許した。ただし地方名 (e.g., 中国地方) および外国の国名・地名など現在の都道府県名に該当しないデータは分析から除外した。また、表記ゆれ等についてはデータセット上で修正を行った。

タグ: Web 検索インターフェイス上において「地域」が「全国」における「代表的な怪異・妖怪」のリストに表示される大項目 12 件 (キツネ・テング・カップ・タヌキ・ヘビ・ダイジャ・ユウレイ・オニ・タタリ・ヒノタマ・ヤマノカミ・ムジナ) を「大カテゴリ」、各大カテゴリに含まれる小項目 (e.g. ウシオニ) を「小カテゴリ」として、それぞれタグとして扱った。事例「ヒノタマ、キツネビ」等、複数のカテゴリに該当するレコードについては両方のタグを付した。タグ情報のみ 2021 年 8 月時点での情報 (2.3.2 版) を用いた。

分析ではまず地域情報のみ、タグ情報のみのデータセットについて集計した。有効な地域情報の存在する事例は 33,744 件、該当するタグをもつレコードは 11,873 件であった。

タグの地理的分布を分析するため、最終的に、上記地域情報およびタグ情報の両方が存在する事例、11,409 件のデータを分析に利用した。データの加工および分析には Python (3.6.8) と R (3.6.3) を用いた。

3. 結果

3.1 都道府県の資料分布

都道府県別の資料 (事例) 数の偏りについて調べるために、地域情報のみのデータセットを用いて都道府県ごとの資料数を集計した。

¹ <https://www.nichibun.ac.jp/YoukaiDB/>

資料数の分布を図に示した(図 3.1.1), ほとんどの都道府県では 300 件から 1000 件程度の間であった。その一方で, 1000 件を超える, または 300 件以下である都道府県もそれぞれ数件ずつあった。資料数がもっとも少ないのは北海道の 73 件, 資料数がもっとも多いのは長野県の 2,968 件であり, 約 40 倍の開きがあった。

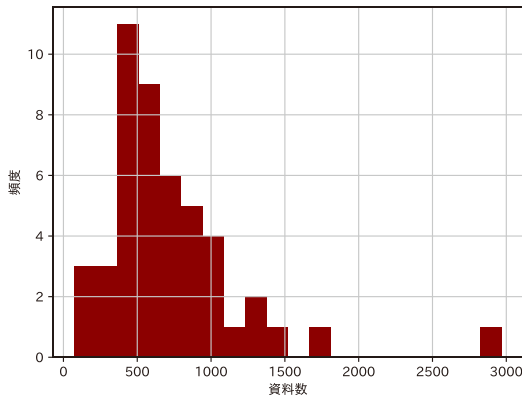


図 3.1.1 資料数の分布 (ヒストグラム) .

各都道府県の資料数を階級区分図として地図上に示した(図 3.1.2). 長野県, 宮城県, 新潟県, 静岡県などでは 1,000 件以上事例がある一方で, 北海道や佐賀県, 福岡県などでは 200 件以下であった。少なくとも件数に関していえば中部や関東・東北, 近畿といった地方の資料は比較的潤沢に採録されているものの, 北海道や九州・中国といった地方の資料数は少ないという結果であった。

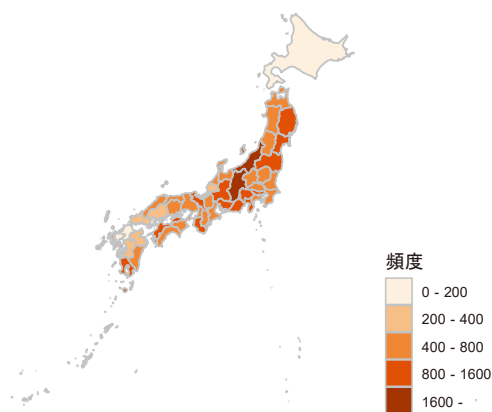


図 3.1.2 都道府県別の資料数.

3.2 典型的な怪異・妖怪の分布

都道府県の事例数の偏りを補正した上でキツネやオニといった大項目カテゴリのタグの分布を分析した。

都道府県によって事例の総数に差があるため, 都道府県ごとに総事例数に対するそれぞれのタグの出現頻度からオッズを算出した(全 12 次元). 出現頻度がゼロであるセルが存在するため Haldane-Anscombe 法²を用いて値を補正した。

それぞれのタグがどの都道府県で出現しやすい／出現しにくいかを把握するために, 各タグの出現オッズについて全都道府県の(幾何)平均オッズとの比を求めた(タグ別の階級区分図については付録を参照のこと)。

都道府県における各タグの対数化出現オッズを用いて, 階層的クラスター分析を行った(図 3.2.1). クラスター数が 7 の時の結果(表 3.2.1)を地図上に示した(図 3.2.2)。

全体的な傾向を見てみると, クラスター 1 は関西とその近隣, クラスター 2 は四国, クラスター 3 は関東及び中部, クラスター 4 は単独で沖縄, クラスター 5 は北海道・東北, クラスター 6 は九州とおおむね地理的な区分を反映しており, 地理的に近いほど典型的な怪異・妖怪の出現傾向が似ているという結果になった。

その一方で, クラスター 7 は九州(長崎・佐賀)・中国(広島・山口・島根)・東北(青森)・近畿(大阪)と飛び地的に分布している。また, クラスター 3 において近畿地方の兵庫県が含まれており, 地域的には隣接していないものの出現傾向が類似している地域があった。

クラスターごとの特徴を把握するために, それぞれのクラスターについてタグの平均出現オッズを求めた(図 3.2.3)。

クラスター 4 (沖縄) はユウレイ・ヘビを除くほとんどのタグについて他のクラスターよりも出現頻度が低かった。これは, 本土において典型的な怪異・妖怪が沖縄ではあまり現れず, 沖縄固有の怪異・妖怪が事例の多くを占めていたと考えられる。

クラスター 1 (関西) および 2 (中国) については, タヌキ, ヘビ, テング, タタリ, ダイジャなどの出現が多かった一方, カップやユウレイ, ヒノタマなどの出現が他のクラスターより少なかった。

クラスター 3 (関東) および 5 (北海道・東北) では, カップおよびキツネ, ムジナの出現が多かった。特にムジナについては他のクラスターではほとんど見られなかった。一方で, ヤマノカミやタヌキの出現は少なかった。

クラスター 6 (九州) については, カップおよびタヌキが他のクラスターよりも多く出現した一方, タタリやダイジャなどの出現は少なかった。

クラスター 7 (その他) については全体の傾向はクラスター 6 に類似しているものの, ヒノタマ, オニ, ユウレイなどの出現が多い, ヘビ, テングの出現が少ないという特徴がみられた。

² すべてのセルの頻度に 0.5 を足して算出する補正手法。

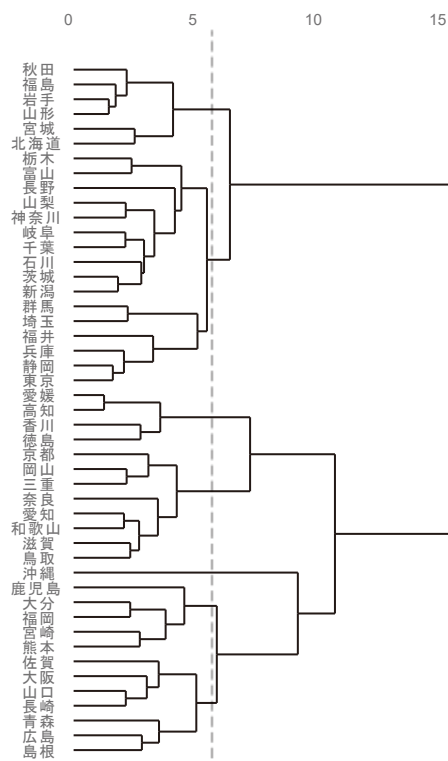


図 3.2.1 代表的な怪異・妖怪の出現パターンのクラスター分析（デンドログラム）。Ward 法による。点線はクラスター数が 7 となる切断レベルを示す。

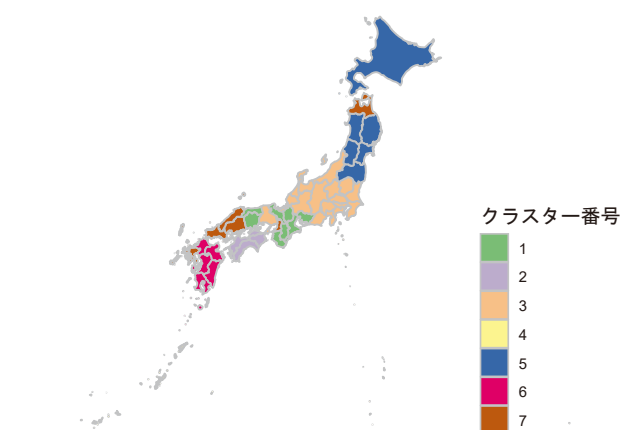


図 3.2.2 代表的な怪異・妖怪の出現パターンによるクラスター。クラスター数が 7 のときの結果を示す。各クラスターに含まれる都道府県については表 3.2.1 を参照のこと。

表 3.2.1 出現パターンによる都道府県のクラスター分類。クラスター数が 7 のときの分類結果を示す。

番号	名前	都道府県
1	関西および近隣	三重, 滋賀, 京都, 奈良, 和歌山, 愛知, 鳥取, 岡山
2	四国	徳島, 香川, 愛媛, 高知
3	関東・中部	茨城, 栃木, 群馬, 埼玉, 千葉, 東京, 神奈川, 新潟, 富山, 石川, 福井, 山梨, 長野, 岐阜, 静岡, 兵庫
4	沖縄	沖縄
5	北海道・東北	北海道, 岩手, 宮城, 秋田, 山形, 福島
6	九州	福岡, 熊本, 大分, 宮崎, 鹿児島
7	その他	青森, 大阪, 島根, 広島, 山口, 佐賀, 長崎

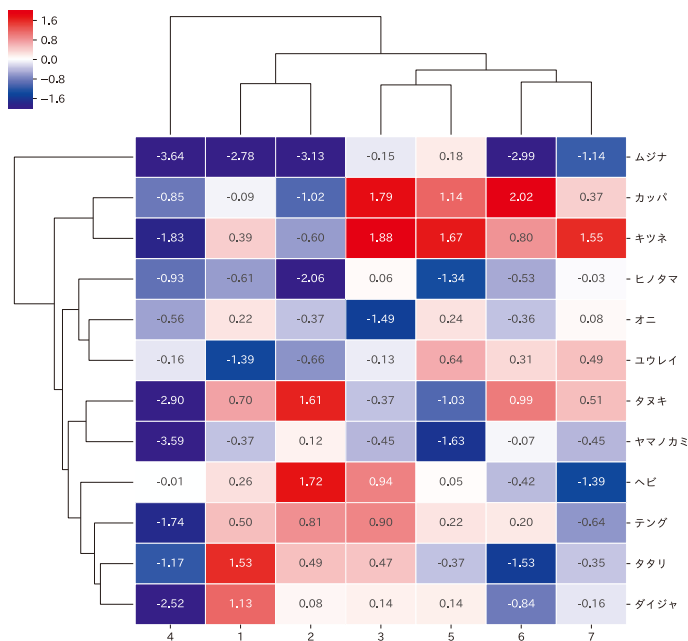


図 3.2.3 出現パターンによるクラスターの特徴。クラスター単位で各タグの出現オッズの幾何平均を算出し、全都道府県の（幾何）平均オッズとの比をとった。値は対数化したもの。Python の seaborn モジュール(0.9.0)の clustermap()関数を用いて行／列のクラスターリング（平均距離法）・並び替えを行った。

4. 考察

4.1 まとめと考察

本研究では、「日本怪異・妖怪伝承データベース」における資料の分布とタグの分布について分析し可視化した。

第一に都道府県ごとの資料数の分布を検討したところ、都道府県ごとに偏りがあり、最大値(長野県)と最小値(北海道)の間で約40倍の開きがあった。

当該データベースを用いた計量的な分析を行う場合、用途によってはこのような地域的偏りについて注意する必要がある。例えば特定の怪異・妖怪(e.g. キツネ)の事例に関するテキストを単純に集約した場合、単に事例数の多い地域のデータに結果が影響される可能性がある。このため分析者側で目的に応じて対応する必要がある。

なぜこのような偏りがあるかについては本稿の結果のみからはわからないが、例えば地域的な要因(民俗資料の採取活動に熱心/協力的であるか)、研究者・収集者側の要因(熱心な研究者・収集者がいるか)、あるいはデータベース側の要因(採録の選定基準)など、複数の可能性が考えられる。特に前者二つの観点からは今後、科学計量学的・計量文献学的な研究も可能であろう。本稿では扱わなかったが、当該データベースでは事例の報告された文献情報も含まれており、それらを分析することで文献の観点からさらなる全体の描像を明らかにすることができるだろう。

第二に、上記の偏りを補正した上で、典型的な怪異・妖怪についてそのタグの分布を求めた上、クラスター分析を行った。

この結果、「怪異・妖怪」のタグのクラスターはランダムに分布しているのではなく、四国地方・東北地方というように地理的に近接した地方区分と部分的に一致する結果が得られた。その一方で、クラスター7のように地理的に必ずしも隣接していない地域で類似したパターンが見られてもいた。このことから地理的な近接性が重要な説明要因である一方で、それ以外の説明要因が存在することが示唆される。

また、本稿では詳細に検討できていないが、怪異・妖怪側の分布にも系統的な傾向があることが推察される。キツネやタヌキのように動物の生態分布から地理的な分布を説明できる可能性のあるものがある一方で、タタリやユウレイ、オニなど背景となる要因が定かではないものもある。

これらの分布については、方言の類似性、空間的距離、植生や動物の生態、人口移動など言語学・地理学・生物学・歴史学等様々な領域で蓄積されたデータを用いて文化伝播の経路を検討することによって、本研究で示したような分布がどのようにして形成されるのかを調べることができるだろう。

4.2 本研究の限界

本研究でタグの分析に用いたレコードは11,873件であり、全事例33,744件に対して三分の一程度しかなかった。これは当該データベース上で「代表的な怪異・妖怪」として項目化されていた情報を用いてタグとして利用したためであり、項目化されていない非典型的な事例については扱えなかったためである。

この影響が顕著に現れたのが沖縄であり、図3.2.3が示すように、多くの代表的な怪異・妖怪のタグの出現オッズが沖縄では低く、今回用いた「代表的な怪異・妖怪」のタグでは沖縄の怪異・妖怪が捉えられていない可能性が高い。

ただし、非典型的な事例の場合、類似事例の少ないものが大量に存在し、仮にそのままタグを付与するとしてもそのまま集約するという形では分析できないと考えられる。

そのため、事例の本文ないしは要約テキストに対して自然言語処理を行い、自動タグ付けや内容分析を行うというのが妥当な方向性であろう。

4.3 意義と展望

本研究では「俯瞰的分析」として、特に事例数の分布を用いて予備的な分析を行った。

その結果、地域的な偏り、すなわち資料の多い地域と少ない地域とで数の上で大きな隔りがあることがわかった。このような地域的な偏りについては、分析者・研究者側で対応する必要があるものの、本質的に問題であるということではない。むしろ、そのような差異がなぜ生じたのかということ自体が新たな研究につながる糸口である可能性がある。

本研究の意義のひとつは、予備的ではあるものの、データベース全体を対象とした分析を行うことで、分析上の知見を提供したことである。

偏りが存在すること自体は単体で学術上の知見と呼べる代物ではないものの、一方で分析者が不適切な分析に陥ってしまう原因となりうる。

公開データベースの場合、分析者が必ずしもデータベース自体に知悉しているとは限らない。もし各人がそれぞれ個別に自身でデータベースの特性や注意点などを都度「発見」しながら分析しなければならぬとすれば、分析者は本来知りたい結果を得るためには不必要に多い障害を乗り越えていかねばならないことになる。

そうではなく、このようなデータベースの持つ制約や特性についての知見を研究者間で蓄積・共有していくことによって、対象のデータベース自体への理解が進み、より容易かつ安全に多様な分野の研究者が計量分析を行うことが可能になるだろう。

本稿では事例数という観点から分析を行ったが、当該データベースのレコードに含まれる情報は他にもある(e.g. 文献情報、事例の要約)。当該データベースの特性を明らかにするためには

それら今回扱わなかった変数についても分析していく必要があるだろう。

また本研究では「代表的な怪異・妖怪」の都道府県別の出現頻度からその分布パターンについて検討した。結果からは、①地理的な近接性が主要な要因でありおおむね地方区分に従ってクラスターが別れること、②地理的近接性では説明できない要因が存在する可能性があること、が示唆された。

これらの結果が民俗学上の通説や知見とどの程度合致しているかどうかまでは確認できていない。今後、より詳細な検討を進めて行くにあたり、当該分野で蓄積されてきた知見とデータの示す計量的な結果との異同を明らかにしつつ、分析結果の妥当性について、都度評価していく必要があるだろう。

近年のデータベース化・電子テキスト化により人文学の蓄積してきた叢智とも呼ぶべき膨大な資料が次々と電子的にアクセス可能になっている。社会科学や認知科学においてもこれら資源を活用する取り組みが始まってはいるものの、いまだ十分に活用されているとはいえない。

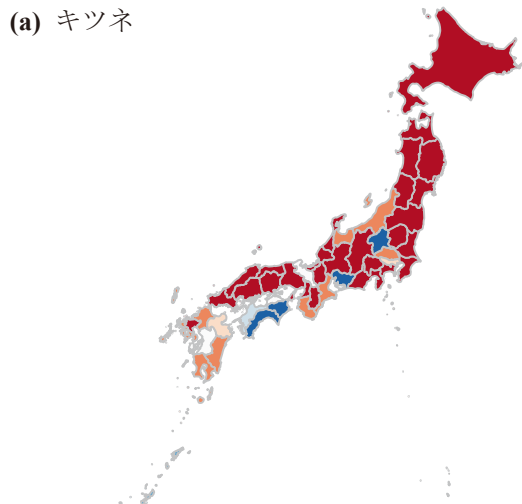
今後のさらなる活用と研究が望まれる。

参考文献

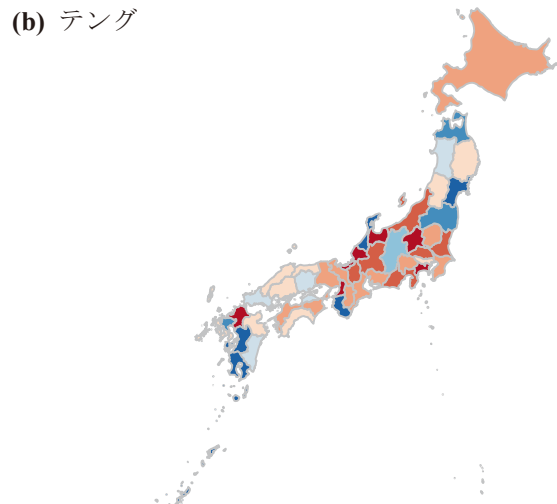
- [1] J. Abello, P. Broadwell, and T. R. Tangherlini. “Computational folkloristics,” *Commun. ACM*, vol. 55, no. 7, p. 60, 2012, doi: 10.1145/2209249.2209267.
- [2] 山田奨治. “怪異・妖怪呼称の名彙分解とその計量,” 妖怪文化研究の最前線, 小松和彦, せりか書房, 2009, pp. 269–284.
- [3] 小松和彦, 常光徹, 山田奨治, 中山和久. “異界へのいざない: 怪異・妖怪伝承データベースの試み,” 総研大ジャーナル, no. 3, pp. 42–43, Mar. 2003.
- [4] J. Oda. “Description of Structure of the Folktale: Using a Bioinformatics Multiple Alignment Program,” *Senri Ethnol. Stud.*, vol. 55, pp. 153–174, Mar. 2001.
- [5] S. G. da Silva and J. J. Tehrani. “Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales,” *R. Soc. Open Sci.*, vol. 3, no. 1, p. 150645, 2016, doi: 10.1098/rsos.150645.
- [6] J. J. Tehrani. “The phylogeny of little red riding hood,” *PLoS One*, vol. 8, no. 11, 2013, doi: 10.1371/journal.pone.0078871.
- [7] J. D’Huy. “Le conte-type de Polyphème: essai de reconstitution phylogénétique” *Mythol. française*, vol. 248, pp. 47–59, 2012.
- [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, vol. 40, pp. 103–128, 2009.
- [9] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds. “The emotional arcs of stories are dominated by six basic shapes,” *EPJ Data Sci.*, vol. 5, no. 1, pp. 1–12, 2016, doi: 10.1140/epjds/s13688-016-0093-1.
- [10] Y. Nakawake and K. Sato. “Systematic quantitative analyses reveal the folk-zoological knowledge embedded in folktales,” *Palgrave Commun.*, vol. 5, no. 1, p. 161, Dec. 2019, doi: 10.1057/s41599-019-0375-x.
- [11] 小方 孝. “プロップから物語内容の修辞学へ: 解体と再構成の修辞を中心として,” 認知科学, vol. 14, no. 4, pp. 532–558, 2007.
- [12] T. Meder, F. Karsdorp, D. Nguyen, M. Theune, D. Trieschnigg, and I. E. C. Muiser. “Automatic enrichment and classification of folktales in the Dutch folktale database,” *J. Am. Folk.*, vol. 129, no. 511, pp. 78–96, 2016, doi: 10.5406/jamerfolk.129.511.0078.
- [13] P. Boyer, *Religion explained: the evolutionary origins of religious thought*. Basic Books, 2001.
- [14] J. L. Barrett, “Coding and quantifying counterintuitiveness in religious concepts: Theoretical and methodological reflections,” *Method Theory Study Relig.*, vol. 20, no. 4, pp. 308–338, 2008, doi: 10.1163/157006808X371806.
- [15] A. Norenzayan, S. Atran, J. Faulkner, and M. Schaller. “Memory and mystery: The cultural selection of minimally counterintuitive narratives,” *Cogn. Sci.*, vol. 30, no. 3, pp. 531–553, 2006, doi: 10.1207/s15516709cog0000_68.
- [16] E. Dubourg and N. Baumard. “Why Imaginary Worlds?,” *Behav. Brain Sci.*, pp. 1–52, Jul. 2021, doi: 10.1017/S0140525X21000923.
- [17] J. Carroll, M. Clasen, and E. Jonsson. Eds., *Evolutionary Perspectives on Imaginative Culture*. Cham, Switzerland: Springer, 2020.
- [18] 高橋綾子. “妖怪に対する社会心理学的手法を用いた探索的研究,” 東洋大学大学院紀要, vol. 55, pp. 21–30, 2019.
- [19] 太田貴大. “怪異・妖怪伝承データベースを用いた計量分析の試み: 伝承呼称数と島嶼環境特徴の関係性に注目して,” 環境共生, vol. 33, pp. 25–30, 2018.

付録

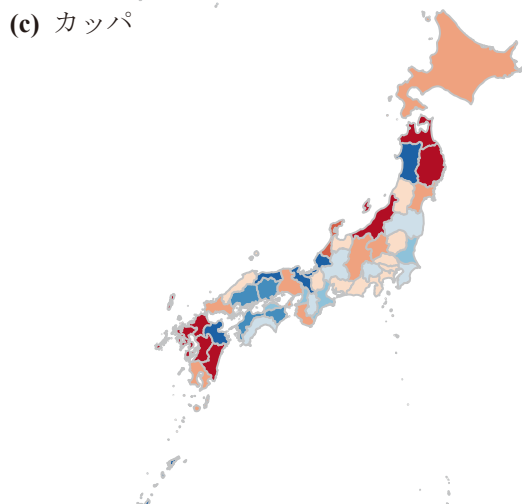
(a) キツネ



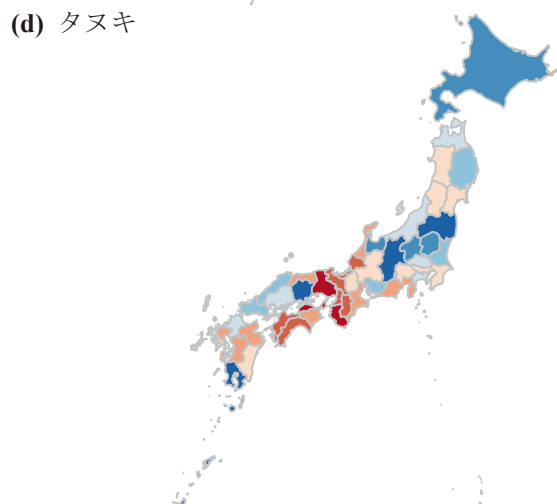
(b) テング



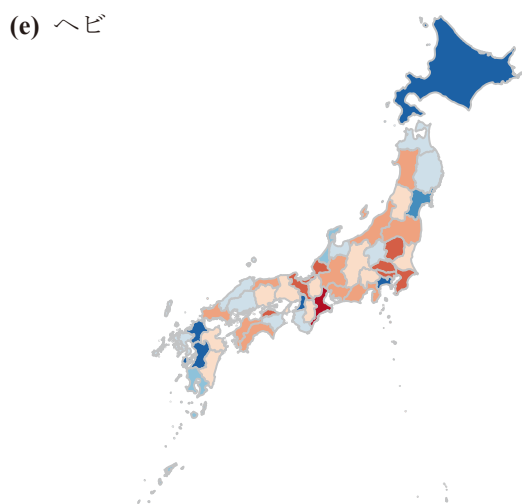
(c) カッパ



(d) タヌキ



(e) ヘビ



(f) ダイジャ

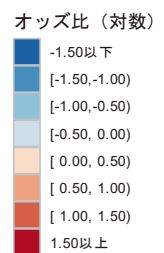
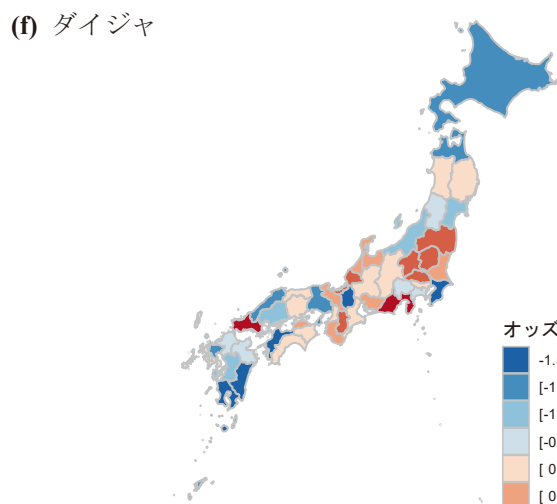
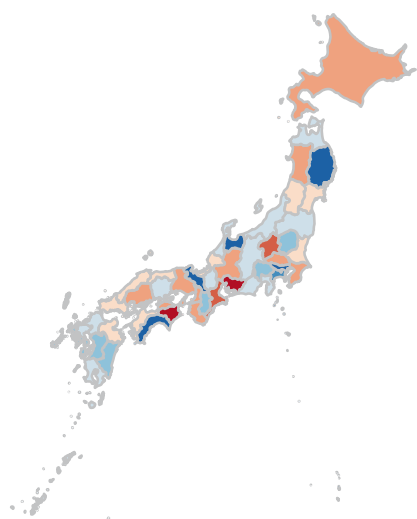
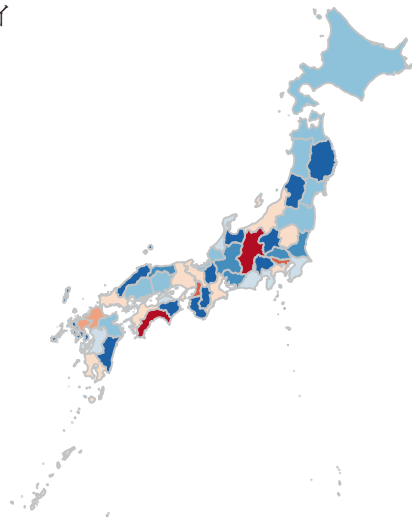


図 SI1 「代表的な怪異・妖怪」の出現パターン. タグそれぞれについて各都道府県における出現オッズを算出し、全都道府県の（幾何）平均オッズとの比をとった。値は対数化したもの。

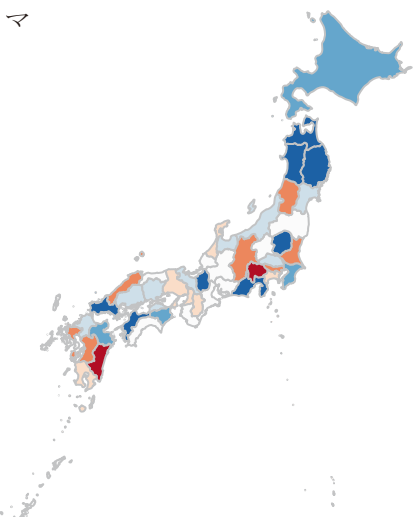
(g) オニ



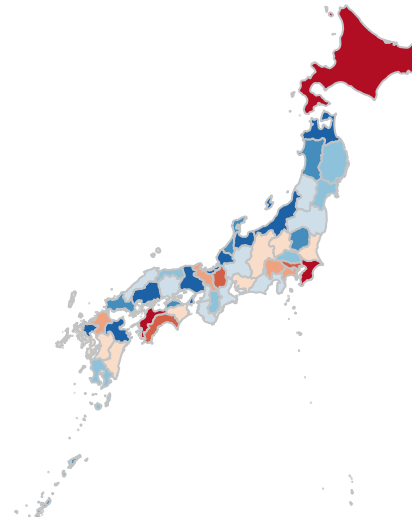
(h) ユウレイ



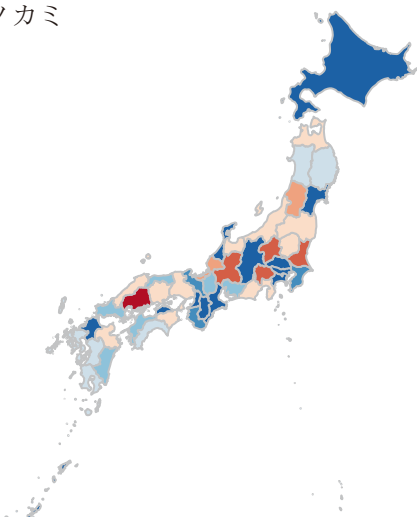
(i) ヒノタマ



(j) タタリ



(k) ヤマノカミ



(l) ムジナ

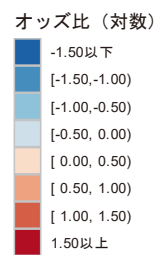
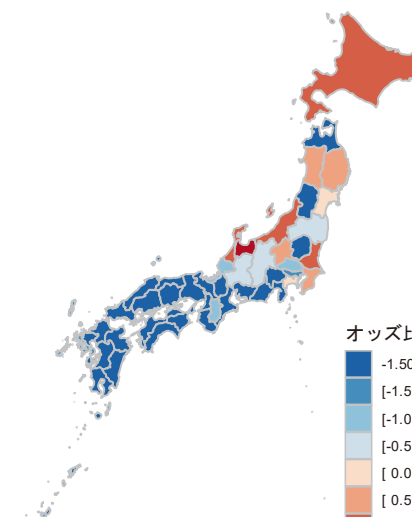


図 SI2 「代表的な怪異・妖怪」の出現パターン (承前) .