

歴史災害史料からの自動地名抽出に向けた 自然言語処理システムの性能評価

武内 樹治（立命館大学 文学研究科）

大内 啓樹（奈良先端科学技術大学）

東山 翔平（国立研究開発法人情報通信研究機構）

概要：歴史災害について記述された歴史史料から、地名や位置情報といった地理的情報を抽出し、データベースや地図に統合・可視化することで、災害の詳細な状況の調査・分析を高度化・効率化できると期待できる。本報告では、自然言語処理技術を用いて、近世の歴史災害に関する史料から地名を抽出する取り組みを報告する。人手で地名をアノテーションしたデータセットを作成し、地名抽出に利用可能な既存の自然言語処理システム、GiNZA および ChatGPT（GPT-3.5, GPT-4）の抽出精度の評価と誤り事例の分析を行った。GiNZA の現代語向けモデルでは十分な精度が得られず、GPT-4 では期待の持てる結果が得られたものの、コスト上の課題があることを確認した。今後は事前学習済みモデルをファインチューニングする方法などを試し、コスト・精度ともに実用性の高いモデルの実現を目指す。

キーワード：歴史災害、自然言語処理、地名、歴史史料

Performance Evaluation of Natural Language Processing Systems for Automatic Place Name Extraction from Historical Disaster Documents

Mikiharu Takeuchi (Graduate School of Letters, Ritsumeikan University)

Hiroki Ouchi (Nara Institute of Science and Technology)

Shohei Higashiyama (National Institute of Information and Communications Technology)

Abstract: In this paper, we describe a practice of using natural language processing technology to extract place names from Japanese historical documents related to historical disasters in the early modern period. We created a dataset manually annotated with place names, evaluated the extraction accuracy of existing natural language processing systems available for place name extraction, GiNZA and ChatGPT (GPT-3.5, GPT-4), and analyzed error cases. The model for modern languages of GiNZA did not provide sufficient accuracy. On the other hand, the GPT-4 model gave promising results, but we confirmed that there were some cost issues. In the future, we will experiment with methods such as fine-tuning of pre-trained models, and aim to realize a model that is practical in terms of both cost and accuracy.

Keywords: Historical Disasters, Natural Language Processing, Place Name, Japanese Historical Documents

1. はじめに

過去に発生した歴史災害—たとえば地震、火災、冷害、渇水、疫病、旱魃、洪水—について、歴史史料を読み解くことで、災害の発生日時、発生場所、被災の場所や広がり方の様子といった、災害の地理的・時間的な情報を明らかにできる可能性がある。また、多くの歴史災害に関する時空間情報をデータベース上に整備・集約することができれば、関連資料や情報、たとえば発掘調査で出土した災害の痕跡や絵図・古地図などと突き合わせて、統合的・多角的な分析をすることも容易となる。

こうした背景の下、歴史災害に関する史料のデータ化や分析に関する研究が進められている。渡邊らは、東京大学史料編纂所データベースの歴史史料より京都における火災に関する記録を収集し、位置情報の特定・地理情報システム（GIS）

データ化を行い、時空間分析によって当時の火災の発生地域や季節性などを考察している[1]。加納らは、『地震史料集テキストデータベース』から地名を抽出し、地理情報を付与している[2]。橋本は、翻刻された災害史料のテキストから構造化データを効率的に抽出する方法の開発を試みている[3]。その試作である「みんなでマークアップ」[4]は、安政江戸地震を対象に歴史史料に登場する情報を市民参加型でマークアップする仕組みであり、「場所」もそのマークアップの対象とされ、歴史災害の場所情報がデータ化されつつある。

大量の史料に含まれる情報を構造化し分析するためには、自然言語処理・固有表現抽出技術を活用し、地名や災害関連情報を自動抽出することが必要となる。そうした自動化を視野に入れている研究[3]もあり、現代語テキストを対象とする既存のシステムを、前近代のテキストに適用しても十分な精度が見込めない点も指摘されている。し

かし、既存の固有表現抽出システムを日本語前近代テキストに実際に適用し、災害関連情報の抽出精度や課題を評価・報告した研究は我々の知る限りない。したがって、前近代テキストの正解データを作成し、ゼロからモデルを学習させるの方法や、現代語向けのモデルを前近代テキストに適応させる追加の学習を行う方法などが考えられる中で、どの方法が最適であるかは明らかでない。

我々の研究の長期的なゴールとして、深層学習に基づく最新の自然言語処理技術を活用し、実用的な解析精度で、歴史史料から災害関連情報を抽出し、さらに位置情報を推定し、歴史地名データベースと紐づけるシステムを構築することを目標とする。その最初のステップとして、近世災害史料に含まれる歴史地名の抽出に焦点を当て、本研究では、正解データの構築と、それをを用いた既存の自然言語処理システム、GiNZA および ChatGPT (GPT-3.5, GPT-4) の精度評価と誤り事例の分析を行った。GiNZA の現代語向けモデルでは十分な精度が得られず、GPT-4 (5-shot 学習) では期待の持てる結果が得られたものの、コスト上の課題があることを確認した。地名を含む歴史災害の諸情報の抽出を試みている橋本の研究 [3] との違いとして、①人手アノテーションデータ作成における作業員間一致率の算出、②複数の自然言語処理システムの精度比較の点で、定量的な評価に重点を置く。

2. アノテーションデータの構築

2.1. 史料の選定

本研究の対象データとして、歴史災害に関する記述のある近世の史料を想定する。今回、正解データ作成のためのアノテーションを行うにあたり、翻刻済みのテキストデータが再利用可能な形式で公開されている「みんなで翻刻」の成果物である「みんなで翻刻データ」[5]の中から史料の選定を行った。具体的には、近世の災害史料である「諸国大地震大津波一代記」「信越地震水記」等のテキストデータを用いる (表 1)。

「みんなで翻刻データ」では、史料の原本の見開きページごとに 1 テキストファイルとなっており、原本の 1 行に対してテキストファイル上でも 1 行が当てられている。そのため、文や単語の途中で改行が施されている場合がある。アノテーション段階では、原本と見比べやすいためダウンロードしたままのテキストファイルを用いて行うこととした。なお、「みんなで翻刻」における翻刻の品質は、100 文字あたり 1.5 文字の割合で誤刻や表記の揺れが含まれていることが明らかになっている [5,7]。

2.2. アノテーション基準

本研究では、史料に含まれる災害関連情報の中で、地名、特に地域名の抽出に焦点を当てる。どのような地名の表現をアノテーションするかについて、「みんなで翻刻」のマークアップ作業マニュアル [8] で規定されている「場所」のマークアップ方針を基にアノテーション基準を定義した。

「みんなで翻刻」作業マニュアルの「場所」のマークアップ方針では、指示語を伴う場所表現、人名が含まれる場所表現、階層を伴う場所表現など、歴史史料に出現する場所表現のパターンに対してマークアップの方法を示している。本研究でも、同方針を踏襲しつつ、2 点異なる方針を設けた。1 点目は、本研究では場所を表す表現の中で最も重要度の高い地域名に焦点を当て、地形名や施設名は現時点では対象外とした点である (地域名、地形名、施設名の区別は、作業マニュアル [8] の定義に従う)。2 点目は、範囲を示す場所表現について、範囲表現全体 (例: 「[本所より深川まで]」) ではなく、個別の地名 (例: 「[本所]より[深川]まで」) にアノテーションを行う点である。これは、データベース化や検索などの応用を想定した場合、個々の地名を抽出・構造化の単位とする方が扱いやすいと考えたためである。

2.3. アノテーション作業

アノテーション作業は、テキストアノテーションツール brat [9] を用いて、選定した 5 史料に対

表 1 対象史料の概要

史料 ID	史料名	文字数	災害名 (地震・津波)
L000086	諸国大地震大津波一代記	31,340	宝永地震 (宝永 4 年 10 月 4 日) 伊賀上野地震 (安政元年 6 月 15 日) 安政東南海地震 (安政元年 11 月 4,5 日) 安政江戸地震 (安政 2 年 10 月 2 日)
L000105	諸国海辺地震津波書	13,157	安政東南海地震 (安政元年 11 月 4,5 日)
L000169	信越地震水記 全	64,256	善光寺地震 (弘化 4 年 3 月 24 日)
L000305	享保十三年江戸洪水記	23,987	
L001060	地震記 下	53,362	安政東南海地震 (安政元年 11 月 4,5 日)
計		186,102	

史料 ID・史料名・災害名 (地震・津波) は、「みんなで翻刻」の掲載資料の書誌情報リスト [6] に対応している。

して筆頭著者（作業員 1）が行った。作業員 1 は考古学・歴史 GIS を専門とする博士課程後期課程の大学院生である。作業の結果、アノテーションされた地名数は 5 史料合計で 980 件となった（表 2 の「地名数（行またぎ有）」）。

表 2 アノテーションデータの行数と地名数

史料 ID	行数	地名数 (行またぎ有)	地名数 (行またぎ無)
L000086	807	231	231
L000105	195	134	129
L000169	1,813	280	271
L000305	668	102	98
L001060	1,413	233	231
全体	4,896	980	960

2.4. 作業員間一致率

アノテーション基準および作業員 1 のアノテーション内容の妥当性を確認するため、別の作業員 2 にアノテーション作業を依頼し、作業員間一致率の測定と差分の分析を行った。具体的には、5 史料のうちの一部に相当する、アノテーション未実施のテキストに対して、作業員 2 によるアノテーションを実施した。作業員 2 は、日本史学(中世)を専門とする博士課程後期課程の大学院生である。本作業に用いたテキストは、いずれかの史料中の連続するページのまとまり 5 つ分に相当するデータであり、全体の 3.5% 程度の文字数のデータである。作業員 1 の作業結果から、100 件程度の地名が出現していることを想定して作業量を決定した。

作業員 2 名の地名アノテーション件数・一致数と、一致率として適合率 (P)、再現率 (R)、F1 値を表 3 に示す。適合率および再現率は、作業員 1 の結果を正解とみなしたときの値である。

表 3 アノテーション数と作業員間一致率

アノテーション数		一致数	P	R	F1
作業員 1	作業員 2				
104	148	82	0.554	0.788	0.651

再現率 (0.788) はある程度高いものの、適合率 (0.554) はやや低い。つまり作業員 1 のアノテーション結果の 8 割近くについて作業員 2 もアノテーションしている一方で、作業員 2 のアノテーション結果のうち作業員 1 もアノテーションしたものは半数強にとどまった。アノテーション数については作業員 2 の方が多く、作業員 1 の結果にアノテーション漏れがあることが示唆される。

作業員 1, 2 のアノテーションの相違点について調査を行った。分かりやすい形で地名が記述されている箇所では、アノテーション結果が一致していることが多かった。例えば史料 ID : L001060 では、災害の被害状況を箇条書きの形式で列挙し

ている箇所があり、該当する数ページについては作業員 1, 2 ともに出現していた地名 22 件に対して同一のアノテーションを行っていた。しかし、他の文書では両者とも単純な地名の取りこぼしがあった。他には、文章の中で地名を列挙している場合など、地名同士の区切り位置に相違がみられた。例えば、「伊勢崎町冬木町北新川芝井町京橋加治町」(史料 ID:L00086 の 52 ページ目[10]) という文章内で地名を列挙している箇所では、「北新川」は地名か、それとも「北新川芝井町」で一つの地名となるのかといった点でアノテーションが異なっていた。この例で示した「芝井町」は、「柴井町」の表記ゆれと思われるため、「[芝井町]」でアノテーションするのが正解であると考えられる。「北新川」は地形名とも考えられるが、今回対象としている地域名との区別が自明でない。したがって、地形名(川や山など、自然の地形と同等の地理的範囲を持つもの)と地域名(人工的に区切られた地理的範囲を持つ土地)をより明確に区別できるアノテーション基準を定めることが今後必要である。なお、今回は「北新川」については地形名とみなしている。

3. システム性能評価

3.1 データセット

作業員 1 がアノテーションを行った 5 史料のうち、L000086 を開発データ、L000105, L000169, L000305 の 3 史料をテストデータとした。本データには、文のような区切りの単位が付いていないため、行を単位とし、解析対象テキスト 1 行ごとにシステムに入力する設定とした。精度評価時、行をまたがってアノテーションされている地名は正解事例から除外した。したがって、テストデータ中の地名数は、表 2 の「地名数(行またぎ無)」の値を 3 史料について合計した 498 件となる。史料 L001060 については本実験では使用しなかった。

3.2 対象システムと実験設定

日本語の地名抽出が可能な複数の自然言語処理システムの精度評価を行った。具体的には、日本語自然言語処理ライブラリ GiNZA v5.1.3 [11] および対話型大規模言語モデル ChatGPT (OpenAI Chat completions API) を評価対象とした。

GiNZA では、日本語 BERT [12] モデルに基づく ja_ginza_bert_large を用いて固有表現抽出を行った。拡張固有表現ラベル[13]で出力される GiNZA の解析結果に対する後処理として、地域名に相当するラベル (Location_Other, GPE の各サブカテゴリ, Region の各サブカテゴリ, Postal_Adr ess) を本研究で用いた LOC_NAME に変換した上で評価を行った。ChatGPT では、GPT-3 (gpt-3.5-turbo-0613) および GPT-4 (gpt-4-0613) の 2 種類のモデルを用いた。各モデルについて、0-shot および 5-shot 学習 (in-context learning) による推

論を実行した。使用したプロンプトは図 1 に示したものであり、図 1 の 2 段目の「例題」(few-shot examples) 部分を削除したものが 0-shot、「例題」部分に 5 事例分の入力文・正解情報の対を当てはめたものが 5-shot に当たる。0-shot, 5-shot とともに、「入力文」の「(INPUT_SENTENCE)」の箇所には実際に入力テキストが当てはめられたものとなる。5-shot の「例題」には、開発データの先頭から地名が出現する 5 事例 (5 行) を採用し、いずれの推論実行でもこれらの事例は固定した。本プロンプトは、Han ら[14]が英語固有表現抽出に使用したプロンプトを参考に作成したものである。Han らのプロンプトとの差分は、出力対象と出力形式を本研究に合わせて変えている点と、同一文字列の地名を重複して出力する指示を加えている点である。ChatGPT の実行時ハイパーパラメータについては、出力の安定性を高めるため temperature=0 とし、その他ハイパーパラメータではデフォルト値を用いた。GPT-3 では同一実験設定 (shot 数) につき 3 回の実行を行い、GPT-4 では API 利用料金が高いことから、同一実験設定で 1 回のみ実行を行った。

指示文	Given the list of entity types ["LOCATION"], read the given sentence and find out all words/phrases that indicate the above types of named entities. Answer in the format ["entity_name","entity_type";"entity_name","entity_type";...] in the order of appearance without any explanation. Extract all instances even when named entities with the same expression appears multiple times. If no entity exists, then just answer "[]".
例題	Sentence: (EXAMPLE_SENTENCE1) Answer: (EXAMPLE_ANSWER_1) Sentence: (EXAMPLE_SENTENCE_2) Answer: (EXAMPLE_ANSWER_2) ...
入力文	Sentence: (INPUT_SENTENCE) Answer:

図 1 ChatGPT に入力したプロンプト

3.3 実験結果

各モデルの精度を表 4 に示した。F1 値では、GPT-3.5 (0-shot) が 0.14 で最も低く、次いで GPT-3.5 (5-shot), GPT-4 (0-shot), GiNZA が 0.28-0.30 程度で近い値となり、GPT-4 (5-shot) が 0.38 で最も高い精度を達成した。

2 つの GPT モデルで 0-shot と 5-shot 推論の結果を比較すると、5-shot にすることで適合率・再現率が最大 0.12 ポイント程度向上し、5 事例という少量でも例題をプロンプトに含めることの有効性を確認した (ただし GPT-3.5 では再現率が低下した)。特に GPT-4 (5-shot) では、約 0.64 という比較的高い再現率が得られた。指示文の調整や例題事例数の増加といったプロンプトの工夫によってさらに精度が向上し得る点を考慮すると、精度上は有望な結果と捉えられる。ただし、GPT-4 は API 利用料金などの推論コストが高いため、実用上の課題は残る。なお、GPT-3.5 の 3 回の実行結果については、実行ごとに多少の出力内容や精度の違いは生じたものの、表 4 の標準偏差 (sd) で示したように、精度上の大きな差は見られなかった。

今回の GiNZA の結果から、現代日本語の固有表現抽出向けに構築されたモデルをそのまま近世史料テキストに適用しても、十分な精度が得られないことを確認した。実用的な精度を得るには、近世史料にアノテーションを施した学習データを構築し、BERT 等の事前学習済みモデルをファインチューニングする方法が、確実性が高いと想定する。必要な学習データ量の調査は今後の課題である。

3.4 各システムの誤り事例

F1 値上位の 2 システム, GPT-4 (5-shot) と GiNZA の抽出結果について誤り事例の調査を行った。

GPT-4 (5-shot) では、宿や勘定所、山門などの施設名や、河川名である「犀川」など地形名が多くアノテーションされていた。その他、階層を伴う場所表現について、1 つの地名として抽出している場合と、複数の地名に分割して抽出している

表 4 各システムの精度 (GPT-3.5 については 3 回の実行の平均, 標準偏差 sd)

System	正例数	予測数	一致数	P	R	F1
GiNZA	498	674	174	0.258	0.349	0.297
GPT-3.5 (0-shot)	498	3,042	247	0.081 (sd 0)	0.496 (sd 0.0016)	0.140 (sd 0.00047)
GPT-3.5 (5-shot)	498	1,056	216	0.205 (sd 0.00082)	0.435 (sd 0.0034)	0.278 (sd 0.0012)
GPT-4 (0-shot)	498	1,279	259	0.203	0.520	0.292
GPT-4 (5-shot)	498	1,177	318	0.270	0.639	0.380

場合があった。また、動作の方向を表す助詞「江」(真仮名)について、「江」までを含めて地名として抽出しているものが多かった(例:「去ル十五日中ノ[条村江]参着仕」, 正解:[中ノ条村], 図2)。このように, GPT-4 (5-shot) は必ずしも地名の正確な範囲を特定できず, また本研究で対象外とした地形名・施設名も抽出していたことから, 適合率が 0.27 と低い値になったと考えられる。ただし, 正解地名の多くに対して, 部分的に範囲が一致する形で地名として検出できている傾向が見られた。

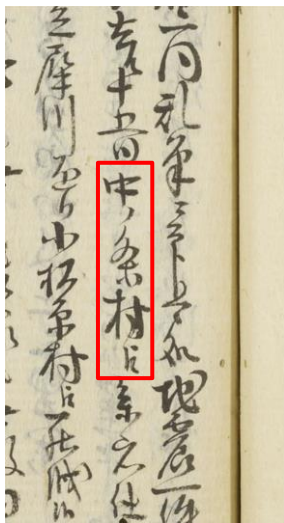


図2 地名の後に真仮名「江」のつく事例

GiNZA についても, GPT-4 (5-shot) と同じく, 施設名や地形名に相当するものが地名として検出されていた。施設名については, 地名が含まれる施設名(例:「西国橋」「嶋田宿」「小石川御門」)が抽出されている場合が多かった。その他, 「中」「辺」「之方」など, 地名の後に空間的な意味を補足するような単語があった場合に, それも含めて地名として検出している事例(例:「一九月二日には[江戸中]の高き所は」, 正解[江戸])や, 実在する地名と合致する文字列を地名として誤検出している事例があった(例1:「此子を助けて[呉]よといへは」, 例2:「西海枚並木の所山[大分]崩大地ほれ当分往還留り候」, どちらも地名でない)。また, 史料 ID : L001060 では, 「一」という文字の後に被害の内容を記し, 災害の被害状況を箇条書きで列挙している一つ書き箇所がある(例:「一宮一社 傷 一蔵一軒 潰」, 図3)。これに対し, 「一」を範囲に含めて地名として検出しているものが見受けられた(例:[一宮]一社 傷)。さらに, 1つの地名を複数に分解しているもの(例: 夫方[飯田][町]之方へ, 正解:[飯田町])や, 地名(例:「幡多郡」)を周辺の単語とつなげて地名ではなく自然災害名として検出しているもの(例:[幡多

郡惣破損縮」, 正解:[幡多郡])などがあった(この場合, 自然災害名は評価対象に当たらないため, 正解地名「幡多郡」に対する False Negative となる)。

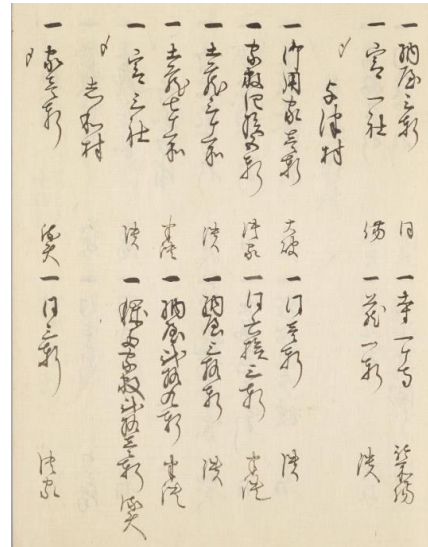


図3 一つ書きの事例

4. まとめと今後の展望

本研究では, 歴史災害史料に地名を人手アノテーションしたデータセットを用いて, 既存の自然言語処理システムの精度評価と誤り事例の分析を行った。GPT-4 での 5-shot 学習の実験では, ごく少量の学習データでもある程度の精度が達成できるという期待の持てる結果も得られたものの, 推論コストが大きい点が問題となる。

今後は, コスト・精度ともに実用性の高いモデルの実現を目指す。まず, 確実性の高い方法として, BERT 等の事前学習済みモデルをファインチューニング(追加学習)する方法を試す予定である。他にも, 翻刻済みテキストでは失われている文字画像の情報も活用できる可能性がある。たとえば, 真仮名の「江」が他の文字よりも小さく書かれていることや, 「一」(一つ書き)が行の先頭で箇条書きの記号として使われていることなどは, 古典籍画像上では容易に確認できる(図2, 3)。

その他の発展として, 抽出対象の情報を地域名以外の場所表現や災害情報に拡大することや, 抽出した場所の地理座標を特定するジオコーディングモデルを開発することを考えている。特に後者のジオコーディングについては, 地理情報処理基盤 GeoLOD[15]や『日本歴史地名大系』関連データセット[16]などのデータベ

ースと連携することで、地図上への表示や、GISを用いた高度な分析が可能となる。これらの取り組みはデジタル・ヒューマニティーズ分野におけるデータ駆動型研究の一例となり、関連分野における更なる研究の推進にも貢献すると考えている。

謝辞

「みんなで翻刻」での成果などについて加納靖之氏、橋本雄太氏、大邑潤三氏にご教授いただいた。またアノテーションの一致率比較には濱野未来氏にご協力頂いた。この場をお借りして御礼申し上げます。本研究は、JSPS 科研費 22H03648 の助成によるものである。

参考文献

- [1] 渡邊泰崇・塚本章宏・赤石直美・松本健太郎・吉越昭久・片平博文. GIS を用いた歴史災害の時空間分析—12 世紀平安京の火災を事例に—, 人文科学とコンピュータシンポジウム論文集, 131-138, 2007.
- [2] 加納靖之・大邑潤三. 地震史料集テキストデータへの地理情報の統合, 研究報告人文科学とコンピュータ (CH), 2023-CH-131(3), 1-3, 2023.
- [3] 橋本雄太. 歴史災害資料のマークアップシステムの試作, 研究報告人文科学とコンピュータ (CH), 2023-CH-131(2), 1-6, 2023.
- [4] みんなでマークアップ【安政江戸地震】. <https://markup.honkoku.org/welcome/#page-top>, (参照 2023-08-31)
- [5] みんなで翻刻データ. <https://github.com/yuta1984/honkoku-data>, (参照 2023-08-31)
- [6] みんなで翻刻データ 書誌情報リスト. <https://github.com/yuta1984/honkoku-data/blob/master/v1/entries.csv> (参照 2023-11-04)
- [7] 橋本雄太. 市民参加型史料研究のためのデジタル人文学基盤の構築, 京都大学大学院文学研究科現代文化学専攻博士論文, 2018. <https://repository.kulib.kyoto-u.ac.jp/dspace/handle/2433/233817>, (参照 2023-11-04)
- [8] みんなで翻刻ウィキ 「場所」のマークアップ. <https://wiki.honkoku.org/doku.php?id=annotation>, (参照 2023-08-31).
- [9] Pontus Stenetorp, Sampo Pyysalo, Goran Topic, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. brat: a web-based tool for NLP-assisted text annotation. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp. 102–107, 2012.
- [10] <https://github.com/yuta1984/honkoku-data/blob/master/v1/L000086/052.txt>, (参照 2023-11-04)
- [11] 松田寛・大村舞・浅原正幸. 短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習, 言語処理学会第 25 回年次大会発表論文集, 201-204, 2019.
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee,

Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171-4186, 2019.

- [13] 「関根の拡張固有表現階層」定義. https://nlp.cs.nyu.edu/ene/version7_0_0.html, (参照 2023-11-04)
- [14] Han, R., Peng, T., Yang, C., Wang, B., Liu, L., Wan, X.: Is Information Extraction Solved by ChatGPT? An Analysis of Performance, Evaluation Criteria, Robustness and Errors, *Computation and Language*, 2023.
- [15] 北本朝展. 地名情報基盤 GeoLOD による歴史地名の共有に向けて. 日本地球惑星科学連合 (JpGU) 2022 年大会, no.MIS22-04, 5 月 2022.
- [16] ニュースリリース 歴史的地名の「行政区画変遷」を大規模オープンデータ化～『日本歴史地名大系』を平凡社地図出版との協働により機械可読データとして強化～. <https://www.nii.ac.jp/news/release/2023/1018.html>, (参照 2023-11-04)