様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ８ 月 ２８ 日現在

機関番号：８４６０４
研究種目：若手研究(A)
研究期間：2016～2019
課題番号：１６Ｈ０５８８１

研究課題名（和文）日本考古学国際化のための考古学関係用語シソーラス構築と自動英語化の研究

研究課題名（英文）Internationalizing Japanese archaeology: Constructing a Japanese archaeology thesaurus equipped with the function to translate terminology into English automatically

研究代表者

高田　祐一（Takata, Yuichi）

独立行政法人国立文化財機構奈良文化財研究所・企画調整部・研究員

研究者番号：５０７０８５７６

交付決定額（研究期間全体）：（直接経費）　　14,800,000円

研究成果の概要（和文）：本研究では、文化財関係用語シソーラスおよび日英対訳データベースを構築する。全国の発掘報告書の全文データを格納している「全国遺跡報告総覧」システムを拡張開発し、文化財関係用語シソーラスを組み込むことで日本考古学の国際化に資することを目的とした。検索時の用語の類語自動変換、報告書ワードマップ（頻出用語俯瞰図）、各都道府県版　報告書特徴語ワードマップ、旧石器遺跡の遺跡関係性ネットワーク図などの機能を公開した。

研究成果の学術的意義や社会的意義
全国遺跡報告総覧システムが保有する19億文字のテキストデータに対し英語で検索できるようになり、海外研究者や市民が発掘成果にアクセスできるようになった。アクセスが困難な日本の文化財情報について言語のハードルをさげることで、日本研究のためのハードルを下げることが可能となった。また膨大な成果の提供を基盤にして、新しい需要の掘りおこしも可能となった。日本研究の進展は、日本文化の理解が進むことである。考古学分野では国外から国内に成果が還元されることで、今後日本考古学の発展につながると予想される

研究成果の概要（英文）：The aim of this project was to expand the functionality of the Comprehensive Database of Archaeological Site Reports in Japan (the full-text, nationwide database of excavation site reports) to facilitate the internationalization of Japanese archaeology. For this purpose, we constructed a thesaurus on cultural assets and a Japanese-English bilingual glossary. The system is now capable of automatically converting synonyms of search terms. Visualizations of the most common terminology nationwide, the characteristic terminologies of each prefecture, and the network of Paleolithic sites are available as well.

研究分野：人文情報学

キーワード：データベース　考古学　シソーラス　考古学ビッグデータ　発掘調査報告書　デジタルアーカイブ　多言語化　自然言語処理

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

日本考古学の国際化が叫ばれ、少しずつ論文の英語発表が増えつつある。しかし論文の根拠であるはずの発掘調査報告書の英語化は進んでいない。論文発表は研究成果であって、査読や結果検証のためにも、その根拠となった報告書について外国語でもアクセスできる必要があろう。日本考古学の蓄積は、先進的な研究事例があるにも関わらず世界的に十分に認知されているとは言い難い。日本考古学の国際化を推進し、情報発信の裾野を広げるために、外国語による研究・論文発表の次なる段階は、報告書の外国語対応であるといえる。

２．研究の目的

本研究は、日本考古学の基礎的学術成果である発掘調査報告書を英語化することで、海外研究者の日本研究へのハードルを下げ、日本考古学の国際化に資することを目的とする。報告書の英語化においては、報告書本文全文を翻訳することは現実的でない。利用者が必要とする報告書に言語の壁を越えてアクセスできることこそが重要である。必要とする報告書を探す際は、テキスト検索が一般的である。その際、検索時に入力するのは専門用語のキーワードである。そこで、専門用語について、日英対訳を整備し、テキスト検索時に英語→日本語で自動変換できれば、情報アクセスに資することができる。

本研究の情報資源は、全国遺跡報告総覧である。全国の報告書全文データ約 25000 件の報告書データが登録されており、日本語は約 19 億文字である。人間が１件ずつ読むことは不可能な量である。そのため、報告書の内容をある程度、自動で類推し、利用者が膨大な情報から適切に情報アクセスできる技術が必要である。本研究で開発する文化財関係用語シソーラスを組み込むことで、内容要約や報告書間の類似度算出に使用できる。

３．研究の方法

（１）文化財関係用語シソーラスの構築

文化財関係用語シソーラスとして専門語彙の辞書を作成する。通常の統計的自然言語解析で使用される辞書では、考古学・歴史学・文化財等の専門語彙をカバーしていないため、専用に作成する必要がある。辞書作成に当たっては、刊行物の辞書の見出し語や報告書本文から語彙を収集する。

（２）シソーラスを活用した研究

文化財関係用語シソーラスの語彙情報が 19 億文字のテキストデータを解析する際の基盤となる。頻出用語や特徴語の算出することができる。さらに共起関係の解析等が可能となる。

４．研究成果

（１）文化財関係用語シソーラスの整備

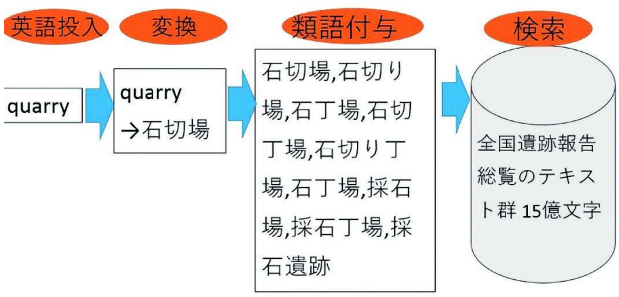
収集整理した用語数は次の通り。語彙数 ： 190228、英語対訳数 ： 8494、韓国語対訳数 ： 694、簡体字対訳数 ： 695、よみ数 ： 64791、類義語数 ： 5088、関連語数 ： 13381、説明数 ： 126948、表記ゆれ数 ： 59736 である。語彙に対する網羅率ではなく単純に登録件数である。

（２）検索時の自動変換機能

日本考古学の成果に関心を示す海外の研究者には、言語の壁や報告書を手にとって閲覧できないという、情報アクセスの問題がある。日本語を習熟しても、日本の考古学関係用語には多くの類語がある。遺跡総覧では、日英の考古学用語の対訳と日本語の考古学用語の類語をデータベース化し、英語自動変換機能を実装した、文化財関係用語シソーラスを構築した。英語の用語を投入するだけで日本語用語に自動変換し、類語を自動で付与する。簡単に言語の壁を超えて網羅的な検索が可能である。

（３）報告書ワードマップ（頻出用語俯瞰図）

現在遺跡総覧に登録されている報告書類 19 億文字に対して、どういった考古学関係用語が頻出しているかを可視化したもので、遺物関係（桃色）・遺構関係（黄色）・その他（青色）の 3 つの種別を付与している。土器に関する用語が頻出しているのが特徴的である。



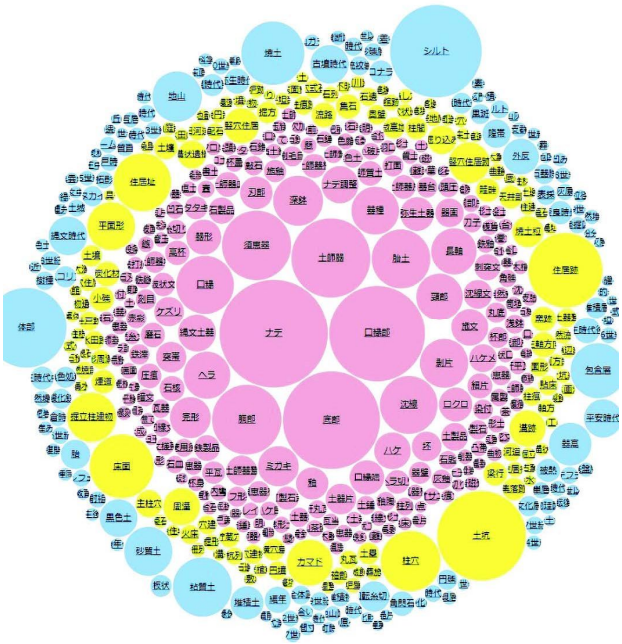
用語の英語自動変換機能

（４）各都道府県版　報告書特徴語ワードマップ

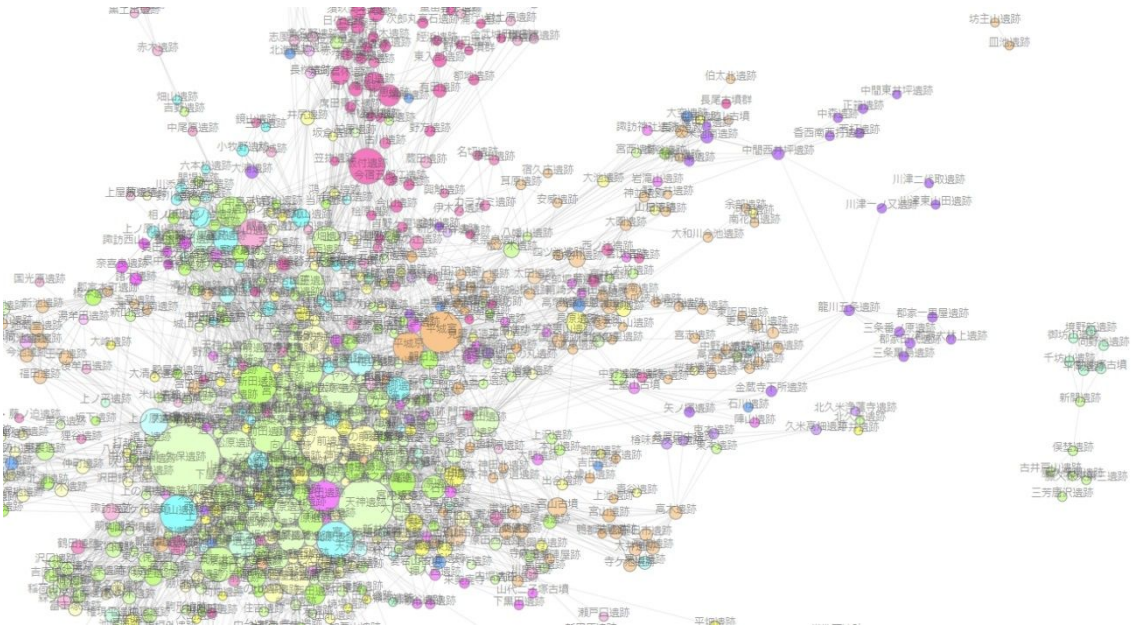
当該都道府県内にて頻出する用語、かつ他都道府県では出現頻度が低い希少用語は重要であることを勘案することで、当該都道府県の強い特徴を示す用語を可視化した。自然言語処理技術のベクトル空間モデルの TF（索引語頻度）と IDF（逆文書頻度）を組み合わせた TF-IDF にて算出している。その地域で特徴的な用語を提示することで、キーワードの入力に頼らない検索方法を提供している。専門用語を知らなくとも検索する事が可能になる。

（５）旧石器遺跡の遺跡関係性ネットワーク図

報告書では、当該遺跡への言及がある。それは当該遺跡を評価するために、周辺遺跡や類例などに言及するためである。この出現の組み合わせをカウントし、ネットワーク図として可視化した。すべての遺跡を対象にした場合、処理量が膨大となるため、旧石器遺跡で試行した。旧石器遺跡は、日本旧石器学会が公開しているデータベース『日本列島の旧石器時代遺跡』を活用した。現時点において、日本全国の報告書が遺跡総覧に登録されていないため、データとしては不備がある。また OCR の誤読や、同一遺跡名などうまく処理できないという課題がある。しかし、このネットワークの関係性を検索結果の表示順等に活用することで、確認すべき報告書の優先度設定などに資すると考えられる。



報告書ワードマップ（頻出用語俯瞰図）



旧石器遺跡の遺跡関係性ネットワーク図

５．主な発表論文等

〔雑誌論文〕　計9件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　3件）

１．著者名 高田 祐一	４．巻 1
２．論文標題 発掘調査報告書の電子公開による情報発信とその新たな可能性	５．発行年 2019年
３．雑誌名 デジタル技術による文化財情報の記録と利活用	６．最初と最後の頁 73-78
掲載論文のDOI（デジタルオブジェクト識別子） http://doi.org/10.24484/sitereports.33189	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 高田祐一	４．巻 337
２．論文標題 全国遺跡報告総覧における学術情報流通と活用の取り組み	５．発行年 2018年
３．雑誌名 カレントアウェアネス	６．最初と最後の頁 15-19
掲載論文のDOI（デジタルオブジェクト識別子） 10.11501/11161999	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 高田　祐一	４．巻 2017-CH-115(8)
２．論文標題 歴史的文字に関する既存知の集積と分析	５．発行年 2017年
３．雑誌名 研究報告人文科学とコンピュータ（CH）	６．最初と最後の頁 1-4
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 禾代 誠仁 ,高田 祐一 ,井上 幸 ,方 国花 ,馬場 基 ,渡辺 晃宏 ,井上 聡	４．巻 59(2)
２．論文標題 字形画像をキーとした情報検索による古文書デジタルアーカイブ活用への効果	５．発行年 2018年
３．雑誌名 情報処理学会論文誌	６．最初と最後の頁 351-359
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 高田祐一	４．巻 2
２．論文標題 歴史的文字に関する経験知のデータ化と共有手法	５．発行年 2016年
３．雑誌名 漢字字体史研究	６．最初と最後の頁 319-330
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 高田祐一	４．巻 13
２．論文標題 「全国遺跡報告総覧」プロジェクトの状況-発掘調査報告書のデジタルアーカイブの実現に向けて-	５．発行年 2016年
３．雑誌名 遺跡学研究	６．最初と最後の頁 188-191
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 高田祐一	４．巻 2016
２．論文標題 発掘調査報告書のデジタルアーカイブ	５．発行年 2016年
３．雑誌名 奈良文化財研究所紀要	６．最初と最後の頁 14-15
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Akihito Kitadai , Yuichi Takata, Miyuki Inoue, Guohua Fang, Hajime Baba, Akihiro Watanabe , Satoshi Inoue	４．巻 -
２．論文標題 A Web Based Service to Retrieve Handwritten Character Pattern Images on Japanese Historical Documents	５．発行年 2016年
３．雑誌名 JADH2016: “Digital Scholarship in History and the Humanities”	６．最初と最後の頁
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 高田祐一，金田明大，Dessislava Veltcheva	4．巻 2019
2．論文標題 Prospects and potential for the comprehensive database of archaeological site reports in Japan	5．発行年 2019年
3．雑誌名 The ARIADNE Impact	6．最初と最後の頁 175-185
掲載論文のDOI（デジタルオブジェクト識別子） http://doi.org/10.5281/zenodo.3476712	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計10件（うち招待講演　1件／うち国際学会　1件）

1．発表者名 高田　祐一・永島　幹大
2．発表標題 考古学ビッグデータの可視化技術とアクセス性向上の実践例-機械学習による画像認識と統計的自然言語処理技術を用いて-
3．学会等名 文化財方法論研究会
4．発表年 2018年

1．発表者名 高田祐一，昌子喜信，矢田貴史
2．発表標題 行政における文化財情報の電子化と発信：埋蔵文化財行政のデジタル技術活用の動向
3．学会等名 デジタルアーカイブ学会第3回研究大会
4．発表年 2019年

1．発表者名 高田祐一・国武貞克
2．発表標題 全国遺跡報告総覧における旧石器関係シソーラスの構築
3．学会等名 2017年度日本旧石器学会 総会・研究発表・シンポジウム
4．発表年 2017年

１．発表者名
高田　祐一

２．発表標題
歴史的文字に関する既存知の集積と分析

３．学会等名
第115回　人文科学とコンピュータ研究会　企画セッション「歴史的文字に関する情報と経験知の共有」

４．発表年
2017年

１．発表者名
高田　祐一

２．発表標題
全国遺跡報告総覧と考古学ビッグデータ

３．学会等名
奈良文化財研究所　第９回東京講演会「デジタル技術で魅せる文化財　奈文研とＩＣＴ」

４．発表年
2017年

１．発表者名
高田祐一

２．発表標題
全国遺跡報告総覧の公開とその役割について

３．学会等名
平成28年度 兵庫県埋蔵文化財調査成果連絡会（招待講演）

４．発表年
2016年

１．発表者名
高田祐一

２．発表標題
全国遺跡報告総覧の現況―システムの機能改善の状況―

３．学会等名
全国遺跡報告総覧シンポジウム 「文化遺産の記録をすべての人々へ！―全国遺跡報告総覧のメリットと公開までのハードル―」

４．発表年
2016年

１．発表者名
高田祐一,森本晋

２．発表標題
Publishing the Comprehensive Database of Archaeological Site Reports in Japan for the Purpose of Increasing Information Accessibility

３．学会等名
8th World Archaeological Congress（国際学会）

４．発表年
2016年

１．発表者名
高田祐一

２．発表標題
Prospects and potential for the national digital repository of archaeological site reports

３．学会等名
ADS workshop

４．発表年
2017年

１．発表者名
高田祐一

２．発表標題
Prospects and potential for the national digital repository of archaeological site reports

３．学会等名
The Digital Repository of Japanese Archaeological Site Reports:background and prospects for collaboration

４．発表年
2017年

〔図書〕　計2件

１．著者名	４．発行年
高田　祐一（担当:監修）	2019年
２．出版社	５．総ページ数
奈良文化財研究所	96
３．書名	
デジタル技術による文化財情報の記録と利活用	

１．著者名 高田　祐一（担当:監修）	４．発行年 2020年
２．出版社 奈良文化財研究所	５．総ページ数 245
３．書名 デジタル技術による文化財情報の記録と利活用2 オープンサイエンス・データ長期保管・知的財産権・GIS	

〔産業財産権〕

〔その他〕

全国遺跡報告総覧　報告書ワードマップ
https://sitereports.nabunken.go.jp/ja/visualization/term
全国遺跡報告総覧　報告書ワードマップ
https://sitereports.nabunken.go.jp/ja/visualization/term
全国遺跡報告総覧　考古学関係用語英語自動変換機能
http://sitereports.nabunken.go.jp/en
報告書ワードマップ
http://sitereports.nabunken.go.jp/ja/visualization/term

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考