

令和 6 年 6 月 5 日現在

機関番号：32710

研究種目：基盤研究(C) (一般)

研究期間：2020～2023

課題番号：20K00619

研究課題名(和文) 超長期運用を想定した文書型データベースによる言語ドキュメンテーションの研究

研究課題名(英文) A study of language documentation with development of a text-based database system for long-term use and maintenance

研究代表者

大矢 一志 (Ohya, Kazushi)

鶴見大学・文学部・教授

研究者番号：80386911

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究は、言語学者個人の研究活動を支援する言語ドキュメンテーション活動向けデジタル環境構築の一環として、言語資料向けデータ形式の策定と、文書型DBを導入した環境構築を目標とした4年計画の活動である。研究成果として、コロナ禍による世界の変化に対応した新たな価値観の誕生を目撃し、すなわち新しい研究目標の発見(ICOM-C1DOC2023で発表)と、これまでの研究成果として得られLREC2022で発表したデータ形式(MDM)はこの新たな世界観にも通用することを確認した。

研究成果の学術的意義や社会的意義

言語学者個人の研究活動、とりわけ言語ドキュメンテーション活動を支援するデジタル環境を構築するためにこれまで取り組んできたことは、コロナ禍による社会や研究が目標とする価値観の大きな変化においても、意義が失われていないだけでなく、今後の社会動向(とりわけGLAM, Digital GLAMにおいて)重要な知見・成果を持つものであったことが確認できた。

研究成果の概要(英文)：This study had aimed at establishing a digital environment to support personal research activities of language documentation especially with Document or Hierarchical Database system, e.g. MUMPS or MongoDB, or without RDB that has been a de facto database system, since we regard the RDB itself as a serious obstacle to realize a digital environment to support personal research activities of language researchers. As a result of this study, we have found that after the pandemic of Wuhan/China/New Corona-virus disease, targets of research communities around digitization activities such as Digital Libraries, Archives, Humanities, Museums, etc. changed rapidly and determinedly, i.e. all the activities will become the same that has not have the name but could be called Digital GLAM(or D-GLAM). As one of the result of this study, a new data format MDM has been proposed at LREC 2022, and which seems to be a useful data format even in coming new targets in D-GLAM communities.

研究分野：言語学，人文情報学，電子資料論

キーワード：言語ドキュメンテーション Digital Humanities D-GLAM ユカギール語 イテリメン語

1. 研究開始当初の背景

言語ドキュメンテーション活動は、人文学研究にコンピュータ利用が浸透してきた時代の変化と共に、2000年代には研究活動中一貫して対象資料をコンピュータ上で扱うことが期待されてきた。ところが、そのようなデジタル環境上で閉じたデータ管理を可能にする、言語研究者個人の研究活動を支えるシステムは提案されていない。現状は、アーカイブやレポジトリという社会基盤(システム)で利用されるデジタル環境が構築されるのみで、それらに提供される言語資料を作成する言語研究者個人の活動を支えるシステムは提案されて来なかった。言語研究者は、特定アプリケーションで作成されたファイルを単位としたデータ管理に留まっているのが現状である。まさにこの現状が「言語ドキュメンテーション」といういち領域を言語学で確立させた背景にもなっている。本研究では、個人研究者の活動を長期に支援するデジタル環境の構築を目指し、その中の一つの要素技術であるデータベースそのものの選択を課題・解決すべき問題と捉えた。

データベース利用を歴史をみると、現在主流の関係型データベース(RDB)よりも、文書型データベース(文書型DB)の方が開発された時期は古く、また共に現在まで現役で使い続けられてきたことから、使用年限も文書型DBの方が長い。例えば、文書型DBのMUMPSは、1966年に開発されてから今日まで、病院のカルテなど文字を処理対象とした場面で使われ続け、また世界規模のクレジットカード会社の基幹システムとしてはMUMPSの後継としてあるgtmが使われている。今日DBの主流であるRDBとこの文書型DBを比べてみると、文書型DBはRDBと比べて、①柔軟にデータ構造が変更できること(schema-less)、②蓄積したデータを永く使えること(歴史上、RDBよりデータ寿命は長い)、③テキスト分析に長けている(RDBは数値処理が得意)、という特徴があり、言語資料を扱うデータベースとしてRDBよりも相応しい特徴を備えている。但し、現在のDBの主流はRDBであり、文書型DBに関する情報は人文学のみならず世間一般にもその利用に関する情報は流通されていない。まして言語学については皆無の状況である。

2. 研究の目的

本研究の目的は、既に医療情報データベースとして実績のある文書型DBであるMUMPSを言語研究向けに導入するために必要な性能評価と利用環境を開発することにある。言語研究者向けに文書型DBを取り入れた言語ドキュメンテーション活動を支援するデジタル環境の構築に取り組む活動を知らないことから、本研究は、利用者、使用する技術、完成・運用される環境など、全てのの点で新たに検討を始めるという、実用を目的としながらも、基礎研究の色濃い研究活動となる。

3. 研究の方法

本研究では以下の具体的な活動を計画している。

1. 言語ドキュメンテーションの実践
2. 動向調査
3. データ構造の検討・評価
4. デジタル環境選定の検討・評価
5. デジタル環境の構築
6. データベース利用のサポート環境作成ならびに利用案内資料の作成

1) 言語ドキュメンテーション活動として、シベリア地方におけるフィールド調査と、そこで得られた言語資料をデジタル環境上で整理する活動を実施する。また言語研究に相応しいデータベースの策定のため、言語ドキュメンテーションの実践活動からの要求に基づく性能評価を行う。2) 同時に、作成された言語資料は共時的・通時的共有を目指すものであることから国際的な研究コミュニティの動向調査は欠くことができない。規格化されたデータ形式であっても、採用されなければ価値はなく、また大型プロジェクトであっても継続されなければ、それに合わせる意味はない。本研究は、業績リストの項目を増やすことを目的とするものではなく、実質的な成果として言語研究者個人の研究活動を支援する成果を求めることから、国際動向に敏感であることが求められる。従って、頻繁かつ継続した動向調査は欠かせない。

これらの基礎活動を元に 3. データ構造・モデル評価, 4. データベース環境評価, ならびに 5. その構築, さらには 6. 完成したデータベースのサポート環境の整備とその利用を促す解説類資料の作成に取り組む。

4. 研究成果

3 年計画の初年度 (2020 年) に新型コロナウイルスによる活動自粛と海外渡航の禁止により, 研究計画書あった内容は大幅に変更せざるを得なかった。また当初の 3 年計画を 1 年延長, 都合 4 年間の研究活動となった。

言語ドキュメンテーション活動では, 2020 年から始まるコロナ禍による海外渡航禁止と, その後のロシアによるウクライナ戦争により, シベリア地方をフィールドとする本研究のフィールド調査は実施できなかった。代わりに, 以前の科研費事業で取り組んできたフィールドワークで得られている言語資料を, デジタル環境上で整理するドキュメンテーション活動に取り組み, ユカギール語では補部節のアノテーションを充実, supine という動詞形式の用法とその歴史的变化を記述, さらにはフォークロアテキストに形態分析と多言語訳 (英語, 日本語) アノテーションを加えたデータを印刷物ならびに web 上で公開した。イテリメン語では, 統語情報のアノテーションを充実させ, FLEx 上でのデータ整理に取り組み, それらの成果論文は言語学で最も権威ある国際雑誌 "Language" に採択・掲載された。

データ構造の検討・評価については, LREC2022 年でその概要と定義を発表, 同時にその利便性を体感できる Python モジュールを pip でインストールできる環境で公開した。LREC の発表論文は多いことから, 発表論文への評価は必ずしも体感できるものではないが, 査読では極めて高い評価を受けた。

動向調査としては, 副産物的ではあるが重要な知見として, オンライン会議は論文発表程度の情報しか得られない, 時間縛りのある非効率的な方式であること, 一方で移動を伴う非効率的とされてきた対面会議では, 発表に付帯する情報共有の有効性であることが改めて確認された。これで, これまでの科研費事業で取り組んできた, 発表すること無しでも国際会議に参加し, 情報収集するという取り組み・姿勢は正しかったことが確認できた。また, 2022 年に参加した 3 つの国際会議 (言語処理系の LREC2022, 博物館学系の ICOM2022, G-GLAM 系の DLF2022) の会議から, 日本国外, とりわけアメリカでは, コロナ禍においても研究活動やさらに社会活動は停滞することなく, むしろある面では加速化していた現実を知ることになった。日本はただただ研究活動も含めて社会活動を停止させていた期間に, 言語ドキュメンテーション活動, 広くは人文学の電子化活動は世界では進められていた。この領域でコロナ禍前にすでに周回遅れにあった日本は, もう追いつきもできないほど欧米から取り残された事実を 2022 年の動向調査で知ることになった。この結果から, もはや日本における研究活動は, 根本的な問から始める基礎研究から取り組み直す研究を始めることでしかもはや欧米の研究には肩を並べることができないと考えた。この反省による新たな研究目標の設定は, 本研究に続く科研費事業のテーマとなり, 幸運にも, 2024 年度から新規 3 年計画で取り組むことになった。

策定したデータ構造と, 国際動向調査の結果をもとに, 文書型 DB を使う言語研究者個人の研究活動で使うデジタル環境の策定・構築の取り組みについては, 文書型 DB として MUMPS と MongoDB を対象に, その環境構築の実態, 特性, 本事業の成果として策定したデータ構造の記述・入力方法の検討に取り組み, 未発表ながら成果報告はできるまでのそれなりの情報はまとめられたが, 先に記した 2022 年での動向調査から, (1) もはや言語ドキュメンテーションという狭い領域だけで個人研究者のデジタル環境を策定できる時代ではなくなり, D-GLAM と呼ぶべき, より広い視点での検討が必要であること, またそのようなより応用対象が広がったとしても, (2)LREC2022 での提案したデータ形式とその利用モデルの有効性は変わらないことも確認できた。従って, 本研究の大きな成果は, (a) コロナ禍による価値観の変化を動向調査から知り, (b) これまでの科研費事業や本研究で検討してきたデータ形式の成果の有効性を発表・確認できたことである。

5. 主な発表論文等

〔雑誌論文〕 計11件（うち査読付論文 5件 / うち国際共著 0件 / うちオープンアクセス 9件）

1. 著者名 Kazushi Ohya	4. 巻 1
2. 論文標題 An Architecture of resolving a multiple link path in a standoff-style data format to enhance the mobility of language resources	5. 発行年 2022年
3. 雑誌名 Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)	6. 最初と最後の頁 2865-2873
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 オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 長崎郁	4. 巻 13
2. 論文標題 北東シベリア諸言語の数詞における加法の表現	5. 発行年 2023年
3. 雑誌名 北方言語研究	6. 最初と最後の頁 171-191
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 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 Chiikako Ono	4. 巻 1
2. 論文標題 Subgrouping of Paleoasian Languages	5. 発行年 2022年
3. 雑誌名 Linguistic Atlas of Asia and Africa	6. 最初と最後の頁 3-3
掲載論文のDOI (デジタルオブジェクト識別子) 10.5281/zenodo.7118188	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 Chikako Ono	4. 巻 1
2. 論文標題 'Mouse' in Chukotko-Kamchatkan	5. 発行年 2022年
3. 雑誌名 Linguistic Atlas of Asia and Africa	6. 最初と最後の頁 32-32
掲載論文のDOI (デジタルオブジェクト識別子) 10.5281/zenodo.7118188	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 Chikako Ono	4. 巻 2
2. 論文標題 'Wheat' in Chukotko-Kamchatkan	5. 発行年 2023年
3. 雑誌名 Linguistic Atlas of Asia and Africa	6. 最初と最後の頁 5-5
掲載論文のDOI (デジタルオブジェクト識別子) 10.5281/zenodo.7754469	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 Chikako Ono	4. 巻 2
2. 論文標題 Stop series in Chukotko-Kamchatkan	5. 発行年 2023年
3. 雑誌名 Linguistic Atlas of Asia and Africa	6. 最初と最後の頁 154-155
掲載論文のDOI (デジタルオブジェクト識別子) 10.5281/zenodo.7754469	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 長崎郁	4. 巻 12
2. 論文標題 コリマ・ユカギール語の「迂言的過去」	5. 発行年 2022年
3. 雑誌名 北方言語研究	6. 最初と最後の頁 167-183
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 小野智香子	4. 巻 187
2. 論文標題 イテリメン語のアスペクト接辞 -skne	5. 発行年 2022年
3. 雑誌名 北海学園大学学園論集	6. 最初と最後の頁 35-41
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 Kazushi Ohya	4. 巻
2. 論文標題 An online course system easy to make, preserve, and promote critical thinking	5. 発行年 2020年
3. 雑誌名 DH2020 Book of Abstracts	6. 最初と最後の頁 PDF版編集中
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 長崎郁	4. 巻 11
2. 論文標題 コリマ・ユカギール語の Supine : 統語機能と言語接触	5. 発行年 2021年
3. 雑誌名 北方言語研究	6. 最初と最後の頁 17-35
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 小野智香子	4. 巻 22
2. 論文標題 イテリメン語におけるロシア語由来の動詞	5. 発行年 2020年
3. 雑誌名 千葉大学ユーラシア言語文化論集	6. 最初と最後の頁 67-76
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 オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計4件 (うち招待講演 0件 / うち国際学会 2件)

1. 発表者名 Iku Nagasaki
2. 発表標題 On the contact of Kolyma Yukaghir with Russian
3. 学会等名 7th ANNUAL MEETING of the Slavic Linguistics Society (SLS-17) (国際学会)
4. 発表年 2022年

1. 発表者名 長崎郁
2. 発表標題 コリマ・ユカギール語の「迂言的過去」
3. 学会等名 日本北方言語学会 第4回大会
4. 発表年 2021年

1. 発表者名 長崎郁
2. 発表標題 19世紀末以降のコリマ・ユカギール語における文法変化: supine の使用頻度から
3. 学会等名 「シベリア先住民諸語の歴史と類型」科研費主催研究会
4. 発表年 2021年

1. 発表者名 Chikako Ono
2. 発表標題 Aspectual suffix -skne in Itelmen
3. 学会等名 The second Conference on Uralic, Altaic and Paleoasiatic languages (国際学会)
4. 発表年 2020年

〔図書〕 計1件

1. 著者名 Iku Nagasaki	4. 発行年 2023年
2. 出版社 科研費基盤研究 (C)「シベリア先住民諸言語のテキスト トコーパス構築と文法及びその構造的変化に関する研究」、科研費基盤研究 (B)「シベリア先住民諸語の方言に関する基礎的研究と語彙データ ベース構築」研究班	5. 総ページ数 274
3. 書名 Kolyma Yukaghir Texts	

〔産業財産権〕

〔その他〕

-

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
研究分担者	小野 智香子 (Ono Chikako) (50466728)	北海学園大学・工学部・教授 (30107)	
研究分担者	長崎 郁 (Nagasaki Iku) (70401445)	名古屋大学・人文学研究科・特任講師 (13901)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関