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

平成　２９　年　　６　月　１３　日現在

機関番号：　３２７１０
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2016
課題番号：　２６３７０５１２

研究課題名（和文）シベリア少数言語コーパス構築に向けた理論的・実践的ドキュメンテーション研究

研究課題名（英文）A study of language documentation for establishing a corpus system of endangered languages in Siberia

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）　　3,500,000円

研究成果の概要（和文）：言語ドキュメンテーションについて、シベリアの少数言語を対象としたフィールド調査やデータ処理ソフトの開発といった実践的な面と、言語資料に相応しいデータ形式の研究やデータ変換についての理論的な研究の、２つの側面を同時に実践する総合的な研究に取り組む。なお，この手法は人文情報学でその効果が確認されている新しい学際研究のスタイルである．また、国内外の動向調査にあたり、理論や仕様の策定にそれらを反映させ将来の実用的な活動にも配慮した基礎研究に取り組む。また同時に、言語ドキュメンテーションそのものの見直しにも取り組む。

研究成果の概要（英文）：This study is a synthetic research with practical and theoretical activities, which has been regarded as an ideal research style but hard to be implemented, but is now recognized as a new way of collaboration style in an academical community of Digital Humanities. As a practical activity, we engaged in fieldwork at Siberia, Russia, and implementing an experimental software for data conversion and sound data handling. As a theoretical activity, we sought an appropriate data format that ensures bi-directional reference between sound data and text data, and a way of data conversion. In addition to them, we scrutinized the format itself with information gathered from international conferences and activities of international institutions, and checked a process of the language documentation itself.

研究分野：　言語学 Digital Humanities

キーワード：　言語ドキュメンテーション　危機言語　コリマ・ユカギール語　イテリメン語　セリクープ語　データ変換　時間情報

１．研究開始当初の背景

本研究を申請した当時の背景として，当該時点までに取り組んできた言語ドキュメンテーションの研究(科研費課題番号 23240125)から，言語ドキュメンテーションのアーカイブ・レポジトリの多くが国際規格をベースに提案されているが，その肝心の国際規格が言語研究者の要求を十分に分析したものではなかったことが分かった．具体的には，公開方法の検討はされてはいても，公開の対象となる言語データそのものの作成に関する研究・分析は十分ではないことがわかった．そこで本研究では
(1)危機言語のフィールド調査で得られた音声データとそれを文字化したデータとの関連付けの手法を探ること，
(2)フィールド調査におけるデータの収集・記録からその公開までをより簡易に行うことのできる環境を探ること，
を大きな目的とした．具体的には，音声データをテキストデータと連携をしたときの関連性の保証方法と，言語データがフィールドで記録される段階から，最終的に公開されるまでの様子を分析し，そこから必要となるシステムの仕様書案をまとめることを目標とした．

２．研究の目的

上記にある当初の背景から本研究を始めた当初は，(1)音声データとテキストデータの関連性の研究と，(2)データ作成手順を反映した仕様書案の作成を研究目標としたが，後述するように，本研究の成果として，音声データとテキストデータの関連性をとるデータ書式を策定し，それが国際学会で評価され，さらに試験的なソフトを開発しその実効性を確認できたことにより，(1)の目標は一定レベルの達成をみたことから，(2)の目標として掲げたシステムの仕様書案を作るという，いわばすぐにできることに取り組むのではなく，長期的な基礎調査を視野に入れ、(2)の目標にあるもう一つの研究ポイントである(3)「言語ドキュメンテーションの工程そのものの観察」に焦点を当て，その分析を新たな目標として，このプロジェクト期間内に追加設定をした．

３．研究の方法

上記の研究目標を達成するために，具体的には 7 つの活動に取り組んだ(申請書にある計画上の手法ではなく，研究の進行に伴い修正された上記の目標達成のための取り組み)．

(1)言語ドキュメンテーション活動の実践，
(2)言語資料向けデータフォーマットの動向調査，
(3)時間情報表現の書式の策定，
(4)上記書式の実行可能性の検証，
(5)データ交換/変換の仕組みの検討，
(6)上記の実証的な検討を支援するツール開発，
(7)言語ドキュメンテーションの工程そのものの分析．

４．研究成果

(1) 言語ドキュメンテーションの実践活動は，研究分担者である長崎と小野が担当している．
長崎は，これまでのフィールド調査から得られたコリマ・ユカギール語の民話などの原文テキストを FLEx（FieldWorks Language Explorer）に取り込んで形態素分析し，形態情報，露訳，和訳の注釈付けを行った．またセリクープ語ナリム方言に関しては，母語話者への調査で得られた文例を FLEx に取り込み，分析と注釈の付与を行った．またセリクープ語ナリム方言（ロシア連邦トムスク州）のフィールド調査を行い，その母語話者と共に，話者自身が母語で文を書く際の表記法について検討をかさね，いくつかのテキスト（ロシア語訳を含む）のチェックを行った．

小野は，これまでのフィールド調査から得られたイテリメン語の言語データを FLEx のデータとして作成し，さらに ELAN を使い音声や映像による言語の一次資料からテキストを文字データとして起こし，音声とテキストデータに時間情報を付与することで，言語資料の価値を高めた．また，これまでに使用したフィールドノートを電子化し，その中から言語データと関連情報を文字データとして変換した．フィールド調査としては，ロシア連邦カムチャツカ地方チギリ地区においてイテリメン語の調査にあたり，文法項目の調査，母語話者によるテキストデータおよびロシア語訳を作成した．この調査で得られた音声・映像データは，ELAN を使い文字データに起こし，音声とテキストデータに時間情報を付与することで，その言語資料の価値を高め，また FLEx を使い形態素分析を行った．

(2)動向調査については，DH2014，JCDL2014，TEI2014，ICLDC4，WWW2015，DHRA2015，DLF2015，LREC2016，DH2016，ARS Conference 2016，DocENg2016，ICLDC5 に参加し，最新の研究動向を調査した．このうち，TEI2014，ICLDC4，LREC2016，ICLDC5 では成果発表もしている．本研究の事業期間における新しい動向としては，大型アーカイブ・レポジトリプロジェクトは，当初の計画通りには運用されなかったこと，その背景としてはデータ提供者へのサービスが明らかに不足していたこと，ISO や TEI 等の国際規格を採用する動きが少ないこと，組織的なデータ集積よりも個々の研究者の活動支援や，得られた言語デ

ータをどのように当該言語コミュニティへ還元してゆくのか，その手法や手順を計算機に限定せずに模索する研究活動が盛り上がってきたこと，などを特記することができる．

(3)時間情報を表現する書式については，GIST(General Information of Sub-Time for Linguistics)を策定し，これを ICLDC4 で発表，同時に(4)の実行可能性を検証するためのツールを開発し，公開した．時間表現の書式についてはこれを使うこととし，続いて必要となる検討課題は，ドキュメンテーション全体の工程における当該書式の位置づけの検討となった．

(5)言語データの変換サービスが重要であることは，先の科研費事業(科研費課題番号23240125)で確認した項目であるが，初年度(2014 年)に，その実験を支援するソフト，具体的には SGML 系の各種のマークアップ言語の書式に対応可能なパーサを Java で実装した．これによりオジリナルの文法で書かれたマークアップデータも処理できるようになった．なお，このパーサで出力されたオリジナルのデータ形式によるデータを元に，XPath を独自に改良した書式を使うことで，データを変換するツールも同時に開発している．本研究の初年度に，このツール開発が修了したことから，(6)のツール開発は一応の達成を見たと評価し，続いて必要となる検討課題として，言語ドキュメンテーション全体の工程を観察し，それに必要なデータ変換の機能を検討することにした．これは，(3)で挙げたものと同じ研究活動が必要となることとなった．

(7)上記に挙げられたように，本研究では最終的には，言語ドキュメンテーションの工程そのものの分析が重要であることを確認し，その初期研究に取りかかった．その結果として「記録が生まれる課程のモデル化」が必要であることがわかった．この内容は LREC2016 で発表をしている．また，この研究をいっそう深めるために，科研費に応募し，H29(2017)年度の基盤研究(C)として採択された．

５．主な発表論文等

〔雑誌論文〕（計 10 件）

Kazushi Ohya, Data Formats and Management Strategies from the Perspective of Language Resource Producers -- Personal Diachronic and Social Synchronic Data Sharing --, Proceedings of the Tenth International Conference on Language Resources and Evaluation(LREC2016), 査読有り, 2016, 3243-3248

Iku Nagasaki, Relative clauses and nominalizations in Kolyma Yukaghir, Crosslinguistics and Linguistic Crossings in Northeast Asia: Papers on the Languages of Sakhalin and Adjacent Regions, 査読有り, Vol.117, 2016, 137-151

小野智香子, 「イテリメン語における動詞の名詞化について--派生名詞と準動詞の名詞性--」『ひろがる北方研究の地平線：中川裕先生還暦記念論文集』, 査読なし, 2017, 95-103

Kazushi Ohya, Corpus Sharing Strategy for Descriptive Linguistics, Journal of Japan Association of Digital Humanities, 査読有り, Vol.1, 2015, 68-85

長崎郁, 「コリマ・ユカギール語の非定型節における能格性」『北方言語研究』, 査読有り, 6 巻, 2016, 25-42

Chikako Ono, Labile verbs and their argument structure alternations in Itelmen, Acta Linguistica Petropolitana, 査読有り, Vol.13, 2017,(印刷中)

長崎郁, 「コリマ・ユカギール語の否定と他動性」『北方言語研究』, 査読有り, 5 巻, 2015, 15-24

長崎郁, 「北東ユーラシア諸言語における否定構造」『北方言語研究』, 査読有り, 5 巻, 2015, 1-4

小野智香子, 「イテリメン語の否定の構造」『北方言語研究』, 査読有り, 5 巻, 2015, 39-53

小野智香子, 「イテリメン語西部語南北方言とチュクチ・コリャーク諸語--語彙から見た接触・系統関係の再検討」『千葉大学ユーラシア言語文化論集』, 査読なし, 16 巻, 2014, 217-230

〔学会発表〕（計 13 件）

Kazushi Ohya, Two Main Contents in a Syllabus for Language Documentation: the Learning Data Models and an Assignation of Data Conversion, the 5th International Conference of Language Documentation and Conservation(ICLDC), 2017 年 3 月 2 日, University of Hawaii (Honolulu, USA)

長崎郁, 「ユカギール語資料に見られる動物のイメージ」, 北方の言語と文化にかんするシンポジウム「北方の人と動物」, 2017 年 1 月 28 日, 北海道大学（北海道札幌市）

Iku Nagasaki, "Ergative in Kolyma Yukaghir Niominalizations", International Conference Northeast Asia and the North Pacific As Linguistic Area, 2015 年 8 月 20 日 21 日, 北海道大学 (北海道札幌市)

Iku Nagasaki, "The Structural Properties of Kolyma Yukaghir's Kakarimusubi-like Construction, International Workshop Kakarimusubi from a Comparative Perspective, 2015 年 9 月 5 日 6 日, 国立国語研究所 (東京都立川市)

Kazushi Ohya, "Unit-based Scheme Connection Between TEI and Original Scheme To Promote Data Sharing Beyond Cultual Diversities" TEI 2014, 2014 年 10 月 22 日, Northwestern University (Evanston, USA)

Kazushi Ohya, "A General Format for Time Information to the First-class Data of General Linguistics", ICLDC4, 2015 年 2 月 27 日, Ala Moana Hotel (Honolulu, USA)

長崎郁, 「ユカギールごとユカギールの人々」, 言語で巡るシベリアの旅--北方の人々のことばと暮らし, 2014 年 7 月 5 日, 新潟大学 (新潟県新潟市)

Iku Nagasaki, "Relativization and nominalization functions of JE verb forms in Kokyma Yukaghir", The conference System Changes in the Languages of Russia, 2014 年 10 月 17 日, Institute for Linguistic Studies, Russian Academy of Sciences (Sankt-Petersburg, Russiya)

長崎郁, 「コリマ・ユカギール語の否定と他動性」, 日本言語学会, 2014 年 11 月 16 日, 愛媛大学 (愛媛県松山市)

Iku Nagasaki, "Nominalization and related functions in Kolyma Yukaghir", International Symposium (HALS Field Seminar 2), 2014 年 11 月 29 日, University of Helsinki (Helsinki, Finland)

長崎郁, 「V.I.ヨヘリソンのユカギール語テキスト」, 北方の言語と文化にかんする国際シンポジウム, 2015 年 1 月 24 日, 北海道大学 (北海道札幌市)

小野智香子, 「イテリメンごとイテリメンの人々」, 言語で巡るシベリアの旅--北方の人々のことばと暮らし, 2014 年 7 月 5 日, 新潟大学 (新潟県新潟市)

小野智香子, 「イテリメン語の否定と法」, 日本言語学会, 2014 年 11 月 16 日, 愛媛大学 (愛媛県松山市)

〔図書〕(計 4 件)

大矢一志, 神奈川新聞社, 『人文情報学読本 胎動期編 Digital Humanities Reader - The Quickening Period』, 2017 年, 175

永山ゆかり, 長崎郁(編者), 東海大学出版部, 『シベリア先住民の食卓』, 2016 年, 220

小野智香子(執筆者), 東海道大学出版部, 『シベリア先住民の食卓』, 2016 年, (10-16, 82-86, 140-145)

長崎郁(執筆者), 三省堂, 『明解言語学辞典』, 2015 年, (41, 71, 164-165, 225)

６. 研究組織

(1)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

(2)研究分担者
長崎郁 (NAGASAKI, Iku)
国立国語研究所・理論・対象研究領域・非常勤研究員
研究者番号：70401445

(3)研究分担者
小野智香子 (ONO, Chikako)
千葉大学・文学部・特任研究員
研究者番号：50466728