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　４ 年　５ 月 ２０ 日現在

機関番号：１０１０１
研究種目：基盤研究(B)（一般）
研究期間：2018～2021
課題番号：１８Ｈ００６６１

研究課題名（和文）研究職を離れた言語研究者が保持する言語データの適正再資源化のための基盤確立研究

研究課題名（英文）Fundamental studies for proper utilization of linguistic data preserved by retired linguists

研究代表者
加藤　重広（KATO, Shigehiro）

北海道大学・文学研究院・教授

研究者番号：４０２８３０４８

交付決定額（研究期間全体）：（直接経費）　　13,200,000円

研究成果の概要（和文）：記述言語学のデータが現在喪失の危機にあることを踏まえて，その保存と利活用の方策を広範に開発すること，また，基礎的方針に従って，実際に言語データを保存する作業を進めることが，本研究における二本の柱となる。前者は，どのようにデータは危機にあり，どのような方法で喪失の危機を回避できるかを広く実態調査することと，法務的な問題や研究倫理上の問題が生じないかを検討することに，後者は，実際にデータをデジタル化して保存していく作業の中で，どのような手順や工程が必要となり，それらにどの程度の資源が必要かを明らかにすること，将来的に利活用に供しやすい形式や体制を構築するには何が必要かの解明に主眼がある。

研究成果の学術的意義や社会的意義
本研究では，実際に数件の言語データの喪失を回避し，電子データ化しており，一定の制約の中で利活用できる状態に整備している。同種の問題は，国際的に生じているが，日本における言語研究に特異な問題もあり，それらの全体像がおおよそ把握できるようになった。実際の言語データの保存とデジタル化は，個別のケースごとの問題があり，事例集として集積している。言語学のデータは，部分的にマルチメディア化が先行している面もあり，類似の事例が生じる他の研究領域への知見を提供できるほか，学術研究を持続可能な形態にする上で何が必要かを明確にすることは，言語学や関連領域，科学研究全体に貢献できるものと考えている。

研究成果の概要（英文）：Our research gives priority to preservation of linguistic data, which had been collected about a half century ago. Since these data consist chiefly of handwritten documents, analogue-recorded tapes, and some materials including cultural items, we must digitize them for the sustainable utilization of these original data. For these digitized data to be properly used, a systematic management and eternal preservation is necessary, which must be designed for observing the concerned laws and protecting the copyrights of linguists and consultants. We have converted several sets of linguistic data into digital forms and submitted scientific and ethical investigations about the concerned problems.

研究分野：言語学

キーワード：言語学　音声学　言語データ　方言学　デジタル化　アーカイブ化

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

１．研究開始当初の背景

現在の言語研究には表面化していない危機がある。それは，個々の言語研究者が有する言語データがいずれ無に帰すという危機である。定年退職などで研究職を離れた研究者（本研究では，RLR=Retired Linguistic Researcher の略称を用いている）はすぐに言語研究をやめてしまうわけではなく，収集済みのデータから新しい知見を得るべく研究を続けることが多い。それでも，研究環境や研究資金，また加齢に伴う状況変化のために，現役時代と同じ水準で研究を継続することは困難である。研究職を離れた研究者が保有する言語データには，危機言語に数えられる言語や方言のデータも含まれ，現時点では言語調査が困難で入手不可能なものもあると想定される。また，半世紀前の言語データはその後の変動の激しい時代を考慮すると通時的なデータとしても貴重なデータである。これらのデータは現在の研究者が手を尽くしても入手できないものであり，今後も入手できる見込みのないデータであり，その喪失は言語研究全体にとって取り返しのつかない損失となる。

しかし，現状では，研究者個人が保有・管理するデータを大学などの研究機関に委託したいと考えても引き受けるしくみが整っていない。個人的に知己のある研究者に託すことはできても，データを託された研究者が研究職を離れたときに同様の問題が，ときに拡大した形で生じる。問題を先送りすると，結果的に深刻さが見えないまま肥大化する。言語データの形態という，現実的かつ原始的な問題は，早めに対応しないと徐々に対応が難しくなると考えられる。

研究開始時点で想定していた問題としては以下 3 点が挙げられる。

(1)言語データの管理にかかる方法論的問題･･･言語や方言の記述を行うのは研究者個人であり，データの知的所有権はデータを作成した研究者個人にあるが，言語音を録音する際に協力した母語話者・母方言話者（以下「調査協力者」と呼ぶ）にも隣接的権限が考えられる。言語調査に調査協力者は諒諾したはずだが，自分の肉声による音声を他の研究者が二次的に利用したり，電子的に公開したりすることまで許諾したとは言えない。調査協力者には既に物故した方，海外などで再度連絡を取ることが難しい方もあり，今後説明と了解の機会が得られないケースもある。また，研究者のデータには，フィールドノートなども含まれるが，これらは公的な研究資金による調査であっても，個人のメモという性質が強く，公的研究資源として共有すべきと安易に判断できない。つまり，言語データの委譲や管理委託については，保有する言語研究者の意向を尊重すべきで，隣接する調査協力者の権限にも配慮を要する。そこでは，データ管理にかかる倫理的・法的・技術的問題に総合的に対応できるスキームと基盤的方針が必要であり，しかも，それらは実務性や現実性を有するものでなければならない。

(2)データ形式と保存にかかる技術的問題･･･言語の筆記データ類は古くからあるが，言語音を録音したものは 1960 年代からの 1/4 吋オープンリール，その後のカセットテープ，平成以降は種々のデジタルメディアが混在している。本研究が現時点で対象とするデータはカセットテープが多く，デジタル化自体は専門業者も利用可能で容易だが，その記録方式の選択や原データの劣化などの問題，加えて有効な研究資源とするためのデータの保持形式の問題もある。例えば，単語リストを続けて読む録音データなら，単語や一定のシンタグマごとに 1 つのデータとし，Praat などでの分析データを付すことで活用価値は大きく増す。音調が rising か high rising かなどはピッチのデータを付せば視覚的な確認が容易になる。これらは，音声学や分析ソフトに通暁した院生や PD の協力を得て進める。この班は，連携研究者・中川裕（東京外大）が差配し，研究分担者が管理運営も含めて。総合的な支援をおこなう。

生データを単にデジタル化するだけではなく，フィールドノートや論文類と照合して，形態素や句あるいは文のレベルでデータを切り分け，他の情報を付加することで「質の高いデータ」が得られ，情報通信技術が進んだ現代の利点を活かした研究が可能になる。

(3)研究文化に関わる個別の問題と全体的課題･･･言語調査を行い，そのデータを蓄積して分析・記録していく，いわゆる記述的な研究では，対象言語の類型的特性や研究者の関心，また地理的特性などにより，言語データの種別や粒度はさまざまである。例えば，日本語の地域方言について言えば，音声データは分節音の差異（母音無声化や中舌母音，円唇性など，破擦音化中和や有声化，声門閉鎖など）に着目する場合もあるが，もっぱら高低アクセントに着目する研究も多い。また，危機言語など調査協力者が限定されるケースもあれば，有力言語など大規模な社会言語学的調査が可能なケースもあり，方言研究では言語研究者自身が内省をデータ化することもあり得る。海外での言語調査においても，音声や形態に重点を置くケースも，句レベルのシンタグマや文形態のデータを意図的に集めるケースもあり，言語類型論のテーマや通言語学的な関心がデータの種別に影響を与えることもある。対象とする言語・言語群や語族，地域により研究者の関心は異なるためデータが満遍なく均等に存在することはなく，調査協力者との関係は個々に異なるため変異への配慮やエリシテーションの質など，記述の粒度も違いが想定される。総じて，地域の言語文化の風土に研究風土も影響を受けることは考えておかなければならない。

２．研究の目的

本研究だけで，すべての危機的な言語データの問題を解決できるわけではないにしても，以下のような研究手順と目的をあらかじめ定めた。

(1) 現状の正確な認識と可能な対処の確認
(2) データの発掘・救出・永続化に関わる技術的な問題点の確認
(3) データの発掘・救出・永続化に関わる倫理的，また，法務的な問題点の確認
(4) データの保存・利活用・永続化のためのシステム設計に必要なものの確認

以上のような実用的な目的に沿って，可能な限り，データの発掘と永続化の作業を進め，そのなかで，どのような問題が出来るか，また，それにはどう対処すべきかなどの経験知を積み上げることを研究の方針とした。

３．研究の方法

研究全体を大まかに 2 つの領域に分け，データのデジタル化・永続化とそのデータベース設計やアーカイブ化に関わる作業と研究を，東京外国語大学の中川裕研究室で，塩原朝子の協力のもとで進めることとした。また，それ以外の実務や倫理・法務に関わる研究を北海道大学の加藤重広研究室で担当することにした。前者では，本研究の実務を進める知識と能力を備えた大学院生らの協力が得られることから，データ処理を進めながら，関連するアプリケーションの開発や問題点の解消などをおこない，蓄積を進めた。後者では，研究倫理・法務全般の問題点と，本研究における関連問題のうち，後者に重点を置き，検討を進めることとした。また，最重要の問題でもある。埋もれたままのデータを発掘することについては中川と加藤が知己を頼り，個別に状況確認を重ねていく方法が，地道ではあるものの，むしろ効率的であることがわかり，その方法によって進めることにした。

４．研究成果

本研究では貴重な言語データがアナログ形態のまま埋もれているものを発掘し，それをデジタル化することで永続化し，保存するシステムや制度の設計を推進するための研究をおこなったが，そのなかで実際のデータの処理も多数手がけ，それによって，データ永続化に要するリソース（人的資源・研究費・時間など）をおおよそ把握することができた。中には，ノート類のデジタルスキャンなど専門知識を要しない作業など，民間業者を利用した方が効率がよいものもあるが，言語学を専門とする大学院生や学部生に作業してもらっても経費そのものは大きく変わらず，教育効果や研究資源への先行投資と考えれば，むしろ後者の方が望ましいと判断されるケースもあった（コロナ禍であったため，確認の方法は制約があったことを付言したい）。

(1)データ化の進捗と計画

本研究では，以下のデータを受け入れることとし，このうち半数以上でデータの永続化作業が完了している（つまり，喪失の危機を免れた）。ただ，データをどのようなレベルまで処理するかによって，利活用の精度は大きく影響を受けるので，この点は重要である（次項に記す）。

#	言語データ提供者	対象言語とデータ収集時期
1	T.S氏	オーストロネシア諸語(1960年代以降)
2	故Y.Y氏	バントゥ諸語（1970年代以降）
3	S.K氏	グイ語(1980年代以降)
4	T.J氏	ガナ語（1960年代以降）
5	U.Z氏	日本語諸方言（1970年代以降）
6	Y.S氏	ビルマ語系諸言語・諸方言
7	K.S氏	テンボ語（1980年代以降）
8	Y.Y氏	オーストロネシア諸語（1970年代以降）
9	Y.T氏	アフロアジア諸語（1970年代以降）

これらのデータのうち，2 のようにバントゥ諸語のパラレルデータを数百言語について収集しているものもあれば，3 のように自然発話データがその量と質において並ぶものがないほどのレベルで収集されているものもある。1 や 6 や 8 では，既に危機言語の記述データを多く含んでおり，中には話者が絶えてしまった言語のデータも含まれている。以上の言語データは，現在でも参照や活用の希望が絶えないものが多く，このことは，半世紀近く前の日本の記述言語学の学術的水準を物語っている。

ただし，今回の研究経費と研究期間ではカバーできなかった言語データもあり，それらを今後データとして失わせないために，作業を継続する必要がある。

(2)データの永続化の段階と方法

本研究を通じて，データを永続的なものとして処理するには，おおむね以下の 3 段階があることが判明した。

	作業内容(activity)	成果(product)　　利活用の 3 工程
第 1 工程	(1)生資料入手 (2)デジタル化 (3)メタデータ収集・集約	(1)アナログ録音・紙ノート (2)デジタル・ファイル (3)一覧表・関連付けした情報
第 2 工程	(4) pdf のテキスト化 (5)テキストデータ校正	(4) データ処理可能なテキストデータ (5)精度の高いテキストデータ
第 3 工程	(6)言語学的な情報付加	(6) グロス・註付きデータ（音声とのリンク）

このうち，第 1 工程(1)(2)では特段の専門知識を要しないものの，(3)と第 2 工程の作業では専門知識を持つ大学院生などが従事する必要があり，第 3 工程では専門の言語研究者が協力して利用しやすい状況に処理しなければならないことがわかった。

これらのデータは処理が完全に済んでも，すぐに利活用されて新たな研究成果に結び着くとは限らず，成果が現れるまでに相当の時日を要するケースも想定される，長期的なプロジェクトとなる。効率化が求められ，即時的な成果を求められる潮流が強い現状では，重視されないかもしれないが，今後とも日本の言語研究の水準とこころざしの高さを守るために，地道に続けていくことが必要である。

なお，本研究では，データの永続化を効率化するためにいくつかのアプリケーションの開発をしており，その一部は学術目的であれば，他にも使用を許している。

(3) データの活用に関する倫理・法務・権利

本研究では，収集された言語データに関わる権利の問題と学術上の価値の問題を研究倫理と法律の観点からも解明しようとした。その多くは，おおむねの方針が立てられる程度に結論が得られているが，国際的な標準と日本の標準の齟齬に文化的な価値観が関わっていること，データ収集当時と現在とで権利意識が大きく異なっていること（これは，今後も変化する可能性があることを示している），など容易に解決し得ない根本的な問題があることもわかっている。

言語データに関わる形態ごとに，考慮すべき権利者と権利形態をまとめると以下のようになる。なお，下表における「データ提供者」はいわゆるコンサルタントのことであり，以前は聴者協力者・インフォーマントと呼ばれていた方のことである。

データ種・活用形態	権利者	権利形態
フィールドノート（メモ類含む）	研究者	著作権
音声データ	データ提供者	著作隣接権（実演）
	研究者	著作隣接権（編集・作成）
動画	データ提供者	著作隣接権（実演）・肖像権
	研究者	著作隣接権（編集・作成）
一般公開	学会・研究者	公衆送信権

研究者もデータ提供者も永遠の生を生きるわけではない。これらの著作権・著作隣接権が厳密に 50 年あるいは 70 年と継続するのであれば，その管理が必要になるが，われわれが扱うものは学術データであって特に経済的利益を生み出すとは想定しないことが多い。しかし，近年はビッグ・データとして活用するために必要とされることがあり，想定外の需要がありうるとわかっている。しかし，倫理的観点や法務上の責任から，容易に，転用を許さない，データ管理の教育や啓蒙が必要だという見解を結論としている。

５．主な発表論文等

〔雑誌論文〕　計6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　1件）

１．著者名 加藤重広	４．巻 161
２．論文標題 言語データの継承と保存に関する課題について	５．発行年 2020年
３．雑誌名 北海道大学文学研究院紀要	６．最初と最後の頁 35-49
掲載論文のDOI（デジタルオブジェクト識別子） 10.14943/bfhhs.161.l35	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 加藤重広	４．巻 2
２．論文標題 心理的文脈と前提を巡る動的心理語用論の構想	５．発行年 2020年
３．雑誌名 動的語用論研究の構築に向けて	６．最初と最後の頁 240-264
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 加藤重広	４．巻 21
２．論文標題 日本語副助詞と世界知識	５．発行年 2020年
３．雑誌名 語用論研究	６．最初と最後の頁 19-37
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 Witzlack-Makarevich, A. and H. Nakagawa	４．巻 －
２．論文標題 Linguistic Features and Typologies in Languages Commonly Referred to as ‘Khoisan’	５．発行年 2019年
３．雑誌名 E. Wolff (ed.) The Cambridge Handbook of African Linguistics	６．最初と最後の頁 382－416
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 Kimura, Kimihiko	4．巻 39
2．論文標題 An excavational investigating of /u/-forming in southern-Pennsylvania English	5．発行年 2019年
3．雑誌名 RANDOM	6．最初と最後の頁 1-23
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 Anthony Jukes, Asako Shiohara and Yanti	4．巻 11
2．論文標題 “Collaborative Project for Documenting Minority Languages in Indonesia and Malaysia	5．発行年 2018年
3．雑誌名 Asian and African languages and linguistics	6．最初と最後の頁 45-56
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 －

〔学会発表〕　計1件（うち招待講演　0件／うち国際学会　0件）

1．発表者名 加藤重広・中川裕・米田信子・塩原朝子・加藤幹治・木村公彦
2．発表標題 最後のアナログ言語調査資料
3．学会等名 日本言語学会
4．発表年 2021年

〔図書〕　計4件

1．著者名 加藤重広・岡墻裕剛	4．発行年 2021年
2．出版社 勉誠出版	5．総ページ数 432
3．書名 日本語文字論の挑戦	

1．著者名 加藤重広、澤田 淳	4．発行年 2020年
2．出版社 研究社	5．総ページ数 290
3．書名 はじめての語用論	

1．著者名 加藤　重広、滝浦　真人	4．発行年 2020年
2．出版社 ひつじ書房	5．総ページ数 308
3．書名 日本語語用論フォーラム	

1．著者名 加藤　重広	4．発行年 2019年
2．出版社 筑摩書房	5．総ページ数 304
3．書名 言語学講義	

〔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	中川　裕 (Nakagawa Hirosi) (70227750)	東京外国語大学・大学院総合国際学研究院・教授 (12603)	
研究分担者	塩原　朝子 (Shiohara Asako) (30313274)	東京外国語大学・アジア・アフリカ言語文化研究所・教授 (12603)	

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