

データのアーカイブからできること：発掘的研究の可能性

木村公彦（東京外国語大学大学院）

1 データのアーカイブと発掘的研究

本発表では、JSPS 科研費 JP18H00661 「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」（代表：加藤重広）のフィールドデータのアーカイブ化プロジェクトが潜在的にもたらす発掘的研究手法を紹介する。

発掘的研究は過去の言語学的一次資料をデータとして用いることで、これまでの研究史の空白を埋め、過去の言語現象に関する新たな知見を得るための研究手法である。一方、発掘的研究を行うためには、豊富に蓄積されたデータが広く利用可能な状態で公開されていることが前提であり、その環境が整っていなければ実行することは困難である。

発表者は上記科研プロジェクトにおいて、菅原和孝氏のグイ語談話資料のアーカイブ化に携わっているが、現時点の目算ではこの資料のアーカイブ化が完了すると、総語数約 56 万語の文字資料と音声資料を結びつけたデータが電子化される。このデータベースが今後公開されることになれば、それは単に過去の記録を保存したものに留まるものではなく、発掘的研究のための土壌を整えるものとなり、今後の研究の幅を広げることにつながる。

以下、発表者がこれまでに他のデジタルアーカイブを利用して行った研究例を紹介し、フィールドデータのデジタルアーカイブ化が可能にする、発掘的研究の重要性に言及する。

2 発掘的研究の事例

発表者が専門とするアメリカ英語方言では、2010 年代半ばから方言変化に関する発掘的研究が盛んに行われるようになってきた (cf. Gordon and Strelluf (2017))。この背景には、20 世紀を通して収集された、アメリカ英語の方言地図・方言辞書編纂のためのフィールドデータが蓄積されていたことがある。データの中には 1930 年代に残されたアルミニウムディスクの録音記録や、1960 年代のカセットテープを用いた録音も含まれていた。このような資料のデータベース化は 20 世紀の後半には開始され (Kretzschmar, 2002)、2010 年代には広く利用可能なウェブサイト¹に公開されていった¹。

発表者は公開されたデータベースのうち、アメリカ方言辞書 (Dictionary of American Regional English; DARE) 編纂プロジェクトの資料を利用して発掘的研究を行った。この方言辞書プロ

¹e.g. アメリカ方言地図プロジェクトの HP : <http://www.lap.uga.edu/>

プロジェクトでは1960年代後半のアメリカ英語方言が対象となっており、上記の科研プロジェクトで構築中の危機言語を含むデータベースとは性格が異なる部分はあるものの、半世紀前のフィールドデータを保存・公開したものであるという点では共通している。DAREプロジェクトでは語彙収集のためのアンケート調査が行われたほか、全米の方言発音の記録・収集のためにフィールドワークの様子が1843本のテープに残された。一部欠損はあるものの、2021年10月の時点では、そのうち1759本分の音声データがウィスコンシン大学図書館のデータベース²から公開されている。録音されたデータは大まかに文章読み上げと自由会話に分けられ、それぞれややフォーマルな発話と通常の状態での発話を比較することを目的に集められたものである (Cassidy & Hall, 1985)。自由会話の内容は多岐にわたり、語彙や発音などの言語学的な情報以外にも、話者の宗教観や当時の教育制度など、文化的背景に関する情報も得られる。

発表者はDAREプロジェクトのデータを用いて、アメリカペンシルベニア州内部での後舌狭母音/u:/の前舌化の地理的拡大の経過について調査した。後舌狭母音の前舌化は、1940年の時点ではペンシルベニア州の東端と西端の2都市の周辺だけで報告されていた現象であったが (Kurath & McDavid, 1961)、1990年代の発音に基づくアメリカ方言地図、Labov, Ash, and Boberg (2006)によると、1990年代にはペンシルベニア州全域への拡大が確認できる。DAREのデータは両時点の中間にあたる1960年代に収集されたものであり、ちょうど研究史の空白を埋めるという目的に合致したデータベースであった。詳細な分析手法と結果についてはKimura (2018)に掲載されているため、ここでは割愛するが、録音音声の音響分析を行った結果、後舌狭母音の前舌化を音響的に示す話者は1940年代に整備された幹線道路沿いに分布する傾向が確認された (図1を参照)。

DAREのデータ自体は方言辞書の編纂を目的に集められたものであったが、この事例のように、当初は想定していなかった研究目的を持つ研究者がデータベースを利用し、研究史の空白を埋めるような新たな知見を得られる可能性がある。先行研究に記載された二次的記録ではなく、現在ではもはや収集することが困難な一次資料を実際に用いて上記のような研究が行える点でも、発掘的研究の重要性は大きいといえる。

3 総括

本発表ではアメリカ英語の研究事例の紹介を通して、フィールドデータのアーカイブが潜在的に持つ発掘的研究の可能性と重要性について言及した。アメリカ英語のようなメジャー言語では、DAREのような大規模なものは滅多にないにしても、過去の記録が利用可能な状態で散見される。しかし、上記科研プロジェクトで扱う言語群のように、過去のデータ蓄積が少ない場合、発掘的研究の土壌が整備されるという点で、データアーカイブの整備と公開をする意義はさらに大きなものになると考えられる。

²<https://search.library.wisc.edu/search/digital>

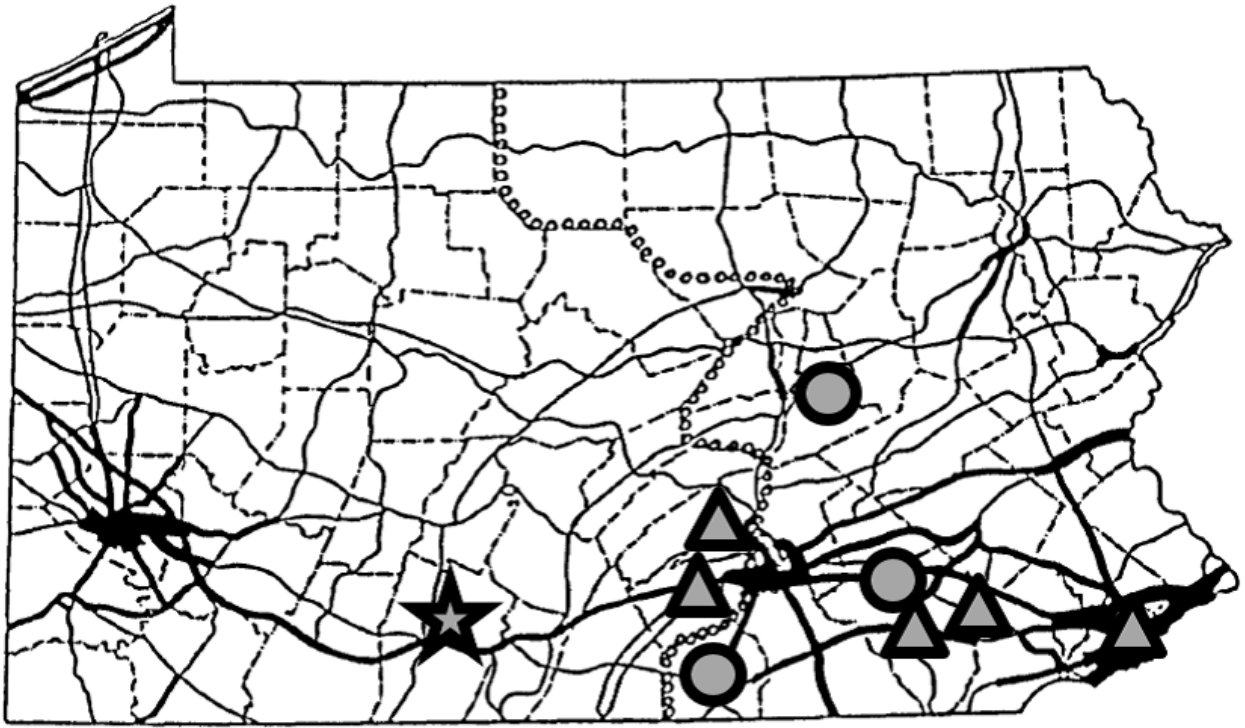


図 1: Kimura (2018) 掲載の文章読み上げ音声の分析結果。星、三角、丸の順で前舌化の度合いが大きいことを示している。地図はHerold (1997) に掲載されていたものを発表者が加工した。地図上の線は太いほど幹線道路の交通量が多いことを表している。

References

- Cassidy, F. G., & Hall, J. H. (Eds.). (1985). *Dictionary of American Regional English, Volume I, A-C*. Cambridge, MA: Belknap Press of Harvard University Press.
- Gordon, M. J., & Strelluf, C. (2017). *Evidence of American Regional Dialects in Early Recordings* (R. Hickey, Ed.). Cambridge: Cambridge University Press.
- Herold, R. (1997). Solving the actuation problem: Merger and immigration in eastern Pennsylvania. *Language Variation and Change*, 9, 165-189.
- Kimura, K. (2018). Excavational investigation of /u/-fronting in southern-Pennsylvania English. *RANDOM*, 39, 1-23.
- Kretschmar, W. A., Jr. (2002). Linguistic Databases of the American Linguistic Atlas Project (ALAP).
- Kurath, H., & McDavid, R. I. (1961). *The pronunciation of English in the Atlantic States: based upon the collections of the linguistic atlas of the Eastern United States*. Ann Arbor: The University of Michigan Press.
- Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Walter de Gruyter.