

GISと言語研究

池田 潤（筑波大学）

キーワード： 地理情報システム、言語地理学、北西セム語

1 はじめに

筆者は文部科学省の科学研究費補助金の助成を受けて¹、地理情報システム（Geographic Information System, 以下 GIS とする）を利用した北西セム語の言語地理学的研究（以下、北西セム語 GIS プロジェクトと略称する）に取り組んでいる。本稿では、GIS を利用した言語研究の可能性を展望したうえで、北西セム語 GIS プロジェクトを例に言語地理学への応用の可能性について考察する。

2 GISと言語研究

2.1 GIS とは何か

国土地理院によると GIS は「地理的位置を手がかりに、位置に関する情報を持ったデータ（空間データ）を総合的に管理・加工し、視覚的に表示し、高度な分析や迅速な判断を可能にする技術である」²。基本的にはコンピュータ上に地図を表示させるシステムなので、行政やビジネスを初めとする多方面で利用価値の高い技術である³。

我が国では、平成7年の阪神・淡路大震災をきっかけに政府が GIS による国土空間データ基盤（NSDI：National Spatial Data Infrastructure）の整備

¹ 基盤研究 C (18520292)「前 2-1 千年紀における北西セム語の等語線の再画定：GIS による言語地理学的研究」（研究代表者：池田潤、平成 18-21 年度）

² <http://www.gsi.go.jp/GIS/whatisgis.html>

³ 日経 BP 社の GIS サイト (<http://premium.nikkeibp.co.jp/gis/>) には自治体や企業における GIS の導入事例が紹介されている。

に乗り出し、2002年2月のGIS関係省庁連絡会議において「GISアクションプログラム2002-2005」と題する行動計画が決定された。これはGISにより豊かな国民生活を実現することを目指したもので、政府の各分野においてGISを有効に活用し、行政の効率化と質の高い行政サービスの実現を図ることを目標としている⁴。

GISを用いると、電子地図を作成し、それを画面に表示したり印刷したりすることが可能となる。電子地図なので、地図の拡大・縮小・移動が容易なうえ、レイヤ（画層）を自在に管理することもできる。レイヤとは一枚の主題図に相当するデータセットの層のことで⁵、GISでは見たいレイヤだけを選んで表示することができる。したがって、同一の空間を対象としたさまざまな主題図を生成したり、いくつかのレイヤを重ね合わせること（オーバーレイ）によって複数の情報の間の相関を可視化したりすることもできる。さらに、ソフトウェアによっては空間的な解析を行う各種機能もある⁶。

2.2 GISを用いた言語研究の可能性

人文系の学問分野の中ではとくに考古学におけるGISの活用が盛んで、ケンブリッジ大学出版局から考古学者向けのGISのマニュアル（Conolly and Lake 2006）が出ており、「GIS考古学」という新領域も存在するようである⁷。

言語研究へのGISの導入も始まっている。たとえば、社会言語学のフィールド調査で得られたデータをGISソフトによって地図化する試み（中井2005）や、既存の言語地図をGISを用いてデータベース化する試み（大西2003、鳥谷2006）がなされている。また、*Ethnologue* 第15版（Gordon 2005）

⁴ <http://www.mlit.go.jp/kokudokeikaku/gis/>

⁵ <http://gisschool.csis.u-tokyo.ac.jp/gisa/index.php3>

⁶ 図形間の演算機能、属性の集計機能、ネットワーク処理機能、圏域分析機能など。詳しくは、町田（2004）の第3章を参照。

⁷ 国際日本文化研究センターのウェブページによると、「GIS考古学は、GIS（Geographical Information System, 地理情報システム）を活用して空間情報の研究を進める考古学の新しい分野である」という（<http://www.nichibun.ac.jp/graphicversion/dbase/iseki.html>）。

の GIS 版も発売された⁸。これらは、いずれもこれまで紙の上でおこなってきた作業を電子化する試みだと言えるが、それにはどんなメリットがあるのだろうか。

ひとつのメリットは、前節で言及したレイヤの重ね合わせ（オーバーレイ）機能にある。国立国語研究所による『日本語地図』や Moseley and Asher (1994) に代表される紙の言語地図にも多くの場合、複数のレイヤが存在する。たとえば、(白地図ではなく) 行政図や地形図に言語データをプロットすれば、地形と言語データという 2 層のレイヤが存在することになる。あるいは、複数の語彙の分布を 1 枚の地図にプロットすれば、そこにも語彙 A と語彙 B という 2 層のレイヤが存在することになる。紙の地図の場合、レイヤを操作するのは、地図の編集者である。地図の使用者が任意のレイヤを選んで組み合わせようと思ったら、複数の地図を OHP 用紙にでもコピーするしかなく、非常に不便であった。それに対して、電子化された地図では自在なレイヤ操作が可能となる。これにより、標高、河川・湖沼、行政区分、交通網、集落分布などの非言語情報と言語情報との相関、ないしは思いつく限りの言語情報の組み合わせの相関をコンピュータの画面上で簡単に可視化することができる。その点で GIS は思わぬ発見を生む可能性を秘めている。

しかし、最大のメリットはデータの共有性にある。インタラクティブなかたちで電子化された言語地図は少数ながら存在する。その代表は Haspelmath et al. (2005) に付属する “interactive CD-ROM” であろう。図 1 はこの CD-ROM を使って筆者が生成した言語地図で、動詞 (V) と目的語 (O) の語順と側置詞の相関関係をプロットしてある。

⁸ Global Mapping International 社の World Language Mapping System, Version 3.2. 詳しくは、同社のウェブページ (<http://www.gmi.org/wlms/>) 参照。

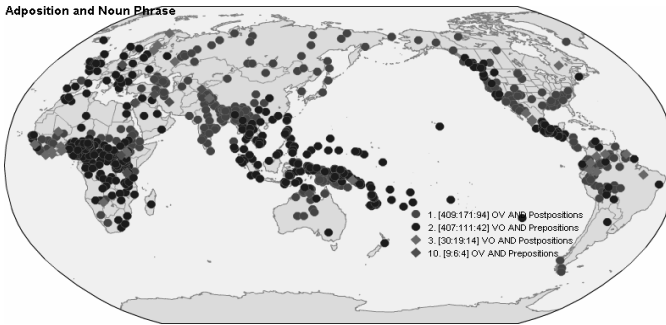


図 1 : 語順と側置詞の相関

また、マンチェスター大学の Yaron Matras 教授が作成した Romani Dialects Interactive⁹には、あらかじめ用意された複数の等語線を任意に組み合わせて表示するモードがある (図 2 参照)。

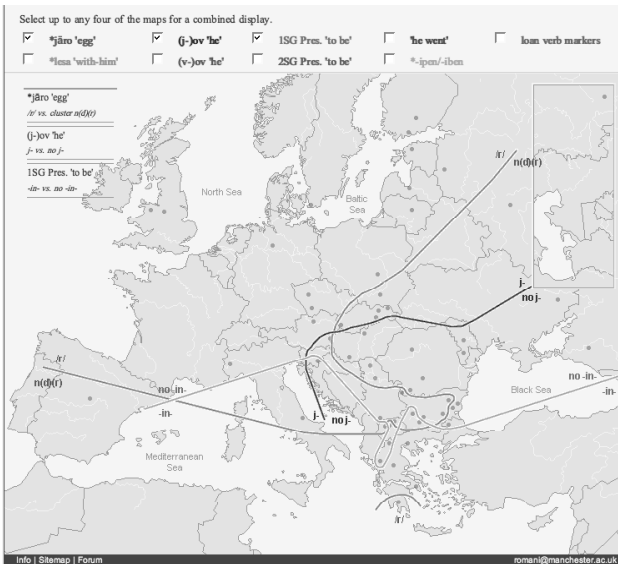


図 2 : Romani Dialects Interactive

⁹ http://www.llc.manchester.ac.uk/Research/Projects/romani/files/31_dialectsinteract.shtml

言語地図を作製するためのソフトも存在する。一例として県立新潟女子短期大学の福嶋秩子教授が開発し、無償で配布している SEAL (System of Exhibition and Analysis of Linguistic Data) をあげることができる¹⁰。SEAL は Windows 上で作動するソフトウェアで、言語地図用の白地図作製、言語データの入力、言語地図の描画、言語データの集計等の機能をもつ。言語地理学の研究者が自分自身の研究上の要請から開発したソフトであるため、非常に魅力的なツールである。そこで筆者は数年前に SEAL version 6.1J を入手し、前1千年紀における北西セム語の等語線の再画定¹¹に利用してみた。しかし、当時は複数の言語地図の重ね合わせができず¹²、また時間的制約や筆者自身の力量不足もあり、残念ながらこのツールを使いこなすことができなかった。

上記の電子地図およびソフトウェアと GIS の間には大きな違いがある。それはデータ共有の問題である。電子地図は基本的に閉じたシステムであり、データにはまったく汎用性がない。SEAL は英語版・中国語版・韓国語版もあり、システム・データ・マニュアルが公開されているため、SEAL ユーザの間ではデータが共有できる。しかし、情報共有手段としては GIS の方が有望であろう。2.1 節でふれたとおり、政府も GIS による国土空間データ基盤の整備に乗り出しおり、「国、地方公共団体及び民間が、各々整備している空間データのうち基盤的なものを・・・電子媒体やネットワークを通じてこれらを提供し¹³」共有する体制作りが進められているという。また、XML をベースとした標準フォーマット案も登場している¹⁴。

言語研究は GIS を活用することによって上記のメリットを享受することができる。空間データを駆使する言語研究としては、フィールド言語学、類型論、言語地理学などが考えられるが、フィールド言語学への応用については本誌所収の乾氏の論考、類型論への応用については本誌所収の山本

¹⁰ 詳しくは、<http://www.nicol.ac.jp/~fukusima/inet/lg3.html> を参照。

¹¹ 科学研究費補助金による基盤研究 C (13610637) 平成 13-14 年度。

¹² Version 7.0J では複数の言語地図の重ね合わせも可能になったようである。

¹³ <http://www.mlit.go.jp/kokudokeikaku/gis/>

¹⁴ (財) データベース振興センターの G-XML や米国で設立された OpenGIS Consortium の GML など。

氏の論考にゆずり、以下、本稿では北西セム語プロジェクトを例に言語地理学への応用の可能性について少しく考えてみたい。

3 北西セム語 GIS プロジェクト

3.1 プロジェクトの概要

前2～1千年紀のシリア・パレスチナでは前2千年紀にアモリ語、アマルナ語、ウガリト語など、前1千年紀にはアラム語、フェニキア語、ヘブライ語、アンモン語、モアブ語、エドム語などの北西セム語が話されていた。このうち前1千年紀における等語線は Garr (1985)によってほぼ画定されている。しかし、その後、聖書ヘブライ語の研究において新展開が見られ、これをもとに Garr によって画定された等語線を根本的に見直す余地が出てきた。

研究代表者はこの点に注目し、平成 13-14 年度に科学研究費の助成を受けて「前1千年紀における北西セム語の等語線の再画定」と題する基盤研究(C)をおこなった。この研究によって、12の言語的特徴について Garr による等語線が再画定され、また前2千年紀のアマルナ語に南北方言の違いがはっきりと認められることが明らかとなった¹⁵。しかし、言語地図を作成するという当初の計画は部分的にしか達成されず、今後の課題として残る結果となった。これらの成果と課題を踏まえ、このたび次の2点を目的とした新たな研究を構想した。

(1) GIS による言語地図の作成：平成 13-14 年度の研究では時間的制約もあり、言語地図を検索や公開に適したかたちで電子化するよい方法が見つからなかった。その後、GIS の存在を知り、言語地理学的研究に応用できる可能性を感じた。GIS を利用すれば、空間データと言語データを組み合わせた検索をおこなったり、XML でマークアップした言語コーパスの検索結果を地図上に表示したりすることが可能となる。

(2) 通時的言語地図の作成：時代の異なる複数の言語地図を重ね合わせると、言語的特長の共時的な分布が通時的にどう変化したかを可視化することができる。時間と空間という2つの次元をダイナミックに制御した言

¹⁵ 詳細については、池田 (2003) 参照。

語地図が作成できれば新たな言語地理学的知見が得られるはずである。本プロジェクトでは、Garr や平成 13-14 年度の基盤研究(C)に基づく前 1 千年紀の言語地図を作成すると平行して、アマルナ語のデータをもとに前 2 千年紀の言語地図を別途作成し、時代の異なる言語地図を効果的にオーバーレイさせる方法を模索する。

本プロジェクトは大きく分けて次の 3 つのステップからなる。

- ① 北西セム語の等語線の再画定のため、筆者が前 2 千年紀、竹内茂夫氏(研究分担者) が前 1 千年紀の言語データの収集・分析・マークアップをおこなう。
- ② このデータベースを、筆者と乾秀行氏(研究分担者) が協力して構築する GIS システムの地図と連動させる。
- ③ ①と②によって作成された前 2 千年紀と前 1 千年紀の地図をオーバーレイさせることにより、筆者と竹内氏が北西セム語の等語線の通時的変遷を明らかにする。

3.2 プロジェクトの可能性

GIS を用いた言語研究は、管見の及ぶかぎり国内外を通じてセム語学の実分野では存在しない。そのため、その初の試みとして北西セム語 GIS プロジェクトのもつ意義は世界的に見ても大きいと言える。北西セム語の等語線に関して既知の情報がインタラクティブなかたちで電子化されるだけでも研究史上大きな意義を有するうえ、自在なオーバーレイによって思わぬ発見も期待できる。4 年後を目指して、世界の研究者が利用できる北西セム語研究の強力なツールの構築、および北西セム語に関する世界的な研究拠点の形成を進めていきたい。

さらに、本プロジェクトは次の 2 点に関して言語地理学一般に対する方法論的貢献を目指している。

- (1) 空間データ、時間データ、言語データを組み合わせた検索が可能な言語地図を作成するにはどうしたらよいか。
- (2) そのような言語地図を Web 上で共有するにはどうしたらよいか。

(1)は、言語コーパスのマークアップの問題に帰着する。本プロジェクトでは、あらかじめ地図上にプロットした言語的情報を電子的に表示するのではなく、マークアップされたコーパスを検索し、検索結果をGISサーバに送り、言語地図をダイナミックに生成したいと考えている。そのためには、言語データに空間データと時間データを組み込む必要がある。目下、この条件を満たし、なおかつセム語の言語的特長¹⁶や文字の特殊事情¹⁷にも適合したタグ構造の検討を進めている¹⁸。タグ構造やスキーマはできるかぎり公開し、他言語の言語地理学的研究への応用を図るつもりである。

(2)は、言うまでもなく本プロジェクトだけで解決できる問題ではない。連携する2つの研究プロジェクト¹⁹と協力しつつ、「世界言語地図をデジタル化し、多くの言語学者がそれを共有できる環境を作る」(乾 2006: 2) ための具体的方策を練っていきたい。

【参照文献】

- Conolly, J., and M. Lake (2006) *Geographical information systems in archaeology*. Cambridge: Cambridge University Press.
- 福盛貴弘・池田潤 (2002) 「文字の分類案：一般文字学の構築を目指して」『一般言語学論叢』 4・5: 33-56.
- Garr, W. R. (1985) *Dialect geography of Syria-Palestine, 1000-586 B.C.E.* Philadelphia: University of Pennsylvania Press.
- Gordon, R. G. (ed.) (2005) *Ethnologue: languages of the world*, 15th ed. Dallas: SIL International.

¹⁶ 形態論的には、とくに語根と語型による不連続な語形成が特徴的である(高橋 2005 参照)。

¹⁷ アマルナ語は楔形文字(アマルナ語式シュメール系音節表語限定文字)で書かれている。前1千年紀の北西セム語はフェニキア系子音文字で書かれている。文字の類型に関しては、福盛・池田(2002)参照。転写の際に特殊文字を多用するため、unicodeのサポートも不可欠である。

¹⁸ この作業に関しては、研究協力者の高橋洋成氏(筑波大学大学院生)と杉江拓磨氏(テルアビブ大学大学院生)の支援を受けている。ここに記して感謝したい。

¹⁹ 基盤研究(B)16401008「オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築」(研究代表者：乾秀行)および基盤研究(C)18520288「地理情報システムによる世界諸言語の言語類型地理論的研究」(研究代表者：山本秀樹)。

- Haspelmath, M., M. Dryer, D. Gil, and B. Comrie (eds.) (2005) *The world atlas of language structures*. Book with interactive CD-ROM. Oxford: Oxford University Press.
- 池田潤 (2003) 『前 1 千年紀における北西セム語の等語線の再画定』平成13-14年度科学研究費補助金基盤研究(C)(2)研究成果報告書.
- 乾秀行 (2006) 「地理情報システム(GIS)によるエチオピアのデジタル言語地図」 乾秀行(編) 『オモ・クシ系少数言語の調査研究及び地理情報システムを用いたデータベース構築 (Cushitic-Omotc Studies 2006)』科学研究費補助金基盤研究(B)実績報告書, 1-7.
- 町田聡 (2004) 『GIS・地理情報システム：入門&マスター』山海堂.
- Moseley, Ch., and R. E. Asher (eds.) (1994) *Atlas of the world's languages*. New York: Routledge.
- 中井精一(編) (2005) 『社会言語学の調査と研究の技法：フィールドワークとデータ整理の基本』おうふう.
- 大西拓一郎 (2003) 「方言学と GIS」 『人文科学と GIS』 61-66. 富山大学人文学部.
- 高橋洋成 (2005) 「セム諸語の語構造に対する一般言語学的アプローチ」 『オリエント』 48.2: 28-46.
- 鳥谷善史(2006) 「GIS を用いた既存言語地図データベースの試み」 『日本のフィールド言語学：新たな学の創造にむけた富山からの提案』日本海総合研究プロジェクト研究報告 4: 189-205. 桂書房.

GIS and Linguistics

Jun IKEDA

This paper describes the methodology and significance of a linguistic-geographic project on Northwest Semitic languages in the second and first millennium B.C. supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (18520292). The crux of the project lies in the use of GIS (Geographic Information System) as the platform. This paper discusses some advantages of GIS over other alternatives and the various potential it embraces for the project in particular and for linguistic research in general.

Doctoral Program in Literature and Linguistics

University of Tsukuba

1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan

E-mail: ji@lingua.tsukuba.ac.jp