

# 言語・文化研究とデータサイエンス

選択 2単位

原 智弘・ブラッドリー ジョフ・山崎 直也・  
大野 雅子・江原 裕美・塩谷 英一郎・  
藤森 弘子

## 1. 授業の概要(ねらい)

言語研究や文化研究、地域研究を行なっている複数の教員によるオムニバス形式の授業である。各教員の多様な専攻分野・地域における文献資料を中心にした研究の中で、データサイエンスがどのように活用され、研究に生かされているのか、そして、その手法に対する批判を含めて検討しつつ概観する。多様な研究分野における活用事例を見ることで、データサイエンスの持つ可能性の大きさを実感すると同時に、批判的な視線を意識し、その限界の一部を垣間見ることが目標である。

## 2. 授業の到達目標

- ① 様々なデータサイエンスの活用事例について広く知り、応用の可能性について考えることができる。
- ② データサイエンスに対して、その限界を知り、批判的な議論もすることができる。

## 3. 成績評価の方法および基準

授業への参加度(50%)  
期末レポート(50%)

## 4. 教科書・参考文献

教科書  
特に指定しない  
参考文献  
各教員から適宜指示がある

## 5. 準備学修の内容

好奇心を持って授業に臨むこと  
積極的に意見表明できるように、関連参考文献などを事前に読み、理解しておくこと

## 6. その他履修上の注意事項

原則すべての授業に出席すること  
受け身の姿勢ではなく積極的に授業に参加し、意見・感想など積極的に発信を行うこと

## 7. 授業内容

【第1回】 ガイダンス

【第2回】 「言語とデータサイエンス」①

言語のデータサイエンスは多岐にわたる。テーマ別・分野別・時代別などの「言語のデータ」のピックアップ方法、大量データベースとしての様々なコーパスの意義・特徴・利用法、データの範囲の絞り方、使用頻度、データの計量の信頼性、言語データの文法的・意味的タグ付け(言語処理の基本いくつか)、意味ネットワークの研究の現状、成句(コロケーション、イディオム)のデータ分析法、意味論・語用論・認知言語学・社会言語学等におけるデータ分析の利用、言語教育や機械学習への応用、主に関日本語と英語を例に、概要を述べる。

【第3回】 「言語とデータサイエンス」②

言語のデータサイエンスは多岐にわたる。テーマ別・分野別・時代別などの「言語のデータ」のピックアップ方法、大量データベースとしての様々なコーパスの意義・特徴・利用法、データの範囲の絞り方、使用頻度、データの計量の信頼性、言語データの文法的・意味的タグ付け(言語処理の基本いくつか)、意味ネットワークの研究の現状、成句(コロケーション、イディオム)のデータ分析法、意味論・語用論・認知言語学・社会言語学等におけるデータ分析の利用、言語教育や機械学習への応用、主に関日本語と英語を例に、概要を述べる。

【第4回】 「日本語教育とデータサイエンス」①

コーパスを使った言語研究の方法は、言語現象への新しいアプローチとして様々な分野から注目されている。言語教育分野においても、学習項目や語彙の選択及びその量の目安など個人の教師の経験と勘に頼るのではなく、科学的な指標として示唆を与えてくれる。本講義では、最近公開された「多言語母語の日本語学習者横断コーパス(I-JAS)」という大規模コーパスを取り上げ、その作成経緯や内容を概観した上で、日本語学習者の言語習得を多角的に捉えられることを計量言語学的分析をとおして実証し、コーパスを用いた研究の利点と課題について考える。

【第5回】 「日本語教育とデータサイエンス」②

コーパスを使った言語研究の方法は、言語現象への新しいアプローチとして様々な分野から注目されている。言語教育分野においても、学習項目や語彙の選択及びその量の目安など個人の教師の経験と勘に頼るのではなく、科学的な指標として示唆を与えてくれる。本講義では、最近公開された「多言語母語の日本語学習者横断コーパス(I-JAS)」という大規模コーパスを取り上げ、その作成経緯や内容を概観した上で、日本語学習者の言語習得を多角的に捉えられることを計量言語学的分析をとおして実証し、コーパスを用いた研究の利点と課題について考える。

【第6回】 「英詩とデータサイエンス」①

英語の詩においては、弱強五歩格(iambic pentameter)が基本とされているが、意識的にそこから逸脱することによって、新たな意味を創り出すことも可能である。この授業においては、ルネッサンスから現代にかけての様々な詩または劇を音の観点から詳細に分析し、計量的な視点から詩の意味にアプローチしたい。たとえば、シェイクスピアの劇作品においてiambic(弱強)の代わりにtrochaic(強弱)がどの程度使われているのか、どのような場面でiambic pentameter以外のリズムが使われるのか、作品ごとにデータをとり、悲劇と喜劇における違い、年代による違いなどを検証する。さらに、20世紀に近づくにつれ、韻を踏まないblank verseが一般的になってくるが、たとえば、T. S. Eliotの詩において韻を踏まない代わりにどのようなリズムが用いられているのかも探求する。

【第7回】 「英詩とデータサイエンス」②

英語の詩においては、弱強五歩格(iambic pentameter)が基本とされているが、意識的にそこから逸脱することによって、新たな意味を創り出すことも可能である。この授業においては、ルネッサンスから現代にかけての様々な詩または劇を音の観点から詳細に分析し、計量的な視点から詩の意味にアプローチしたい。たとえば、シェイクスピアの劇作品においてiambic(弱強)の代わりにtrochaic(強弱)がどの程度使われているのか、どのような場面でiambic pentameter以外のリズムが使われるのか、作品ごとにデータをとり、悲劇と喜劇における違い、年代による違いなどを検証する。さらに、20世紀に近づくにつれ、韻を踏まないblank verseが一般的になってくるが、たとえば、T. S. Eliotの詩において韻を踏まない代わりにどのようなリズムが用いられているのかも探求する。

- 【第8回】 Philosophy, critique and data science 1: world-brain  
Overall, I will propose my perspective on why we need a critical philosophy to understand the dangerous effects of data science in education. I will compare the idea of the World Brain in H.G. Wells to 1) Bernard Stiegler's philosophy and 2) the history of cybernetics. I will apply Deleuze's philosophy and Bernard Stiegler's philosophy to understand our algorithmic/AI era. I will contrast the concept of concept societies in Deleuze with the notion of uncontrollable society in Stiegler.
- 【第9回】 Philosophy, critique and data science 2: collective intelligence  
I will analyse the question the nature of collective intelligence. I consider Bernard Stiegler's concern with the pathological effects of collective intelligence. I question the possibility of emancipation in education vis-à-vis the pharmakon of knowledge ecologies (automatism, augmented intelligence, cooperative learning). I will question the contemporary meaning of the noosphere (De Chardin, Vernadsky, Pierre Lévy).
- 【第10回】 「比較国際教育学とデータサイエンス」①  
健康、経済、知識は人間開発の三要素とされている。知識基盤社会が目指されている今日、各国において教育は、一人一人の潜在的な能力の開花と選択の幅の拡大、同時に経済生産性の向上をももたらし得る重要部門として大きな力が注がれている。教育の普及や学力試験の結果は世界的にモニタリングされ定期的に統計が発表されており、これらは各国の政策に影響を与えている。教育統計にはどのようなものがあるのか、それは教育の発展にどのように貢献しているのだろうか。講義を通じて、教育を研究するためのデータの役割を改めて考えてみたい。BATH(バイドウ・アリババ・テンセント・ファーウェイ)と呼ばれるインターネット企業の急成長が世界的注目を集める中国、COVID-19対策でオードリー・タンIT担当大臣の活躍が注目を集めた台湾の具体的事例を参照しながら、ビッグデータの社会的利用について考える。
- 【第11回】 「比較国際教育学とデータサイエンス」②  
健康、経済、知識は人間開発の三要素とされている。知識基盤社会が目指されている今日、各国において教育は、一人一人の潜在的な能力の開花と選択の幅の拡大、同時に経済生産性の向上をももたらし得る重要部門として大きな力が注がれている。教育の普及や学力試験の結果は世界的にモニタリングされ定期的に統計が発表されており、これらは各国の政策に影響を与えている。教育統計にはどのようなものがあるのか、それは教育の発展にどのように貢献しているのだろうか。講義を通じて、教育を研究するためのデータの役割を改めて考えてみたい。BATH(バイドウ・アリババ・テンセント・ファーウェイ)と呼ばれるインターネット企業の急成長が世界的注目を集める中国、COVID-19対策でオードリー・タンIT担当大臣の活躍が注目を集めた台湾の具体的事例を参照しながら、ビッグデータの社会的利用について考える。
- 【第12回】 「中国・台湾社会とデータサイエンス」①  
BATH(バイドウ・アリババ・テンセント・ファーウェイ)と呼ばれるインターネット企業の急成長が世界的注目を集める中国、COVID-19対策でオードリー・タンIT担当大臣の活躍が注目を集めた台湾の具体的事例を参照しながら、ビッグデータの社会的利用について考える。
- 【第13回】 「中国・台湾社会とデータサイエンス」②  
BATH(バイドウ・アリババ・テンセント・ファーウェイ)と呼ばれるインターネット企業の急成長が世界的注目を集める中国、COVID-19対策でオードリー・タンIT担当大臣の活躍が注目を集めた台湾の具体的事例を参照しながら、ビッグデータの社会的利用について考える。
- 【第14回】 「朝鮮研究とデータサイエンス」①  
朝鮮史の分野では、朝鮮は列強の侵略がなければ自力で近代化を成し遂げ資本主義段階へと到達することができたとする資本主義萌芽論が1960年代以降勃興し、それに伴って朝鮮社会の内に発展を生み出す契機を持っていたとする内在的発展論が強く主張された。このような議論に対し、経済史学者を中心に前近代朝鮮社会の計量的分析を根拠として反論が行われた。しかしこれらの議論は今日いずれも力を失っている。なぜ全ての議論が力を失ったのかを考えることで、人文学においてデータの持つ強みと弱みを考えていきたい。
- 【第15回】 「朝鮮研究とデータサイエンス」②  
朝鮮史の分野では、朝鮮は列強の侵略がなければ自力で近代化を成し遂げ資本主義段階へと到達することができたとする資本主義萌芽論が1960年代以降勃興し、それに伴って朝鮮社会の内に発展を生み出す契機を持っていたとする内在的発展論が強く主張された。このような議論に対し、経済史学者を中心に前近代朝鮮社会の計量的分析を根拠として反論が行われた。しかしこれらの議論は今日いずれも力を失っている。なぜ全ての議論が力を失ったのかを考えることで、人文学においてデータの持つ強みと弱みを考えていきたい。