

Explorative Research with Twitter Data: How Much Public Awareness of Biodiversity People Have

地球循環共生工学領域 08E17006 石田愁 (Shu ISHIDA)

**Abstract:** Biodiversity is deteriorating worldwide due to the human activities. For the biodiversity mainstreaming, the policy levers to promote transforming individual lifestyle is needed and awareness analysis by the use of bigdata is expected to understand the individuals' attitudes towards biodiversity. In this research, I conducted an international comparative analysis between Japanese and English tweets referring to biodiversity and ecosystem services to identify and visualized the commonalities and differences by languages. As the result, I found that Japanese tweets focused on the current affairs such as environmental pollution by oil spills and post-Aichi target, on the other hand English tweets insisted to promote the conservation of biodiversity as the basic opinion.

**Keywords:** Twitter, biodiversity, ecosystem services, conservation, public awareness

## 1. 背景・目的

現在、人間活動により人類も含む生物多様性がかつてない規模で失われており、これを克服するべく生物多様性の主流化 (mainstreaming biodiversity) が求められている<sup>1)</sup>。ライフスタイルを転換するための施策を設計するには、人々が生物多様性に対してどのような意識をもっているかを把握することが必要である<sup>2)</sup>。しかし、個々人の意識調査は世界的に不足しており、Twitter や Google Trends などビッグデータを活用した調査が期待されている<sup>2)</sup>。本研究では Twitter で生物多様性に言及したツイートを収集し、自然言語処理技術によって日本語圏と英語圏でのユーザの発言の特徴を国際比較することで、どのような施策がライフスタイルの転換に有効か検証するための知見を得ることを目的とする。

## 2. 研究手法

### 2. 1 ツイートデータの取得

Twitter の検索エンジンに対して、日本語では「生物多様性 OR 生態系サービス」、英語では「biodiversity OR ecosystem services」を検索した。検索期間は 2020 年 9 月 7 日から 10 月 31 日とし、Twitter Search API を用いて、日本語で 11,847 件、英語は 487,159 件のツイートデータを取得した。

### 2. 2 ツイートデータの分解

次に取得したツイートデータから、URL、リツイート (日本語 9,249、英語 392,423)、絵文字や記号などの不要な文字列を削除し、形態素解析エンジン Mecab を用いて単語に分解した。その後、不要語 (stopword) を取り除き名詞のみを抽出した。さらに単語の出現した回数と場所から、(1) 式により単語の重要度を表す tf/idf 値を算出した。日本語圏と英語圏で出現する単語の共通性と独自性を比較するために、Google Translate で日本語を英語に翻訳し、積集合と差集合を可視化した。

$$\text{tf/idf} = \left( \frac{\text{全ツイート中の単語}x\text{の出現回数}}{\text{全ツイート中の全単語の出現回数の和}} \right) \times \log \left( \frac{\text{全ツイート数}}{\text{単語}x\text{を含むツイート数}} \right) \quad (1)$$

### 2. 3 重要な単語の IPBES 概念的枠組みへのマッピング

日本語圏と英語圏で出現する単語のうち、tf/idf 値が上位 100 個の単語を抽出し、IPBES が提唱する人間生態系と自然生態系のつながりを示す IPBES 概念的枠組み<sup>3)</sup> の各構成要素にマッピングした。



図1 日本語圏と英語圏で出現する単語の共通性と独自性

### 3. 結果と考察

#### 3.1 出現回数で重み付けした単語群の可視化

出現回数で重み付けした単語群を図1に示す。日本語と英語の積集合では、日本語は[人, 世界, 周囲]などの生物多様性を連想しにくい単語, 英語は[nature, loss, climate]などの生物多様性を連想しやすい単語が大きく可視化された。差集合では、日本語は熊が人里に出現する被害やモーリシャス重油流出事故を言及した[被害, 重油]など、その時話題になっているイベント, 英語は生物多様性を直接言及する[wildlife, Nature, protect]が大きく可視化された。本研究ではツイートデータを用いて生物多様性に関して人々が持つ関心の地域と言語別の特徴を明らかにすることができた。

#### 3.2 IPBESの概念的枠組みでカテゴリの分類

tf/idf値の上位100単語をIPBESの示す概念的枠組みのカテゴリに分類した単語群を図2に示す。日本語で「間接変化要因」、英語で「自然」のカテゴリの単語群が最も多く可視化された。生物多様性を論じる際に、日本語圏に比べ英語圏ではより生態系の視点を重要と捉えていると考えられる。また日本語では[愛知, 目標, 重油]といった2020年に目標年を迎えた愛知目標やモーリシャス重油流出事故などの時事的話題が大きく可視化されたことからtf/idf値によって長期的なトレンドは抽出できていないと考えられる。一方英語では気候変動や生物多様性の損失と保全などの生活に関わる生物多様性の話題が大きく可視化されたことからtf/idf値によって重要とされる概念が抽出できた。

### 4. 今後の課題

今後、①ツイートデータの取得期間を長くすることで生物多様性に関する長期的なトレンドを抽出する、②日本語と国連6か国語でツイートを比較することを目指す。

### 参考文献

- 1) IPBES (2019): Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat, Bonn, Germany. 56 pages.
- 2) Secretariat of the Convention on Biological Diversity (2020) Global Biodiversity Outlook 5. 208 pages.
- 3) S. Díaz et al. (2015) The IPBES Conceptual Framework - connecting nature and people, *Current Opinion in Environmental Sustainability*, Vol 14, pp 1-16, <https://doi.org/10.1016/j.cosust.2014.11.002>.

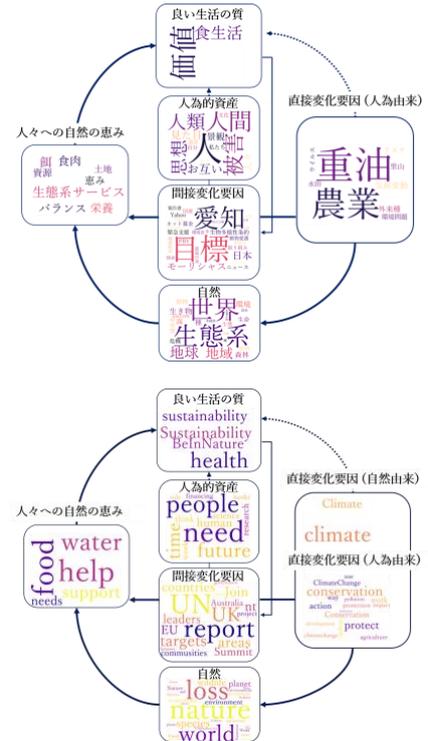


図2 IPBES概念的枠組みへのマッピング