

LGBTに関するG20 各国の Wikipediaテキストの分析

大原圭人, 間中駿介, 赤木茅, 江草遼平, 橋本隆子
千葉商科大学

※本研究は「千葉商科大学・数理データサイエンス教育プログラム」における
「特別講義(データサイエンス)」の一環として実施されている

目次

1.はじめに

2.研究の目的

3.分析手法・
分析結果

4.考察

1 はじめに

1.1 LGBT

1.2 社会情勢と性多様性

1.3 先行研究

1.4 本研究の目的

1.1 LGBT

- LGBT

- Lesbian, Gay, Bisexual, Transgenderの頭文字をとった言葉で、性的マイノリティ（性的少数者）を表す総称のひとつとして使われる（LGBT政策情報センター，2009）
- 性のあり方について，法律上の性別，性自認，性的指向，性表現など様々な要素からの多様性を認める考え（Tokyo Rainbow Pride, 2024）
- LGBTQ, SOGIなどの表現を用いることもある

1.2 社会情勢

- SDGs

- 2030年までに持続可能でよりよい世界を目指す国際目標
- SDGsの目標の中で明確に性的マイノリティを指す言葉は登場しない
- 目標5「ジェンダー平等を実現しよう」、目標10「人や国の不平等をなくそう」、目標16「平和と公正をすべての人に」に関わる（久保, 2020）

- 日本の動向

- 大阪市における調査では4,285人中115人(2.7%)がLGBTに当てはまると回答し、「アセクシュアル」「決めたくない・決めていない」を含めると352人(8.2%)が該当(釜野ら, 2019)
- 2023年6月にLGBT理解増進法が施行されるも、当事者団体、支援者からの批判も見られる（村上, 2023）

1.3 先行研究

- SNSの言説空間における道徳的分断に関する分析では、投稿テキストのネットワーク分析により両言語において道徳的類似によるコミュニティの形成が見られ、LGBTについて扱われる道徳問題として集団への脅威が挙げられていることが示されている（笹原・森, 2019）
- 労働政策研究・研修機構によるLGBTの就労状況に関する調査では、従来より同問題に関する議論が行われてきていた米，英，独，仏を対象に調査が行われ，各国の先進的な取り組みや現状の課題が示されている（労働政策研究・研修機構, 2016）

→LGBTに関する諸問題を全世界的な課題として考える時，研究対象が個別的または少ない組み合わせでの研究である点に課題がある．

2 本研究の目的

2 本研究の目的

- 国毎のLGBTに関する認識に関して、テキストデータからその共通性・差異について明らかにすることを目的とする
 - Wikipediaにおける「LGBT」の記事データ及び世界経済フォーラムが公開するジェンダー・ギャップ指数（2022年データ）
 - 対象国は先進国と新興国を含むG20

3 分析手法・分析結果

3.1 対象データ

3.2 分析手法

3.3 類似度比較

3.4 BERTを用いたクラスタリング

3.1 対象データ

- Wikipedia

- ボランティアの共同作業で執筆及び作成されるフリーのインターネット百科事典
- 多くの言語で同一内容の記事が作成されており，比較可能
- 内容の詳細さや記述の厚みがその言語を使用して記事を作成する作成者グループによって異なることから，各言語における認識について一部を明らかにすることができる
- 各言語のテキストは一度英語に機械翻訳し，その後日本語に機械翻訳する

- ジェンダー・ギャップ指数（2022）

- 世界経済フォーラムが毎年公表する経済活動，政治への参画度，教育水準，出生率，健康寿命などから算出される男女格差を示す指標
- 男女に関する指標ではあるが，ダイバーシティの推進の一側面を表すデータでありLGBTに関する認識と関係が深いと考えられる

3.1 対象データ

各言語と該当するG20の国一覧

言語	国名
German	ドイツ
Spanish	メキシコ, アルゼンチン
France	フランス
English	アメリカ, イギリス, カナダ, オーストラリア, 南アフリカ
Portuguese	ブラジル
Italian	イタリア
Indonesian	インドネシア
Korean	韓国
Chinese	中国
Japanese	日本
Hindi	インド
Turkish	トルコ
Arabic	サウジアラビア
Russian	ロシア

ジェンダー・ギャップ指数（2022）のランキング

※複数国が該当するときは最も人口が高い国のデータを使用

言語	総合	経済	教育	健康	政治
German	6	88	82	64	5
Spanish	33	110	62	49	15
France	40	51	1	76	39
English	43	21	59	78	63
Portuguese	57	86	73	1	56
Italian	79	104	60	95	64
Indonesian	87	87	106	73	81
Korean	105	114	104	46	88
Chinese	107	45	123	145	114
Japanese	125	123	47	59	138
Hindi	127	142	26	142	59
Turkish	129	113	99	100	118
Arabic	131	130	87	114	131
Russian					

3.2 分析手法

- 類似度比較

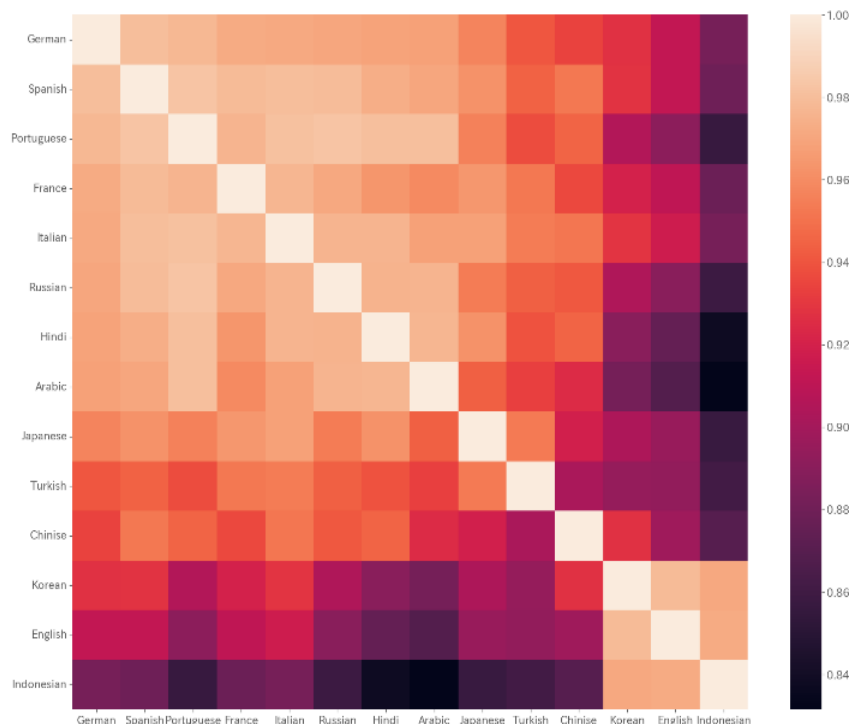
- 自然言語処理モデルであるBERTを用いて記述データをベクトル化
- 各言語の記事間の類似度を言語ベクトルのコサイン距離によって求める
- コサイン距離は、ジェンダーギャップ指数の各項目において最も順位が高い国を基準として算出する

- クラスタリング

- PCAを用いて次元削減し、ward法で階層クラスタリングを行う
- クラスタ毎に頻度の高い名詞・形容詞・動詞についてワードクラウドを作成
- 各クラスタの差を明瞭化するために、50%以上の言語で共通に使用されているワード（例:LGBT）を除外する

3.3 類似度比較

Germanを基準とした文章類似度のヒートマップ



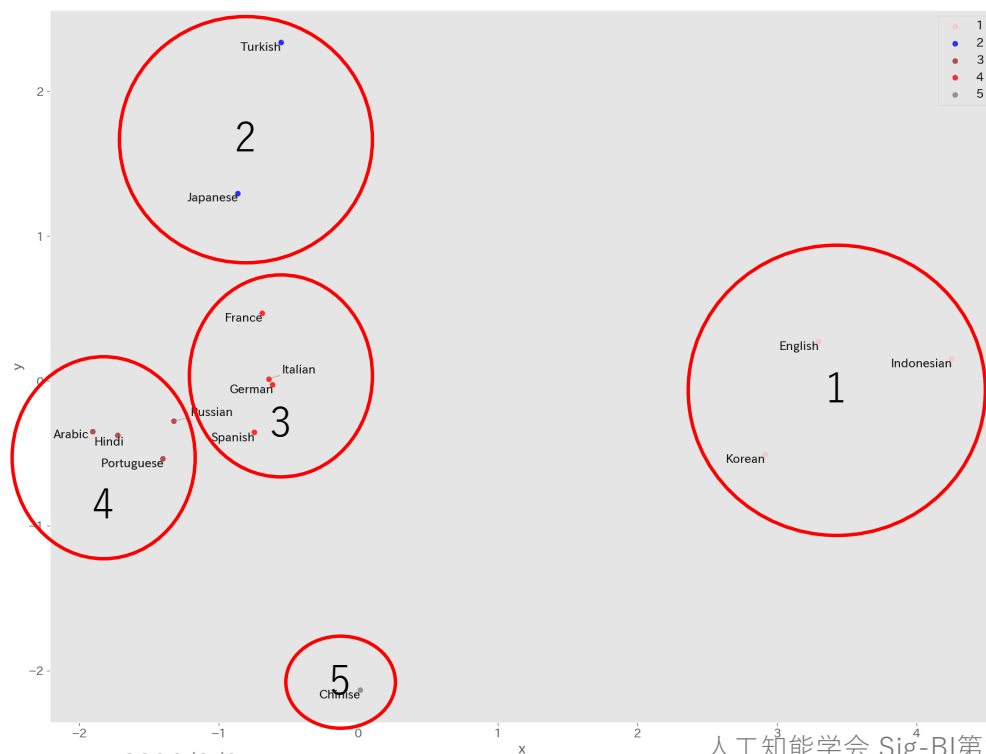
類似度との相関

*各ランキングトップの国との類似度による

ジェンダーギャップ指数 項目	類似度との相関係数
総合	-0.312
経済	-0.498
健康	-0.024
教育	-0.604
政治	-0.426

3.4 BERTを用いたクラスタリング

クラスタリング図



クラスタリング番号と該当言語

クラスタ番号	言語名
1	English, Korean, Indonesian
2	Turkish, Japanese
3	Portuguese, Hindi, Russian, Arabic
4	German, Spanish, France, Italian
5	Chinese

3.4 BERTを用いたクラスタリング

各クラスターの言語の記事における記述をもとにしたワードクラウド



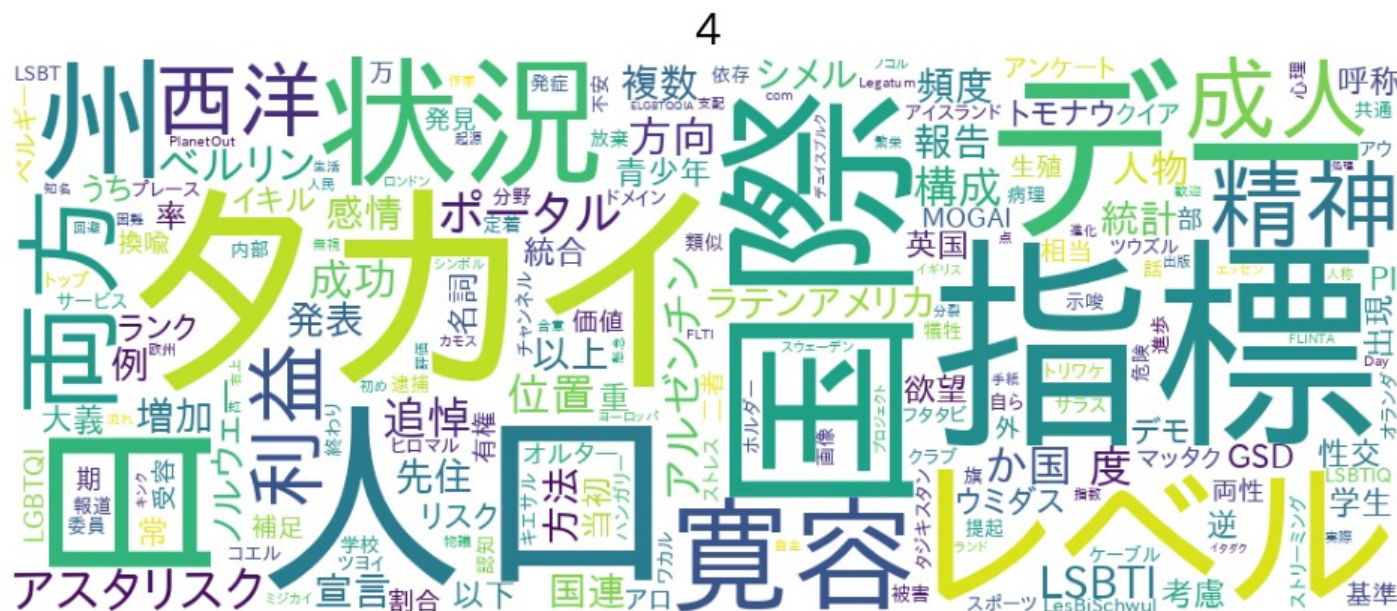
3.4 BERTを用いたクラスタリング

**German, Spanish, France,
Italian**

「国際」「デー」「日」は国際行動デーを指しており、Wikipediaの項目として年間におけるイベント、記念日をリストしていることから強調されて含まれたと推察される

「人口」，「指標」，「高い（タカイ）」は，自国及び諸外国の近年の統計資料の掲載によるものと考えられる

現在の状況やアクションに関する記述が強く現れており、先進的な取り組みや受け入れに関する議論



3.4 BERTを用いたクラスタリング

English, Korean, Indonesian

「切り離す」は、主にEnglishで言及されており白人主導のLGBTコミュニティに異議を唱えた集団が、自分たちをそのコミュニティから切り離して考えるといった意味で使われている。

「フラッグ」は、レインボーフラッグを指す

「アライ」は、LGBTでは無いが
LGBTの人たちの活動を支持し、支援
している人たちのことを指し、
SGMの当事者に含めるかの議論が
ある。

LGBTを取り巻く社会問題について 熱く記述があると見られる



3.4 BERTを用いたクラスタリング

Turkish, Japanese

「ヨーロッパ」、「キリスト」、「古代」、「犯罪」など、LGBTの歴史や外国の実情に関する記述、項目が記事の文章を多く占めていることがわかる。



考察

- Wikipediaにおける「LGBT」というタイトルの記事と「ジェンダー・ギャップ指数」，特に教育には相関がみられた
 - 男女のジェンダー平等性と，LGBTに関する記事内容との関係が見られた
 - 男女共同参画の推進とLGBTに関する認識に何らかの関係があることがわかる
 - ただし，因果については分析できていない
- クラスタリング，ワードクラウドの分析から，クラスタ間のLGBTに関する認識の差異について記述から特徴の可視化が可能であることが示唆された
 - 各言語における認識について，Wikipediaの記述から特徴を示すという目的を達成した
 - Wikipediaの記述の更新を追跡することで変化を見ることができると期待できる
 - 今後の課題として，対象言語，国や採用する記事の種類，Wikipedia以外の比較可能な記述データについてなど，分析をより深める

参考文献

- LGBT政策情報センター(2009):働く人の LGBT 入門 ハンドブック, Retrieved on from https://lgbtjapan.org/PDF/LGBT_handbook.pdf (最終閲覧2024/03/09)
- Tokyo Rainbow Pride(2024): About LGBT, Tokyo Rainbow Pride, Retrieved from <https://tokyorainbowpride.com/lgbt/> (最終閲覧 2024/03/09)
- 久保桂子(2020): 家庭科の家族・家庭生活領域の学習とSDGs, 日本家庭科教育学会誌, 63, 1, p. 33-36.
- 釜野さおり, 小山泰代, 千年よしみ, 布施香奈, 山内昌和, 岩本健良, 藤井ひろみ, 石田仁, 平森大規, 吉仲崇 (2019):大阪市民の働き方と暮らしの多様性と共生に かんするアンケート」 結果速報および Q&Aより, 人口問題研究, 75, 3, 248-253.
- 村上富美(2023.06.15): 「今のLGBT法案は廃案に！」 松中権氏が語る当事者たちの懸念,日経ビジネス, Retrieved from <https://business.nikkei.com/atcl/gen/19/00474/061500014/> (最終閲覧2024/03/09)
- 笹原和俊, 杜宝発(2019): ソーシャルメディアにおける道徳的分断：LGBTツイートの事例, 社会情報学, 8, 2, 65-77,.
- 独立行政法人, 労働政策研究・研修機構, 「諸外国のLGBTの就労をめぐる状況」 ,(2016.05.31) https://www.jil.go.jp/foreign/report/2016/0531_01.html
- 世界経済フォーラム, 「ジェンダー・ギャップ指数2022」 ,(2022.08) https://www.gender.go.jp/public/kyodosankaku/2022/202208/202208_07.html