

# ソーシャルリスニングの日中韓言語展開における課題

白土由佳\*1・華金玲\*2  
yuka@bunkyo.ac.jp

\*1: 文教大学情報学部

\*2: 慶應義塾大学総合政策学部

◎Key Words      ソーシャルメディア, データ分析, ソーシャルリスニング

## 1. はじめに

ソーシャルメディアのデータ分析を通じた社会の理解をソーシャルリスニングと呼ぶ。近年、ソーシャルリスニングは社会を理解するための調査手法として浸透しつつあり、従来型の社会調査などと組み合わせることは有効だと考えられる。

このような状況において、ソーシャルリスニングにおける多言語間の比較が課題の一つとなっている。例えば昨今話題となった生成系AIは世界中で同時多発的に流行を席卷したが、各国の状況を比較するためには、分析対象となるメディアやキーワードの選定など、様々な障壁が立ちだかる。

本研究では、日中韓言語の比較分析を行った研究事例に基づき、ソーシャルリスニングにおいて言語間比較を行うための課題を整理する。

## 2. ソーシャルリスニングとは

ソーシャルリスニングとは、ソーシャルメディア上などで自然に発生した人々の会話や行動などを研究し、人々の生活に基づいた意見を読み取ることである<sup>(1)</sup>。日単位・週単位で調査を行える柔軟性があり、時として調査者の想像の及ばないデータまでもが得られることもあり、仮説探索的なアプローチが可能となる。従来型の調査と相互補完的に用いることで、社会の理解に役立てることができる。

ソーシャルリスニングの実践には、分析対象に関するドメイン知識、データ取得のためのプログラミング、データ分析スキルが求められる。加えて、実践のプロセスには繰り返しが欠かせないことから、筆者はこれまでソーシャルリスニング実践において繰り返しを組み込んだフレームワークを提案し<sup>(2)</sup>、教育の現場で活用している。

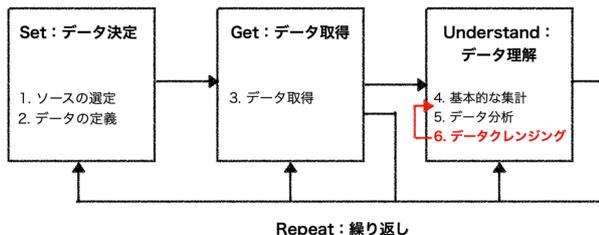


図1 ソーシャルリスニング実践プロセスフレームワーク“SGUR”<sup>(2)</sup>

## 3. ソーシャルリスニングの多言語展開

ソーシャルリスニングの利点の一つに、プラットフォーム上で交わされるコミュニケーションに言語の壁がないという点が挙げられる。例えばTwitterには様々な言語の投稿が蓄積されている。

言い換えると、ソーシャルリスニングは技術的には多言語展開のしやすいアプローチであるとも捉えられる。例えば、任意のテーマについてTwitter上でどのような投稿がなされているのか、英語や日本語、韓国語など言語を指定してデータを取得することそれ自体は容易い。

しかし、同一プラットフォーム上で多言語展開をすればそのままソーシャルリスニングの実現になるかと言えば、そうではない。本研究では、事例を通じてソーシャルリスニングにおける多言語展開の課題について検討し、多言語展開の実践における知見を蓄積したい。

## 4. 研究事例

本研究では、筆者らのChatGPTに対する社会的関心の日中韓比較<sup>(3)(4)</sup>を事例に日中韓言語比較における課題を整理する。この事例は、昨年末に世界中で注目を集めた対話型AIのChatGPTについて、それぞれの言語における投稿の比較を通じて社会的関心の共通点と相違点を示している。

以降では、図1に示したソーシャルリスニング実践プロセスの各段階における日中韓言語展開の困難と課題を整理する。

### 4.1 Set: データ決定

ソーシャルリスニングの第一段階はデータ決定である。事例では、ChatGPTに対する社会的関心を明らかにすることを目的にソースの選定を行った。日本語および韓国語については技術的な話題が共有されやすいTwitterを対象とした。中国語については、中国国内で国外のサービス利用が禁止されていることから、Twitterではなく同様の短文型ソーシャルメディアWeiboを対象とした。

このように、データ決定の際には各言語の置かれた社会的状況によって分析対象とできるプラットフォームが異なる可能性がある。そのような場合には、できる限りメディア特性が類似した対象を吟味し、比較に適切であるかを検討する必要がある。

### 4.2 Get: データ取得

ソーシャルリスニングの第二段階はデータ取得である。

データ取得における困難は、キーワード選定と取得方法の異なりに起因する。

事例では、ChatGPT について述べる投稿を分析対象とした。従って、当初は一律で投稿に「ChatGPT」を含む投稿を対象としたが、データ取得後に韓国語データの不十分さが指摘され、同様の内容を意味する「챗 GPT」も追加キーワードとして追加のデータ取得を行った。

このように、各言語の特徴によって同一キーワードでのデータ取得では不十分な場合がある。それは、単にキーワードを翻訳すればいいわけではないという意味でもある。そのため、言語間比較においては、目的に応じ、言語的背景を鑑みた柔軟なキーワード選定が求められる。

#### 4.3 Understand : データ理解

ソーシャルリスニングの第三段階であるデータ理解には、データクレンジングとデータ分析のそれぞれにおいて困難がある。

データクレンジングは多言語展開において最も難しい点の一つである。事例では、ローデータのクレンジングとテキストマイニングにおけるストップワードの処理という2つの場面で困難が生じた。

ローデータのクレンジングについては、ソーシャルメディア特有の問題であるスパムの扱いが課題となる。例えば日本語においてはパスワードである ChatGPT を付与しただけで全く内容に関係のない投稿はスパムとして判定したが、中国語のデータにおいてそのようなスパムは極めて稀であり、比較分析という点において日本語のみクレンジングの基準を厳しくすることには慎重にならなければならない。

ストップワードの処理については言語それ自体に対する知識の課題がある。事例では投稿本文のテキストマイニングを行ったが、日本語であれば助詞や助動詞で意味をなさない語はストップワードとして分析対象外とする。中国語および韓国語でも原理的には同様だが、どの語をストップワードとするかには言語に対する知識が必要であり、ここに多言語展開の高い壁が見てとれる。語の意味を調べるための翻訳よりも難易度の高い判断が必要とされるためである。

データ分析における大きな困難は技術的な側面である。分析には、KH Coder<sup>6)</sup>や Microsoft Excel を使用したが、データの行き来をする過程で文字化けが生じることが多々あった。瑣末な点であるものの、多言語に関する知識がないと対応できないことからソーシャルリスニングの実践においては重要な課題であると考えられる。

また、データ分析においては少なくとも分析対象の2言語に精通している必要性が認められた。事例では、著者が中国語ネイティブかつ日本語もネイティブレベルに堪能であったことから語の翻訳や分析結果の解釈が行えた。特に事例のテーマが ChatGPT であったことから新語などへの知識も欠かせず、単に機械的な翻訳ではこのような研究水準には達することができなかつたと考えられる。したがって、ソーシャルリスニングの多言語展開においては、分析結果を外部に翻訳依頼する形で取り組むのではおそらく不十分であり、研究者に一定以上の言語レベルが求められる。

さらに、データ分析結果の考察におけるドメイン知識として、諸外国の企業やサービス名、動向といった産業および社会的知識が求められた。対象とした ChatGPT は、新規サービスが次から次へと登場し非常に話題性のある事例のため、サービス提供側と産業構造、その社会文化的背景といった日中韓3カ国、それに加えて世界の関連動向の把握が必要となった。

#### 4.4 Repeat : 繰り返し

ソーシャルリスニングの実践において重要なのは繰り返しのプロセスである。いずれの段階からも繰り返しを行うことで探索的に分析を進めることができる。

事例では、1人の著者が全てのデータ分析を担ったため、繰り返しの困難は分析手続きという物理的な負荷の大きさによるものであった。しかし、もし言語ごとに別の研究者が分析を担当する場合、細かな繰り返しプロセスによる分析の精緻化が必ずしも他の言語に展開できる保証はない。一方で、言語間において分析プロセスの統一が求められることから、繰り返しをどのように行うかについては研究者間で密なコミュニケーションを取り続ける必要がある。

#### 5. 考察

本研究では、事例を通じてソーシャルリスニングの日中韓言語展開における困難と課題を整理した。多言語展開においては、複数言語に対する高度な理解、および分析対象に関連するドメイン知識、分析手続きにおける技術的なスキルが求められる。言語の理解やドメイン知識は一般的な社会調査における多言語展開と同様の要請である一方、技術的なスキルについては言語間の知見の蓄積に乏しいことから、具体的なプロセスの共有化が求められる。

#### 6. 今後の展望

ソーシャルリスニングは仮説探索的な社会調査のアプローチの多言語展開の方向性としての価値を秘めている。実際に多言語比較を行う上での知見を蓄積し続け、グローバルな社会における緻密な理解の方法論的貢献を目指したい。

#### 参考文献

- (1) Rappaport, Stephen D., 電通ソーシャルメディアラボ: “リッスン・ファースト!: ソーシャルリスニングの教科書”, 翔泳社 (2012).
- (2) 白土由佳: “ソーシャルリスニングのフレームワーク SGUR (サグール) の提案と実践”, CIEC 春季カンファレンス論文集, Vol.12, pp.9-16 (2021).
- (3) 華金玲・白土由佳: “対話型 AI に対する社会的反響の日中韓比較—ソーシャルメディアへの ChatGPT に関する投稿から—” 日本メディア学会 2023 年度春季大会 (2023).
- (4) 華金玲・白土由佳: “日中韓における対話型 AI 「ChatGPT」の社会的関心—ソーシャルメディアとマスメディアのタイムラインから—” 情報文化学会誌, Vol.30, No.1 (2023). (掲載予定)
- (5) 樋口耕一: “テキスト型データの計量的分析—2つのアプローチの峻別と統合—” 理論と方法, Vol.19, No.1, pp.101-115 (2004).