

## 内容中心のコーパス分析の可能性 —住みやすい国コーパスの分析例に基づいて—

村田裕美子 (ミュンヘン大学)

李在鎬 (早稲田大学)

<共同研究者>スルダノヴィッチ, イレーナ (プーラ大学)

<共同研究者>トリチコヴィッチ, ディヴナ (ベオグラード大学)

### 1. 背景と主張

これまでの日本語教育分野におけるコーパス研究では、学習者コーパスは学習者の言語実態を解明するために用いられ、学習者が用いる「言語」に注目した研究が多かった。このようなアプローチに対して、本発表では「内容」に注目した分析の方法論を提案し、コーパスの分析結果を異文化間教育のための発見学習の材料として活用できることを主張する。本発表は、「内容」に注目したコーパス研究であり、「言及度」という指標で出現語彙を整理したあと、個々の用例を詳細に観察記述していくことを目指すものである。こうしたアプローチは学習者一人一人の声を拾うものであり、言語が持つ「伝えるツール」としての側面を重視するものでもある。この主張の妥当性を示すため、「住みやすい国コーパス」(<https://sumiyasui.jpn.org/>, 村田 2021) を用いた研究事例を紹介する。

### 2. 先行研究

これまでの学習者コーパス研究の多くは「頻度モデル」に基づく研究であった。というのは、これまでの学習者コーパス研究では、データを品詞分解し、語の頻度をもとに学習者の過剰・過少使用の特徴や誤用、母語干渉などの言語の使用実態を明らかにしてきたからである(野田・迫田(編) 2019, 李 2020, 石川 2021)。こうした研究を推進するためのコーパス開発も進んでおり、「I-JAS 多言語母語の日本語学習者横断コーパス」(迫田・石川・李(編) 2020)は、12の母語を持つ1000名の学習者の産出データを収録した大規模コーパスとして一般に公開され、学習者の言語使用の実態解明に大きく貢献している。こうした「頻度モデル」に基づくコーパス研究は、使用頻度のような計量的指標で一般化するものが多く、多様性よりは普遍性を明らかにする研究が主であった。

一方、私たちの研究グループでは、「内容モデル」に基づくコーパス研究を提案する。このアプローチは、言語とは思考を表現したものであるという前提のもとで、意味内容に注目したコーパス研究である。具体的には、共起ネットワークでコーパス全体の語の分布を確認した上で、特定の語を取り上げ、KWIC (keyword in context) データのコンコーダンスラインを詳細に観察し、一人一人の思考の多様性を明らかにする(村田・トリチコヴィッチ・李 2020, 2022)。このようにコンコーダンスラインを丁寧に観察すると、一人一人異なる考え方や感情、評価基準が表れ、育った環境やその人がこれまで積んできた経験が影響

していることが明らかになっている。本研究では、「内容モデル」に基づく学習者コーパスのケーススタディを示し、学習者たちの思考の断片を明らかにするとともに、異文化間教育の教材としてのコーパス利用の可能性を検討する。

### 3. 本研究の課題

本研究では、社会的背景の異なる3つの国の大学生が同じテーマ「住みやすい国の条件とその理由」で書いた意見文をデータとして用いて、3か国で共通して出現している語彙を数値化して確認し、その語がどのような文脈で用いられているのかを用例を見ながら明らかにしていく。具体的には、以下2つの課題を明らかにする。

課題1：それぞれの国で共通して用いていた語にはどのようなものがあるか。

課題2：共通して用いていた語がどのような文脈で用いられているのか。

### 4. データと分析方法

#### 4-1. データ

本発表では、2022年12月に公開された「住みやすい国コーパス」に収録されている「住みやすい国コーパス書き言葉編」を使用する。「住みやすい国コーパス書き言葉編」には、ヨーロッパの日本語学習者が「住みやすい国の条件とその理由」というタイトルで書いた意見文が収集されている。ここでは、ヨーロッパの学習者が現在の社会をどのように捉え、これからどのような社会を作るべきかなどの考えが日本語で述べられている。本コーパスは、日本語学習者たちが日本語学習を通して、自分たちの社会を批判的観点から捉え、それを言語化したコーパスであると言える（村田 2021）。

本発表では、公開されているドイツ（20名）、セルビア（20名）、クロアチア（20名）の中級学習者のデータを利用する。なお、学習者のレベルは、意見文とともに提供されている日本語の客観テスト「SPOT90 (Simple Performance-Oriented Test, 小林 2015)」の結果に基づくものである。

以下、例としてドイツ語話者の中級学習者の作文を紹介する。データサイズは表1のとおりである。

表1 コーパスサイズ

国 (参加者)	延べ語数	一人当たりの延べ語数	一人当たりの異なり語数
ドイツ (20)	8131	406.55	148.55
セルビア (20)	5421	271.05	108.80
クロアチア (20)	5962	298.10	114.10

解析辞書 UniDic と形態素解析エンジン MeCab の解析結果に基づき算出

【学習者作文例】

楽に暮らせる国といえば、諸々の条件がある。しかし、楽に暮らすこととは具体的に何であろうか。私の意見では悩み無しで、平和で安全で安心して暮らすということである。

まずは当然のことであるが人々沢山が同じ国に住んでいるなら、食料品が十分であり、衛生が良いのが必要不可欠である。食糧不足、悪い衛生や治療がないのであれば、安心して暮らせないのであろう。そのため、入りやすい健康保険があると良いと思う。

他にとても大切な点は教育である。成長する時に教育を受けることが大事なので、人々が重圧無しできちんと学ぶことができれば、楽に暮らせる。そして、各自に合っており、尊重される仕事が十分であったり、守られる労働法もあつたりすれば暮らしやすいと思う。各自に合っている仕事とは自分が毎回職場に行くのを楽しむことができ、自分の能力を使える仕事である。

教育、仕事や生活様式といえば諸々の方法があり、自由に選択できると良いと思う。やはり、一番大切な条件が重圧無しで、自分と他人の思想や宗教の関係無しで、個性的にありのままに暮らせることである。国ではそうであると、安心できるし楽に暮らせる。

#### 4-2. 分析方法

本研究の分析では、「KH Coder」(<https://kncoder.net/>, 樋口 2014)で「住みやすい国コーパス」を格納し、形態素解析済みデータを作成した。そのあと、共起ネットワークを作成し、使用語彙の分布を確認した。特に3か国で共通する語彙に注目し、以下に示す「言及度」という指標で、言及の度合いを評価した。

$$\text{言及度} = (\text{事柄に言及した人の人数} / \text{母集団の総数}) * \text{使用頻度}$$

次に3か国に共通して同程度に出現していた「お金」に注目し、コンコーダンスラインの観察を通して詳細な分析を行った。

### 5. 結果

#### 5-1. 課題1の結果

3か国において共通して用いられていた語彙は「生活」「仕事」「教育」「人々」「自分」「お金」「国民」「社会」「経済」「人間」「人生」であった。さらに、これらの語彙の「言及度」を調べた(表2)。その結果、3か国で同程度の「言及度」で用いられていた語は「お金」と「人生」であった。

表2 共通語彙の「言及度」(「言及度」の高さ順にソート)

言及度	生活	仕事	教育	人々	自分	お金	国民	社会	経済	人間	人生
ドイツ	19.80	12.00	4.80	3.50	12.50	3.50	3.85	5.60	1.40	2.70	1.35
セルビア	44.20	22.10	30.60	8.00	4.00	3.00	1.05	1.80	4.20	2.00	1.40
クロアチア	8.50	5.40	0.80	11.40	2.00	2.40	3.50	1.05	0.75	0.90	1.50
全員	66.30	37.33	25.13	22.05	16.48	8.87	8.00	7.73	5.60	5.37	4.40

## 5-2. 課題2の結果

課題1の結果をもとに、3か国で同程度に用いられていた「お金」に注目し、用例分析を行った。「お金」を選んだ理由は、「人生」よりも価値観の違いが顕在化しやすいと考えたからである。分析の結果、「お金」は、以下の3つの文脈で用いられていることが明らかになった。1) 個人の生活や経済活動を支えるもの、2) 個人が人間らしく生きるために必要なもの、3) 国や社会を豊かにするために必要なもの、の3つである。

### 1) 個人の生活や経済活動を支えるもの

1つ目は「仕事」や「給与」と共起する「お金」であり、作文には「お金」は個人の生活や経済活動を支えるものであるという考えが表れている。

(例1) 普段のシステムが決まったら、好きな仕事をもって、お金を稼いで、いろいろな生きるための必要な物を買うことができます。結局は、商業も必要です。

(例2) 給与はいつか、分からないですが、本当に大切じゃありません。年を追うごとにお金はだんだんと無意味になります。物価と値段だけを見たら世界は虚しくなります。

### 2) 個人が人間らしく生きるために必要なもの

2つ目は学校へ行くための「学費」や病院へ行くための「保険」と共起する「お金」であり、作文には「お金」は個人が人間らしく生きるために必要であるという考えが表れている。

(例3) 学校はみんなが参加できるのは大事です。そのため学費がなければいいと思います。なぜならもし学費が高かったら、あまりお金がない家族の子供が学校に参加できないからです。

(例4) 保険証を持っていない人は医者で高い代金をしなければなりません。そうならあまりお金を持っていない人は病気のときに医者へ行けません。

### 3) 国や社会を豊かにするために必要なもの

3つ目は国の貧困に対する「お金」であり、作文には「お金」は社会を豊かにするために必要であるという考えが表れている。

(例5) 6時45分にアパートから出てホームレスなしが容器の中から食べ物の残骸を探し、お金を「稼ぐ」ためにペットボトルを集めているを見ない。

(例6) 国はいい経済があればお金が多いです。それから払わなければならない精勤〔税金:発表者補足〕は安くて社会扶助は多いです。

## 6. 考察

これまでの学習者コーパス研究は、頻度をもとに表面的な一般化を目指すものが主であった。しかし、言語は思考のツールであるため、内容にも注目しなければならない。こうした問題意識のもとで、「言及度」という指標で学習者コーパスの語彙使用を観察していくと、語彙使用の背景には、文脈という膨大な情報が隠れていることがわかる。

本研究で取りあげた「お金」は程度こそ人それぞれであるが、生きるためになくはならないものの1つである。そのため、異なる国で日本語を学ぶ大学生に「住みやすい国の条件とその理由」というタイトルで意見文を書いてもらったとき、国を問わず「お金」がその条件や理由として共通して出てくるのはそれほど驚くことではない。しかし、「お金」の価値や機能は1つではない。そのため、「お金」のどのような価値が重要であるか、「お金」のどのような機能が重要であるかという考えは多岐にわたる。本研究では、学習者の意見文を「お金」という語で絞り込み、その文脈を読むことで、彼らの「お金」に対する考え方、特に「お金」のどのような機能が重要であるかという考え方が明らかになった。

## 7. 実践に向けた提言：コーパスを用いた教育実践

本研究のアプローチは、同じ日本語を学ぶ学習者たちでありながら、自分と異なる環境で生きる人々の考えを知る学習や、自分自身の考えを深める学習として、異文化間教育の、特に議論の場を提供する教材として役立てられると考えている。例えば、作文の課題として、自分の学習者に同じテーマで作文を書いてもらい、他の国の学習者の作文と照らし合わせ、内容を比較し、類似点や相違点を探す活動、また他者が書いた意見に対する自分の意見を問いただす活動などが可能である。もちろん、語彙や文型に注目し、日本語の表現方法を学ぶ活動としても可能である。そこには、これまでのような教育のために加工した日本語ではなく、生(なま)の日本語だからこそ得られる発見学習ができると考えられる。

以上から、学習者コーパスは、一般的には研究のリソースとして認識されているが、筆者らは、コーパスデータは教材としても活用できると考えている。研究で扱っている「住みやすい国コーパス」の場合、多様性と普遍性が混在した意見文のコーパスであると言える。というのは、国や育った環境が異なっているため、考え方や物事に対する評価の基準も異なる一方で、国や育った環境の影響を受けないものもあるはずだからである。こうした特性を活かすことで、学習者コーパスは他者の考えを知り、自分の考えを深める教材としても利用できると考えている。とりわけ、近年、注目されている CLIL (Content and Language Integrated Learning, 奥野ほか編著 2018) の学習コンテンツになると考えている。

## 8. まとめ

本研究では、従来の「頻度モデル」による研究とは別に「内容モデル」に基づくコーパス研究の事例を紹介した。具体的には量的分析によって、国単位で共通する、または共通しない語彙の特徴が発見できることを示した。その結果をふまえて、質的分析を行い、同

ビテーマで書いているからこそ見えてくる, そこに住んでいる人々の考えの相違点, 類似点に着目できることを示した。量的分析においては, 「言及度」という指標を提案しており, ドイツ, セルビア, クロアチアの学生が共通して用いていた語「お金」をもとに質的分析を行なった。質的分析の結果, 学習者一人一人が考える「お金」に対する考え方の多様性が明らかになった。そして, この結果をふまえて教育実践の異文化間教育の場で活用できることを示した。

自分と同じ立場である学習者の意見文を読み, 話し合う活動を通して, 他者の考え方に触れることができ, 自分の価値観の内省を促すことができる。日本人が書いた標準的なテキストだけが教材ではなく, 隣の友人が書いた作文もまた思考力を促進する教材として有効である。それがこれからの多様な社会の真正性であると考えられる。

### 参考文献

- (1) 石川慎一郎 (2021) 『ベーシックコーパス言語学 第2版』 ひつじ書房.
- (2) 奥野由紀子 (編) (2018) 『日本語教師のための CLIL (内容言語統合型学習) 入門』 凡人社.
- (3) 小林典子 (2015) 「SPOT」 李在鎬 (編) 『日本語教育のための言語テストガイドブック』 110-126, くろしお出版.
- (4) 迫田久美子・石川慎一郎・李在鎬 (編) (2020) 『日本語学習者コーパス I-JAS 入門』 くろしお出版.
- (5) 野田尚史・迫田久美子 (編) (2019) 『学習者コーパスと日本語教育研究』 くろしお出版.
- (6) 樋口耕一 (2014) 『社会調査のための計量テキスト分析—内容分析の継承と発展を目指して—』 ナカニシヤ出版.
- (7) 村田裕美子 (2021) 「小規模コーパスの構築方法」 李在鎬 (編) 『データ科学×日本語教育』 34-53, ひつじ書房.
- (8) 村田裕美子・トリチコヴィッチ, ディヴナ・李在鎬 (2020) 「異文化間能力の育成を目指す計量テキスト分析—ドイツ・セルビア・日本の学生を対象に—」 『第64回大会計量国語学会予稿集』, 43-48.
- (9) 村田裕美子・トリチコヴィッチ, ディヴナ・李在鎬 (2022) 「ドイツ・セルビア・日本の大学生が考える「住みやすい国」とは何か—複言語・複文化能力の構築を目指す作文活動—」 『ヨーロッパ日本語教育』 第25号, 275-286.
- (10) 李在鎬 (2020) 「日本語教育学の課題に対して計量分析は何ができるか」 『計量国語学』 32(7), 372-386.

### 資料

- (1) 「住みやすい国プロジェクト」 version1.0  
< <https://sumiyasui.jpn.org/> > (2023年3月15日参照).