

# 文化差事例の収集手法の提案と評価

吉野 孝<sup>1,a)</sup> 宮部 真衣<sup>2,b)</sup>

**概要:** 多言語間コミュニケーションにおいて、同一の単語を用いて会話をしている場合でも、相手の文化について十分に理解していないために、誤解が生じる可能性がある。我々は、文化差の検出手法を提案し、文化差可視化システムの開発を行っている。文化差可視化の目的は、文化差理解支援である。文化差の可視化の方法としては、お互いの言語における語句の説明を、文字（説明文）を用いる方法や画像を用いる方法が考えられる。我々はこれまでに、文化差理解支援における、説明文および画像による文化差の可視化効果の検証を行った。検証実験の結果、画像による文化差の可視化は、説明文に比べ、短時間での文化差の判定を可能にし、文化差判定の主観的な負荷が低いことを示した。しかし、「説明文」の提示および「画像」の提示は、いずれも文化差理解支援において、十分に貢献しない可能性があることもわかった。そこで、低負荷かつ確かな文化差理解の支援手法を実現するためのリソースとして、Web上に存在する「文化差の理解を促す事例」に着目した。本稿では、文化差事例の収集サービスを提案し、サービスの実現に必要な文化差事例の収集手法および収集した文化差事例の適切さについて述べる。

**キーワード:** 多言語間コミュニケーション, 文化差, 文化差事例, Wikipedia, 可視化, Web サービス

## Proposal and Evaluation of Cultural Difference Cases Collection Method

**Abstract:** There is a possibility that the misunderstanding is caused in multilingual communications, because people cannot understand enough other culture even when talking by using the same word. We propose cultural difference detection methods and develop a cultural differences visualization service. The purpose of the cultural differences visualization is to support cultural differences understanding. There are two methods to visualize cultural differences using words and phrases and using images for each word in each language. We verified the effect of visualization of cultural difference understanding supports between using words and phrases and using images. From the result of the verification experiments, we found that the method of cultural differences by images enabled the judgment of a cultural difference in a short time compared with that of cultural differences by words and phrases. However, the accuracy rate of cultural difference understanding is low. Both methods may not fully contribute from a viewpoint of cultural difference understanding support. Therefore, we focus on the “case example of a cultural difference” on the Web. This paper proposes a collection service of a cultural difference case. Then, we present the collection method of cultural difference cases and the collected appropriateness of cultural difference cases.

**Keywords:** multilingual communication, cultural difference, cultural difference case, wikipedia, visualization, Web service

### 1. はじめに

多言語間コミュニケーションにおいて、同一の単語を用いて会話をしている場合でも、相手の文化について十分に理解していないために、誤解が生じる可能性がある [1]。文化の違いに基づく誤解を回避するためには、用いている単語に文化差があることを、話者に認識させる

<sup>1</sup> 和歌山大学 システム工学部  
Faculty of Systems Engineering, Wakayama University,  
Wakayama-shi, Wakayama 640-8510, Japan

<sup>2</sup> 東京大学 知の構造化センター  
Center for Knowledge Structuring, The University of  
Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan

a) yoshino@sys.wakayama-u.ac.jp

b) mai.miyabe@gmail.com

必要がある。しかし、相手の文化に関する十分な知識が必要となるため、文化差の有無の判断し、違いを理解することは容易ではない。

我々はこれまでに、まず、文化差検出に関する検討を行った。語句に関する文化差の有無については、Wikipediaの項目数に基づく手法、記事内の国名・言語名数に基づく手法、記事における執筆者の意図に基づく手法を統合することにより、比較的高い精度での判定を可能にした[2]。さらに、日本語版および中国語版 Wikipediaにおける相互言及の有無に着目することで、従来手法よりも精度よく文化差を検出でき、文化差の検出を日本語と中国語以外へ適用できる手法を提案した[3]。

さらに、我々は、文化差理解支援のために、説明文および画像による文化差の可視化効果システムの構築を行い、その効果の検証を行った[4]、[5]。検証実験の結果、画像による文化差の可視化は、説明文に比べ、短時間での文化差の判定を可能にし、文化差判定の主観的な負荷が低いことを示した。しかし、現時点の「説明文」の提示および「画像」の提示は、いずれも文化差理解支援において、十分に貢献しない可能性があることもわかった[4]。

そこで、低負荷かつ的確な文化差理解の支援手法を実現するためのリソースとして、Web上に存在する「文化差の理解を促す事例」に着目した。「文化差の理解を促す事例」とは、誰かの経験などをもとに作成された、文化差が明確に記述された文章である。「文化差の理解を促す事例」は、文化の違いを端的に説明しており、これらを収集しておくことで、文化差理解支援における重要なリソースとして利用できる可能性がある。異文化間における誤解などの事例は、従来から様々な研究者が紹介している[6]。しかし、多様な語句への対応は困難であり、情報技術を用いて収集する必要がある。

本稿では、文化差事例の収集サービスを提案し、サービスの実現に必要な文化差事例の収集手法および収集した文化差事例の適切さについて述べる。

## 2. 関連研究

関連研究として、まず、異文化間コミュニケーションにおける、文化差に関する研究を示す。Choらは、異文化話者らがコンピュータとネットワークを介してコミュニケーションを行う際に用いる絵文字に着目した。絵文字は、異文化間で普遍的に解釈されないという問題がある。そこで、その問題を解決するために、解釈に文化差のある絵文字の検出における工学的な手法の適用可能性について検討した[7]。検討の結果、従来の工学的な手法では、人の文化差判定を近似することは困難であることを示した。Kodaらは、アバターを介したコミュニケーションにおける、異文化間での表情の解釈に着目した。アバターの表情に関するユーザの解釈について実験を行った結果、表情の解釈が文化によって大きく異なるこ

とを示した[8]。文化差に関しては、これまでにいくつかの検討が行われているが、文化差判定は容易ではない。

Wikipediaの複数の言語版を利用した研究を示す。藤原らは、自国の言語版のWikipediaだけでは情報量が不足する場合の補完のために、多言語版のWikipediaに対して、リンク構造解析を用いることで、差異情報を抽出する方法を提案している[9]。松浦らは、日本語と外国語での同一ニュースに関する変遷を分析するために、Wikipediaを用いている[10]。吉岡は、機械翻訳システムの精度向上のために、Wikipediaの言語間リンクを用いた中日の翻訳辞書の作成方法を提案している[11]。

このように、Wikipediaは知識抽出分野で資源として注目を集めており、様々な利用が検討されている。しかし、これまでに、Wikipediaの多言語データを利用した文化差検出に関する試みは行われていない。異文化間における誤解などの事例は、従来から様々な研究者が紹介している[6]、[12]、[13]、[14]。しかし、多様な語句への対応は困難であり、情報技術を用いて収集する必要がある。

## 3. 文化差の定義

本章では、本研究で検出・理解支援の対象とする「文化差」について定義する。

文化差を定義するためには、まず、「文化」の定義が必要である。「文化」(Culture)の定義は、日本と欧米では異なり、一概に定義することは困難である[12]。例えば、今日欧米で用いられる「文化」は、「知識、信仰、芸術、道徳、慣習、その他社会の一員としての人間によって獲得される能力や習慣を包含する複合体である」と定義づけられている[12]。このような「文化」を単純に「測る」ことは困難であるが、コミュニケーション支援に文化差検出手法を適用するためには、何らかの尺度を考える必要がある。そこで本稿では、特に「知識」の面から「文化」を捉えることとし、形式知化された知識の違いで文化差を測ることとした。

次に、「第1種の文化差」と「第2種の文化差」を定義する。「第1種の文化差」のある内容は、一方の文化圏で発生したり、存在したりしている「もの」や「こと」で、基本的には、もとの文化圏の内容を指しているが、伝わっている知識が限定的であり、もとの文化圏における解釈が完全には再現されないものである。例えば、日本の地名のいくつかは、海外にも伝わっているが、その地名の持つ背景(歴史的あるいは文化的)などは正確には理解されない。「第2種の文化差」のある内容は、どちらの文化圏にも存在するが、それぞれの文化圏で意味の異なるものである。例えば、「醤油」は日本と中国のどちらにも存在するが、日本の「醤油」と中国の「醤油」は異なる\*1。

\*1 中国の醤油は、味やにおいが日本の醤油よりも強い。中華料理での使われ方は、香りや味よりも、色づけに重点を置いてい

## 4. 文化差事例の収集

### 4.1 文化差理解支援において提供すべき文化差事例の要件

我々のこれまでの実験結果から、「説明文」<sup>\*2</sup>の提示および「画像」<sup>\*3</sup>の提示は、文化差の理解に関して、十分に貢献しない可能性のあることがわかった [4]。そこで、これまでの検証実験をもとに検討した、文化差理解支援において提供すべき文化差事例の要件を以下に示す。

#### (1) 機械翻訳を利用せず、母語の情報を提示

文献 [4] では、機械翻訳を用いて他国の語句の説明を行った。しかし、ぎこちない機械翻訳結果の解釈に大きな負担があり、十分に文化差の理解を促すことができないことがわかった。そのため、母語で流暢に記述された情報の提示が必要である。

#### (2) 具体的な違いの事例を提示

文献 [4] では、語句の説明として、Wikipedia の概要を提示した。しかし、同じ語句を説明する各言語版 Wikipedia の概要の大部分は類似しており、母国と他国での内容の両方を提示しても、文化差の理解が容易でないことがわかった。そのため、具体的な違いに関する説明を提示する必要がある。

### 4.2 文化差事例の収集手法

我々は、誰かの経験などをもとに作成された「文化差の理解を促す事例」は、4.1 節で述べた要件を満たすと考えた。そこで、「文化差の理解を促す事例」(以降、文化差事例と呼ぶ)の収集手法について検討を行った。

本研究では、文化差事例の収集手法として、次の Web 検索エンジンを用いた手法を提案する。

- (1) 「日本では」「中国では」<検索語>の3語を、完全一致検索を用いて、Web 検索を行う<sup>\*4</sup>。
- (2) 検索結果に記載された Web ページを直接参照し、そのページに含まれるテキストデータを取得する。
- (3) テキストデータから、「日本では」「中国では」「日本の<検索語>」「中国の<検索語>」のいずれかが含まれる文を、「文化差事例の候補」とする。
- (4) 「文化差事例の候補」全体の中に、<検索語>が1個以上含まれている場合には、「文化差事例の候補」全体を、その Web ページにおける<検索語>の「文化差事例」として蓄積する。

今回、検索語として、「日本では」「中国では」を用いた。これらを用いた理由は、執筆者が文化差を意図的に記述する際に、これらの語句が用いられる可能性が高い

る。日本の醤油は、中国から日本に伝わり、日本で独自に改良されたと言われており、製造方法も異なる。また、世界各国に輸出あるいは現地生産されており、広く受け入れられている。

<sup>\*2</sup> 説明文の提示では、日本語と中国語の Wikipedia の説明文を用いた。中国語は、機械翻訳を用いて日本語にした。

<sup>\*3</sup> 画像の提示では、Google 画像検索の検索結果を用いた。

<sup>\*4</sup> Web 検索エンジンとして、日本語版の Google 検索を用いた。

と考えたからである [15]。

## 5. 検証用のデータセット

提案手法を評価するために、提案手法の検証用のデータセットが必要となる。また、従来手法と比較するために、文化差を説明する説明文、語句の文化差を表す画像が必要となる。本章では、提案手法の検証用のデータセットの作成手順および比較に用いる文化差を説明する説明文、語句を表す画像の作成方法について述べる。

### 5.1 検証用の語句の収集

本評価で用いる検証用の語句は、著者らが独自に選定・収集したものをを用いた。検証用の語句は、次の手順で収集した。

Step 1 本学のシステム工学部および大学院システム工学研究科の学生に依頼し、下記に該当する日本語の語句を収集した。

- 日本独特のもの(と思っている語句)
- 日本にも海外にもあるが、日本と海外とは違うもの(と思っている語句)

ただし、地名、人名、固有名詞は、ほとんど第1種の文化差のある語句に該当するため、入力しないように依頼した。

Step 2 収集した語句が Wikipedia の日本語版と中国語版の両方に、記事として存在するか調べた。

Step 3 日本に3年以上住んでいる本学の中国人留学生5名に依頼し、Step 1 で収集した語句のうち Wikipedia に記事(日本語版と中国語版の両方)として存在するもの(Step 2 で調べた結果によって抽出したもの)を、次の3種類に分類してもらった。

- 中国にはない(第1種の文化差)
- 中国にもあり、日本と同じ(文化差なし)
- 中国と日本のものは違う(第1種の文化差または第2種の文化差)

「中国と日本のものは違う」を選択した場合には、どのように違うかを簡単に記述してもらった。

Step 1 で、日本語の語句が、2,200 語句以上集まった。Step 2 で、入力語句の5%程度について、入力された順に Wikipedia の記事の有無を調べ、114 語句をデータセットとした。

表 1 に「第1種の文化差」に分類された語句(42語)を、表 2 に「第2種の文化差」に分類された語句(41語)を、表 3 に、「文化差なし」に分類された語句(31語)を示す。本稿では、Step 3 において、「第1種の文化差」と「第2種の文化差」と分けられた語句を、まとめて「文化差あり」としても扱う。また、表 4 に、「第1種の文化差」および「第2種の文化差」に分類された語句の理由を示す。なお、語句の分類は、分類作業を行った各中国人留学生の経験に基づいた意見をもとにしてい

表 1 「第 1 種の文化差」に分類された語句 (42 語)

Table 1 Words set classified as “the cultural difference of the first kind” (42 words)

おせち料理	こたつ	和菓子	赤飯	舞妓	侍
おにぎり	ちくわ	牛丼	節分	萌え	能
カラオケ	ニート	剣道	相撲	漫才	
すき焼き	わさび	刺身	団子	味噌	
パチンコ	七五三	寿司	忍者	妖怪	
ひな祭り	親子丼	神社	納豆	浴衣	
羽根突き	天ぷら	神主	俳句	落語	
おたく	天津飯	神道	柏餅	祭	

語句の分類は、日本に 3 年以上滞在している中国人留学生 5 名が行った。各自の経験に基づいた意見をもとにしているため、事実と異なる可能性がある。

表 2 「第 2 種の文化差」に分類された語句 (41 語)

Table 2 Words set classified as “the cultural difference of the second kind” (41 words)

バレンタインデー	タクシー	マナー	元日	煎餅	茶
インターネット	ラーメン	年賀状	昆布	大人	
カレーライス	公衆便所	貨幣	七夕	地震	
ファッション	就職活動	干支	首相	豆腐	
クリスマス	中華まん	漢字	宗教	漫画	
自動販売機	アニメ	教員	祝日	老人	
アイドル	ちまき	空気	醤油	饅頭	
コンビニ	ビール	軍隊	政府	禅	

語句の分類は、日本に 3 年以上滞在している中国人留学生 5 名が行った。各自の経験に基づいた意見をもとにしているため、事実と異なる可能性がある。

表 3 「文化差なし」に分類された語句 (31 語)

Table 3 Words set classified as “no cultural difference” (31 words)

ようかん	ケーキ	時刻表	観光	新聞	母親	虹
携帯電話	ゲーム	障害者	教授	戦争	友情	夫
固定電話	外国人	映画	仕事	発酵	鬼	
新婚旅行	公務員	花火	子供	父親	銃	

語句の分類は、日本に 3 年以上滞在している中国人留学生 5 名が行った。各自の経験に基づいた意見をもとにしているため、事実と異なる可能性がある。

るため、事実と異なる可能性がある。表 4 に示した分類結果は、5 名の留学生による分類結果を集計し、最も多い分類結果 (2 名以上が分類) を代表値とした場合の結果をもとに分類した。「日本と中国との違いについて」の内容は、留学生らの記述をもとにしている。

## 5.2 比較用の説明文の作成

提案手法との比較で用いる比較用の説明文は、Wikipedia の説明文を用いている。比較用の説明文は、5.1 節で収集した文化差判定済みの 114 語を対象として、次の手順で作成した。

表 4 「文化差あり」に分類された語句とその理由の一部

Table 4 A part of experimental words set and the reason classified to have a cultural difference.

ID	語句	分類結果	日本と中国との違いについて
6	干支	日本とは違う	日本では、亥(いのしし)だが、中国では豚。
28	刺身	中国にはない	中国では、あまり生の食べ物は食べない。
45	煎餅	日本とは違う	中国では煎餅は主食。日本ではお菓子。
53	天津飯	中国にはない	中華料理に、ご飯の上に卵を載せる料理はない。
57	ラーメン	日本とは違う	日本のラーメンとは味が違う。
100	パチンコ	中国にはない	中国ではギャンブルは禁止されている。
108	納豆	中国にはない	中国に納豆はない。
112	醤油	日本とは違う	中国の醤油は塩辛い。
113	おにぎり	中国にはない	中国では、冷めたご飯は食べない。
114	饅頭	日本とは違う	日本の饅頭の中には、餡がある。

「中国にもあり、日本と同じ」と分類されたものは省いている。分類は、日本に 3 年以上滞在している中国人留学生 5 名が行った。

判定結果は、留学生各自の経験に基づいた意見をもとにしているため、事実と異なる可能性がある。2 名以上が分類したものを結果とした。

「日本と中国との違いについて」の内容は、留学生らの記述および Wikipedia の内容をもとにしている。

**Step 1** 該当語句の Wikipedia 記事から、導入部<sup>\*5</sup>を抽出する。今回の比較用の説明文の評価者は日本人であるため、中国語版は機械翻訳を用いて日本語へ翻訳する<sup>\*6</sup>。

**Step 2** 導入部がない記事の場合には、Wikipedia の解説本文の最初の部分を中心に抜き出す。最初の部分が、記事の内容を示すものとして不適切な場合には、他の部分を抜き出す。

**Step 3** 日本語の説明文と中国語の説明文 (機械翻訳結果) を比較して、文章量に大きく差がある場合には、同程度 (差が 2~3 倍程度以内) になるように、文章量を調整する<sup>\*7</sup>。

表 5 に、各言語の説明文の分量を示す。平均文字数は中国語の方が約 20% 多い。また、最長文字数および最短文字数も中国語の長い。今回、収集した語句は、もともと「日本独特のもの (と知っている語句)」を想定して集めた。日本に関する記事の場合、日本語の Wikipedia の

\*5 [http://ja.wikipedia.org/wiki/Wikipedia:スタイルマニュアル\\_\(レイアウト\)](http://ja.wikipedia.org/wiki/Wikipedia:スタイルマニュアル_(レイアウト))

\*6 言語グリッド [16] を介して J-Server を利用した。

\*7 多くの場合は、日本語の説明が中国語の説明に比べて少ないため、追加した。

表 5 各言語の説明文の分量

Table 5 Number of explanation characters in each language.

	日本語の 説明文	中国語の説明文 (機械翻訳)
文字数の平均(文字)	276	335
文字数の標準偏差(文字)	174	228
最長文字数(文字)	898	1388
最短文字数(文字)	28	52

説明の導入部は、比較的簡潔に記載されており、解説本文が充実している記事が多い。逆に、中国語の Wikipedia において、日本に関する記事の場合は、解説本文そのものしかない場合が多いため、分量が多くなる傾向があった。

### 5.3 比較用の画像の収集

提案手法との比較で用いる検証用の画像は、Google 画像検索の結果を用いている。比較用の語句を表す画像は、5.1 節で収集した、文化差判定済みの 114 語を対象として、次の手順で収集した。

Step 1 該当語句を Google 画像検索<sup>\*8</sup>を用いて検索し、画像を 10 件取得する<sup>\*9</sup>。中国語の検索は、検索言語を中国語(簡体字)にして、同様に Google 画像検索<sup>\*10</sup>を用いて検索し、画像を 10 件取得する。日本語の語句に対応する中国語は、予め Wikipedia の多言語リンクを利用して抽出する。

Step 2 すべての画像を確認し、不適切な画像<sup>\*11</sup>を削除する。また、該当語句の意図する内容以外の意味を表す画像については除外した。

Step 3 該当語句を表す画像を 3 件選択する。選択の基準は、類似の画像を選択するのではなく、可能な限り多様性のある画像を選択する<sup>\*12</sup>。該当語句を表す画像が、画像検索結果に含まれていない場合には、検索範囲を広げて該当語句を表す画像を選択する。日本語の語句を表す画像と中国語の語句を表す画像が同じにならないように、異なる画像を選択する。

## 6. 実験方法

5 章で述べた検証用のデータセットを用いて、文化差事例の収集実験を行った。本実験では、Web 検索エンジンの上位 20 件を利用して文化差事例を収集した<sup>\*13</sup>。

提案手法により、114 語中 111 語<sup>\*14</sup>に関して、731 件の文化差事例を収集した。1 語句あたり平均 6.4 件の文化差事例となった。表 6 に、収集された文化差事例の例を示す。表 6 の例では、「日本では」「中国では」「インターネット」が含まれた文が抽出されている。

<sup>\*8</sup> <http://www.google.co.jp/imghp?hl=ja>

<sup>\*9</sup> リンク先に画像がない場合に取得できない場合もあった。

<sup>\*10</sup> <http://www.google.co.jp/imghp?hl=zh-cn>

<sup>\*11</sup> 公序良俗に反する画像や刺激の強い画像。

<sup>\*12</sup> 多様性がない場合には類似の画像とする。

<sup>\*13</sup> 2013 年 1 月 8 日に収集を行った。

<sup>\*14</sup> 「友情」「浴衣」「落語」に関する文化差事例は 0 件であった。

表 6 収集された文化差事例の例(「インターネット」)

Table 6 An example of a collected cultural difference case ("Internet")

日本ではブロードバンドが普及し快適なインターネット接続環境が整備されました。中国では企業にも ADSL が普及しておりますが、固定 IP アドレスは日本に比べ高価です。中国では固定 IP アドレスが高価な地域もありますのでコスト削減になります。中国のインターネットは、キャリアの事情によりある特定の経路のみ遮断されることがあります。

収集された文化差事例の適切さ(文化差事例の提示により、文化差の有無を適切に判定可能か)を評価するために、本学のシステム工学部および大学院システム工学研究科の学生 10 名に評価を依頼した。各語句に対する文化差事例を見て、各語句の文化差の有無を次の 5 種類で評価してもらった。

- 文化差があると思う
- 少し文化差があると思う
- 文化差はないと思う
- 全て文化差とは無関係の内容だと思う
- わからない

## 7. 実験結果と考察

### 7.1 文化差事例の正解率について(文化差あり/文化差なし)

表 7 に、正解データの「文化差あり」「文化差なし」の件数、収集された文化差事例の件数、文化差事例の正解率を示す。なお、上述した 5 種類への評価結果については、「文化差があると思う」「少し文化差があると思う」と判定された文化差事例を「文化差あり」、それ以外を「文化差なし」とした。

「正解率(個別)」は、収集された個別の文化差事例による平均正解率である。正解率は、「文化差あり」「文化差なし」で、それぞれ 0.52, 0.45 であり、収集された文化差事例の約半分に、適切に文化差判定可能な内容が含まれていることがわかった。

「正解率(各語句)」は、各語句の文化差事例(複数)に対する評価結果を統合し、正解かどうか判定した場合の平均正解率である。各語句の個別の文化差事例の評価結果の中に、一つ以上「文化差あり」と判定された文化差事例がある場合に、「文化差あり」と判定した。正解率は、「文化差あり」「文化差なし」で、それぞれ 0.81, 0.23 であった。「文化差あり」の文化差事例の多くを、適切に「文化差あり」と評価できることがわかった。なお、「文化差なし」の語句については、正解率が 0.23 であり、適切な評価ができていない。本研究では「文化差なし」を誤って「文化差あり」と判定する場合には、大きな問題がないと考えているが、「文化差なし」の精度は十分にでないため、検討が必要であると考えている。

本稿で提案した文化差事例の収集手法は、個別の文化

表 7 文化差事例の収集結果 (文化差あり / 文化差なし)

Table 7 Collection results of cultural difference cases (with a cultural difference/no cultural difference)

	正解データ (件)	収集された文化差事例 (件)	正解率 (個別)	正解率 (各語句)
文化差あり	83	547	0.52	0.81
文化差なし	31	184	0.45	0.23
合計	114	731	0.50	0.65

表中の正解率は、評価者 10 名の平均値である。

差事例では、約半分に文化差判定可能な内容が含まれる。特に、複数の文化差事例 (今回は最大 20 件) をまとめた場合には、約 8 割の語句は適切に文化差判定が可能であり、本手法により、適切な文化差事例を十分に収集可能であると考えられる。

### 7.2 文化差事例の正解率について (第 1 種 / 第 2 種 / 文化差なし)

表 8 に、正解データの「第 1 種」「第 2 種」「文化差なし」の件数、収集された文化差事例の件数、文化差事例の正解率を示す。評価結果の「文化差があると思う」「少し文化差があると思う」と判定された文化差事例を、「第 1 種」「第 2 種」として正解と判定している。また、表 8 の「第 1 種」「第 2 種」は、表 7 の「文化差あり」の内訳となっている。「文化差なし」は、同一である。

「正解率 (個別)」は、収集された個別の文化差事例による平均正解率である。正解率は、「第 1 種の文化差」「第 2 種の文化差」「文化差なし」で、それぞれ 0.50, 0.54, 0.45 であり、表 7 と同様に、約半分の語句の文化差に、適切に文化差判定可能な内容が含まれていることがわかった。

「正解率 (各語句)」は、各語句の文化差事例 (複数) に対する評価結果を統合し、正解かどうか判定した場合の平均正解率である。各語句の個別の文化差事例の評価結果の中に、一つ以上「文化差あり」と判定された文化差事例がある場合に、「第 1 種の文化差」「第 2 種の文化差」が正しく判定されたこととした。正解率は、「第 1 種の文化差」「第 2 種の文化差」「文化差なし」で、それぞれ 0.70, 0.93, 0.23 であった。特に、「第 2 種の文化差」の正解率は高く、適切に「文化差あり」と評価できることがわかった。

### 7.3 説明文および画像との比較

表 9 に、説明文、画像および事例 (提案手法) の正解率を示す。なお、説明文および画像は、文献 [4] における実験結果である。

表 9 の「文化差あり」の正解率をみると、説明文、画像はそれぞれ、0.18, 0.17 であるが、事例 (提案手法)

表 8 文化差事例の収集結果 (第 1 種 / 第 2 種 / 文化差なし)

Table 8 Collection results of cultural difference cases (with the first kind/the second kind/no cultural difference)

	正解データ (件)	収集された文化差事例 (件)	正解率 (個別)	正解率 (各語句)
第 1 種の文化差	42	207	0.50	0.70
第 2 種の文化差	41	340	0.54	0.93
文化差なし	31	184	0.45	0.23
合計	114	731	0.50	0.65

表中の正解率は、評価者 10 名の平均値である。

表 9 文化差の説明文、画像、事例との正解率の比較

Table 9 Comparison with accuracy of 'explanatory notes', 'images' and 'cases.'

	正解データ (件)	正解率		
		説明文	画像	事例 (提案手法)
文化差あり	83	0.18	0.17	0.81
文化差なし	31	0.77	0.87	0.23

表中の「事例」の件数は、評価者 10 名の平均値である。

は 0.81 と高い。表 9 の「文化差なし」の正解率をみると、説明文、画像はそれぞれ、0.77, 0.87 であるが、事例 (提案手法) は 0.23 と低い。

各正解率から、説明文と画像を見た場合、「文化差なし」と判断することが多いが、事例を見た場合には、「文化差あり」と判定することが多いことがわかる。

### 7.4 文化差の判定のしやすさ

表 10 に、文化差の判定のしやすさに関するアンケート結果を示す。アンケートは、各実験終了後に回答して評価者に回答してもらった。評価者数はそれぞれ 10 名である。説明文、画像、事例 (提案手法) の間で、Kruskal-Wallis の検定を用いた結果、 $p = 0.012^{*15}$  となり、有意差が見られた。多重比較の結果、説明文と事例 (提案手法) の間で、 $p = 0.015$  となり、有意差が見られた。事例 (提案手法) は、従来の説明文に比べて文化差の判定が容易であることが分かった。

表 11 に、文化差の評価実験における評価者の「文化差の判断しやすさ」の理由を、表 12 に、文化差の評価実験において評価者が気がついたことに記述された内容を示す。

文化差に関する記述が適切に含まれている事例に関しては、判断がしやすかったことがわかる (表 11(A), (B), (C), (G), 表 12(G))。しかし、今回、Web 上から文化差事例に該当する部分を自動的に切り出したため、不完全な文章が含まれており、特に、文章の読みづらさが評

\*15 有意水準は、 $p < 0.05$  としている

表 10 文化差の判定のしやすさに関するアンケート結果 (5段階評価)  
Table 10 Result of questionnaire survey on the ease of judgment of a cultural difference (5-point Likert scale).

質問項目	回答グループ	評価値(人)					中央値	最頻値	有意確率
		1	2	3	4	5			
文化差の判定はしやすかった.	説明文	2	8	0	0	0	2	2	0.012
	画像	0	5	2	3	0	2.5	2	
	事例(提案手法)	0	4	4	2	0	3	2, 3	

・5段階評価の評価値: 1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する  
・有意差の検定には, Kruskal-Wallis の検定を用いた.

価を下げていることがわかる(表 11(D), (F), (H), (I)). 文化差に関する記述が適切に含まれていない事例も含まれており, さらに, 文化差の抽出方法の精度の向上が必要であると考えられる(表 11(E), (I), (J)).

今回の評価実験では, 「文化差なし」の正解率が低い(表 7). 表 12 から「文章次第で評価結果が変わる」(表 12(E))の記述があり, 評価者は, 事前に文化差があると感じていなかった内容についても, 提示された文化差事例によって, 「文化差あり」と判断している. また, 「日本と中国の文化差は至る所に転がっているんだと感じた」(表 12(C))との記述からも, 少しの文化差の違いを「文化差あり」と評価した可能性もある. そもそも今回, 文化差判定に用いている「事例」は, 「違いについて言及した文」を収集したものである. そのため, これまで「文化差なし」に分類されていた語句も, 「文化差あり」に判定されてしまった可能性がある.

### 7.5 今後の方針

今回の実験では, 適切な文化差事例が収集されている例もあったが, 不適切なものも多く含まれていることが分かった. 表 7 から, 語句に関する記事全体を見た場合, 正解率は比較的高いが(表 7「正解率(各語句)」), 「正解率(個別)」があまり高くないことから, 個別の記事に関する精度が不十分であることがわかる. 上記については, 評価者の文化差の判断しやすさの理由(表 11)および今回の実験に関するコメント(表 12)からも分かり, 個人の主観が強すぎる点やジョークサイトの結果が表示されている点が, 評価を下げる原因の一つであることが分かった(表 12(F)). なお, 「文化差事例」そのものは, 文化差判定において, 語句の「説明文」「画像」を提示するよりも効果が高いと考えている.

今後, 今回適切な「文化差事例」と判断された Web ページの傾向(質問サイト, ブログサイト等)について調査し, どのような傾向を持つ Web ページから, 適切な文化差事例の収集が可能かについて検討を行う. 今回提案した「文化差事例」を提示する手法は, 「文化差あり」に判定されてしまう傾向があるため, 適切に「文化

表 11 文化差の評価実験における評価者の「文化差の判断しやすさ」の理由

Table 11 Estimators' judgment reasons in the experiment the ease of of judgment cultural difference.

- (A) 普段からニュースなどを見たり, 中国人の方々と接していても, 日本と中国の文化差を感じていたので文化差があるという判断はしやすかったです(評価 4).
- (B) 「文化差とは無関係の内容だと思う」\*16があったので, 悩むことが少なかったと思う. また, 精読するよりも, ブログの記事を見ている感覚で判断していた.(評価 4).
- (C) 判断しやすいものもあったが, それと同じくらい判断しづらいものがあったから(評価 3).
- (D) ブログやウェブページの一部分のみが抜き出されたものなどが時々あり, 文章の途中で終わっているものがあったので判断に困りました(評価 3).
- (E) 日本では, 中国ではという対比構造で書かれていた文章が少なく感じたから(評価 3).
- (F) 文章が読みづらく, 判断しにくいものが多かったから. 読みながら語句に関係ないところで文化差を感じ, 頭が混乱しがちだったから. ほんの少し足りないといった文章が多かったから(評価 3).
- (G) 判断しやすいものと判断しにくいものがあり, 判断しにくいものの方が多いと思ったため. 具体的にどのように違うなど書かれていればわかりやすいが, 書かれていないものが多く難しかった. また, 個人の経験や主観で述べているものも多く, 「中国のものの方が美味しい」など文化差として判断していいかわかりにくかった. 文章が整っていないと読みにくいものもあった(評価 2).
- (H) 文章の調子が変わったりするため, 混乱したりして判断しづらかった. 一つの語句に対して, 中国と日本のそれぞれの特徴を別の観点から説明しているものがあり, 比較しづらいことが多々あった(評価 2).
- (I) 日本語がおかしいもの, 文章が途切れているもの, 改行がなく文章として読みにくいものが多々あったため. また, 片方の国における記述がないものが多かったため(評価 2).
- (J) 文章が乱雑であったり, 文化差はあるが対象の単語でない場合があった(評価 2).

記述は, 原文のままである.

各文の最後の「評価」の値は, その記述を行った評価者の「文化差の判定のしやすさ」の評価値(表 10)である.

差なし」と判断できる事例について分析し, 理解支援のための手法を再検討する必要がある.

\*16 今回の評価実験で用いた評価項目「全て文化差とは無関係の内容だと思う」のことである.

表 12 文化差の評価実験において気がついたこと

Table 12 Estimators' comments in the experiment of judgment cultural difference.

<p>(A) 単語について、「教えようとしている文章」「個人の意見を述べる文章」「他人と意見をかわす文章」にわかれていると思う。「教えようとしている文章」は比較的読みやすく、文化差がわかりやすかった。「個人の意見を述べる文章」はブログなどのように、自分の経験で書いているので、そのものの文化差の一部のみを表していることが多かった。「他人と意見をかわす文章」は文化差を探るのに適切でないと思った。「 について教えてください」に対する意見も、正しいのかわからない情報に見えたので、文化差を判断するのに用いていいの疑問に思った。文章が読みにくいものが多くて精神的につらかった。</p> <p>(B) 掲示板のレスのような文章もあり、文化差がわかりやすいのとわからないのが極端に感じた。</p> <p>(C) 日本と中国の文化差は至る所に転がっているんだと感じた。文化の違いが人の考え方、接し方などの違いも引き起こしているのだと思います。</p> <p>(D) 広告みたいな文章がそのまま反映されてたり、重複した文章があったのが非常に見づらい要因になっていた。中国から日本に伝わったものや、日本から中国に伝わったものは、それぞれの国の文化に合わせて変化させているものが多いと感じた。</p> <p>(E) 事前知識を使うと評価が変わりそう。今回、日本文化に関しては事前知識無しとして評価した、出てきた文章だけで判断すると、文章次第で評価結果が変わるのがふしぎ。ブログっぽい文章が多かったのが気になった。カルチャーショックを記事にしていることが多いのだろうか。</p> <p>(F) アンサイクロペディアのようなジョークサイトと思われるサイトからの引用があった。「中国ではない」といったヒットの仕方をしている場合があった。個人の思想あるいは政治的思想が強いといった内容が偏っている場合があった。</p> <p>(G) 出てくる文章は、うまく文化差を表記しているものもあったが、その単語に関する別の語句(夫だったら、夫婦とか結婚とか)についての文化差を表示しているものがあったように思う。</p> <p>(H) 概念的な文化差は文章であってもわかりづらいんだなあ、と思いました。</p> <p>(I) 日本由来のものに関しての文化差はあまり無いように感じましたが、中国発祥で日本に伝わったものは、文化差があるように思います。天津飯や中華まん等。</p> <p>(J) 趣旨をわかっている、文章中の文化差と語句の文化差がまざり、判断するのに時間がかかった印象があります。文章自体の信ぴょう性が低い印象を受け、それが文化差の判断に影響した気がします。</p>
---

記述は、原文のままである。

## 8. おわりに

本稿では、Web上に存在する「文化差の理解を促す事例」を文化差事例として収集する手法を提案した。評価の結果、複数の文化差事例をまとめた場合には、文化差のある語句の約8割については適切に文化差を判定可能であり、本手法により、適切な文化差事例を収集できると考えられる。今後は、文化差事例の収集精度の向上について検討する。

## 謝辞

本研究の一部は、独立行政法人科学技術振興機構研究成果最適展開支援事業(A-STEP)探索タイプ「検索エンジンと機械翻訳を用いた多言語用語間における文化差検出サービス」および日本学術振興会科学研究費基盤研究(B)(22300044)の補助を受けた。

## 参考文献

- [1] 藤井薫和,重信智宏,吉野 孝:機械翻訳を用いた異文化間チャットコミュニケーションにおけるアノテーションの評価,情報処理学会論文誌,Vol.48, No.1, pp.63-71 (2007).
- [2] 吉野 孝,宮部真衣,諏訪智大:Wikipediaを用いた文化差可視化サービスの開発,情報処理学会,マルチメディア,分散,協調とモバイル(DICOMO2011)シンポジウム(2012).
- [3] 吉野 孝,宮部 真衣:日本語版・中国語版 Wikipediaにおける相互言及の有無を用いた文化差検出方法の提案,第5回 Web とデータベースに関するフォーラム(WebDBフォーラム2012), B2-2, pp.1-8 (2012).
- [4] 吉野 孝,宮部真衣:文化差理解支援における可視化効果の検証,情報処理学会研究報告,グループウェアとネットワークサービス研究会,Vol.2013-GN-86, No.36, pp.1-8 (2013).
- [5] 諏訪 智大,宮部 真衣,吉野 孝:Web ページにおける文化差可視化システムの開発,平成24年度情報処理学会関西支部大会, F-01, pp.1-4 (2012).
- [6] 久米昭元,長谷川典子:ケースで学ぶ異文化コミュニケーション—誤解・失敗・すれ違い,有斐閣(2007).
- [7] Cho Heeryon,石田 亨,山下直美,稲葉利江子,高崎俊之,神田智子:絵文字解釈における人間の文化差判定,ヒューマンインタフェース学会論文誌,Vol. 10, No. 4, pp. 427-434 (2008).
- [8] Tomoko Koda and Toru Ishida: Cross-cultural study of avatar expression interpretations, SAINT 2006, pp. 130-136 (2006).
- [9] 藤原裕也,灘本明代:Wikipediaの言語間比較による差異情報抽出手法の提案,情報処理学会研究報告,Vol. 2011-DBS-152, No. 3, pp. 1-8 (2011).
- [10] 松浦愛美,江口浩二:時系列対訳トピックモデルを用いた言語横断トレンド分析,情報処理学会研究報告,Vol. 2010-DD-75, No. 11, pp. 1-5 (2010).
- [11] 吉岡真治:Wikipediaを用いた中日カタカナ翻訳辞書の作成と言語グリッドへの応用,電子情報通信学会技術報告,人工知能と知識処理,Vol. 109, No. 424, pp. 43-46 (2010).
- [12] 西田ひろ子:異文化間コミュニケーション,創元社(2000).
- [13] 古田 暁,石井 敏,岡部朗一,平井一弘,久米昭元:異文化コミュニケーションキーワード,有斐閣(1990).
- [14] 多文化関係学会編:多文化社会日本の課題,明石書店(2011).
- [15] 吉野 孝,宮部真衣:日本語版 Wikipediaにおける執筆者の意図に着目した日中間の文化差検出手法の検討,電子情報通信学会技術報告,言語理解とコミュニケーション(第3回集合知シンポジウム),Vol.111, No.427, NLC2012-58, pp. 13-18 (2012).
- [16] Toru Ishida. Language Grid: An Infrastructure for Intercultural Collaboration. IEEE/IPSJ Symposium on Applications and the Internet (SAINT 2006), pp.96-100, keynote address (2006).