

絵文字コミュニケーションにおけるセマンティクス

CHO Heeryon^{†a)} 稲葉利江子^{††} 石田 亨[†] 高崎 俊之^{†††}
森 由美子^{†††}

Semantics in Pictogram Communication

Heeryon CHO^{†a)}, Rieko INABA^{††}, Toru ISHIDA[†], Toshiyuki TAKASAKI^{†††}, and
Yumiko MORI^{†††}

Abstract. 絵文字メッセージをやり取りして人間同士が意思疎通するには、コミュニケーションに参加している人間たちが同じ絵文字解釈（意味理解）を持たなければならない。絵文字コミュニケーションシステムの潜在利用者である日本と米国の子供たちが、同じ絵文字解釈を持つかを調べるために、システムで利用される 120 個の絵文字の意味を問うアンケート調査を行った。結果、ジェスチャー、色と性別、時間、空間、馴染みのもの、顔／表情の六つの領域に関係する 19 個の絵文字に対して、両国の子供が異なる解釈を持つことが分った。ここでは、多様な解釈を持つ絵文字を、絵文字コミュニケーションシステムでうまく利用するための方法として、絵文字の解釈語リストと頻度情報を利用し、たくさんの絵文字の中から検索語に関連する絵文字を検索する方法を提案する。提案手法は、絵文字の意味的なあいまい性を考慮した検索を実現し、文化的な解釈の差を扱う検索システムの土台となる。

Keywords. 絵文字、セマンティクス、異文化間コミュニケーション

1. ま え が き

国境を越えた交流は、今や政府や企業だけでなく、民間レベルでも盛んに行われている。その一例として、世界の子供たちの、絵文字を用いたコミュニケーションがある [1]。

絵文字コミュニケーションでは、二人の人間がそれぞれ一つ以上の絵文字を組み合わせてメッセージを作成し、そのメッセージをやり取りすることで意思伝達を図る [2]。このときメッセージを構成する絵文字は、言語非依存な手書き画像であるため、言葉の違う人間同士でもやり取りできる利点を持つ。しかし、人間の絵文字解釈（絵文字の意味理解）は、万人において同じわけではないので、メッセージをやり取りしている人間同士が、共通の絵文字解釈を持たない場合、絵文

字メッセージの理解に食い違いが生じる。

異なる文化や言語を背景に持つ人間同士が、システムを介して通信するとき、システムは、言語や文化の違いに起因するメッセージ理解のズレ（誤解）を発見し、それをシステム利用者に通知し、修正することが望まれる。このような誤解の自動検出は、文化や言語の異なるシステム利用者たちのコミュニケーションを促進し、相互理解を構築する上で重要である。

異文化間コミュニケーションを支援するシステムの構築に向けて、これまで様々な研究が行われてきた。例えば [3] は、異言語話者たちが機械翻訳を介して意思疎通する際、相手にわかりやすい翻訳文を作成できるようにするために、機械翻訳システムが自ら翻訳品質を推定できるよう、翻訳品質と強い相関を持つ類似度の計算手法を考案した。[4] は、多言語 BBS のログを解析し、メッセージの返信構造を手がかりとした形式的なスレッド (syntactic thread) と、メッセージ内で用いられている語の相関に基づいた意味的なスレッド (semantic thread) が乖離している場合、異言語話者間で誤解が生じやすいことを解明した。また [5] は、大規模なオンライン調査を通して、アバタ表情解釈の文化差を明らかにした。

[†] 京都大学大学院情報学研究科社会情報学専攻, 〒 606-8501 京都市左京区吉田本町

^{††} 独立行政法人情報通信研究機構 (NICT) 知識創成コミュニケーション研究センター言語グリッドプロジェクト, 〒 619-0289 京都府相楽郡精華町光台 3-5

^{†††} 特定非営利活動法人バンゲア 京都 R&D センター, 〒 600-8411 京都市下京区水銀屋町 620 COCON 烏丸 4F

a) E-mail: cho@ai.soc.i.kyoto-u.ac.jp

(注 0): 本稿の著作権は (社) 情報処理学会にある。

[3]、[4]、[5]の研究で繰り返し取り上げられている問題は、異文化間コミュニケーションでやり取りされる(メッセージの)「意味(セマンティクス)の違い」である。[3]は折り返し翻訳前後の入出力文の意味の違いを、[4]は異言語話者たちのメッセージ理解のズレを、[5]は多国籍回答者たちのアバタ表情解釈の差を、扱っている。

本稿では、絵文字を用いた異文化間コミュニケーションに焦点を当て、絵文字コミュニケーションシステムの潜在利用者である日本と米国の子供たちの、絵文字理解について考察する。そのために、日米子供たちにオンライン絵文字アンケートを実施し、絵文字コミュニケーションシステムで用いられる絵文字120個の、日米二カ国の絵文字解釈を比較分析する。

また、絵文字コミュニケーションを支援する一方法として、絵文字メッセージ作成に関連する、たくさんの絵文字の中から必要とする絵文字を検索する問題に取り組む。この問題は、国家間の絵文字解釈差を扱う検索を実現する前に必要となる、絵文字解釈の意味的なあいまい性を考慮した検索を扱っている。提案する意味的なあいまい性を考慮した検索手法は、絵文字アンケート調査から得た解釈語とその頻度情報を適合率の計算に利用する。

以降、2章では絵文字アンケート調査の概要を述べ、3章では日米子供たちの絵文字解釈の差について詳しく報告する。そして、4章で、意味的なあいまい性を考慮した絵文字検索システムを提案し、5章で展望を述べ、本稿をまとめる。

2. 絵文字アンケート調査

日本と米国の子供たちを対象に、2005年10月1日から2006年8月10日まで、オンライン絵文字アンケート調査を実施した^(注1)。

2.1 目的

絵文字アンケート調査は、日米の子供たちが、絵文字をどのように理解しているかを調べるために行う。具体的には、日米の子供たちが異なる解釈を持つ絵文字を発見し、その内容を比較することを目的とする。

2.2 調査方法

調査対象の絵文字は全部で120個あり、10個が1セットになっている(全12セット)。図1に、日本語アンケート調査ページの例を示す。アンケート回答者

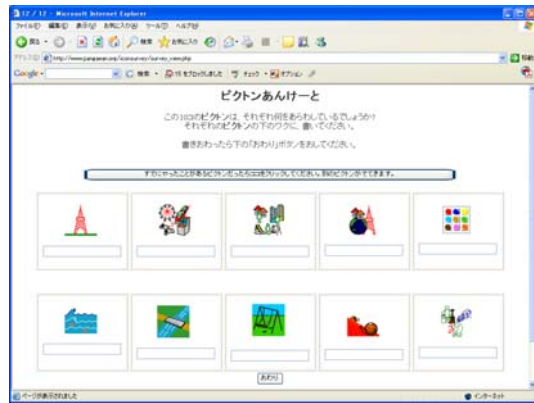


図1 絵文字アンケート調査ページのスクリーンショット
Fig. 1 A screenshot of online pictogram survey page

はランダムに12セットの絵文字の中から1セットを選び、一画面10個の絵文字を見て、それぞれが何を表すかを自由記述形式で、絵文字の下のテキストボックスに記入する。10個の絵文字の意味を回答し終わったら、別の絵文字セットに移動できる。同じ回答者は、1セットから12セットまで自由に回答できる。

2.3 データ

日本語と英語アンケート調査ページで収集された回答データはそれぞれ1,107と1,047回答セットである(1回答セットに12個の絵文字に対する回答が記述されている)。回答者が絵文字を見て自由に記述した解釈語を、絵文字ごとにまとめ、解釈語のリストを生成した。次に、解釈語のリストから固有語を取り出し、固有語が出現する頻度を集計した。図2のような絵文字から得られた米日解釈語の、実際の集計結果を表1に示す。

表1を見ると、米国と日本の子供の多くが、図2の絵文字を見て「shopping, 買い物」と記述したことが分る。一番上の「shopping」の右に記されている頻度「23」は、図2の絵文字を見て「shopping」と答えた米国外児が23人いることを示す。調査対象となった120個の絵文字の解釈語の頻度合計は、68~108(個々の絵文字の総回答者数が68人~108人)となった。

3. 絵文字解釈の日米差

絵文字120個の解釈語リストを集計し、その結果を二人の人間が分析し、日本と米国の子供の間で、絵文字の解釈に差があるかどうかを調べた。結果、19個の絵文字に、日米解釈差があることが分った。解釈の差

(注1): URL は、<http://www.pangaean.org/iconsurvey/>



図 2 アンケート調査に使われた絵文字の一つ
Fig. 2 One of the 120 pictograms used in the online pictogram survey

表 1 図 2 の絵文字の米日解釈語の集計結果
Table 1 U.S.-Japan interpretation words and frequency for Fig.2

解釈語 (米国)	頻度	解釈語 (日本)	頻度
shopping	23	買い物	40
grocery shopping	17	かいもの	6
groceries	12	お買い物	4
grocery store	3	おかいもの	3
shopping cart	3	買物	3
gorceys	2	カート	2
healthy	2	スーパーで買い物	2
shop	2	う	1
shopping for food	2	お買い物 (食べ物)	1
buting groceries	1	かいものかご	1
curt	1	かご	1
filling	1	ぐちゃぐちゃ	1
food	1	しょくひんカート	1
food shopping	1	はげしい	1
full shopping cart	1	へるしー	1
good	1	やさい	1
graocery shopping	1	ショッピング	1
store	1	スーパー	1
vegetables	1	スーパーに買い物	1
		スーパーのかご	1
		スーパーマーケット	1
		食事の買い物	1
		買い物かご	1
		買い物に行く	1
		買い物籠	1
		買う	1
		野菜を買う	1
合計	76	合計	80

は、(1) ジェスチャー、(2) 色と性別、(3) 時間、(4) 空間、(5) 馴染みのもの、(6) 顔/表情の、六つの領域で発見された。それぞれの詳細を、実際の絵文字とともに見る。

3.1 ジェスチャーの文化差

頭の上で両手の指先をあわせ、大きく輪を描く図 3 の左の絵文字を、日本の子供たちは「まる、オッケー、正解、いいよ、あたり」と解釈したが、米国の子供たちは「exercise, jump rope, stretch, dance, aerobics」といった運動に関連するしぐさと解釈した。回答者(日:95人,米:67人。以降,国:回答者の総数)のほとんどがこの傾向を示した。



図 3 ジェスチャーの解釈差が顕著に現れた絵文字
Fig. 3 U.S.-Japan interpretation of gesture was different in these pictograms



図 4 「赤は女、青は男」という色と性別の相関は、万国共通でない(左は赤、右は青)

Fig. 4 Correlation between color and gender - red denotes female and blue denotes male - is not universal (Left human figure is red, and right human figure is blue)

同様に、両腕を前で交差している図 3 の真ん中の絵文字を、日本の子供たちは「だめ、ばつ、はずれ、まちがっている、いいえ」と解釈したが、米国の子供たちは「angry, mad, frustrated, selfish」といった否定的な感情表現と解釈した。米国の子供たちはおそらく、絵文字に描かれた表情に注目したり、交差した腕を、腕組みと解釈したと思われる。回答者(日:95人,米:68人)のほとんどが似た回答を返した。

最後に、図 3 の右の絵文字では、両国回答者(日:72人,米:91人)の多くが「話す(22人,30.6%)、talking(26人,28.6%)」という共通の解釈を出したが(おそらく人の右に描かれている風船に注目したものである)、一方で、日本の子供は「ありがとう、お願い(10人,13.9%)」という回答を、米国の子供は「praying(20人,22%)」という回答を出した。これは、手を胸の前であわせるジェスチャーの読み違いに起因するものと思われる。

3.2 色と性別の相関

「赤は女を表し、青は男を表す」という、色と性別に関する常識(もしくは、固定観念)は、必ずしも万国共通でないことが分った。回答者(図 4 の左右の絵文字ともに同じ総回答者数:日本:96人,米国:93人)のうち、日本の子供の 99%(一人を除く)は、赤い線で描かれた図 4 の左の絵文字を「女性、女の人、女、おかあさん、おねえさん」といった女性を表す人間として理解した。それに対し、米国の子供は、全体の 28



図 5 時間に対する解釈が異なった絵文字．それぞれ左から，米国の回答は「late, on time, early」，日本の回答は「未来，現在，過去」．

Fig.5 Pictograms with different interpretation on time. U.S. respondents answered, “late, on time, early” whereas Japanese respondents answered, “future, present, past.” (each from left)

% (26人)が，赤い線の人を「man, dad, boy, male」と理解した．残りの67人は，日本の子供同様，赤い人を女性と理解した．しかし，米国外の子供の一部にとっては，「赤い男性」は有り得る．

同様に，青い線で描かれた図4の右の絵文字も，日本の子供にとっては，99%が「男性，男の人，男，おとうさん，兄貴」を意味するが，米国外の子供の一部(18人，19%)にとっては「woman, mom, big girl, female, old women」を意味する．色と性別の認識が，米国外の子供にとっては絶対的でないことが分かった．

3.3 時間概念

絵文字から読み取る時間の解釈も違うことが分かった．図5の左の，三時を指している時計が，三時十分になる絵文字を見て，両国回答者(日：101人，米：104人)の一部は「10分後，10 minutes later」という共通解釈を出したが，各国において最も多かった回答は，日本の場合「未来(15人，14.3%)」，米国の場合「late(14人，13.5%)」だった．ちなみに，日本側の回答で「late」に近い回答は「遅れる(1人)，遅刻(1人)」，米国側の回答のうち「未来」に近い回答は「future tense(1人)」だった．

図5の真ん中の絵文字の場合，両国回答者(日：95人，米：105人)の最も多かった回答は，日本が「現在，今(25人，26%)」，米国の「on time(20人，19%)」だった．日本の回答で「on time」に近い回答は「ちょうど(1人)，オンタイム(今)(1人)」，米国の回答で「現在，今」に近い回答は「now(4人)，present(1人)，present tense(1人)」だった．

最後に，図5の右の絵文字の場合，両国回答者(日：99人，米：104人)の最も多かった回答は，日本が「過去(26人，26%)」，米国の「early(14人，13.5%)」だった．日本の回答で「early」に近い回答は「早く着いちゃった(1人)」，米国の回答で「過去」に近



図 6 空間に関連する解釈が異なった絵文字．それぞれ左から，米国の回答は「up, down」，日本の回答は「あそこ，ここ」．

Fig.6 Pictograms with different interpretation on space. U.S. respondents answered, “up, down” whereas Japanese respondents answered, “there, here.” (each from left)

い回答は「past(4人)，past tense(1人)」だった．一方，両国の子供の共通した解釈は「10分前，10 min before」だった．

まとめると，時計が描かれている絵文字を見て，日本の子供は「未来，現在，過去」といった，時間の前後関係を表す概念を思い浮かべるのに対し，米国外の子供は「late, on time, early」といった，約束に関係する時間概念を思い浮かべることが分かった．

3.4 空間認識

空間的要素を含む絵文字の解釈にも違いがあることが分かった．図6の絵文字の，指の指す方向に対し(左右の絵文字ともに総回答者数は同じ．日：104人，米：92人)，米国外の子供は，主な解釈として，垂直軸に遠近感を伴う「up(21人)，up/above(2人)，top(1人)」(図6の左の絵文字．合計24人，26.1%)や「down(24人)，down/below(2人)，bottom(1人)」(図6の右．合計27人，29.3%)を出したのに対し，日本の子供は，水平軸に遠近感を伴う「あそこ(33人)，あれ(28人)，あっち(9人)」(図6の左．合計70人，67.3%)や「ここ(38人)，これ(26人)，こっち(6人)」(図6の右．合計70人，67.3%)という解釈を，主な回答として出した．

一方，米国外で「あそこ，あれ」という回答に近い回答は「there(14人)，that(1人)」(合計15人，16.3%)，「ここ，これ」に近い回答は「here(15人)，this(1人)」(合計16人，17.4%)だった．日本側で「up」に近い回答は「上(5人，5%)」，「down」に近い回答は「下(5人，5%)」だった．

3.5 馴染みのもの

絵文字から，普段見慣れているものを読み取る例も観察された．図7の三つの絵文字から，日米の回答者は違うものを読み取った．まず図7の，左の赤い塔を見て(日：85人，米：76人)，日本の子供は半数近くが



図7 「東京タワー」や「運動会」は、日本で馴染みのものである反面、「party」は米国で馴染みのものである
 Fig.7 “Tokyo tower (left)” and “school sports festival (middle)” are familiar concepts in Japan whereas “party (right)” is a familiar concept in the U.S.

(38人, 44.7%)「東京タワー」と答えたのに対し、米国の子供は半数近くが(34人, 44.7%)「eiffel tower」と答えた。一方、米国の子供で「tokyo tower」と答えた子供はいなかったが、(代わりに「paris/tokyo」と回答した子供が一人いた)、日本の子供には「エッフェル塔」と答えた子供が7人いた。

図7の真ん中の絵文字の場合、両国回答者(日:74人, 米:91人)はともに「一番(6人), first place(5人)」「優勝(5人), winning(5人)」といった、勝利を表す回答を共通解釈として返したが、日本の子供の場合、最も多かった回答は「運動会(25人, 33.8%)」だった。米国の子供の場合「winner(34人, 37.4%)」という回答が最も多かった。おそらく「運動会」は、日本特有の行事であり、米国には類似する行事がないと思われる。

最後に、複数の人が集まって話している図7の右の絵文字(日:69人, 米:91人)の場合、日本の子供は、黒い人に注目し「うそつき, うそをつく, かげぐち, 嘘」(19人, 27.5%)と回答したのに対し、米国の子供は「party(10人), friends(10人), laughing(2人), playing(2人)」といった、にぎやかな様子を表す解釈語を回答として返した。これは、米国では、パーティが日常的な行事として定着しているからだと思われる。

3.6 顔 / 表情

図8のような顔/表情に関する絵文字は、解釈が実に多様であった。上の段の左の絵文字から見てみよう。両国回答者(日:90人, 米:97人)のうち、日本の子供の半数近くが図8左上の絵文字を「寒い(44人, 48.9%)」と解釈したのに対し、米国の子供は「scared(19人, 19.6%), sad(11人, 11.3%), cold(10人, 10.3%)」と解釈した。日本の子供でも「怖い」という回答(19人)は多かったが、「悲しい」という回答は一人しかいなかった。



図8 顔や表情に関する解釈が日米で異なった絵文字
 Fig.8 Facial expressions were interpreted differently by U.S. and Japanese respondents

図8上の段の真ん中の絵文字(日:87人, 米:96人)の場合、日本の子供は「ばかにする, あざ笑う, 軽蔑, 嘲笑(14人, 16.1%)」という解釈が多かったが、米国の子供は、ずるいという意味を持つ「sneaky(14人, 14.6%), sly(9人, 9%)」という解釈が多かった。

図8上の段の右の絵文字(日:71人, 米:90人)の場合、日本の子供は「しらんぷり(20人, 28.2%)」という回答が多かったが、米国の子供は「whistling, whistle(28人, 31.1%)」という回答が多かった。

図8の下段に移ると、日本の子供たちは、外見を評価する解釈語、たとえば「かっこいい(28人, 30.8%, 左の絵文字), かわいい(38人, 41.8%, 真ん中), きれい(14人, 16.5%, 右)」などを主な解釈語として返しているが、米国の子供たちは、そのようなまとまった評価語を返さず「mean(6人, 6.3%, 左), happy(24人, 25.3%, 真ん中), happy(8人, 8.4%, 右)」といった内面の状態を表した解釈を回答として出している。

4. 解釈語を用いた絵文字検索システム

これまで日本と米国の絵文字解釈の違いについてみてきたが、実際のところ絵文字の解釈は同じ国の人間同士でも違ってくる。日本語のアンケート結果だけを見ても、唯一の解釈を持つ(回答者が全員同じ解釈語を記述した)絵文字は一つもなかった。ここでは、絵文字コミュニケーションを支援する方法として、たくさんの絵文字の中から検索語に適合する絵文字を検索する問題について考える。絵文字検索の支援は、絵文字メッセージ作成の支援につながる。

国家間の解釈差を扱う絵文字検索を実現するためには、まず、絵文字解釈のあいまい性を考慮した絵文字

検索を先に実現しなければならない。絵文字解釈の意味的なあいまい性は、今回取り上げる絵文字システムに内在する問題であり、国家間の解釈差を扱う絵文字検索を実現するための先決課題となる。そこでここでは、先決課題を解決するという意味で、絵文字解釈のあいまい性を考慮した検索方法を提案する。

4.1 定式化

人間が一つ以上の絵文字を組み合わせてメッセージを作成し、その絵文字メッセージを二人の人間がやり取りし、意思疎通する問題を考える。一つの絵文字は一つの画像を持ち、その画像は多様な解釈、すなわち、一つ以上の意味を持つ。多様な意味を持つ絵文字を用いて、二人の人間が絵文字コミュニケーションを行うには、メッセージに用いられている絵文字の画像と、その画像の持つ意味が、コミュニケーションに参加している人間ごとに、正しく対を成して (pairing) いなければならない。例を用いてこの「画像 - 意味」のペアリングについて詳しく見る。仮に「おいしいケーキを食べた」というメッセージを、二人の人間、AとBが、絵文字を用いて作成したとする。

図9に、AとBが作成した絵文字メッセージの例を示す。図9の、上の段には人間Aが作成した絵文字メッセージが、下の段には人間Bが作成した絵文字メッセージが表示されている。図9から分かるように、AとBが選んだ絵文字(画像)の組み合わせは、完全には一致していない。例えば「おいしい」と「食べる」の意味を持つ絵文字に、AとBは異なる絵文字(画像)を用いている。このような多様な画像選択の可能性は、特定の意味(セマンティクス)を持つ絵文字(例えば「おいしい」と解釈される絵文字)が複数あるために起きる。この、特定の意味を持つ画像が複数存在する問題を、ここでは、絵文字メッセージ作成(encoding)における「意味 - 画像」の多重ペアリング(meaning-image multiple pairing)と定義する。

図9で人間Aが作成した絵文字メッセージを、人間Cが受け取ったとき、人間Aが意図した意味通りに、人間Cが読み取るとは限らない。もしかすると人間Cは、人間Aの絵文字メッセージを見て「すっぱいケーキを食べた」と解釈(decode)するかも知れない。このようなメッセージ理解の食い違いは、個々の絵文字が複数の意味(多様なセマンティクス)を持つために起きる。この、特定の画像が複数の意味(セマンティクス)を持つ問題を、ここでは、絵文字メッセージ解読(decoding)における「画像 - 意味」の多重ペアリ

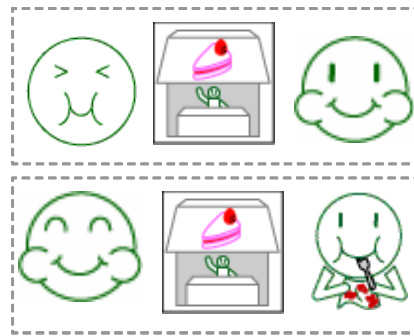


図9 「おいしいケーキを食べた」という意味を持つ二つの絵文字メッセージ(上と下)

Fig.9 Two pictogram messages meaning, "I ate delicious cake" (top and bottom)

ング(image-meaning multiple pairing)と定義する。

以上で絵文字コミュニケーションにおける問題を、(1) 絵文字メッセージ作成時の「意味 - 画像」多重ペアリング問題と、(2) 絵文字メッセージ解読時の「画像 - 意味」多重ペアリング問題として定義した。

ここからは、絵文字メッセージ作成時の「意味 - 画像」多重ペアリング問題に焦点を当て、特定の意味を持つ絵文字を、複数の絵文字の中から探しだす問題について考える。具体的には、解釈語が明示的に付与されている絵文字の中から、特定の意味を持つにふさわしい絵文字を探しだす問題を取り上げる。そのために、前述のアンケート調査から得た、解釈語データを利用する。

4.2 適合率の定義

図2と表1のように、ある絵文字(の画像)は、解釈語のリストと、解釈語ごとに、その解釈語を記入した回答者の数を頻度として持つ。ある絵文字の、ある解釈語の割合(その解釈語の頻度を、その絵文字のすべての解釈語の頻度合計で割った数値)は、その絵文字が、どれくらい多くの割合の人たちに、その解釈語で支持されているかを示す。つまり、解釈語の割合が高いほど、より多くの人がその絵文字を、その解釈語の意味として理解している、と捉えることができる。

任意の検索語が与えられたとき、個々の絵文字がその検索語にどれくらい適合するかを、適合率として定義する。適合率は、絵文字アンケート調査から取得した解釈語と頻度情報を用いて計算する。

絵文字 e の解釈語を w_1, w_2, \dots, w_n とし、その解釈語の出現率を $P(w_1|e), P(w_2|e), \dots, P(w_n|e)$ とする。図2のような絵文字の場合、解釈語「shopping」の出

現率を $P(\text{shopping}|e) = 23/76$ (表 1 参照) のように計算できる. w_i を検索語として用いたときに, 絵文字 e を提示することが妥当であることを示す最も簡単な指標は以下になる.

$$P(w_i|e) \tag{1}$$

このとき, 解釈語間の類似度は考慮されない. 例えば, 「うれしい」を検索語としたとき「うれしい」は含まないがそれと類似の意味を持つ「しあわせ」を解釈語として持つ絵文字は, 出現率だけを考慮すると検索結果からもれてしまう. これを解決するため, 解釈語の間の類似度, $similarity(w_i, w_j)$, を何らかの形で定義する. これを用いて, 絵文字 e を提示することが妥当であることを示す指標を以下で計算できる.

$$\sum_j P(w_j|e)similarity(w_i, w_j) \tag{2}$$

$similarity(w_i, w_j)$ の定義にはいくつか考えられる. w_i が解釈語となる絵文字の集合を E_i とすると, w_i と w_j の類似度は以下で定義することもできる. $|E_i \cap E_j|$ は, w_i と w_j 両方を解釈語として持つ絵文字の要素数を, $|E_i \cup E_j|$ は, w_i か w_j のどちらか一つでも解釈語として持つ絵文字の要素数である.

$$similarity(w_i, w_j) = |E_i \cap E_j| / |E_i \cup E_j| \tag{3}$$

以上を組み合わせると, w_i を検索語として用いたときに, 絵文字 e を提示することが妥当であることを示す指標, すなわち適合率は, (2) と (3) から以下のようになる.

$$\sum_j P(w_j|e)|E_i \cap E_j| / |E_i \cup E_j| \tag{4}$$

以上で, 意味的なあいまい性を扱った絵文字検索のための, 適合率の計算方法を述べた.

4.3 実装

解釈語リストと頻度情報を持つ 26 個の絵文字を用いて, 上記で提案した適合率の計算方法を利用する絵文字検索システムを実装した. 図 10 にシステムのスクリーンショットを示す. 例えば, システムに「夜」という検索語を入力すると, 左側に「夜」という解釈語を持つ絵文字のリストを, 真ん中に適合率を導入した絵文字のリストを, 右側に適合率にカットオフ値を適用した絵文字のリストを, それぞれ表形式で表示する.



図 10 解釈語を利用し, 適合率を計算する絵文字検索システムのスクリーンショット

Fig. 10 A screenshot of prototype pictogram retrieval system

絵文字	適合率	解釈語
	0.878030303030303	おやすみ(1) 夜(1B) 朝(1) 夜明け近く(1) にわとり(1)
	0.348214285714286	おやすみ(5) ぐーぐー(1) 就寝(1) 夜(1) ねむる(7) しいゆめみたい(1) すやすや(3) 夢の中(1) ばん(1) おねんね(1) ねている(5) ほしぞら(1)
	0.347701149425287	夜(1) 宇宙旅行(1) 夜遊び(1) 会話(3) こんばんは(23)
	0.276666666666667	ねごと(4) おやすみ(17) ねむる(6) 夢を見ている(1) 睡眠(1) 夢(1)
	0.051666666666667	おやすみ(2) お泊り(2) こんごは(2) おはよう(19) 朝(1) 会話(2) あいさつ(2)

図 11 「夜」を検索語とした図 10 の拡大画面
Fig. 11 A search result to a query “night”

図 11 は, 図 10 の画面を拡大した結果を表示している. 「夜」という解釈語を持たない絵文字でも, 意味的にふさわしいと考えられる場合, 関連する絵文字として返されることが分る(下の二つの絵文字). なお, 現在のところカットオフ値はヒューリスティックで設定している.

5. むすび

絵文字は多様な解釈を持つが, 言語非依存なため, 異文化間コミュニケーションの道具として利用できる可能性を持つ. 絵文字を用いて意思疎通するには, コミュニケーションに参加している人たちの絵文字解釈が一致しなければならない. そのためにはまず, どの

ような絵文字解釈が存在するのかを把握しなければならない。

本稿では、絵文字コミュニケーションシステムの潜在利用者である日本と米国の子供たちに、絵文字アンケート調査を行い、ジェスチャー、色と性別、時間、空間、馴染みのもの、顔/表情の六つの領域において絵文字の日米解釈差があることを明らかにした。

また、絵文字コミュニケーションを支援する一つの方法として、検索語に関連する絵文字を、たくさんの絵文字の中から探しだす問題を取り上げた。国家間の絵文字の解釈差を扱う絵文字検索を実現するためには、まず、絵文字解釈に現れる、意味的なあいまい性を考慮した絵文字検索を実現しなければならない。そこで、絵文字アンケート調査から取得した解釈語のリストと頻度情報を利用し、関連絵文字の適合率を計算する方法を提案することで、あいまい性を考慮した絵文字検索を実現した。

これまで異文化間コミュニケーションを支援し、異文化コラボレーションを促進させる試みはあったが（[6]に代表される）、文化差のセマンティクスを定式化できないがために、文化差を仲立ちするエージェントの構築が難しかった。今回、絵文字という、抽象的ではあるが、文化差が観測できる媒体を詳しく調べることで、異文化間交流を仲立ちできるエージェントの構築に一歩近づけたと思う。今後、更なる研究を重ね、[7]で取り上げているような、異文化をまたがるエージェントの構築に挑みたい。

謝辞 研究を進めるにあたって、京都大学情報学研究科社会情報学専攻の小山聡先生と NTT コミュニケーション科学基礎研究所の山下直美さん、そして、石田研究室の皆様（特に、八槇博史先生と中島悠さん）から有意義なコメントと助言をたくさんいただきました。ここに記して感謝いたします。本稿に掲載された絵文字の著作権は、特定非営利活動法人バンゲアに帰属します。

文 献

- [1] 森由美子, “子供たちの異文化間コミュニケーション,” 情報処理, 47 巻 3 号, pp.276-282, 2006.
- [2] T. Takasaki, “PictNet: Semantic Infrastructure for Pictogram Communication,” The Third International WordNet Conference (GWC-06), pp.279-284, 2006.
- [3] K. Uchimoto, N. Hayashida, T. Ishida, and H. Isahara, “Automatic Rating of Machine Translatability,” 10th Machine Translation Summit (MT Summit X), pp.235-242, 2005.
- [4] N. Yamashita and T. Ishida, “Automatic Prediction of Misconceptions in Multilingual Computer-Mediated Communication,” International Conference on Intelligent User Interfaces (IUI-06), pp.62-69, 2006.
- [5] T. Koda and T. Ishida, “Cross-cultural Comparison of Interpretation of Avatars' Facial Expressions,” IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp.130-136, 2006.
- [6] 野村早恵子, 石田亨, 船越要, 安岡美佳, and 山下直美, “アジアにおける異文化コラボレーション実験 2002: 機械翻訳を介したソフトウェア開発,” 情報処理, 44 巻 5 号, pp.503-511, 2003.
- [7] T. Ishida, “Communicating Culture,” IEEE Intelligent Systems, Special Issue on the Future of AI, Vol.21, No.3, pp.62-63, May/June 2006.