

特別研究報告書

Webサーベイを用いた絵文字解釈における  
文化差の検出

指導教員 石田 亨 教授

京都大学工学部情報学科

朱里 秀作

平成19年2月9日

## Webサーベイを用いた絵文字解釈における文化差の検出

朱里 秀作

### 内容梗概

近年インターネットの急速な普及に伴い、民間レベルのコミュニケーションが国境を越えることも珍しくなくなった。このような民間レベルのコミュニケーションの1つに、子どもたちによる絵文字コミュニケーションがある。

絵文字コミュニケーションでは、コミュニケーションに参加している子どもたちが、絵文字だけを組み合わせるメッセージを作成し、そのメッセージを他国の子どもに送り、意思疎通を図る。絵文字は言語に依存しないため、外国語を知らない子どもたちでも他国の子どもたちとコミュニケーションを図れる利点を持つ。しかし、一方で、次のような問題が生じる。

#### 1. 絵文字解釈の不一致によるメッセージ理解の食い違い

絵文字コミュニケーションにおいて、コミュニケーションを図ろうとする者同士の作成したメッセージに用いられた絵文字すべてが、お互いの間で共通の解釈を持った場合、絵文字自体は言語に依存していないので、言語の異なる者の間でもコミュニケーションできる可能性がある。しかし、お互いの間で1つでも解釈の異なる絵文字が用いられた場合、メッセージの理解に食い違いが生じ、正確なコミュニケーションができなくなる可能性が高い。したがって、絵文字コミュニケーションにおいて意思疎通を図るためには、お互いの使用する絵文字が共通の解釈を持つかどうかを知ることが重要である。

#### 2. 絵文字メッセージ作成の際に相手の絵文字が解釈不能

相手の絵文字メッセージに対して絵文字で返答するには、まず、相手の絵文字メッセージを解釈する必要がある。しかし、相手のメッセージが解釈できない場合は返答に困惑し、対話が途切れてしまうという欠点がある。したがって、円滑なコミュニケーションを図るためには、相手が理解できそうにない絵文字を知る必要がある。

ところで、絵文字の解釈は、人間がこれまでの生活環境で吸収してきた記憶に大きく依存するため、同じ文化や言語を背景に持つ者の間ではある程度はまとまっていると考えられる。この点に着目して、上記の問題を解決するために、日本と米国の子どもたちを対象とした絵文字アンケート調査の結果である日米

の解釈語リストを利用して、文化差を自動的に検出する方法を提案した。本研究の貢献は次の2点である。

1. 文化差のある絵文字の検出

日本と米国の子供たちが、システムで使用できる120個の絵文字についてどのように解釈しているかを調査した結果を利用し、各絵文字について国ごとにまとめられたデータを1つの集合とみなし、類似概念を用いて、各絵文字について国名の異なる2つの集合間の類似度を計算する手法を提案した。システムがこの類似度を利用者に通知することにより、利用者がその情報を参照し、類似度の小さな絵文字、すなわち、お互いの解釈の異なる可能性が高い絵文字の使用を控えることができるようになる。

2. 絵文字検出システムの実装と検証

上記の提案を実現するために、日米間において文化差のある絵文字を検出するシステムを実装する。そして、このシステムによって検出された絵文字が本当に文化差のあるものかを検証し、システムの妥当性・問題点について考察した。また、実装した検出システムの出力のうち、妥当性が確認できたデータを用いて、絵文字メッセージを作成する際に、使用したい絵文字を検索するための絵文字検索システムにおいては、国内における絵文字解釈の差異を考慮するだけでなく、文化差も考慮した絵文字検索が可能となる。

提案した手法により、各絵文字の類似度を計算し、アンケート調査で用いた120個の絵文字を文化差が小さい順に並べた。その結果、下位40個の絵文字のうちの約半数が、人間が文化差があると判断したものと一致した。

## Detecting Cultural Difference in Pictogram Interpretation Using Web Survey

Shusaku AKARI

### Abstract

With the increasing development of the Internet, communication between common people of different nations has been more and more popular, among with there is communication between children using pictograms.

In a pictogram-based communication, children from different nations exchange message with each other by using the pictograms only to reach mutual understanding. Since pictograms are independent with languages, children from different nations are able to communicate with each other even without grasping any foreign language. However, the following problems occur.

1. Misunderstanding in messages caused by disagreement of pictogram interpretations

In a pictogram-based communication, there is a possibility that people with different languages are able to communicate with each other because pictograms are independent with languages when all pictograms used in messages have common interpretations with each other. However, it may become impossible to communicate accurately because of misunderstanding in messages when only a pictogram which don't have a common interpretation are used with each other. Therefore, it is important to know whether pictograms used with each other have a common interpretation to communicate with each other in a pictogram-based communication.

2. Other party's pictogram should be not able to be interpreted

To reply with pictograms, first of all, it is necessary to interpret the other party's pictogram-based messages. However, there is a fault that users are embarrassed to reply, and the conversation becomes interrupted when messages can't be interpreted. Therefore, it is necessary to know the pictogram that the other party seems not to understand to communicate smoothly.

By the way, it is thought that pictogram interpretation between people who have same culture or language is common in some extent according to the memory that people absorbed in their environment of living. We pay attention

to this respect, to solve the above research issues, we propose an approach to automatic detection of cultural difference using the result of the web-based survey of Japanese and the US children. The contributions of this research include the following two aspects.

1. Detection of pictogram with cultural difference

Using the result of web-based survey that Japanese and the US children how to interpret 120 pictograms used in the system, we consider the data of each country brought together about each pictogram to be one set, and we propose an approach calculating the similarity between two sets about each pictogram. Users refer this similarity and come to be able to refrain from the use of the pictogram that has a small similarity, that is, the pictogram that won't have common interpretation with each other. when the system notifies the users the similarity.

2. Implement and verification of pictogram-detection system

To achieve the above proposal, we implements a system that detects a pictogram has cultural difference between Japan and the US. And, we verified the pictogram detected by this system truly has a cultural difference, and considered about validity and problems of that.

Using this proposal, we calculate the similarity of every pictogram, and arranged 120 pictograms used in the survey in order with a small cultural difference. As a result, some half the number of pictograms of 40 subordinate positions agreed with the pictograms judged that they have cultural differences.

# Webサーベイを用いた絵文字解釈における文化差の検出

## 目次

第1章	はじめに	1
第2章	絵文字解釈の文化差	2
2.1	絵文字解釈アンケート調査	3
2.2	絵文字解釈の日米比較	4
2.2.1	ジェスチャー	4
2.2.2	慣習	5
2.2.3	暗黙の了解	6
第3章	解釈語を用いた文化差の検出	6
3.1	文化の定義	7
3.2	絵文字解釈における文化差	8
3.3	文化差を扱うための枠組み	10
3.3.1	検出の方針	10
3.3.2	問題の解法	10
第4章	提案手法の実装と実験	14
4.1	文化差検出システムの構成	14
4.2	文化差検出実験	15
4.3	検出結果についての考察	18
4.3.1	検出結果が妥当な絵文字の例	18
4.3.2	検出結果が妥当でない絵文字の例	20
第5章	おわりに	23
	謝辞	24
	参考文献	24
	付録	A-1
A.1	ソースコード	A-1
A.1.1	restruct_edict.rb	A-1
A.1.2	comparison.rb	A-4
A.2	解釈語リスト	A-7

A.2.1	図 10 の絵文字の解釈語リスト . . . . .	A-7
A.2.2	図 11 の絵文字の解釈語リスト . . . . .	A-8
A.2.3	図 12 の絵文字の解釈語リスト . . . . .	A-9
A.2.4	図 13 の絵文字の解釈語リスト . . . . .	A-10
A.2.5	図 14 の絵文字の解釈語リスト . . . . .	A-11
A.2.6	図 15 の絵文字の解釈語リスト . . . . .	A-12
A.3	日米の次元を統一した特徴ベクトル . . . . .	A-13
A.3.1	図 13 の次元を統一した特徴ベクトル . . . . .	A-13
A.3.2	図 14 の次元を統一した特徴ベクトル . . . . .	A-14
A.3.3	図 15 の次元を統一した特徴ベクトル . . . . .	A-15
A.4	システムで使用された絵文字 . . . . .	A-16

## 第1章 はじめに

近年，パンゲア<sup>1)</sup>の絵文字コミュニケーション [1][2][3] をはじめとして，民間レベルでの国境を越えるコミュニケーションが増えていく傾向にある．しかし，文化・言語の異なるもの同士が意思の疎通を図ることは決して容易なことではない．そこで，コミュニケーションに，共通の解釈を持つ可能性のある，画像で描かれた絵文字を用いるという考えが生まれた．実際，パンゲアの絵文字コミュニケーションでは，コミュニケーションを図ろうとする者同士がそれぞれ絵文字だけを並べ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によって意思の疎通を図る，というコミュニケーション方法がとられている．このコミュニケーション方法には，メッセージに使用される絵文字が言語に依存しないため，言語の異なる者の間でもコミュニケーションできる可能性がある，という利点がある．しかし，コミュニケーションに絵文字だけを用いた場合，次のような問題が生じる．

### 1. 絵文字解釈の不一致によるメッセージ理解の食い違い

絵文字コミュニケーションにおいて，コミュニケーションを図ろうとする者同士の作成したメッセージに用いられた絵文字すべてが，お互いの中で共通の解釈を持った場合，絵文字自体は言語に依存していないので，言語の異なる者の間でもコミュニケーションできる可能性がある．しかし，お互いの中で1つでも解釈の異なる絵文字が用いられた場合，メッセージの理解に食い違いが生じ，正確なコミュニケーションができなくなる可能性が高い．したがって，絵文字コミュニケーションにおいて意思疎通を図るためには，お互いの使用する絵文字が共通の解釈を持つかどうかを知ることが重要である．

### 2. 絵文字メッセージ作成の際に相手の絵文字が解釈不能

相手の絵文字メッセージに対して絵文字で返答するには，まず，相手の絵文字メッセージを解釈する必要がある．しかし，相手のメッセージが解釈できない場合は返答に困惑し，対話が途切れてしまうという欠点がある．したがって，円滑なコミュニケーションを図るためには，相手が理解できそうにない絵文字を知る必要がある．

---

<sup>1)</sup> <http://www.pangaeaan.org/>

さらに、この絵文字コミュニケーションの利用者として想定されているのは、やっと1つの言語を習得したばかりの子供たち、あるいは、1つの国でしか生活したことのない子供たちであり、異言語・異文化間のコミュニケーションにおいて、物事の意識や理解に違いがあることを認識していない可能性が高い。そのため、絵文字コミュニケーションでは、コミュニケーションが成立するように、解釈の異なる可能性のある絵文字を検出する機能がシステムに求められる。ところで、絵文字の解釈は、人間がこれまでの生活環境で吸収してきた記憶に大きく依存するため、同じ文化や言語を背景に持つ者の間ではある程度はまとまっていると考えられる。この点に着目して、これら問題に対処するために、本研究では、この絵文字コミュニケーションにおいて、メッセージを送信する前に、解釈が異なる可能性のある絵文字を文化差のある絵文字として自動検出し、これをメッセージの作成者に通知する手法を提案する。これにより、メッセージ送信前にお互いの間で解釈の異なる絵文字を検出できれば、メッセージの誤解によるコミュニケーションの不成立を未然に防ぐことができる。

そこで、本研究では絵文字解釈についてのアンケート調査の結果と類似概念を利用し、上記の問題点を解決する手法を提案する。また、この手法を実現するシステムを実装し、実際に問題の解決を図る。

以降、本稿の構成は以下のとおりである。まず、第2章で、絵文字解釈のアンケート調査について述べ、その結果の分析について述べる。次に、第3章では、文化について述べ、文化差のある絵文字の検出アルゴリズムについて説明する。そして、第4章で、提案手法を実現するために開発したシステムについて述べ、本システムを用いた検出結果の妥当性・問題点について考察する。最後に、第5章にまとめを記し、本稿を総括する。

## 第2章 絵文字解釈の文化差

「絵文字が、どの国の子供たちにも同じような意味に見えるかどうか？」について知るために、NPO法人「Pangaea」のオンラインアンケート<sup>1)</sup>により調査された。調査結果のうち、日本の子供たちと米国の子供たちの回答のみを、日米の文化に通じた2人の人間が分析したところ、解釈の異なる理由が文化の違いに起因すると推測できる絵文字を19個発見した [4][5]。

---

<sup>1)</sup> 2007年1月現在でも、<http://www.pangaeaan.org/>内において実施されている。



図 1: アンケート調査で使した絵文字の 1 つ

表 1: 図 1 の絵文字の日米解釈語と頻度

解釈語 (日本)	頻度	解釈語 (米国)	頻度
おはよう	60	good morning	39
こんにちは	6	morning	8
おやすみ	4	hello	7
あいさつ	2	friends	6
おはなし	2	hi	5
おみまい	2	talking	4
かいわ	2	good night	3
朝	2	play	3
あさ	1	conversation	2
いすに座る	1	day	2
おしゃべり	1	goodmorning	2
おははししよう	1	sleep over	2
...	...	...	...
合計	101	合計	117

## 2.1 絵文字解釈アンケート調査

絵文字解釈アンケート調査は、日米の子供たちが絵文字をどのように解釈するかを調査するために、2005年10月1日から2006年11月30日まで、オンラインで絵文字アンケート調査を実施した。この調査において、回答者は120個の絵文字についての意味を回答する。この回答した意味を絵文字の「解釈語」と呼んでいる。

このアンケートによって得られた絵文字の解釈語を各絵文字について国別にまとめ、解釈語のリストを作成した [4][5]。また、各絵文字についてその解釈語が出現した数を解釈語の頻度とし、リストに加えた。例として、図1の絵文字に関して得られた解釈語のリストを表1に示す。ただし、表が長いため、一部を省略してある。

表1において、最初の行は項目名である。「解釈語(日本)」と「解釈語(米

国)」（以後，それぞれ日本の解釈語，米国の解釈語と呼ぶことにする。）の列には，図1の絵文字についての解釈語が現れている．そして，この解釈語に対応する頻度が，すぐ右側の「頻度」の列に示されている．また，最終行には絵文字についての解釈語の頻度の合計数，すなわち，解釈語の回答者数がやはり国別に表示されている．表の2行目は「おはよう」「60」「good morning」「39」と読むことができるが，これは図3の絵文字を見て「おはよう」と解釈した日本人が60人，「good morning」と解釈した米国人が39人であったことを意味する．また，最終行は，この絵文字の解釈語の回答者のうち，日本の子供たちが101人，米国の子供たちが117人であったこと意味する．なお，解釈語の表記ゆれや回答者の入力間違いなどに対しては，何も変更を施さずにそのまま解釈語として集計している．

## 2.2 絵文字解釈の日米比較

実施した絵文字解釈についてのアンケート調査において，集計された結果のうち，回答者を日本と米国の子供たちに限定して分析が行われた [4][5]．この結果，コミュニケーションシステムに使用された絵文字に，解釈が異なると判断できる絵文字，すなわち文化差ある絵文字が19個含まれており，これらがどのような性質の文化差に起因するかが，ジェスチャー・性別・時間概念・空間認識・顔文字の解釈・馴染みのものという6つの領域に分類されている．本節では文化差のある絵文字の一部について解説し，他の絵文字については割愛する．また，本論文では既存の分析結果を再分析し，「馴染みのもの」という領域を「慣習」と「暗黙の了解」という2つの領域に細分類した．なお，文化差があるとされた絵文字は，付録の絵文字一覧において背景が黄色のものである．

### 2.2.1 ジェスチャー

本節では，ジェスチャーの使い方の違いにより日米の解釈に差が出たと思われる絵文字について分析する．

図2の左側のような，人間が頭の上で，両腕で円を描いた絵文字の解釈語として，日本の解釈語には「ok」や「まる」「正解」のような肯定の意見が，85%と圧倒的に多く見られたのに対し，米国の解釈語には「exercise」や「jump rope」，「yoga」など運動に関連する語が約45%と多く見られた．逆に，日本で運動に関連する語を答えたのは1人(1%)だけで，米国で正解に関連する語を答えたのは4人(3%)だけであった．



図 2: ジェスチャーに関する絵文字



図 3: 慣習に関する絵文字

また、図 2 の右側のような、人間が顔の下付近で両手を交差している絵文字の解釈語として、日本の解釈語には「だめ」や「ダメ」、「ばつ」のような否定の意見が、85%と圧倒的に多く見られたのに対し、米国の解釈語には「mad」や「angry」、「stubborn」など否定的な感情を表す語が多く、中でも「怒り」に関する語が70%近くも見られた。

いずれの絵文字についても、日本ではジェスチャーとして解釈した子供が多数派であったが、米国では少数派であった。これは、日本では、これらのジェスチャーが一般的であるが、米国ではそうでないことに起因すると予測できる。

#### 2.2.2 慣習

本節では、慣習や伝統の違いにより解釈に差が現れた絵文字について分析する。図 3 のような、表彰台に立つ人間と周りで踊っている人間が描かれた絵文字の解釈語として、日本の解釈語には「運動会」(40%) が最も多かったが、意味的にまとめると「一番」や「優勝」などの意味の解釈語 (45%) が多く見られた。一方、米国の解釈語には「winner」や「winning」、「champion」など勝利に関連する語が70%と圧倒的に多く見られ、日本の運動会に相当する「sports」と回答した人は1人(1%未満)であった。この解釈の違いは、運動会は、日本では一般的な学校行事であるが、米国ではあまり行われないうことに起因するものと推測できる。



図 4: 暗黙の了解に関する絵文字

### 2.2.3 暗黙の了解

本節では、暗黙の了解における解釈の違いにより差が現れる絵文字について分析する。

まず、暗黙の了解とは、ある特定集団において、普通の一般名詞が固有名詞を意味することである。例えば、日本の古典においては暗黙の了解として「山」といえば「比叡山」を意味し、「祭り」といえば「葵祭り」を意味する。

図 4 のような、赤い塔が描かれている絵文字の解釈語として、日本の解釈語には「東京タワー」が 50% と多く見られ、残りの 50% は一般名詞としての「タワー」と「エッフェル塔」など東京タワー以外の塔であった。一方、米国の解釈語には「eiffel tower」が 40% と多く見られ、次に多かったのは、eiffel tower から連想される「paris」や「france」など (25%) であった。日米いずれにおいても、「一般名詞の塔」と解釈するより、「ある固有名詞の塔」を回答した子供の方が多かった。これは、暗黙の了解として、日本では塔といえば東京タワーを意味し、米国では塔といえばエッフェル塔を意味することに起因するものであると推測できる。

## 第 3 章 解釈語を用いた文化差の検出

絵文字コミュニケーションにおいて、お互いの解釈が異なる絵文字を用いた場合、コミュニケーションが成立しない可能性がある。したがって、意思疎通を図るためには、お互いの使用する絵文字が共通の解釈を持つかどうかを知ることが重要である。

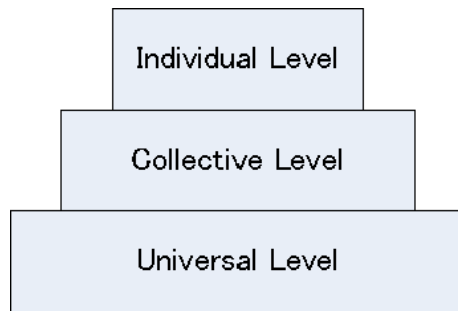


図 5: メンタル・プログラムの階層

### 3.1 文化の定義

「文化」とは実に複雑な言葉であり、多数の文献によりさまざまな定義がなされている。例えば、岩波書店の広辞苑第 5 版では、次のように書かれている。

(culture) 人間が自然に手を加えて形成してきた物心両面の成果。衣食住をはじめ技術・学問・芸術・道徳・宗教・政治などの生活形成の様式と内容とを含む。文明とほぼ同義に用いられることが多いが、西洋では人間の精神的な生活にかかわるものを文化と呼び、技術的発展のニュアンスが強い文明と区別する。

また、Sitaram の文献 [6] では、次のように定義されている。

一つの社会システムであり、世代から世代へと受け継がれてきた学習される行動、芸術的伝統、技術的成果、コミュニケーション技能、宗教的信条、哲学的概念、および遺伝的特徴からなっているもの。

このような多様な定義の中で、本研究における文化の定義として参考とするのは、オランダの社会心理学者である Geert Hofstede のモデルである [7][8]。このモデルにおいて、人間には行動を制御するメンタル・プログラムがあると想定されており、このメンタル・プログラムの差異が人間の認識や状況判断、行動の違いとなって現れる、とある。メンタル・プログラムは、世代から次の世代へと受け継がれていくものと考えられており、その継承方法によって 3 つの階層に分けられている。この階層を図 5 に示す。下の階層ほど原始的で、上の階層ほど高等なプログラムである。

一番下の階層には「Universal Level」があり、これは感情表現など人間の生態的活動を司るプログラムであると考えられている。この階層のプログラムは、主に遺伝的に次の世代へと継承され、全人類が共有しているものであると考えられている。

真ん中の階層には「Collective Level」があり、これは衣食住の方法など人間の生活形成のための活動を司るプログラムであると考えられている。この階層のプログラムは、主に学習を通じて継承され、ある特定の集団の中でのみ共有されるものであると考えられている。Hofstede は、主観的な意味で文化と呼ばれるものは、すべてのこの階層に所属するものであると言っている。

一番上の階層には「Individual Level」があり、これは人間の個性を司るプログラムであると考えられている。この階層のプログラムは、部分的に受け継がれていくものの、その説明が難しいとされており、個人により異なるものであると考えられている。

このモデルのメンタル・プログラムを持った人間が、絵文字を解釈することを考える。まず、「Universal Level」で絵文字を解釈した場合を考える。この階層では全人類がプログラムを共有しているため、解釈の違いは現れないと想定される。次に、「Collective Level」で絵文字を解釈する場合を考える。この階層では特定の集団で同じプログラムを共有しているため、この集団内では解釈の違いは現れないと想定される。しかし、集団ごとにメンタル・プログラムが異なっているため、異なる集団に属する個人の間では解釈が異なる可能性がある。最後に、「Individual Level」で絵文字を解釈する場合を考える。この階層では一般に個人ごとにメンタル・プログラムが異なっているため、解釈は異なる可能性がある。以上において、絵文字解釈に差異が現れるのは、「Collective Level」あるいは「Individual Level」で解釈した場合であるが、本研究では、前者の違いを文化差、後者の違いを個人差として捉えることにする。したがって、この Hofstede のモデルにおける「Collective Level」のメンタル・プログラム、すなわち「世代から世代へと受け継がれていく人間の活動を司るもののうち、主に学習を通じて継承されるもの」を文化の定義とする。

### 3.2 絵文字解釈における文化差

前節では文化の定義を行ったので、本節では絵文字解釈における文化差について改めて考える。文化とは、「主に学習を通じて継承される人間の活動を司るもの」であることはすでに述べた。この定義に従うと、文化差とは「異なる集団間における『主に学習を通じて継承される人間の活動を司るもの』の差異」ということになる。学習を通じて継承されるものには言語表現なども含まれるため、絵文字アンケート調査での解釈語の言語表現の違いは文化差ということに



図 6: 靴屋の絵文字

表 2: 図 6 の絵文字の解釈語と頻度

解釈語	頻度 (日本)	頻度 (米国)
「靴屋」「shoe store」など靴に関連する店舗的なもの	97	62
「靴」「shoe」など靴に関連する店舗的でないもの	8	85
その他	1	5
合計	106	152

なる。今回のアンケート調査では、絵文字の解釈を解釈語という言語表現で回答してもらう手法を採用したが、絵文字の解釈を言語表現に変換する過程において、前述の意味での文化差が発生している。本研究の「文化差のある絵文字の検出」において使用した「文化差」は、このような文化差ではなく、「コミュニケーションを図ろうとする者の間での解釈の違い」という意味で使用した。

また、ある言語で表現できる語が、他の言語では簡単に表現できない場合がある。例えば、英語の「identity」は日本語に翻訳しにくい単語とされているし、日本語の「柔道」に相当する概念自体がかつての英語圏にはなかった（現在では、柔道という概念が英語圏にも広まったため、「judo」という新たな語が作られている。）また、図 6 の日本の子供たちの絵文字の解釈語の 1 つである「靴屋」に相当する語は、一般的な和英辞典では「a shoe store」や「a shoe shop」とされており、米国の子供たちの解釈語にも冠詞を除いた「shoe store」や「shoe shop」として現れていた。しかし、回答者の 55% は 1 語で「shoe」や「shoes」と回答した。図 6 の解釈語リストを「靴屋」のように店舗として捉えた解釈語と「靴」のように店舗的でないものとして捉えたものと解釈語を区別して、解釈語を意味的にまとめた新たなリストを作成したところ表 2 のようになった。この表を見ると図 6 の捉え方に大きな差があることが分かる。この差は、絵文字の解釈の違いではなく、表現方法の複雑さによる違いであると考えられる。

したがって、「絵文字解釈における文化差」には、解釈語における言語表現の

違いや微妙なニュアンスの違いを含まないものとする。また、第2章での分析においても図6の絵文字は文化差のない絵文字とされていたため、本節で述べた「絵文字解釈における文化差」の捉え方は、アンケート調査の分析者の捉え方と概ね一致すると考えられる。

### 3.3 文化差を扱うための枠組み

日米間の絵文字解釈の文化差を検出するために、本研究では、類似概念を用いて日米間の解釈の類似度を計算することを考える。なお、類似度は絵文字ごとに計算する。

#### 3.3.1 検出の方針

第2章のアンケート調査において、日米の子供たちが絵文字をどのように解釈しているかを調べるために解釈語リストを作成した。これを用いて、まず、ベクトル空間モデルの概念を用いて、ある絵文字についての、日本の解釈語リストと米国の解釈語リストをそれぞれ、解釈語の頻度をベクトルの成分とするV次元（Vは各解釈語リストに出現する解釈語の総数）の特徴ベクトルとして捉える。次に、類似概念を用いて、日米両方の解釈語リストの特徴ベクトル（以後、これら2つのベクトルを「日本の特徴ベクトル」および「米国の特徴ベクトル」と呼ぶことにする。）の類似度を求めることを考えるが、解釈語の言語が異なるため、日本の解釈語と米国の解釈語の意味的な対応を考える必要がある。例えば、日本語の「おはよう」という解釈語は英語では「good morning」に対応する。本研究では、この対応に和英辞書を用いた。和英辞書を用いて、日本の特徴ベクトルにおいて、翻訳可能な（辞書に見出しとして含まれる）すべての日本の解釈語を英語に翻訳する。そして、この解釈語が翻訳された日本の特徴ベクトルと、米国の特徴ベクトルとの類似度を計算する。この類似度の値から絵文字に文化差があるかどうかを判断する。

#### 3.3.2 問題の解法

日米の解釈語リストを、2つの特徴ベクトル

$$I_j = ((w_{1,j}, f_{1,j}), (w_{2,j}, f_{2,j}), \dots, (w_{m,j}, f_{m,j}))$$

および

$$I_e = ((w_{1,e}, f_{1,e}), (w_{2,e}, f_{2,e}), \dots, (w_{n,e}, f_{n,e}))$$

として捉える。  $w_{k,j}, w_{k,e}, f_{k,j}, f_{k,e}$  は、それぞれ以下のように定める。

- $w_{k,j}$ : 日本の特徴ベクトル  $I_j$  の解釈語の 1 つ .
- $w_{k,e}$ : 米国の特徴ベクトル  $I_e$  の解釈語の 1 つ .
- $f_{k,j}$ : 解釈語  $w_{k,j}$  の頻度 .
- $f_{k,e}$ : 解釈語  $w_{k,e}$  の頻度 .
- $m$ : 日本の特徴ベクトル  $I_j$  の次元数
- $n$ : 米国の特徴ベクトル  $I_e$  の次元数

また ,  $S_{w,j}, S_{w,e}$  をそれぞれ以下のように定める .

- $S_{w,j}$ : 日本の特徴ベクトル  $I_j$  の解釈語  $w_{k,j}$  の集合 .
- $S_{w,e}$ : 米国の特徴ベクトル  $I_e$  の解釈語  $w_{k,e}$  の集合 .

例えば , 1 の解釈語リストを特徴ベクトルとしてみなすと ,

$$I_j = ((\text{おはよう}, 60), (\text{こんにちは}, 6), \dots, (\text{病院}, 1))$$

$$I_e = ((\text{good morning}, 39), (\text{morning}, 8), \dots, (\text{woke up}, 1))$$

となり ,  $S_{w,j}, S_{w,e}$  は ,

$$S_{w,j} = \{ \text{おはよう}, \text{こんにちは}, \dots, \text{病院} \}$$

$$S_{w,e} = \{ \text{good morning}, \text{morning}, \dots, \text{woke up} \}$$

となる .

一般の和英辞書では , 日本語の見出し語 1 つに対して , 1 つ以上の意味が英語で書かれている . 英語の意味が 1 つしか載っていない日本語に限定すれば , 和英辞書は日本語から英語への写像であると考えることができる . 英語の意味が複数の場合でも , 何らかの規則を用いて意味をただ 1 つ指定できれば , 和英辞書に載っているすべての見出し語において , 和英辞書は写像であると考えることができる . 本研究では , 和英辞書を用いた日本語の解釈語から英語への写像 ( 翻訳 ) を次のように定義する .

定義 1 . 日本語の解釈語から英語への写像 写像  $translate : w_{k,j} \rightarrow w_{k,j2e}$  を次の 2 つに場合分けして定義する .

1. 和英辞書に日本語の解釈語  $w_{k,j}$  に一致する見出し語がある場合 , 見出し語に対応する最初の英語 ( 単語あるいは熟語 ) を  $w_{k,j2e}$  とする .
2. 和英辞書に日本語の解釈語に一致する見出し語がない場合 , 恒等写像として日本語をそのまま  $w_{k,j2e}$  とする .

この写像を用いて，特徴ベクトルの解釈語の言語を翻訳する写像を次のように定義する．

定義 2．定義 1 の特徴ベクトルへの適用 写像  $translate\_all : I_j \rightarrow I_{j2e}$  を次のように定義する．

- $\forall w_{k,j} \in S_{w,j}$  について，写像  $translate$  を実行する．すなわち，

$$I_{j2e} = ((w_{1,j2e}, f_{1,j}), (w_{2,j2e}, f_{2,j}), \dots, (w_{m,j2e}, f_{m,j}))$$

となる．

また， $S_{w,j2e}$  を以下のように定める．

- $S_{w,j2e}$ : 解釈語が翻訳された日本の特徴ベクトル  $I_{j2e}$  の解釈語  $w_{k,j2e}$  の集合．  
例えば， $I_j = ((おはよう, 60), (こんにちは, 6), \dots, (病院, 1))$  に定義 2 の写像  $translate\_all$  を適用すると，

$$I_{j2e} = ((good\ morning, 60), (hello, 6), \dots, (hospital, 1))$$

$$S_{w,j2e} = \{good\ morning, hello, \dots, hospital\}$$

となる．

解釈語が翻訳された日本の特徴ベクトル  $I_{j2e}$  と米国の特徴ベクトル  $I_e$  の次元数は，一般に異なる．類似度を計算するには，2 つのベクトルが同次元でなければならないため，次元を統一する必要がある．ここでは，ベクトルの次元数を 2 つの解釈語の集合の和である  $S_w$  の要素数に統一する．すなわち， $S_w$  は以下の式で定義される．

$$S_w = S_{w,j2e} \cup S_{w,e}$$

また， $S_w$  の要素数  $l$  とする． $l$  には次のような制限がある．

$$\begin{aligned} & \max(m, n) \leq l \leq m + n \\ & \left( \text{ただし } \max = \begin{cases} m(m \geq n) \\ n(m < n) \end{cases} \right) \end{aligned}$$

$S_w$  の 1 要素  $w_k$  と要素数  $l$  を用いて，2 つのベクトルから次元を統一したベクトルへの写像を次のように定義する．

定義 3．ベクトルの次元の統一 写像  $dim\_unify : (I_{j2e}, I_e) \rightarrow (I'_{j2e}, I'_e)$  を次のように定義する．

- $I_{j2e}$  を  $I'_{j2e}$  へ ,  $I_e$  を  $I'_e$  へ移す . ただし ,

$$S_w = \{w_1, w_2, \dots, w_l\}$$

$$I'_{j2e} = ((w_1, f'_{1,j}), (w_2, f'_{2,j}), \dots, (w_l, f'_{l,j}))$$

$$I'_e = ((w_1, f'_{1,e}), (w_2, f'_{2,e}), \dots, (w_l, f'_{l,e}))$$

とするとき , 次の 2 つの条件をみたすものと定義する .

1.  $w_{x,j2e} \in S_{w,j2e}$  をみたす  $\forall x$  について ,  $w_{x,j2e} = w_y$  のとき  $f_{x,j} = f'_{y,j}$  をみたすように  $w_y, f'_{y,j}$  を定める .
2.  $w_{x,e} \in S_{w,e}$  をみたす  $\forall x$  について ,  $w_{x,e} = w_y$  のとき  $f_{x,e} = f'_{y,e}$  をみたすように  $w_y, f'_{y,e}$  を定める .

例えば ,

$$I_{j2e} = ((\text{good morning}, 60), (\text{hello}, 6), \dots, (\text{hospital}, 1))$$

$$I_e = ((\text{good morning}, 39), (\text{morning}, 8), \dots, (\text{woke up}, 1))$$

に定義 3 の写像  $dim\_unify$  を適用すると ,

$$I'_{j2e} = ((\text{good morning}, 60), (\text{hello}, 6), (\text{morning}, 3), \dots, \\ (\text{朝の挨拶 (おはよう)}, 1), (\text{hospital}, 1), \dots, (\text{woke up}, 0))$$

$$I'_e = ((\text{good morning}, 39), (\text{hello}, 7), (\text{morning}, 8), \dots, \\ (\text{朝の挨拶 (おはよう)}, 0), (\text{hospital}, 0), \dots, (\text{woke up}, 1))$$

となる .

以上より 2 つのベクトルが同次元に統一されたので , 類似度を計算することができる . 類似度は次のコサイン相関値を用いて定義する .

定義 4 . 類似度 類似度  $sim(I'_{j2e}, I'_e)$  を次式により定義する .

$$sim(I'_{j2e}, I'_e) = \frac{\sum_{k=1}^l f'_{k,j} f'_{k,e}}{\sqrt{\sum_{k=1}^l f'_{k,j} + \sqrt{\sum_{k=1}^l f'_{k,e}}}}$$

ベクトル空間と類似度の関係を図 7 に示す . これは , 図 1 の絵文字についての , 解釈語が翻訳された日本の特徴ベクトル (赤いベクトル) と , 米国の特徴ベクトル (青いベクトル) を 「good morning」 , 「hello」 , 「morning」 の 3 つの次元に限定してグラフに表したものである . 類似度を求めることは , 2 つのベクトル

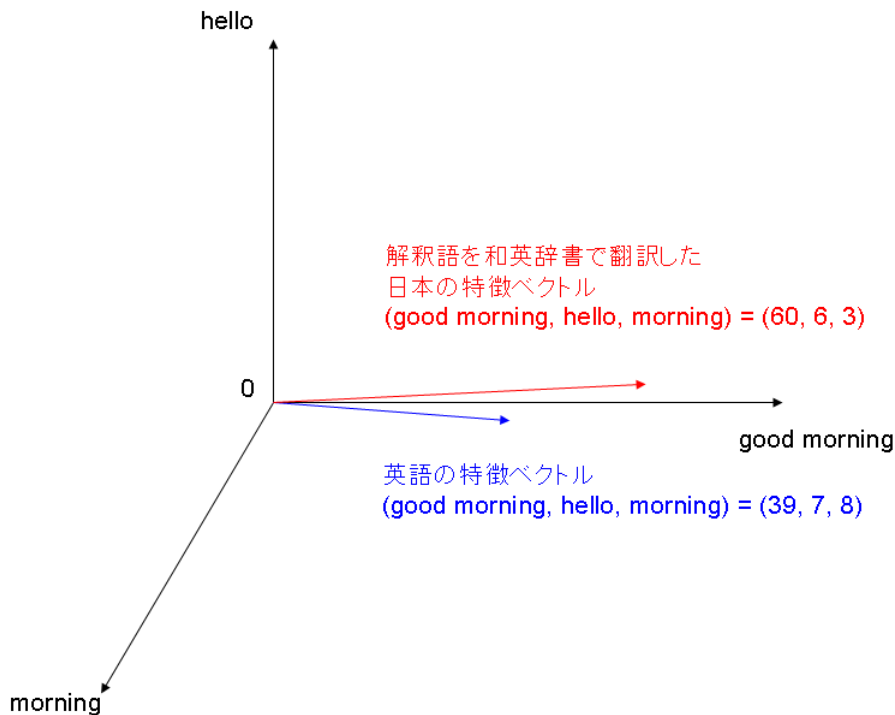


図 7: ベクトル空間と類似度

のなす角を求めることに相当する。

定義 1~4 および閾値を表すパラメータ  $\alpha$  ( $0 \leq \alpha \leq 1$ ) を用いて、文化差のある絵文字を検出する問題を、次式をみたす特徴ベクトル  $I_j, I_e$  の解釈語リストの絵文字を求める問題として定義する。

$$\text{sim}(I'_{j2e}, I'_e) < \alpha$$

## 第 4 章 提案手法の実装と実験

本章では、第 3 章で述べた手法に基づいて実装した文化差検出のシステムについて説明し、システムの出力についての妥当性を吟味する。

### 4.1 文化差検出システムの構成

本システムは、絵文字検索システムにおいて、文化差のある絵文字を検出するために実装された。絵文字検索システムとは、利用者が必要な絵文字を、検索語と第 2 章のアンケート調査の集計結果を用いて検索するシステムである。絵文字検索システムおよび文化差検出システムの構成図を図 8 に示す。

図 8 において、文化差検出システムは、まず、データベース内の絵文字と解

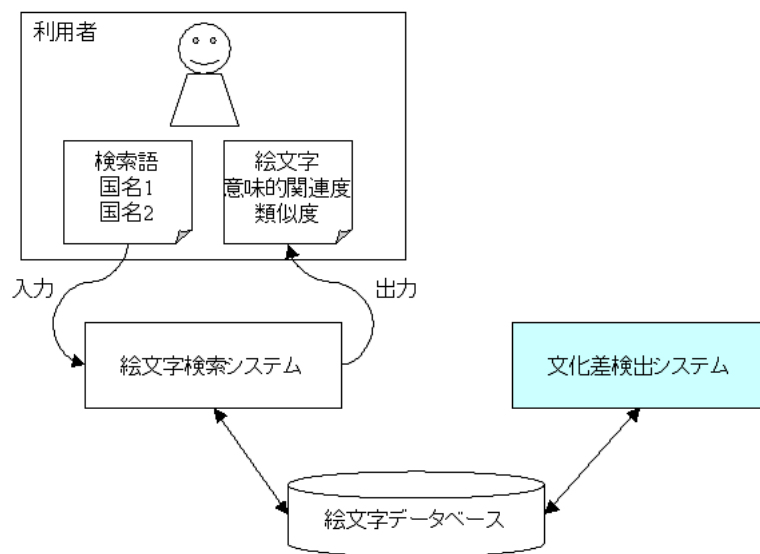


図 8: 絵文字検索システムおよび文化差検出システムの構成図

訳語リストを入力として受けつける。次に、第 3 章によって提案した解法によって、日米間の解釈の類似度を各絵文字について計算し、出力する。最後に、この出力をデータベースに格納する。文化差検出システムのシーケンス図を図 9 に示す。

絵文字検索システムでは、利用者は、伝えたいメッセージに用いる絵文字を、単語レベルの検索語の入力を繰り返すことによってメッセージを作成する。1 つの絵文字を検索する際に、利用者は最初に検索語、および、自分とメッセージの送信相手の国名情報を入力する。この入力に対して、システムはデータベース内の絵文字の解釈語リストを用いて、検索語に関連する絵文字をすべて探し出し、それらを意味的関連度と呼ばれる指標を用いて、検索語に近い順に並べ替えを行った上で、利用者に提示する。絵文字検索システムでは、文化差検出システムの出力結果を用いるため、利用者が絵文字を検索する前に、文化差検出システムを実行する必要がある。

## 4.2 文化差検出実験

絵文字アンケート調査から収集した絵文字 120 個の日米解釈語リストを入力として文化差検出システムを用い、文化差のある絵文字を検出した。検出に用

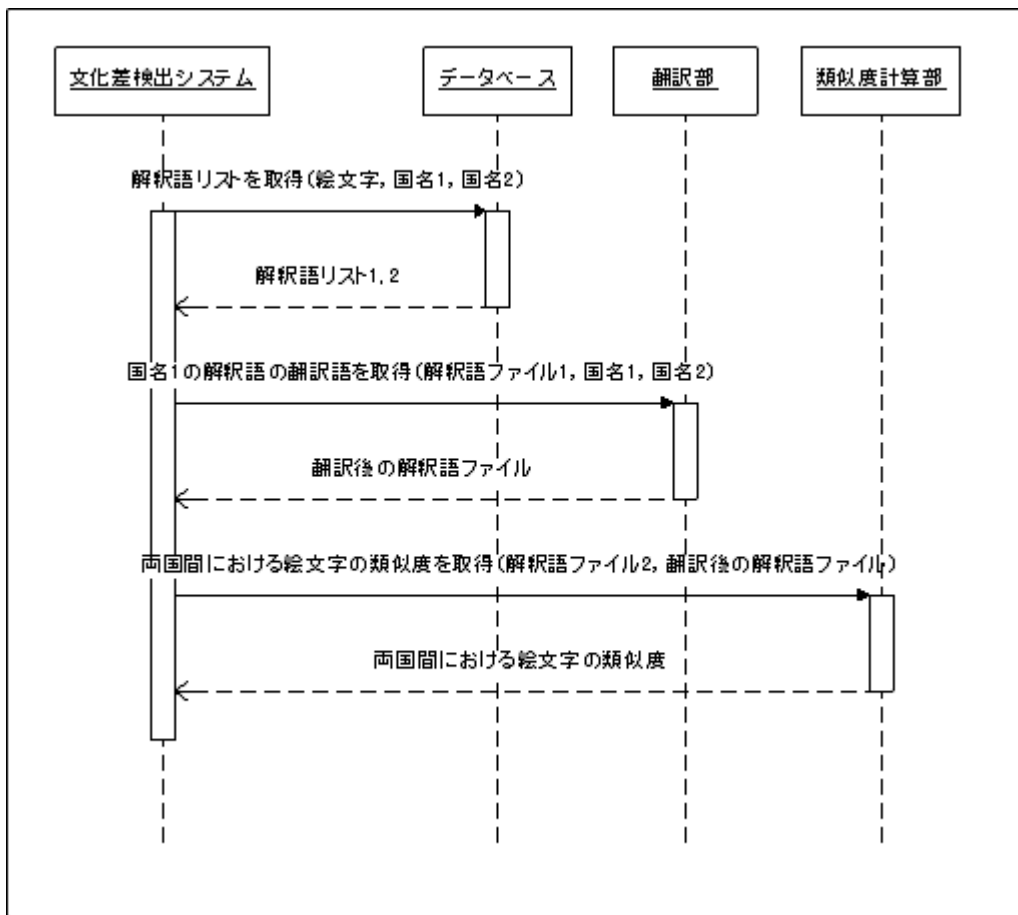


図 9: 文化差検出システムのシーケンス図

表 3: 絵文字 120 個の日米絵文字解釈語を用いた文化差判定コサイン類似度の計算結果 (降順にソート)

番号	類似度	番号	類似度	番号	類似度	番号	類似度
21	0.9989326	95	0.7358348	18	0.2066466	46	0.0406229
11	0.9973791	8	0.7061836	* 88	0.1768083	101	0.0397866
29	0.9906155	104	0.6990132	79	0.1764161	60	0.039227
38	0.9904288	41	0.5734594	32	0.1397863	30	0.0341964
33	0.9886472	48	0.5023211	* 78	0.1381676	75	0.033719
22	0.9880708	120	0.4982087	119	0.1249264	90	0.0337173
65	0.9850251	* 36	0.4862806	62	0.1191116	108	0.0291137
23	0.9804605	17	0.4480486	39	0.1170575	92	0.0289192
6	0.9784825	102	0.4308068	47	0.1096754	87	0.0235808
96	0.9777913	109	0.4259612	94	0.1063516	* 77	0.0226125
117	0.9768221	16	0.4219149	50	0.1048301	58	0.0188512
31	0.9429818	112	0.4048521	114	0.1036768	* 72	0.0186724
1	0.9419528	113	0.3851261	3	0.1032576	59	0.0184302
15	0.9272528	73	0.3673519	103	0.0996132	84	0.0128153
24	0.9169242	*111	0.349399	105	0.0931841	55	0.0106724
12	0.906803	82	0.3353841	9	0.0868782	* 89	0.0103205
13	0.9050622	54	0.3296177	116	0.0854233	81	0.0095609
25	0.8862193	115	0.3155664	98	0.0843883	106	0.0091359
64	0.8825686	52	0.2876146	110	0.0715901	* 5	0.00905
66	0.870776	56	0.2788978	7	0.070318	28	0.0088578
* 26	0.8630772	* 68	0.2632271	99	0.0694427	91	0.0073715
45	0.8491546	43	0.2521031	14	0.0670665	49	0.0045472
35	0.8317453	2	0.2436303	51	0.058613	100	0.0035594
34	0.8170683	71	0.241252	* 37	0.057404	40	0.0033931
63	0.8112086	10	0.232357	* 67	0.0571662	93	0.0033837
* 27	0.798863	19	0.2188715	74	0.0562244	85	0.0009005
97	0.7959037	20	0.2153297	* 69	0.0558075	42	0
44	0.7907515	* 80	0.2106237	53	0.0553016	* 76	0
61	0.7472094	118	0.2104891	* 4	0.0484549	* 86	0
57	0.737765	* 83	0.2077716	70	0.0482434	107	0

いた閾値  $\alpha$  は 0.058 としたが，これは，すべての絵文字について提案手法で計算した類似度を見比べ，検出結果が最もよくなるように設定した．表 3 に類似度の計算結果を示す．また，第 2 章で日米間において文化差のある絵文字と判断したものは，該当する番号の前に\*を付してある．

表 4: 図 10 の絵文字の解釈語リスト



図 10: 001.gif

解釈語 (日本)	頻度	解釈語 (米国)	頻度
おはよう	60	good morning	39
こんにちは	6	morning	8
おやすみ	4	hello	7
あいさつ	2	friends	6
おはなし	2	hi	5
おみまい	2	talking	4
かいわ	2	good night	3
朝	2	play	3
...	...	...	...
合計	101	合計	117

### 4.3 検出結果についての考察

文化差検出システムによって検出された絵文字について、妥当であるかを考察する。本システムは、第 2 章で文化差があると人間が判断した絵文字の約半数を検出し、検出された絵文字の 3 割程度が妥当な検出であった。

#### 4.3.1 検出結果が妥当な絵文字の例

まず、文化差の小さな絵文字の例をあげる。今、絵文字「001.gif」(図 10)に関して、表 4 のような解釈語リストが得られている。これは図 10 の絵文字に関する解釈語とその頻度を国別に集計した 2 つの解釈語リストである。ただし、本章の解釈語リストにおいて、頻度の少ない解釈語は割愛し、完全な解釈語リストは付録とした。

このデータを用いて、文化差検出システムで文化依存の絵文字解釈の類似度を計算すると「0.9419528」となり、検出されなかった。この類似度は 0 以上 1 以下の実数で、値が大きいほど文化差が小さいことを意味する。

一方で、表 4 を人間が見て、文化差があるかどうかを判断することを考える。各国の解釈語において頻度が最大の単語が同じような意味であり、また、その頻度が他の解釈語のものに比べて比較的大きければ、文化差は小さいと判断できる。この例では、頻度最大の解釈語は「おはよう」と「good morning」であり、これらは同義である。また、これらの頻度は他の解釈語に比べて比較的大きいので、人間も文化差が小さいと判断できる。さらに、第 2 章での分析においても図 10 は文化差のない絵文字とされていた。したがって、この絵文字についてのシステムの検出結果は妥当であるといえる。

次に、文化差の大きな絵文字の例を挙げる。絵文字「004.gif」(図 11)、「005.gif」

表 5: 図 11 の絵文字の解釈語リスト



図 11: 004.gif

解釈語 (日本)	頻度	解釈語 (米国)	頻度
ok	19	exercise	22
まる	10	jump rope	6
正解	9	happy	5
オッケー	8	exercising	4
いいよ	6	yoga	4
はい	5	circle	3
せいかい	4	dance	3
おっけー	2	stretch	3
丸	2	stretching	3
...	...	...	...
合計	101	合計	115

表 6: 図 12 の絵文字の解釈語リスト



図 12: 005.gif

解釈語 (日本)	頻度	解釈語 (米国)	頻度
だめ	27	mad	37
ダメ	12	angry	35
ばつ	6	no	5
バツ	4	stubborn	4
いいえ	3	anger	2
はずれ	3	frustrated	2
やだ	3	i'm angry	2
考える	3	upset	2
不正解	3	anger; cutoff	1
...	...	...	...
合計	101	合計	117

(図 12) に関しても表 5, 表 6 のような解釈語リストが得られている。先例と同様に類似度を計算すると, 値は順に「0.0484549」, 「0.0090500」となり, 文化差のある絵文字として検出された。

やはり先例と同様に人間が文化差を判断することを考える。日米それぞれの頻度最大の解釈語は, 図 11 では「ok」および「exercise」, 図 12 では「だめ」および「angry」であるが, これらの意味は異なる。また, 図 11 において「ok」に次いで頻度の大きな「まる」や「正解」なども「exercise」とは異なる意味であり, 図 12 においても「angry」に次いで頻度の大きな「mad」は「だめ」と異なる意味である。さらに, 第 2 章での分析においても, どちらの絵文字も文化差のある絵文字とされていた。したがって, 文化差は大きいと判断でき, と



図 13: 005.gif

表 7: 図 13 の絵文字の解釈語リスト

解釈語 (日本)	頻度	解釈語 (米国)	頻度
おもちゃ	69	toys	100
玩具	5	baby toys	2
こども	1	playing	2
...	...	...	...
合計	81	合計	128

もに結果は妥当である。

#### 4.3.2 検出結果が妥当でない絵文字の例

本節では、システムの検出結果と人間が下した判断が異なる例を挙げる。初めに、人間は文化差の小さな絵文字と判断したが、文化差が大きいと検出された例をあげる。絵文字「107.gif」(図 13)に関して、表 7 のような解釈語リストが得られている。このデータを用いて類似度を計算すると「0.0000000」となり、文化差のある絵文字として検出された。しかし、表 7 において頻度最大の解釈語は「おもちゃ」と「toys」であり、これらは同じ意味である。また、この 2 つの解釈語はそれぞれ全体の 8 割程度を占め、文化差は小さいように見える。さらに、第 2 章での分析においては、図 13 は文化差のない絵文字とされていた。したがって、検出結果は妥当でないといえる。

次に、人間は文化差の大きな絵文字と判断したが、文化差が小さいと検出された例をあげる。また、絵文字「26.gif」(図 14)、「27.gif」(図 15)に関して、表 8, 9 のような解釈語リストが得られている。同様に類似度を計算すると、順に「0.8630772」と「0.7988630」となり、検出されなかった。解釈語リストにおいて、日本では、約 99% が図 14 を女性(「女の人」、「おかあさん」などの解釈語を含む)、図 15 を男性(「男の人」、「おとうさん」などの解釈語を含む)と解釈したのに対し、米国では、図 14 を女性、図 15 を男性と解釈したのは 8 割程度にとどまり、約 2 割の回答者が図 14 を男性、図 15 を女性という正反対の解釈をした。どちらの絵文字も大多数の解釈は一致しているものの、米国では 2 割という少なくない人たちが多数派と正反対の解釈をしていることを考慮すると、文化差は大きいと言える。さらに、第 2 章での分析においても、どちらの絵文字も文化差のある絵文字とされていた。したがって、検出結果はともに妥当でないといえる。

それでは、妥当でない結果が算出された原因について考察する。類似度を計

表 8: 図 14 の絵文字の解釈語リスト

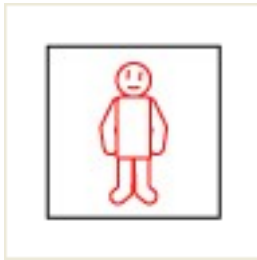


図 14: 026.gif

解釈語 (日本)	頻度	解釈語 (米国)	頻度
女の人	28	woman	43
女性	27	man	17
女	10	mom	13
おかあさん	5	dad	11
お母さん	3	adult	8
成人女性	3	mother	7
おとな	2	women	4
おんな	2	father	3
おんなのひと	2	person	3
...	...	...	...
合計	100	合計	152

表 9: 図 15 の絵文字の解釈語リスト



図 15: 027.gif

解釈語 (日本)	頻度	解釈語 (米国)	頻度
男の人	28	man	51
男性	27	dad	15
男	10	woman	12
おとうさん	4	adult	9
お父さん	4	mom	8
成人男性	3	father	7
おとこ	2	boy	3
おとこのひと	2	mother	3
おとな	2	person	3
...	...	...	...
合計	101	合計	152

表 10: 図 13 の絵文字について，次元を統一した特徴ベクトル

解釈語	日本	米国
toys	0	100
toy	74	0
baby toys	0	2
playing	0	2
dream	1	0
乳児の遊び	1	0
トイ	1	0
幼児玩具	1	0
子供のおもちゃ	1	0
child	1	0
小さい子	1	0
children toys	0	1
la la land	0	1
babies toys	0	1
...	...	...

表 11: 図 14 の絵文字について，次元を統一した特徴ベクトル

解釈語	日本	米国
woman	69	43
man	2	17
mother	9	7
mom	0	13
adult	3	8
dad	0	11
women	0	4
成人女性	3	0
father	0	3
person	0	3
girl	1	1
teenager	0	2
big man	0	2
female	0	2
adult woman	0	2
...	...	...

表 12: 図 15 の絵文字について，次元を統一した特徴ベクトル

解釈語	日本	米国
man	44	51
male	27	2
father	8	7
dad	0	15
woman	1	12
adult	2	9
mom	0	8
boy	1	3
成人男性	3	0
mother	0	3
person	0	3
adult man	0	2
...	...	...

算する過程で，辞書を用いて日本語の解釈語を英訳する．ただし，辞書にない語は日本語のままである．この翻訳後の解釈語に次元を統一したリストを，表 10, 11, 12 に示す．

表 10 を見ると、図 13 の類似度が小さかった原因が、日米間で同じ解釈語を持っていないためであることが分かる。米国の解釈語「toys」に翻訳された日本の解釈語がなかったためである。「toy」と「toys」のような語形変化のある語は、人間は同じ意味として解釈するが、システムは1文字でも異なると別の意味として解釈する。解釈語の頻度が大きいと類似度に影響するため、システムが類似度を計算する前に、語形変化や表記ゆれのある語を同じ次元として扱えるようにする必要がある。

また、表 11, 12 を見ると、それぞれの類似度が大きくなった要因が、日米両方の「woman」および「man」の頻度がかなり大きいためであることが容易に理解できる。類似度の値そのものについては、おおむね（「man」と「male」など同じ意味の語は、同じ次元にまとめる必要があるという点を除いては）正しいものの、ここで問題であるのは、日本人と正反対の解釈をする約2割の米国人間のコミュニケーションでは意思の疎通を図ることができないだけでなく、正反対の意味という大きな誤解が生じる点である。コミュニケーションにおいて相手の意図が分からないと判断した場合は、聞き返すなど相手の意図を確認することによりコミュニケーションを続けることができるが、相手の意図を誤解した場合は、話に食い違いが生じるまで誤解であることが分からない。したがって、システムには、類似度という概念以外にこれらの絵文字を検出するような工夫が必要になる。

## 第5章 おわりに

本研究では、序章で述べた問題点の解決策として次の手法を提案した。

### 1. 文化差のある絵文字の検出

絵文字の解釈は、人間がこれまでの生活環境で吸収してきた記憶に大きく依存するため、同じ文化や言語を背景に持つ者の間ではある程度はまとまっていると考えられる。そこで、本研究では、日本と米国の子供たちが、システムで使用できる120個の絵文字についてどのように解釈しているかを調査した結果を利用し、各絵文字について国ごとにまとめられたデータを1つの集合とみなし、類似概念を用いて、各絵文字について国名の異なる2つの集合間の類似度を計算する手法を提案した。システムがこの類似度を利用者に通知することにより、利用者がその情報を参照し、類似度の小さ

な絵文字，すなわち，お互いの解釈の異なる可能性が高い絵文字の使用を控えることができるようになる．

## 2. 絵文字検出システムの実装と検証

上記の提案を実現するために，日米間において文化差のある絵文字を検出するシステムを実装した．そして，このシステムによって検出された絵文字が本当に文化差のあるものかを検証し，システムの妥当性・問題点について考察した．実装した検出システムの出力のうち，妥当性が確認できたデータを用いて，絵文字メッセージを作成する際に，使用したい絵文字を検索するための絵文字検索システムにおいては，国内における絵文字解釈の差異を考慮するだけでなく，文化差も考慮した絵文字検索が可能となる．

今後の課題としては，検出の精度を向上させることが最重要である．現在のシステムは，文化差のない絵文字まで誤検出してしまうため，コミュニケーションに使うことができる絵文字が少なくなってしまう．文化差のある絵文字を検出できるという性能を失わないような解法を模索していく必要がある．また，今回利用したアンケート調査では，絵文字の解釈を解釈語という言葉表現で回答してもらう手法を採用してあった．このため，絵文字の解釈と言語表現は一対一対応ではないことがあり，絵文字と解釈語の間にズレが生じている可能性がある．今後は，このズレに関しても調査していく必要がある．

## 謝辞

本研究を行うにあたり，熱心なご指導，適切なお助言を賜りました石田亨教授に厚く御礼申し上げます．また，親身なご指導と多大なるご協力を行ってくださいました Cho Heeryon 氏には心の底より深謝いたします．そして，日頃から多くのお助言とご協力を行ってくださいました石田研究室の皆様にも心より感謝いたします．

最後に，本稿で使用したすべての絵文字の著作権は特定非営利活動法人パンゲアに帰属します．

## 参考文献

- [1] 森由美子: 子ども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] Takasaki T.: PictNet: Semantic Infrastructure for Pictogram Communication, *The 3rd International WordNet Conference (GWC-06)* pp. 279-284 (2006).
- [3] Takasaki T.: Design and Development of a Pictogram Communication System for Children around the World, *The 1st Int'l Workshop on Intercultural Collaboration (IWIC-07)* (2007).
- [4] Cho Heeryon, 稲葉利江子, 石田 亨, 高崎俊之, 森由美子: 絵文字コミュニケーションにおけるセマンティクス, 合同エージェントワークショップ&シンポジウム (*JAWS-06*) (2006).
- [5] Cho H., Inaba R., Takasaki T., Mori Y., and Ishida T.: *Culturally-Situated Pictogram Retrieval*, *The 1st Int'l Workshop on Intercultural Collaboration (IWIC-07)* (2007).
- [6] K. S. Sitaram: 異文化間コミュニケーション, 御堂岡潔訳, 東京創元社, p. 33 (1985).
- [7] Hofstede Geert: Culture's Consequences, *Sage Publications* (1984).
- [8] 朝川哲司: 文化的差異に対する認識の向上を支援する教育システムの開発, pp. 8-11 (2003).

# 付録

## A.1 ソースコード

### A.1.1 restruct\_edict.rb

```
require 'jcode'
$KCODE = 'sjis'

data=[
  [{"cntry_code"=>"jp"},
  {"昼のあいさつ"=>1, "かいわ"=>2, "お話 会話"=>1, "おやすみ"=>4, "いすに座る"=>1, "こんにちわ"=>1, "こんにち"=>6, "お泊り"=>1, "おやすみなさい"=>1, "おはよ"=>1, "おはなし"=>2, "total"=>101, "やあ・こんにちわ"=>1, "お泊り会"=>1, "おやすみのあいさつ"=>1, "朝"=>2, "会話する"=>1, "会話"=>1, "おはよう"=>60, "病院"=>1, "朝の挨拶(おはよう)"=>1, "おはよー"=>1, "おははししよう"=>1, "おしゃべり"=>1, "あさ"=>1, "あいさつ"=>2, "朝のあいさつ"=>1, "寝る前のあいさつ"=>1, "オハヨー"=>1, "おみまい"=>2, "file"=>"001.gif"},
  {"よる"=>2, "おやすみ"=>1, "夜"=>2, "宇宙"=>2, "夜の挨拶(こんばんわ)"=>1, "合言葉"=>1, "宇宙旅行"=>1, "total"=>100, "火星人"=>1, "夜遊び"=>1, "夜のあいさつ"=>3, "会話する"=>1, "会話"=>3, "夜の会話"=>1, "さようなら"=>1, "ほしをみよう"=>1, "ばいばーい"=>1, "今晚は"=>1, "こんばんわ"=>24, "こんばんは"=>49, "おしゃべり"=>1, "コン"=>1, "file"=>"002.gif", "やあ・こんばんは"=>1},
  {"眠るところ"=>1, "よる"=>1, "ねむり"=>1, "ねごと"=>4, "ぐーぐーぐー"=>1, "おやすみ zzzz"=>1, "おやすみ"=>44, "夜"=>1, "入院してる子供"=>1, "寝言"=>5, "寝る"=>3, "もう寝ちゃった"=>1, "ねむる"=>1, "ねむっている"=>1, "ねます"=>1, "おやすみなさい"=>16, "寝ている"=>1, "total"=>101, "寝言・おやすみなさい"=>1, "家の夜"=>1, "夢を見ている"=>1, "ゆめ見ている"=>1, "ねむい"=>1, "睡眠"=>2, "寝るときの挨拶(おやすみなさい)"=>1, "ゆめ"=>1, "ねる"=>1, "file"=>"003.gif", "夢"=>5, "あいうえお"=>1, "zzz(寝ている)"=>1},

  (ハッシュデータが長いため中略)

  {"かぜ"=>1, "揺れ"=>1, "前後"=>1, "total"=>92, "揺れる"=>2, "そよそよ"=>1, "公園"=>1, "ブランコが動く"=>1, "ブランコ"=>59, "風"=>4, "韓国製"=>2, "ゆらゆら"=>1, "file"=>"118.gif", "ゆれる"=>4, "ゆれ"=>1, "ぶらんこ"=>13},
  {"ボウリング"=>3, "ドッジボール"=>1, "転がるボール"=>2, "よっこいしょ"=>1, "転がり"=>1, "回転運動"=>1, "転がる"=>17, "坂"=>7, "バスケ"=>1, "ぼーる"=>2, "total"=>83, "回る"=>1, "バスケットボール"=>3, "地球の自転"=>1, "上下"=>1, "坂道"=>1, "ボールの速さ"=>1, "まわる"=>1, "さかみち"=>1, "ころころ"=>2, "ころがるボール"=>1, "転がりでの"=>1, "さか"=>1, "ころがり"=>1, "ぐるぐる"=>1, "発想の転換"=>1, "転がす"=>1, "転がしている"=>1, "回転"=>2, "コロコロ"=>1, "ころがる"=>8, "あっちからこっちから"=>1, "衝突"=>1, "ボール"=>5, "ぼうるがおちる"=>1, "ぶつかる前"=>1, "落ちる"=>1, "file"=>"119.gif", "不明"=>1, "ボーリング"=>3, "バスケット"=>1, "わからん"=>1},
  {"合奏"=>2, "カラオケ"=>1, "うた"=>6, "ラララ"=>1, "コミュニケーション"=>1, "伴奏つきの歌"=>1, "歌声"=>1, "オペラ"=>1, "うたう"=>2, "演奏"=>4, "total"=>90, "独唱会"=>1, "歌を歌う"=>3, "音楽を楽しむ"=>1, "歌う"=>10, "発表会"=>3, "歌を歌っている"=>1, "コンサート(歌)"=>1, "歌"=>11, "音楽隊"=>1, "コンサート"=>1, "えんそう"=>1, "おんがく"=>3, "合唱"=>2, "音楽会"=>1, "音楽"=>25, "ハーモニー"=>1, "file"=>"120.gif", "えんそうかい"=>1, "うたをうたう"=>1, "karaokeの海外普及"=>2}
  ],
  [{"cntry_code"=>"us"},
  {"happy"=>1, "goodmorning"=>2, "friendly"=>1, "day and night"=>1, "talk in your room"=>1, "hi"=>5, "games"=>1, "woke up"=>1, "play"=>3, "mom waking child up"=>1, "going to bed"=>1, "talking early"=>1, "friends"=>6, "total"=>117, "talking"=>4, "sleepover"=>1, "morning"=>8, "making friends"=>1, "friendship"=>1, "communication"=>1, "bubbles"=>1, "waking"=>1, "wake up"=>1, "talk doing the day"=>1, "kids playing"=>1, "good morning"=>39, "bed time chat"=>1, "saying/good night"=>1, "day"=>2, "talking during day"=>1, "sleep over"=>2, "sdfg"=>1, "kids talking"=>1, "hello"=>7, "greeting"=>1, "conversation"=>2, "people saying hello"=>1, "mother waking child"=>1, "daytime friends"=>1,
```

```

"chat at home"=>1, "spam"=>1, "saying hello"=>1, "people talking"=>1, "goodnight"=>1,
"good mornng"=>1, "file"=>"001.gif", "sharing"=>1, "good night"=>3},
{"saying goodnight"=>1, "nighttime talk"=>1, "giveing high fives"=>1, "talk"=>2,
"people saying goodbye"=>1, "glad"=>1, "by"=>1, "talk at night"=>1, "surprise"=>1, "play"=>1,
"night friends"=>1, "night"=>10, "talking at night"=>4, "people saying hi"=>1, "friends"=>1,
"earth"=>1, "date"=>1, "total"=>117, "talking"=>5, "sdfgh"=>1, "science"=>1, "good bye"=>1,
"friendship"=>1, "friends talking"=>2, "friends at night"=>1, "speaking"=>1, "good morning"=>1,
"good evening"=>9, "dancing"=>1, "camping"=>1, "i like you"=>1, "waving in the street"=>1,
"two people talking"=>1, "speak"=>1, "on a date"=>1, "nighttime friend"=>1, "kids talking"=>1,
"hello"=>3, "greeting"=>1, "conversation"=>4, "saying goodbye"=>1, "goodniht"=>1,
"chat at night"=>1, "talking late"=>1, "spam"=>1, "goodnight"=>4, "don't know"=>1,
"file"=>"002.gif", "party"=>1, "night walk"=>1, "goodbye"=>2, "good night"=>33},
{"zzzz"=>2, "snore"=>2, "sleeping. zzzz"=>1, "person sleeping"=>1, "i'm tired"=>1, "sleep"=>25,
"bed time"=>1, "sleepy"=>1, "dreaming"=>14, "have a great dream"=>1, "dream"=>6, "total"=>116,
"zzz"=>1, "snoring"=>2, "sleeping"=>32, "child sleeping"=>1, "asleep"=>1, "a dream"=>1,
"camping"=>1, "tired"=>2, "someone sleeping"=>1, "i slept"=>1, "bedtime"=>6, "zzzzz"=>1,
"passed out"=>1, "i am sleeping"=>1, "spam"=>1, "goodnight"=>1, "file"=>"003.gif",
"good night"=>7},

```

(ハッシュデータが長いため中略)

```

{"swing set"=>5, "motion"=>1, "jj"=>1, "wind"=>5, "swing-set"=>1, "pump"=>1, "play"=>1,
"nuclear winter"=>1, "how to swing"=>1, "swings"=>8, "park"=>2, "total"=>132, "swingset"=>7,
"swinging"=>14, "swing time"=>1, "swinging/movingswing"=>1, "playground"=>4, "back and forth"=>3,
"play yard"=>1, "to swing"=>1, "the swing"=>1, "swing"=>68, "fun"=>1, "file"=>"118.gif",
"windy"=>2, "to and fro"=>1},
{"roll"=>16, "ramp"=>4, "pumpkins"=>1, "physics"=>1, "motion"=>1, "jj"=>1, "bowling"=>4,
"rolling uphill"=>1, "roal"=>1, "holder"=>1, "hill"=>1, "games"=>1, "courts"=>1, "baasketbaal"=>1,
"uphiill"=>1, "to roll"=>1, "depth perception"=>1, "basetball"=>1, "balls going down"=>1,
"balls"=>3, "ball"=>4, "rolling balls"=>1, "rolling a ball"=>1, "rollin up"=>1, "basket ball"=>1,
"balls rolling up and"=>1, "ball rolling up and"=>1, "total"=>130, "rolling basketball"=>2,
"hoops"=>1, "fulcrum"=>1, "up and down"=>3, "up"=>1, "rolling"=>15, "bb practice"=>1,
"basketball"=>18, "sports defy physics"=>1, "rollball"=>1, "playing ball"=>2, "ok"=>1,
"how to roll ball"=>1, "dibble a basketball"=>1, "circling"=>1, "bowling ball place"=>1,
"steep hill"=>1, "incline"=>3, "free throw"=>1, "downhill uphill"=>1, "balls collide"=>1,
"b-ball race"=>1, "round and round"=>1, "rolling ball"=>4, "roll up"=>1, "ball rolling"=>1,
"sports"=>1, "sport"=>1, "rotate"=>1, "over the frnce"=>1, "over"=>1, "collision course"=>1,
"bouncing"=>2, "file"=>"119.gif", "uphill"=>1, "pass"=>1, "ball rolling up hill"=>1},
{"roll"=>1, "jj"=>1, "singers"=>1, "perform"=>1, "people singing"=>1, "musical family"=>1,
"audition"=>1, "band"=>4, "total"=>131, "the talented"=>1, "sing"=>18, "music arts"=>1,
"in a band"=>1, "a boy singing"=>1, "sining and intuments"=>1, "singing solo"=>1,
"performance"=>1, "music"=>34, "a child singing"=>1, "singer"=>1, "musci"=>1,
"american idol"=>1, "singing performance"=>1, "music class"=>1, "chior"=>2, "sound"=>1,
"singing"=>38, "person singing"=>1, "choir"=>4, "someone singing"=>1, "solo"=>1,
"sing/music"=>1, "arts"=>1, "file"=>"120.gif", "song"=>1, "sining"=>1, "concert"=>2, "chorus"=>1}
]
]

```

```

def out_dict(file, dict)
  f = open(file, "w")
  dict.each {|k, v|

```

```

    f.puts k+, "+v
  }
  f.close
end

##### M A I N #####

#edict の jp 各単語のうち解釈語に含まれているものを抜き出し, jp-us の対応をつける

dict_jpus = {}
File.foreach("edict.dat") {|line|
  word_with_rubi, word_us = line.split(/\s*\s*/)

  #jp 単語の抽出
  if(/.*\.[.]\./ =~ word_with_rubi)
    word_jp, rubi = word_with_rubi.split(/\s*\[|\]/)
  else
    word_jp = word_with_rubi
  end

  #us 単語の抽出
  word_us.gsub!(/ *\(.*) */, "")
  if(word_us == "") #空になる場合の処理 (eg こんにちは)
    next
  end
  word_us.downcase!

  #word_jp の処理
  #重複の処理
  unless(dict_jpus.has_key?(word_jp))
    data[0].each {|picton|
      #新たな単純辞書ファイルへ書き込む
      if(picton.has_key?(word_jp))
        dict_jpus[word_jp] = word_us
        break
      end
    }
  end

  #rubi の処理
  #重複の処理
  unless(dict_jpus.has_key?(rubi))
    data[0].each {|picton|
      #新たな単純辞書ファイルへ書き込む
      if(picton.has_key?(rubi))
        dict_jpus[rubi] = word_us
        break
      end
    }
  end
end
}

```

```
# p dict_jpus
out_dict("edict_for_picton.dat", dict_jpus)
```

## A.1.2 comparison.rb

```
require 'jcode'
$KCODE = 'sjis'
require 'mathn'
include Math
```

```
ata=[
  [{"cntry_code"=>"jp"},
  {"昼のあいさつ"=>1, "かいわ"=>2, "お話 会話"=>1, "おやすみ"=>4, "いすに座る"=>1, "こんにちわ"=>1, "こんにちは"=>6, "お泊り"=>1, "おやすみなさい"=>1, "おはよ"=>1, "おはなし"=>2, "total"=>101, "やあ・こんにちわ"=>1, "お泊り会"=>1, "おやすみのあいさつ"=>1, "朝"=>2, "会話する"=>1, "会話"=>1, "おはよう"=>60, "病院"=>1, "朝の挨拶(おはよう)"=>1, "おはよー"=>1, "おははししよう"=>1, "おしゃべり"=>1, "あさ"=>1, "あいさつ"=>2, "朝のあいさつ"=>1, "寝る前のあいさつ"=>1, "オハヨー"=>1, "おみまい"=>2, "file"=>"001.gif"},
```

(ハッシュデータが長いため中略)

```
{"合奏"=>2, "カラオケ"=>1, "うた"=>6, "ラララ~"=>1, "コミュニケーション"=>1, "伴奏つきの歌"=>1, "歌声"=>1, "オペラ"=>1, "うたう"=>2, "演奏"=>4, "total"=>90, "独唱会"=>1, "歌を歌う"=>3, "音楽を楽しむ"=>1, "歌う"=>10, "発表会"=>3, "歌を歌っている"=>1, "コンサート(歌)"=>1, "歌"=>11, "音楽隊"=>1, "コンサート"=>1, "えんそう"=>1, "おんがく"=>3, "合唱"=>2, "音楽会"=>1, "音楽"=>25, "ハーモニー"=>1, "file"=>"120.gif", "えんそうかい"=>1, "うたをうたう"=>1, "karaokeの海外普及"=>2}
```

```
],
```

```
[{"cntry_code"=>"us"},
{"happy"=>1, "goodmorning"=>2, "friendly"=>1, "day and night"=>1, "talk in your room"=>1, "hi"=>5, "games"=>1, "woke up"=>1, "play"=>3, "mom waking child up"=>1, "going to bed"=>1, "talking early"=>1, "friends"=>6, "total"=>117, "talking"=>4, "sleepover"=>1, "morning"=>8, "making friends"=>1, "friendship"=>1, "communication"=>1, "buubbles"=>1, "waking"=>1, "wake up"=>1, "talk doing the day"=>1, "kids playing"=>1, "good morning"=>39, "bed time chat"=>1, "saying/good night"=>1, "day"=>2, "talking during day"=>1, "sleep over"=>2, "sdfg"=>1, "kids talking"=>1, "hello"=>7, "greeting"=>1, "conversation"=>2, "people saying hello"=>1, "mother waking child"=>1, "daytime friends"=>1, "chat at home"=>1, "spam"=>1, "saying hello"=>1, "people talking"=>1, "goodnight"=>1, "good molrning"=>1, "file"=>"001.gif", "sharing"=>1, "good night"=>3},
```

(ハッシュデータが長いため中略)

```
{"roll"=>1, "jj"=>1, "singers"=>1, "perform"=>1, "people singing"=>1, "musical family"=>1, "audition"=>1, "band"=>4, "total"=>131, "the talented"=>1, "sing"=>18, "music arts"=>1, "in a band"=>1, "a boy singing"=>1, "sining and intuments"=>1, "singing solo"=>1, "performance"=>1, "music"=>34, "a child singing"=>1, "singer"=>1, "musci"=>1, "american idol"=>1, "singing performance"=>1, "music class"=>1, "chior"=>2, "sound"=>1, "singing"=>38, "person singing"=>1, "choir"=>4, "someone singing"=>1, "solo"=>1, "sing/music"=>1, "arts"=>1, "file"=>"120.gif", "song"=>1, "sining"=>1, "concert"=>2, "chorus"=>1}
```

```
]
```

```
]
```

#一方の解釈語(ハッシュ)を他国言語に翻訳する(今はjp->us)

```
def translate(picton)
  #辞書ファイル edict_for_picton.dat からハッシュを作る
```

```

jp2us = {}
File.foreach("edict_for_picton.dat") {|line|
  jp, us = line.chop.split(/\s*,\s*/)
  jp2us[jp] = us
}
# p jp2us

#辞書にある解釈語は英訳，ないものはそのまま
picton_translated = {}
picton_translated.default = 0
picton.each {|k, v|
  if(jp2us.has_key?(k))
    picton_translated[jp2us[k]] += v
  else
    picton_translated[k] = v
  end
}
return picton_translated
end

#2 つのベクトルからコサイン類似度を求める
def cal_cossim(vector1, vector2)
  cossim = 0
  length1 = length2 = 0
  (0..vector1.size).each {|i|
    cossim += vector1[i]*vector2[i]
    length1 += vector1[i]*vector1[i]
    length2 += vector2[i]*vector2[i]
  }
  return cossim/(sqrt(length1)*sqrt(length2))
end

#2 ハッシュを比較して *文化類似度を求める
def cal_culsim(picton1, picton2)
  #頻度を同じ key 順に並べ替え，配列にする．
  freq_ary1 = []
  freq_ary2 = []
  keys = (picton1.keys | picton2.keys) - ["file", "total"]
  # p keys
  keys.each {|key|
    if(picton1[key] == nil)
      freq_ary1 << 0
    else
      freq_ary1 << picton1[key]
    end
    if(picton2[key] == nil)
      freq_ary2 << 0
    else
      freq_ary2 << picton2[key]
    end
  }
  # p keys
  # p freq_ary1

```

```

# p freq_ary2
return cal_cossim(freq_ary1, freq_ary2)
end

#culsim をファイルに書き込む
def out_culsim(file, culsim)
  f = File.open(file, 'w')
  f.puts culsim
  f.close
end

##### M A I N #####

culsim = []
data_transposed = data.transpose
data_transposed[1..data_transposed.size].each {|picton|

  #日本語の解釈語を英語に翻訳
  picton_translated = translate(picton[0])
  # p picton[1]
  # p picton_translated

  #*文化類似度を返す
  culsim << cal_culsim(picton[1], picton_translated)
  # p culsim
}

#類似度をファイルに書き込む
out_culsim("culsim.dat", culsim)

```

## A.2 解釈語リスト

### A.2.1 図 10 の絵文字の解釈語リスト

解釈語 (日本)	頻度	解釈語 (米国)	頻度
おはよう	60	good morning	39
こんにちは	6	morning	8
おやすみ	4	hello	7
あいさつ	2	friends	6
おはなし	2	hi	5
おみまい	2	talking	4
かわい	2	good night	3
朝	2	play	3
あさ	1	conversation	2
いすに座る	1	day	2
おしゃべり	1	goodmorning	2
おはししよう	1	sleep over	2
おはよ	1	bed time chat	1
おはよー	1	buubles	1
おやすみなさい	1	chat at home	1
おやすみのあいさつ	1	communication	1
お泊り	1	day and night	1
お泊り会	1	daytime friends	1
お話 会話	1	friendly	1
こんにちわ	1	friendship	1
やあ, こんにちは	1	games	1
オハヨー	1	going to bed	1
会話	1	good molrning	1
会話する	1	goodnight	1
寝る前のあいさつ	1	greeting	1
昼のあいさつ	1	happy	1
朝のあいさつ	1	kids playing	1
朝の挨拶 (おはよう)	1	kids talking	1
病院	1	making friends	1
		mom waking child up	1
		mother waking child	1
		people saying hello	1
		people talking	1
		saying hello	1
		saying, good night	1
		sdfg	1
		sharing	1
		sleepover	1
		spam	1
		talk doing the day	1
		talk in your room	1
		talking during day	1
		talking early	1
		wake up	1
		waking	1
		woke up	1
合計	101	合計	117

## A.2.2 図 11 の絵文字の解釈語リスト

解釈語 (日本)	頻度	解釈語 (米国)	頻度
ok	19	exercise	22
まる	10	jump rope	6
正解	9	happy	5
オッケー	8	exercising	4
いいよ	6	yoga	4
はい	5	circle	3
せいかい	4	dance	3
おっけー	2	stretch	3
丸	2	stretching	3
まる	1	all	2
・ok	1	around	2
ok	1	exercising	2
o.k	1	jumping jacks	2
ok/いいですよ	1	together	2
ok です	1	aerobics	1
あいうえお	1	all about me	1
あたり	1	angel	1
あってる	1	awake	1
いいですよ	1	ballerina	1
オウケイ	1	bally	1
オーケー	1	cheerful	1
おーけー	1	circlar	1
オーライ	1	complete	1
げんきだよ	1	connect	1
ごうかく	1	content	1
すばらしいいいよ	1	everyone up	1
とってもいいね	1	exercise	1
なわとび	1	exercising	1
まゝ	1	exercising	1
マル	1	exercize	1
まる ok yes	1	hello	1
まる.ok	1	help	1
まる・ok	1	hulohoop	1
マル・オッケー	1	i don't know	1
みんなのわ	1	joy	1
よい	1	jumping jax	1
よくできました	1	jumpng jacks	1
りょうかい	1	jumprope	1
許容	1	jump-rope	1
合ってる	1	let's exercise	1
世界に広げよう・友達の輪	1	love	1
世界は一つ	1	making sun	1
大丈夫	1	me	1
平和	1	moon	1
良い	1	over my head	1
		personal space	1
		pick me up	1
		play	1
		playing	1
		poise	1
		reach for the sky	1
		relaxed	1
		round	1
		sdfg	1
		so big	1
		someone stretching	1
		spam	1
		straight	1
		strech	1
		streching	1
		stretch/dance	1
		wake up	1
		welcome	1
		whole world	1
		wholeness	1
合計	101	合計	114

### A.2.3 図 12 の絵文字の解釈語リスト

解釈語 (日本)	頻度	解釈語 (米国)	頻度
だめ	27	mad	37
ダメ	12	angry	35
ばつ	6	no	5
バツ	4	stubborn	4
いいえ	3	anger	2
はずれ	3	frustrated	2
やだ	3	i'm angry	2
考える	3	upset	2
不正解	3	anger; cutoff	1
ng	2	belligerent	1
だめですよ	2	big meanie	1
駄目	2	bored	1
no	1	crabby	1
x ばつ	1	dissapointed	1
x, 駄目	1	don't talk to me	1
x. ダメ	1	don't touch me	1
あ	1	emotion pout	1
いけません	1	grumpy	1
いやだ	1	i refuse	1
うーん	1	i'm mad	1
ううむ...	1	im angry	1
おこる	1	mad kid	1
かんがえちゆう	1	mad person	1
こまりましたねー	1	math	1
してはだめ	1	no way	1
そうだな	1	play	1
それはできません	1	pouting	1
だめえ・	1	pouty	1
だめです	1	puzzled	1
ちがう	1	refusal	1
はずれている	1	selfish	1
ばつ no	1	someone whos mad	1
ばつ, だめ	1	spam	1
ばつ. だめ	1	stand off	1
ばつ・ダメ	1	stuborn	1
まちがってる	1	vy angrey	1
やっちゃだめ	1		
ダメ, 通行禁止	1		
ブ, プー, だめー	1		
違う	1		
禁止	1		
残念	1		
地球の危機	1		
合計	101	合計	117

#### A.2.4 図 13 の絵文字の解釈語リスト

解釈語 (日本)	頻度	解釈語 (米国)	頻度
おもちゃ	69	toys	100
玩具	5	baby toys	2
こども	1	playing	2
トイ	1	babies toys	1
子供のおもちゃ	1	baby	1
小さい子	1	bn	1
乳児の遊び	1	childish toys	1
夢	1	children	1
幼児玩具	1	children toys	1
		childrens toys	1
		day care	1
		dfgfdg	1
		fun	1
		i like toys	1
		la la land	1
		learning	1
		little kids toys	1
		my favorite toys	1
		plau	1
		playing with trucks	1
		playtime	1
		rainbow maker	1
		rainbow truck	1
		toddler	1
		toddler toys	1
		toy land	1
		wonder toys	1
合計	81	合計	128

## A.2.5 図 14 の絵文字の解釈語リスト

解釈語 (日本)	頻度	解釈語 (米国)	頻度
女の人	28	woman	43
女性	27	man	17
女	10	mom	13
おかあさん	5	dad	11
お母さん	3	adult	8
成人女性	3	mother	7
おとな	2	women	4
おんな	2	father	3
おんなのひと	2	person	3
人	2	adult woman	2
おとな (おんな)	1	big man	2
おとなのおんな	1	female	2
おとなのおんなのひと	1	teenager	2
おねえさん	1	18 year old	1
おほほ	1	a giant man	1
おんなのひと (おとな)	1	adolt girl	1
かあさん	1	adult (man)	1
デカ歩いてる	1	adult dad	1
女のひと	1	angry parent	1
女の子	1	aunt	1
女の大人	1	ayanna	1
女子トイレ	1	bastard	1
大人	1	big	1
大人 (女)	1	big boy	1
大人の女性	1	boy	1
男子	1	concerned person	1
		evil man	1
		fat man	1
		girl	1
		grown man	1
		grown up	1
		grownup	1
		hgkh	1
		kassan	1
		mad man	1
		male	1
		mama	1
		man red	1
		mum	1
		older adult	1
		oldmen	1
		parent	1
		person winking	1
		red grown up	1
		sad	1
		spam	1
		strong	1
		tall man	1
合計	100	合計	152

## A.2.6 図 15 の絵文字の解釈語リスト

解釈語 (日本)	頻度	解釈語 (米国)	頻度
男の人	28	man	51
男性	27	dad	15
男	10	woman	12
おとうさん	4	adult	9
お父さん	4	mom	8
成人男性	3	father	7
おとこ	2	boy	3
おとこのひと	2	mother	3
おとな	2	person	3
人	2	adult man	2
えっへん	1	male	2
おとこのひと (おとな	1	a relaxed man	1
おとな (おとこ)	1	adolt boy	1
おとなのおとこ	1	adult (man)	1
おとなのおとこの人	1	adult mom	1
おにいさん	1	annoyed person	1
とうさん	1	big girl	1
デカ歩いてる 2	1	big sad man	1
兄貴	1	big woman	1
女子	1	bitch	1
大人 (男)	1	blue	1
大人の男	1	blue grown up	1
大人の男性	1	bored	1
男のひと	1	daddy	1
男の子	1	dat	1
男の大人	1	female	1
男子トイレ	1	good man	1
		grown up	1
		grownup	1
		happy man	1
		ljhlj	1
		mama	1
		man blue	1
		men	1
		mens	1
		myanna	1
		oldwomen	1
		parent	1
		person winking	1
		sad parent	1
		skinny man	1
		small	1
		spam	1
		tall man	1
		teenager	1
		uncle	1
		women	1
		younger man	1
合計	101	合計	152

### A.3 日米の次元を統一した特徴ベクトル

#### A.3.1 図 13 の次元を統一した特徴ベクトル

解釈語	頻度 (日本)	頻度 (米国)
toys	0	100
toy	74	0
baby toys	0	2
playing	0	2
dream	1	0
乳児の遊び	1	0
トイ	1	0
幼児玩具	1	0
子供のおもちゃ	1	0
child	1	0
小さい子	1	0
children toys	0	1
la la land	0	1
babies toys	0	1
bn	0	1
childish toys	0	1
childrens toys	0	1
playtime	0	1
toddler	0	1
toddler toys	0	1
rainbow truck	0	1
baby	0	1
wonder toys	0	1
playing with trucks	0	1
toy land	0	1
i like toys	0	1
dfgfdg	0	1
learning	0	1
little kids toys	0	1
rainbow maker	0	1
children	0	1
fun	0	1
my favorite toys	0	1
plau	0	1
day care	0	1

### A.3.2 図14の次元を統一した特徴ベクトル

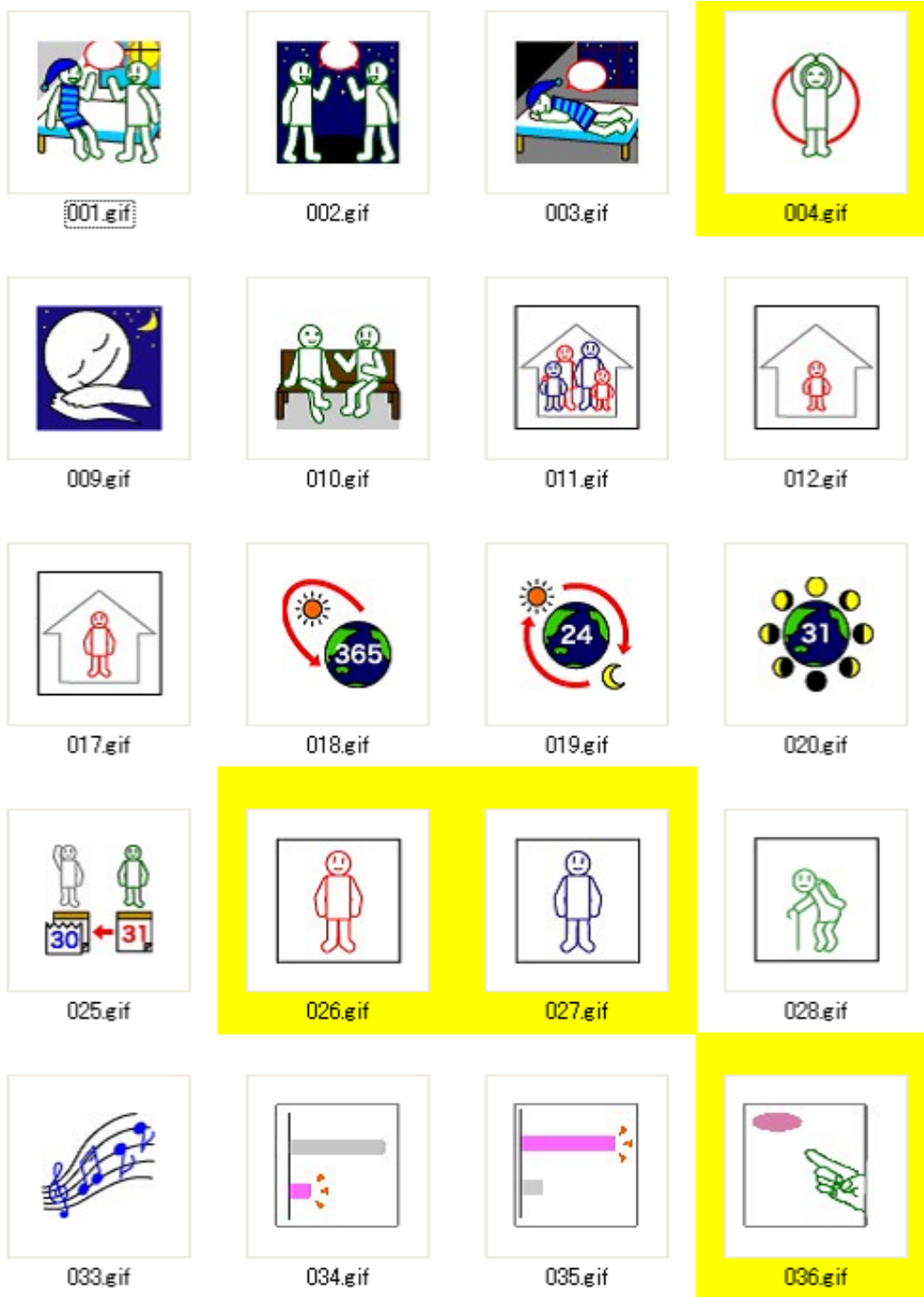
解釈語	頻度(日本)	頻度(米国)
woman	69	43
man	2	17
mother	9	7
mom	0	13
adult	3	8
dad	0	11
women	0	4
成人女性	3	0
father	0	3
person	0	3
girl	1	1
teenager	0	2
big man	0	2
female	0	2
adult woman	0	2
youth	1	0
大人(女)	1	0
女の大人	1	0
おほほ	1	0
女のひと	1	0
女子トイレ	1	0
おとなのおんなのひと	1	0
大人の女性	1	0
デカ歩いてる	1	0
おんなのひと(おとな)	1	0
older sister	1	0
おとなのおんな	1	0
おとな(おんな)	1	0
a giant man	0	1
big	0	1
concerned person	0	1
grown up	0	1
grownup	0	1
sad	0	1
big boy	0	1
boy	0	1
fat man	0	1
mama	0	1
angry parent	0	1
ayanna	0	1
grown man	0	1
kassan	0	1
older adult	0	1
man red	0	1
red grown up	0	1
adult dad	0	1
mum	0	1
strong	0	1
hgkh	0	1
male	0	1
18 year old	0	1
evil man	0	1
tall man	0	1
adolt girl	0	1
aunt	0	1
bastard	0	1
oldmen	0	1
person winking	0	1
spam	0	1
adult (man)	0	1
mad man	0	1
parent	0	1

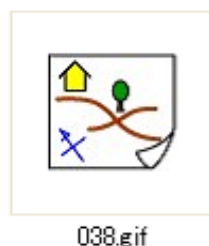
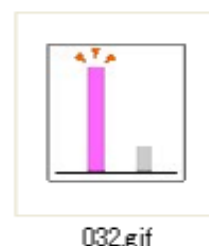
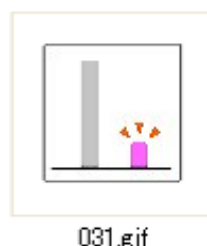
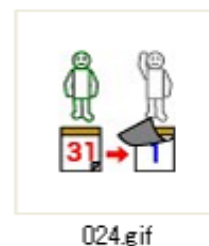
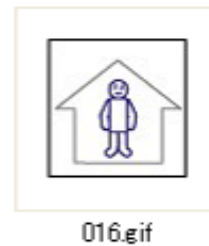
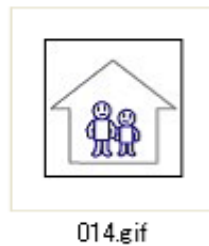
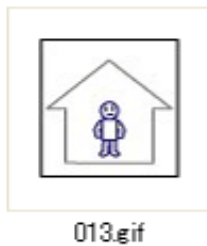
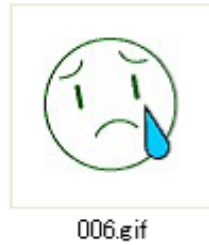
### A.3.3 図 15 の次元を統一した特徴ベクトル

解釈語	頻度 (日本)	頻度 (米国)
man	44	51
male	27	2
father	8	7
dad	0	15
woman	1	12
adult	2	9
mom	0	8
boy	1	3
成人男性	3	0
mother	0	3
person	0	3
adult man	0	2
男の大人	1	0
大人の男	1	0
one's senior	1	0
男のひと	1	0
大人の男性	1	0
デカ歩いてる 2	1	0
おとなのおとこの人	1	0
おとな (おとこ)	1	0
おとこのひと (おとな	1	0
bankruptcy	1	0
大人 (男)	1	0
男子トイレ	1	0
おとなのおとこ	1	0
older brother	1	0
えっへん	1	0
annoyed person	0	1
grown up	0	1
grownup	0	1
men	0	1
small	0	1
dat	0	1
mama	0	1
myanna	0	1
women	0	1
skinny man	0	1
teenager	0	1
uncle	0	1
adolt boy	0	1
blue	0	1
bored	0	1
happy man	0	1
man blue	0	1
big sad man	0	1
female	0	1
good man	0	1
oldwomen	0	1
big woman	0	1
blue grown up	0	1
daddy	0	1
bitch	0	1
mens	0	1
tall man	0	1
ljhlj	0	1
person winking	0	1
adult mom	0	1
sad parent	0	1
spam	0	1
younger man	0	1
a relaxed man	0	1
adult (man)	0	1
big girl	0	1
parent	0	1

#### A.4 システムで使用された絵文字

システムで使用した 120 個の絵文字を以降のページに掲載する。120 個すべての絵文字の著作権は、特定非営利活動法人パンゲアに帰属する。







041.gif



042.gif



043.gif



044.gif



049.gif



050.gif



051.gif



052.gif



057.gif



058.gif



059.gif



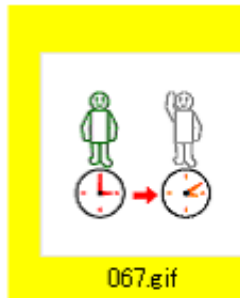
060.gif



065.gif



066.gif



067.gif



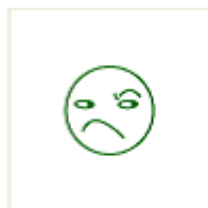
068.gif



073.gif



074.gif



075.gif



076.gif



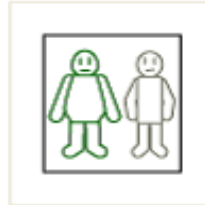
045.gif



046.gif



047.gif



048.gif



053.gif



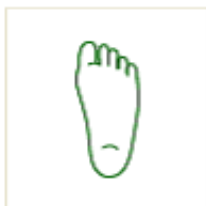
054.gif



055.gif



056.gif



061.gif



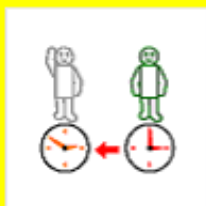
062.gif



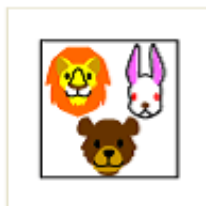
063.gif



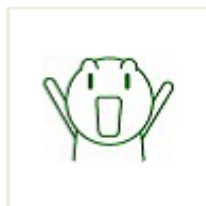
064.gif



069.gif



070.gif



071.gif



072.gif



077.gif



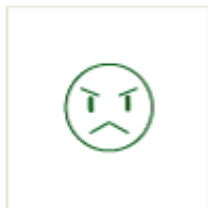
078.gif



079.gif



080.gif



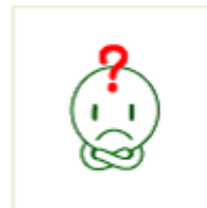
081.gif



082.gif



083.gif



084.gif



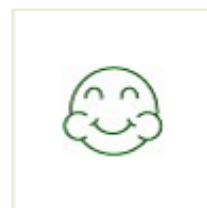
089.gif



090.gif



091.gif



092.gif



097.gif



098.gif



099.gif



100.gif



105.gif



106.gif



107.gif



108.gif



113.gif



114.gif



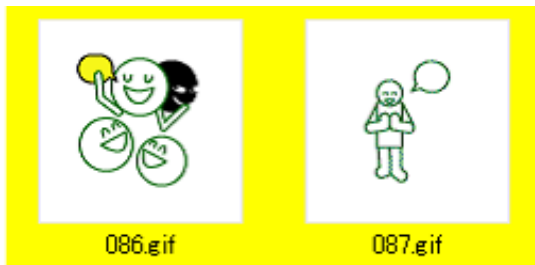
115.gif



116.gif



085.gif



086.gif



087.gif



088.gif



093.gif



094.gif



095.gif



096.gif



101.gif



102.gif



103.gif



104.gif



109.gif



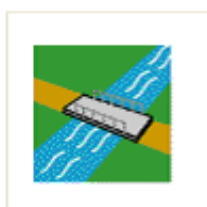
110.gif



111.gif



112.gif



117.gif



118.gif



119.gif



120.gif