

携帯メールにおける絵文字の登場位置

山崎由佳（京都大学大学院人間・環境学研究科）*

The Positions of Emojis in Mobile E-mails

Yuka Yamazaki

(Graduate School of Human and Environmental Studies, Kyoto University)

要旨

本研究は「加藤安彦ケータイメールコーパス」における携帯メール上の絵文字の登場位置と使用傾向について報告する。本コーパスを用いた絵文字の研究には田中・林 (2021), 三宅 (2022) などがあるが、本研究は異なる観点を導入するものである。まず傾向を分析するために絵文字の分類を行う。その上で第一に、絵文字の位置に着目する。各位置の登場割合を示し、頻度としては低い文頭部での絵文字の使用傾向が全体の傾向とは異なることを指摘する。第二に、文末部での絵文字の連続使用に着目する。ここでは絵文字の連続ペアを取り上げ、各カテゴリの末尾側（右側）・手前側（左側）への「来やすさ」を分析するとともに、カテゴリ間の結びつきの強さについて論じる。さらに、送受信での「携帯会社名」が異なる場合のメール件数の経年変化や、年別のカテゴリの割合等のデータを示して分析し、先行研究の議論を補う。

1. はじめに

本稿で分析の対象とする「加藤安彦携帯メールコーパス」（英語名 Kato Yasuhiko Mobile-mail Corpus）は、加藤安彦氏が 2004 年度～2010 年度の専修大学文学部のゼミ生のメールデータに基づいて構築したコーパスを、さらに数名の研究者が整備したものである（加藤 2007, 三宅 2022, 宮崎 2023a,b）。現在は、言語資源協会が配布可能な言語資源に含まれている⁽¹⁾。

本コーパスの対象となるメールデータは、携帯電話で送信・受信された電子メールとなる⁽²⁾。具体的なメールの送受信時期は 2001 年 3 月 1 日から 2010 年 11 月 6 日まで（三宅 2022:56）のおよそ 10 年間（2001～2010 年）にわたり、送受信者は大学生を中心としている。

構築されたコーパスにおいて、メール本文データは、一文一行の体裁をとっている⁽³⁾。さらに、絵文字・顔文字等にタグ付けが行われている。

また、本コーパスには、画像ファイルとして、絵文字を表すファイル 916 個と、「現在不

* yk.ymazaki+ct<@>gmail.com

(1) 言語資源協会 (GSK) の該当ページ：<https://www.gsk.or.jp/catalog/gsk2023-b/>

(2) 送受信端末の一方が PC であるものも含まれる（詳細は 4.2 節）。

(3) ただし「一文」は、メール原文での改行による「一行」と必ずしも一致するわけではない。コーパスの readme.pdf(宮崎 2023b) では、「一文末は「句点・絵文字・顔文字・記号・改行」が挿入されていた箇所と判断し入力されている」と述べられ、さらに脚注部に「絵文字・顔文字・記号が文中の語として捉えられる場合はそれを一文の終了部とせず、改行はしない」「ただし、一行の文章があまりにも長すぎる場合は、読点もしくは意味の区切れで改行としてよい」「句点と見なされる空白はそれを終端記号としてそこで改行とする」「上記以外の空白（行頭、読点代わりなど）はそのまま入力して、終端記号とはしない」加藤 (2007:5, 一部用語改訂).」（宮崎 2023b:5）とある。

明.gif)⁽⁴⁾が付属する。絵文字を表す画像のファイル名は、「(絵文字のコード部) 絵文字の名前_携帯会社等を示す文字列.gif」という形となっている。動画を示す「_動」が携帯会社名の略称の前に入るケースもある。この接尾に現れているのは「au」「au(web) (と (web) が全角表記のもの)」「dc」「dc (web)」「sb」である。



図1 絵文字画像とファイル名の例

ここで、本コーパスにおける絵文字の用法を調べる意義について述べる。本コーパスは、日本でスマートフォンや LINE のようなメッセージアプリが普及する前のメールアドレスを集めたものである。2024 年時点では、スマートフォンにおいても PC においても、メッセージアプリや SNS 上で、Unicode に組み込まれた絵文字 (Davis・Holbrook (編) 2023) を使うことができる。ただし、現行の Unicode 上の絵文字は、日本の携帯メールで 2010 年までに用いられていたものを継承したものである。実際、2014 年の Unicode の Tech Report である Unicode Emoji の提案ドラフト (Davis・Holbrook (編) 2014) は、“Emoji became available in 1999 on Japanese mobile phones.” と述べ、日本の携帯キャリアで用いられる絵文字を追加した経緯を記している。このため、本コーパスでの絵文字の用法を調査することは、絵文字の使用に関する通時的な議論の基礎づけにも寄与すると考えられる。

本コーパスを用いた絵文字についての先行研究に、コーパスの整備者のうち二名による論文である田中・林 (2021) がある。この研究は、絵文字について次のようなデータを解析・提示している。まず、出現した 904 種類の絵文字の累積使用率を BCCWJ 教科書コーパス小学校 (国語・数学・理科・社会／保健体育の二種類に分けている) と比較し、同じ種類のものが (やや／かなりの程度) 繰り返し使用される傾向にありそうだと論じる。さらに、年度⁽⁵⁾ごとの絵文字の使用度数と 1 行あたりの使用割合、年度別契約会社比率を提示し、高頻度絵文字についての分析を展開する。具体的には、契約会社 (docomo・au・vodafone⁽⁶⁾) ごとに、頻度上位

(4) コーパスの readme.pdf (すなわち宮崎 (2023b)) に「Unicode が入力されているが調べると不明だったものは、「現在不明」画像にリンクさせた。」とある。

(5) この「年度」は送受信年 (2001～2010 年) ではない。また範囲は 2004 年度～2010 年度である。

(6) 現 SoftBank (本稿では基本的に SoftBank と記す)。

10位の絵文字と年度別の出現度数および割合を提示し、その出現割合を契約会社横断で順位づけした総合上位10位を示す。この総合上位10位の絵文字を、「絵文字の特徴や性質という点から」(田中・林 2021:116)タイプ分類し、「感情を表すタイプの絵文字」「ダイヤルボタン等で直接入力可能な一般的な記号の絵文字」「顔の絵文字」の3タイプを提示する。その上で、タイプの使用傾向等の変化を指摘する。

また、三宅(2022)は、絵文字と顔文字や「カッコ文字((笑)など)」等を合わせて「ヴィジュアル表現」と呼び、その使用の経年変化に関する議論を行うとともに、さらに田中・林(2021)が抽出した上位10位の絵文字を「アイコン」「インデックス」「シンボル」の角度から分析している。

本研究は、別の角度からの分析を行う。第一に、絵文字の文中(行中)での位置に着目する。これは、「その位置に絵文字が来る頻度は低い」が「その位置では高頻度に見られる絵文字や使用傾向がある」という可能性が考えられるためである。本稿では絵文字の位置と登場割合を示し、その上で、比較的低頻度であった文頭部の絵文字の用法について述べる。第二に、連続する絵文字に着目する。絵文字の連続にも、同じ絵文字が反復する場合と、異なる絵文字の組み合わせとなる場合がある。このパターンを解明することは、絵文字の文法的な性質を明らかにすることに関わるであろう。本稿では文末での連続使用について分析する。また、これらの分析にあたって、傾向性を調べるため、絵文字を5つのカテゴリに分類する。

2. 絵文字の分類

本節では、本研究で行った絵文字の分類について述べる。

2.1 分類法的设计

本研究では、絵文字の使用傾向について論じるために、コーパスに付属する絵文字画像群に対しての分類法を設計した。

先行研究である田中・林(2021)は、本コーパスに高頻度に見られた絵文字を3つに分類している。それは、「感情を表すタイプの絵文字」(あせあせ、きらきら、ぴかぴか、ひらめき系)、「ダイヤルボタン等で直接入力可能な一般的な記号の絵文字」(感嘆符・疑問符・上向き矢印系)、「顔の絵文字」(「わーい(嬉しい顔)」、「うれしいカオ」)の3つである⁽⁷⁾。本研究は、より絵文字の範囲を広げた議論をするために、この3分類を拡張した分類法を設計する。このために、既存の包括的な分類を参照する。

ここで参照するのは、現行のUnicodeに組み込まれた絵文字(Davis・Holbrook(編)2023)の分類と、ケータイメールの絵文字に対する分類である。

まず、現行のUnicodeに含まれる絵文字(v15.1)に対しては、Unicodeコンソーシアムが付与した9分類(Smileys & Emotion, People & Body, Animals & Nature, Food & Drink, Travel & Places, Activities, Objects, Symbols & Signs, Flags)⁽⁸⁾がある。

(7) 全絵文字を分類したものではない。

(8) 具体的な割り当ては <https://unicode.org/emoji/charts/full-emoji-list.html> で閲覧することができる

一方で、ケータイメールの絵文字については、「au」「docomo」「SoftBank」の3つの携帯会社（キャリア）が次のような分類を示している。まず au が「絵文字メール | サービス・機能 | au」という Web ページ⁽⁹⁾で、「au / docomo 共通絵文字」に対して「顔・表情」「気持ち・からだ」「生き物」「食べ物・飲み物」「自然・季節」「ファッション・遊び」「乗り物・建物・地図」「道具」「記号・文字」という9分類を提示し、docomo も「絵文字 | メール / デコメールの作成・送受信 | ドコモメールアプリでのご利用 | ドコモメール | サービス・機能 | NTT ドコモ」というページ⁽¹⁰⁾で、「au / docomo 共通絵文字」に対しては au の分類名をさらにまとめたような分類を提示している。ただし、「共通絵文字」ではない絵文字については、扱いが異なる。docomo は少なくとも上述のページで「ドコモ絵文字」の分類は提供しておらず、au は「au 絵文字」については「生き物」ではなく「生き物・星座」という分類名を提示している⁽¹¹⁾。また SoftBank は「絵文字（ケータイ） | スマートフォン・携帯電話 | ソフトバンク」というページ⁽¹²⁾で、「他社ケータイへの絵文字変換」との記載の下に、「顔」「気持ち・装飾」「季節・イベント」「キャラクター」「食べ物」「日常」「ツール」「趣味」「文字・記号」という、au・docomo とは多少異なる分類名を示している。ここで、ケータイメールの絵文字の多くが（「ドコモ絵文字」「au 絵文字」の多くも）他キャリアとの間で変換可能になっている⁽¹³⁾ 即ち互換性があるといえることと、上記キャリアが提供する資料には他キャリアでの絵文字イメージも示されていることを補足する。以降「(キャリア名)の(分類名)」というような書き方をするが、その分類に入る絵文字には、当該キャリアの分類資料に掲載された他キャリアの絵文字も含める。

本研究では、本コーパスについての先行研究である田中・林 (2021) の分類を拡張しつつ、上記の包括的な複数の絵文字分類をつなぐような、5つの分類を設けた。

本分類と絵文字の対応は、二段階で行う。第一段階として、カテゴリ1に au の「顔・表情」と Unicode 絵文字の“Smileys & Emotion”に共通するようなもの、カテゴリ2に au の「顔・表情」「気持ち・からだ」「ファッション・遊び」と“People & Body”に共通するようなもの、カテゴリ3に au の「記号・文字」と“Symbols & Signs”に共通するようなもの、カテゴリ4に SoftBank の「気持ち・装飾」に入る⁽¹⁴⁾がここまでのカテゴリに当てはまらないようなもの、カテゴリ5にその他の絵文字を割り当てることを考える。しかし、この基準では、実態を

る。絵文字の構造と大分類についての統計は <https://unicode.org/emoji/charts/emoji-counts.html> を参照のこと。

(9) <https://www.au.com/mobile/service/emoji/>, 2024/08/10 アクセス。

(10) https://www.docomo.ne.jp/service/docomo_mail/appli/deco/pictograph/, 2024/08/10 アクセス。

(11) 「au 絵文字」の「生き物・星座」分類の資料では、「魚座」等の星座について、天文記号にあたる抽象的な画像と、au 独自のより写実的な画像が掲載されている。ただし、この星座の天文記号は「au / docomo 共通絵文字」では「記号・文字」に分類されている。

(12) <https://www.softbank.jp/mobile/service/mail/3g/pictogram/>, 2024/08/10 アクセス。

(13) 絵文字コードつきの変換表は次のような携帯会社公式以外のウェブサイトも作成している。「オートビズ」の「絵文字対応表」: https://powerbiz.jp/psm/wp-content/uploads/2010/01/emoji_code.html “<https://code.cside.com/>” の「ドコモ絵文字コード・タイトル（意味）一覧」: <http://code.cside.com/i-mode/emojimode/>

(14) 補足すると、au の「気持ち・からだ」と SoftBank の「気持ち・装飾」は、名前だけではなく内容も異なる。たとえば「爆弾」「炎」「羽の生えたお札」は au では「気持ち・からだ」に入るが、SoftBank では「気持ち・装飾」ではなく「ツール」に入る。ここでは、より狭いと思われた SoftBank の分類を使った。

捉えるのに不十分であると思われる。たとえば「ドコモ絵文字」の一部（例：「」（E6D2）iモード（枠付き）_dc」は携帯会社の分類が使えず、Unicodeにも入っていない。だが、上記基準で分類されるカテゴリ5には、生物や具体的な物体・場所・建物等を表した絵文字が多いのに対し、「」は「記号・文字」でカテゴリ3に入る「」などに性質が近いのではないかと考えられる。このため、本研究では上記を基準として各カテゴリに分類される絵文字の傾向を観察し、中心的なものを抽出し、各カテゴリの意味合いを捉え直して名付け（表1）た。このようにしてできた5つのカテゴリ「表情系」「身体系」「文字符号系」「漫符系」「物体生物系」を用い、カテゴリの意味合いも考慮して、改めて絵文字をカテゴリに割り当てた。

表1 分類名と中心的なもの

番号	分類名	中心的なもの
1	表情系	表情にバリエーションがある顔
2	身体系	体の一部、人間の全身像、 無表情な顔、職業を示す服飾品のある顔
3	文字符号系	文字・数字、 くぎり符号・つなぎ符号・矢印類・商用記号 (小学館辞典編集部 2007a)
4	漫符系	漫符 (気持ちや状態を表す)
5	物体生物系	類像的 (iconic) に表された、 (人間以外で非情動的な) 動物・植物・物体・建物・場所

各絵文字画像について、その画像と絵文字画像に付与された名称（ファイル名）を視認して、分類を行った。カテゴリとその例は表2のようになる。

表2 携帯会社名と内訳

	例
表情系	 (E410) びっくり_sb  (E72A) うれしい顔_dc  (E753) ウッシッシ_dc  と  の動画 (E663) 泣き笑い (ネコ)_動.au
身体系	 (E691) 目_dc  (E4F9) おやゆびサイン.au  (E50) 見ざる.au  (E018) サッカー_sb
文字符号系	 (E482) !_au.png  (E6E2) 1_dc  (E6F5) グッド (上向き矢印)_dc  (E6D6) 有料_dc  (E651) 魚座_dc
漫符系	 (E5B1) あせあせ (飛び散る汗)_au  (E6FA) ぴかぴか (新しい)_dc  (E6EF) ハートたち (複数ハート)_dc  (E708) ダッシュ (走り出すさま)_dc  (E6F6) るるん_dc  (E6FB) ひらめき_dc ⁽¹⁵⁾
物体生物系	 (E664) ビル_dc  (E4E0) 鳥.au  (E688) 携帯電話_dc  (E656) サッカー_dc  (E49A) 星座 (うお座)_au

これを田中・林 (2021) の 3 つの分類と対応させると、表 3 のようになる。

表 3 本分類と田中・林 (2021) の 3 分類の対応

田中・林 (2021) の高頻度絵文字の分類	本分類
「顔の絵文字」	表情系
「ダイヤルボタン等で直接入力可能な一般的な記号の絵文字」	文字符号系
「感情を表すタイプの絵文字」	漫符系
—	身体系・物体生物系

ここで、カテゴリ「漫符系」について説明する。このカテゴリは、SoftBank の「気持ち・装飾」に入る絵文字と互換性のある絵文字のうち、人間・動物の外見的なパーツ（例：手、目）を含まず、「くぎり符号の使ひ方〔句読法〕（案）」（文部省教科書局調査課国語調査室 2007b）で定められたような旧来の記号（例：感嘆符、疑問符）からできたものとは異なる絵文字から成る。具体的には、「あせあせ（飛び散る汗）_au : 」、ハート類（「ハートたち（複数ハート）_dc : 」「黄ハート_au」等）⁽¹⁶⁾、「(E 6 F A) ぴかぴか（新しい）_dc : 」「(E 6 F 6) るんるん_dc : 」「(E 6 F B) ひらめき_dc : 」「(E 7 0 8) ダッシュ（走り出すさま）_dc : 」といった絵文字である。このような絵記号は、特に漫画の中で、人物や物体の周囲に配置され、その感情等の状態を形容する記号として使われている。これらは「漫符」(comic symbols) と呼ばれる（赤井ほか (2014, 2017) を参照）ため、このような分類名とした。

2.2 留意点

上述のように、本研究での分類は、各カテゴリの意味合いを考えて行ったものとなっている。しかし、複数のカテゴリにみられる傾向性を備えている絵文字がある。ここではそのような絵文字を複数取り上げ、どのように分類したかについて述べる。

2.2.1 動物の絵文字と「表情系」

カテゴリを「意味」の観点で考えたとき、まず、表情が数パターンある動物の顔の絵文字の扱いが問題となりうる⁽¹⁷⁾。

本研究では、表情に数パターンある猫の絵文字のうちで無表情ではないものは、「顔・表情」と“Smileys & Emotion”に分類されている（Unicode の分類は図 2 参照）のを引き継いで、「表情系」に分類した。無表情の動物の顔は au の「生き物」と Unicode 絵文字の“Animals”に分類されている（図 3）のを引き継ぐように、「物体生物系」に分類した。

「顔・表情」と“Smileys & Emotion”に分類されている人間以外の動物の顔には、3 パターンのジェスチャーを伴う猿（「見ざる・聞かざる・言わざる」の「三猿」）もある。ここで、人間のジェスチャーに関する Unicode の絵文字群が、“person-gesture”という“People & Body”

(15) ハートの絵文字でも、「スベード」「ダイヤ」「クラブ」と並んでいるハートの絵文字は、docomo/au/SoftBank で「記号・文字」に分類されていることも受け、「文字符号系」に分類した。

(16) ハートの絵文字が心臓に由来するものであったとしても、黄色いハートや複数のハートは、具体的な「心臓」の表示から離れていると考えられる。

(17) 動物として捉えるか、情動を表すと捉えるか。

の下位分類に入っている（図4）ことを踏まえると、これらに揃えるように「身体系」に分類することにも合理性があるのではないかといえよう。本研究では最終的に、これまでの資源に沿うような「表情系」に分類したが、異なる分類指針も考えられる。

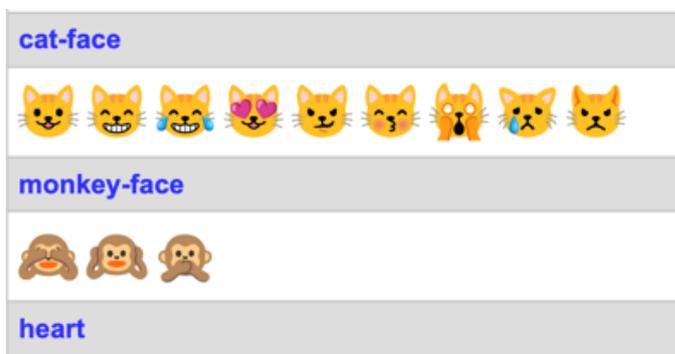


図2 猫の顔と「見ざる・聞かざる・言わざる」の三猿（Unicode）

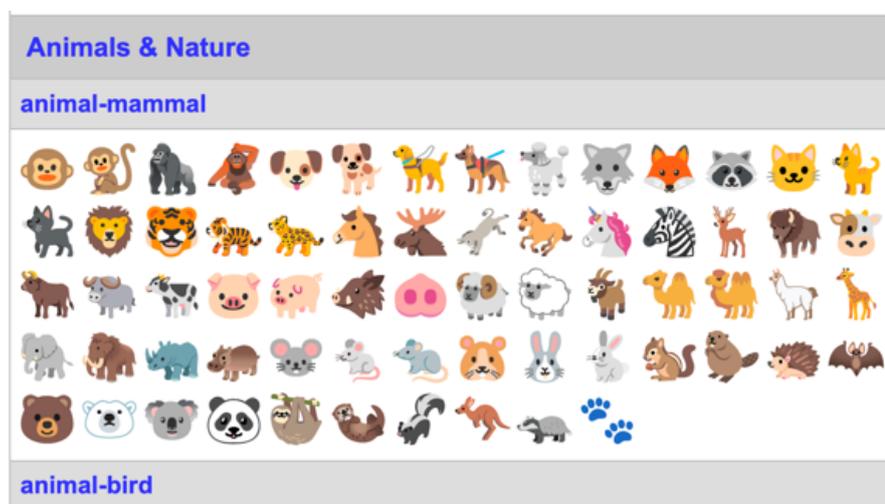


図3 動物（哺乳類）分類の絵文字（Unicode）

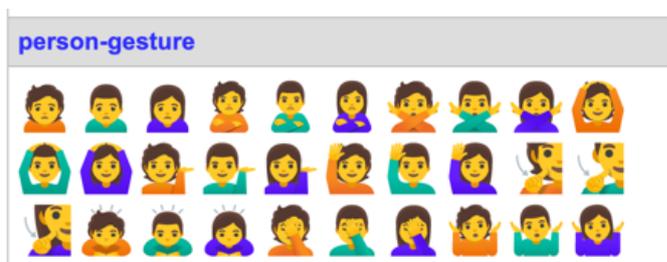


図4 person-gesture の絵文字（Unicode）

2.2.2 画像内容を見て分類を考えた例

■**同じ名前に異なるカテゴリ** 本研究では、同じ名前を持ち、キャリア間の変換機能で相互に変換される絵文字であっても、画像内容によっては別の分類を割り当てた。その例が「サッカー」の絵文字である。これについては、サッカーボールのみを描く絵文字と、人間を大きく描く絵文字がある。(au: , docomo: ) と (SoftBank: ⁽¹⁸⁾)。本研究では前者は「物体生物系」に分類し、後者は「身体系」に分類した⁽¹⁹⁾。

■**同じような画像で異なる名前に同じカテゴリ** 上の場合とは異なり、異なる名前でも、同じような画像の絵文字に対して、同じ分類を割り当てた例がある。豆電球を表す絵文字には、「 (E 6 F B) ひらめき_dc」の他に「(E 4 7 6) 電球_動_au」「(E 1 0 F) 電球_動_sb」(それぞれ3枚と6枚の静止画像でできた動画)がある。これらは皆似たような内容の静止画・動画であり、またキャリア間で変換されると考えられ、SoftBankの「気持ち・装飾」に分類されているように見えたことから、本研究ではこの3つをすべて「漫符系」に分類した⁽²⁰⁾。

3. データの整理

本節ではデータの整理方法(前処理)について述べる。

3.1 データの整理と絵文字の抽出

まず、複数の表形式(xlsx形式)のファイルで提供されている本コーパスにおいて、メールアドレスを一通ごとにまとめることを試みた。具体的には、各ファイルにおいて、「管理ID/題名/本文」列の値が「管理ID」である行を区切りとして、メールを一件ごとにまとめるとともに、“title”列と“body_combined”列を新たに設けた。「管理ID/題名/本文」列の値が再び「管理ID」になるまでの各行に対して、「管理ID/題名/本文」列の値が「題名」である場合は「本文」列の値を以て“title”列の値とし、「本文」である場合は「本文」列の値を“body_combined”列に(改行を挟んで)追加するように、処理を行った⁽²¹⁾。さらに、こうしてコーパス中の複数のファイルから得られたデータを結合し、「年度」列と「通しNo.」が重複するデータを削除した。

メールからの絵文字データの抽出は以下のように行った。コーパス中のデータには「ごめんね、<人名>は無理そうです%%絵(E 6 F 2) %%%絵(E 6 F 2) %%」のようにして、絵文字にタグが付与されている。本発表では、「%%絵(」で始まり「) %%」(ただし、%と丸括弧は半角も可とする)で終わる最短の部分抽出し、絵文字として扱うことにした。な

⁽¹⁸⁾ ただし、現在公開されているPDFでは、SoftBankもサッカーボールの絵文字を使用しているようである。https://www.SoftBank.jp/mobile/set/data/service/mail/3g/pictogram/pdf/pictogram_hobby.pdf

⁽¹⁹⁾ なお、これに対応するUnicodeの絵文字は“Activities”の下位分類“Sports Equipment”に入っている。

⁽²⁰⁾ もっとも、別の分類法として、絵文字の名前およびUnicodeで「」が「」とともに“Objects”(下位分類“Visual and Lighting Equipment”)に分類されていることを考慮して、「電球」という名前の絵文字を「物体生物系」に分類するというやりかたも考えられる。

⁽²¹⁾ ただし、「管理ID/題名/本文」列の値「管理ID」である行数と「題名」である行数は本来一致すべきだが、三宅(2022)によるとの行数に多少の異なりが見られたとのことである。ここでメールをまとめたことが、後述する絵文字個数計算においての、先行研究との絵文字数十個分の違いに関係する可能性がある。

お、この丸括弧で挟まれている部分を、本稿では「(絵文字) コード」と呼ぶ⁽²²⁾。

3.2 絵文字コードと絵文字データの対応

各絵文字コードがどのような絵文字に該当するのかを調べるため、コーパスに付属する絵文字画像ファイル群（一部は図5）のファイル名から、先頭の「(」に挟まれているコード部を元にして、（半角全角の表記ゆれに対応する形で）絵文字コードと絵文字ファイルの対応を試みた⁽²³⁾。



👋	(F 8 F 6)	手 (パー)_dc(web).gif
🚇	(F 7 E C)	地下鉄_au(web).gif
✉️	(F 6 F A)	封筒_au(web).gif
🔒	(F 6 F 5)	錠前_au(web).gif
📷	(F 6 F 0)	ビデオカメラ_au(web).gif
👁️	(F 6 B 7)	ぶた_au(web).gif
♨️	(F 6 9 5)	温泉_au(web).gif
🏀	(F 6 9 3)	野球ボール_au(web).gif
👒	(F 4 8 3)	かたつむり_au (w e b).gif
👖	(F 4 7 B)	ジーンズ_au(web).gif
💓	(F 4 7 9)	ドキドキしているハート_動_au(web).gif

図5 絵文字画像とファイル名の例

その結果、一対一の対応ができなかったケースが存在した。まず、コーパスから抽出した絵文字コードに対して、対応画像ファイルが見つからないケースが、2269個（323コード）存在した。このような例には、コードが空であるもの、コードは記載されているが絵文字画像が不明であるもの（「E 6 0 B」等）、絵文字コードを挟む「%」が1つしかないために後続する文字列がふくまれたものがある。この一方で、絵文字画像ファイル群から抽出された906個の絵文字コードについては、画像ファイルが一意ではないコード（例：「E 5 0 2」に「(E 5 0 2)アート_sb.gif」と「(E 5 0 2) テレビ_au.gif」の2つの画像ファイルが対応する）が、10コード20画像ファイル分存在した。

3.3 絵文字コードへのカテゴリ割り当て

絵文字画像群に含まれている各絵文字コードに対し、絵文字画像と名前を経由して、先述のカテゴリを割り当てた。ただし、同じコードに複数の画像が結びついた場合は、それらのカテゴリが同じであればそのカテゴリを割り当て、異なれば「_Undefined」を割り当てた。その結果、各カテゴリに該当する絵文字コードの数は、表4のようになった。また、各カテゴリ名を略記して出現割合とともに図示する（図6）。

⁽²²⁾ 加藤 (2007) 等において、当該箇所は「絵文字のユニコード」と呼ばれているが、現行の Unicode に組み込まれた「絵文字」(参考：<https://www.unicode.org/emoji/techindex.html>)の「コードポイント」との混同を避けるため、本稿では「コード」と呼ぶ。

⁽²³⁾ コーパスの表形式データには、絵文字画像へのリンクを出現順に与える列があるが、今回は処理の都合上、これらの列は参照していない。

表4 各カテゴリの該当コード数

分類名	表情系	身体系	文字符号系	漫符系	物体生物系	_Undefined
該当コード数	97	65	185	54	499	6

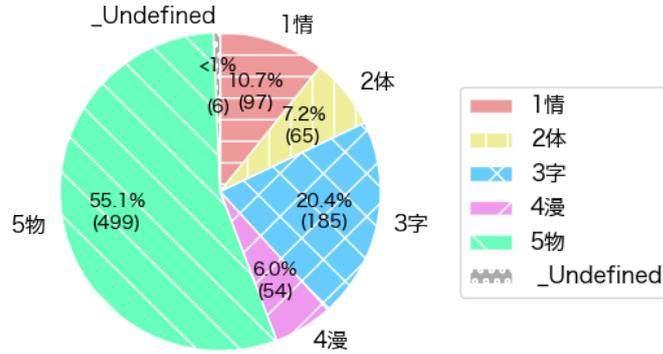


図6 各カテゴリの該当コード数と割合

4. データの概観

本節では、得られたデータの構成、携帯会社名が同じ場合と異なる場合の絵文字使用の変化、絵文字の位置タイプとその割合、本文文末部での絵文字連続使用傾向の変化、絵文字のカテゴリの変化について述べる。

4.1 データの構成

まず、メールの件数および、本文と件名で空ではないものの個数を示す（表5、図7）。空ではない件名（題名記入行）の総数は（三宅 2022:47）との間に数十件程度の差があるが、これは割合としては0.3%以下に過ぎない。異常のあるデータのまとめ方に伴う差である可能性がある。

表5 メールの件数と、空ではない本文・件名の個数の変化

送受信年 ⁽²⁴⁾	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	合計
件数	215	223	5868	11016	14406	15744	6072	15322	17016	13855	99737
本文（非空）	215	223	5857	10972	14336	15720	6060	15278	16925	13831	99417
件名（非空）	6	25	1099	2008	3167	1275	317	1134	955	887	10873

(24) 「送受信年月日」列のうち、年月日として読み取れるデータから、年を抽出した。

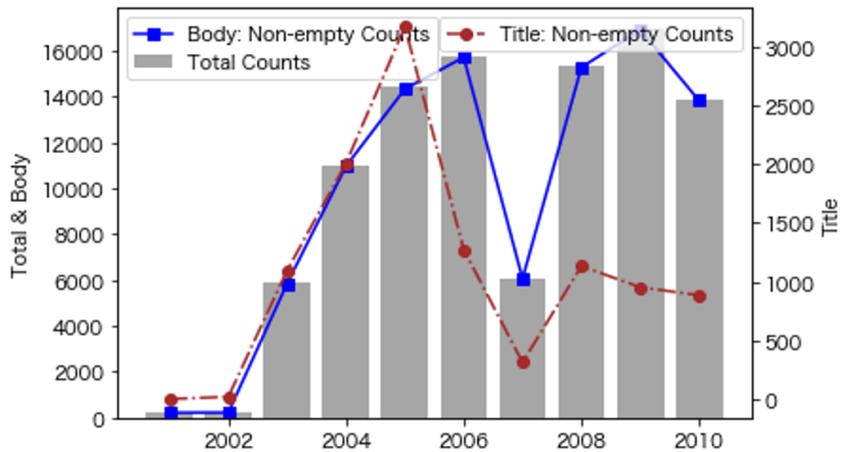


図7 メール件数と、空ではない本文・件名の個数の変化

ここで、メール送信者の生年月日のデータを調べる。本コーパスには、送受信主体者と（送受信の）相手の生年月日を表す列がある。本稿では、メールの送受信の別を表す行をもとにして、送信者の生年月日列を作成した。この中で送信者年月日として妥当だと考えられたもの⁽²⁵⁾を重複なく抽出し、得られた1325通りの値をヒストグラム（図8）にした。1980～1992年が多いが、1950～1960年前後にも小さな山が存在すると見られる⁽²⁶⁾。

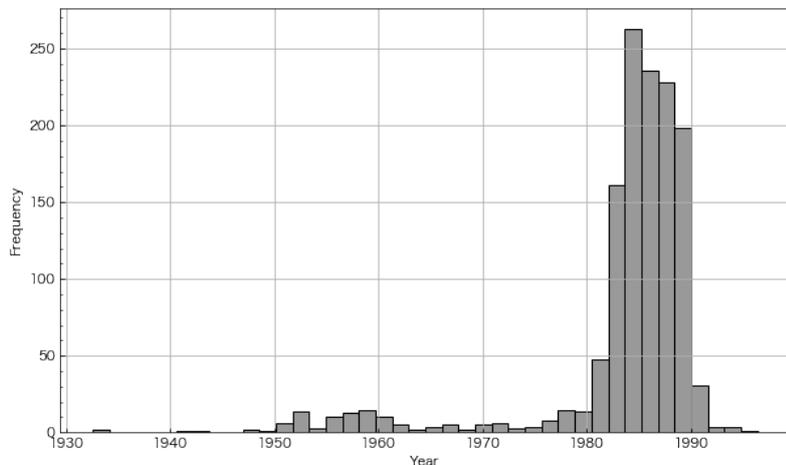


図8 送信者生年月日の分布（ヒストグラム）

次に、絵文字の個数を示す（表6, 図9）⁽²⁷⁾。そもそものメール件数が10,000件を超えた2004年以降、絵文字は、件名では一通あたり平均で0.03～0.2個程度、本文では平均1.1～2.5個程度登場している。なお、件名では2005年に突出して多く絵文字が用いられている。

⁽²⁵⁾ ここでは、8桁であり、数値化して1900000以上、年月日に変換可能であることを基準とした。

⁽²⁶⁾ 1950～1960年前後の山については、データを提供した（加藤2007:2）大学のゼミ生たちと、保護者や教師などとのやりとりによるものが含まれるのではないかと考えられる。

⁽²⁷⁾ 田中・林（2021）は調査の「年度」ごとに値をまとめているが、本研究では送受信年ごとの値を示す。

表 6 絵文字の個数と一通あたりの使用率の変化（本文・件名）

送受信年	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
絵文字個数（本文）	60	95	4058	15058	32229	18218	12195	37082	34527	32719
絵文字個数（件名）	0	10	398	908	2770	697	304	767	605	374
個数/件（本文）	0.279	0.426	0.692	1.367	2.237	1.157	2.008	2.420	2.029	2.362
個数/件（件名）	0.000	0.045	0.068	0.082	0.192	0.044	0.050	0.050	0.036	0.027

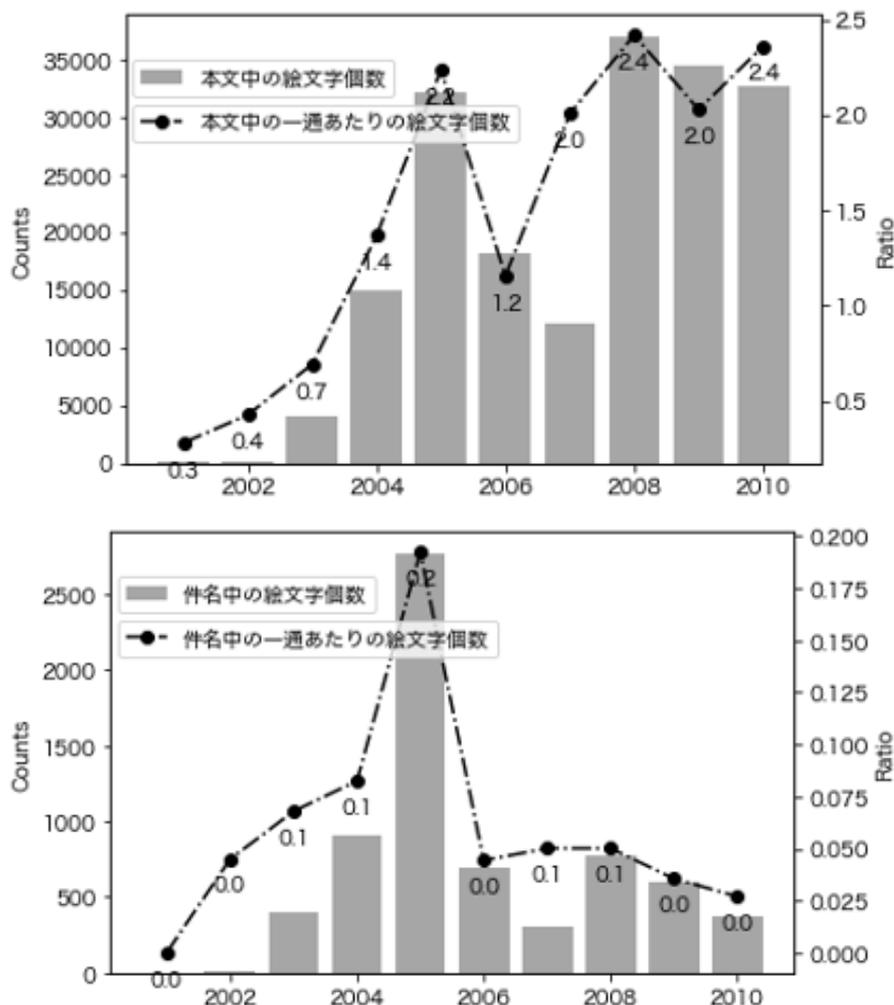


図 9 絵文字の個数と一通あたりの使用率の変化（本文・件名）

4.2 携帯会社名が異なる場合の絵文字の使用の変化

ここで、メールを送受信する携帯電話会社の同異と絵文字の使用率について見ておきたい。田中・林 (2021) は、「2006 年～2008 年にかけて絵文字使用が広がった」(ibid:115) と述べ、後押しの要因として、「2006 年に開始された異なるキャリア間の絵文字変換」(安岡 2007) を指摘する。ただし、au と Vodafone (2006 年以降 SoftBank) の使用度数の上昇傾向は論じているが、異なるキャリア間での絵文字使用数が増加したかどうかの計測に踏み込んではいない。

このため、本研究では、実際に携帯会社名が送信者と受信者で異なる場合の絵文字使用の変化を調べ、メール送信件数とともに図にした（図 10）。なお、表 7 に示すように、「携帯会社名」にはメーリングリストや不特定多数を表すものも含まれているが、割合は小さい。

表 7 携帯会社名と内訳

携帯会社名	D	E	V	D I V	ML	W	P	H	不明	その他	合計
説明	docomo	au/ TU-KA	Vodafone/ SoftBank	不特定 多数	メーリング リスト	WILLCOM	PC	PHS	不明	半角スペース 入りのE ⁽²⁸⁾	
送信側件数	57496	31960	9507	236	176	157	152	15	50	1	99750
受信側件数	58618	31118	9303	206	122	158	94	2	129	0	99750

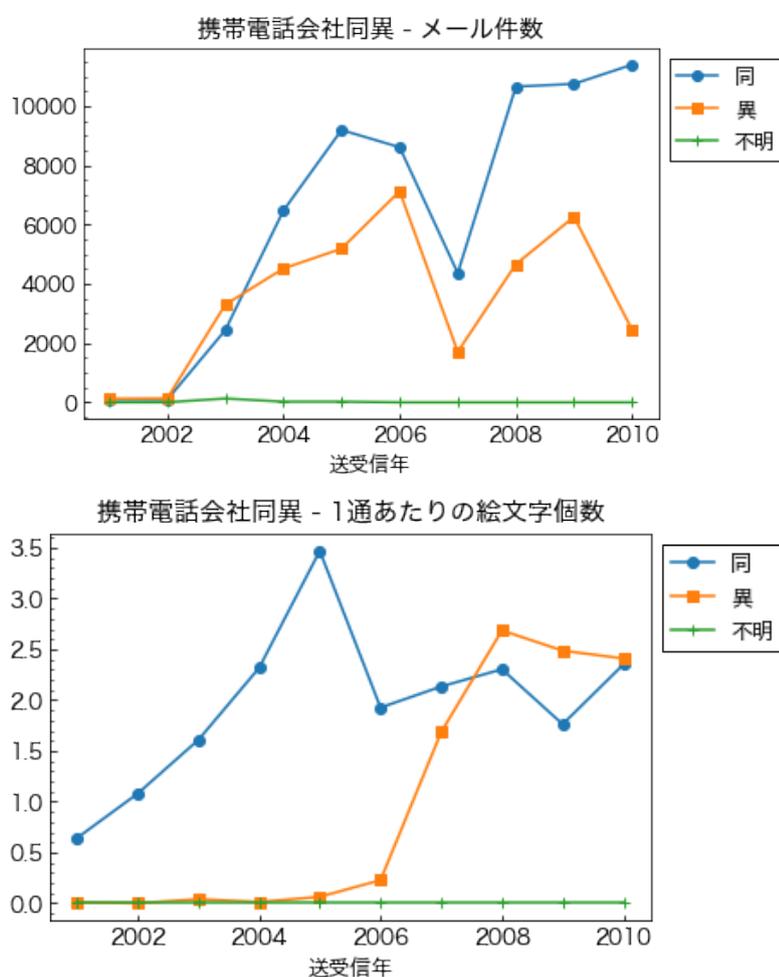


図 10 「携帯会社名」が同じ・異なる場合のメール件数と一通あたりの絵文字個数

たしかに、2006～2008 年にかけて、異なる携帯会社名間でのメールについての一通あたりの絵文字使用個数の上昇が見られ、同一名称間と同じような水準にまで上がっていることがう

(28) 実質的には「E」だと考えられるが、ここでは統合しない。

かがえる。

4.3 絵文字の位置タイプとその割合

本コーパスがメールを文ごとに区切っていることを受けて、本研究では絵文字の位置タイプを、大まかに文頭・文末・文の内部に分けて扱うこととした。

具体的には、まず、各絵文字に対して、その絵文字を含む絵文字のみでなる最長の塊を探し、その塊が一文全体 (the whole sentence) に一致する場合 (「w」)、文全体には一致せず塊が文頭 (the beginning of the sentence) にある場合 (「b」)、文末 (the end of the sentence) にある場合 (「e」)、どちらでもない、すなわち文内部 (in the of the sentence) にある場合 (「m」) の4つの場合に分けた。さらに、文頭 (b) と文末 (e) の場合は、文頭の塊の先頭にその絵文字があるか (「b0」) ないか (「b1」)、文末の塊の末尾にその絵文字があるか (「e0」) ないか (「e1」) の2通りに分けて、ラベル付けした。なお、文内部にある場合 (m) には便宜上、絵文字に「m0」というラベルを割り当てた。絵文字の塊が文全体に一致する場合は、文 (すなわち絵文字の塊) がその絵文字1文字のみでできている場合 (「wo」)、そうではなくその絵文字が文頭にある場合 (「wb」)、文末にある場合 (「we」)、文内部にある場合 (「wm」) の4通りに分けて、ラベル付けした。まとめると、位置タイプのラベルは b0, b1, e0, e1, m0, wb, we, wm, wo の9つである。

この基準に則って、コーパスの本文中・件名中の絵文字の位置タイプごとの割合を算出し、表と図で示した (表8、図11、図12)。

表8 絵文字の位置タイプごとの割合

位置ラベル	文頭部		文末部		文内部	文全体			
	b0	b1	e0	e1	m0	wb	we	wm	wo
本文中での割合	0.43%	0.11%	67.72%	24.34%	7.14%	0.07%	0.07%	0.06%	0.06%
件名での割合	2.33%	0.29%	64.22%	24.66%	4.39%	1.17%	1.17%	0.86%	0.91%

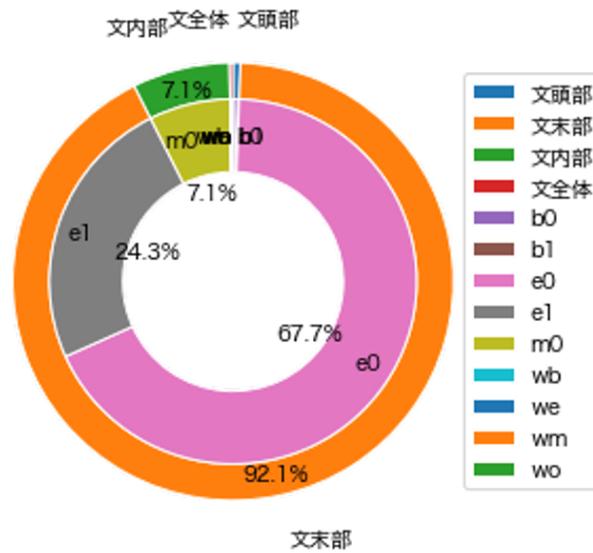


図 11 絵文字の位置タイプごとの割合（本文）

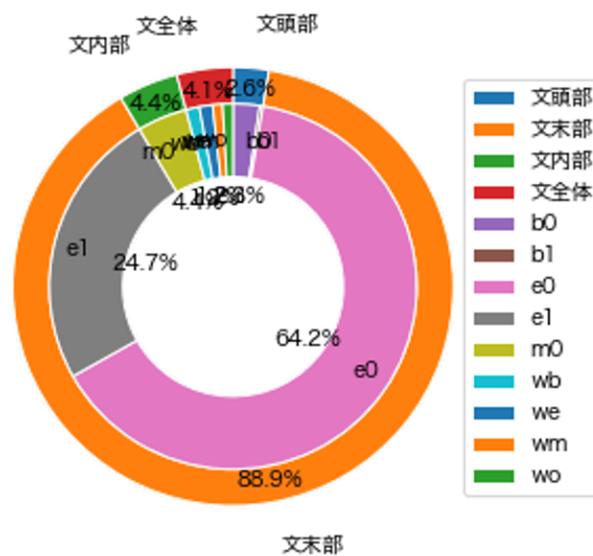


図 12 絵文字の位置タイプごとの割合（件名）

本文中では、文末部が 92% 程度と圧倒的に多く、文頭部は 0.54% 程度であった。また件名中でも、文末部が 89% 程度と圧倒的に多く、文頭部は 2.6% 程度であった。

件名中では本文中に比べて、一文が絵文字のみでなる割合が大きい。中には以下のような例があった（絵文字部分を山括弧を用いて表記し、コードに対応する画像を表示する）。(1) 絵文字 1 つのみの使用：「< (E 4 7 5) . 📌 📌 📌 の 3 枚でなる動画>」(2) 同じ絵文字の反復：「< (E 7 2 A) 🍷 を 3 回>」「< (E 6 F A) 🎮 を 9 回>」(3) 異なる絵文字の組み合わせ：「< (E 6 8 3) 🍷 🎮 >」「< (E 7 0 4) !! >」（本+二重感嘆符）

4.4 経年変化 - 本文文末部絵文字連続使用傾向の変化

位置タイプごとの割合の変化を図示する（図 13）。2004 年までに対して、2005 年以降は本文で e1 の割合が大きくなっている傾向が見受けられる⁽²⁹⁾。e1 は文末部であるが末尾ではないという位置のラベルであるため、この割合の増加は、絵文字を文末で連続使用する傾向の強まりを示唆する。

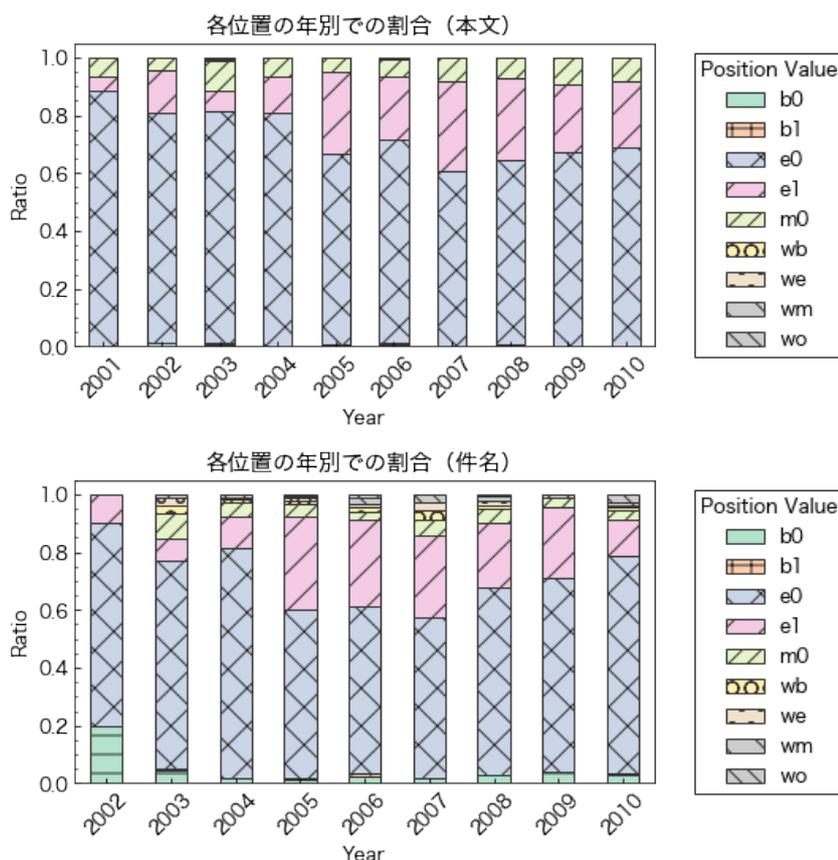


図 13 絵文字の位置タイプごとの割合の変化

ただし、本コーパスは、送信者を固定して、その絵文字使用を 10 年間追跡したものではない。そのため、同じ送信者たちに対しても e1 の割合が増加したかどうかを調べたい。ここで、2004 年と 2006 年で共通する送信者生年月日のうち、8 桁で 19000000 以降のもの（39 個）に対し、2006 年の e1 率から 2004 年の e1 率を引いたものを計算した。分布の平均値は 0.0569 で、中央値は 0.0388、分散は 0.185 となった。この分布のヒストグラムを図 14 に示す。やや正の側にあることがうかがえる。この分布に対し、ウィルコクソンの符号順位検定で、対立仮説を「2006 年の e1 率 - 2004 年の e1 率は 0 より大きい」とした片側検定を実施したところ、 $p = 0.0398 (< 0.05)$ が得られた。（補足：e1 の割合 ÷ (e0 か e1 の割合)、すなわち文末部の絵文字のうちで末尾以外にある割合について、その分布のヒストグラムを図 15 に示した。この

⁽²⁹⁾ 件名も似ているが、2010 年で割合がかなり小さくなっているように見える。

分布に同様に片側検定を実施したところ、 $p = 0.0330 (< 0.05)$ が得られた)

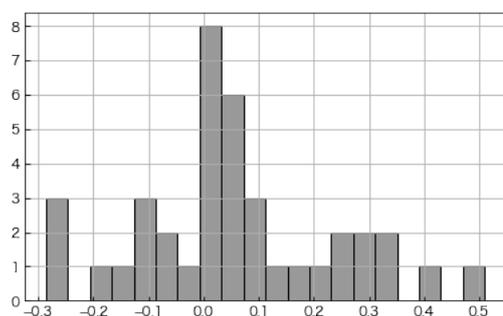


図 14 送信者が同じ場合の e1 割合の 2006 年と 2004 年の差

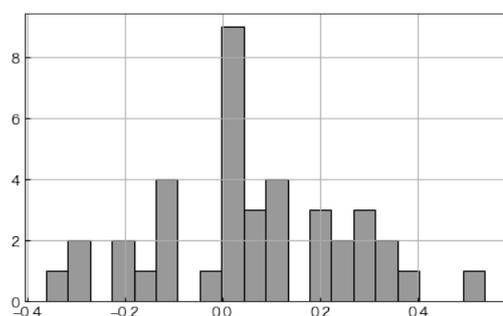


図 15 送信者が同じ場合の e1 割合 / (e0 割合 + e1 割合) の 2006 年と 2004 年の差

4.5 経年変化 - 絵文字のカテゴリとその割合および絵文字と感情表出

さて、絵文字を用いた感情表出について述べる。田中・林 (2021) は、高頻度絵文字の出現割合を示して、「本コーパスにおいて、(1)「感情を表すタイプ」はおおむね 2006 年度ごろから増加傾向を示しているのに対し、(3)「顔」の中でも最も多く現れる絵文字「わーい (嬉しい顔) (F 6 F 0)」は 2004 年度から安定的に使われている。このことから、絵文字使用の黎明期には顔の絵文字によって表現された「感情」が次第に (1) のような顔に比べて間接的な表現の絵文字に移行した可能性も指摘できる。」(ibid:115) と述べている。この「顔」と「感情を表すタイプ」の絵文字は本稿ではそれぞれ「表情系」と「漫符系」に包含される。このためカテゴリ割合の変化を調べ、図 16 に示したが、ここからは、カテゴリとして表情系から漫符系に移行したと言い切るのには難しそうであった。

そこで、「感情を表すタイプ」として提示された「あせあせ」等の絵文字の増加に関する別の説明を考えるため、使われにくくなった絵文字に着目した。個別の絵文字の順位を 2003 年を基準に算出すると、2003 年の時点で 6 番目に多かった「E 7 0 7 (たらーっ (汗) 🥲)」、同率 8 位だった「E 7 0 0 (パッド (下向き矢印) 📵)」「E 7 2 2 (冷や汗 🥵)」、10 位だった「E 6 4 3 (台風 🌀)」といった絵文字が、他と比べて順位を落とし⁽³⁰⁾、全体で 14~32 位と

⁽³⁰⁾ 順位の値だけを見るのではなく、他と比べるのは、2003 年では docomo 系が高順位だったが、2009~2010 年に

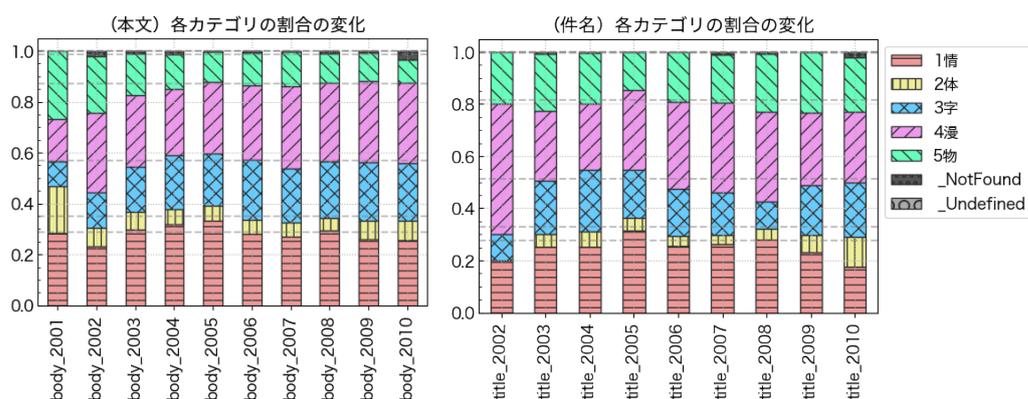


図 16 送受信年別の各カテゴリの割合

なっていることがわかる (表 9)。このこと、「E706 (あせあせ[👉])」は多く使われ続けていることを考慮すると、同じカテゴリの中でも、否定的な感情を「下向きの」図像で表す絵文字が減ったのではないかという仮説が立てられる。さらに、「否定的な感情の表し方が、直接的な表示から間接的な表示に移行したのではないか」という問いも立てられる。本稿では検証しないが、さらに、否定的感情を直接表す顔の絵文字と、「ニコニコ+あせあせ」のような複数の絵文字を用いて表す絵文字との、出現割合の変化を比較することからも、このような問いに関する調査と語用論的な分析が行えるだろう。

表 9 2003 年に本文中で高順位だった絵文字の本文中での順位の変化

EmojiCode \ 本文中での順位	2003	2004	2006	2009	2010	合計
E6F0	1	1	2	11	25	1
E6FB	2	3	6	25	32	6
E706	3	4	3	5	34	3
E6FA	4	5	4	6	22	4
E703	5	2	1	10	17	2
E707	6	15	25	44	213	32
E704	7	6	9	7	29	10
E700	8	12	7	58	127	18
E722	8	8	12	44	213	14
E643	10	14	17	52	78	26

なお、図 16 からは、本文中での物体生物系の割合が 2001~2004 年にかけて減少していることが認められる。

5. 文頭における絵文字使用

ここでは、文頭における文頭における絵文字の使用について、用例、高頻度絵文字の傾向、絵文字のカテゴリ、という三つの観点から分析する。これを通じて、文頭における絵文

なると au 系の絵文字が特に高順位に来ているためである。

字の使用には、先行研究で指摘されているような絵文字の用法や、全体的な傾向とは異なる性質があることを示す。

5.1 用例

まず、名詞または動詞の語幹として機能し、文の一部になっていると読めるものがある。「<(E 6 8 8) 携帯電話_dc: 📞>について。」「<(E 6 6 3) 不明>着いた?」「<(E 7 1 6) パソコン_dc: 💻>がきたら」「<(E 6 E 7) 6_dc: 🕒>時」「<(E 6 8 7) 電話_dc: 📞>していい?」「<(E 7 4 8) 桜_dc: 🌸>後、少して咲きそうだよ」のようなものである。上の例は件名からとったものであるが、ここにおける絵文字には、表現の文字数を減らすとともに、該当箇所をより目立たせるような効果が考えられる。

ただし、本文での絵文字には絵文字と説明部の重複があるものがある。例に「<(E 6 4 0) 雨_dc: 🌧️>雨降ってて、やっぱ無理って感じ<(E 6 F 0) わーい(嬉しい顔)_dc: 😄>」がある。この例の雨の絵文字では文字数は減らないが、強調の効果は考えられる。また、絵文字でテキストを挟むものがある。たとえば「<(E 6 4 3) 台風_dc: 🌀>集会<(E 6 4 3) 台風_dc: 🌀>」という件名である。このメールは本文で「今回は3人で集合しましょ」と持ちかけていた。この件名での左右対称的な絵文字使用には装飾的な効果が考えられる。

```
%%絵 (E 4 A C) %%この前のディップのレシピ
材料：
%%絵 (E B 7 D) %%クリームチーズ：1箱
%%絵 (E B 7 D) %%キャンベルクリームマッシュルームスープ：1缶
%%絵 (E B 7 D) %%マヨネーズ：1カップ
%%絵 (E B 7 D) %%ゼラチン：大さじ1.5(1袋でも可)(4倍の水でふやかす)
%%絵 (E B 7 D) %%玉ねぎ：1/2カップ
%%絵 (E B 7 D) %%セロリ：1～2本
%%絵 (E B 7 D) %%万能ねぎ：2～3本
%%絵 (E B 7 D) %%かに缶：小さいの1缶(缶汁ごと)
手順：
①：野菜類をみじん切り
②：ふやかしたゼラチンをレンチン
```

図 17 レシピと絵文字：絵文字を用いた材料の箇条書き

他方で、箇条書きを通じたメッセージ内容の構造化の機能を持つと見られる例もある。

たとえば、同じ絵文字を行頭で反復的に用いて、レシピを表すものがある（メッセージの一部を図 17 に表示）。ここで (E 4 A C) は「ゴハン1.au: 🍴」、反復する (E B 7 D) は「双葉.au: 🌿」に対応する。ここで絵文字は、対等な項目を並列する際の各項目のマーカースとして使われているといえよう。また、項目の並列において、絵文字が項目の内容や性質を表すと考えられるものもある。例えば、「<(E 6 8 8) 携帯電話_dc: 📞><電話番号>(改行)<(E 6 D 3) メール_dc: 📧><アドレス>」では、電話番号とアドレスに対応する絵文字が先頭でマーカースとなっているといえる。他に、「<(E 5 9 5) ハート.au: ❤️>あるもの」の後に箇条書きを複数行続け、「<(E 4 7 7) ハート2(われている)_動.au: 📺>ないもの」の後にまた箇条書き複数行を続ける、という例がある。ここ

では「あるもの」と「ないもの」をハートと割れたハートに対応させているといえる。

上述した簡条書き等の用例は、絵文字が情報の構造化にも用いられることを示す。さらにはどの絵文字を用いるかが、前後の文脈に関係しているケースもある（ハートと割れたハートの使い分けなど）。このような構造化に関わる例は、三宅 (2012) が提示した絵文字の分類「①事物そのものを表す」「②事象の表現に前置き/後置き」「③身体動作」「④プロソディー」「⑤前接文に合わせた感情を示す」「⑥前接文とは異なる感情を示す」「⑦装飾・雰囲気、リズムとり、句読点、トピック転換」には収まらないと考えられる⁽³¹⁾。

5.2 高頻度絵文字

ここでは文頭部における高頻度の絵文字の特徴を述べる。

まず、件名と本文をまとめた本コーパス全体で高頻度に出現した絵文字を表 10 に示す。この表での頻度は、田中・林 (2021) がキャリアごとに上位 10 位ずつ高頻度絵文字を抽出して示した度数とは数十件程度の差が見られたが、度数（頻度）での上位 14 位まで⁽³²⁾は一致している。頻度の差は、データの統合や絵文字の判定といった処理の違いに関わる可能性がある。

表 10 本文+件名全体での高頻度絵文字（上位 20 位まで）

Rank	Frequency	Image	Name
1	8926	😊	わーい (嬉しい顔) _dc
2	7912	!?	exclamation&question_dc
3	7873	💦	あせあせ (飛び散る汗) _dc
4	7115	🆕	びかびか (新しい) _dc
5	6471	💧	あせあせ (飛び散る汗) _au
6	5287	💡	ひらめき_dc
7	5027	📈	グッド (上向き矢印) _dc
8	4623	👉	きらきら_動_au
9	4513	!!	exclamationx2_dc
10	4367	?	?_au
11	4020	😄	うれしい顔_dc
12	3714	😄	顔 1 (うれしいカオ) _動_au
13	3492	❤️	黒ハート_dc
14	2872	💧	冷や汗_dc
15	2764	😄	がまん顔_dc
16	2736	👉	ウッシッシ_dc
17	2604	!!	「!!」_au
18	2472	ㇿ	ー (長音記号 2) _dc
19	2456	🎵	るんるん_dc
20	2425	📉	バッド (下向き矢印) _dc

次に、文頭部 (b0+b1) における高頻度絵文字を本文と件名に分けて示す (表 11)。全体とは異なり、顔の絵文字が順位を落とす一方、数字、決定 (OK)、天気、場所・建物、メールや

⁽³¹⁾ 山崎・西村 (2024) で X (旧 Twitter) の用例を取り上げて述べたことが、本コーパスでも当てはまっている。

⁽³²⁾ 度数で取ると、全体の上位 14 位までで docomo の上位 10 位が尽くされてしまうため、15 位以降の比較はできなかった。

電話機を示す絵文字が上位に来ている。これらは直接的には⁽³³⁾全体で高頻度に見られた絵文字のタイプ「感情を表すタイプの絵文字」「ダイヤルボタン等で直接入力可能な一般的な記号の絵文字」「顔の絵文字」(田中・林 2021:116)に当てはまるとは言い難い。ここからも、文頭部の絵文字には、全体とは異なる機能が推定される。

表 11 文頭部における高頻度絵文字（上位 20 位まで）（本文（左）、件名（右））

Rank	Frequency	Image	Name
1	109	✦	きらきら_動_au
2	50	☑	決定_dc
3	49	✉	メール_dc
4	43	🏠	びかびか（新しい）_dc
5	30	✉	封筒_au
6	27	🏠	家_dc
7	19	☔	雨_dc
8	18	1️⃣	1_dc
9	15	!!	exclamationx2_dc
10	15	🎓	学校_dc
11	15	2️⃣	2_dc
12	14	💻	ラブレター_dc
13	13	♥️	ハート（トランプ）_au
14	13	🚗	満室・満席・満車_dc
15	12	📞	携帯電話_dc
16	11	🚗	車（セダン）_dc
17	11	4️⃣	4_dc
18	11	✉	封筒2_au
19	10	🚗	電車_dc
20	10	3️⃣	3_dc

Rank	Frequency	Image	Name
1	14	♥️	黒ハート_dc
2	13	☀️	晴れ_dc
3	13	✦	きらきら_動_au
4	11	🏠	家_dc
5	8	✉	メール_dc
6	8	🏠	びかびか（新しい）_dc
7	8	🎉	パースデー_dc
8	6	☑	決定_dc
9	5	🐼	ウマ_dc
10	4	2️⃣	2_dc
11	4	😊	わーい（嬉しい顔）_dc
12	4	📞	電話_dc
13	4	🍷	本_dc
14	3	🚗	電車_dc
15	3	🌿	クローバー_dc
16	3	💻	パソコン_dc
17	3	📞	携帯電話_dc
18	3	🌸	桜_dc
19	3	💻	ラブレター_dc
20	3	🍷	ひよこ_dc

5.3 絵文字のカテゴリと文頭部

本節では、最後にカテゴリの観点からも分析する。本文・件名における各位置タイプでの各カテゴリの割合を図示する（図 18）。図 18 から、文頭部（b0・b1）では文末部・文内部に比べて、表情系の割合がかなり小さいといえる。また、特に先頭（b0）では漫符系の割合も小さくて物体生物系の割合が大きい（過半数となる）こと、2文字目以降（b1）では漫符系の割合が大きいことがうかがえる。なお、今回はカテゴリとして「文字符号系」にまとめているが、文字符号系の中でも、文頭部では文末部より数字類の割合が大きかった。このことから、文頭では、全体の傾向と異なり、事物を表す絵文字が用いられやすいと推測される。

⁽³³⁾ 文脈によっては天気絵文字が気分を表す、といったことは考えられる。

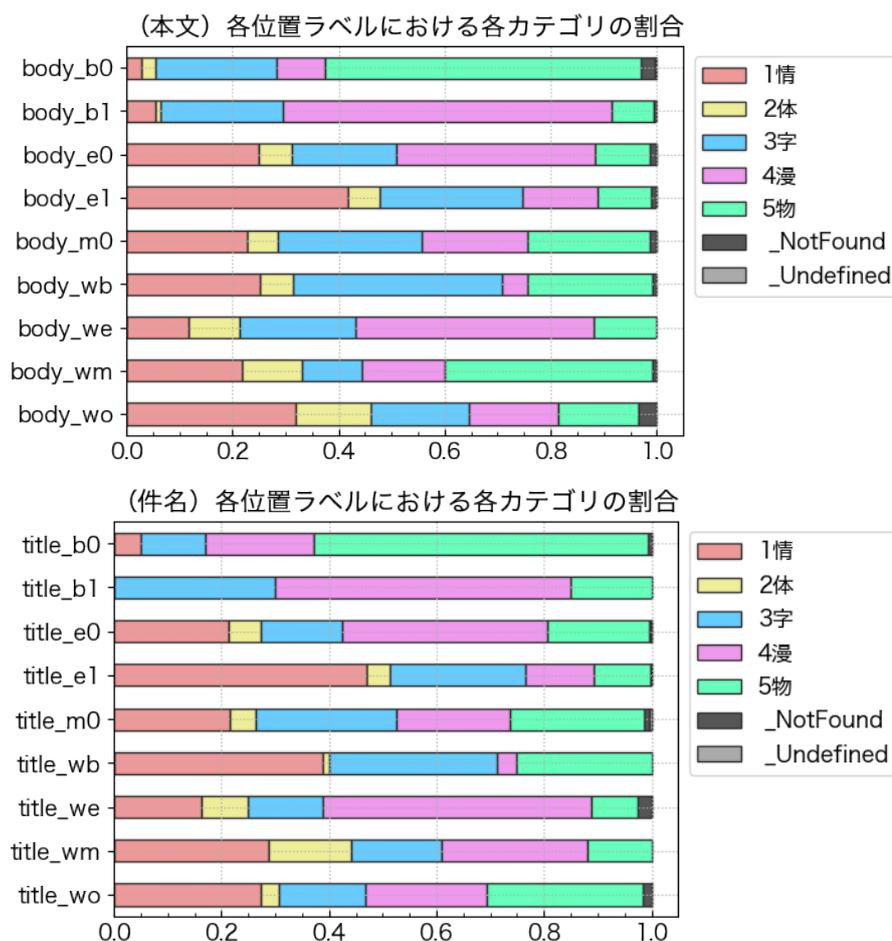


図 18 本文・件名での各位置タイプにおける各カテゴリの割合

6. 文末部での絵文字の連続使用について

本節では、文末部での絵文字の連続使用について議論する。

6.1 絵文字チャンクの長さ

末尾ではない絵文字が文末部全体の絵文字に占める割合は、本文で 26.4% 程度、件名で 27.7% 程度である（表 8 を参照）。ここで、文末部において絵文字を連続して何個用いていたかの分布を、片側対数表示のグラフで示す（図 19、20）⁽³⁴⁾。

⁽³⁴⁾ (補足) 指数関数でのフィッティングが「まあまあ」良いように見え、このことは、単に文末部から一定の確率で絵文字が追加されていく、というモデルとの「近さ」を思わせる。ただし本文のグラフでは、回帰曲線（片対数グラフで直線）からのずれを見るに、そのモデルと比べると、チャンク長 2 が 3~10 に対して多いように思われる。

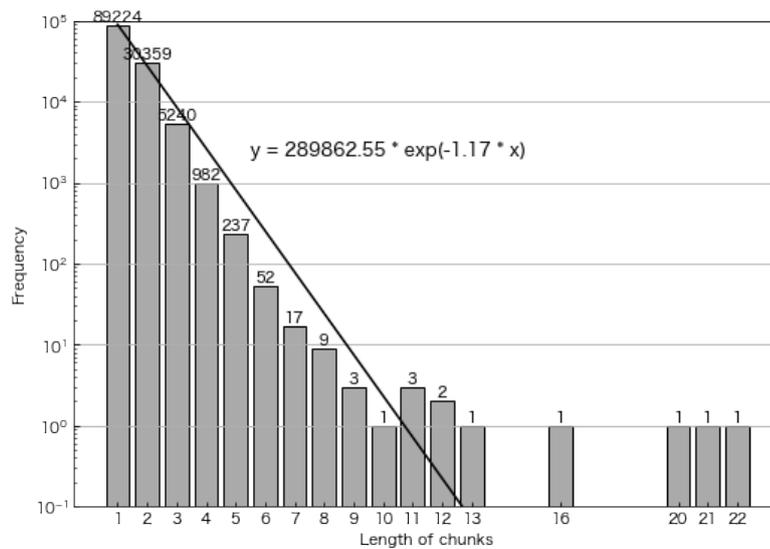


図 19 本文での文末絵文字チャンク長の分布

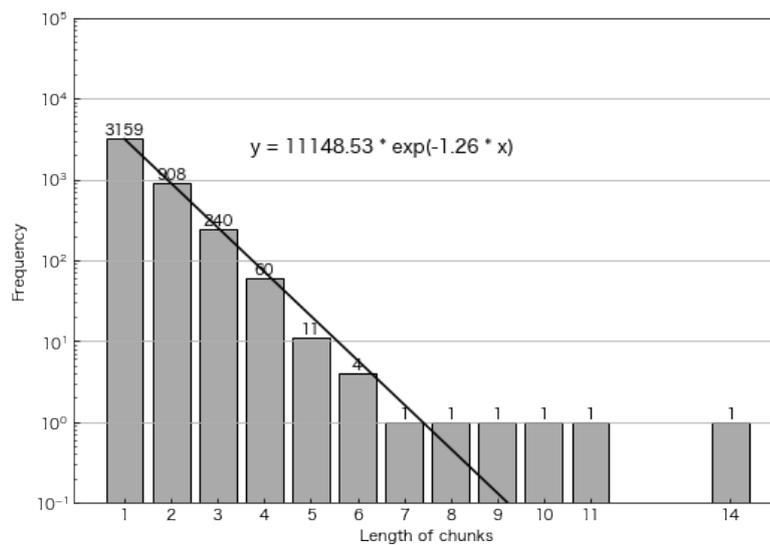


図 20 件名での文末絵文字チャンク長の分布

本文では 1 個だけ絵文字を使う場合が 8 万 9 千回以上あった一方で、22 個連続して使う場合もあったことがわかる。また、件名では最大で 14 個連続して使われていた。

6.2 文末部での用例

6.2.1 文末部に高頻度に出てくる組み合わせ

ここでは、まず接続する 2 つの絵文字の組に着目する。本文文末部でのこのような組 (ペア) の、キャリアを問わない頻度上位 15 位までを、e1-e0 と、e1-e1 それぞれについて示す⁽³⁵⁾。

⁽³⁵⁾ ここでは、絵文字コードから絵文字画像名を通じて計算した絵文字名を表示に使っている。絵文字名が不明のコードには空白を割り当てたが、高頻度ペアの表には登場していない。また、件名の文末部では、最大頻度が e1-e0 で

表 12 本文 e1-e0 での高頻度ペア

頻度	左	右 (左と異なる場合のみ記す)	絵文字同異	左カテゴリ	右カテゴリ
867	(E706) あせあせ (飛び散る汗) _dc	-	同	漫符系	漫符系
860	(E5B1) あせあせ (飛び散る汗) _au	-	同	漫符系	漫符系
780	(E483) ? _au	-	同	文字符号系	文字符号系
725	(E6F5) グッド (上向き矢印) _dc	-	同	文字符号系	文字符号系
478	(E703) exclamation&question_dc	-	同	文字符号系	文字符号系
447	(E6F0) わーい (嬉しい顔) _dc	(E6FA) ぴかぴか (新しい) _dc	異	表情系	漫符系
432	(E704) exclamation × 2_dc	-	同	文字符号系	文字符号系
417	(E6FA) ぴかぴか (新しい) _dc	-	同	漫符系	漫符系
414	(E72B) がまん顔_dc	(E706) あせあせ (飛び散る汗) _dc	異	表情系	漫符系
392	(E471) 顔1 (うれしいカオ) 動.au	(E5B1) あせあせ (飛び散る汗) _au	異	表情系	漫符系
390	(E700) バッド (下向き矢印) _dc	-	同	文字符号系	文字符号系
367	(E2D) 上向きカーブ矢印.au	-	同	文字符号系	文字符号系
351	(EAA B) きらきら 動.au	-	同	漫符系	漫符系
328	(E72A) うれしい顔_dc	(E6FA) ぴかぴか (新しい) _dc	異	表情系	漫符系
308	(E6F0) わーい (嬉しい顔) _dc	(E6FB) ひらめき_dc	異	表情系	漫符系
294	(E708) ダッシュ (走り出すさま) _dc	-	同	漫符系	漫符系
266	(E72A) うれしい顔_dc	(E6FF) ムード_dc	異	表情系	漫符系
254	(E2E) 下向きカーブ矢印.au	-	同	文字符号系	文字符号系
252	(E727) 指で OK_dc	(E6FA) ぴかぴか (新しい) _dc	異	身体系	漫符系
238	(E30) 「!!」_au	-	同	文字符号系	文字符号系

表 13 本文 e1-e1 での高頻度ペア

頻度	左	右 (左と異なる場合のみ記す)	絵文字同異	左カテゴリ	右カテゴリ
289	(E6F5) グッド (上向き矢印) _dc	-	同	文字符号系	文字符号系
263	(E703) exclamation&question_dc	-	同	文字符号系	文字符号系
227	(E704) exclamation × 2_dc	-	同	文字符号系	文字符号系
180	(E5B1) あせあせ (飛び散る汗) _au	-	同	漫符系	漫符系
173	(E72A) うれしい顔_dc	(E706) あせあせ (飛び散る汗) _dc	異	表情系	漫符系
156	(E70A) ー (長音記号2) _dc	(E703) exclamation&question_dc	異	文字符号系	文字符号系
120	(E70A) ー (長音記号2) _dc	(E6F0) わーい (嬉しい顔) _dc	異	文字符号系	表情系
117	(E70A) ー (長音記号2) _dc	(E72A) うれしい顔_dc	異	文字符号系	表情系
97	(E700) バッド (下向き矢印) _dc	-	同	文字符号系	文字符号系
90	(E706) あせあせ (飛び散る汗) _dc	-	同	漫符系	漫符系
84	(E30) 「!!」_au	-	同	文字符号系	文字符号系
82	(E31) くり.au	(E483) ? _au	異	文字符号系	文字符号系
78	(E72B) がまん顔_dc	(E706) あせあせ (飛び散る汗) _dc	同	表情系	漫符系
76	(E703) exclamation&question_dc	(E72A) うれしい顔_dc	異	文字符号系	表情系
73	(E6EC) 黒ハート_dc	-	同	漫符系	漫符系
72	(E757) げっそり_dc	-	同	表情系	表情系
70	(E471) 顔1 (うれしいカオ) 動.au	(E483) ? _au	異	表情系	文字符号系
65	(E483) ? _au	-	同	文字符号系	文字符号系
60	(E729) ウィンク_dc	(E727) 指で OK_dc	異	表情系	身体系
60	(E72A) うれしい顔_dc	(E6F5) グッド (上向き矢印) _dc	異	表情系	文字符号系

この高頻度絵文字ペアの表においては、上位 4 位は、同じ絵文字 (コード) が反復するペア

も e1-e1 でも 50 未満と小さかったため、ここでは提示を略す (上位 15 位までに登場するペアやカテゴリは似た傾向にあった)。

である。なお、本文文末部全体としては、同じ絵文字（コード）が反復する場合よりも、異なるコードが続く場合のほうが多い（表 14）⁽³⁶⁾。

表 14 本文文末部、連続絵文字コードが同じ場合と異なる場合

本文文末部ペア	同の個数	異の個数	異の割合
e1-e0	9435	27475	74.4%
e1-e1	2706	5714	67.9%

高頻度絵文字ペアの用例を見ると、同じ絵文字の反復には、表情系（げっそり）、漫符系（あせあせ、ぴかぴか、きらきら、ダッシュ）、文字符号系（感嘆符、疑問符、上下の向きの矢印）が見られる。これらについては、類像性 (Givón 1985) の観点から、同じ表現の重複による強調、特に、感情・ムード・驚きや疑問に思う気持ちの強調効果が考えられる⁽³⁷⁾。他方で、異なる絵文字の接続には、表情系＋漫符系（うれしい顔（3種）＋ぴかぴか・あせあせ）、身体系＋漫符系（指で OK ＋ぴかぴか）、表情系＋身体系（ウィンク＋指で OK、など）といった組み合わせが見られる。「ぴかぴか」が右に来る 2 通りには、肯定的なムードを付け加えているという解釈が考えられる。「うれしい顔＋あせあせ」は、前に来ている顔だけでは完結しない気持ちの状態（顔は笑っているが、ヒヤヒヤしている、など）を表現している可能性がある。また、「ウィンク＋指で OK」は「は～い< (E 7 2 9) >< (E 7 2 7) >」「木曜にでも資料を渡すよ< (E 7 2 9) >< (E 7 2 7) >」といった例文に現れた。ここでの指の絵文字は、相手を受容・肯定する態度が顔だけでは完結しておらず、それを補うものではないかと考えられる。また、文字符号系＋表情系に見られた「!?!+うれしいカオ」は、文末記号のあとに顔の絵文字がついたものだと解釈できる。

上記の例から、右側に来る絵文字は、前接絵文字までから受け手が予想・解釈する内容（特にモダリティ的な情報）に対する、強化・補足・（予想の）修正を行い、また「指で OK」のように受け手との距離感を調整するように配置される傾向があるのではないかと推測される。

6.2.2 長いチャンクでの用例

絵文字を 3 つ以上連続させた際の組み合わせ方についても調べるため、連続する絵文字の長さとともにそのような文の例を 6 つ挙げる（表 15）。

⁽³⁶⁾ 本稿では計算しないが、全ての絵文字に対してその（文末部での）登場率の 2 乗を計算したものの和を求め、実際に同じ絵文字が反復して現れている率と比較することは、「同じ絵文字が（ランダムな場合よりも）続きやすいかどうか」を評価する 1 つの方法となりうるだろう。

⁽³⁷⁾ 疑問符が二つ連続しても、疑問に対する疑問を表すというわけではない。

表 15 絵文字チャンクの長さ为例

絵文字チャンクの長さ	例
(本文) 4	昨日は仕事行って帰りの電車は停電でとまって、誕生日どころじゃなかったよ < (E 6 4 0) 🌧️ が 3 回 > < (E 6 4 4) 🌧️ >
(本文) 9	H A P P Y B I R T H D A Y < (E 6 8 6) 🎂 が 7 回 > < (E 6 D 1) 🕯️ > < (E 6 F A) 🎉 >
(本文) 22	うおおおめでとう < (E B 3 0) 🎉 が 13 回 > < (E A A B) 静止画像 🎉 等でなる動画絵文字が 9 回 >
(件名) 4	どしたのコレ < (E 7 2 6) 🧡 > < (E 7 1 B) 🌀 > < (E 7 0 3) !? > < (E 6 E F) 📉 >
(件名) 10	ハウル < (E 7 2 6) 🧡 が 10 回 >
(件名) 14	え < (E 7 0 4) 🎉 が 14 回 >

この表で同じ絵文字が反復するケースを見ると、第一に、漫符、表情のある顔・感嘆符が入った絵文字を重ねることで、感情やムードを強く伝えたりしているのではないかと推測される。また「雨」という天気を示す絵文字を重ねることに、発話者にとっての天気の印象や影響の強さを示す効果の可能性が考えられる。

その一方で、異なる絵文字を組み合わせた例もある。「昨日は仕事行って帰りの電車は停電でとまって、誕生日どころじゃなかったよ < (E 6 4 0) 🌧️ が 3 回 > < (E 6 4 4) 🌧️ >」では「雨」3つの後に「霧」の絵文字を加えている、この霧の絵文字については、状況をより詳細に伝えるものであるという解釈も、霧のかかった気持ちを表すという解釈も考えられる。また、「どしたのコレ < (E 7 2 6) 🧡 > < (E 7 1 B) 🌀 > < (E 7 0 3) !? > < (E 6 E F) 📉 >」については、「発話者」+「話題の対象 (=コレ)」+「気持ち」といった構造が推測される。ここでは絵文字「🌀」が、前接テキスト中の直示表現「コレ」についての不足している情報を補うのではないかと考えられる。絵文字の連続で視覚的な状況を構成し、その一部として「🌀」を提示しているのである⁽³⁸⁾。

さらに、同じ絵文字の反復と異なる絵文字の追加を組み合わせており、その個数が特別な意味をもたらすとみられる例がある。本文での長さ 9 の例では「(E 6 8 6) パースデー_dc : 🎂」を 9 つ反復した後に「(E 6 D 1) i モード_dc : 🕯️」の絵文字を重ねている。このメールの送受信年月日と受信者の生年月日から、この受信者は 22 歳の誕生日を迎えたものと推測された⁽³⁹⁾。ここでは、「🎂」の絵文字の蠟燭 3 本を 9 つと、「🕯️」の 1 つの縦棒（蠟燭に似ていると思われようか）を合わせて、視覚的に 22 本を実現しているものだと考えられる。「🎂」の絵文字は、誕生日であることと、3 本あることを、ともに示す機能を担うのである。

以上から、絵文字を多く連続して用いている例には、一般的に多く見られるというよりも、その話題・背景でこそその用い方をしている、というものがあるといえよう。

(38) ここで、視覚的な状況のほうに着目すれば、「どしたのコレ」は人物のセリフのようにも読めるかもしれない。

(39) 実際、前後のメールデータにはそれを明示したものもあった。

6.3 絵文字連続ペアとカテゴリ

上では個別の例について述べたが、ここから、カテゴリ同士の組み合わせの傾向について分析する。件名での頻度が低いことを考慮し、本文のみ取り扱う。はじめに、本文の文末部での各カテゴリごとの組み合わせの頻度を表で示す。

表 16 本文文末部、末尾とその1つ前 (e1-e0) での各カテゴリの組み合わせの頻度

左側\右側	表情系	身体系	文字符号系	漫符系	物体生物系	(NF)	(Undef.)	左側での合計
表情系	867	1598	2217	9765	1366	200	2	16015
身体系	257	153	229	1412	279	48	1	2379
文字符号系	1781	180	4415	1768	447	75	1	8667
漫符系	400	50	278	4531	139	47	0	5445
物体生物系	194	140	492	2116	972	107	0	4021
(NF)	17	23	17	116	12	194	0	379
(Undef.)	1	0	0	1	0	1	1	4
右側での合計	3517	2144	7648	19709	3215	672	5	36910

表 17 本文文末部、非末尾とその一つ前の (e1-e1) での各カテゴリの組み合わせの頻度

左側\右側	表情系	身体系	文字符号系	漫符系	物体生物系	(NF)	(Undef.)	左側での合計
表情系	437	440	650	1160	188	11	0	2886
身体系	34	84	66	142	37	3	0	366
文字符号系	1224	106	1729	260	250	19	0	3588
漫符系	81	14	44	738	35	3	0	915
物体生物系	47	42	110	133	250	2	0	584
(NF)	6	2	6	22	0	41	0	77
(Undef.)	1	0	0	0	0	0	3	4
右側での合計	1830	688	2605	2455	760	79	3	8420

表 16 と表 17 から、まず、漫符系は、連続使用で末尾 (e0) に来る頻度が全体の過半数を占め、他のカテゴリと比べて多い (他の各カテゴリの倍以上ある) ことがいえる。特に、e0 自体では漫符系の割合は表情系の 2 倍には達していなかった (図 18) が、連続使用の e0 では 5 倍以上ある。ただし、漫符系は、e1-e0 の左側 (e1 側) に来る頻度でいうと、全体の $\frac{1}{6}$ 以下で、カテゴリとしての順位も 3 番目となっている。さらに e1-e1 でも右側に来る頻度は文字符号系に次ぐ 2 位、左側に来る頻度は 3 位である。ここから、漫符系は、特に末尾 (e0) に来やすいといえる。

次に、表情系は、e1-e0 で右側 (e0) に来る頻度が全体の $\frac{1}{10}$ 以下で、順位も 3 位であるが、左側 (e1) に来る (末尾から 2 番目の絵文字となる) 頻度は全体の $\frac{1}{3}$ 以上で順位は 1 位である。ただし、e1-e1 での右側に来る頻度は全体の $\frac{1}{4}$ 以下で順位も 3 位となる。e1-e1 のパターンは、末尾から 3 つ以上絵文字が続く場合に限られている。ここから、表情系は、特に末尾から 2 つだけ絵文字が連続する場合の、末尾の 1 つ手前に多いのではないかと推測される。

この 2 つの表で、「合計」において 1 位をとったカテゴリの残る 1 つが、文字符号系である。これは e1-e0 では右側でも左側でも頻度 2 位となっている (割合は $\frac{1}{5} \sim \frac{1}{4}$)。ただし、他のカテゴリが左にある場合には漫符系が右 (e0) に来る頻度が最も大きいのに対し、文字符号系が左

にある場合は同じ文字符号系が続く頻度が他に優っている。右 (e0) になる場合についても見ると、身体系・漫符系・物体生物系が右になる場合は表情系が左に来る頻度が最も大きいのに対し、文字符号系が右になる場合は同じ文字符号系が左に来ることが最も多い。とはいえ、表情系が右になるときも、文字符号系が左に来ることが最も多い。ここから、文字符号系-文字符号系という組み合わせの強さと、表情系-文字符号系という組み合わせの強さが推測される。また、e1-e1 では、文字符号系は右側でも左側でも頻度 1 位である (割合は $\frac{1}{4} \sim \frac{1}{3}$)。e1-e0 と同様に、文字符号系同士の組の頻度が他に優っている。e1-e1 での「合計」の順位や相対的な割合が e1-e0 より高いことから、文字符号系は比較的、末尾から 3 つ以上連続する絵文字の塊にも現れやすいと推測される。

ここで、同じカテゴリ同士の組の中に異なる絵文字の組がどの程度あるかについて述べる。

表 18 同じカテゴリ同士の組に異なる絵文字の組がどれだけあるか

	e1-e0 で			e1-e1 で		
	カテゴリが 同じ組	絵文字も 同じ組	同分類で絵文字が 異なる割合	カテゴリが 同じ組	絵文字も 同じ組	同分類で絵文字が 異なる割合
表情系	867	788	9.1%	437	410	6.2%
身体系	153	136	11.1%	84	79	6.0%
文字符号系	4415	4004	9.3%	1729	1346	22.2%
漫符系	4531	4091	9.7%	738	691	6.4%
物体生物系	972	348	64.2%	250	141	43.6%

カテゴリが同じで絵文字が異なる割合は、e1-e0 では物体生物系 (60% を超える) 以外、どれも 10% 前後 (9.1~11.1%) である。ただし、e1-e1 では、文字符号系でも 2 割以上異なる絵文字の組み合わせが生じている。表 13 に示したが、e1-e0 では長音記号絵文字 (画像は^ㇿ) が左に出現することが多く、これの右に感嘆符と疑問符のセット絵文字 (画像は^{!?}) が来る組が、高頻度である⁽⁴⁰⁾。

ここで、同じカテゴリでも異なる絵文字のペアが多く見られた物体生物系について、そのようなペアの例を見ると、鳥と猫を組み合わせるなど、その場に応じた複雑な景色の創造を行っているのではないかと考えられるものがあった。その一方で、鳥と太陽、鳥と四つ葉、鳥と花、という組もあったが、これらについては、太陽や四つ葉や花が「絵文字で表される物体そのもの」ではなく「晴れやかな気持ち」「幸運を願う気持ち」や「なにかポジティブな気持ち」を表すために用いられているという可能性も考えられた⁽⁴¹⁾。

6.3.1 カテゴリ間の結びつき - 左右の非対称性

さて、カテゴリの「右に来やすさ」「左に来やすさ」について述べる。文末部の末尾に至る 2 連続絵文字 (e1-e0) に関して、各カテゴリの組み合わせ (X, Y) に対して、X が左で Y が右で

⁽⁴⁰⁾ ただし長音記号^ㇿの反復の組は、e1-e1 でも^{!!}と^{!?}の反復が高順位なのに対し、同率 40 位 (32 回) に留まる。このことは、長音記号絵文字と感嘆符・疑問符絵文字の反復に関する性質の差を示唆するものかもしれない。

⁽⁴¹⁾ たとえば「太陽」については、篠原・松中 (2010) のように、天候の視覚的な記号表現が感情解釈に影響を与えることを支持するような研究がある。このため、発話者が感情表現に使う可能性も考えられる。

ある頻度と、その逆である頻度の比を求めた。なおここではカテゴリ未定義・エラーは除く。

その結果は、次のようなものであった。(1) 表情系は、他のどのカテゴリが相手でも左に来るほうが多く、各頻度比は 1.2 以上であった。(2) 身体系は、文字符号系・漫符系・物体生物系の 3 カテゴリに対しては左寄り（頻度比 1.3、28.2、2.0）であったが、表情系に対してのみは右寄り（表情系-身体系の頻度は逆の 6.2 倍程度）であった。(3) 文字符号系は、表情系・身体系・物体生物系の 3 カテゴリに対しては右寄り（頻度比 1.2、1.3、1.1）であったが、漫符系に対してのみは左寄り（文字符号系-漫符系の頻度は逆の 6.4 倍弱）であった。(4) 漫符系は、他のどのカテゴリが相手でも右に来るほうが多く、各頻度比は 6 以上であった。特に「表情系-漫符系」「身体系-漫符系」は、逆の 24 倍以上あった。(5) 物体生物系は、中間的な振る舞いであった。漫符系・文字符号系に対しては左に来やすく（頻度比 15.2 と 1.1）、表情系・身体系に対しては右に来やすかった（頻度比 7.0 と 2.0）。

ここから、互いに異なるカテゴリのペアにおける「左に来やすいか」「右に来やすいか」と矛盾しないような、「(左) 表情系 - 身体系 - 物体生物系 - 文字符号系 - 漫符系 (右)」という並びができる。

6.3.2 カテゴリ間の結びつき - 存在感

「カテゴリ A (左側) とカテゴリ B (右側) の結びつき」と「カテゴリ A (左側) とカテゴリ C (右側) の結びつき」のようなものを比較したいとする。本研究では、その指標として、「カテゴリ B が右側に来る中での、カテゴリ A-カテゴリ B の割合」と「カテゴリ C が右側に来る中での、カテゴリ A-カテゴリ C の割合」を用いることを考えた。いわば「存在感」の比較ともいえよう。

実際に、左側のカテゴリを固定して (X)、右側に来る各カテゴリ (Y_i) に対し、 Y_i が右に来る全ての組での X が左側となる率を計算した⁽⁴²⁾。この値を線の太さに対応させ、さらに、 X を固定したときにこの値が大きい順に Y_i を並べた図を示す (図 21、カテゴリ名称は略記)。

(42) つまり、全体のうち、カテゴリ $X - Y$ という組み合わせの割合を $e_{X,Y}$ で表すと、各 Y_i について $\frac{e_{X,Y_i}}{\sum_Z e_{Z,Y_i}}$ を計算したことになる。

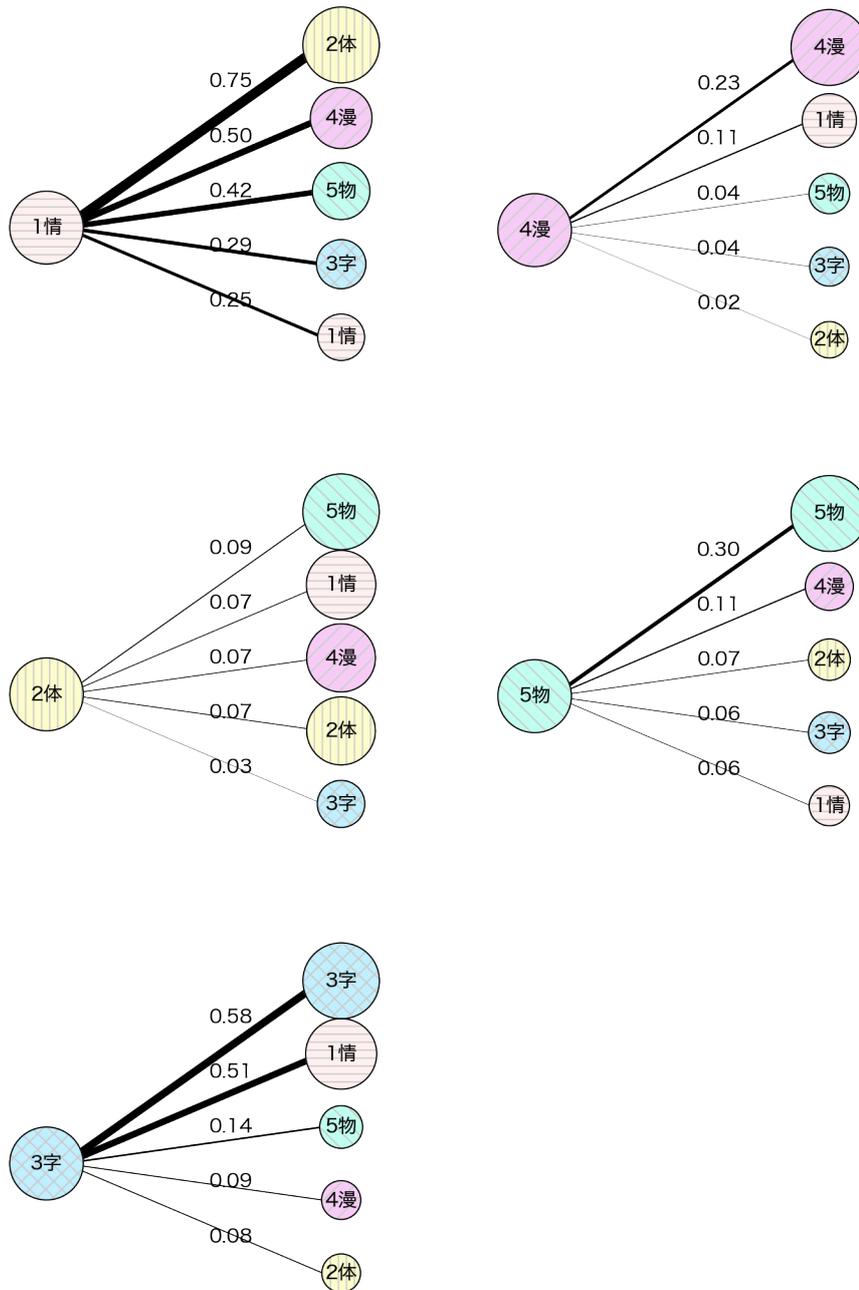


図 21 本文 e1-e0 での存在感比較 (始点固定・各終点での存在感を比較)

たとえば、右側に「身体系」が来る場合においては、左側が「表情系」であるものが75%(0.75)程度である。これを左側の「表情系」の視点で見た時には、「表情系」が左に来る割合が75%もあるカテゴリは他にないため、この「1情」ノードから発する線の並び順では「身体系(2体)」が1位となる。左側が「身体系」では、1位の「物体生物系」ですら9%しかない(つまり「身体系」が左に来ることが少なかった)が、他のカテゴリと比べれば値が大きいため、1位となっている。

逆に、右側のカテゴリを固定し、各左側における存在感を比較したものが、図 22 となる。漫符系が右側に多いことが、集まる線の太さからもうかがえる。

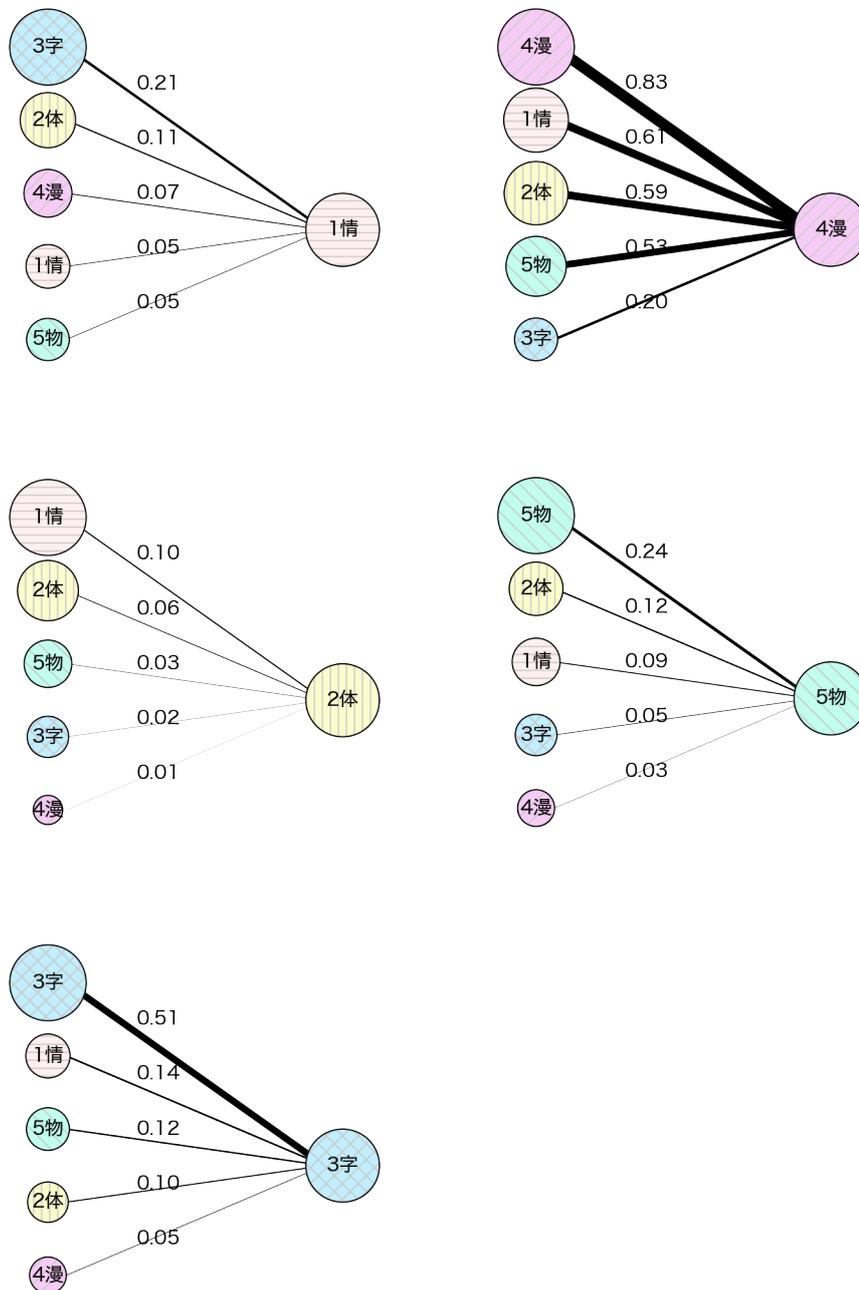


図 22 本文 e1-e0 での存在感比較（終点固定・各始点での存在感を比較）

二つの図から、カテゴリ間の結びつきの強さについて論じる。まず、線の終点（右側）での存在感を始点（左側）で可視化した図 21 からは、「表情系→身体系」「文字符号系」→「表情系・文字符号系」「漫符系→漫符系」「物体生物系」→「物体生物系」という結びつきが、始点（左側）において比較的強いことがうかがえる。次に、線の始点（左側）での存在感を終点（右

側) で可視化した図 22 からは、「文字符号系→表情系」、「表情系・身体系」→「身体系」、「文字符号系」→「文字符号系」、「(文字符号系以外)」→「漫符系」、「物体生物系→物体生物系」という結びつきが、比較的強いことがうかがえる。両者に共通するものを総合すると、「表情系→身体系」と「文字符号系→表情系」、文字符号系・漫符系・物体生物系同士の連続、となる。

最後に、このカテゴリ間での結びつきの強さを始点視点・終点視点で表すやりかた（並べ替えは行わない）を、e1-e1 にも適用した図（図 23）と、さらに絵文字が異なる場合に絞った図（図 24）を示す。

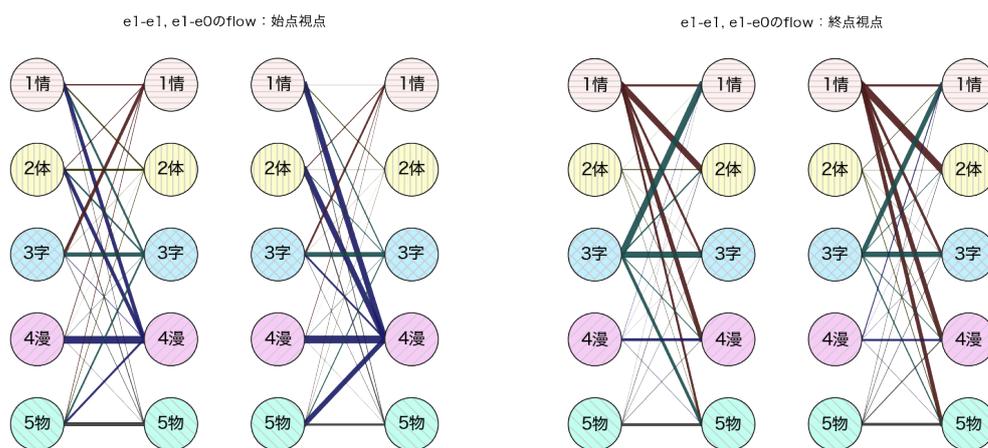


図 23 e1-e1 と e1-e0 での結びつきを始点視点（左）と終点視点（右）で図示したもの

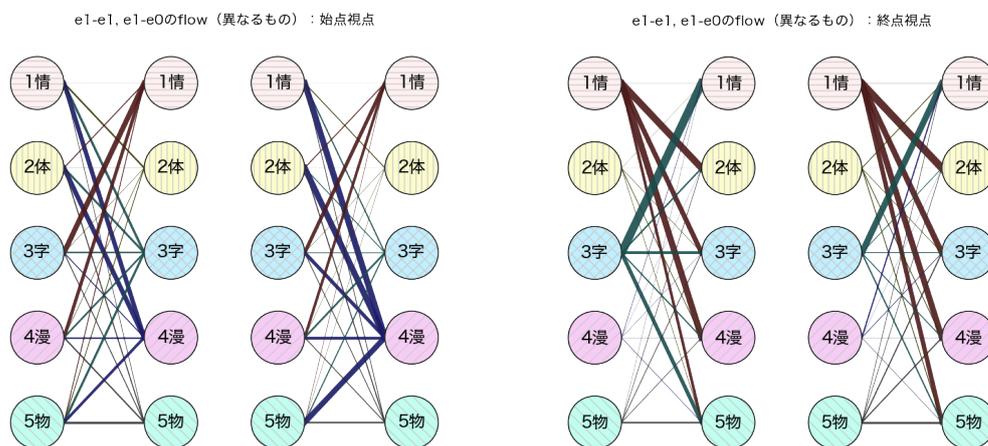


図 24 連続絵文字が異なる場合に絞って、e1-e1 と e1-e0 での結びつきを始点視点（左）と終点視点（右）で図示したもの

図 23 の左側（始点視点）では、漫符系に集まる線の太さが見受けられ、特に「表情系・身体系・物体生物系→漫符系」は e1-e1 より e1-e0 で太いことがうかがえる。逆に終点視点では、e1-e0、e1-e1 ともに「文字符号系-表情系」のフローが見出される。実際、高頻度用例の表（表

12 と表 13) に入らない順位のものも含めて、「感嘆符・疑問符・長音の絵文字+表情の絵文字」といった組み合わせが多く存在していた。

7. おわりに

■**まとめ** 本研究は「加藤安彦携帯メールコーパス」を用いて、特に文中での絵文字位置に着目した分析を行い、以下の結果を得た。

まず、絵文字の位置を「文頭部・文末部・文内部・一文が絵文字のみで成る場合」の四つに分類すると、メール本文中・件名中とも文末部に現れる絵文字が 88% 以上を占めていた。絵文字の連続使用により真の末尾以外に現れたものが全体の 2 割以上あり、その割合は 2005 年前後に上昇傾向が見られた。

絵文字が文頭部に位置する率は本文中で 1% 未満、件名中で 2.6% 程度であったが、文頭での絵文字の使用には全体の傾向 (田中・林 2021) と異なる点があった。頻度上位の絵文字には具体的な物体を表すものや数字を含むものが多く、カテゴリで見ても特に先頭では「表情系」「漫符系」の割合が低く、「物体生物系」の割合が高かった。用例からは、箇条書きのマーカールのようにして情報の構造化に寄与することがうかがえ、そのマーカールとしての絵文字選択の一部は前後の文脈に関わると考えられた。

文末部での絵文字の連続使用については、次のような結果が得られた。まず、末尾と一つ前の接続ペアでは、末尾側 (右側) に来る傾向は「漫符系」が強く、手前側 (左側) に来る傾向は「表情系」が強いといえた。さらに、この接続ペアでのカテゴリ間の結びつきでは「表情系→身体系」と「文字符号系→表情系」、文字符号系・漫符系・物体生物系同士の強さが観察された。

また、「携帯会社名」が送受信で異なる場合の絵文字使用の増加について、田中・林 (2021) の裏付けとなるデータを示した。

■**今後の展望** 今回は、送信年が 2010 年までの携帯メールにおける絵文字を対象とした解析・分析を行った。この結果を各種 SNS 等での結果と比較し、媒体の特徴の違いといった観点から比較することは意義があるだろう⁽⁴³⁾。また、右周辺部・左周辺部 (Beeching and Detges 2014) の観点から、絵文字・カテゴリの右に来やすさ・左に来やすさをさらに分析することが考えられる。さらに、絵文字の句読点的な機能の強さが、カテゴリによって異なる可能性も考えられることから、句読点との類似点と相違点を今後さらに分析することも考えられる。

付 録

本文・件名における各位置タイプでの各カテゴリの割合を示した図 18 の対として、本文・件名での各カテゴリにおける各位置タイプの割合を示す図を載せる。

⁽⁴³⁾ 他メディアでの絵文字の位置に関する報告には、たとえば、絵文字の位置の数え方 (一文におけるものか/メッセージ全体におけるものか、連続絵文字を個別に数えるか/まとめるか) の点では異なるが、WhatsApp のチャットデータを対象とした Sampietro (2016) などがある。WhatsApp はケータイメールとは「メッセージがチャット形式で表示されていく」「件名・本文のような区別がない」といった相違点がある。このようなメッセージアプリでのデータと、本文・件名での位置を比較していくことが考えられる。

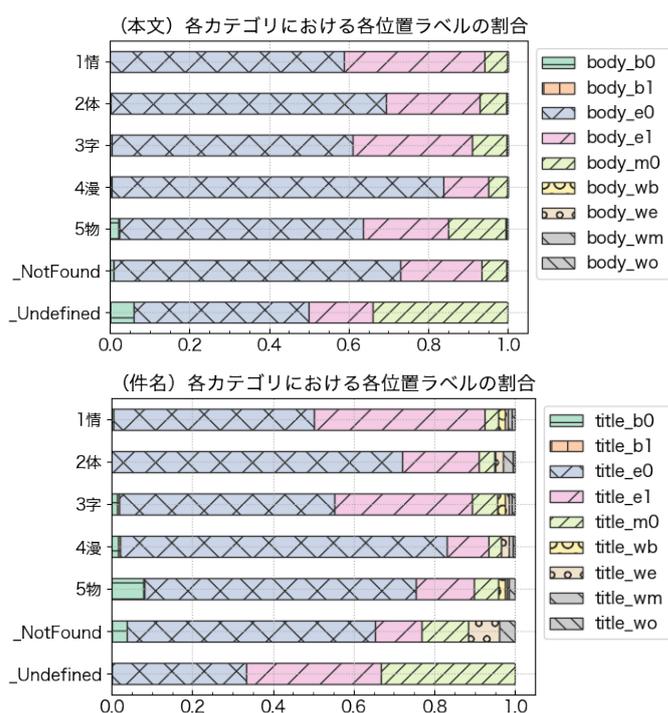


図 25 本文・件名での各カテゴリにおける各位置タイプの割合

謝 辞

本稿では「加藤安彦ケータイメールコーパス」中の絵文字画像の表示に、コーパスに付属する画像ファイルを用いた。また、Unicode の絵文字画像の表示に、Twitter 絵文字ライブラリ (Twemoji) (CC-BY 4.0: (<https://creativecommons.org/licenses/by/4.0/>) Copyright Twitter, Inc and other contributors) を、 $\text{T}_{\text{E}}\text{X}$ の $\text{B}\text{X}\text{c}\text{o}\text{l}\text{o}\text{r}\text{e}\text{m}\text{o}\text{j}\text{i}$ ・ $\text{t}\text{w}\text{e}\text{m}\text{o}\text{j}\text{i}\text{s}$ パッケージ経由で使用した。

本研究の過程では西村綾夏氏、谷口研究室の諸氏を含めた方々に有益なコメントをいただいた。感謝を述べる。

本研究は JST 次世代研究者挑戦的研究プログラム JPMJSP2110 の支援を受けたものである。

文 献

- 田中ゆかり・林直樹 (2021). 「「打ちことば」コミュニケーションにおける絵文字使用: 「加藤安彦ケータイメールコーパス」を用いた分析」 語文, 170, pp. 111–124.
- 三宅和子 (2022). 「2000 年代のケータイメールの実態を捉える: 「加藤安彦ケータイメールコーパス」を利用した研究の可能性」 日本文学文化, 21, pp. 43–56.
- 加藤安彦 (2007). 「ケータイメールにおける顔文字と記号の出現頻度とその関係- ケータイメールコーパスの紹介とともに」 専修国文, 81, pp. 1–17.

- 宮寄由美 (2023a). 「句点」の文末使用の経年変化：加藤安彦ケータイメールコーパスの整備を通して」 専修国文, 112, pp. 1–15.
- 宮寄由美 (2023b). 『加藤安彦ケータイメールコーパス』. <https://www.gsk.or.jp/files/catalog/GSK2023-B/readme.pdf> (2024/07/17 アクセス)
- Mark Davis・Ned Holbrook (編) (2023). “Unicode® Technical Standard #51 Unicode Emoji.” Technical report, Unicode, Inc. Version 15.1, <https://www.unicode.org/reports/tr51/tr51-25.html> (2024/08/12 アクセス)
- Mark Davis・Ned Holbrook (編) (2014). “Proposed Draft Unicode Technical Report #51 Unicode Emoji.” Technical report, Unicode, Inc. Version 1.0 (draft 4), <https://www.unicode.org/reports/tr51/tr51-1-archive.html> (2024/08/12 アクセス)
- 小学館辞典編集部 (編) (2007a). 『句読点、記号・符号活用辞典。』 東京: 小学館
- 文部省教科書局調査課国語調査室 (2007b). 「くぎり符号の使ひ方〔句読法〕(案)」 小学館辞典編集部 (編) 『句読点、記号・符号活用辞典。』 東京: 小学館 pp. i–viii.
- 赤井友紀・山下諒・上間大生・松下光範 (2014). 「漫符を利用した登場キャラクターへの感情付与に関する基礎検討」 電子情報通信学会 HCG シンポジウム 2014 論文集, pp. 294–299.
- 赤井友紀・盛山将広・松下光範 (2017). 「漫符を利用したコミック内の心情・行動描写検索システム」 Technical Report 42, 関西大学大学院総合情報学研究科, 関西大学大学院総合情報学研究科, 関西大学総合情報学部.
- 安岡孝一 (2007). 「ケータイの絵文字と文字コード」 情報管理, 50:2, pp. 67–73.
- 三宅和子 (2012). 「ケータイの絵文字：ヴィジュアル志向と対人配慮」 日本語学, 31:2, pp. 14–24.
- 山崎由佳・西村綾夏 (2024). 「SNS 上の絵文字位置パターンの分析と Levenshtein 距離を用いたサンプル抽出」 言語処理学会第 30 回年次大会発表論文集, pp. 121–126. 言語処理学会.
- Thomas Givón (1985). “Iconicity, Isomorphism, and Non-arbitrary Coding in Syntax.” John Haiman (Ed.), *Iconicity in Syntax*. pp. 187–220.
- 篠原和子・松中義大 (2010). 「視覚的に表出された概念メタファーの解釈：天候記号が感情の推測に及ぼす影響」 日本認知科学会大会発表論文集 27 巻, pp. 745–751.
- Agnese Sampietro (2016). “Exploring the Punctuating Effect of Emoji in Spanish WhatsApp Chats.” *Lenguas Modernas*, 47, pp. 91–113.
- Kate Beeching, and Ulrich Detges (Eds.) (2014). *Discourse Functions at the Left and Right Periphery: Crosslinguistic Investigations of Language Use and Language Change*. Vol. 12. *Studies in Pragmatics*.: Leiden: Brill.

関連 URL

加藤安彦ケータイメールコーパス <https://www.gsk.or.jp/catalog/gsk2023-b/>