

# 感情コーパス作成支援システムの開発

松本 和幸      David B . Bracewell      任 福継      黒岩 眞吾

徳島大学

近年の情報処理技術の発達に伴い、情報処理の分野ではあまり取り扱われることの無かった人間の感性をコンピュータで処理する研究が盛んになってきている。擬人化エージェントや感性ロボットが人のように振舞うためには、人間の感性を認識し、自らの感情を表出することが必要である。感情を認識し、表出する感性ロボットには、ifbot [1] などがある。我々は、感性ロボットに応用するための感情認識技術について研究している [2]。しかし、感情認識の研究は始まったばかりであり、感情認識のために利用できる言語コーパスが少ない。また、そのようなコーパスは人手により作成する必要があるが、感情情報の付与手法やデータのフォーマットなどが統一されておらず、コーパスの構築を行い研究を進めるための環境としては不十分だと考えられる。我々は、感性情報処理の研究のための言語コーパスの作成を支援するシステムの開発を行っている。本稿では感情コーパス作成支援システムの開発概要について述べる。

キーワード：感情コーパス構築，感情認識，自然言語処理

## Development of an emotion corpus creation support system

Kazuyuki Matsumoto      David B. Bracewell

Ren Fuji      Shingo Kuroiwa

Department of Information Science and Intelligent Systems,  
the University of Tokushima

*Abstract*—In recent years, computer automation have developed in various types of industries, making research about processing human sensibility more active. Emotion recognition and expression technologies are needed to create anthropomorphic agents and sensibility robots that behave like humans. The "ifbot" is an example of a sensibility robot which expresses emotions and recognizes emotions. However, language corpora for emotion recognition are small because emotion recognition is still in the primitive stage of research. We need to construct emotion corpora manually in order to progress the research efficiently, but there doesn't exist a unified format or methods for constructing such emotion corpora. We are developing a support system for constructing a large emotion corpus. In this paper, we propose a system which supports making a natural language corpus of tagged emotion information and describe the outline of the system development.

Keywords : emotion corpus, emotion recognition, natural language processing

## 1 はじめに

近年の感性情報処理研究の進展に伴い、感情認識のための情報資源の必要性が高まっている。一方で、感情認識技術に関する一定の評価基準はまだ無く、利用可能なコーパスも限られている。ゆえに、研究者は感情認識技術の研究開発の基盤となる感情コーパスを構築する必要がある。感情コーパスとしては情緒注釈のタグをコーパスに付与する研究 [3] がある。また、タグ付きコーパスを管理するツールとしては、リレーショナルデータベースに文章の形態素解析結果を格納して管理するツール「茶器」 [4] がある。品詞タグ付けのツールとしては形態素解析器「茶筌」 [5] などが利用できるが、感情情報の付与に関してはテキストからの感情抽出に関する研究の歴史がまだ浅く、言語資源などの環境が整っていないためタグ付けツールのようなものが存在しない。我々の研究グループでは、感情認識のための多言語対応訳感情コーパスの構築について研究しており、その第一歩として日英対応感情コーパスの分析 [6] を行っている。この分析において文の形態素解析結果に対して、形態素または単語、イディオムへの感情の種類を付与するという基本的な感情タグ付けが行われている。このような感情コーパスを効率的に作成するには、感情情報タグ付けの見本となる例を準備すべきである。また、基本的な感情語などに対して自動で感情情報を付与することで作業者の負担を軽減させる必要がある。そこで感情情報自動付与のための基準となる感情語辞書の構築も行っている。しかし、文脈に依存する語彙の意味定義を辞書のみで特定することは困難であるので、自動付与結果に対して人手による修正を施す必要がある。そこで我々は感情コーパスの作成手順に一定の基準を設け、基本的な感情コーパスの作成手法を提案する。また、提案手法に基づく一連の作業を複数作業者によって行える感情コーパス作成支援システムを開発する。

## 2 感情コーパス

本研究では、話者または書き手の感情を含む会話文や日記文のような文のコーパスに対して、感情情報を付与したコーパスのことを感情コーパスと定義する。文に対して文全体から読み取れる感情の種類を付与し、単語やイディオムに対しては、それらが文中で示している感情の種類を付与する。さらに、感情を示す単語やイディオムを修飾する語に対しては修飾タイプを付与する。

### 2.1 感情コーパスの基本仕様

ここでは、本研究で構築する感情コーパスの基本的な仕様について述べる。感情コーパスは、文ごとに形態素解析した結果を格納する「形態素解析結果テーブル」と、形態素解析する基の文を格納する「文テーブル」に分類することができる。形態素解析結果と元になる文は、形態素解析結果テーブルの文 ID (Sentence ID) とリンクしている (図 1)

次に、本システムにおいて使用している感情情報の種類について説明する。まず、形態素に対して付与するタグの種類は、表 1 に示す通りである。これらの感情の種類は感情コーパスの種類によって変更可能になっている。また、連続する複数の形態素がひとまとまりになって意味を持つような「イディオム」であることを示すために「イディオムラベル」を設定する。イディオムラベルは、イディオムの先頭であることを示す「Beginning」、また、イディオムの終端であることを示す「End」、Beginning と End の間にある要素であることを示す「Inner」、1 つの形態素で構成されるイディオムを示す「Single」の 4 種類を定義することができる。 [7]

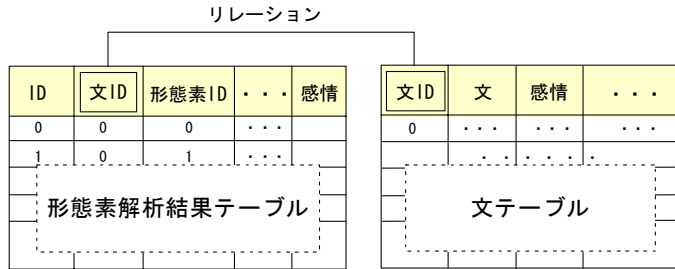


図 1: 形態素解析結果テーブルと文テーブルのリンク

表 1: 基本タグの種類 (一部)

感情の種類	感情タグ	修飾タイプ	修飾タグ
喜び	joy	程度変化なし	neudeg
怒り	anger	程度強め	plusdeg
悲しみ	sadness	程度弱め	minusdeg
恥	shame	否定	negdeg

## 2.2 基本感情辞書

感情情報自動付与の際、感情語と感情イディオムの基本辞書を参照する。基本感情語辞書は「日本語語彙大系」[8]などのシソーラスや辞書から感情に関する語 3000～5000 語程度を抽出し、感情の種類ごとに分類して構成する。感情イディオムの基本辞書は、「感情表現辞典」[9]などから、感情を表現するイディオムで頻出の 2000 種類程度を感情の種類ごとに分類して構成する。

## 3 システムの特徴

ここでは、感情コーパス構築の上で問題となる点を挙げ、本システムではそれらをどのように解決するかについて述べる。まず、タグ修正・付与の作業は、コーパスの量が増えるとコーパス作成作業への負担が大きくなり、コーパス品質の低下を招くという問題が挙げられる。また、感情タグの付与は個人差が出やすく、どのタグを付与すべきか迷ってしまうという問題がある。そこで、本システムではこれらの問題を次のような機能を実装することで解決する。

- 感情タグの自動付与
- GUI を用いた簡単操作
- 複数作業によるコーパスの共同作成

以下、開発中の感情コーパス作成支援システムが持つ特徴について説明する。

### 3.1 感情情報自動付与

本システムでは、感情コーパスへのタグ付け作業の前処理として、基本感情辞書を参照してタグの自動付与を行う。作業者は、自動タグ付与結果が誤っている箇所を修正する。これにより、一からタ

グ付けする必要が無くなり、作業の負担が軽減される。

### 3.2 GUIを用いた操作

本システムのインターフェースには GUI を採用し、簡単な操作性を目指している。(図 2)



図 2: システムの GUI

### 3.3 複数作業による共同作業

本システムでは、コーパスデータベースをネットワーク上のサーバに置くことで、複数作業者が感情タグの付与を共同で行える。共同で作業を行う利点として、作業者が他の作業者の付与した結果を参考にして作業を進めることができるという点がある。システムでは「感情タグ付与履歴提示機能」として実装される。これは、どのような語にどのような感情タグの付与を行ったかという履歴を保存することで、タグ付け頻度のデータなどを算出して作業者に提示する機能である。したがって、作業員または他の作業員が作業時にこのデータを閲覧し、タグ付けの参考にすることが可能である。

## 4 システムの基本構成

本システムは、「データ格納モジュール」、「タグ自動付与モジュール」、「感情コーパス分析モジュール」、「タグ修正・付与モジュール」の4つの処理モジュールから構成される。システム構成図を図3に示す。

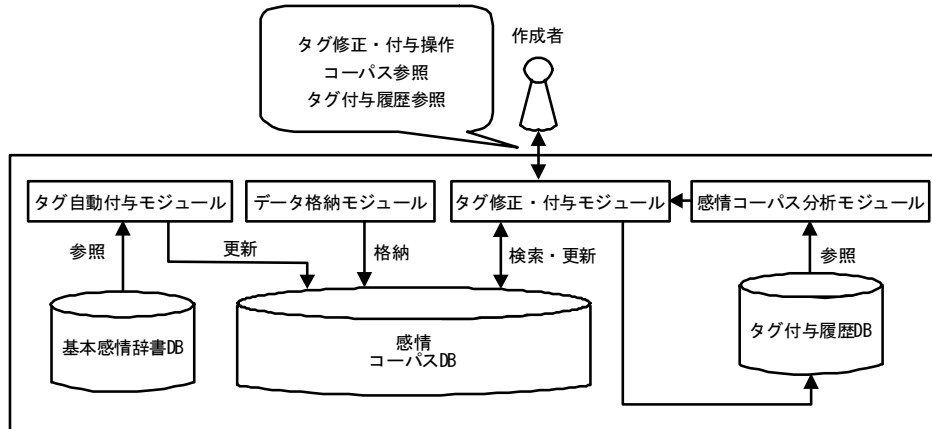


図 3: システム構成

以下，システムの基本構成について説明する．

1. データ格納モジュール

文と文の形態素解析結果をそれぞれ文テーブルと形態素解析結果テーブルに格納する処理を行う．

2. タグ自動付与モジュール

基本感情辞書を用いて感情情報タグを形態素解析結果テーブルに自動付与する処理を行う．

3. 感情コーパス分析モジュール

感情タグの付与されたコーパスに対して統計的な分析の処理を行う．具体的には，ある単語（またはイディオム）に対してどのようなタグが何回付与されたかなどのデータを算出する．感情タグが付与された語・イディオムは，表 2 に示すような頻度テーブルに蓄積されていく．

表 2: 頻度テーブルの例

感情語・イディオム	joy	anger	sadness	shame	regret
嘲笑	0	1	2	5	0
後ろ髪を引かれる	0	0	3	0	10

4. タグ修正・付与モジュール

コーパス作成者がタグ付与履歴を参照しながら，システムによって自動付与された感情情報の修正，また，新たな感情情報の付与を行う操作などを受け付ける．

## 4.1 感情コーパス作成の流れ

提案システムを用いて感情コーパスを作成する流れについて述べる。まず、感情コーパスの元になる文のテキストデータを準備する。このデータを形態素解析し、出力結果をさらにコーパス格納用の形式に変換する。この変換結果をリレーショナルデータベースシステム上の形態素解析結果テーブルに格納し、タグ自動付与モジュールにより前処理としてタグの自動付与が行われる。次に、作業者が自動タグ付与された結果を GUI 上で参照しながら修正を施していく。修正・付与を施した部分はタグ付与履歴として履歴テーブルに格納される。この履歴は作業時に参照することができる。

## 5 まとめ

本稿では、感情情報タグ付きコーパスの作成を支援する感情コーパス作成支援システムの開発概要について述べた。今後の課題を以下に示す。

- 複数感情の付与への対応
- 多言語対訳コーパスへの対応

また、本システムにより構築したコーパスの評価を行い、感情推定システムに応用したいと考えている。

## 参考文献

- [1] 加納 政芳, 吉田 宏徳, 加藤 昇平, 伊藤 英則: 「感性会話型ロボット『ifbot』の表情制御の感情空間へのマッピング」, 情報処理学会第 66 回全国大会, (2000)。
- [2] Kazuyuki Matsumoto, Junko Minato, Ren Fuji, Shingo Kuroiwa: "Estimating Human Emotions Using Wording and Sentence Patterns", IEEE ICIA2005 International Conference, June 27-July 3, (2005)。
- [3] 徳久良子, 乾健太郎, 徳久雅人, 岡田直之: 「言語コーパスにおける感情生起要因と感情クラスの注釈づけ」, 人工知能学会研究資料, SIG-SLUD-A003-2, pp.9-16, (2001)。
- [4] 松本 裕治, 高岡 一馬, 浅原 正幸, 乾 健太郎, 投野 由紀夫, 大谷 朗, Edson T Miyamoto: 「タグ付きコーパスの格納/検索ツール『茶器』」, 言語処理学会第 10 回年次大会発表論文集, pp.405-408, (2004)。
- [5] 松本裕治 他: 「日本語形態素解析システム『茶釜』 version 2.0 使用説明書 第二版」, NAIST Technical Report, NAIST-IS-TR99012, (1999)。
- [6] Junko Minato, David B. Bracewell, Ren Fuji, Shingo Kuroiwa: "Automatically Identifying Emotion from Bilingual Text", 電気関係学会四国支部連合大会論文集 (2005)。
- [7] 田中裕紀, 高村大也, 奥村学: 「文字ベースのコミュニケーションにおける顔文字に関する研究」, 言語処理学会 第 10 回年次大会発表論文集 (2004)。
- [8] 池原 悟, 宮崎 正弘, 白井 諭, ほか: 「日本語語彙大系 CD-ROM 版」, 岩波書店 (1999)。
- [9] 中村 明: 「感情表現辞典」, 東京堂出版 (1993)。