

# 情緒生起情報付き結合価パターン辞書の開発

田中努 徳久雅人 村上仁一 池原悟  
鳥取大学工学部知能情報工学科

{ttanaka,tokuhisa,murakami,ikehara}@ike.tottori-u.ac.jp

## 1 はじめに

言語の意味理解の1つとして、発話された文から話者や登場人物の情緒を推定する技術の実現が期待されている。これに対して既に文中の副詞や形容詞の感情表現性に注目した解析や [1]、動詞の深層格関係に注目した解析 [2] が提案されているが、今後さらに表現の構造に着目した解析方法の実現が期待される。

本稿では、表現の構造から語義を解析する能力が「結合価パターン」にあること [3] に着目して、日本語語彙大系に収録されている全ての結合価パターンに対して情緒の生起に関する情報を付与した情緒推定用の結合価パターン辞書の開発を目的とする。

本パターン辞書の開発における問題点は、1) 情緒生起の原因と語義の関係を定める基準を設ける必要があること、2) 情緒生起の判断には、文脈情報や世界知識を用いて解析しなければならない部分があり、こうした部分と語義を結びつけておく必要があること、3) 作成した辞書の精度と網羅性を確認する必要があることがあげられる。

1) に関しては、[4] の情緒生起の特徴に着目する。[4] では情緒生起の原因が端的な記述で示されており、語義と情緒の対応付けが比較的安定して実施できると予想される。2) に対しては、世界知識や文脈情報を取り込むための条件を辞書に設けることで、語義の解析に続いて意味理解の観点を含んだ情緒推定を行うことを目指す。また、3) の対策として、新たに3名の分析者が分担で辞書の検査を行い、校正を行う。

本稿では、こうした作業を3段階に分けて実施する。そして、辞書の精度を計測するために推定実験を行う。

## 2 情緒推定の方法

### 2.1 情緒推定の原理

本稿における情緒推定とは、文の表す状況において登場人物に「喜び」や「悲しみ」という情緒が生じていることを読み手（計算機）が推定することである。情緒を推定する原理は、情緒の生起する原因を、文の表す状況から解析することである。

ここで、[4] は8つの基本的な情緒である「喜び/悲しみ、好ましい/嫌だ、驚き、期待、恐れ、怒り」について、原因となる状況の特徴分析をしている。

例えば、「喜び」の生起する状況には「現状態は前状態よりも好都合である」という特徴がある。これは「最上位の特徴」である。特徴の下位分類をみると、例えば「獲

(喜び:現状態は前状態よりも好都合である

生理的(内的な快,外的な快)

心理的(目標実現(情報収集(思惑どおり,発見,判明)

計画(立案)

実行結果(完遂,獲得,有効))

対人関係(仲間意識(同意,同感,協力,仲直り)

優劣関係(優越,賞讃,服従,厚遇,保護)))

その他)

図1: 情緒生起の特徴フレーム

得、すなわち、目標実現に必要な物を手に入れた」という具体的な特徴がある。これは「最下位の特徴」と [4] では呼んでいる。

### 2.2 結合価パターンを用いた情緒推定

一般に結合価パターンは、文と適合することで文の表す言語規範としての意味を解析することができる。具体的に [3] は、用言を見出し語とした文型パターンに、解析で得られる意味をあらかじめ付随させている。

そこで、結合価パターンを用いた情緒推定を行うために、結合価パターンに情緒生起の特徴を付随させておく。こうすることで、文と結合価パターンの照合により、文の表す状況に情緒生起の原因が含まれていることが解析できるようになる。そのような結合価パターンは、例えば図2のような情報を持つ。次章では図2のような辞書の作成を行う。

見出し語:	入手する
意味属性:	所有的移動
パターン:	{N1(3主体)がN2(533具体物...)を N3(*)から/より入手する}
前提条件:	{情緒主Frが目標Gをもつ} {目標GはプランPで実現可能} {プランPには対象Nが必要} {対象Nの入手は評価Eが高い}
情緒主:	N1
情緒対象:	N2
原因:	獲得
情緒名:	《喜び》

図2: 情緒属性付き結合価パターン

本稿では、「情緒主」、「情緒対象」、「原因」、「情緒名」、「前提条件」をまとめて「情緒属性」と呼ぶこととする。

### 2.3 前提条件

情緒生起の特徴の認識には、文脈情報や世界知識を用いて解析しなければならない部分がある。これは、結合価パターンによる意味解析を越える範囲である。そのた

め、その部分は意味理解により判断情報を得る必要がある。そこで、情緒推定という意味理解のプロセスで焦点をあてなければならぬ部分を本パターン辞書では前提条件の形式で持つこととする [5]。

そこで、条件の定義は情緒生起特徴に基づいて行う。情緒生起特徴はフレーム構造で形式化されている。例えば図 1 を用いて言えば、最上位特徴である《喜び》の下には中位特徴として 心理的 があり、更に 目標実現、計画 と続き最下位特徴として 立案 がある。まず、効率的に前提条件を語義に対応付けるために、フレーム構造の各段階に前提条件を割り当てる。また、フレーム構造において下位の特徴は上位の特徴を含んでいるので、前提条件も上位の条件を含んでいるものとして作成する。

例として、最下位特徴 立案 の満たすべき前提条件の内容は、次の 4 つである。目標実現 の条件は 2 つ、(1) [情緒主  $Fr$  が目標  $G$  をもつ]、(2) [目標  $G$  はプラン  $P$  で実現可能] である。その下位特徴 計画 の条件は (3) [プラン  $P$  に向けて行動  $Ac$  を実行] である。そして、最下位特徴である 立案 の条件は (4) [行動  $Ac$  の評価  $E$  が非常に高い/高い] である。ここで用いる「評価」とはコストの事を指し、費用や時間の高さや長さ、人力の多さや一般的な価値観など多義に渡る。

このように、中位特徴から最下位特徴までそれぞれに条件を定義し、情緒特徴別に見て全 68 種類の前提条件を作成した。図 3 にその一部を掲載する。

中位特徴	最下位特徴	条件 ID
1 目標実現		情緒主 $Fr$ が目標 $G$ をもつ (1)
		目標 $G$ はプラン $P$ で実現可能 (2)
1.1 目標設定		情緒主 $Fr$ がプラン $P$ をもつ
	1.1.1 自由	行動 $Ac$ はプラン $P$ に有益
	1.1.2 制限	行動 $Ac$ はプラン $P$ に無益
	1.1.3 強制	プラン $P$ は正当
1.2 情報収集		プラン $P$ には対象 $N$ が必要
	1.2.1 判明	対象 $N$ の入手は評価 $E$ が高い
	1.2.3 教訓	行動 $Ac$ はプラン $P$ に有益
	1.2.4 迷信	行動 $Ac$ はプラン $P$ に無益
	1.2.5 嘘	プラン $P$ は正当
1.3 計画		プラン $P$ に向けて行動 $Ac$ を実行 (3)
	1.3.1 立案	行動 $Ac$ の評価 $E$ が非常に高い/高い (4)
	1.3.3 成算	行動 $Ac$ はプラン $P$ に有益
1.4 実行結果		プラン $P$ には対象 $N$ が必要
	1.4.1 獲得	対象 $N$ の入手は評価 $E$ が高い

図 3: 前提条件 (一部)

なお、情緒推定の際の条件判定は、本稿では対象外である。条件判定の機械化については、文脈情報や世界知識から節間キーワードを用いて試行している [6]。

### 3 辞書作成の過程

具体的な辞書作成の過程は、第一段階として「前提条件」以外の情緒属性の付与基準を設け、結合価パターン辞書に情緒属性の付与を試行する。第二段階では、「前提条件」の付与と校正を行う。第三段階は、辞書の精度を高めるために情緒分析の経験のある 3 名の分析者による

検査を行う。

#### 3.1 第一稿

##### 3.1.1 情緒属性の付与基準

第一稿では、情緒属性付きパターン辞書作成の基礎として結合価パターンに情緒属性を割り当てる基準を設ける。

1. 情緒主となる人物がパターン中の変数で現れること。
2. 1 の人物の立場に立ち、用言の語義を考え、生起特徴の中心的な動作・状態を比較すること。
3. 結合価パターンに指定のない限り、様相は標準形で考えること。

ただし、1 つの結合価パターンに複数の情緒属性が考えられるときは、各場合に分けて付与する。

##### 3.1.2 情緒属性の付与結果

以上の条件を基に、14,819 個の結合価パターンに情緒属性を付与したところ、付与基準を満たす結合価パターンは、5,606 個であり、付与した情緒属性は、7,320 個となった。これより、結合価パターンのうち、38% が情緒生起の原因に関連することがわかった [7]。

#### 3.2 第二稿

第二稿では前提条件の付与と共に、第一稿での付与ミスや過不足などを修正するための校正を行う。

#### 3.3 第三稿

##### 3.3.1 検査方法

第一稿および第二稿では一人の作業者が情緒属性の付与を行った。そこで、別の分析者によって辞書の偏りを検査する。分析者は、大規模な情緒注釈つきコーパスの構築に携わった者で、タグ付けの正解率が高かった上位 3 名である [8]。3 名には、第二稿の辞書の内容を 3 分割して、一人当たり 4,940 件のパターンを検査してもらう。分析者には、「何らかの情緒が生起すると思われる箇所」あるいは「すでに付与された情緒属性に疑問を感じた箇所」を指摘してもらう。その後、作業者と分析者との検討で詳細な情緒属性を定義していき、第三稿として編纂する、という手順をとる。

##### 3.3.2 検査結果

結合価パターン辞書の中から 2,863 個の修正必要個所が指摘され、そのうち 2,494 箇所修正を加えた。

369 箇所の用言は、修正を加えなかった。その原因は主に 3 つに分類される。内訳と詳細を以下に示す。

・用言意味属性〔5 属性〕

修正の加えなかった用言のうち、307 箇所は情緒属性付与の難しい用言意味属性〔5 属性〕の用言であった。〔5 属性〕の用言の結合価パターンは例えば、「 $N1(3$  主体) は金遣いが荒い」、「 $N1(3$  主体) は機転が利く」、「 $N1(3$  主体) に決断力が無い」などである。これらの結合価パターンは、情緒主  $N1$  に何らかの情緒が生起する原因となるかもしれないが、この表現に対する  $N1$  の持つ認識によるところが大きいと判断されるため、本稿の範囲では情緒主  $N1$  についての情緒の定義を保留とした。

・感情の直接表現

直接表現と判断したパターンは 25 箇所あり、パター

ン例は、「N1(4人)がN2(\*)に狂喜する」「N1(4人)がN2(\*)に照れる」などである。これらのパターンは感情を直接的に表現したものであるため、元々本稿では対象外とする用言であった。

・その他の原因の用言

その他の特殊な用言としたパターンは37箇所あり、「N1(4人)が死去する」「N1(4人 535 動物)が/の命を落とす」など、情緒主の意識が失われている状態を表す用言である。このようなパターンは本稿では対象外とした。

以上のように、保留あるいは対象外とした箇所とした範囲が数点あるが、基本的な部分は完成したと考える。

## 4 辞書改訂の結果

### 4.1 最上位特徴の分布の推移

基本情緒ごとの情緒属性付きパターンの分布を表1に示す。最終稿である第三稿で全ての情緒が大きく数を増し、14,819個の結合価パターンのうち、52%が情緒属性を持つこととなった。なお、上昇率は 第三稿 / 第一稿 × 100% である。

表1: 最上位特徴パターン分布

	第一稿	第二稿	第三稿	上昇率
喜び	1,887	2,204	2,298	122
期待	717	1,149	1,827	255
恐れ	462	763	1,109	240
驚き	64	76	86	134
嫌だ	1,271	1,556	1,905	150
好ましい	1,656	1,998	2,644	160
怒り	193	457	494	256
悲しみ	1,070	1,250	1,361	127
なし	9,213	8,310	7,060	77

最終的には、「怒り」、「期待」、「恐れ」の3つの情緒の上昇度が特に向上している。この原因は、第一稿では判断の基準が結合価パターンのみで、この3つの情緒が生起する状況の判断が困難であったが、第二稿では前提条件を設けたことで他の情緒よりも情緒が生起する状況が明確になったためだと思われる。

### 4.2 用言意味属性ごとの分布

結合価パターンは日本語の用言の語義を網羅的に収集しており、「用言意味属性」という語義の分類コードが付与されている。このコードは36個あり、深さが4段の木構造を形成している。

用言意味属性別に集計をして、第一稿から密度が急激に増加したものの上位7つを表2に示す。ここでAは当該意味属性をもつ結合価パターンの数を、Bはその中で情緒属性の付与されたパターンの数をそれぞれ表す。また、B/Aは密度を示す。

ここで、密度が増加した〔13 心的状態〕の具体例を見る。第二稿から第三稿で新たに4件の情緒を付与した。4件の結合価パターンを詳しくみると、a)「N1(4人)が/の手をこまねく」b)「N1(4人)が/の腕をこまねく」、c)「N1(3主体)がN2(1000抽象)に直面する」、d)「N1(3

表2: 用言意味属性別

用言意味属性	A	第一稿	第二稿	第三稿	密度の上昇度
		B/A	B/A	B/A	
13 心的状態	9	0.33(3)	0.44(4)	0.89(8)	0.56
35 開始	99	0.24(24)	0.31(31)	0.52(51)	0.28
23 身体動作	1,811	0.28(504)	0.33(602)	0.55(995)	0.27
24 利用	121	0.40(48)	0.57(69)	0.65(79)	0.25
12 思考状態	25	0.52(13)	0.56(14)	0.76(19)	0.24

主体)がN2(1000抽象)に当面する」であった。

a)とb)は類似した用言である。第三稿の段階で付与した情緒属性は無計画による《悲しみ》と共通のものであった。また、c)とd)の用言も類似性があり、思惑違いによる《驚き》など共通していた。第三稿の分析者が見落とすことなく修正の指摘ができていたといえる。

## 5 推定実験

### 5.1 実験の目的

第二稿と第三稿の辞書を用いて、結合価パターンによる推定の精度を計る。

### 5.2 実験対象

情緒推定の対象は、[9]における日記文とする。この文献には、「嬉しかったことを日記に書く」、「羨ましいことを日記に書く」など情感でテーマ分けされた日記文が紹介されている。推定目標文は、以下の条件で選択する。

1. 日記文著者が格要素に現れること(ゼロ型でもよい)。
2. 感情の直接表現や反応の表出の表現を用いた文は対象外。
3. 結合価パターンの存在しない文は対象外。

ただし、「～と言う」、「～と伝える」などは一文として判定する。

以上の条件で、日記文の8テーマから298文を抽出し、そのうちの137文を推定目標文とした。

### 5.3 実験手順

推定目標文と結合価パターンを照合する。その後、情緒属性の前提条件の判定を行う。条件を満たしていれば、その情緒を出力する。そうして、出力情緒と情緒正解との比較を行う。以上を手作業で行う。

### 5.4 情緒正解の作成および正答の基準

実験を行うにあたって、正解基準を定める必要がある。しかし、一般に情緒には正解とよべる単一の答えを定めることは容易ではない。

そこで、情緒の正解として、情緒分析の経験者7名が独立に、各推定目標文にそれぞれの場面で生起すると思われる情緒を、「喜び/悲しみ、好ましい/嫌だ、驚き、期待、恐れ、怒り」の8つの基本情緒と情緒《なし》から選択してもらう。また、異なる情緒が同時に生起する場合もあるので複数回答も認める。

そのうち、正答とする回答は、最も多く票が集まった情緒の上位2位までとする。

### 5.5 実験過程

具体的に実験の様子を図4の3文を用いて解説する。

- 1) 蒲郡のおばさんからみかんが1箱届い<sup>(1)</sup>た。
- 2) とても甘くて
- 3) おいしかった。

図4: 日記文の一例

実験条件から、第1文が実験対象となる。第1文の情緒正解は、《喜び》に5票、《なし》に2票、《好ましい》に1票であったので、3位以下は不当とし、上位2位の《喜び》と《なし》を正解とする。下線部(1)の結合価パターンを本辞書で照合すると、「N1(1647 慶 764 贈り物)がN2(3 主体)に/へ届く」である。情緒属性は、第二稿、第三稿の辞書共に、情緒主N2に生起する情緒は獲得による《喜び》である。しかし、この時点ではまだ情緒を出力しない。

次に、前提条件の判定に移る。獲得の前提条件は図3より、1) [情緒主 Fr が目標 G をもつ]、2) [目標 G はプラン P で実現可能]、3) [プラン P には対象 N が必要]、4) [対象 N の入手は評価 E が非常に高い/高い]、の4つである。情緒主 Fr から対象 N までの変数を手作業で確認する。この例では、全ての条件を満たしているため、辞書に付与した《喜び》を出力する。よって、両辞書とも正解となる。

### 5.6 推定結果

正答率を  $\text{正答数} / \text{回答数}$  で求める。

第二稿の正答率は74%(=102/137)となり、第三稿では正答率は67%(=92/137)となった。

また、A者からG者までの7名の各正解作成者の正答率を他作成者回答との適合数 / 作成者回答数で計算する。以上を表3および図5に示す。

表3: 正答率の比較

推定者	正答率	(正答数/回答数)
第二稿	74%	(102/137)
第三稿	67%	(92/137)
A者	78%	(121/156)
B者	80%	(111/139)
C者	70%	(124/178)
D者	74%	(96/129)
E者	77%	(102/133)
F者	85%	(119/140)
G者	73%	(104/142)

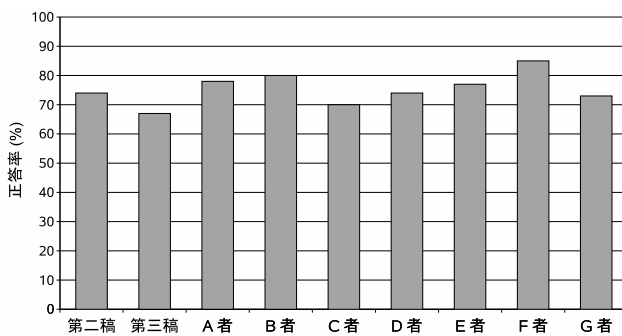


図5: 正答率の比較

## 6 考察

正答率が第二稿と比較して、第三稿の結果が低下した点について考察する。実験対象文の中で、第二稿との出力結果が異なった文数は15文あった。第三稿はその15文の中で、何らかの情緒を出力し、4文は情緒正解と適合し、正解であったが、11文は不正解であった。例えば、「来月から人事部に異動だと言われた。」の1文がある。ここでの情緒正解は《怒り》と《なし》であるが、出力では《嫌だ》となり、不正解であった。しかし、この後の文章は「あそこだけには行きたくないのに…」と続いており、日記著者が嫌だと感じていても不自然ではない。従って、第三稿の辞書の精度が著しく悪いものではないことが推察できる。

## 7 おわりに

文の表現構造に着目した情緒推定の実現に向けて、情緒生起の原因に根ざした情緒属性を結合価パターンに付与した。日本語の主な用言の語義に相当する14,819個の結合価パターンのうち、52%が情緒属性を持つことがわかった。また、前提条件を設けることで世界知識や文脈情報の判断の手がかりとすることが可能となり、文章中の場合に応じた情緒の出力分けが可能となった。

また、日記文を用いた推定実験を行うことにより、情緒属性付き結合価パターン辞書の校正と前提条件の強化を重ねていくことで情緒推定の可能性が向上することを確認できた。

今後の課題は、前提条件で定義した変数と結合価パターン中の変数番号のレベルでの関連づけを行うことである。

謝辞 本研究は科研費(若手(B):17700151)の下で行いました。辞書分析並びに正解作成に協力頂きました研究室メンバーに感謝します。

## 参考文献

- [1] 佐伯他:副詞および形容詞による感情表現性の判定, FIT2003(2), pp.117-118, 2003.
- [2] 目良他:語の好感度に基づく自然言語発話からの情緒生起手法, 人工知能学会論文誌, 17(3), pp.186-195,2002.
- [3] 池原他:日本語語彙大系, 岩波書店, 1997.
- [4] 徳久他:パターン理解的手法に基づく知能エージェントの情緒生起, 情報処理学会論文誌,39(8), pp.2440-2451,1998.
- [5] 徳久他:情緒推定の手がかりについて, 言語処理学会第11回年次大会発表論文集, S3-2, pp.652-655, 2005.
- [6] 古塩他:テキストを対象とした情緒推定のための事象間関係の解析に向けて, 言語処理学会第12回年次大会発表論文集, 2006(発表予定).
- [7] 田中他:結合価パターンへの情緒生起情報の付与, 言語処理学会第10回年次大会発表論文集, pp.345-348, 2004.
- [8] 徳久他:漫画における表情に着目した情緒注釈付き対話コーパスの作成, 電子情報通信学会, 総合大会, AS-4-6, 2006(発表予定).
- [9] 石原:英語で日記を書いてみる, ベレ出版, 2002.