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
鳥取大学大学院工学研究科
{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 [7] 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を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，こうした部分と語彙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為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．

```
(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

たとえば，《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」(LLF; Lowest Level Feature) 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 [7] 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 $N1$ が目標 G を持つ 目標 G はプラン P で実現可能 プラン P には対象 $N2$ が必要 対象 $N2$ の入手はコストが高い		
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の範囲を越えるため，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ない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成化されている．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．各特徴段階において図 3 のように前提条件を定義する．また，最下位特徴は中位の特徴を含んでいるので，前提条件の内容も全て含んでいるものとする．例として，最下位特徴¡立案¿の満たすべき前提条件内容は，" 情緒主 N が目標 G をもつ" かつ" 目標 G はプラン P で実現可能" かつ" プラン P に向けて行動 Act を起こす" かつ" 自己の行動 Act の入手 / 実行はコストが高い" の 4 つである．ここで用いるコストとは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．

中位特徴		最下位特徴	条件 ID
1 目標実現			情緒主 N が目標 G をもつ 目標 G はプラン P で実現可能
→	1.1 目標設定		プラン P は情緒主 N にとって必要なものである
	→	1.1.1 自由	自己/他者の行動 Act はプラン P に有益
	→	1.1.2 制限	他者の行動 Act はプラン P に無益
	→	1.1.3 強制	プラン P は正当なものである
→	1.2 情報収集		プラン P には対象 N が関係している
	→	1.2.1 判明	対象 N の入手はコストが高い
	→	1.2.3 教訓	自己/他者の行動 Act はプラン P に有益
	→	1.2.4 迷信	他者の行動 Act はプラン P に無益
	→	1.2.5 嘘	プラン P は正当なものである
	→	1.2.6 思惑通り	対象 N は目標 G の達成と結び付くものである
→	1.3 計画		プラン P に向けて行動 Act をおこす
	→	1.3.1 立案	自己の行動 Act の入手/実行はコストが高い
	→	1.3.3 成算	自己/他者の行動 Act はプラン P に有益
	→	1.3.5 絶体絶命	自己/他者の行動 Act はプラン P の達成を阻害するものである
→	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 の入手はコストが高い
	→	1.4.3 完遂	自己の行動 Act の実行はコストが高い
	→	1.4.5 有効	自己/他者の行動 Act はプラン P に有益
	→	1.4.8 妨害	他者の行動 Act はプラン P に無益
	→	1.4.9 成功	プラン P の実行はコストが高い
	→	1.4.b ずる	プラン P は正当なものである

図 3: 前提条件 (一部)

なお，条件の判定は機械的に行うが，本稿では対象外であるため手動で行う．条件判定の機械化については，[8] で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．

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，世界知識や文脈情報を扱うための「前提条件」を作成する．第三段階は，辞書の精度を高めるために情緒分析能力に優れた 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 [3]．

3.2 第二稿

情緒推定を行うにあたり，第一稿で作成したパターン辞書のみでは世界知識や文脈情報を扱うことができず，より深い情緒をすることは困難である．

そこで，第二稿では情緒推定時に結合価パターンの適合を確定する前に，情緒が生起する条件や状況を限定するための前提の条件を設け，結合価パターンに付与する．

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 [6]．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「すでに付与してある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，といった手順となる．

3.3.2 検査結果

これによって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 箇所に修正を加えた．

369 箇所の用言は，本稿の対象外や分析者のミスといった原因から修正を加えなかった．原因は主に 3 つに分類

され，その内訳と詳細は以下に示す．

・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」などであり，これらの結合価パターンは，何らかの情緒を生起するが，結合価パターンが示す属性情報が他者からの視点によるところが大きいと判断される為，本稿の範囲では情緒主 N1 についての情緒が定義することができなかった．

・(感情) 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り，新しく情緒が生起すると考えにくいため，本稿では作業対象外とした．

・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におけるものであり，このようなパターンも本稿では対象外とした．

以上のように，今回の作業では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，今後は前提条件を更に拡張・強化することでカバーできる範疇であると考えられる．

4 辞書の改定過程

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．各稿による数の推移も同時に掲載しているが，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，上昇率は第一稿から第三稿を見た時の分布数の上昇度を表す．

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

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．

表 2: 用言意味属性別

		第一稿		第二稿		第三稿		
用言意味属性	A	B	B/A	B	B/A	B	B/A	上昇度
13 心的状態	9	3	0.33	4	0.44	8	0.89	0.56
35 開始	99	24	0.24	31	0.31	51	0.52	0.28
23 身体動作	1,811	504	0.28	602	0.33	995	0.55	0.27
24 利用	121	48	0.40	69	0.57	79	0.65	0.25
12 思考状態	25	13	0.52	14	0.56	19	0.76	0.24
18 物理的移動	511	88	0.17	122	0.24	207	0.41	0.24
16 行動	159	84	0.53	84	0.52	120	0.75	0.22

ここで，〔13 心的状態〕の密度が増加した点について着目してみると，第二稿から第三稿で新たに 4 件ほど情緒を付与した結果となっている．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 主体) が N2(1000 抽象) に当面する」であった．

a と b の用言は類似したものであり，第三稿の段階で，付与した情緒属性も¡無計画¿による《悲しみ》と共通のものであった．また，c と d の用言も類似性を持ち，¡思惑違い¿による《驚き》など共通していたことから，第三稿の分析者の修正指摘が概ね正しいことがわかった．

5 推定実験

5.1 実験の目的

本辞書の実際の精度を計る為に，前提条件の判定は人手で行うという条件下で，第三稿を用いて推定実験を行う．実験は以下の 2 点を主に重視して行う．

- 結合価パターンによる推定の精度
- 前提条件の導入による出力結果の変化

5.2 実験方法

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．本研究では，8 テーマから 301 文を抽出し，推定実験の対象とする．

また，本稿の実験は，

(A) 非情緒生起文も実験対象に含める
(B) 「-」と言う，「-」と伝えるなどは一文として判定

と，定義して実験を行う．(A) は非情緒生起文であっても情緒が生起しないことを正しく出力できるかチェックするものであり，(B) は結合価パターンの構造上一文として判断することとする．

しかし ,本稿の付与基準では情緒主となる変数がパターン中に現れることを原則として付与を行ったので，日記著者が動作主でない文や，所有格が動作主である文 (体重が... など) は結合価パターンの枠を逸脱するために実験対象外とする．また，結合価パターンそのものが存在しない為にパターン辞書と適合しない文や，(情緒) 直接表現文も本稿の作業外の範囲である為，対象外とする．

これらを除いた 134 文を推定目標文とした．

5.3 情緒正解の作成

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，情緒研究に携わる 5 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，最も多く票が集まった上位のうち最大で 3 位までの情緒を正解とする．

5.4 実験過程

実際に情緒推定を行うに当たり，推定目標文の用言からパターンの照合後に ,前提条件の判定を行った上で ,情緒属性情報を出力し ,情緒正解との比較を行う．なお ,今回の実験では第三稿の結合価パターン辞書を用いて行うが，精度の比較として第二稿も用いて行う．

具体的に実験の手順を図 4 の 5 文を用いて解説する．

1) 今日，英語の小テストがあった
2) そして午後は歴史のテストを受け，
3) 科学の小テストが返って来た
4)1 日に何度も自分が判断されている
5) なんだかそんな感じだ

図 4: 日記文の一例

情緒正解を定めるに当たり ,1 行目と 3 行目，及び 5 行目は実験対象外の文章となり，推定目標文となるのは 2 行目と 4 行目である．

正解作成の回答では ,2 行目で生起する情緒は ,《嫌だ》だとする回答が 5 票 ,《恐れ》であるとする回答が 1 票であった．回答の上位 3 位を正解とするので，ここでの情緒正解は《嫌だ》あるいは《恐れ》であるとする．4 行目の正解はこれに習って《嫌だ》である．

文数	情緒正解	第三稿出力	第二稿出力
1	対象外	-	-
2	嫌だ (5) 恐れ (1)	恐れ	なし
3	対象外	-	-
4	嫌だ (4)	なし	なし
5	対象外	-	-

図 5: 推定結果 (一部)

2 行目と 4 行目に対して情緒推定を行う．まずは，2 行目の用言「受ける」からパターンの照合に入る．2 行目に適合する結合価パターンは「N1(4 人) が N2(1426 試験) を受ける」であり，情緒属性は情緒主が N1，生起する情緒が¡成算¿による《期待》と¡試行¿による《恐れ》の 2 種類である．

ここで ,前提条件の判定を行い ,文脈情報から判断して¡成算¿の条件はこれを満たさず ,¡試行¿の条件はこれを満たすため，ここでの出力は¡試行¿による《恐れ》となる．

正解との比較を行うと ,2 行目でのパターン辞書からの出力は第二稿では《なし》であるが，第三稿では《恐れ》となっているので第三稿では正解である．

また ,4 行目では第二稿・第三稿ともに《なし》と出力され，不正解である．

図 5 に例の結果を掲載する．

5.5 推定結果

推定目標である 134 文を判定したところ，第二稿のパターン辞書を用いた実験では 72 文が適合した．なお，正解率は適合文数/推定目標文数で計算する．

よって第二稿の正解率は 72/141=51% となる．第三稿を用いると，適合した文数は 83 文となり，正解率は 83/134=62% となる．

6 考察

実験の結果 ,情緒出力と正解とが異なる場合に ,用言 +「できる」など助詞による情緒の変化や，接続詞によって前文からの情緒が継続されるもの，または全く逆の情緒が出力されるものがあった．

このことから見て，今後は結合価パターン辞書の強化に加えて，助詞や接続詞が情緒に及ぼす影響を調査することが必要であると言える．

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．同時に ,助詞や接続詞 ,前提条件により深い定義付けを行うことで更なる数字の向上が期待できるものと思われる．

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との関連づけを行うことである．

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, 人知論誌,17,(3),pp.186-195,2002.

[3] 田中他:結合価パターンへの情緒生起情報の付与, 言語処理大会,2004.
[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, 情処論誌,39,(8),pp.2440-2451,1998.
[5] 徳久他:,NLP,2005.
[6] 徳久他:, 信学会,2006.
[7] 池原他:日本語語彙大系, 岩波書店,1997.
[8] 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, 言語処理大会, 発表予定,2006.
[9] 石原:英語で日記を書いてみる, ベレ出版,2002.