

## 対応分析とFrameNetを用いた共起語の意味分析：状態変化を表す連結動詞を例に

内田, 諭  
九州大学

<https://hdl.handle.net/2324/1932609>

---

出版情報 : The Institute of Statistical Mathematics Cooperative Research Report. 381, pp.39-50, 2017-03-27. The Institute of Statistical Mathematics

バージョン :

権利関係 :

# 対応分析と FrameNet を用いた共起語の意味分析：

## 状態変化を表す連結動詞を例に

### Semantic Analysis of Collocations Using Correspondence Analysis and FrameNet: A Case Study on Linking Verbs

内田 諭  
九州大学

Satoru UCHIDA  
*Kyushu University*

#### 1. はじめに

共起語（コロケーション）の分析は、コーパスの進化とともに多くの研究が行われてきた。Sinclair (1991), Stubbs (1995)などの言語研究を皮切りに、言語教育（Lewis (ed.) 2000など）、文学研究（田畑2009など）、言語処理（松尾・石塚2002など）など、様々な分野に渡って研究が進められている。その重要性はますます高まっていくと考えられるが、共起語の意味の側面については十分に研究が進められているとは言えない。

本研究は、特定の単語の共起語を意味的に分析し、その特徴を明らかにすることを目的としている。具体的には状態変化を表すとされる *become*, *fall*, *get*, *grow*, *turn* の5つの連結動詞を例に取り、これらと共起する形容詞が意味内容の点でどのように異なるか（あるいは類似するか）を明らかにする。統計的手法として多変量解析の一つである対応分析を用いて共起語の分布と特徴を考察する。本研究の特徴は、FrameNet(<https://framenet.icsi.berkeley.edu/fndrupal/>)を用いて共起語が喚起するフレームを特定し、それらをフレームによってグルーピングすることで共起語の意味特徴（フレームの特徴）を明らかにするという点である。以降、2節でフレーム意味論とFrameNetについて概観し、3節でフレームを用いた対応分析の手法と結果を提示する。4節はまとめである。

## 2. FrameNet 概要

フレーム意味論は Fillmore (1982, 1985 など)によって提唱された言語理論で、意味理解はフレームと呼ばれる背景知識を下敷きにして行われると仮定する。例えば、「買う」という単語の理解を達成するには、「売り手」が「買い手」に「代金」と引き換えに「商品」を移譲するという一連の商取引に関する知識が必要である（フレームに登場する「売り手」や「買い手」などの要素はフレーム要素と呼ばれる）。同様のことは「売る」という単語にも当てはまり、「買う」と「売る」は同一のフレームを下敷きにした視点の異なる表現である。

フレーム意味論では語句がフレームを喚起 (evoke) すると考える。それぞれの語句が「どのフレームを喚起するか」ということを記しているのが FrameNet である。FrameNet はオンラインで構築が進められているいわば「フレームの辞書」で、2017年2月現在約1200のフレームとそれらを喚起する約13000の Lexical Unit(おおよそ語義に相当する)が収録されている。詳細な説明は Ruppenhofer et al. (2010)、内田(2015)に譲るが、最大の特徴は単語がフレームの観点から記述されているということである。例えば、ready (adj)は Activity\_ready\_state フレームを喚起すると規定され、このフレームは他に prepared (adj), set (adj)などによっても喚起される。つまりこれらの単語はフレームを介してグルーピングでき、さらに「活動の準備が整った状態を表す形容詞」という意味的な分類名を付与することができる。

フレームによるグルーピングの特徴は、反意語が同じグループに属するという点である。これはフレームがそれに含まれるフレーム要素によって規定されているからであり、例えば increase (v)と decrease (v)は Change\_position\_on\_a\_scale フレームを喚起すると規定されているが、このフレームは Item, Initial\_value, Final\_value, Difference などのフレーム要素を含む。数量が増える場合も減る場合もこれらの要素は共通であるため、増減を表す表現はこのフレームを喚起するということができる（他に drop (v), fall (v), grow (v), jump (v)など）。

フレームはフレーム間関係によって互いに結び付けられており、この関係をたどることでさらに一般化した単語のグルーピングが可能となる。フレーム間関係で最も結びつきが強い関係は Inheritance で、この関係を利用することでフレームの意味タイプを特定することが可能となる（その他の関係については Ruppenhofer et al. 2010 を参照）。例えば、Activity\_ready\_state フレームは図1が示すように上位に Process\_initial\_state フレームを持っており、さらにこのフレームはその上位に State

フレームを有する。したがって、Activity\_ready\_state フレームを喚起する単語は、より一般的には State を表すと言うことが可能になる。

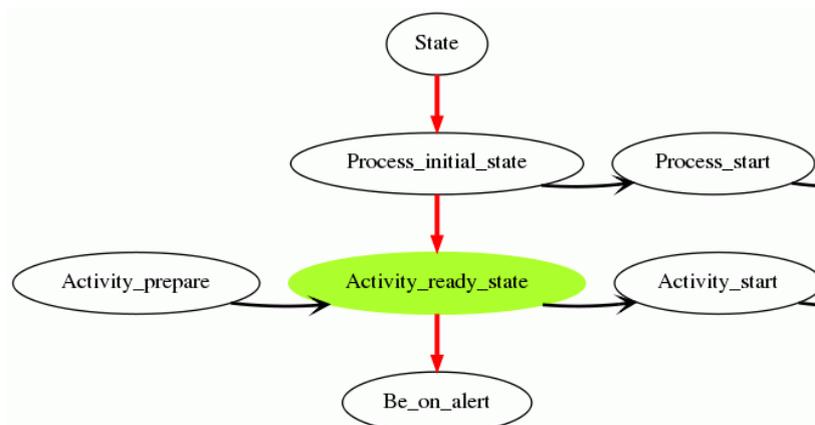


図1 Activity\_ready\_state フレームのフレーム間関係 (FrameGrapher より)

### 3. 共起語の分析

#### 3.1 分析対象

本研究の対象は状態変化を表す become, fall, get, grow, turn の5つの連結動詞である。これらの動詞は学校文法では第2文型を表す動詞に類されている。井上(2016)はコーパスでの検索結果を基に fall と become の補語になる名詞と形容詞について考察し、fall は「出来事の成立が主語の意思に関わらないことを暗示する文脈で用いられる」と指摘しており、『ウィズダム英和辞典』（第3版, 2012 三省堂）には fall の項でこの内容の注記が記載されている。また、大谷(2015)は同様にコーパスでの観察から fall, grow, turn の共起形容詞の違いを指摘しており、fall は「静かで力がない状態を表す形容詞」、grow は「体に起こる変化」や「目に見える体積の増加を表す形容詞」、turn は「色の変化を表す形容詞」とそれぞれ共起する傾向があることを指摘している。

本研究では、become, fall, get, grow, turn の5つの動詞と共起する形容詞をフレームの観点から分析し、先行研究の指摘の正当性を意味の観点から立証を試みる。フレームを用いることのメリットは、意味内容を具体化できるということである。また、FrameNet を用いることで一定の基準に基づいた体系的な分析が可能となる。

### 3.2 データと手法

本研究で用いるデータは2016年版の *Corpus of Contemporary American English* (以降 COCA) である。COCA は 1990 年から 2015 年の各年の出版物やオンラインの記事などをそれぞれ約 2000 万語ずつ含む汎用コーパスで、総語数は 5 億 2000 万語以上に及ぶ。本論文では *become, fall, get, grow, turn* をレマ形で検索し、その直後にくる共起形容詞 ([j\*]) を集計した。それぞれの動詞の頻度を基に相対化し、1 万語あたり 50 以上の頻度のあるものを分析の対象とした (111 件)。

分析対象は 111 (共起語) × 5 (分析対象の連結動詞) の分割表であるが、変数が多く目視での特徴抽出は難しい。本研究では多変量解析の手法の一つである対応分析を用いてそれぞれの動詞と共起語の関係性を見出すことを試みる。対応分析は言語研究で広く用いられている手法であり、その有効性が示されている (後藤 2007、田畑 2009 など。主成分分析との違いについては水本 2009 に詳しい)。視覚的な出力が可能で、関連する項目が 2 次元の座標軸上に近接して配置される。

### 3.3 フレームの付与手順

共起語の意味内容を分析するため、それぞれの形容詞に対して FrameNet の情報を基にフレームを付与した。例えば、*angry* は FrameNet では *Emotion\_directed* フレームを喚起すると規定されているため、フレーム名に置き換えて記録した。この他、*bored, excited, upset* なども *Emotion\_directed* フレームを喚起するため、これらの単語をこのフレーム名のもとに分類することが可能となる。多義語についてはノードとなる動詞との組み合わせで意味を類推し、フレームを決定した。例えば、*tough* は難易度を表す *Difficulty* フレームと物の丈夫さを表す *Level\_of\_force\_resistance* フレームを喚起するが、*get* とのコロケーションであれば前者のほうが適切だと判断できる。なお、フレームの決定にあたっては FrameNet のアノテーションレポートなども参照し、慎重に判定した (*tough* の場合、*Difficulty* フレームのアノテーションに *get tough* の用例が含まれている)。なお、FrameNet に登録のない形容詞についてはフレーム分析の対象外とした。例えば、*golden* は FrameNet に登録がない。この単語は *black* などと同様に *Color* フレームを喚起すると考えられるが、一貫性を保つため、分析対象からは除外した。その他、*available, obsessed, pregnant* など 15 単語が対象外となった。

### 3.4 結果と考察

分析対象となるデータに対して、統計ソフトの R (ver. 3.3.1) を用いて対応分析を行った。対応分析の計算には MASS ライブラリの `corresp` 関数を用い、描画には `wordcloud` パッケージの `textplot` 関数を使用した。以下、(1)単語ベースでの分析、(2)フレームベースでの分析、(3)become, get, grow の 3 語に絞ったフレーム分析の順に結果を提示し、さらに 3.5 節で上位フレームレベルでの分析結果について考察する。

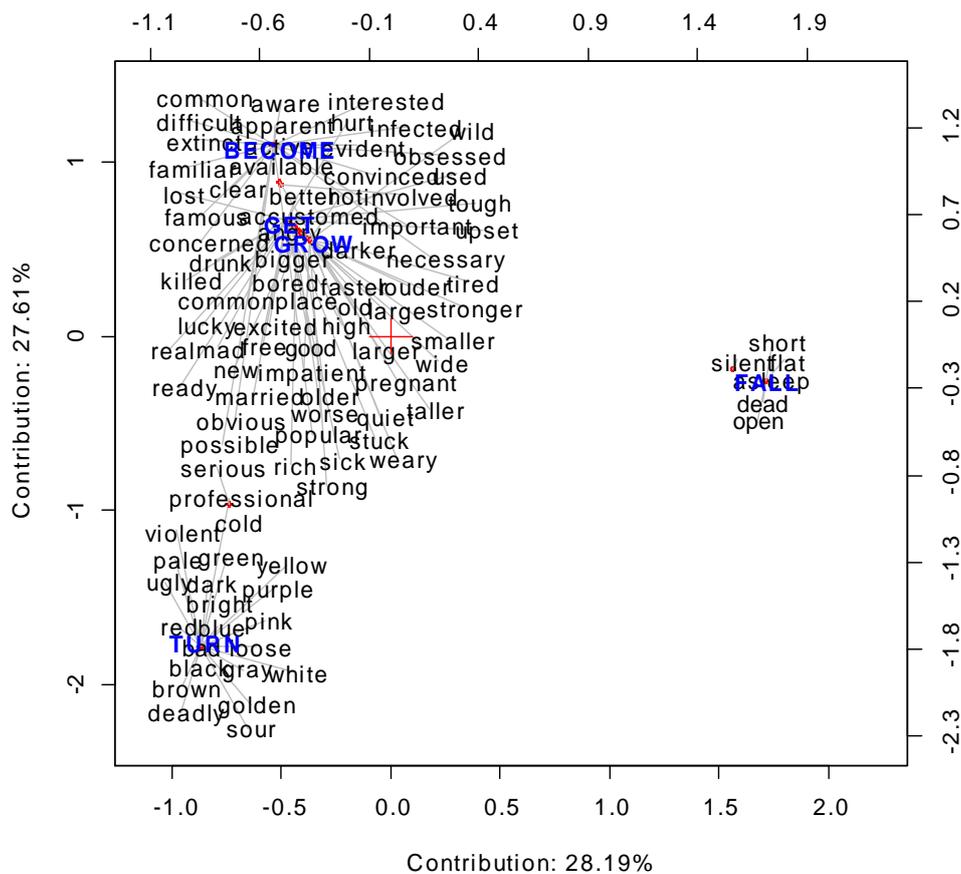


図 2 単語単位での対応分析の結果(nf=5)

図 2 は単語の相対頻度による対応分析の結果で、中心語とコロケーションの関係性を描画したものである(第 1 次元の寄与率は 28.19%、第 2 次元は 27.61%)。この結果から読み取れることは、fall と turn の特異性と become, get, grow の類似性である。特に fall のみが第 1 次元がプラスになっており、もっとも特異な振る舞いをするのが読み取れる。その共起語は silent, short, dead, asleep, flat, open などである。fall flat は転倒することを、fall open 啞然として口が開くこと

を意味するが、それぞれ慣用的な側面が強い。turn は brown, pink, white, yellow などの色を表す形容詞が目立つ。become, get, grow についてはこの図からは判別が付きにくいいため、後半でこれらを独立した対応分析を行い議論する。

単語をそのまま使う場合のメリットは、中心語と実際の共起語を対応付けることができるということである。一方で単語の意味内容の検証については分析者の主観に委ねられる。また、項目数が多いため、視覚的に判別しにくいというデメリットも挙げられる。

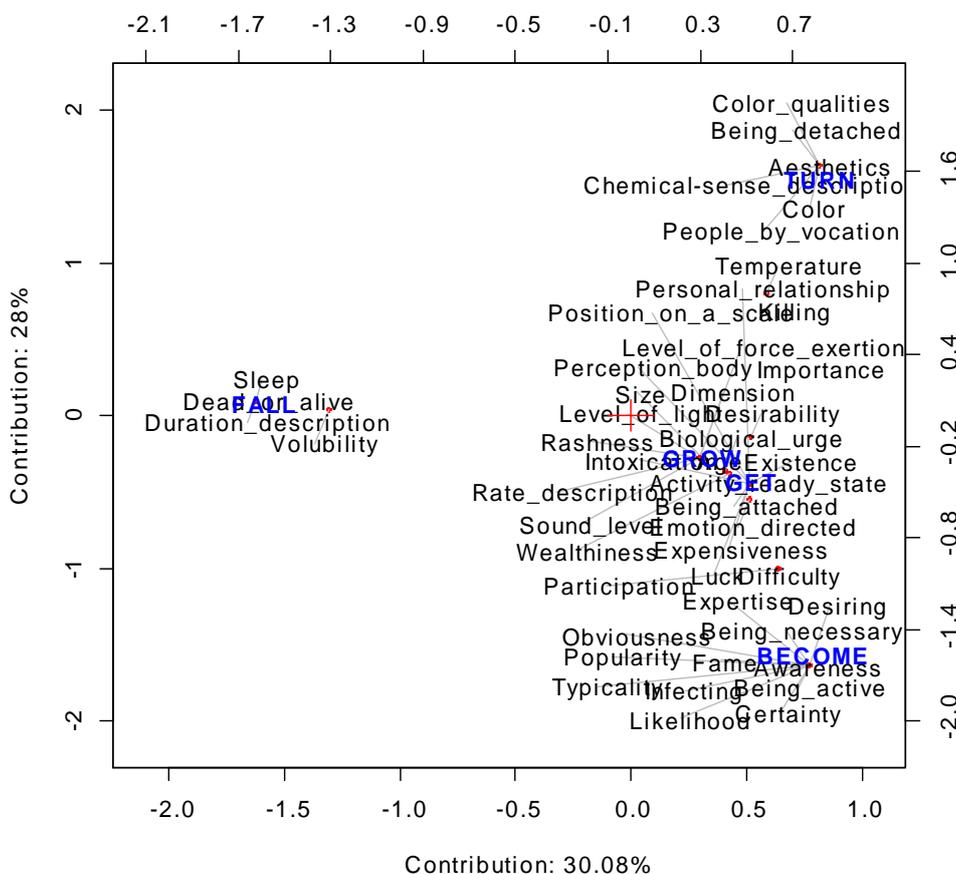


図3 フレーム単位での対応分析の結果(nf=5)

図3は単語に付与したフレームをベースに対応分析を行った結果である。寄与率は第1次元が30.08%、第2次元が28%となり、僅かではあるが説明率の向上が見られる。図2と上下左右が反転した形であるが、中心語の配置関係に変化はなく、フレームベースの対応分析においても中心語間の関係性は保持されているといえる。単語ベースの場合と同様、fallの特異性が目を引くが、そのフレームはSleep, Volubility, Dead\_or\_aliveなどであるとわかり、その意味内

容がより具体的にわかる。turn に注目すると、Color\_qualities, Color が近接することから色を表す形容詞が特徴的であることが読み取れる。他に Aesthetics (ugly), Chemical-sense\_description (sour), Being\_detached (turn loose で解放するという意味の慣用表現)などが turn と関連性が強い。

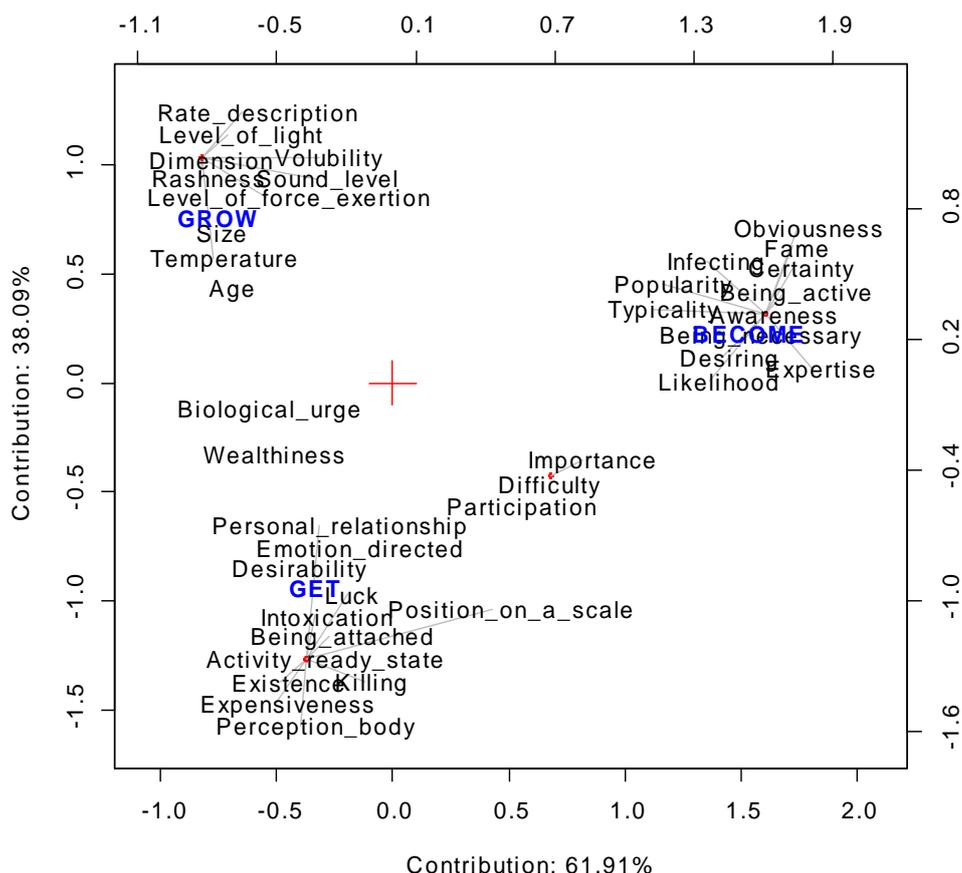


図 4 become, get, grow の共起フレームによる対応分析の結果(nf=3)

図 4 は become, get, grow を取り出してフレームベースで分類したものである(寄与率は第 1 次元 61.91%、第 2 次元 38.09%)。図 2、3 では近接していた 3 つの語だが、この分析によりそれぞれの共起語が喚起するフレームによって峻別できていることが読み取れる。become は特に Certainty, Likelihood, Obviousness フレームとの近接から蓋然性や判断に関わる形容詞が多いことがわかる。get で特に注目すべき特徴的なフレームは Emotion\_directed、Desirability など感情や願望を表す意味内容が多いことが示唆される。その他、Perception\_body, Expensiveness フレームなどが get ラベルの付近に観察される。

また、grow では Size, Temperature, Age, Rate\_description などスケールを伴う形容詞と特徴的に共起することが読み取れる。

### 3.5 上位フレームでの分析

これまでの分析の結果、フレームレベルの分析においても単語の関係性が十分に維持され、かつ意味内容をより具体的に検討できることがわかる。本節では、さらに共起語を大きな分類で考察するため、上位フレームのレベルで分析を行った。フレームの一般化には Inheritance の関係のみを用い、可能な限り上位のフレームをたどった。例えば、Age フレームは Measurable\_attributes -> Gradable\_attributes -> Attributes と上位にたどることができるので Attribute タイプとして分類した。なお、Inheritance の関係で上位フレームがないフレーム (Emotion\_directed フレームなど) に関しては下位のフレーム名をそのまま用いた。分析に用いたデータを表 1、結果を図 5 に示す。

表 1 上位レベルでのフレームの集計結果

Super frames	BECOME	FALL	GET	GROW	TURN
Attributes	2710	3205	2945	5654	5848
Being_detached	0	0	0	0	228
Being_necessary	129	0	0	0	0
Dimension	0	0	0	246	0
Emotion_directed	100	0	706	137	0
Expensiveness	0	0	103	0	0
Experiencer_focused_emotion	198	0	0	0	0
Importance	145	0	128	0	257
Mental_activity	663	0	0	0	0
Objective_influence	119	0	130	0	152
Participation	423	0	484	0	0
People	0	0	0	0	157
Perception_body	0	0	256	0	0
Personal_relationship	0	0	102	0	0
State	0	5286	1192	0	0

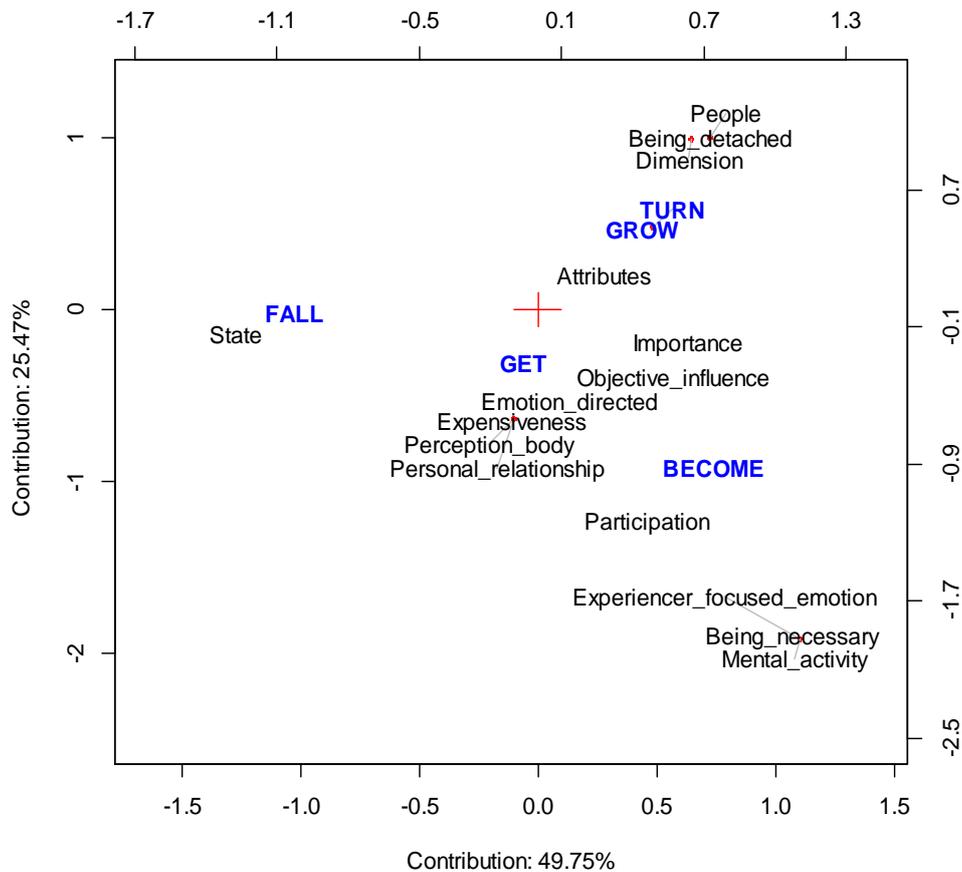


図5 上位フレーム単位での対応分析の結果(nf=5)

単語のグルーピングが大きくなったことで図5は図2, 3などと比べて可読性が高く、寄与率も非常に高い（第1次元：49.75%、第2次元25.47%）。表1をみると、どの動詞も **Attributes** タイプのフレームが多く、このラベルはそれぞれの動詞の中間点に配置されていることがわかる。このことは **become, fall, get, grow, turn** の全ての動詞が属性を表す形容詞と共起することを示し、これらの動詞に共通性があることが示唆される。

次に、それぞれの動詞の特徴を見ていくと、**fall** に関しては **State** タイプのフレームともっとも密接な関係にあることがわかる。**get, become** に関しては前述の観察とほぼ同様の結果を読み取ることができる。**grow, turn** については解釈に注意が必要である。これらは **Being\_detached, People, Dimension** に近接しているが、表1から前者2つについては **turn** のみ、**Dimension** については **grow** のみに出現することがわかる。この結果は変数を圧縮したことにより大きな値を持つ項目が出現し、その影響が大きくなったことによる。ここでは **Attributes** の値が

grow : 5654 件、turn : 5848 件と近い値になっており、Dimension などのフレームにおける頻度の違いよりこちらが重みを持つ結果になっている。なお、図 6 は Attributes フレームを排除した結果である。fall, get, become の結果は同様の解釈ができるが、grow は Dimension タイプのフレームとの親和性がもっとも高く、他の動詞とは異なる振る舞いを示すことがわかる。

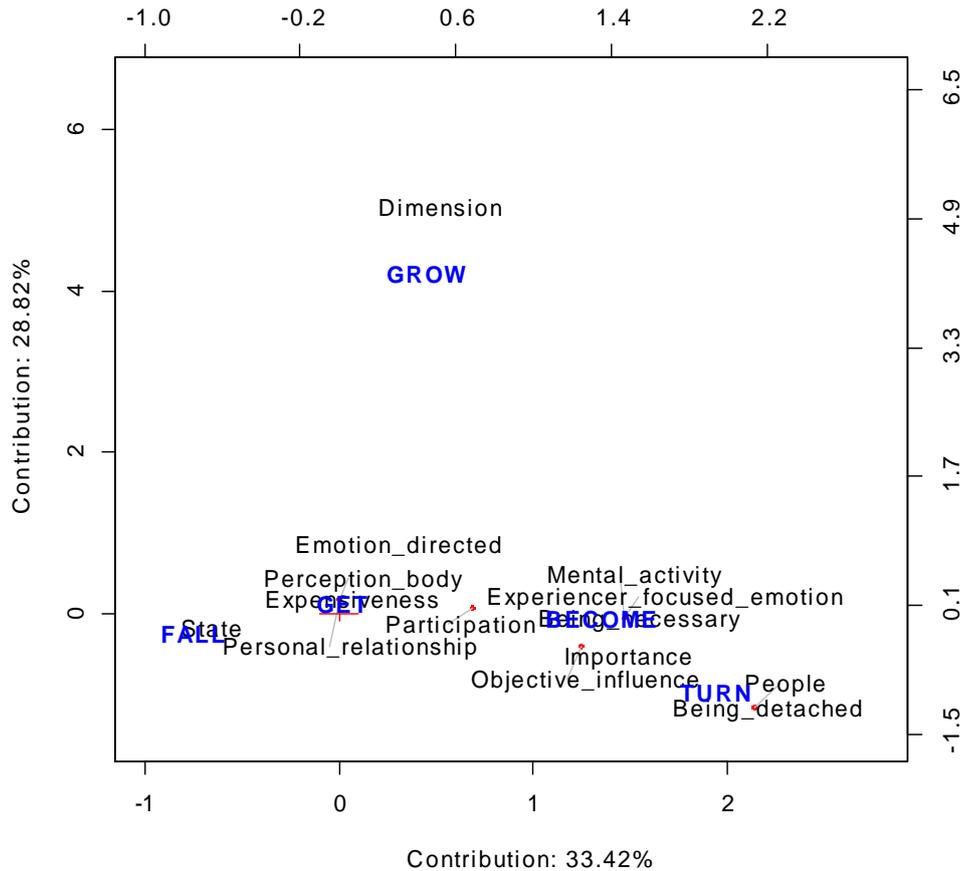


図 6 Attributes を排除した場合の対応分析の結果(nf=5)

#### 4. まとめ

本論文では状態変化を表す 5 つの連結動詞 become, fall, get, grow, turn を例に、その共起語の意味分析をフレーム意味論の観点から行った。フレームを用いることで、体系的な分析が可能となり、意味内容の具体化が可能となった。また、単語のグルーピングに FrameNet を用いることで対応分析等の統計的手法を用いる際に項目を集約して変数を減らすことが可能になることを示した。上位フレームによる分

析は、より一般的な意味タイプで分析できることが明らかになったが、一方で過剰な一般化の危険性があることも指摘した。

今後の課題として、本研究の手法を他の単語の共起語分析に応用可能かどうかを明らかにすること、フレームを用いた単語のグルーピングの際にどのレベルが最適かを検証すること、意味タイプの一般化で *Inheritance* 以外のフレーム間関係の有効性を調査することなどが挙げられる。

## 参考文献

- Fillmore, C. J. (1982) Frame semantics. In Yang, I. (ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981*. pp.111-137. Seoul: Hanshin.
- Fillmore, C. J. (1985) Frames and the semantics of understanding. *Quaderni di Semantica*, 6 (2). pp.222-254.
- 後藤一章 (2007)「統語機能別頻度分布に基づく名詞の特徴的コロケーションの発見」『多変量解析を用いたテキスト分析研究』統計数理研究所共同研究レポート 201. pp.1-23.
- 井上永幸 (2016)「語法研究の要—副詞：コーパスを活用した辞書編集の立場から」『英語語法文法研究』第 23 号. pp. 20-35.
- Lewis, M. (ed).(2000) *Teaching collocation: Further development in the lexical approach*. Hampshire: HEINLE Cengage Learning.
- 松尾豊・石塚満(2002)「語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム」『人工知能学会論文誌』17(3). pp.217-223.
- 水本篤 (2009)「コーパス言語学研究における多変量解析手法の比較—主成分分析 vs. コレスポンデンス分析—」『コーパス言語学研究における量的データ処理のための統計手法の外観』統計数理研究所共同研究レポート 232. pp.53-64.
- 大谷直輝(2015)「認知言語学を用いてコーパスから意味を探る：入門編」『Lingua』研究社 WEB マガジン <<http://www.kenkyusha.co.jp/uploads/lingua/prt/15/OtaniNaoki1502.html>> [Accessed: Feb. 7, 2017].
- Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. R. Johnson and J. Scheffczyk (2010) *FrameNet II: Extended Theory and Practice*. <<https://framenet2.icsi>.

berkeley.edu/docs/r1.5/book.pdf> [Accessed: Feb. 7, 2017].

Sinclair, J. (1991) *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stubbs, M. (1995) Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1). pp.23-55.

田畑智司 (2009) 「Gentleman in Dickens—多変量アプローチで見る文体意匠としてのコロケーション—」 『多変量アプローチによるテキストの計量研究』 統計数理研究所共同研究レポート 231. pp.1-22.

内田諭 (2015) 『フレーム意味論に基づいた対照の接続語の意味記述』 福岡:花書院.