# FrameNet からの動詞階層構造の抽出

岩本蘭 [1]　小原京子 [2,3]
[1] 富士フイルムビジネスイノベーション株式会社　[2] 慶應義塾大学 [3] 理研 AIP
ran.iwamoto.ih@fujifilm.com　ohara@hc.st.keio.ac.jp

## 概要

単語の階層構造は含意判定や情報抽出にとって有用な情報である。しかし応用タスクでの使用に適した動詞の階層構造に関する語彙資源は少ない。本研究では言語表現の意味をフレームを用いて記述する語彙資源 FrameNet のフレーム間関係の 1 つである Inheritance に着目し、フレームの階層構造を可視化した。そして、フレームを喚起する英語/日本語の動詞を抽出し、動詞階層構造のデータセットを作成した。また、階層構造の表現に適した Poincaré embedding を用いて英語/日本語動詞ペア間の階層構造の有無を判定するタスクを解き、70%以上の F 値を達成した。

## 1　はじめに

言語表現の背景にある常識や知識は自然言語処理の応用タスクでの性能向上に無くてはならないものである。その中でも、単語の階層構造、上位語下位語に関する知識はオントロジー構築 [1] や質問応答 [2]、意味検索 [3] にとって有用である。単語の階層構造を考慮したモデルを作成するため、単語の抽象性に関する語彙資源の作成 [4, 5, 6] と階層構造を埋め込むベクトル表現の開発 [7, 8]、つまりデータと手法の両面から研究が進められている。

単語の中でも名詞の上位下位性に関するデータセットは数多く存在するが、動詞については整備が進んでいない。動詞の階層構造のデータセットとして代表的なものとして WordNet [4, 5] がある。WordNet では、*communicate* と *talk*、*talk* と *whisper* のように、ある動詞と、その動詞より具体的な概念を表す動詞 (troponym) の関係性が記述されている。しかし、WordNet の troponymy 関係では、関係構造がより複雑な部分ほど階層構造が正しく設定されていないという問題が指摘されている [9]。WordNet の名詞、動詞の階層構造を用いて語義曖昧性解消を行った研究 [10] では、名詞に比べて動詞の語義曖昧性解消の性能が明らかに低い結果となっており、応用タスクに用いることができる動詞の階層構造の語彙資源としては十分ではないことが分かっている。

そこで我々は動詞の階層構造のデータセットを作成するため、語彙資源として FrameNet [6] を用いた。本研究の貢献は次の 3 つである。

- FrameNet のフレーム間関係の 1 つである Inheritance に着目し、フレーム同士の階層構造を可視化し大域的に観察した。
- フレーム間関係をもとに、英語/日本語フレームネットから動詞の階層構造のデータセットを作成した。
- 作成したデータセットを用いて、動詞ペア間の階層構造の有無を判定する、動詞階層構造の埋め込みに関する新しいベンチマークタスクを考案した。

## 2　FrameNet

### 2.1　フレーム意味論

FrameNet は、言語表現が言語使用者の意識上に喚起する背景知識（フレーム）を用いて言語表現の意味を記述するフレーム意味論 [11] という言語学的枠組みに基づき、言語表現の意味をフレームを用いて記述した語彙資源である。FrameNet は様々な言語で作成されており、英語の他にも日本語 [12] やドイツ語 [13, 14]、ブラジルポルトガル語 [15] などのフレームネットが存在する。本論文では、フレーム意味論の枠組みに基づき作成された語彙資源全体を FrameNet と定義し、ある特定の言語の語彙資源を指す場合はフレームネットと記述する。

「誰が」「何を」「どうする」のような、人が経験するさまざまな状況や事象、物体に関する背景知識をフレーム (Frame) と呼び、特定のフレームを喚起する特定の語義を語彙単位 (Lexical Unit: LU) と定義している。フレームは、それぞれのフレームに即

した具体的な意味役割であるフレーム要素（Frame Element: FE）を持つ。また、それぞれのフレーム同士の関係はフレーム間関係 (Frame to Frame Relation) として定義されている。フレーム間関係は 9 種類存在し、あるフレームの要素を全て引き継ぐ、つまりフレーム同士がある状況とより具体化した状況の関係である Inheritance や、ある状況を別の視点から見た (例えば、売買を売り手と買い手の側から見た)2 つのフレーム間の関係である Perspective on などがある。

### 2.2 フレーム間関係 Inheritance

ここでは動詞の階層構造を抽出する際の手がかりとなるフレーム間関係 Inheritance(継承) について説明する。あるフレームが別のフレームのより具体的な状況を表している場合、2 つのフレームの間にはフレーム間関係 Inheritance が存在する。その際、親フレームについての定義は全て子フレームについても当てはまる。

FrameNet の構造と Inheritance 関係を持つフレームの例を図 1 に示す。ここでは `Getting` フレームと `Commerce_buy` フレームに着目している。商取引の中の一場面を買い手の立場から表現した `Commerce_buy` フレームは Buyer, Money などのフレーム要素を持つ。`Commerce_buy` フレームは `Getting` フレームを継承 (Inherits from) している。つまり、`Getting` フレームは `Commerce_buy` フレームの親フレームである。`Commerce_buy` フレームを喚起する語彙単位は、動詞の *buy* や *purchase* などである。これらの動詞が主動詞として文中に出現するとき、その文は `Commerce_buy` フレームに関する意味を持つ。つまり、商取引を買い手の立場から表現した意味を持つ。`Getting` フレームを喚起する語彙単位として *acquire* や *get* などが挙げられる。

本研究では、ある 2 つのフレームが Inheritance 関係 (継承、上位下位関係) を持つとき、それらのフレームを喚起する語彙単位同士も上位下位関係を持つという仮定を置いた。その仮定をもとに英語/日本語フレームネットから動詞階層構造を抽出した。

## 3 FrameNet からの階層構造抽出

本章では FrameNet のフレーム間の Inheritance 関係を用いて動詞の上位下位関係を抽出する方法について説明する。ここでは対象として英語/日本語フレームネットを用いているが、現状では他言語のフレームネットの多くが英語フレームネットと同じフレームとフレーム間関係を採用しているため、本論文の手法は多言語にも適用可能である。

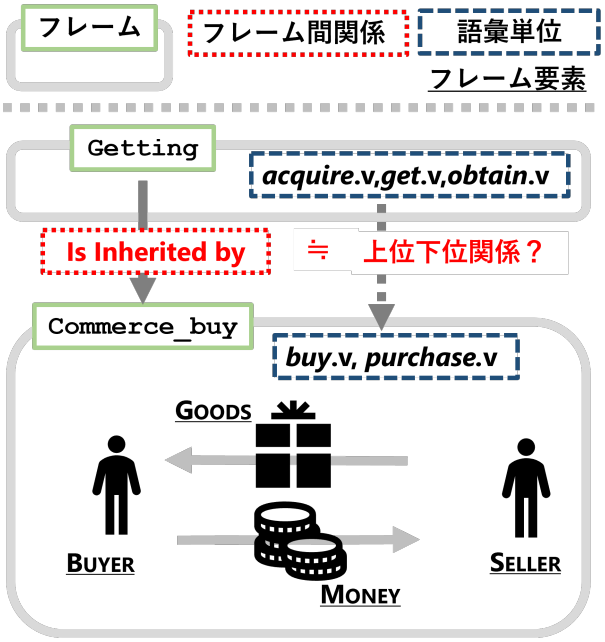


**図 1** FrameNet の構造と Inheritance 関係を持つフレームの例。`Getting` フレームと `Commerce_buy` フレームが Inheritance 関係を持つとき、それぞれの語彙単位 (LU) 同士が上位下位関係を持つと仮定する。

### 3.1 継承関係を持つフレームの抽出

英語フレームネットから Inheritance 関係でつながっているフレームを抽出した様子を図 2 に示す。英語フレームネットのバージョン 1.7 を使用し、グラフの可視化ツールとして Cytoscape [16] を用いた。ノードが英語フレームネットの個々のフレームを表し、エッジが Inheritance 関係を表す。

図 2 を見ると、上部に一番大きくかつ連結な部分グラフが 1 つあり、下部に 2 つまたは 3 つのフレームが繋がった小さな部分グラフが多数確認できる。大きな連結グラフの存在から、FrameNet 内ではフレーム同士は個々のフレームペアの継承関係のみを持つのではなく、あるフレームペアのさらに上位のフレームにわたって多段の階層構造を形成していることが分かる。これは FrameNet が多数の動詞間の上位下位関係を定義していること、つまり動詞階層構造の語彙資源として活用できそうだということを示す。

また、FrameNet 内のフレーム間関係の 1 つに着目して全体を可視化、大域的に観察した研究は著者の

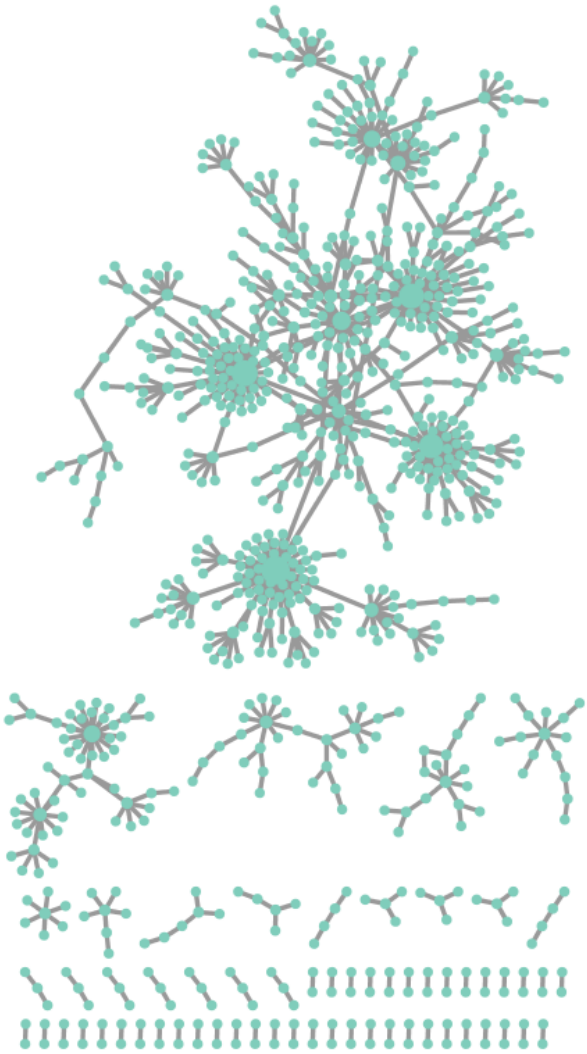

**図 2** Inheritance 関係でつながっているフレーム。ノードがフレーム、エッジが Inheritance 関係を表す。FrameNet 内にはフレームの多段の階層構造が存在することが分かる

知る限り存在しない。継承関係でつながっているフレームを可視化したことによって、部分グラフごとにアノテータに似たフレームのアノテーションを依頼するなど、既存のフレームのアノテーションの確認にも活用できると考えている。

図 2 の上部、大きな連結グラフの中心部分を、Inheritance 関係でつながったフレームの例として図 3 に示す。ここでは `Event` フレームが `Getting` フレームや `Coming_to_be` など多数のフレームの親フレームになっていることが分かる。また、`Event` フレームから `Getting` フレーム、`Commerce_buy` フレームと多段の継承関係があることが分かる。グラフ化することによって抽象的なフレームをより見つけやすくなっている。

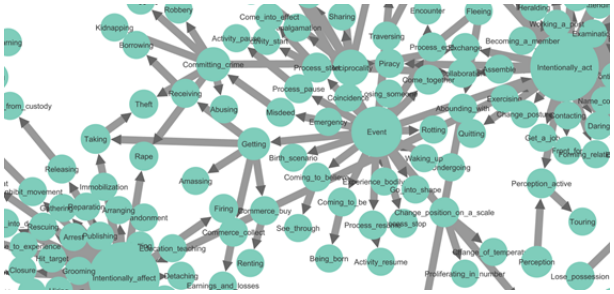


**図 3** Inheritance 関係でつながったフレームの例。`Event` フレームは多数のフレームの親フレームである。

### 3.2 動詞階層構造抽出

節 3.1 で作成した Inheritance 関係を持つフレームのネットワークから、一番大きな連結部分グラフを取り出した。そして、それぞれのフレームの語彙単位 (ここでは動詞) を抽出して動詞階層構造のデータセットを作成した。例えば、Inheritance 関係を持つ `Getting` フレームと `Commerce_buy` フレームの語彙単位 *get* と *buy* をノードとし、エッジでつなげることによりデータをグラフ形式に変換した。

ここで、フレーム同士の上位下位関係が存在するとフレーム内の語彙単位同士も上位下位関係を持つという仮定を定性的に検証した。Inheritance 関係を持つフレームのペアごとに語彙単位の例を挙げたものを表 1 に示す。*yell* と *say* や、“教える” と “言う” など、継承関係を持つフレームの多くが、語彙単位同士も上位下位関係を持つことが分かる。

抽出したフレーム数と語彙単位 (動詞) 数、上位語下位語のペア数を表 2 に示す。抽出した動詞を用いて実験を行う。

## 4 動詞階層構造埋め込み

本章では、作成した動詞の階層構造のデータセットを Poincaré embedding [7] に埋め込み、学習した埋め込みを用いて英語/日本語フレームネットの語彙単位間の上位下位関係の有無を判定する link prediction タスクを解く。

実験の目的は次の 2 つである。

- フレームネットから作成した動詞の階層構造データセットが機械学習可能な構造を持つデータであることを確認する。
- 動詞の階層構造埋め込みのベンチマークタスクを提案する。

表 1　英語/日本語フレームネットから抽出した、継承関係を持つフレームとその語彙単位 (動詞) の例

| 言語 | 子フレームの語彙単位 | 親フレームの語彙単位 | 子フレーム | 親フレーム |
|---|---|---|---|---|
| 英語 | *walk* | *go* | Self_motion | Motion |
| 英語 | *yell* | *say* | Communication_noise | Communication |
| 英語 | *wash* | *do* | Grooming | Intentionally_affect |
| 英語 | *jump* | *do* | Attack | Intentionally_affect |
| 英語 | *eat* | *take* | Ingestion | Ingest_substance |
| 日本語 | 売る | 与える | Commerce_sell | Giving |
| 日本語 | 教える | 言う | Telling | Statement |
| 日本語 | 飛ぶ | 行う | Self_motion | Motion |
| 日本語 | 作り上げる | 作る | Intentionally_create | Creating |
| 日本語 | 走り去る | 動く | Self_motion | Motion |

表 2　英語/日本語フレームネットから抽出したデータ数

| | フレーム | 動詞 | 上位語下位語ペア |
|---|---|---|---|
| 英語 | 307 | 1996 | 28857 |
| 日本語 | 117 | 437 | 4774 |

表 3　英語/日本語フレームネット単語ペアの上位下位関係を判定する二値分類タスクの実験結果

| | 精度 | 適合率 | 再現率 | F 値 |
|---|---|---|---|---|
| 英語 | 0.741 | 0.782 | 0.666 | 0.719 |
| 日本語 | 0.826 | 0.895 | 0.741 | 0.811 |

### 4.1　モデル/評価

動詞階層構造を埋め込むために、双曲空間を利用し低次元の分散表現を作成する手法である Poincaré embedding を使用した。WordNet を始めとした教師あり上位語下位語判定タスクでよく用いられる。英語/日本語フレームネットそれぞれから抽出した動詞の上位下位語ペアで 2 次元の Poincaré embedding を学習した。学習ライブラリとして gensim [17] を使用、epoch 数は 500–50000 の間でチューニングした。

評価タスクとして、単語をノードとするグラフから単語を 2 つ取り出し、単語ペアがエッジで結ばれているか (上位下位関係を持つか) を予測する link prediction タスクを用いた。本研究では英語/日本語フレームネットで link prediction を行う。ある単語ペアが分散表現内でノルムの差が一定以上かつ類似度が一定以上の時、上位下位関係を持つとした。検証/テストデータとして英語フレームネットは正例負例 10000 ペアずつ、日本語は 3000 ペアずつを無作為に抽出し、評価尺度として F1 スコアを用いた。

### 4.2　実験結果

実験結果を表 3 に示す。英語/日本語フレームネット共に、単語ペア間の上位下位関係の有無を 7 割以上の確率で予測できたことから、作成した動詞の階層構造のデータセットは機械学習可能な構造になっていることが確認できた。

## 5　結論

本研究では FrameNet のフレーム間関係の一つである Inheritance に着目し、フレーム同士の階層構造を可視化した。また、それらのフレームを喚起する動詞を用いて英語/日本語の動詞階層構造のデータセットを作成した。作成したデータセットで Poincaré embedding を学習し、ある動詞ペアが上位下位関係を持つかを判定する 2 値分類タスクで両言語共に 7 割以上の精度を達成した。このことから、フレームネットから作成したデータセットは機械学習が可能な構造のデータであること、また動詞の上位下位関係を分散表現に埋め込めていることが確認できた。フレーム意味論では、特に日常語に関しては言語や文化に依存しないフレームが多く存在するとしているため、本手法は英語、日本語以外の多言語フレームネットにも適用可能である。

今後の展望として、動詞の階層構造の知識の応用タスクへの適用が挙げられる。名詞の階層構造の情報が応用タスクにとって有用であるという研究 [10] は存在するが、動詞に関しては階層構造データの整備が不十分なため知識を応用タスクに埋め込む段階まで至っていない。本研究で作成した動詞の階層構造のデータセットと、埋め込みの精度を測るためのベンチマークを手掛かりに、動詞知識活用の研究を加速させていきたい。

# 参考文献


[1] Paola Velardi, Stefano Faralli, and Roberto Navigli. OntoLearn Reloaded: A Graph-Based Algorithm for Taxonomy Induction. **Computational Linguistics**, Vol. 39, No. 3, pp. 665–707, 09 2013.

[2] Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, and Gerhard Weikum. Robust question answering over the web of linked data. In **Proceedings of the 22nd ACM International Conference on Information amp; Knowledge Management**, CIKM '13, p. 1107–1116, New York, NY, USA, 2013. Association for Computing Machinery.

[3] Johannes Hoffart, Dragan Milchevski, and Gerhard Weikum. Stics: Searching with strings, things, and cats. In **Proceedings of the 37th International ACM SIGIR Conference on Research amp; Development in Information Retrieval**, SIGIR '14, p. 1247–1248, New York, NY, USA, 2014. Association for Computing Machinery.

[4] Christiane Fellbaum. A Semantic Network of English: The Mother of All WordNets. **Comput. Humanit.**, Vol. 32, No. 2-3, pp. 209–220, 1998.

[5] George A. Miller. Wordnet: A lexical database for english. **Commun. ACM**, Vol. 38, No. 11, p. 39–41, nov 1995.

[6] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In **Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1**, ACL '98/COLING '98, p. 86–90, USA, 1998. Association for Computational Linguistics.

[7] Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.

[8] 岩本蘭, 小比田涼介, 和地瞭良. 極座標を用いた階層構造埋め込み. 言語処理学会第 27 回年次大会予稿集, 2021.

[9] Tom Richens. Anomalies in the WordNet verb hierarchy. In **Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)**, pp. 729–736, Manchester, UK, August 2008. Coling 2008 Organizing Committee.

[10] Satanjeev Banerjee and Ted Pedersen. Extended gloss overlaps as a measure of semantic relatedness. In **Proceedings of the 18th International Joint Conference on Artificial Intelligence**, IJCAI'03, p. 805–810, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.

[11] Charles J. Fillmore. Frame semantics and the nature of language*. **Annals of the New York Academy of Sciences**, Vol. 280, No. 1, pp. 20–32, 1976.

[12] 小原京子, 河原大輔, 笹野遼平, 関根聡. 集合知を用いた大規模意味的フレーム知識の構築. 言語処理学会第 27 回年次大会予稿集, 2021.

[13] Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. The SALSA corpus: a German corpus resource for lexical semantics. In **Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)**, Genoa, Italy, May 2006. European Language Resources Association (ELRA).

[14] Alexander Ziem and Hans C Boas. Towards a constructicon for german. In **2017 AAAI Spring Symposium Series on Computational Construction Grammar and Natural Language Understanding**, Vol. SS–17–02, pp. 274–277, 2017.

[15] Tiago T. Torrent, Maria Margarida M. Salomão, Fernanda C. A. Campos, Regina M. M. Braga, Ely E. S. Matos, Maucha A. Gamonal, Julia A. Gonçalves, Bruno C. P. Souza, Daniela S. Gomes, and Simone R. Peron. Copa 2014 FrameNet brasil: a frame-based trilingual electronic dictionary for the football world cup. In **Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations**, pp. 10–14, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.

[16] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome research**, Vol. 13, No. 11, pp. 2498–2504, 2003.

[17] Radim Rehurek and Petr Sojka. Gensim–python framework for vector space modelling. **NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic**, Vol. 3, No. 2, 2011.