

動詞における連想と共起の差異性

寺岡 丈博†

†東京工科大学

石崎 俊‡

‡慶應義塾大学

1 はじめに

従来の自然言語処理研究では、コンピュータが使用する知識としてコーパスやシソーラスなどが統計的手法や機械学習を用いた手法とともに利用されることで、様々なタスクに関して一定の精度向上が果たされてきた。コーパスの多くは新聞や Web 文書などから構成されており、抽出できる単語間の関係性は共起に基づいている。そのため、これまでに提案されてきた有用な手法でコーパスを用いている場合は、言葉の共起情報を知識として利用したといえる。一方で、言葉の連想に基づいた知識が一部の自然言語処理研究 [1, 2] で利用されている。さらに先行研究では、このような連想情報を換喩解析に応用して手法の有用性を示すと同時に、ベースラインとして使用した共起情報の特長を確認している [3, 4]。

今日、深層学習を利用したシステムの高精度化が期待されているが、学習として使用されるデータの違いが問われている。このことを踏まえ、本研究では動詞における連想情報と共起情報で共通する語について分析し、差異性を明らかにすることを目的としている。

2 動詞連想概念辞書

本研究では、先行研究 [4] から拡張した動詞連想概念辞書の連想情報を用いるため、以下で動詞連想概念辞書の概要を述べる。

動詞連想概念辞書は、人間を対象とした大規模な連想実験のデータから構築され、2017 年現在は刺激語の動詞 773 語に対して連想語数が約 295,000 語、異なり語数が約 57,000 語となっている。刺激語の動詞は、日本語基礎辞典 [5] と日本語動詞用法辞典 [6] の見出し語の動詞を全て含んでいるため、基本的な動詞は網羅されているといえる。連想実験で刺激語と一緒に呈示される連想の課題は 10 種類の主な深層格であり、「動作主」、「対象」、「始点」、「終点」、「時点」、「場所」、「手段」、「様相」、「理由」、そして「目的」である。なお、移動や感情を表す動詞や自動詞と他動詞などで各刺激語と課題の組み合わせは異なっている。実験参加者は

表 1: 動詞連想概念辞書のデータ例

連想の課題	「運ぶ」の連想語 (連想距離)
動作主	私 (3.60), 引越し業者 (4.21)
対象	荷物 (1.36), 人 (4.45), 家具 (7.78)
始点	家 (1.45), 学校 (3.81), 部屋 (5.80)
終点	家 (1.92), 駅 (3.73), 目的地 (4.02)
場所	倉庫 (3.73), 工場 (4.50), 会社 (6.22)
手段	車 (1.62), トラック (2.34), 手 (3.47)

刺激語動詞 1 語につき 40 人であり、大学生と大学院生を対象に延べ約 2,900 人となっている。また、刺激語動詞と連想語の連想距離 [1] が、連想の頻度と順位の間数として表されており、表 1 のように連想距離の値が小さいほど動詞「運ぶ」に対して連想しやすいという関係を示している。

3 連想情報と共起情報の差異性の分析

3.1 共起情報

動詞連想概念辞書の連想情報と比較する対象として、5 年分の新聞コーパス約 6,170,000 文から共起情報を次の手順で抽出した。まず、文中で動詞連想概念辞書の刺激語 (動詞) が含まれている文を全て抽出した上で、それらの動詞と係り受けの関係がある文節から格情報を取得する*。この格情報は、格助詞と格要素のペアを指している。取得する格助詞は「が」、「を」、「から」、「まで」、「で」の 5 種類とし、文節内でともに使用されている名詞を格要素として取得して動詞に対する共起頻度を求める。このようにして、動詞連想概念辞書の刺激語 (動詞) に関する格情報をコーパスから共起情報として抽出した。

表 2: 連想情報と共起情報の対応関係

連想情報	共起情報
連想の課題	格助詞
「動作主」	「が」
「対象」	「を」
「始点」	「から」
「終点」	「まで」
「場所」「手段」	「で」
連想語	格要素
連想距離 (昇順)	共起頻度 (降順)

Difference between Association and Co-occurrence in Verbs

†Takehiro TERAOKA ‡Shun ISHIZAKI

†Tokyo University of Technology

‡Keio University

*形態素解析と構文解析には、それぞれ MeCab 0.996 と CaboCha 0.69 を使用

表 3: 連想語と格要素で共通する単語 ($n \geq 5$) がある動詞の語数と順位相関係数の平均

	動作主 - が	対象 - を	始点 - から	終点 - まで	場所 - で	手段 - で
動詞の語数	396	375	51	20	270	263
順位相関係数の平均	0.260	-0.018	0.298	0.363	0.082	0.306

表 4: 「捕まえる」の「対象 - を」に対する連想語と格要素の順位関係

	連想語	格要素
上位	犯人, 泥棒, 動物, 虫	カエル, カブトムシ, トンボ, ネズミ, 蚊
下位	ネズミ, カブトムシ, 蚊, カエル	犯人, 動物, 泥棒, 虫

3.2 分析方法

表 2 は, 動詞連想概念辞書の連想情報と抽出した共起情報において対応する項目をまとめたものである. 連想の課題は深層格ではあるが, コーパスで使用される格助詞がそれぞれ表 2 に示される. また, 連想語と連想距離 (昇順) は, 動詞と連想の課題に対して連想された単語とその連想されやすさを指しているため, 共起情報では動詞と格助詞に対して格要素になる単語とその共起しやすさを示す共起頻度 (降順) がそれぞれ対応している. これらの関係を用いて, 連想語と格要素の間で共通する単語が 5 語以上ある動詞において, 共通する単語の各順位に対して Spearman の順位相関係数を求めることで, 連想しやすい語と共起しやすい語がどのように異なるのかを分析する.

3.3 結果と考察

表 3 は, 連想の課題と格助詞の組み合わせごとに Spearman の順位相関係数の平均をまとめたものである. この結果から, 連想語と格要素で共通する単語に関して全体的には順位の相関がほとんどないことから, 動詞に対して連想しやすいことと共起しやすいことが必ずしも一致していないことがいえるが, 一部の動詞では正の相関や負の相関が確認できた.

そこで, 連想の課題が「対象」と格助詞「を」の組み合わせを事例として挙げると, 「悼む」($n = 6$)や「勤める」($n = 5$)は連想語と格要素で共通する単語の順位がほとんど同じだった (ただし, n は共通する単語の数を指す). これは, どちらの動作もある程度限定された状況で行われるため, 連想しやすい内容と共起しやすい内容が似通っていると考えられる.

その反対に, 共通する語の順位が逆になっている動詞としては「捕まえる」($n = 19$)が挙げられる. 表 4 は, 「捕まえる」に対する連想語と格要素で順位が上位と下位にそれぞれ該当する語を記したものであり, 順位が上位と下位で入れ代わっている. 連想語の順位の上位が「動物」や「虫」で下位が「ネズミ」や「カブトムシ」であることから, 上位語にあたる単語がより連

想しやすいと考えられるのに対し, 格要素は逆に下位語が共起しやすい. これは, 共起情報の元が新聞コーパスのため, 「動物」や「虫」よりも「ネズミ」や「カブトムシ」のような具体的な情報が記されなければ記事の情報量としては一般的に不十分だからである. この共通する語の順位が入れ代わる傾向は, 他の動詞や他の組み合わせにおいても該当するため, これが一部の動詞にとっては連想と共起の差異性を表していると考えられる.

4 おわりに

共起情報は動詞と格要素の組み合わせで換喩表現や慣用表現を多く含む内容がある一方で, 動詞連想概念辞書の連想情報からは得られ難い「押し入れの“中”」や「テーブルの“上”」のような方向の情報を抽出できる [4]. 本研究で示した一部の動詞が該当する差異性を考慮し, 目的に応じて連想情報と共起情報を使い分けることで, より高度な応用が期待できる.

謝辞

本研究の一部は JSPS 科研費 15K16787 の助成を受けたものである.

参考文献

- [1] 岡本潤, 石崎俊. 概念間距離の定式化と既存電子化辞書との比較. 自然言語処理, Vol. 8, No. 4, pp. 37-54, 2001.
- [2] 寺岡丈博, 東中竜一郎, 岡本潤, 石崎俊. 単語間連想関係を用いた換喩表現の自動検出. 人工知能学会論文誌, Vol. 28, No. 3, pp. 335-346, 2013.
- [3] 寺岡丈博, 石崎俊. 単語間連想関係を考慮した換喩表現の解析. 第 78 回全国大会講演論文集, 第 2016 巻, pp. 1-2, 2016.
- [4] Takehiro Teraoka. Metonymy Analysis Using Associative Relations between Words. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pp. 4614-4620, 2016.
- [5] 森田良行. 基礎日本語辞典. 角川書店, 1989.
- [6] 小泉保, 船越道雄, 本田晶治, 仁田義雄, 塚本秀樹. 日本語動詞用法辞典. 大修館書店, 1989.