

和語動詞を基本単位とした複合動詞と事態性名詞の予備調査

田村 花綾^{†1} 渋谷 英潔^{†2} 永山 翔滋^{†1} 阪本 浩太郎^{†1} 森 辰則^{†1} 神門 典子^{†2}^{†1}横浜国立大学 ^{†2}国立情報学研究所

E-mail: {tamura-k,nagayama,sakamoto,mori}@forest.eis.ynu.ac.jp {shib,kando}@nii.ac.jp

1 はじめに

近年のフェイクニュースなどの問題から虚偽情報に対する信憑性判断支援が世界的に注目されている [1]。我々は、NTCIR QALab-PoliInfo[2] として政治情報に関する shared task を開催する一方、虚偽情報の処理という観点から人狼知能プロジェクト [3] に参加し、人狼ゲームプラットフォーム LiCOS[4] の開発に取り組んでいる。人狼ゲームとは、対話を通して「村人」の中に潜伏した「人狼」を見つけ出す対戦型の多人数ゲームであり、「人狼」となったプレイヤーは、他のプレイヤーに正体を悟られないよう、普通の「村人」のふりをして「偽の推理」を披露したり、「古い師（人狼の正体を知ることができる）」を騙って他のプレイヤーを扇動したりすることで勝利を目指す。そのため、プレイヤーの言動間にある矛盾や齟齬などを手がかりに真実を導かなくてはならない。

こういった文（イベント）間の真偽関係を理解するためには、「狼が村人を殺した」が「村人が死んだ」を包含するかといった含意関係認識 [5] を行う必要がある。一般に、含意関係認識のためには「殺す」が「死ぬ」を包含するといった語彙的な知識が必要である。しかしながら、日本語においてイベントの核となる述語には単独動詞以外にも「食い殺す」といった複合動詞や「襲来（する）」といった事態性名詞など多くの語彙が存在する。これらの語彙間の含意関係を辞書に全て登録するには多大な労力がかかる。一方で、人間にとって「食い殺す」と「殺す」の間に含意関係が成り立つことは容易に推測できる。また、事態性名詞を構成する漢字には「襲う」や「来る」といった和語の動詞となる漢字が含まれていることが多く、「襲来」と「襲う」の間の含意関係も容易に推測できる。したがって、和語動詞を述語の基本単位として複合動詞や事態性名詞を定義できれば、比較的少数の語彙で述語を表現できると考えられる。これらは、表意文字である漢字を媒介に推測していると考えられるが、「狂い咲く」と「狂う」、「履行」と「履く」のように共通の漢字があることが必ずしも含意関係を意味しているわけでは

ない。したがって、和語動詞およびその語幹となる漢字により含意関係を推測するという基本思想の下、本稿では複合動詞と事態性名詞を構成する動詞（漢字）間の関係を分析する。なお、本稿で分析対象とする複合動詞は動詞連用形+動詞型の語彙的複合動詞であり、事態性名詞はサ変動詞接続可能な二字熟語である。

2 関連研究

複合動詞に関しては、複合動詞レキシコン¹が参考となる。複合動詞レキシコンには、動詞連用形+動詞型の語彙的複合動詞 2,759 語が収録されており、構成要素となる 2 つの動詞の間に以下の 4 つの関係を設定している。

- C1 VV（動詞+動詞）
- C2 Vs（動詞+補助的な動詞）
- C3 pV（接頭辞化した動詞+動詞）
- C4 V（一語化）

この関係は、複合動詞を構成する動詞の単独での意味がどの程度残っているかによるものである。例えば、C1 の複合動詞「(人狼が) 食い殺す」は前項動詞「食う」と後項動詞「殺す」で構成されており、動詞テ形+動詞形にした「食って殺す」とほぼ同義であるのに対し、C2 の複合動詞「(子供が) 走り出す」は「走って出す」とは異なる関係である。本稿でも同様の考えに基づいて分析を行う。ただし、複合動詞レキシコンは言語学の観点から分析されており、我々が目的とする工学的応用については考慮されていない。

我々は、事態性名詞を構成する 2 つの漢字の間にも同様の関係が成り立つと仮定した。すなわち、事態性名詞「襲来」は前項漢字「襲（う）」と後項漢字「来（る）」で構成され、「襲い来る」や「襲って来る」と同義の関係にあるのに対し、「履行」は「履いて行く」と

¹<https://db4.ninjal.ac.jp/vvlexicon/>

異なる関係である。事態性名詞を扱った従来研究 [6, 7] は、こういった構成漢字の関係に焦点をあてていない。また、複合動詞の関係を事態性名詞に応用し統一的な枠組みでイベントを解釈しようとする試みも我々が初めてである。

3 対象データ

BCCWJ コーパス [8] のコアデータ²を用いて、以下の手順で分析対象とするデータを抽出した。品詞が「動詞-一般」で、最初の4文字が「漢字-ひらがな-漢字-ひらがな」になっている単語を複合動詞として抽出した。また、品詞が「名詞-一般-サ変可能」で、直後に原形が「する」となる単語が接続し、漢字2文字で構成される単語を事態性名詞として抽出した。結果として、複合動詞として抽出した異なり単語が1,228語、事態性名詞の異なり単語が2,430語となった。上記の方法で全ての複合動詞と事態性名詞が抽出されるわけではないが、動詞(漢字)間の関係にどのような類型があるかを調査するのに大きな問題はないと判断した。

4 分析方法

抽出された複合動詞と事態性名詞に対して、第一著者が以下の手順でアノテーションを行った。

4.1 複合動詞

複合動詞レキシコンの4つの関係は、構成要素となる動詞の単独での意味がどの程度残っているかによるものである。したがって、複合動詞を構成要素となるそれぞれの動詞で換言できるかにより、以下の4つの類型に分類する。

V1 前項動詞、後項動詞ともに換言できる。

V2 前項動詞のみ換言できる。

V3 後項動詞のみ換言できる。

V4 前項動詞、後項動詞ともに換言できない。

例えば、「人狼が食い殺した」という文は「人狼が食った」や「人狼が殺した」に換言できるのでV1、「村に辿り着いた」は「村に着いた」に換言できるが「村を辿った」に換言できないためV3に分類する。文脈によっては「飛び出す」が「飛んで出す」という解釈になることがあるため必ずしも一意に分類できるわけで

²UniDicの短単位での形態素解析結果を利用した。

はないが、本稿では最も一般的と思われる用法での解釈で判断した。

4.2 事態性名詞

事態性名詞として抽出された単語は二字熟語の一種である。教科書や漢字検定³などにおける二字熟語の構成の種類には以下のものがある。

1. 同じ漢字を重ねたもの(例、「日々」、「青々」)
2. 似たような意味の漢字を組み合わせたもの(例「思考」、「進行」)
3. 反対の意味や対の意味の漢字を組み合わせたもの(例「勝負」、「遠近」)
4. 前項漢字が後項漢字を修飾する言葉になっているもの(例「青空」、「再会」)
5. 前項漢字が主語、後項漢字が述語にあたるもの(例、「人造」、「国営」)
6. 前項漢字が動作、後項漢字が対象などを表しているもの(例、「消火」、「着席」)
7. 前項漢字が否定接頭辞で後項漢字の意味を打ち消すもの(例、「不足」、「無用」)
8. 後項漢字が接尾辞で前項漢字の様子を表したり、意味を添えたりするもの(例、「知的」、「美化」)
9. 長い熟語の略語(例、「国連」、「入試」)

本研究でも、最終的に上記の種類に準ずる分類にしたいと考えている。しかしながら、抽出された単語には「入金(金を入れる)」のように、和語動詞の語幹とならない漢字で構成されているものがある。そのため、4.1で述べた複合動詞の分類をそのまま適用できない。したがって、最初に、事態性名詞を構成する漢字が和語動詞の語幹となるか否かによって、以下の4つの類型に分類する。

N1 前項漢字、後項漢字ともに和語動詞の語幹となる。

N2 前項漢字のみ和語動詞の語幹となる。

N3 後項漢字のみ和語動詞の語幹となる。

N4 前項漢字、後項漢字ともに和語動詞の語幹とならない。

ここでN1と分類されたものに対しては複合動詞と同様の分類を行うことができると考えられる。

³<https://www.kanken.or.jp/>

表 1: 抽出された複合動詞 (Top20)

複合動詞	頻度	換言可能		類型
		前項動詞	後項動詞	
取り組む	176	×	×	V4
繰り返す	82	×	×	V4
受け入れる	80	×	×	V4
思い出す	79	○	×	V2
引き続く	54	○	○	V1
振り返る	48	○	○	V1
盛り込む	47	×	×	V4
取り上げる	47	○	○	V1
付き合う	40	×	×	V4
受け取る	40	×	○	V3
落ち着く	38	×	×	V4
打ち出す	35	×	○	V3
取り入れる	34	○	○	V1
持ち込む	33	○	×	V2
引っ張る	32	○	○	V1
呼び掛ける	31	○	×	V2
呼び出す	28	○	×	V2
茹で上がる	28	○	×	V2
立ち上がる	28	○	×	V2
取り出す	28	○	○	V1

5 結果と考察

執筆時点では、アノテーション作業が継続中であり、複合動詞の上位 1,000 件と事態性名詞の上位 113 件のみ付与している。表 1 と表 2 に、BCCWJ コーパスのコアデータにおける出現頻度順での複合動詞と事態性名詞の上位 20 語をそれぞれ示す。また、アノテーションを付与した範囲における種類の割合を表 3 に示す。表 3 から、現時点では、事態性名詞の中に和語動詞の語幹となる漢字が必ず含まれており、和語動詞を基本として含意関係の推測を行うための前提は満たされていることが分かる。

6 まとめ

本稿では、和語動詞およびその語幹となる漢字を手がかりとした含意関係認識を目的として、BCCWJ コーパスを用いた複合動詞と事態性名詞を構成する動詞（漢字）間の関係を分析した。今後、アノテーション対象を拡大して分析を継続するとともに、人狼 BBS⁴のゲー

⁴<http://www.wolfg.x0.com/>

ムログなどを用いて同様の分析を行う予定である。

謝辞

本研究の一部は、JSPS 科研費 JP16H01756 の助成を受けたものである。

参考文献

- [1] P. Nakov, A. Barrón-Cedeño, T. Elsayed, R. Suwaileh, L. Márquez, W. Zaghouni, P. Atanasova, S. Kyuchukov, and G. Da San Martino. Overview of the CLEF-2018 CheckThat! Lab on Automatic Identification and Verification of Political Claims. Proceedings of the Ninth International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Lecture Notes in Computer Science, 2018.
- [2] Y. Kimura, H. Shibuki, H. Ototake, Y. Uchida, K. Takamaru, K. Sakamoto, M. Ishioroshi, T. Mitamura, N. Kando, T. Mori, H. Yuasa, S. Sekine and K. Inui, “Overview of the NTCIR-14 QA Lab-PoliInfo Task,” Proceedings of the 14th NTCIR Conference, 2018.

表 2: 抽出された事態性名詞 (Top20)

事態性名詞	頻度	対応和語動詞		N 類型	V 類型
		前項漢字	後項漢字		
実施	359	—	施す	N3	—
推進	257	推す	進める	N1	V1
利用	195	—	用いる	N3	—
発表	147	発つ	表す	N1	V1
対応	132	—	応える	N3	—
説明	124	説く	明らむ	N1	V1
使用	122	使う	用いる	N1	V1
開催	120	開く	催す	N1	V1
活用	119	活かす	用いる	N1	V1
提供	118	提げる	供える	N1	V1
紹介	114	紹ぐ	介ける	N1	V1
確認	114	確かめる	認める	N1	V1
参加	112	参る	加わる	N1	V1
設置	111	設ける	置く	N1	V1
発生	108	発つ	生む	N1	V1
結婚	107	結ぶ	—	N2	—
期待	107	—	待つ	N3	—
増加	105	増す	加わる	N1	V1
支援	103	支える	援ける	N1	V1
検討	98	検める	討ねる	N1	V1

表 3: 類型の件数と割合

複合動詞			事態性名詞	
V1	626	(0.626)	N1	87 (0.770)
V2	23	(0.023)	N2	10 (0.088)
V3	162	(0.162)	N3	16 (0.142)
V4	189	(0.189)	N4	0 (0.000)
合計	1,000	(1.000)	合計	113 (1.000)

- [3] Y. Kano, C. Aranha, M. Inaba, H. Osawa, D. Katagami, T. Otsuki, and F. Toriumi, “Overview of the AIWolf-Dial 2019 Shared Task: Competition to Automatically Play the Conversation Game “Mafia”,” Proceedings of the 1st International Workshop of AIWerewolf and Dialog System (AI-WolfDial 2019), the 12th International Conference on Natural Language Generation (INLG 2019), 2019.
- [4] 阪本浩太郎, 永山翔滋, 石下円香, 渋谷英潔, 森辰則, 神門典子. 人狼ゲームプラットフォーム LiCOS を用いた欺瞞対話コーパスのためのゲームログの収集. 言語処理学会第 25 回年次大会発表論文集, pp. 1411–1414, 2019.
- [5] S. Matsuyoshi, Y. Miyao, T. Shibata, C.-J. Lin, C.-

W. Shih, Y. Watanabe and T. Mitamura, “of the NTCIR-11 Recognizing Inference in TExt and Validation (RITE-VAL) Task,” Proceedings of the 11th NTCIR Conference, 2014.

- [6] 小町守, 飯田龍, 乾健太郎, 松本裕治. 名詞句の語彙統語パターンを用いた事態性名詞の項構造解析. 自然言語処理, vol. 17, no. 1v pp. 141-159, 2010.
- [7] 黒田航. 事態性名詞の項構造と動詞の項構造の統合 PMA を使った日本語の支援動詞構文の分析とその含意. 言語処理学会第 14 回年次大会発表論文集, pp. 1128–1131, 2008.
- [8] 国立国語研究所コーパス開発センター『現代日本語書き言葉均衡コーパス』DVD 版 (Version 1.1), 国立国語研究所, 2015.