

# 依存型意味論による日本語複合動詞の意味合成と推論

佐久間香那 富田朝 戸次大介  
お茶の水女子大学

{sakuma.kana,tomita.asa,bekki}@is.ocha.ac.jp

## 概要

日本語複合動詞は、これらを構成する動詞の関係に応じて全体の項構造が変化する。Isono らは、日本語複合動詞に対して統語的・意味的に適切な分析を与えるために、組合せ範疇文法 (Combinatory Categorical Grammar, CCG) に空範疇 glue element を導入する手法を提案した。本研究では、CCG 統語解析器 lightblue 上で glue element を実装し、複合動詞を含む文の統語構造・意味表示を自動解析できるように拡張するとともに、独自に作成した自然言語推論データセットを用いて、推論による妥当性の評価を行った。

## 1 はじめに

日本語複合動詞は、二つの動詞が結合することで、単純な動詞句では表現できない意味関係や項構造を形成する。このような複合動詞の統語的・意味的分析については、影山 [1] によって、並列関係、右側主要部の関係、補文関係からなる記述の分類が与えられているが、厳密な統語構造、意味合成の分析は長らく与えられていなかった。

これに対し Isono et al. [2] は、組合せ範疇文法 (Combinatory Categorical Grammar, CCG) [3] [4] と分散形態論 (Distributed Morphology, DM) [5] を組み合わせた手法により、複合動詞の統語構造と意味合成の形式的に厳密な分析を提案した。特に、glue element と呼ばれる空範疇の導入によって、日本語複合動詞における動詞間の意味的關係や項構造の共有・分担を表現することに成功している。

そこで本研究では、日本語 CCG 統語解析器 lightblue [6]<sup>1)</sup> 上で glue element を実装し、日本語複合動詞に対する統語構造・意味合成の分析を、計算的な手法で検証することのできる推論システムを構築する。このシステムは Isono et al. [2] の分析の単なる実装に留まらず、分析自体の実装面からの再検討

を含むものである。

さらに、複合動詞にまつわる経験的事実を自然言語推論 (Natural Language Inference, NLI) のデータセットとして構築し、glue element 導入前後での lightblue の精度をデータセット上で比較する。それによって、glue element による日本語複合動詞の分析の有効性を実証的に評価する。

## 2 先行研究

### 2.1 影山 [1] による日本語複合動詞の分類

影山 [1] は日本語複合動詞における 2 つの動詞の関係を、以下の 5 通りに分類している。なお、日本語複合動詞において、V1 は複合動詞の前半部分を、V2 は後半部分を構成する動詞を指す。

#### 1. 並列関係

例: 太郎が泣き叫ぶ

#### 2. 右側主要部の関係

(a) 複合動詞の項構造は V2 から引き継がれる

例: 太郎がドアを押し開ける

(b) V1 と V2 の内部引数は所有関係を持つ

例: 太郎が服の汚れを洗い落とす

(c) 複合動詞の項構造は V1 と V2 の両方から成る

例: 太郎が夏の一夜を思い出を語り明かす

#### 3. 補文関係

例: 太郎が歌を歌い上げる

並列関係では、V1, V2 が同一の項構造を持つ。右側主要部の関係では、V1 が V2 の修飾部となる。Theme は意味役割の一つであり、動詞によって影響を受ける対象を指すが、2a では、V1 と V2 が同じ Theme を共有する。「押し開ける」では、「押す」と「開ける」の Theme はどちらも「ドア」である。2b では、V1 の対象である Theme1 と V2 の対象である Theme2 が所有格の関係を持つ。「洗う」の Theme

1) <https://github.com/DaisukeBekki/Lightblue>

は「服」であり、「落とす」の Theme は「汚れ」である。このとき、「服」と「汚れ」には所有関係が成立している。複合語全体の Theme としては、「服の汚れ」を一つの項として取る。2c では、V1 と V2 が別々の Theme を持つ。「語り」の Theme は「思い出」、「明かす」の Theme は「一夜」であり、複合語全体の Theme はこの両方を取る。補文関係では、複合動詞の項構造は V2 のものをそのまま継承し、足りない Theme を V1 から補う。

## 2.2 Isono et al. [2] の分析

Isono et al. [2] は、日本語複合動詞の項構造を分析するために、DM の枠組みに基づいた分析を CCG に基づいて形式化した。この手法では、CCG に次の glue element を追加することで、複合動詞の構造的・意味的關係の合成を可能にした。glue element は、空範疇の一種であり、V1 と V2 の統語的・意味的關係を明示する役割を果たす。

$$\emptyset \vdash V_{[1]}^{stem} \$ / (V_{[1]}^{stem} \$) \backslash (V_{cont} \$) :$$

$$\lambda P.\lambda Q.\lambda \bar{x}.\lambda e.\exists e_1.\exists e_2.P(\bar{x})(e_1) \wedge Q(\bar{x})(e_2) \wedge parallel(e_1)(e_2)(e)$$

$$\emptyset \vdash V_{[1]}^{stem} \$ / (V_{[1]}^{stem} \$) \backslash (V_{cont} \$) :$$

$$\lambda P.\lambda Q.\lambda \bar{x}.\lambda e.\exists e_1.\exists e_2.P(\bar{x})(e_1) \wedge Q(\bar{x})(e_2) \wedge by(e_1)(e_2)(e)$$

意味表示内の *parallel* は、V1 と V2 のイベントが並行して発生することを記述している。これを用いることで、並列関係に分類される複合動詞において、V1 と V2 のイベントが同時に発生することを表現できる。*by* は、V1 が V2 のイベントの原因であることを記述している。これを用いることで、右側主要部の関係に分類される複合動詞において、V1 が V2 の直接の原因となることを表現できる。補文関係については、今回導入した glue element は使用せず、V2 が V1 を補文として処理することで合成される。

## 2.3 lightblue

lightblue は、CCG と依存型意味論 (Dependent Type Semantics, DTS) [7] に基づいて設計された、統語解析・意味解析・自動推論を統合的に行う自然言語推論システムである。lightblue は、文の統語構造を CCG により解析し、得られた構造に基づいて DTS の意味表示を構築する。さらに、構築された意味表示に対して定理証明を行うことで、自然言語推論までを一貫して処理するパイプラインを備えている。現在の lightblue では、複合動詞の分析において、図 1 のように *cont-mod* (連用節導入空範疇) を用いて

結合している。*cont-mod* は、V1 を V2 の連用修飾句として結合するための空範疇であり、V1 は V2 の動作様態や付随状況を表す副詞的要素として扱われる。これにより、V1 は V2 を副詞的に修飾する形となり、因果関係や結果関係といった動詞間の意味的依存関係は明確に表現されない。

$$\frac{\frac{\text{押し}}{S \backslash NP_{ga}} \quad \frac{\text{cont-mod}}{S/S \backslash S}}{S/S \backslash NP_{ga}} < B \quad \frac{\text{開ける}}{S \backslash NP_{ga}}}{S \backslash NP_{ga}} > S_x$$

$$\lambda x_0.\lambda x_1. \left[ \begin{array}{l} u_0 : \left[ \begin{array}{l} e_0 : evt \\ \text{押し/おす/ガ}(e_0, x_0) \end{array} \right] \\ u_2 : \left[ \begin{array}{l} e_1 : evt \\ u_3 : \text{開ける/あける/ガ}(e_1, x_0) \\ x_1(e_1) \end{array} \right] \\ cont(u_2, u_0) \end{array} \right]$$

図 1 lightblue による「押し開ける」の統語構造と意味表示

「太郎がドアを押し開ける」を例に挙げると、「押し開ける」という複合動詞は、「押す」という行為によって「開ける」という行為を達成することを意味する。図 1 の意味表示では、「押す」が連用節として扱われているため、そのような因果関係を表現することができない。すなわち、これらの動詞の関係は、「押すことで開ける」ではなく、「押しながら開ける」「押し、そして開ける」といった副詞的な解釈に限定される。

## 3 実装

### 3.1 提案手法

lightblue 上で複合動詞を適切に分析するために、Isono et al. [2] の glue element を実装する。\$記法について、 $X/\$$  は  $X, X/Y, X/Y/Z, \dots$ ,  $X \backslash \$$  は  $X, X \backslash Y, X \backslash Y \backslash Z, \dots$  といった統語範疇を一般化して表したものである。glue element に含まれる 3 つの \$ は同一の統語範疇を受け取らなければならないが、lightblue の実装では最終的な統語範疇が同一でなければならないという制限があるため、Isono et al. [2] による glue elements をそのまま実装することはできない (詳細については 5 節で述べる)。そのため、glue element から \$ を削除し、以下のような (レベル 2 の) 順関数交差置換規則を追加する。

$$\frac{(X/Y) \backslash Z \backslash W : f \quad Y \backslash Z \backslash W : g}{X \backslash Z \backslash W : \lambda x.\lambda y.(fxy)(gxy)} > S_{x2}$$

Isono et al. [2] において\$が行っていた項構造の共有は、順関数交差置換規則が行う。なお、(レベル2) 順関数交差置換規則は本研究の複合動詞の分析にのみ使用されるアドホックな規則ではなく、一般的な複文構造にも現れる規則である。

上記の方針に従って実装したところ、lightblue で glue element は以下のようになった。

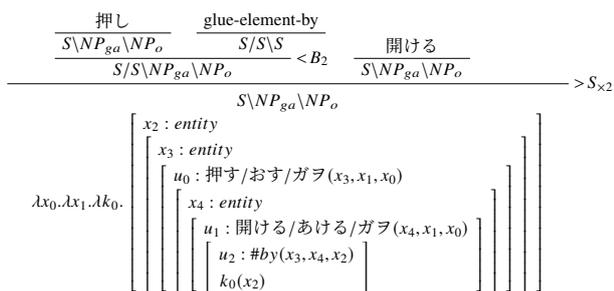
glue-element-parallel ⊢ S/S\S :

$$\lambda x_0.\lambda x_1.\lambda k_0. \left[ \begin{array}{l} x_2 : \text{entity} \\ x_0 \left( \lambda x_3.x_1 \left( \lambda k_1. \left[ \begin{array}{l} u_0 : \#parallel(x_3, k_1, x_2) \\ k_0(x_2) \end{array} \right] \right) \right) \end{array} \right]$$

glue-element-by ⊢ S/S\S :

$$\lambda x_0.\lambda x_1.\lambda k_0. \left[ \begin{array}{l} x_2 : \text{entity} \\ x_0 \left( \lambda x_3.x_1 \left( \lambda k_1. \left[ \begin{array}{l} u_0 : \#by(x_3, k_1, x_2) \\ k_0(x_2) \end{array} \right] \right) \right) \end{array} \right]$$

また、glue element 追加後の lightblue による「押し開ける」の解析結果を図 2 に示す。意味表示には、動作主が対象を「押す」ことによって「開いて」いることが記述されている。



が、最後に V2 が合成されている。

一方で、glue element を実装した lightblue による「語り明かす」の分析では、図 4 のように、まず V1 と V2 が glue element を介して合成され、その項と Theme1 が合成されている。この図では、V1 と V2 が合成された複合動詞はガ格名詞句を取って文になるが、理論的には、Theme2 を取るためにヲ格名詞句を取ったのちにガ格名詞句を取る必要がある。そのため、この分析では、複合動詞が Theme2 を取るができないという課題がある。

$$\frac{\frac{\text{語り}}{S \backslash NP_{ga} \backslash NP_o} \quad \text{glue-element-by} \quad \frac{S \backslash S}{S \backslash S} < B_2 \quad \frac{\text{明かす}}{S \backslash NP_{ga} \backslash NP_o}}{\frac{\text{思い出を}}{T_1 / (T_1 \backslash NP_o)} \quad \frac{S \backslash S \backslash NP_{ga} \backslash NP_o}{S \backslash NP_{ga} \backslash NP_o} > S_{\times 2}}{S \backslash NP_{ga}} >$$

図 4 本研究における「語り明かす」の分析

\$ 記法による統語範疇の定義は、Steedman [4] (p.23, (42)), 戸次 [8] (p.27, 定義 2.5.5) において与えられている。α を統語範疇とすると、α/\$ は（そして α/\$ も同様に）以下のような集合として再帰的に定義されている。

1.  $\alpha \in (\alpha/\$)$
2. 任意の統語範疇  $\beta, \gamma$  について、 $\beta \in (\alpha/\$) \implies (\beta/\gamma) \in (\alpha/\$)$

この定義は「/\$」部単独の定義ではなく、統語範疇 α に対して (α/\$) を統語範疇の集合として定義するものである。また、この α は統語素性まで含むものであることにも注意が必要である。

しかしこの定義には、実際の用例に照らすと二つの問題がある。

1. 用例においては、一つの統語範疇内に複数の (α/\$) が現れた場合、それらは同じものでなければならない。たとえば

$$S/NP \in (S/\$), S/NP/NP \in (S/\$)$$

であるが、

$$(S/NP)/(S/NP/NP) \in (S/\$)/(S/\$)$$

ではない。つまり (S/\$)/(S/\$) においては、左の (S/\$) と右の (S/\$) は同じものでなければならない。

2. \$記法は一般化関数合成規則 (Steedman [4] (p.35, (81)), 戸次 [8] (p.26, 定義 2.5.4)) のような規則を記述するために使われている。しかしこの規則では、たとえば順規則では前件部に  $Y/\$/Z$ , 帰結部に  $X/\$/Z$  が現れるが、\$ 部については共

有されていることが想定されている (Y と X は別範疇であるかもしれないにも関わらず)。したがって、一般化関数合成規則に現れる \$ は、上述の \$ 記法の定義においては未定義である。

この二つの問題を踏まえると、再帰的に定義可能な \$ 記法の定義が求められており、そして lightblue における T 範疇の実装の仕様は、この問題に対する理論的な答えの一つとして提示されているといえる。なお、lightblue では一般化関数合成規則は実装されておらず、その例化である  $>B, <B, >B_2, <B_2, >B_3, <B_3$  が直接与えられている。4 項以上を持つ述語は存在しないため、 $n \geq 4$  の場合は必要ない。よって、関数合成規則を一般の n について定義する必要はない。

一方、Isono et al. [2] の glue element は、一般化関数合成規則と同じ方針で \$ 記法を用いている。したがってこれを統合的な分析とするには、Steedman [4]・戸次 [8] の \$ 記法の定義を拡張する必要がある。

## 6 おわりに

本研究では、Isono et al. [2] による複合動詞の統語構造と意味合成の形式的な分析、特に glue element について、lightblue 上で実装を行い、複合動詞に関わる言語現象を自然言語推論データセットとして再構築した上で、NLI タスクによる評価を行った。その結果、glue element を実装した lightblue では、複合動詞を含む文の推論において顕著な性能向上がみられた。一方で、一部の日本語複合動詞には、現行の lightblue における \$ 記法とは合致しない glue element を必要とする場合があることも示した。また、2 種類の glue element は、統語範疇が完全に一致しており、意味表示についても、それぞれに出現する *parallel, by* 述語以外は一致している。そのため、複合動詞の分類に応じて、これら 2 種類の glue element を使い分けるための機構を実装することも、今後の課題の 1 つである。

## 謝辞

本研究の一部は JST CREST JPMJCR2565 および JSPS 科研費 JP23H03452 の支援を受けたものである。

## 参考文献

- [1] 影山太郎. **文法と語形成**. 日本語研究叢書 ; 第 2 期 第 4 卷. ひつじ書房, 1993.
- [2] Shinnosuke Isono, Takuya Hasegawa, Kohei Kajikawa, Koichi Kono, Shiho Nakamura, and Yohei Oseki. Formalizing argument structures with combinatory categorial grammar. 11 2022.
- [3] Mark Steedman. **The Syntactic Process**. MIT Press, 2000.
- [4] M. Steedman. **Surface Structure and Interpretation**. Linguistic Inquiry Series. MIT Press, 1996.
- [5] M Halle and Alec Marantz. **Distributed morphology and the pieces of inflection**, pages 111–176. The MIT Press, 1993.
- [6] Daisuke Bekki and Ai Kawazoe. Implementing variable vectors in a ccg parser. In Maxime Amblard, Philippe de Groote, Sylvain Pogodalla, and Christian Retoré, editors, **Logical Aspects of Computational Linguistics. Celebrating 20 Years of LACL (1996–2016)**, pages 52–67, Berlin, Heidelberg, 2016.
- [7] Daisuke Bekki and Koji Mineshima. **Context-Passing and Underspecification in Dependent Type Semantics**, pages 11–41. Studies in Linguistics and Philosophy. Springer Science and Business Media B.V., 2017.
- [8] 戸次大介. **日本語文法の形式理論 — 活用体系・統語構造・意味合成** —. くろしお出版, 2010.

## A Appendix

表 2 glue element 導入前の lightblue による推論結果

		正解ラベル			
		Yes	No	Unk	Other
推論結果	Yes	8	0	0	0
	No	0	0	0	0
	Unk	5	11	20	0
	Other	6	8	13	0

表 3 glue element 導入後の lightblue による推論結果

		正解ラベル			
		Yes	No	Unk	Other
推論結果	Yes	11	0	0	0
	No	0	4	0	0
	Unk	7	12	29	0
	Other	1	3	4	0