

日本語比較表現のための論理推論システムの構築

三上 燿輔¹ 松岡 大樹¹ 谷中 瞳¹

¹ 東京大学

{ymikami, daiki.matsuoka, hyanaka}@is.s.u-tokyo.ac.jp

概要

比較表現を含む自然言語推論 (NLI) は、比較対象となる性質が成り立つ程度を文から読み取り推論を行う必要があるため、挑戦的なタスクである。このような推論を頑健に扱える推論システムの実現に向けて、合成的意味論に基づく論理推論システムが提案されている。しかし、日本語の比較表現に注目したシステムの開発は十分に組み込まれていない。そこで本研究は、形式意味論に基づいた日本語比較表現のための論理推論システムを提案する。提案システムの評価には比較表現を含む日本語 NLI データセットを用いる。既存の LLM との正答率の比較により本提案の有効性を示す。

1 はじめに

自然言語推論 (Natural Language Inference; NLI)、または含意関係認識は、前提が仮説を含意するか否かを認識するタスクである。特に比較表現を含む NLI には (1) のような例がある。

- (1) P_1 : 太郎は次郎より重い。
 P_2 : 次郎は 70kg より重い。
 H : 太郎は 70kg より重い。(含意)

この推論を正しく導くには、70kg という数量表現、太郎と次郎の重さの程度の比較関係を捉え、大小関係の評価を行う必要がある。

このような挑戦的な言語現象に対して頑健な推論を実現するため、合成的意味論に基づく論理推論システムが提案されている [1, 2, 3, 4]。特に英語の比較表現を組合せ範疇文法 (Combinatory Categorical Grammar, CCG) [5] と degree semantics [6, 7] に基づいて扱った推論システム ccgcomp [8] は深層学習モデル BERT を大きく上回る性能を示している。しかしながら、日本語の比較表現に特有の現象を CCG と degree semantics でどの程度体系的に分析できるかは自明ではない。

そこで、本研究では CCG と degree semantics に基づいた日本語の比較表現のための論理推論システムを提案する。提案システムは、CCG に基づく統語解析とラムダ計算に基づく意味解析を用いて前提文と仮説文の意味表示を導出し、自動定理証明器で文間の含意関係を判定する。特に、前提 (presupposition) を伴う表現に対して multidimensional semantics [9] に基づく意味表示を導出するなど、比較表現に関して英語にはないが日本語にはある現象を体系的に扱えるように統語解析と意味解析の実装を行う。

評価には比較表現を含む日本語 NLI データセット JSeM [10] を用い、提案システムの有効性を示す。比較対象として GPT-4o¹⁾ を用いる。提案システムは研究利用可能な形で公開予定である。

2 形式意味論における比較表現

本研究では degree semantics という枠組みに基づいて比較表現の分析を行う。この枠組みにおいて、段階的な形容詞は二項述語で表される。例えば「重い」という形容詞は $\text{heavy}(x, d)$ と表され、 x の重さが d 以上である場合に真、それ以外の場合に偽となる。この述語を用いることで、太郎の体重が 70kg 以上であることは $\text{heavy}(\text{taro}, 70\text{kg})$ と表すことができる。

また、本研究では A-not-A 分析 [7] と呼ばれる分析に基づいて、「太郎は次郎より重い。」のような表現を以下 (2) のような一階述語論理式で表す。この論理式は、太郎の体重が d 以上、次郎の体重が d 未満であるような d の存在によって両者の大小関係を記述している。

- (2) $\exists d. (\text{heavy}(\text{taro}, d) \wedge \neg \text{heavy}(\text{jiro}, d))$

この枠組みに基づいた基本的な比較表現の意味表示の例を表 1 に示す。

1) <https://openai.com/index/gpt-4o-system-card/>

比較表現を伴う文	意味表示
太郎は重い.	$\text{heavy}(\text{太郎}, \theta)$
太郎は 70kg より重い.	$\exists d. (\text{heavy}(\text{太郎}, d) \wedge d > 70\text{kg})$
太郎はすべての学生より重い.	$\forall x. (\text{student}(x) \rightarrow \exists d. (\text{heavy}(\text{太郎}, d) \wedge \neg \text{heavy}(x, d)))$

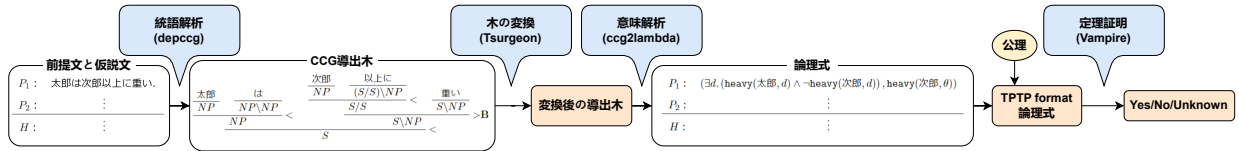


図 1 提案システムの概観

3 提案システム

3.1 システムの概観

図 1 に本研究で提案する推論システムの概観を示す。本節では推論の各ステップについて説明する。

統語解析 このステップではまず、CCG パーサの depccg [11] を用いて入力文から CCG 導出木を構築する。次に Tsurgeon [12] を用いて必要に応じて 3.2.1 で導入する cmp トークンの追加、導出木の修正を行う。

意味解析 このステップでは統語解析で得た CCG 導出木の各葉に対して意味テンプレートで定義された意味表示を割り当てる。そして、CCG の組合せ規則に従って導出木の葉から根方向に向かって文全体の意味を合成する。意味表示の割り当て、合成には ccg2lambda [13] を用いる。

定理証明 このステップでは、意味解析で得た論理式と、degree semantics や同等比較構文に関連する 5 種類の公理を自動定理証明器である Vampire [14] に入力してシステム全体の出力を決定する。追加した公理の 1 つである対義語に関する公理 ANT を表 2 に示す。P は long などの正の形容詞、N は short などの負の形容詞を示している。その他の公理は付録 A を参照。公理を追加した上で、これらの論理式を Vampire が採用している Thousand of Problems for Theorem Provers (TPTP) [15] フォーマットの型付き一階述語論理式に変換する。そして、前提から仮説が証明できた場合は“yes” (含意)、前提から仮説の否定が証明できた場合は“no” (矛盾)、いずれでもない場合は“unknown” (中立) を出力する。

表 2 追加した公理 (例は公理の導入により解けるようになる推論例)

ANT	$\forall x d. (\mathbf{P}(x, d) \leftrightarrow \neg \mathbf{N}(x, d))$
例	P_1 : 太郎は次郎より重い。 P_2 : 次郎は重い。 H : 太郎は軽い。 (矛盾)

3.2 意味合成の詳細

以下では日本語の比較表現に特有の現象及び本研究におけるそれらの現象の扱いについて説明する。

3.2.1 明示的な比較形態素の欠如

英語の比較級には -er や more といった比較を明示する形態素が存在する。一方で、(3) に示すように、日本語の形容詞には比較級がない。

- (3) a. 太郎は次郎より重い。
- b. 太郎は重い。

そのため、(3a), (3b) の形容詞「重い」には同じ意味表示が割り当てられる。両方の文で同じ意味表示の形容詞を用いつつ文全体の意味表示を区別するために、発音されない記号 (空範疇) を導入する。具体的には、(3b) のように「... より」や「... 以上に」といった比較表現がない場合は、代わりにカテゴリ S/S の空範疇 cmp を挿入する。これにより (3b) の「重い」の意味表示に先述の比較の基準 θ を導入することができる。

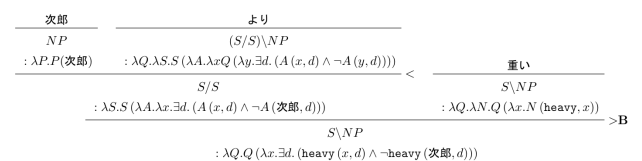


図 2 「次郎より重い」の意味合成

cmp	重い	
S/S	$S \setminus NP$	
$: \lambda S.S(\lambda A.\lambda x.A(x, \theta))$	$: \lambda Q.\lambda N.Q(\lambda x.N(\text{heavy}, x))$	
$S \setminus NP$		$>B$
$: \lambda Q.Q(\lambda x.\text{heavy}(x, \theta))$		

図3 「cmp 重い」の意味合成

3.2.2 同等比較構文

同等比較構文は、2つの対象が同程度にある性質を満たす場合に使われる構文である。例えば“John is as heavy as Bob.”は「ジョンの体重はボブの体重以上である」ということを意味し、既存研究 [8] はこの文の意味表示を以下 (4) のようにしている。

$$(4) \quad \forall d. (\text{heavy}(\text{Bob}, d) \rightarrow \text{heavy}(\text{John}, d))$$

一方で、日本語の同等比較構文を用いた「太郎は次郎と同じくらいの重さだ」という表現には二者の重さの大小関係に関する含意はない²⁾。そこで、本研究では日本語の同等比較構文に対して以下 (5) のような意味表示を提案する。これは、直感的には両者の重さの差が δ 未満であることを示している。

$$(5) \quad \forall d_1 d_2. ((\neg (\text{heavy}(\text{taro}, d_1) \leftrightarrow \text{heavy}(\text{jiro}, d_1)) \wedge \neg (\text{heavy}(\text{taro}, d_2) \leftrightarrow \text{heavy}(\text{jiro}, d_2))) \rightarrow (|d_1 - d_2| < \delta))$$

また、(6) のような推論を可能にするため、 θ と δ の関係について以下の公理 (7) を導入する。こちらは直感的に、 θ から絶対値 δ 以内の範囲では heavy の真偽が変化しない程度に δ は小さいことを表している。

$$(6) \quad P_1: \text{太郎は次郎と同じくらいの重さだ。}$$

$$P_2: \text{太郎は重い。}$$

$$H: \text{次郎は重い。 (含意)}$$

$$(7) \quad \forall x. (\text{heavy}(x, \theta - \delta) \leftrightarrow \text{heavy}(x, \theta + \delta))$$

3.2.3 節比較構文

節比較構文とは、従属節を伴う比較構文であり、(8a) のようなものがある³⁾。また、それに関連して (8b), (8c) のような文も扱う。

- (8) a. 太郎は花子買ったより高い本を買った。
 b. 太郎は花子買ったのより高い本を買った。
 c. 太郎は花子より高い本を買った。

これら3つの文には degree semantics に基づいて (9) のような意味表示を割り当てる。

$$(9) \quad \exists d. (\exists x. (\text{book}(x) \wedge \text{expensive}(x, d) \wedge \exists e. (\text{bought}(e) \wedge (\text{Nom}(e) = \text{taro}) \wedge (\text{Acc}(e) = x))) \wedge \neg \exists x. (\text{book}(x) \wedge \text{expensive}(x, d) \wedge \exists e. (\text{bought}(e) \wedge (\text{Nom}(e) = \text{hanako}) \wedge (\text{Acc}(e) = x))))$$

3.2.4 前提 (presupposition)

日本語の比較表現には前提 (presupposition) という特殊な意味内容を伴うものがある [17, 18]。前提とは、特定の表現に伴う、否定やモダリティの影響を受けない意味のことである [19]。例として、英語の“know”という述語を考える。以下の (10) の場合、ボブが走ったことが前提となる。

$$(10) \quad \text{John knows that Bob ran.}$$

ここで、(11) のように文を否定するとジョンが知っていることは否定されるが、ボブが走ったことは否定されない。

$$(11) \quad \text{John does not know that Bob ran.}$$

したがって、文 (10) の意味表示として“John knows that Bob ran.”の意味表示と“Bob ran.”の意味表示を単に論理積で結んだものを採用するのは適切でない。

日本語の比較表現にも同様に前提を生じさせる表現がある。以下の3つの文はいずれも二者の比較であるが、すべての文に「次郎は重い。」という前提が生じている。また、文 (11) と同様に、これらの文に否定やモダリティを足しても前提はその影響を受けない。

- (12) a. 太郎は次郎以上に重い。
 b. 太郎は次郎と同じくらい重い。
 c. 太郎は次郎ほど重くない。

このような前提を伴う比較表現に意味表示を与えるため、本研究では multidimensional semantics [9] という枠組みを採用する。この枠組みでは文全体の意

2) 英語と同じ解釈を可能にするためには「太郎は次郎と少なくとも同じくらいの重さだ」のように表す必要がある。

3) 日本語には本質的に節比較構文は存在しないという主張もある [16] が、本研究では簡単のためにこれを節比較構文として扱う。

味表示を、主たる意味表示と前提の意味表示の組で表す。例えば文 (12a) の意味表示は以下 (13) のようになる。

$$(13) \quad \left(\exists d. (\text{heavy}(\text{taro}, d) \wedge \neg \text{heavy}(\text{jiro}, d)), \right. \\ \left. \text{heavy}(\text{jiro}, \theta) \right)$$

これにより、意味合成の過程で否定やモダリティが主たる意味表示にしか影響しないようにすることが可能になる (4)。

重い	というわけではない
$S \setminus NP$	$S \setminus S$
$: \lambda Q. \lambda N. (Q(\lambda x. N(\text{heavy}, x)), T)$	$: \lambda S. \lambda N. (\neg (S(N))_1, (S(N))_1)$
$S \setminus NP$	
$: \lambda Q. \lambda N. (\neg Q(\lambda x. N(\text{heavy}, x)), Q(\lambda x. N(\text{heavy}, x)))$	

<B

図 4 (12a) の否定の意味合成の一部

4 実験

4.1 データセット

本研究では評価用データセットとして JSeM データセットの Comparatives セクションを用いる。テンス・アスペクトの問題は本研究では扱っていないため除外した。モダリティも本研究では扱っていないが、JSeM のモダリティを含む問題には前提に影響を与えないという性質を扱った問題のみが収録されているため、これらの問題についてはモダリティを否定に置き換えた問題に差し替えて評価を行った。例えば、「太郎は次郎以上に早起きかもしれない。」という文は「太郎は次郎以上に早起きというわけではない。」という文に変更した。データセットの問題数は全部で 71 問、正解ラベルの分布は (yes/no/unknown)=(42/8/21) である。表 3 にデータセット中の問題例を示す。

提案システムの比較対象として、GPT-4o を用いる。使用したプロンプトは付録 B を参照。

4.2 結果と分析

実験結果を表 4 に示す。Majority は最多数のラベルである yes を常に回答した場合の正答率である。表に示す通り、提案システムの正答率は GPT-4o を上回っている。GPT-4o は表 3 の jsem-620 のように前提の存在を考慮しなければ正答できない問題に対して前提の存在を無視した場合と同じ解答を出力していた。一方で、提案システムは正答できていた。

表 3 JSeM データセットの問題例

jsem-569	
前提 1	PC-6082 は ITEL-XZ より速い。
前提 2	ITEL-XZ は速い。
仮説	PC-6082 は速い。
正答	Yes
jsem-620	
前提 1	太郎は花子以上に早起きだ。
仮説	花子は早起きだ。
正答	Yes

表 4 JSeM データセットでの精度

問題数	Majority	GPT-4o	提案手法
71	.592	.760	.845

提案システムが正答できなかった問題の一つに「曲がっている」のような絶対的形容詞を含む問題 (14) があった。絶対的形容詞は事物の性質を表す形容詞であり、この問題例のように、「重い」、「賢い」などの相対的形容詞では成立しない含意関係が成立する場合がある。そのため、絶対的形容詞と相対的形容詞を区別するための意味テンプレートの改良と公理の導入が必要である。

- (14) *P*: この棒はあの棒より曲がっている。
H: この棒は曲がっている。(含意)

提案システムは、最上級表現を含む問題 (15) も正答できなかった。これに対応するには「...の中で」や「一番」に対する意味表示の導入が必要になる。

- (15) *P*: 太郎はこのクラスの学生の中で一番背が高い。
H: 太郎はこのクラスの他のどの学生よりも背が高い。(含意)

5 おわりに

本研究では CCG と degree semantics に基づいて、日本語比較表現のための論理推論システムを構築した。実験では比較表現を含む日本語 NLI データセットを使用し、提案システムが GPT-4o より高い正答率を示すことを確認した。今後の課題として、絶対的形容詞や最上級を含む表現に関わる推論に対応することなどが挙げられる。

謝辞

本研究は JST さきがけ JPMJPR21C8 の支援を受けたものである。

参考文献

- [1] Lasha Abzianidze. Natural solution to fracas entailment problems. In **Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (* SEM 2016)**, pp. 64–74. Association for Computational Linguistics (ACL), 2016.
- [2] Lasha Abzianidze. A tableau prover for natural logic and language. In **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**. Association for Computational Linguistics (ACL), 2015.
- [3] Koji Mineshima, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki. Higher-order logical inference with compositional semantics. In **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, pp. 2055–2061, 2015.
- [4] Pascual Martínez Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. On-demand injection of lexical knowledge for recognising textual entailment. In **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers**, pp. 710–720, 2017.
- [5] Mark Steedman. **The syntactic process**. MIT press, 2001.
- [6] Max J Cresswell. The semantics of degree. In **Montague grammar**, pp. 261–292. Elsevier, 1976.
- [7] Ewan Klein. A semantics for positive and comparative adjectives. **Linguistics and philosophy**, Vol. 4, pp. 1–45, 1980.
- [8] Izumi Haruta, Koji Mineshima, and Daisuke Bekki. Implementing natural language inference for comparatives. **Journal of Language Modelling**, Vol. 10, No. 1, p. 139–180, Nov. 2022.
- [9] Lauri Karttunen and Stanley Peters. Conventional Implicature. In **Presupposition**, pp. 1–56. Brill, 1979.
- [10] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. An inference problem set for evaluating semantic theories and semantic processing systems for japanese. In Mihoko Otake, Setsuya Kurahashi, Yuiko Ota, Ken Satoh, and Daisuke Bekki, editors, **New Frontiers in Artificial Intelligence**, pp. 58–65, Cham, 2017. Springer International Publishing.
- [11] Masashi Yoshikawa, Hiroshi Noji, and Yuji Matsumoto. A* ccg parsing with a supertag and dependency factored model. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 277–287. Association for Computational Linguistics, 2017.
- [12] Roger Levy and Galen Andrew. Tregex and tsurgeon: Tools for querying and manipulating tree data structures. In **LREC**, pp. 2231–2234. Genoa, 2006.
- [13] Pascual Martínez-Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. ccg2lambda: A compositional semantics system. In **Proceedings of ACL-2016 System Demonstrations**, pp. 85–90, 2016.
- [14] Laura Kovács and Andrei Voronkov. First-order theorem proving and vampire. In **International Conference on Computer Aided Verification**, pp. 1–35. Springer, 2013.
- [15] Geoff Sutcliffe. The tptp problem library and associated infrastructure: from cnf to th0, tptp v6. 4.0. **Journal of Automated Reasoning**, Vol. 59, No. 4, pp. 483–502, 2017.
- [16] Yasutada Sudo. Invisible degree nominals in japanese clausal comparatives. In **Proceedings of the 5th Workshop on Altaic in Formal Linguistics**. MITWPL, Cambridge, Mass, pp. 285–295, 2009.
- [17] J-R Hayashishita. Izyoo (ni)-and gurai-comparatives: Comparisons of deviation in japanese. **GENGO KENKYU (Journal of the Linguistic Society of Japan)**, Vol. 132, pp. 77–109, 2007.
- [18] Yusuke Kubota. The presuppositional nature of izyoo (-ni) and gurai comparatives: A note on hayashishita (2007). **GENGO KENKYU (Journal of the Linguistic Society of Japan)**, Vol. 141, pp. 33–47, 2012.
- [19] Christopher Potts. Presupposition and implicature. **The handbook of contemporary semantic theory**, pp. 168–202, 2015.

A 公理

表 2 で紹介した公理 ANT 以外にも以下のような公理を追加した。

表 5 表 2 以外に追加した公理

CP	$\forall xy. ((\exists d. (\mathbf{A}(x, d) \wedge \neg \mathbf{A}(y, d))) \rightarrow \forall d. (\mathbf{A}(y, d) \rightarrow \mathbf{A}(x, d)))$
UP	$\forall xd. (\mathbf{P}(x, d) \rightarrow \forall d'. (d' \leq d \rightarrow \mathbf{P}(x, d')))$
DOWN	$\forall xd. (\mathbf{N}(x, d) \rightarrow \forall d'. (d' \geq d \rightarrow \mathbf{N}(x, d')))$
DELTA	$\forall x. (\mathbf{A}(x, \theta - \delta) \leftrightarrow \mathbf{A}(x, \theta + \delta))$

CP は一貫性の公理であり、直感的には x の程度が y の程度より大きい場合、 x の程度は y の程度以上であることを表している。この公理の導入により (1) のような推論が可能になる。具体的には CP と P_1 から図 5 に示す証明木が得られ、これに P_1, P_2 を合わせることで H が導出される。

$$\frac{\frac{\frac{\forall xy. ((\exists d. (\text{heavy}(x, d) \wedge \neg \text{heavy}(y, d))) \rightarrow \forall d. (\text{heavy}(y, d) \rightarrow \text{heavy}(x, d)))}{(\exists d. (\text{heavy}(\text{太郎}, d) \wedge \neg \text{heavy}(\text{次郎}, d))) \rightarrow \forall d. (\text{heavy}(\text{次郎}, d) \rightarrow \text{heavy}(\text{太郎}, d))} (\forall \mathbf{E})}{\exists d. (\text{heavy}(\text{太郎}, d) \wedge \neg \text{heavy}(\text{次郎}, d))} (\rightarrow \mathbf{E})}{\frac{\forall d. (\text{heavy}(\text{次郎}, d) \rightarrow \text{heavy}(\text{太郎}, d))}{\text{heavy}(\text{次郎}, \theta) \rightarrow \text{heavy}(\text{太郎}, \theta)} (\forall \mathbf{E})} (\rightarrow \mathbf{E})$$

図 5 証明木の一部

UP/DOWN はそれぞれ正の形容詞・負の形容詞の単調性を表している。つまり UP は d が $\mathbf{P}(x, d)$ を満たす場合にはそれ以下の程度 d' も $\mathbf{P}(x, d')$ を満たすことを表しており、DOWN は d が $\mathbf{N}(x, d)$ を満たす場合にはそれ以上の程度 d' も $\mathbf{N}(x, d')$ を満たすことを表している。これらの公理と公理 DELTA を合わせることで (6) の含意関係を認識することができる。具体的には以下のように仮説を証明する。まず、 P_2 と UP および DELTA により $\text{heavy}(\text{taro}, \theta)$, $\text{heavy}(\text{taro}, \theta + \delta)$ が導出される。次に、 P_1 の意味表示 (5) において d_1, d_2 をそれぞれ $\theta + \delta, \theta$ とし、対偶を考えることによって $\text{heavy}(\text{taro}, \theta + \delta) \leftrightarrow \text{heavy}(\text{jiro}, \theta + \delta)$ または $\text{heavy}(\text{taro}, \theta) \leftrightarrow \text{heavy}(\text{jiro}, \theta)$ が言える。いずれの場合も仮説 $\text{heavy}(\text{jiro}, \theta)$ が導かれる。

B プロンプト

GPT-4o を用いたベースライン実験では以下のようなプロンプトを用いた。

表 6 使用したプロンプト例

system	前提文と仮説文が与えられます。前提文が仮説文を含意しているか答えてください。 「含意」、「矛盾」、「中立」のいずれかで答えてください。
user	前提 1 : PC-6082 は ITEL-XZ より速い。 前提 2 : ITEL-XZ は速い。 仮説 : PC-6082 は速い。