言語学的パイプラインに基づく妥当性の検証：金融テキストを対象として

松原舞 [1]　外園康智 [2]　角田充弘 [2]　田村光太郎 [2]
大洞日音 [1]*　富田朝 [1]　戸次大介 [1]
[1] お茶の水女子大学大学院　[2] 野村総合研究所
{matsubara.mai,daido.hinari,tomita.asa,bekki}@is.ocha.ac.jp
{y-hokazono,m1-tsunoda,k9-tamura}@nri.co.jp

概要

本稿では理論言語学に基づく自然言語推論システムである言語学的パイプラインを実社会タスクに応用する試みとして，金融テキストを対象としたファクトチェックシステムを提案する．具体的には，自然言語の文を高階論理の意味表現に変換し，自動定理証明を行うことで，顧客アンケートとエキスパートの回答の整合性を検証する．本手法では検証結果として証明図を伴うため，その結果に信頼性を持たせることができる．実験の結果，Accuracy は 76.7% に達し，金融テキストを対象とする妥当性の検証手法として本手法の有効性を示した．

1　はじめに

人工知能および自然言語処理の分野において，大規模言語モデル (Large Language Models: LLM) の進化は加速の一途を辿っている [1]．LLM の急速な発展に伴い，社会経済や科学研究における LLM への依存度は今後さらに高まると予想される．しかし，実社会への実装が進むにつれて，現在の LLM を組み込んだシステムは二つの本質的な課題に直面している．

検証手法の問題　LLM による要約や翻訳，調査結果の正確性を定量的に評価することは極めて困難である．人手による正解データの作成はコストが高く，BLEU や COMET といった既存の自動評価指標も生成文の論理的妥当性を測るには不十分である．

改善手法の問題　出力結果に誤りや不満があった場合，ブラックボックスである LLM に対してどのように修正を適用し，精度低下をモニタリングすべきかという解決策は非自明である．

これらの問題は，LLM の推論過程が不透明であることに起因しており，LLM 同士による相互評価や内部表現分析といったアプローチも，評価者が LLM である以上，循環論法に陥る危険性を孕んでいる．

これに対し，自然言語理解の言語学的研究には，形式統語論や形式意味論，数理論理学に基づき長年研究されてきた「意味の理論」の蓄積がある．特に 2000 年代中盤以降，これら理論言語学の知見を実装した「言語学的パイプライン」は著しい進展を遂げている．言語学的パイプラインは，人間の言語機能の分析や意味原理に基づいて推論を行うため，計算履歴がそのまま言語理論による説明となり反証可能性を持つ，という LLM にはない特徴を有する．

本研究では，この「説明可能な」言語学的パイプラインを実社会タスクに応用する試みとして，金融テキストを対象としたファクトチェックに取り組んだ．高度な専門知識と論理的整合性が求められる実務において，LLM のみを用いたアプローチではハルシネーションや論理矛盾のリスクが払拭できない．そこで本研究では，対象文書を言語学的パイプラインによって高階論理の意味表現へ変換し，整合性や矛盾を自動検出する手法を提案する．本手法では，整合性が確認された場合には「証明図」が出力され，判定結果に高い信憑性が与えられる．本論文では，この LLM と言語学的パイプラインの協働によるファクトチェックシステムの構築手法と，その有効性と課題について論じる．

* 本研究は著者が独自に実施したものであり，Amazon の見解や立場を反映するものではない．

2 先行研究

本節では先行研究として，言語学的パイプラインを用いた推論システムを紹介する．

言語学的パイプラインのさきがけとなったのは Bos [2] による Boxer である．談話表示理論 (Discourse Representation Theory: DRT) [3] による意味表示を CCG 統語解析器 [4] を用いて計算する．得られた DRS に対して，λ 計算を行うことで一階論理 (First-Order Logic: FOL) の論理式に変換し，標準的な定理証明器を用いることで自動推論を実現している．

Abzianidze [5] では，CCG 統語解析器 C&C parser [6] により計算された統語構造からラムダ論理形式 (Lambda Logical Forms: LLF) を生成し，それをもとに Natural Tableaux [7] を用いて推論を行う一連の自動推論のパイプラインが提案された．

しかし，Bos [2] と Abzianidze [5] には一般化量化子などの複雑な言語現象を正しく扱うことができないという共通の課題がある．これは意味論において自動証明が可能な体系を採用していることに起因する．

Chatzikyriakidis and Luo [8] では現代的型理論 (Modern Type Theory: MTT) の一種である統一依存型理論 (Unified Theory of dependent Types: UTT) [9] に基づく MTT 意味論 (MTT Semantic) [10] をもとに形式的な意味を表現し，すでに MTT が実装されている証明支援系 Coq [11] を用いる半自動推論が提案された．MTT は高階論理 (Higher-Order Logic: HOL) であるため量化子などの複雑な言語現象を扱うことができるが，統語解析器を使用しておらず，自然言語から形式的な意味への変換は手動で行われている．

この点を改善したのが Mineshima et al. [12] による ccg2lambda[1] である．ccg2lambda は CCG 統語解析器 C&C parser [6] を証明支援系 Coq [11] と接続することで，英語テキストにおいて複数の前提文と仮説文の間の含意関係を自動で判定するシステムである．統語解析の結果をもとに HOL に基づいた意味解析を行い，HOL による意味表示を用いて推論が行われる．また，Mineshima et al. [13] では CCG 統語解析器 Jigg [14] と証明支援系 Coq [11] との接続を提案し，これにより ccg2lambda の日本語対応が実現した．

Chatzikyriakidis and Luo [8] や Mineshima et al. [12], Mineshima et al. [13] では HOL を採用しているため，一般化量化子などの FOL では表現できない言語現象も扱うことができる．その一方で HOL はその記述力の高さから決定不能であるため，一部の単純な証明を除く多くの複雑な証明は手動で行われるか [8]，もしくはタイムアウトとして扱われている [12, 13]．

また，いずれの提案においても SICK [15] や FraCaS [16]，JSeM[2] [17, 18] など標準的なデータセットを用いた評価が行われている．

3 提案手法

3.1 推論システム

本システムで採用している言語理論について紹介した後，推論システムのパイプラインの概要について述べる．

3.1.1 組合せ範疇文法

組合せ範疇文法 (Combinatory Categorial Grammar: CCG) [19] は理論言語学の統語論としての説明的妥当性と、頑健で高速な統語解析器の実装を併せ持つ．

3.1.2 依存型意味論

依存型意味論 (Dependent Type Semantics: DTS) [20, 21] は，依存型理論 (Dependent Type Theory: DTT) [22] に基づいた自然言語のための証明論的意味論である．DTT には項に依存した型を記述できるという特徴があり，DTS は DTT の型として文の意味を記述する．$S_1, \ldots, S_n$ において，S_n に含まれる照応の先行詞を探索する際に $S_1, \ldots, S_{n-1}$ の意味を与えることで，$S_1, \ldots, S_{n-1}$ の意味に依存した S_n の意味を記述することが可能である．

3.1.3 推論システムのパイプライン

本研究では CCG 統語・意味解析器 lightblue [23, 24] と自動定理証明器 wani [25, 26] を組み合わせた推論システム [27, 28] を用いる．この推論システムのパイプラインの概要を図 1 に示す．

入力された日本語文に対して，CCG 統語・意味解析器 lightblue による日本語 CCG[24] に基づいた構文解析を行う．構文解析時に使用される組合せ規則は統語的な振る舞いだけでなく，語の意味合成の計算

1) https://github.com/mynlp/ccg2lambda

2) https://github.com/DaisukeBekki/JSeM

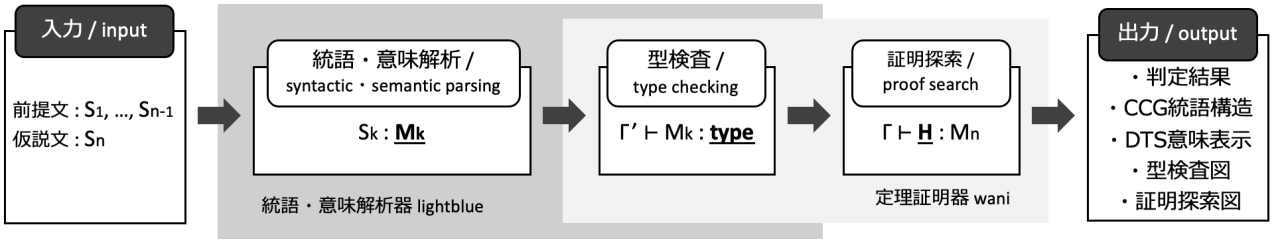

図 1　推論システムのパイプライン．$S_k\colon M_k$ は自然言語の文 S_k の意味表示が M_k であることを意味し，前提文と仮説文の間に含意関係があれば yes，矛盾関係があれば no，どちらでもなければ unknown，型検査に失敗し証明探索が行われない場合は other と判定する．$1 \leqq k \leqq n$，$\Gamma' \equiv S_1\colon M_1, \ldots, S_{k-1}\colon M_{k-1}$，$\Gamma \equiv S_1\colon M_1, \ldots, S_{n-1}\colon M_{n-1}$ とする．

方法も指定するため，構文解析の結果をもとに DTS に基づく意味合成を行うことが可能である．

また，lightblue には型検査という解析結果の整合性を保証するための機能がある．具体的には，DTS の文の意味の型は type でなければならないという制約を満たしているかを検査する．さらに，文 S_k の意味を表す意味表示 M_k の中に，聞き手にとって未知の意味を記述する演算子である未指定項 (underspecified term) が含まれている場合には，型検査を行う際に自動定理証明器 wani を呼び出し，先行詞を探索する照応解決も同時に行われる．

入力されたすべての文が型検査に成功したら，含意関係を判定するために自動定理証明器 wani による証明探索が行われる．DTS では文の意味は型で記述されるが，カリー＝ハワード対応により，型理論における環境・項・型はそれぞれ証明論における前提・証明・帰結と対応するため，前提文 $S_1, \ldots, S_{n-1}$ のもとで仮説文 S_n が成り立つか，つまり前提文と仮説文の間に含意関係が成り立つかを判定することは前提文 $S_1\colon M_1, \ldots, S_{n-1}\colon M_{n-1}$ のもとで命題 M_n を持つ証明 H を探索することに帰結する．また，前提文と仮説文の矛盾関係を検知するには，命題 $\neg M_n$ を持つ証明 H を探索する．

この推論システムには大きく分けて統語・意味解析，型検査，証明探索の 3 つの工程があり，それぞれの工程の計算過程を表す CCG 統語構造，DTS 意味表示，型検査図，証明探索図および含意関係の判定結果を出力する．これにより，LLM を用いた推論ではブラックボックスとなってしまう推論過程を追うことが可能となる．また，証明探索により含意関係の判定を行う際には，与えられた前提からは何が言えるのか・証明したい命題が真であるためには何が必要なのかを探索するため，証明を構築する上で判定の根拠は前提のどの部分であるかという説明を伴うという特徴もある．

3.2　データセット

データセットは，資産運用や金融商品投資への考えを問う，アンケートの自由記述と，エキスパートによる回答のペアデータ 200 件を母集団とした．そこから，機微情報（個人を特定し得る情報や特定の機関・商品、秘匿性の高い情報）が含まれないデータのみを対象とし，かつ，顧客の意図や求める支援内容，対応方針が実質的に同一なものを重複として省き，20 件を選定した．

アンケート文および回答文に複数文構成や表現揺れが含まれるため，LLM を用いて表現を正規化した．具体的には，文を 1 文に整形し，主語の省略や語彙の揺れ（同義語・言い換え）を統一することで，後続の論理表現や比較が可能な文へ変換した．この正規化後のアンケート文を A，正規化後の回答を D と定義した．

さらに，A と D だけでは含意関係の分析に必要な要素が不足するため，補助的なデータを付与した．顧客が期待する支援内容については，別途整備された「個人が期待する支援リスト」から該当する項目を人手で選択し，これを B とした．同様に，業務上提供可能な対応手段については，「提供手段リスト」から該当する手段を人手で選択し，これを C とした．いずれも，自由記述を直接ラベル化するのではなく，定義済みリストの中から最も適合する項目を選ぶ運用とすることで，付与基準の一貫性を確保した．

ただし，A・B・C だけでは，D が一意に導かれるとは限らない．そこで，A・B・C から D へ至る推論を可能にするための追加知識を明示的に導入する．具体的には，推論の中間段階を表す条件文を E とし，提供手段 (C) と中間段階 (E) が満たされると

最終行為 (D) が導かれる条件文を F として LLM により生成した．以上により，各事例を A〜F の 6 要素（各 1 文）で表現した 30 組のデータセットを構築した．いずれも正解ラベルは yes（含意）とする．

lightblue はサ行変格活用動詞を含む文の解析に課題が残されており，本稿では構築したデータセットについて，サ行変格活用動詞が含まれる場合は，サ行変格活用動詞ではない類似の語に置き換えて使用した（例：作成する → 作る）．

3.3 推論の構造

§3.1.3 で紹介した言語学的パイプラインを用いて，金融テキストを対象としたファクトチェックを試みる．推論の構造は図 2 の通りである．

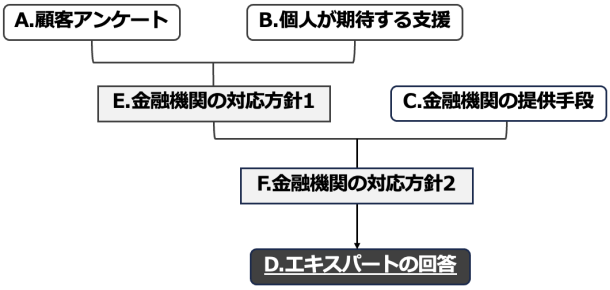

図 2 推論の構造

前提文として「A. 顧客アンケート」「B. 個人が期待する支援」「C. 金融機関の提供手段」を，公理として世界知識である「E. 金融機関の対応方針 1」「F. 金融機関の対応方針 2」を与えた際に，仮説文である「D. エキスパートの回答」を含意しているか判定する．金融機関の方針 1 はある状態の顧客に対する金融機関がとるべき対応、金融機関の方針 2 は金融機関の対応や手段を言い換える公理である．含意関係が認められた際には，エキスパートによる回答は，顧客アンケートの内容に対して金融機関の提供手段や対応方針に沿った妥当な対応案であるとする．

また DTS では「かつ」は単純型付き λ 計算の直積型 $A \times B$ の一般化である Σ 型，「ならば」は関数型 $A \to B$ の一般化である Π 型によって表現される．[3] つまり「ϕ かつ ψ ならば χ」は $(\phi \times \psi) \to \chi$ と表現されるが，$(\phi \times \psi) \to \chi \Leftrightarrow \phi \to (\psi \to \chi)$ であることより，本稿では E および F は $\phi \to (\psi \to \chi)$ として実装した．

3) Σ 型は $\left[\begin{array}{c} x : A \\ B \end{array}\right]$，$\Pi$ 型は $(x\colon A) \to B$ と表記する．

4 結果

本システムでは統語・意味解析や照応解決の曖昧性を加味して複数の解析結果を出力するが，本稿では最上位の解析結果のみを分析の対象として実験を行った．

全 30 題のうち，A〜F の全ての文で型検査に成功したのは 28 題だった．28 題について証明探索を行った結果，含意関係が認められると判定されたのは 23 題 (76.7%) であった．

含意関係が認められた問題の例は以下の通りである．

【A】 顧客は目標に届くか不安だ。
【B】 顧客は進み具合が分かると安心する。
【C】 金融機関は進捗メモを作る。
【D】 金融機関は半年に 1 回進捗メモを顧客に送る。
【E】 顧客は目標に届くか不安かつ進み具合が分かると安心するならば、金融機関は進み具合を知らせる。
【F】 金融機関は進み具合を知らせるかつ進捗メモを作成するならば、半年に 1 回進捗メモを顧客に送る。

推論が不正解だった 5 問は，いずれも妥当な統語・意味解析および型検査が行われており，証明が見つかる（含意関係が認められる）ことが望ましい．これは wani の証明探索に起因するエラーである．型検査に失敗した 2 問は，いずれも時制 (tense) に関する統語・意味解析にエラーの原因があり，これの改善により妥当な推論が可能になると考えられる．

5 おわりに

本稿では，言語学的パイプラインに基づく推論システムを金融テキストのファクトチェックに応用する手法を提案し，ブラックボックスになりがちな LLM 推論とは異なる，説明可能な検証枠組みを実現した．

一方で，統語・意味解析における tense の分析，自動定理証明器における証明探索の失敗など，実用化に向けた課題も明らかになった．

今後は整合性の判定だけでなく，矛盾関係の判定の検証も行い，実社会における高い信頼度を持つ推論システムの応用可能性をさらに検討していきたい．

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] R. Thomas McCoy, Shunyu Yao, Dan Friedman, Mathew D. Hardy, and Thomas L. Griffiths. Embers of autoregression show how large language models are shaped by the problem they are trained to solve. **Proceedings of the National Academy of Sciences**, 121(41):e2322420121, 2024.

[2] Johan Bos. Wide-coverage semantic analysis with Boxer. In **Semantics in Text Processing. STEP 2008 Conference Proceedings**, pages 277–286, 2008.

[3] Hans Kamp. A theory of truth and semantic representation. In **Formal Methods in the Study of Language**, pages 277–322, 1981.

[4] Stephen Clark and James R. Curran. Parsing the WSJ using CCG and log-linear models. In **Proceedings of the 42nd Meeting of the ACL (to appear)**, pages 103–110, 2004b.

[5] Lasha Abzianidze. A Tableau Prover for Natural Logic and Language. In **EMNLP2015**, pages 2492–2502, 2015.

[6] S. Clark and J. R. Curran. Widecoverage efficient statistical parsing with CCG and log-linear models. **Computational Linguistics**, 33(4):493–552, 2007.

[7] Reinhard Muskens. An analytic tableau system for natural logic. In **Logic, Language and Meaning**, pages 104–113. Springer, 2010.

[8] Stergios Chatzikyriakidis and Zhaohui Luo. Natural Language Inference in Coq. **Journal of Logic, Language and Information**, 23:441–480, 2014.

[9] Luo and Zhaohui. **Computation and Reasoning: A Type Theory for Computer Science**. Clarendon Press, 1994.

[10] Zhaohui Luo. Formal Semantics in Modern Type Theories with Coercive Subtyping. **Linguistics and Philosophy**, 35(6), 2012.

[11] Yves Bertot and Pierre Castéran. **Interactive Theorem Proving and Program Development Coq' Art: The Calculus of Inductive Constructions**. Springer, 2004.

[12] 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**, pages 2055–2061, 2015.

[13] Koji Mineshima, Ribeka Tanaka, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki. Building compositional semantics and higher-order inference system for a wide-coverage Japanese CCG parser. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pages 2236–2242, 2016.

[14] Hiroshi Noji and Yusuke Miyao. Jigg: A Framework for an Easy Natural Language Processing Pipeline. In **the 54th Annual Meeting of the Association for Computational Linguistics**, pages 103–108, 2016.

[15] Marco Marelli, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, and Roberto Zamparelli. A SICK cure for the evaluation of compositional distributional semantic models. In **Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)**, pages 216–223. European Language Resources Association (ELRA), 2014.

[16] Robin Cooper, Dick Crouch, Jan van Eijck, Chris Fox, Josef van Genabith, Jan Jaspars, Hans Kamp, David Milward, Manfred Pinkal, Massimo Poesio, and Steve Pulman. A strategy for building a framework for computational semantics(the way forward). In **FraCaS: A Framework for Computational Semantics**, 1996.

[17] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. A framework for constructing multilingual inference problem sets: Highlighting similarities and differences in semantic phenomena between english and japanese. In **1st International Workshop on the Use of Multilingual Language Resources in Knowledge Repre- sentation Systems (MLKRep2015)**, 2015.

[18] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. An inference problem set for evaluating semantic theories and semantic processing systems for japanese. In **the 17th International Workshop on Logic and Engineering of Natural Language Semantics (LENLS12)**, pages 67–73, 2015.

[19] Mark Steedman. **The Syntactic Process**. MIT Press, 2000.

[20] Daisuke Bekki. Representing Anaphora with Dependent Types. In Nicholas Asher and Sergei Soloviev, editors, **Logical Aspects of Computational Linguistics (8th international conference, LACL2014, Toulouse, France, June 2014 Proceedings), LNCS 8535**, pages 14–29. Springer Berlin Heidelberg, 2014.

[21] Daisuke Bekki and Koji Mineshima. Context-passing and underspecification in dependent type semantics. In Stergios Chatzikyriakidis and Zhaohui Luo, editors, **Modern Perspectives in Type-Theoretical Semantics**, volume 98, pages 11–41. Springer, 2017.

[22] Per Martin-Löf and Giovanni Sambin. **Intuitionistic type theory**, volume 17. Bibliopolis, 1984.

[23] Daisuke Bekki and Ai Kawazoe. Implementing Variable Vectors in a CCG Parser. In Christian Retoré and Sylvain Pogodalla, editors, **Logical Aspects of Computational Linguistics (9th international conference, LACL2016, Nancy, France, December 2016 Proceedings)**, pages 52–67. Springer, Heiderburg, 2016.

[24] 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.

[25] Hinari Daido and Daisuke Bekki. Development of an automated theorem prover for the fragment of DTS. In **the 17th International Workshop on Logic and Engineering of Natural Language Semantics (LENLS17)**, 2017.

[26] 大洞日音. DTS の部分体系を用いた定理自動証明器への等号型の導入. Master's thesis, お茶の水女子大学, 2022.

[27] 松原舞 富田朝, 戸次大介. CCG 統語解析器 lightblue と定理証明器 wani による JSeM Verbs データセットの自動推論. In **言語処理学会第 31 回年次大会 (NLP2025)**, Mar 2025.

[28] Asa Tomita, Mai Matsubara, Hinari Daido, and Daisuke Bekki. Natural language inference with CCG parser and automated theorem prover for DTS. In Timothée Bernard and Timothee Mickus, editors, **Proceedings of the Second Workshop on the Bridges and Gaps between Formal and Computational Linguistics (BriGap-2)**, Sep 2025.