

LFG 解析を利用した日本語含意関係判定における機能表現の取り扱い

外池 昌嗣[†] 梅基 宏[†] 大熊 智子[†] 増市 博[†][†] 富士ゼロックス株式会社 研究技術開発本部

1. はじめに

テキスト間の論理的な含意関係を求める研究が、Recognizing Textual Entailment (RTE) Challenge [☆] などを通じて近年さかんに行われている。RTE は、解析対象の言語を英語とした評価型のワークショップであり、前提となるテキスト T と仮説である文 H との組に対して、T から H が推論できるかどうかを決めるタスクが設定されている。ここで行われるような含意関係の判定は、質問応答、情報抽出、要約、機械翻訳などのさまざまな応用に役立つと同時に、意味解析を評価する指標になると考えられている^{1),2)}。

含意関係を判定するシステムのうち、高精度な推論を実現するものとして、Lexical Functional Grammar (LFG) 解析を用いてテキストの論理的含意と矛盾を検出する推論システムが提案されている³⁾。このシステムは、カバー率は低いものの、QA 型のタスクで平均精度が 90-95% を示したことが第 3 回 RTE で報告されている³⁾。提案システムは、対象のテキストに LFG 解析を施し、抽出した f(unctional)-structure に基づいて意味解析を行い、テキストを論理形式に変換する。

ところで、意味解析の入力である f-structure は、言語によらず普遍的な性質と考えられる文法役割や時制などの素性から構成される。したがって、f-structure が言語間で十分に類似していれば、f-structure を利用する応用が言語を横断して利用できると考えられる⁴⁾。すなわち、英語と日本語の間で f-structure が十分に類似していれば、英語の意味解析およびテキスト推論は日本語 LFG の解析結果に対しても利用できると考えられる。

著者らは、多言語で類似した解析を実現する、実用的な LFG を構築することを目的とした Parallel Grammar Project (ParGram) に参画し⁴⁾、処理速度や解析精度、カバー率などの点で実用化の段階にある日本語 LFG を構築した⁵⁾。近年、ParGram に並行して始まった Parallel Semantics Project (ParSem) ^{☆☆} にも加わり、LFG 解析結果の f-structure を入力とした日本語意味解析の構築を行っている⁶⁾⁷⁾。

著者らのこれまでの研究⁷⁾における課題の一つに、機能表現 (複合辞とも呼ばれる) の意味を考慮した含意判定が実現できていないということがあった。機能表現とは、「をめぐり」や「によって」のように、複数の語が連なって、それぞれの語が本来持つ意味を失って、全体として、

表 1 日本語解析の流れ

解析モジュール	出力
文分割	文の文字列
文正規化	正規化された文の文字列
形態素解析	品詞が付与された単語列
機能表現検出	品詞が付与された単語列 (機能表現は 1 形態素にまとめられている)
LFG 前処理	LFG 語彙情報が付与された単語列
LFG 解析	c-structure および f-structure
XFR 意味解析	意味表現 (文の意味的な論理式)

助詞や助動詞のようにふるまうものを言う。例えば、「契約に関する紛争」と「契約をめぐる紛争」という二つの文は、内容としては同じことを言っているにもかかわらず、「に関する」や「をめぐる」といった機能表現の意味を考慮していなかったため、これら 2 つの文の間の含意判定が実現できていなかった。そこで本研究では、先行研究のシステム⁷⁾に、文献 8)9) で提案されている機能表現検出器を取り入れ、さらに、機能表現に意味役割を与えることにより、機能表現の意味を考慮した含意関係判定を実現する。

2. システムの概略

2.1 日本語解析の流れ

日本語文の意味表現を求める解析処理は、文分割、文正規化、形態素解析、機能表現検出、LFG 前処理、LFG 解析⁵⁾、XFR 意味解析¹⁰⁾の順に進める (表 1)。

形態素解析には茶筌 (ChaSen) と IPA 品詞体系辞書を用い、曖昧性を含まない単一の解を求める。機能表現検出に関しては、3 章で説明する。LFG 解析および XFR 意味解析は XLE (Xerox Linguistic Environment) ^{☆☆} 上で動作し、解析した結果生じる曖昧性はチャートの形で、異なる選択空間 (choice space) の中に閉じ込められて (packed) 表現され、複数の曖昧性をもつ解は展開されず閉じ込められた形のまま効率的に処理される。意味解析は XLE の XFR システム上に実装し、f-structure を入力として一連の書き換え規則を順次適用し、同時に概念辞書などを参照することで意味表現を出力する。

2.2 含意関係の判定方法

含意関係は、質問対象のテキスト (Passage; P) と質問文 (Query; Q) とからそれぞれ意味表現を求め、P の意味表現から Q の意味表現が論理的に含意される場合 YES、そうでない場合は NO と判定する。具体的には、P と Q の意味表現のうち、意味役割を示す項に着目して判定を行う。P と Q の両者に共通する意味役割があれば、

[☆] <http://pascallin.ecs.soton.ac.uk/Challenges/RTE/>

^{☆☆} <http://www2.parc.com/isl/groups/nlft/pargram/>

^{☆☆☆} http://www2.parc.com/isl/groups/nlft/xle/doc/xle_toc.html

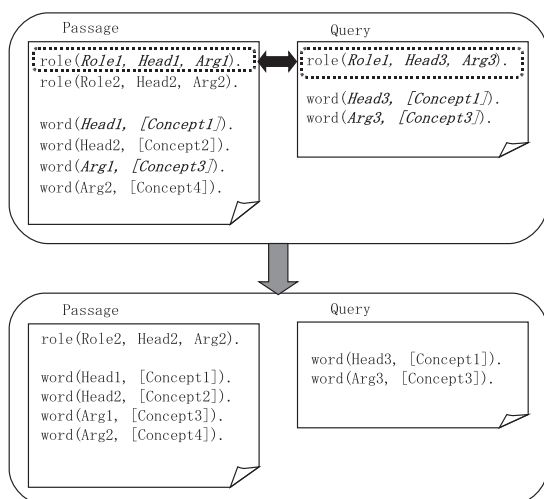


図 1 含意関係の判定

その項に含まれる語彙同士を照合する。語彙が照合できた場合は、その項を取り除き、いずれかの選択空間の中で Q の項がすべて取り除かれて空になったとき、P が Q を含意するとみなす。

図 1 に、含意関係を判定する様子を模式的に示す。図 1 の上下にある角の丸い四角は判定する前と後を、角の丸い四角の内側にある左側と右側は P と Q の意味表現を表す。意味表現の `role(Role1, Head1, Arg1)` は、主辞 `Head1` と引数 `Arg1` が意味役割 `Role1` の関係にあることを表す。また、`word(Head1, [Concept1])` は、主辞 `Head1` が概念 `Concept1` をもつ語であることを表す。まず、P および Q の意味表現の一行目の `role` 項が同じ意味役割 `Role1` をもつことに着目する。つぎに、語彙の照合を行うと、P の主辞 `Head1` と Q の主辞 `Head3` は、単語の概念を表す `word` 項から、同じ概念 `Concept1` をもつことが分かる。そして引数に関しても、P と Q の引数 `Arg1, Arg3` で同じ概念 `Concept3` をもつ。したがって、P と Q の第 1 項が取り除かれる。結果として、図 1 の下側の角の丸い四角が示すように Q には意味役割を含む `role` 項が残っていないため、P が Q を含意するとみなす。

3. 機能表現検出

1 章で、「をめぐる」は機能表現であると述べた。しかしながら、「をめぐる」という文字列が、必ず機能表現として用いられるとは限らない。例えば、「話題のケーキ屋をめぐるイベント」という文においては、動詞「めぐる」は、機能表現の一部としてではなく、動詞本来の意味で用いられている。したがって、機能表現と同じ表層を持つ文字列が文中に現れたとき、それが、機能表現として振る舞っているか、動詞本来の意味で振る舞っているかを判別する必要がある。本稿では、前者の振る舞いを「機能的」、後者の振る舞いを「内容的」と呼ぶことにする。また、機能表現と同じ表層を持つ文字列を「候補文字列」と呼ぶことにする。そして、機能表現の候補文字列が機

能的に振る舞っていると判定されたならば、形態素解析結果において、機能表現を構成する形態素を 1 つの形態素にまとめてやり、機能表現の持つ文法役割にふさわしい品詞を与えてやる必要がある。

土屋らは、MUST と呼ばれる機能表現データベースを構築している¹¹⁾。また、MUST に含まれる機能表現のうち、収録されている用例の正例と負例の割合がおおよそ半々である、59 個の機能表現に対して、YamCha^{*}を利用した機械学習により、機能表現の検出器を学習する方法を提案している⁸⁾⁹⁾。本研究では、この機能表現検出器を利用するものとした。

本研究では、さらに、MUST に収録されている機能表現の中から、内省により、機能的にのみ用いられる機能表現を 96 個抽出し、これらも機能表現として検出できるようにした。例えば、「についての」という文字列は、「ストラップについてのお願い」という文の様に、対象や関連を表す連体化助詞として働く例は見つかるが、動詞「つく」本来の意味で用いられる用例は見つけることができなかったため、機能的にのみ用いられると判定した。

なお、機能表現には、文献⁹⁾を参考にし、助動詞型機能表現には助動詞を割り当て、助詞型機能表現には、格助詞、連体化助詞、接続助詞などの品詞を割り当てた。これら以外の品詞に関しては、4 章で述べる。

4. LFG 文法

機能表現を LFG 文法の枠組みに取り入れるためには、3 章で述べた以外の品詞を定義する必要がある語が存在した。特別な品詞を付与して他の機能表現と区別を行った語の一覧を、表 2 に示す。本節では、これらの機能表現について述べる。

「助詞-並立」は「のみならず」に割り当てられる品詞である。MUST において、「A のみならず B」という形式をとる用法が紹介されているが、「B」の部分が省略された場合を考慮して、接続詞と格助詞両方の振る舞いをする語として定義した。

「助詞-ト類」は「にしても」、「として」などに割り当てられる品詞である。接続助詞と格助詞に相当する振る舞いをする語として分類した。「にしても」は MUST において、複文を形成する接続助詞としての用法が紹介されている。しかし、コーパスには文ではなく名詞に接続して連用修飾成分を形成する事例も存在したため、この品詞を割り当てた。

「助詞-連体化 1」は「上での」と「うえの」に割り当てられる品詞である。MUST では「うえで」の派生形として分類されており、動詞のシタ形、動作的名詞+「の」につくとしてされている。しかし、「上」が名詞であるため、文法的にはこのような限定なしに、一般の名詞が受ける連体修飾はすべて受けることが可能であると考えられる。

^{*} <http://chasen.org/~taku/software/yamcha/>

表 2 特別な品詞を付与した機能表現一覧

品詞分類	文法的な働き	語数	例
助詞-並立	名詞レベルの接続詞相当／格助詞相当	1	のみならず
助詞-ト類	接続助詞相当／格助詞相当	19	にしても、として
助詞-連体化 1	連体修飾を受けることで、連体化助詞として働く 接続助詞相当	2	うえでの、上の
助詞-連体化 2	体言、または埋め込み句に接続して連体化助詞 として働く	3	という、といった、 に応じた
助詞-連用化	連体修飾を受けて連用修飾を行う 連体化埋め込み句に接続して従属節を形成する	14	くせに、あとで

そこで、LFG 規則では連体修飾を受けて連体化を行う特殊な助詞を定義した。

「助詞-連体化 2」は「という」と「といった」に割り当てられる品詞である。格助詞「と」を含んでいるため、体言だけでなく埋め込み句にも接続する連体化助詞として定義した。

「助詞-連用化」は「くせに」、「あとで」などに割り当てられる品詞である。MUST では「あとで」の用法として時間の順番に沿って述べ、順に生起した関係を表す「A あとで B」という文型を紹介している。しかし、構文的には B を修飾する要素が A の前に現れることも想定できることから、接続助詞ではなく、連体修飾を受ける連用修飾成分として定義した。したがって、埋め込み句に接続する場合は、従属節を形成する。

5. 意味解析

5.1 XFR 意味解析における助詞・助動詞の扱い

2 章で述べた XFR 意味解析においては、助詞の意味役割は、以下の 2 つの方法で定義されている。一つは、IPAL 動詞辞書で定義されている述語素が表す動詞と名詞の格関係を、対応する深層格に変換した辞書である。例えば、動詞「正す」に対しては、目的格になる名詞には「対象」、主格になる名詞には「主体」という意味役割が割り当てられている。もう一つは、助詞に対して直接意味役割を定義したリストである。例えば、助詞「ので」に対しては、「理由」という意味役割が割り当てられている。ただし、接続助詞に関しては、LFG で節同士の並置構造として解析されるが、意味役割は付与していない。一方、助動詞に関しては、現状では、意味役割を割り当てることはしていないので、無視されることになる。したがって、以降では、今回検出対象とした機能表現のうち、助詞型機能表現 80 個を対象に論じる。

5.2 機能表現に割り当てる意味役割

本研究では、日本語機能表現辞書¹²⁾に付与されている意味情報を利用して、機能表現に意味役割を付与する仕組みを設計した。

日本語機能表現辞書においては、341 個の見出し語形を最上位の階層として、16,801 個の機能表現を階層的に分類しており、また、341 個の見出し語形に対して、89 種類の意味が付与されている。したがって、この日本語機能表現辞書を意味役割付与に利用すれば、見出し語 341

個の意味役割だけを考えればよいことになる。

今回は、この日本語機能表現辞書を参考に、今回検出対象とした助詞型機能表現 80 個に対して、意味役割を付与した。今後、16,801 個すべての機能表現が検出できるようになったときは、日本語機能表現辞書の階層構造を利用することで、容易に対応が可能である。

ここで、日本語機能表現辞書の 89 種類の意味を、どのように、2 章で述べた XFR 意味解析システムの意味役割階層に取り入れたかということであるが、まず、89 種類の意味のうち、命題的には考慮する必要の無いものを除外した。例えば、機能表現「のみならず」には、日本語機能表現辞書において「添加」という意味が定義されているが、命題的には「A のみならず B」は「A と B」ということなので、「のみならず」に意味役割を付与していない。そして、残った意味は、原則的に、そのままの形で意味役割として取り入れた。ただし、従来システムの意味役割階層で既に定義してある意味役割と重なる場合は、既に定義してある意味役割を利用するものとした。

5.3 同格表現

5.2 節では、従来システムの意味役割階層との整合を取りつつ、機能表現に意味役割を付与することで、これまでできなかった含意関係判定の実現を目指した。しかしながら、これだけでは、利用したい語と語の関係が得られない機能表現もある。その一つが、同格表現を構成する機能表現「という」である。例えば、「バランスファンドという貯蓄がお得だ。」という文は、「バランスファンドがお得だ。」という内容と、「バランスファンドは貯蓄だ。」という内容を含意しているが、単純に「という」に意味役割を付与しただけでは、これらの含意関係を正しく判定することができない。

そこで、同格表現を表す機能表現を含む文「A という B が V だ」に対しては、新たに、「A が V だ」という文と「A は B だ」という文も含意関係の判定に加えるという処理を導入した。このことにより、上の例の含意関係判定が実現できた。

6. 含意関係の判定例と考察

本章では、含意関係の判定を行った結果を、機能表現のタイプごとに示し、考察を行う。

同じ意味役割をもつ機能表現同士の含意関係判定

P1 領有権をめぐる争い

Q1 領有権に関する争い

「をめぐる」と「に関する」は共に同じ意味役割「対象」を持つので、P1 は Q1 を含意し、Q1 は P1 を含意する(つまり、等価である)と判定される。

同じ意味役割をもつ助詞と機能表現間の含意関係判定
P2 契約を調べる。

Q2 契約に関して調べる。

機能表現「に関して」には、既存の意味役割階層に既に定義されている意味役割「対象」が利用できたので、これをそのまま割り当てた。一方、動詞「調べる」の目的語は、意味役割「対象」を持つので、上記の例の場合、P2 は Q2 を含意すると判定される。

助詞-並立の機能表現「のみならず」

P3a 学生のみならず先生も驚く。

P3b 学生のみならず驚く。

Q3a 学生と先生が驚く。

Q3b 学生が驚く。

「のみならず」自体には意味役割を付与していないが、「のみならず」を機能表現として検出し、LFG 文法で並立の関係を捉えられるようにしたことで、P3a が Q3a と Q3b を含意していることを正しく判定できるようになった。また、4 章で述べたように、「A のみならず B」の「B」の部分が省略された場合、格助詞として振る舞うようにしているので、P3b が Q3b を含意していることも正しく判定できる。

埋め込み句+機能表現

P4 決意するとなると、覚悟が必要だ。

Q4 決意する上は、覚悟が必要だ。

「となると」は助詞-ト類なので、埋め込み句に接続する場合、接続助詞とみなされ、意味役割が付与されない。ところが、「上は」は助詞-連用化なので、埋め込み句に接続する場合、従属節とみなされ、意味役割「因状況」が付与される。このため、上記の P4 と Q4 で含意関係の判定をしたとき、P4 には意味役割「因状況」がないので、P4 は Q4 を含意していないと判断されしまう。したがって、接続助詞と従属節との間で意味役割のマッチを取ることを検討する必要がある。

機能表現以外の表現との含意関係判定

P5 子供のくせに生意気だ。

Q5 子供の分際で生意気だ。

P5 と Q5 は意味的には等価である。しかしながら、「くせに」は機能表現であるが、「分際で」は機能表現ではないので、両者の間の含意関係はないと判定されてしまう。この問題に対しては、文献 13) の様に、機能表現の代替表現を収集し、言い換えを行った後、含意関係の判定を行うことが考えられる。

7. まとめと今後の課題

本研究では、機能表現を考慮した日本語含意関係判定を実現する仕組みについて提案し、実際に 80 個の助詞型

機能表現に対して、機能表現を考慮した日本語含意関係判定を実現した。

今後の課題としては、6 章で述べた課題の他に、16,801 種類の機能表現の検出¹⁴⁾の技術を導入し、すべての機能表現を扱えるようにすることが挙げられる。また、機能表現を含む文を対象とした含意関係判定の定量評価のためのテストセットを作成することも今後の課題である。

参考文献

- 1) Dagan, I., Glickman, O. and Magnini, B.: The PASCAL Recognising Textual Entailment Challenge, *Lecture Notes in Computer Science*, Vol. 3944, pp. 177 – 190 (2006).
- 2) Bos, J.: Let's not argue about semantics, *Proceedings of LREC* (2008).
- 3) Bobrow, D., Condoravdi, C., de Paiva, V., Karttunen, L., King, T. H., Nairn, R., L. Price and Zaenen, A.: Precision-focused Textual Inference, *Proceedings of ACL-PASCAL Workshop on Textual Entailment and Paraphrasing* (2007).
- 4) Butt, M., Dyvik, H., King, T. H., Masuichi, H. and Rohrer, C.: The Parallel Grammar Project, *Proceedings of COLING2002, Workshop on Grammar Engineering and Evaluation*, pp. 1–7 (2002).
- 5) 増市博, 大熊智子: Lexical Functional Grammar に基づく実用的な日本語解析システムの構築, 自然言語処理, Vol. 10, No. 2, pp. 79–109 (2003).
- 6) Umemoto, H.: Implementing a Japanese Semantic Parser Based on Glue Approach, *Proceedings of The 20th Pacific Asia Conference on Language, Information and Computation* (2006).
- 7) 梅基宏, 杉原大悟, 大熊智子, 増市博: LFG 解析と語彙資源を利用した日本語含意関係判定, 情報処理学会研究報告 2008–NL–188, pp. 57–64 (2008).
- 8) 土屋雅稔, 注連隆夫, 高木俊宏, 内元清貴, 松吉俊, 宇津呂武仁, 佐藤理史, 中川聖一: 機械学習を用いた日本語機能表現のチャンキング, 自然言語処理, Vol. 14, No. 1, pp. 111–138 (2007).
- 9) 注連隆夫, 土屋雅稔, 松吉俊, 宇津呂武仁, 佐藤理史: 日本語機能表現の自動検出と統計的係り受け解析への応用, 自然言語処理, Vol. 14, No. 5, pp. 167–197 (2007).
- 10) Crouch, D. and King, T. H.: Semantics via F-structure Rewriting, *Proceedings of LFG06 Conference, CSLI On-line Publications* (2006).
- 11) 土屋雅稔, 宇津呂武仁, 松吉俊, 佐藤理史, 中川聖一: 日本語複合辞用例データベースの作成と分析, 情報処理学会論文誌, Vol. 47, No. 6, pp. 1728–1741 (2006).
- 12) 松吉俊, 佐藤理史, 宇津呂武仁: 日本語機能表現辞書の編纂, 自然言語処理, Vol. 14, No. 5, pp. 123–146 (2007).
- 13) 榊田達也, 松吉俊, 佐藤理史: 日本語機能表現から多様な類義表現への言い換え, 言語処理学会第 14 回年次大会, pp. 783–786 (2008).
- 14) 長坂泰治, 宇津呂武仁, 土屋雅稔: 大規模日本語機能表現辞書の階層性を利用した機能表現検出, 言語処理学会第 14 回年次大会, pp. 837–840 (2008).