

# 複文における時間関係の自動推論

理学専攻・情報科学コース 大西舞子 (指導教員：戸次大介)

## 1 はじめに

日本語の複文における従属節の出来事と主節の出来事の時間関係について、それぞれの節のテンス・アスペクトに関連した言語学的分析が行われている (Kaufmann and Miyachi, 2011; Oshima, 2011). Oshima (2011) によるとトキ節では、主節と従属節のどちらかが状態の例 (1a) では同時の解釈で、どちらも非状態かつ主節がタ形、従属節がル形の例 (1b) では、主節の後に従属節が生起する解釈になる。また、節全体のアスペクトは、ル形/テイル形という対立を持つ文法的アスペクト以外に、動詞や副詞といった語彙的アスペクトと複合的に判断される。例えば、同じ述語「家を建てる」と文法形式「テイル」を含む従属節でも、「スデニ」と共起する例 (2a) では、家を建てた後に主節の出来事が生起する解釈になり、「マダ」と共起する例 (2b) では、家を建てている最中に主節の出来事が生起する解釈になる。

- (1) a. ヒロシが本を読んでいたとき、ユミが目を覚ました。  
b. 電車に乗る時、後ろから押されて転んでしまった。
- (2) a. すでに家を建てているけれど、転勤することになった。  
b. まだ家を建てているけれど、家具を買いに行った。

自然言語処理においては、TimeBank (Pustejovsky et al., 2003) などの時間関係がアノテートされたコーパスや、そのコーパスを機械学習や深層学習の手法で学習し、end-to-end で時間関係を予測するモデル (Sakaguchi et al., 2018) 研究が進められている。しかし、これらのモデルで、コーパスの不備に加え、論理推論で導ける時間関係、文の意味計算が考慮されていない。(1,2) のようなテンス・アスペクトを考慮した複雑な時間関係の認識精度には改善の余地があり、

本研究では、言語学と自然言語処理の融合的なアプローチとして、cgc2lambda (Martínez-Gómez et al., 2016; Mineshima et al., 2016) に (a) 構文木の変換、(b) 統語素性の追加、(c) 導出木に基づく構成的意味解析、(d) Coq (The Coq Development Team, 2019) を用いた自動証明という処理を追加した含意関係認識システムの提案と、日本語の時間関係を伴う推論を実現する手法を検討する。そして作成した複文の時間関係を伴うデータセットを使用して深層学習のモデルとの比較を行い、提案手法の性能比較を報告する。

## 2 含意関係認識システム

本システムは前提文と結論文のペアを入力として受け取り、構文解析、構文木の変換、素性の追加、意味解析を経て時間情報を含む意味表示に変換する。得られた意味表示は前提から結論に対する含意判定に使用

される。3つの含意判定を出力し、前提から結論が導ける場合は **yes**、前提から否定の結論が導ける場合は **no**、どちらでもない場合は **unknown** とする。これらの判定に必要な導出は Coq の自動定理証明機能を用いて自動で実現する。

**構文木の導出** CCG パーサとして Jigg (Noji and Miyao, 2016), 及び, depccg (Yoshikawa et al., 2017) を用いる。導出木に外の関係を扱う規則、内の関係を扱う規則が従属節内で適用されている場合、意味解析において主節と従属節の出来事間の時間関係を伴う意味表示を割り当てることができない。そこで Tsurgeon (Levy and Andrew, 2006)<sup>1</sup> を用いてこれらの規則が現れる構文木に対して書き換えを行う。

$$\begin{array}{c} \vdots \\ \frac{S \backslash NP}{NP / NP} \text{ int } NP \\ \frac{NP}{NP} \quad \frac{NP \backslash NP}{S \backslash NP} \\ \hline S \end{array} \quad \Downarrow$$
$$\begin{array}{c} \vdots \\ \frac{S \backslash NP \quad (S / (S \backslash NP)) \backslash (S \backslash NP)}{S / (S \backslash NP)} \quad \frac{(S \backslash NP) / (S \backslash NP)}{S / (S \backslash NP)} \\ \hline S \quad S \backslash NP \end{array}$$

図 1: 内の関係を扱う規則 (int) が適用された導出木の書き換えの例

**素性の追加** 導出された構文木の統語素性に対してテンス・アスペクトに関する素性を追加し、文法的アスペクトと語彙的アスペクトを複合的に判断したアスペクト決定を可能にする。Smith (1991) の状況タイプのアスペクト素性と視点タイプのアスペクト素性を採用し、素性の組み合わせで動詞のアスペクト的性質 (Aktionsarten) と、進行と完了のアスペクトを表す。まず動詞に対して Aktionsarten の素性が付与され、素性が確定していない統語範疇に伝播していくことで節全体のアスペクトを得る。例えば (2) では、最初に「家を建てる」にアスペクト素性が付与され、「テイル」で出来事の一部に焦点が当てられるアスペクトに遷移する。次に、「スデニ」が加わることで、完了のアスペクトに、「マダ」の場合は進行のアスペクトに遷移し、節全体のアスペクトが確定する。

**複文の意味表示** 複文における時間関係の意味は、従属節のテンス・アスペクトが最終的に定まる時点で割り当てられる。具体的には「トキ」「マエ」「ケレド」などの従属節ヘッドで意味表示を与える。複文の多様な従属節ヘッドに対してそれぞれ異なる意味表示が割り当

<sup>1</sup><https://nlp.stanford.edu/software/tsurgeon.shtml>

てられるが、意味表示の役割は主に、(a) 従属節の出来事の成立時を表すこと、(b) 従属節の基準軸の設定、(c) 従属節の基準軸と成立時の時間関係を表すことの3つである。(a)の成立時は出来事のどこに焦点を当てているかを表し、出来事全体とする場合、一部の場合がある。(b)では設定する基準軸によって絶対テンスと相対テンスを表す。(c)は各従属節ヘッドによって異なる。例えば(1a)では、「トキ」に対して時間関係の意味が割り当てられ、それぞれ次のような意味を持つ。(a)従属節の出来事のある一部に焦点が当てられ、そこが成立時である。(b)主節の出来事の成立時を従属節の基準軸とする相対テンスである。(c)従属節の成立時と基準軸は一致する。従属節ヘッドでこれらの意味が与えられることにより、文全体の意味合成の後、従属節の出来事のある一部において主節の出来事が生起するという意味表示が得られる。

### 3 実験

データセットとして時間関係を伴う単文と複文を含む396件を作成し、提案手法の評価に使用した。正解ラベルの内訳は `yes/no/unk=225/134/37` となっている。データセットの含意関係ラベルのアノテーションにおいては、既存の含意関係認識データセット SNLI Bowman et al. (2015) のアノテーションの方法を参考に、前提文と結論文に現れる出来事を同一とみなした。

本システムにおける実験の結果、正答率は33.6%であり、とくに、含意判定が `no` の問題で20.9%正解していた。また、深層学習のモデルとの比較として、汎用言語モデル BERT Devlin et al. (2019) を、日本語の含意関係認識データセットである日本語 Textual Entailment データセット小谷他 (2008), NTCIR-9 RITE 開発用データ<sup>2</sup>の合計 6,884 件 (このうち 1 割を検証データとして使用) でファインチューニングし、評価した。実験の結果、BERT の正答率は41.9%であり、予測の内訳をみると、`no` の判定の再現率が3.2%で、残りの判定が `yes` となり、偏った予測をしていた。本研究で用いたテストセットは、学習に用いた既存のデータセットと比較して、前提文と仮説文とで時間表現に関する語以外で共通している部分が多く、単語の一致率などのヒューリスティックスでは簡単に解けない問題となっている。BERT の偏った予測結果は、BERT が学習データから論理的な推論のパターンではなく、ヒューリスティックスを学習している可能性を示唆している

### 4 結論

統語素性にテンスやアスペクトに関する素性を追加することで、複合的なアスペクトの決定が可能になったことに加え、テンス・アスペクトを考慮した複文における出来事間の時間関係についての言語学的な分析を実装することが可能になった。そして複文における時間関係を伴う意味表示を、従属節全体のテンス・アスペクトの素性が付与された統語範疇に対応させることで、時間関係をローカルに決めず、全体のアスペクトが決定した時に時間関係を定める意味解析が可能となった。自動推論の段階では時間関係の推論を伴う含

意関係認識を実現した。

### 参考文献

- Bowman, Samuel, Gabor Angeli, Christopher Potts, and Christopher Manning (2015) “A large annotated corpus for learning natural language inference,” in *EMNLP*.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019) “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *NAACL*.
- Kaufmann, Stefan and Misa Miyachi (2011) “On the Temporal Interpretation of Japanese Temporal Clause,” *Journal of East Asian Linguistics*.
- Levy, Roger and Galen Andrew (2006) “Tregex and Tsurgeon: tools for querying and manipulating tree data structures,” in *LREC*.
- Martínez-Gómez, Pascual, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki (2016) “ccg2lambda: A Compositional Semantics System,” in *ACL*.
- Mineshima, Koji, Ribeka Tanaka, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki (2016) “Building compositional semantics and higher-order inference system for a wide-coverage Japanese CCG parser,” in *EMNLP*.
- Noji, Hiroshi and Yusuke Miyao (2016) “Jigg: A Framework for an Easy Natural Language Processing Pipeline,” in *ACL System Demonstrations*.
- Oshima, D.Y (2011) “On the interpretation of toki-clauses: beyond the absolute/relative dichotomy,” *J East Asian Linguist*, pp. 1–32.
- Pustejovsky, James, Patrick Hanks, Roser Saurí, Andrew See, Rob Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo (2003) “The TimeBank corpus,” in *Corpus Linguistics*.
- Sakaguchi, Tomohiro, Daisuke Kawahara, and Sadao Kurohashi (2018) “Comprehensive Annotation of Various Types of Temporal Information on the Time Axis,” in *LREC*.
- Smith, Carlota S (1991) *The Parameter of Aspect*. Dordrecht: Kluwer-Reidel.
- The Coq Development Team (2019) *The Coq Proof Assistant: Reference Manual: Version 8.9.0*. INRIA.
- Yoshikawa, Masashi, Hiroshi Noji, and Yuji Matsumoto (2017) “A\* CCG Parsing with a Supertag and Dependency Factored Model,” in *ACL*.
- 小谷通隆・柴田知秀・中田貴之・黒橋禎夫 (2008) 「日本語 Textual Entailment のデータ構築と自動獲得した類義表現に基づく推論関係の認識」、『言語処理学会第14回年次大会発表論文集』。

<sup>2</sup><http://research.nii.ac.jp/ntcir/permission/ntcir-9/perm-ja-RITE.html>