

出来事の成立・不成立の判断をサポートするイベントオントロジー

Event ontology to support reasoning existence/non-existence of events

川添 愛^{*1}
Ai Kawazoe

宮尾 祐介^{*1}
Yusuke Miyao

松崎 拓也^{*1}
Takuya Matsuzaki

横野 光^{*1}
Hikaru Yokono

新井 紀子^{*1}
Noriko Arai

^{*1} 国立情報学研究所
National Institute of Informatics

This paper introduces an event ontology which supports reasoning about existence/non-existence of events described in natural language. This ontology describes 1) some necessary conditions for the existence of events, 2) compatibility/incompatibility between events of same type with different granularity, and 3) relations between events of different types. These are part of implicit knowledge which is used by humans when judging (non-)existence of events. We present an overview of the current state of the ontology.

1. 目的

自然言語文の真偽判断は、機械にとって最も困難なタスクの一つである。まず与えられた文を「機械にとって真偽が問える形式」にするには、照応解決、空要素の意味内容復元、曖昧性の解消などといった処理が必要である上、モダリティ等が入っていた場合はそれらの影響も考慮に入れる必要がある。仮に上記の点がすべて適切に処理されたとしても、真偽判断のための根拠となる十分なデータが存在しなければ、人間と同様、判断を下すことができない(せいぜい、文の発話者/筆者の信頼性に従って、どの程度正しそうかを推測するしかない)。では、文が機械にとって「真偽を問える形式」にあり、真偽判断のために(少なくとも人間にとって)十分なデータが提供されている場合、機械にも人間と同じような真偽判断ができるだろうか。

本論文では、「自然言語文の真偽判断」というタスクの中の一部である、「出来事(イベント)の成立を記述する文」の真偽判断に焦点を置く。すなわち、「桶狭間の戦いで、織田信長が今川義元に勝利した」「ナチスはパリを占領した」のような、「あるイベントが現実起こった」ということを述べる単純な文である。人間は、そのような文に対して、既存の知識・常識と照らして次のような真偽判断を行うことができる。

- 例 1: 高校程度の日本史の知識があれば、「西郷隆盛は織田信長と戦った」という文が偽であると判断できる。
- 例 2: 高校程度の地理の知識があれば、「ナチスが北フランスを占領した」という事実から「ナチスがパリを占領した」ことが真であると判断できる。
- 例 3: 一般常識があれば、「桶狭間の戦いで、織田信長が今川義元に勝利した」という知識から、「桶狭間の戦いで、織田信長が今川義元に敗北した」という文が偽であることが導き出せる。

上のような文の判断を機械で実現する研究としては、国立情報学研究所が推進する「ロボットは東大に入れるか」プロジェクトに関連して、大学入試センター試験の正誤問題の選択肢文を真偽判断するタスクが近年実施されている。[Kanayama 2012]では、英訳した選択肢文を対象に QA システムを利用するアプローチで、四択問題で 65%の正答率(個々の文における真/偽の二値

評価では 79.3%)を実現している。[Miyao 2012]では、日本語の選択肢文を対象とし、判断の根拠となる文を与えた上での含意関係認識により、54%の正答率(四択)を出している。NTCIR 10 においては、教科書および Wikipedia を検索して日本語選択肢文の真偽判断をする shared task (RITE2 Exam-BC subtask) で、最高の正答率(四択)が 32.41%であった[Watanabe 2013]。[狩野 2014]では、RITE2 の最高値よりも高い結果を出している。

上記の研究では、オントロジーのような構造化された知識リソースはほとんど使われていない。しかし、今後さらに精度を高め、教科書や Wikipedia などの自然言語リソースに明示的に書かれている知識を構造化することに加え、人間によるイベントの成立/不成立の判断の過程を明らかにし、そこで関わる知識・情報を形式的に記述することは有用であると思われる。

筆者らは現在、上記のような真偽判断をサポートするために、イベントのクラスに対して以下の属性を記述したオントロジーを構築中である。

- イベントが成立するための必要条件: 参加者が満たしているべき条件、回数に関する制約など
- 同じ種類のイベント間で、参加者の粒度(例えば 北フランス vs パリ)が異なる場合の同時成立関係
- 異なる種類のイベント間の関係(「勝利」と「敗北」の非両立関係など)

これらは、人間がイベントの成立/不成立を判断する際に利用していると思われる「暗黙の知識」の一部であり、先述の人間による判断の例 1~3 にも決定的に関わっている。以下ではこのオントロジーの概要と現在の状況を述べる。

2. オントロジーの概要

2.1 主な構成

当オントロジーは、イベントのクラス(2.2 以降で解説)およびその参加者のクラスを中心とする。上位オントロジーとして YAMATO[溝口 2012]を利用し、その下位にクラスを追加・拡張することでオントロジーの構築を進めている。ここで言うイベントのクラスは、YAMATO の定義する「生起物(occurrent)」以下のクラスに相当する。イベント参加者のクラスのうち主なものは、YAMATO の「弱いエージェント(weak agent)」(人物や組織など)、「人工物(artifact)」、「内容(content)」、「地理オブジェクト

(geographical object)」、「時間(time)」等の下位に拡張されている(詳しくは[Kawazoe 2013]を参照)。

2.2 複合イベントと単純イベント

当オントロジーで定義するイベントのクラスは YAMATO の生起物(occurrent)クラス以下の分類に従っている。YAMATO の「後付け完遂(extrinsic accomplishment)」および「本来的完遂(intrinsic accomplishment)」の下には、それぞれ複合イベント(complex event)と単純イベント(simple event)の分類を設けた。複合イベントの下位クラスとしては、現時点では事件(インスタンス例:「生麦事件」など)、事故(「チェルノブイリ原発事故」など)、災害(「関東大震災」など)、戦争(「百年戦争」、「太平洋戦争」など)、社会運動・暴動(「ラッドライト運動」など)、催し物(「ミュンヘンオリンピック」)などがある。複合イベントのクラスは、大雑把に言えば「固有名を持つイベント」の集合と重なる。これらのイベントのインスタンスは、それ自体複数のイベントインスタンスからなり、その規模や開始時、終了時などは多くの場合、それらの社会的、政治的、あるいは学術的な視点から決定される。例えば「百年戦争」は「クレシーの戦い」のような戦闘や、オルレアンの包囲、ジャンヌ・ダルクの処刑といった部分イベントからなる。当オントロジーでは、複合イベントの各インスタンスに対し、主要部分イベント(subevent)および開始時(starting_time)、終了時(ending_time)を記述する。また可能な場合には生起場所(location)も記述する。複合イベントでは、イベントの参加者が明確に決められない場合が多いため、それらの属性は原則として使用しない。

単純イベントは一般に、「桶狭間の戦いで、織田信長が今川義元に勝利した」のような、動詞を含む文・節によって記述されるイベントである。各単純イベントのクラスに対して、当オントロジーでは開始時・終了時・生起場所に加え、動作主(agent)や対象(theme)のようなイベント参加者およびそれに対するクラス制約も記述する。参加者のラベル名については、[竹内 2011]の動詞項構造シソーラス(<http://cl.it.okayama-u.ac.jp/rsc/data>)の意味役割ラベルを一部参考にして、「イベント参加者」と、動詞の項の「意味役割」に同一のラベルを使用する必然性は必ずしもないが、そうすることによって、イベントのクラスと、そのインスタンスを記述する言語表現とを結びつける際に有益であると考えている。

その他、単純イベントには、イベントの成立/不成立の判断の助けとなるような知識を追加している。以下に詳細を述べる。

2.3 イベントが成立するための必要条件

以下の各属性は主に、イベントが不成立か、それとも成立している可能性があるかという判断の根拠として使える場合がある。詳細は[川添 2013]を参照。

(1) 参加者の存在時間

イベントの成立に必須な、参加者どうしの存在時間におけるオーバーラップ/非オーバーラップの詳細なパターンを、temporal_relation という属性を利用して記述している。多くのイベントにおいては、動作主とそれ以外の参加者の存在時間がオーバーラップしていることが、イベント成立の必要条件となっている。例えば、先述の「織田信長は西郷隆盛が戦った」、また「ジョン・ウィルクス・ブース(注:リンカーンの暗殺者、1838-1865)が JFK を暗殺した」は、イベントの参加者間で存在時間がオーバーラップしていないために不成立と判断できる例である。ただし、すべてのイベントにおいてオーバーラップが要求されるわけではない。

(2) 参加者の位置関係

イベントの成立のために満たされているべき参加者間の位置関係(「接触可能」、「視界に入っている」、あるいは「制約なし」などのパターン)を、closeness_at_st(イベント開始時の位置関係)、closenes_at_et(イベント終了時の位置関係)という属性によって記述している。詳しくは、[川添 2013]を参照。この知識は、人間による日常的な推論において頻繁に使われる(アリバイの実証など)。ただし、当オントロジーは現時点では世界史の試験問題および教科書に頻出するイベントを優先して記述を行っているため、「制約なし」以外の記述のある項目数は少数である。

(3) 再発可能性

同じ参加者インスタンスの組み合わせを持つイベントが、一度しか起こり得ないか、あるいは別の時点・場所において再度起こり得るかを、recurrence という属性で記述する。この条件は、既知の事実と組み合わせることにより、未知のイベントの真偽判断に役立つ場合がある。例えば、「オスマン帝国はバルカン半島に誕生した」ということが偽であることは、「オスマン帝国が(バルカン半島とは地理的に重ならない地域である)アナトリアに誕生した」という知識に加え、「誕生」イベントが再発不可であるという知識から判断できる。

2.4 同種かつ粒度の異なるイベント間の同時成立関係

以下の各属性は、同じ種類のイベントで、記述の粒度が異なる複数のイベントが同時に成立するかどうかを記述するものである。詳しくは、[Kawazoe 2013]を参照。

(1) 参加者の部分-全体関係に伴う同時成立関係

「地理オブジェクト(geographical object)」クラスのインスタンスを参加者に持つイベントが現実には成立している場合、同種のイベントがその部分、あるいはそれを含むより大きな地理オブジェクトにおいても成立するかどうかを、part-whole_implication という属性で記述している。先述の例 2 のように「ナチスが北フランスを占領した」から「ナチスがパリを占領した」が導けるのは、ある対象について占領イベントが成り立つならば、その対象の一部についても成り立つという、占領イベント特有の性質による。

(2) 参加者の組織-リーダー関係に伴う同時成立関係

「組織」クラスのインスタンスを参加者に持つイベントが現実には成立している場合に同種のイベントがそのリーダーたる人物についても成立するかどうか、あるいは「人物」クラスのインスタンスを参加者に持つイベントが成立している場合にその人物がリーダーを務めている組織についても成立するかどうかを、leader-org_implication という属性を用いて記述している。

2.5 異なる種類のイベント間の関係

現在、当オントロジーにおいては、以下の 8 種類の関係を異なるイベントクラス間で定義している。一部の関係は、[兼岩 2011]において定義されているイベント間関係を元にしてしている。また、あるイベントの成立/不成立において、直接的な手掛かりにはならないが、ある程度の推測を可能にするための関係も含めている。

(1) conflicting_event(非両立イベント)

当該イベントが成立している場合には、決して同時に成り立たないイベント。「敗北」に対する「勝利」、「南下」に対する「北上」などが例として挙げられる。[兼岩 2011]で定義されているイベントクラス関係の一つである、「イベント排他関係」を参考にしている。[Kanayama 2012]では、QA アプローチでカバーできなかったケースの一つとして、先述の例 3 のようにターゲット文の

中で述べられているイベント(「参加者 1 の、参加者 2 に対する勝利」)が、実際起こったイベント(「参加者 1 の、参加者 2 に対する敗北」)と相いれないものである場合が指摘されている。その点からも、この関係が実際に成立していないイベントを「不成立」と判断する上で重要な役割を果たすと考えられる。

(2) entailed_event(同時成立イベント)

当該イベントが成立したら、必ず同時期に成立したことになるイベント。「降参」における「敗北」、「風刺」における「描写」、「遠征」における「移動」などが挙げられる。必ずしも当該イベントと開始時・終了時が一致している必要はなく、開始時以降に始まり、終了時以前に終わるイベントも含まれる。また、一つのイベントが複数のイベントを entailed_event として持つことも可能である。そのような例としては、侵攻イベントが成立する場合は、その開始時から終了時の間に侵入イベントと攻撃イベントが必ず成立することが挙げられる。あるイベントの成立/不成立を判断する場合、そのイベントの entailed_event にあたるイベントの不成立がリソース等の記述から判断できれば、当該イベントは「不成立」と判断できる。また、そのイベントを entailed_event として持つ別イベントの成立がリソース等から判断できれば、当該イベントも「成立している」と判断できる。

(3) similar_event(類似イベント)

当該イベントが成立している場合、同時期に成立している可能性が高いイベント。「反対」に対する「批判」、「強制」に対する「義務付け」、「奨励」に対する「推進」などが挙げられる。イベントの成立/不成立の判断において、判断の根拠となる事実を探す際に手がかりを広げるために設定した関係である。無論、entailed_event 関係を利用する場合に比べて確実性は落ちるが、ある程度の推測を可能にする。

(4) preceding_event(先行イベント)

当該イベントの開始前に必ず成立しており、かつ当該イベントの開始時に終了するイベント。「放棄」に対する「所有」、「敗北」および「勝利」に対する「戦い」、「在任」に対する「着任」などが挙げられる。当該イベントの不成立を断定する際に、その preceding_event の不成立はその直接的な証拠となる。

(5) presupposed_event(前提イベント)

当該イベントの開始前に必ず成立しており、かつ当該イベントが開始した後も成立するイベント。「権利の行使」における「権利の所有」、「戦い」における「対立」「敵対」が例として挙げられる。preceding_event と同様に、当該イベントの不成立を断定する際の直接的な手がかりとなる。

(6) succeeding_event(後続イベント)

当該イベントの終了と同時に必ず開始するイベント。[兼岩 2011]においてイベントインスタンス間で定義されている「次イベント関係」を、イベントクラス間の関係に拡張したものである。「発生」における「存在」、「放棄」における「喪失」などが挙げられる。entailed_event と同様に、イベントの成立/不成立の判断に利用できる。

(7) bygone_event(過去イベント)

当該イベントの開始前に必ず成立しており、かつ当該イベントが開始するよりも以前に終了しているイベント。「修繕」に対する(対象物の)「破損」、「復活」に対する「消滅」などが挙げられる。preceding_event、presupposed_event と同様に、当該イベントの不成立を断定する際の手がかりとして利用できる。

(8) future_event(未来イベント)

当該イベントの終了時より後のどこかの時点で起こることが予想されるイベント。「法律の公布」に対する「法律の施行」、「任命」に対する(対象の)「着任」などが含まれる。当該イベントの成立/不成立の直接的な手がかりにはならないが、ある程度の推測を可能にすると考えられる。

2.6 言語表現群

各イベントクラスは、linguistic_realization という属性により、そのイベントを表す言語表現(主に動詞的表現+項構造)の集合と関係づけられている。この点は、日本語 WordNet (<http://nlpwww.nict.go.jp/wn-ja/>) や動詞項構造シソーラスなどと共通している。当オントロジーの言語表現群は、自然言語で書かれたリソース中で、さまざまな表現によって記述されるイベントインスタンスを適切なイベントクラスに関係づけるためのものである。表現の種類としては動詞が多いが、「全盛期を迎える」「影響を及ぼす」などの複合的な表現も含んでいる。

3. オントロジーの現況

オントロジー内のイベントクラスの記述の例を次ページの図 1 に示す。現在は主に、高校世界史分野の試験問題や教科書に頻出するイベントを中心に、クラスの構築と知識の記述を行っている。2014 年 2 月現在では表 1 に示すように、YAMATO に新規に追加したクラスが 769 あり、うち 510 がイベントのクラスである。イベントのクラスに関連付けられた言語表現数は 1203 である。これは、過去 21 年の大学入試センター試験世界史科目の正誤問題の選択肢に現れる動詞出現数の 79.7% をカバーする。また、イベントクラス間で記述されている関係の内訳は表 2 のとおりである。

4. 関連研究

イベントオントロジーおよびそれに類する既存のリソースの多くは、イベントの詳細な内容記述および網羅的な分類を目的とする。これらの研究に対し、当オントロジーの特徴は、イベントの内容(つまり、当該イベントがどのようなイベントか)ではなく、イベントの成立/不成立に関わる外的な要因を直接的に記述している点にあると言える。

既に述べたように、当オントロジーの構築にあたっては、[兼岩 2011]のイベントオントロジー、および[竹内 2011]の動詞項構造シソーラスを一部参考にした。[兼岩 2011]のイベントオントロジーは、セマンティック Web の実現のために、Web 上のイベント関連データの意味構造を機械可読にする目的で構築されたものである。イベントに対しては、構成物による分類と意味機能による分類という二通りの分類がなされており、後者においては「状態変化」「時間上の存在変化」「空間上の存在変化」「基数変化」「比較」「同一性変化」の 6 タイプの意味機能が分類に利用されている。当該の意味機能を利用したイベント述語の内容記述もある。また、イベントインスタンス間およびイベントクラス間の関係も定義しており、中でも因果関係には詳細な下位分類がある。当オントロジーに記述されている「イベント成立のための必要条件」は、明示的にはカバーされていないが、一部はイベント機能の定義にいくつかの自明な前提を組み合わせることで導出可能である。ただし、当オントロジーでは明示的にこれらの必要条件を記述することで、意味処理の迅速化を図ることができると考えている。

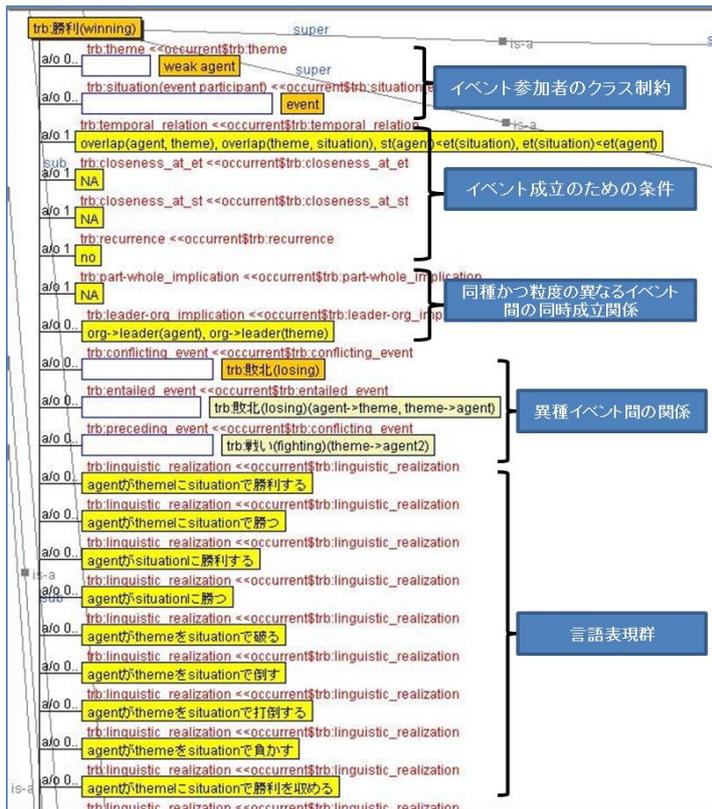


図 1. イベントクラスの記述例

[竹内 2011]の動詞項構造シソーラスは、主に文の言い換え関係を捉えることを目的として構築された動詞の辞書である。このシソーラスでは、動詞の意味内容を、LCS(語彙概念構造)によって記述することで、イベント間の共通点を明示し、類似したイベント群をまとめて網羅的に分類する。ただし、オントロジーではないため、イベント参加者のクラス制約などの情報はない(意味役割のラベルによってカバーされている部分もある(「人」「身体部分」など)。先述した通り、当オントロジーではこのシソーラスの意味役割ラベルをイベント参加者のラベル付けの参考にしてはいるが、完全に一致するものではなく、一部はクラス制約として記述している。

5. 結語

以上では、イベントの成立/不成立に対する判断をサポートするオントロジーの概要を述べた。当オントロジーでは、判断の対象となる文と知識リソースとの間の矛盾の検出や、記述の粒度や焦点の違いを解決することで、文の真偽判断に対する質問応答や含意関係認識によるアプローチを補完することを目指している。現在は人手により構築しているが、今後どのようにして自動的にイベントのクラスや属性を獲得できるかが課題である。

謝辞

本研究の遂行にあたっては、国立情報学研究所人工頭脳プロジェクト「ロボットは東大に入れるか」より助成を受けている。

参考文献

[Kanayama 2012] Kanayama, H., Miyao, Y. and Prager, J: Answering Yes/No Questions via Question Inversion, in *Proceedings of COLING 2012*, 1377-1391, 2012.

全クラス数(YAMATOクラス含む)	1329
新規クラス数	769
新規イベントクラス数	510
イベントクラスに対応する言語表現	1204

表 1. オントロジー(2014年2月現在)内のクラス数および言語表現数

イベント間の関係	項目数
entailed_event	93
similar_event	48
conflicting_event	69
preceding_event	28
presupposed_event	4
bygone_event	6
succeeding_event	43
future_event	4

表 2. オントロジー(2014年2月現在)で記述されているイベントクラス間の関係および項目数

[Kawazoe 2013] Kawazoe, A., Miyao, Y., Matsuzaki, T., Yokono, H., Arai, N: World History Ontology for Reasoning Truth/Falsehood of Sentences: Event Classification to Fill in the Gaps between Knowledge Resources and Natural Language Texts,” in *Proceedings of LENLS 10*, Submission 11, 2013.

[Miyao 2012] Miyao, Y., Shima, H., Kanayama, H. and Mitamura, T: Evaluating Textual Entailment Recognition for University Entrance Examinations, in *ACM Transactions on Asian Language Information Processing Vol 11 Issue 4*, Article No.13, 2012.

[Watanabe 2013] Watanabe, Y., Miyao, Y., Mizuno, J., Shibata, T., Kanayama, H., Lee, C-W., Lin, C-J., Shi, S., Mitamura, T., Kando, N., Shima, H. and Takeda, K: Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10, in *Proceedings of the 10th NTCIR Conference*, 385-404, 2013.

[兼岩 2011] 兼岩憲, 岩爪道昭: 「セマンティック Web のためのイベントオントロジー」, コンピュータソフトウェア, Vol.28, No.3, pp.153-166, 2011.

[狩野 2014] 狩野芳伸: 「大学入試センター試験歴史科目の自動解答」, 人工知能学会第 28 回全国大会 2|3-4, 2014.

[川添 2013] 川添愛, 宮尾祐介, 松崎拓也, 横野光, 新井紀子: 「史実としてありえない」という判断を可能にする世界史オントロジー」, 人工知能学会第 27 回全国大会 2A4-5, 2013.

[竹内 2011] 竹内孔一: 「動詞項構造シソーラスの構築」, 人工知能学会第 25 回全国大会 3H2-OS3-5, 2011.

[溝口 2012] 溝口理一郎: 『オントロジー工学の理論と実践』, オーム社, 2012.