

# 計算機による大学入試問題への解答に向けた世界史オントロジーの設計

## Designing a world history ontology as a basis for answering university entrance examination questions

川添 愛\*<sup>1</sup>  
Ai Kawazoe

宮尾 祐介\*<sup>1</sup>  
Yusuke Miyao

松崎 拓也\*<sup>1</sup>  
Takuya Matsuzaki

横野 光\*<sup>1</sup>  
Hikaru Yokono

新井 紀子\*<sup>1</sup>  
Noriko Arai

\*<sup>1</sup> 国立情報学研究所  
National Institute of Informatics

This paper discusses a usage of a world history ontology in the task of answering university entrance examination questions by computers. We propose a design of an ontology which contains some basic necessary conditions for an event to exist, as well as high-school level knowledge of world history, in order to enable automatic recognition of historical facts, possible events, and unreal ones stated in natural language sentences.

### 1. 目的

国立情報学研究所は 2011 年度より、「人工頭脳プロジェクト—ロボットは東大に入れるか」を推進している。人工知能による大学入試突破を目標とするこのプロジェクトには、タスクごとに細分化した人工知能関連技術の再融合、人間の思考に関するホリスティックな理解など、様々な効果が期待される。その中で、世界史・日本史といった歴史科目の問題は、教科書等書かれた知識を記憶しているかどうかを問う問題が多く、計算機による高い正答率が期待されることである。これらの科目においては、史実を記述している文と、そうでない文との区別を必要とする問題が大部分を占める。このような問題を解く上では、高校教科書レベルの歴史知識を何らかの形式で計算機に与える必要がある。

本研究では、歴史問題、特に世界史問題解答に役立つ知識源としてのオントロジーの設計・利用について論じる。歴史オントロジーの研究には[Ide 2007]、[Kauppinen 2007]、[石川 2008]などがあるが、これらは主に歴史研究のための知識の構造化および検索を想定したものである。ここで提案するのは歴史研究用ではなく、世界史入試問題(特にセンター入試)の選択肢に見られる「比較的簡単な自然言語文」の事実性を判断するという、深い意味処理を実現するためのオントロジーである。歴史問題の選択肢は比較的短文が短く、時制・アスペクトの種類も限られており、複雑な量化やモダリティを含まないことがほとんどである。さらに文脈から切り離された上で真偽が決定できるという特徴も持つため、選択肢の文の意味を形式的な表示で表し、オントロジー上の知識に照らして史実かどうか判断するという試みに適していると考えられる。また、入試問題では公平性と厳密性を保証するため、選択肢において述べられる「史実である出来事」と「史実ではない出来事」との間に明確な線引きがある。また通常のテキストにはあまり出現しない「明らかに起こっていないこと・起こりえないこと」を記述した文が一定量得られるという点も、入試問題を対象とする上で興味深い点である。

本論文では以下、世界史問題の選択肢文を対象とした事実性判断に、オントロジーがどのように貢献できるか、またどのようなオントロジーを設計すればよいかを考察する。以下第 2 節では、オントロジーにおける開世界・閉世界仮説の立場と世界史入試問題解答タスクとの関係について述べ、教科書において記

述される史実と、イベント成立のための必要条件(ある種の常識)の両方を含めるという、大まかなオントロジーの設計について述べる。第 3 節では具体的なオントロジーの設計について述べ、第 4 節では世界史問題解答のための他のアプローチとの関係について考察する。

### 2. 世界史問題解答タスクとオントロジー

#### 2.1 世界史問題解答タスクと開世界仮説・閉世界仮説

従来の論理プログラミングの多くが、明示的に記述されていない知識を偽とみなす閉世界仮説を採用しているのに対し、オントロジー記述言語(OWL)および記述論理においては、明示的に記述されている知識以外の命題について、それらを必ずしも偽としないとする開世界仮説が採用されている。双方の問題点については議論がなされ、両者を統合するための枠組みも提案されてきている([Motik 2006], [小出 2008])。世界史問題における選択肢文の事実性判断というタスクに関して言えば、「史実を述べていない文」の識別が必要となるという点で、開世界仮説とは相性が悪く、閉世界仮説と相性が良いように見える側面が多い。

例えば、以下の文は、センター試験問題世界史 A, B の選択肢に実際に現れた「誤りを含む文」である。

- (1) ピピンは、ランゴバルド王国を滅ぼした。(2009 年度 本試験 世界史 B)
- (2) フランソワ 1 世は、神聖ローマ皇帝オットー 1 世と対立した。(2011 年度 本試験 世界史 B)
- (3) 西アジアに伝わった仏教は、ゾロアスター教の成立に大きな影響を与えた。(1992 年度 本試験 世界史)

これらの文が史実を述べていないことは、当然のことながら、教科書その他の知識リソースには明示的に書かれていない。つまり、「ピピンはランゴバルド王国を滅ぼしていない」「フランソワ 1 世はオットー 1 世と対立していない」のような記述は教科書等には通常存在しない。むしろ、(1)-(3)のような記述が知識リソース上に存在しないことそのものが、これらを偽とする根拠とされることが多いようである。受験参考書等では、これらの文を偽とする根拠として、これらが述べていることと一部異なる史実の存在(例えば、「カール大帝はランゴバルド王国を滅ぼした」「フランソワ 1 世はカール 5 世と対立した」のような史実の存在)を挙げ

ることが多いが、このような説明も、暗に閉世界的な立場を前提としているように見受けられる(論理的には、「カール大帝はランゴバルド王国を滅ぼした」「フランソワ 1 世はカール 5 世と対立した」という史実のみでは(1)(2)を否定する根拠となりえないからである)。このことから、閉世界仮説の立場をとり、オントロジーに明示的に書かれていない出来事や、オントロジー上の知識から演繹できない命題を偽とみなす「失敗による否定」を採用するのが、世界史問題解答には有効であるかに見える。

しかしながら、実際にそのようなアプローチを採用すると以下のような問題が生じる。まず、史実と史実でない出来事に線引きをするために史実を完全に網羅的に記述することが必要となるが、これは現実的に考えて不可能である。また、出来事を記述する表現のバリエーションを適切に認識できなければ、史実を述べている文まで過剰に偽であると判断してしまう可能性がある。さらに本質的な問題として、「史実」は現在明らかになっている事実と過ぎないため、これらのみを事実としてしまうと、将来歴史研究により明らかにされる可能性のある史実までも否定することになる。すなわち、「現時点で史実ではない出来事」のうち、「史実であるかどうか現時点で不明な出来事」と、「明らかに史実ではありえない出来事」の間の線引きができない。

ここでは以上の考察を踏まえ、原則としては開世界仮説を採用し、教科書等で述べられている史実に加え、起こりうる出来事と起こりえない出来事を区別するため、次に述べる「イベント成立のための必要条件」を明示的に記述したオントロジーを提案する。

## 2.2 イベント成立のための必要条件

イベントが成立するための必要条件には様々なものがあるが、世界史問題の選択肢文が史実を述べているかどうかを判断する上で特に重要なのは、時間的条件、より詳しくは「イベントの参加者間の存在期間の関係」である。イベントには、大まかに分けて動作主と他の参加者(対象、起点、着点、道具、時間等)の存在時間が重複していなければならないものとそうでないものがあり、前者と後者それぞれにおいて、参加者間の存在開始時・存在終了時の前後関係に制約がある。例えば「(人・組織・国が)(国を)滅ぼす」というイベントは、動作主の存在期間が、滅ぼす対象の存在終了時を含んでいなければ成立しない。ここから、先に述べた(1)は史実ではありえないことが導ける。というのは、ピピン(大ピピン(580-640))の存在期間が、滅ぼす対象(ランゴバルド王国(774 に滅亡))の存在終了時を含んでいないからである(仮に、「ピピン」が中ピピン(640-714)、あるいは小ピピン(714-768)を指していると仮定しても結果は同じである)。また「(人物が)(人物と)対立する」というイベントを含め、多くのイベントにおいては、動作主とその他の参加者間の存在期間が重複していることが必要である。(2)では、フランソワ 1 世が 16 世紀頃の人物であるのに対し、オットー 1 世は 10 世紀頃の人物であり、生存期間が重なっていないために、両者が「対立した」ことはありえないと判断できる。これらに対し、「(思想・文化様式が)(思想・文化様式の成立に)影響する」というイベントでは、影響する側とされる側の存在期間が必ずしも重複している必要はない。しかし、前者が後者よりも先に存在してはならず、それゆえ(3)は史実ではありえない(仏教が紀元前 5 世紀頃成立しているのに対し、ゾロアスター教はそれよりも早い紀元前 7 世紀頃に存在している)。

世界史入試問題の選択肢文においては、上記のような時間的な条件を明示的に記述することによって「ありえない史実を述

べている」と識別できるケースは多い。その他の条件としては、場所的条件やイベント間の対立に関する条件、特定の役割あるいは国籍を持たなければ実現できない行為などにみられる「参加者の立場に関する条件」などがある。

## 3. オントロジーの概要

### 3.1 上位レベルおよび主要なクラス

ここで提案するオントロジーの上位レベルは、Smith らが開発する Basic Formal Ontology (BFO1.1) を基本とする。このオントロジーは、特定の時点(time instant)における世界の静的なスナップショットを記述する SNAP オントロジーと、特定の期間(time interval)と関連付けられる事物の動的な変化を記述する SPAN オントロジーを含む([Grenon 2003][Grenon 2004])。

主なクラスを(4)の表に示す。各クラスは、BFO1.1 の上位クラスと対応付けられる。

### 3.2 インスタンスの記述

いわゆる現実世界の状況記述は、主にインスタンスレベルで行う。前節で概観した「イベント成立に関わる条件」を記述するため、「イベント」(span:Process)および「もの」(snap:Object)の各インスタンスに対して、以下の情報を指定することにより、それらの存在時間を記述する。

- span:Process (イベント) のインスタンスの開始時と終了時
- snap:Object (もの) のインスタンスの存在開始時と存在終了時

イベント(span:Process)に含まれるインスタンスは、そのプロパティとして、イベントの参加者(動作主(agent), 対象(theme), 着点(goal)など)、場所、開始時間と終了時間などの構成要素を持つ。イベントの存在時間はその開始時から終了時までである(いわゆる achievement タイプのイベントのように、瞬間的なイベントでは開始時と終了時が同じとなる)。

人物や組織、建造物などの「もの」(snap:Object)についても存在時間を指定するが、その決定のために、「イベント」の下位クラスの一つとして、「存在イベント」のクラスを設定する。この中に含まれるイベントは、「現実世界において存在する」という出来事のインスタンスであり、構成要素として、存在物に相当する theme、場所、開始時間と終了時間を持つ。個々の「もの」のインスタンスの存在時間は、それを theme とする存在イベント(process の下位クラス)の開始時から終了時までと規定する。

その他、人物の役割や所属機関、国家の特定の時点での領土、芸術作品の作者なども、それぞれ hasRole (役割を担う) イベントや所属イベント、hasTerritory (領土を持つ) イベント、作成イベントの記述を介して定義する。

### 3.3 イベントの分類とイベント成立に関わる条件の記述

イベントの分類は、BFO の上位の分類に加え、[兼岩 2011]で提案されている意味機能、因果関係など、複数の基準を設ける予定である。その中に、イベント参加者の存在時間に基づく分類(5)も含める。

(4) 主要なクラス

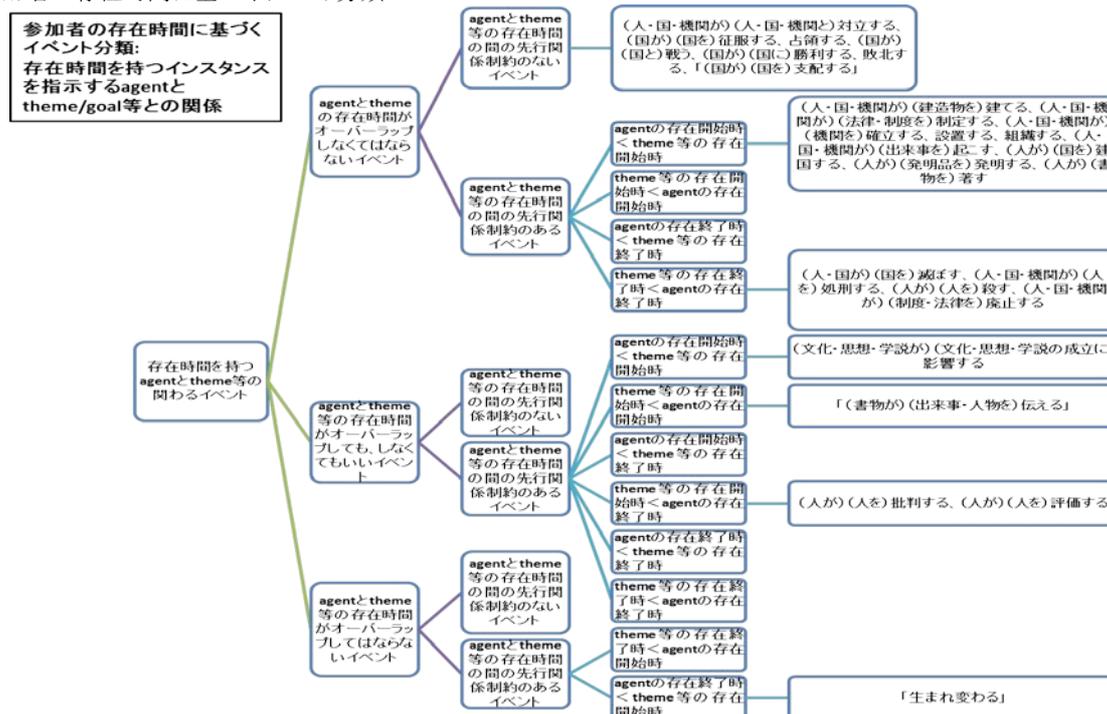
クラス	下位クラスの例、概要	インスタンスの例	主な属性	BFO 上位クラス
イベント(出来事)	存在イベント、所属イベント、hasRole(役割を持つ)イベント、作成イベント、誕生イベント、hasTerritory(領土を持つ)イベント、影響イベント、戦争、革命など	特定のものの存在、パルテノン神殿の建築、フィリップ6世の即位、ファティマ朝の滅亡、ヴェトナム戦争、フランス革命、五・プラハの春、壬申の乱、ブラザ合意、ペスト流行	開始時(starting_time)、終了時(ending_time)、動作主(agent)、対象(theme)、着点(goal)、起点(source)、道具(instrument)、場所(occurred_at)	span:Process
人物		インノケンティウス3世、徳川家康、	国籍、存在期間(存在イベントを介して定義)、役割(hasRole イベントを介して定義)、所属機関(所属イベントを介して定義)	snap:Object
人物集団	家系、民族、その他の集団	ハプスブルク家、フン族、ガリア人、ヴァイキング、インディアン、白蓮教徒、ブラジル移民、パレスチナ難民	国籍、居住地(民族の場合)、存在期間(存在イベントを介して定義)	snap:ObjectAggregate
人物属性-民族		(属性としての)フン族、ガリア人、ヴァイキング、インディアン、漢民族		snap:DependentContinuant
人物属性-役職・身分	国家の首長など	征夷大將軍、太政大臣、守護、地頭、戦国大名、ファラオ、奴隷	国籍、所属する組織(所属イベントを介して定義)	snap:DependentContinuant(Role)
人物属性-その他		(属性としての)白蓮教徒、ムスリム、華僑、ブラジル移民		snap:DependentContinuant
機関・組織	政治団体、宗教団体、企業、公共機関、国際組織など	東インド会社、自由民主党、ヨーロッパ共同体、UNESCO、自衛隊、パリ大学、カトリック教会、	国籍、目的、存在期間(存在イベントを介して定義)	snap:Object
国家・王朝	(機関・組織のサブクラスとしての)国家・王朝	(国家としての)日本、ティムール朝、ドイッ帝国、隋、アケメネス朝ペルシャ	領土(hasTerritory イベントを介して定義)、存在期間(存在イベントを介して定義)	snap:Object
場所	地勢による区分、行政による区分	(場所としての)日本、大阪市、四国、黄河流域、ヨーロッパ、華北、小アジア、バルカン半島		snap:SpatialRegion
制度・法律・条約	制度、法律、条約、協定など	社会主義者鎮圧法、日米和親条約、八色の姓、カースト制度、京都市議定書、荘園整理令	国籍、目的、存在期間(存在イベントを介して定義)	snap:Object
政策・体制		一人っ子政策、55年体制、アパルトヘイト	国籍、目的、存在期間(存在イベントを介して定義)	snap:Object
宗教		ゾロアスター教、大乘仏教、アニミズム	存在期間(存在イベントを介して定義)	snap:Object
思想・学説		三位一体説、王権神授説、イスラム原理主義、染色体説、大陸移動説、95カ条の論題	存在期間(存在イベントを介して定義)	snap:Object
技術・発明		アラビア数字、蒸気機関車(個々の機関車ではなく、作る技術として)、太陽暦、測量術、ラジオ放送、農業技術	存在期間(存在イベントを介して定義)	snap:Object
建造物		タージ・マハル、スエズ運河、アスワン・ハイダム、トリアノン宮殿	存在期間(存在イベントを介して定義)、場所	snap:Object
芸術作品	書物、絵画、彫刻	平家物語、新約聖書、東方見聞録、見返り美人図、金剛力士像、康熙字典	存在期間(存在イベントを介して定義)、作者(作成イベントを介して定義)	snap:Object
美術様式		ドーリア式、ゴシック式、印象派、折衷様	存在期間(存在イベントを介して定義)、場所	snap:DependentContinuant
文化・文明区分		天平文化、インダス文明、イスラム文化、東山文化、貝塚文化	存在期間(存在イベントを介して定義)、場所	snap:Object
時間	世紀、年、月、日、時代区分	1989年、江戸時代、平成元年、中世、		span:TemporalRegion

第2節で述べた「イベントの成立に関わる必要条件」は、(5)のような分類において、イベントのクラスの公理として明示的に記述する。例えば「作成イベント」全般(建造物の建設、制度の制定、作品の作成、機関の設立など含む)については、次のような公理を記述する。(意味:ある作成イベント e が、ある object x を agent とし、ある object y を theme とするとき、y の存在開始時は x の存在開始時より後、存在終了時以前でなくてはならない。)

$$\forall x \forall y [\exists e [\text{作成イベント}(e) \wedge \text{agent}(e,x) \wedge \text{theme}(e,y)] \rightarrow \text{starting\_time}(x) < \text{starting\_time}(y) \wedge \text{starting\_time}(y) = < \text{ending\_time}(x)]$$

イベント成立のための他の条件も、クラス公理として記述していく予定である。例えば「立場による条件」に関しては、agent の役割を参照することで、「特定の役割あるいは国籍を持たなければ実現できない行為」(国家の首長による政策の決定など)を記述する。

(5) 参加者の存在時間に基づくイベント分類



4. 他のアプローチとの関係

計算機による世界史問題解答については、すでに自然言語処理技術の一つである含意関係認識を用いたアプローチが開始されている([宮尾 2012])。これは、Wikipedia や教科書など、自然言語で書かれた知識リソースから得られた文を前提とし、試験問題中の選択肢文がそこから含意されるかどうかを判断するものである。このアプローチの強みは、選択肢文と語彙および構造の面で類似している文が参照先のリソースにある場合の判断である。他方、参照先のリソースに書かれていないような文が史実であるかどうかの判断は困難であると予想される。先に述べたように、教科書等のリソースに書かれているのは史実であり、「実際には起こっていないこと」についての記述はほぼ無いと考えられることから、史実ではない命題を記述している文の事実性を判断するのは難しいと考えられる。

ここで提案するオントロジーにより、史実と、史実かどうかが現時点で不明な出来事、また史実ではありえない出来事の三つに明確な線引きができれば、含意関係認識によるアプローチを補完する枠組みができると考えられる。

5. 結語

以上では、計算機による入試問題解答、特に世界史について述べた文の事実性判断のためのオントロジーの概要を述べた。今後は本稿で述べた設計の方針に基づき、固有表現アノテーション済み教科書データ等のリソースを利用してオントロジーを構築する。同時に選択肢文を簡単な論理的意味表示に変換する手法を確立し、オントロジー上の知識を用いた推論によって自動的に事実性が判定できるか検証する予定である。

参考文献

[Grenon 2004] Grenon, P., and Smith, B. "SNAP and SPAN: Towards Dynamic Spatial Ontology", *Spatial Cognition and Computation*, 4:1, 69-103, 2004.

[Grenon 2003] Grenon, P. "Spatio-temporality in basic formal ontology: SNAP and SPAN, upperlevel ontology, and framework of formalization (part I)," *Technical Report Series*, 05/2003, IFOMIS, 2003.

[Ide 2007] Ide, N. and Woolner, D., "Historical Ontologies." In Ahmad, K, Brewster, C., and Stevenson, M. (eds.), *Words and Intelligence II: Essays in Honor of Yorick Wilks*, Springer, 137-152, 2007.

[Kauppinen 2007] Kauppinen, T., Hyvonen, E.: *Modeling and Reasoning about Changes in Ontology Time Series*, pp. 319-338. In: Kishore et al. (eds.): *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*. Springer-Verlag, 2007/

[Motik 2006] Motik, B., Horrocks, I., Rosati, R., and Sattler, U. "Can OWL and Logic Programming Live Together Happily Ever After?," in *ISWC2006*, Vol. LNCS 4273, pp. 501-514, Springer-Verlag, 2006.

[石川 2008] 石川徹也, 北内啓, 城塚音也. 歴史オントロジー構築のための史料からの人物情報抽出. *自然言語処理* Vol. 15, No. 4, pp. 3-18, 2008.

[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.

[小出 2008] 小出誠二, 武田英明, 「OWL における明示的閉世界と局所閉世界仮説」, *人工知能学会全国大会(第 22 回)論文集*, IF1-04, 旭川, 2008.

[宮尾 2012] 宮尾祐介, 嶋英樹, 金山博, 三田村照子, 「大学入試センター試験を題材とした含意関係認識技術の評価」, *言語処理学会第 18 回年次大会講演論文集(2012)*, C2-9, 2012.