

# 交通オントロジーに基づく質問応答データセットの構築

## Constructing a Traffic Ontology-based QA Dataset

高山 隼矢                      三輪 誠                      佐々木 裕  
 Junya Takayama              Makoto Miwa              Yutaka Sasaki  
 豊田工業大学  
 Toyota Technological Institute

{sd13055, makoto-miwa, yutaka.sasaki}@toyota-ti.ac.jp

### 1 はじめに

自動運転車の法規走行に向けて、交通法規などの知識をオントロジーを用いて知識ベース化する研究が進められている [1]. また、自然言語で書かれた質問文を、オントロジーから知識を抽出するために用いられる問い合わせ言語へ自動変換する手法が研究されている [2, 3]. しかし、既存の変換手法では交通法規のように複文や条件文を持つ複雑なものには対応できておらず、これに対応できるような変換システムを構築する必要がある。

本研究の目的は、複雑な質問文にも対応した変換システムを構築するために、自然言語文と問い合わせ言語文の論理的な関係を明らかにすることである。運転免許試験の正誤問題を題材として、解答に必要な知識を含むオントロジーと、問題文とそれに対応する問い合わせ言語文、それらの中間表現となる論理式のペアを含むデータセットを作成し、問題文の分類を行う。

### 2 関連研究

#### 2.1 ウェブ・オントロジー言語

オントロジーとは、事象や具体物をクラスという単位で分類し、各クラスに対し属性や上位・下位などの関係を定義するための形式的な表現である。図 1 は道路に関するオントロジーの例で、道路の下位クラスとして高速道路と一般道があることが示されている。ウェブ・オントロジー言語 (Web Ontology Language; OWL) とは、ウェブ上の概念を記述するための統一的な枠組みである Resource Description Framework (RDF) を拡張

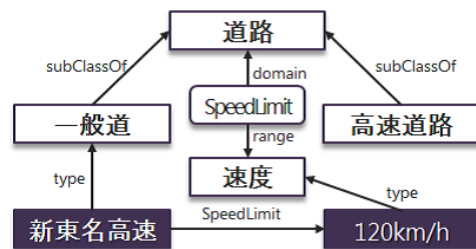


図 1 オントロジーの体系図

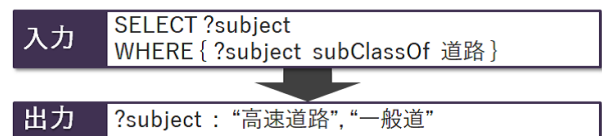


図 2 SPARQL クエリと出力例

した、オントロジーを表現するための言語である。  
 2.2 問い合わせ言語 SPARQL

SPARQL Protocol RDF Query Language (SPARQL) とは、OWL などの RDF データに対する問い合わせ言語の一つである。SPARQL のクエリは、RDF トリプルと呼ばれる「主語・述語・目的語」のセットを用いて記述した WHERE 句の条件に合致するデータに対し、行いたい操作を文として記述する。文には、合致するデータがあるかどうかを返す ASK 文、合致するデータのうち指定した変数の値を取り出す SELECT 文、合致するデータを用いて新たなデータを作成する CONSTRUCT 文、合致したデータの関連情報を提示する DESCRIBE 文の 4 種類がある。図 2 に示す例は「目的語が“道路”で、述語が“subClassOf”となる主語“?subject”を選ぶ」というクエリで、図 1 のオントロジーにこれを問い合わせると“高速道路”と“一般道”が得られる。

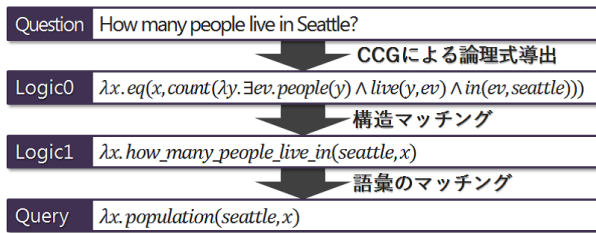


図3 Kwiatkowski ら [2] の手法概要

### 2.3 質問文の問い合わせ言語への変換

Kwiatkowski ら [2] は、英語の質問文を論理式を介してクエリへと変換する手法を提案している。図3に示すように、この手法では CCG 解析を用いて質問文を述語と主語・目的語の関係を示す述語論理式へ変換し、その論理式の構造をオントロジーの文法に合うように変形したのち、論理式の語彙とオントロジー上の語彙のマッチングを行うことで、クエリを生成している。SPARQL 変換においても主語・述語・目的語の情報が必要であり、述語論理式を介する手法は有用だが、日本語の疑問文表現の多様性や英語との文法の違いから、この手法をそのまま日本語に適用することは難しい。日本語に適用する手法も提案された [3] が、これは上位下位関係を問う問題のみが対象であり、運転免許試験問題のように多様な質問文から構成されるもののすべてには適用できていない。

### 2.4 自動車運転免許試験 RTE コーパスの構築

運転免許試験問題の QA データセットの先行研究に、平ら [4] によって構築された、自動車運転免許試験 RTE コーパスがある。テキスト含意認識技術 (Recognizing Textual Entailment; RTE) の質問応答への適用を目的としており、コーパスには問題文間の含意関係のアノテーションがなされている。タスク設定としては、各問題に対し、過去の類題や交通教則文などのテキストとの含意関係から解答を導出するものとなっている。

## 3 データセットの構築

### 3.1 構築手順

運転免許試験の正誤問題文と解答のペア、論理式と SPARQL クエリを含む質問応答データセットを作成し、SPARQL クエリや論理式の構造に基づいた問題分類を行う。データセットの作成手順を図4を用いて説明する。まず、(1) 交通教則文の知識を基に、オントロジーを作成する。さらに、(2) 運転免許試験の正誤問題例の問題文・解答 (○

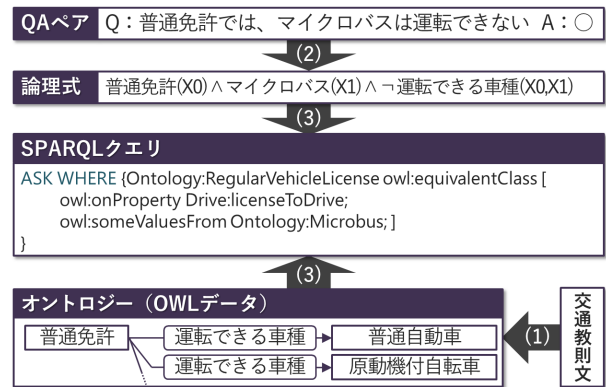


図4 本研究におけるデータセットの構築手順

か ×) ペアから、問題文とクエリの中間表現となるような論理式と、(3) 問題文の内容と合致する知識をオントロジーから問い合わせるための SPARQL クエリを導出する。これらを問題文ごとにまとめたものをデータセットとする。

### 3.2 構築環境

本実験における知識ベースとなる交通オントロジーは、国家公安委員会「交通の方法に関する教則」\*1の教則文を基に構築した。なお、オントロジーの編集に用いたソフトウェアは Protégé \*2である。構築したオントロジーのクラスとオブジェクトプロパティの例を図5に示す。語彙数は、主語や目的語として用いるクラスが 276、述語として用いるオブジェクトプロパティが 87 である。知識はこれらの語彙を用いて、クラスに関する定義を示すクラス公理として記述した。

また、データセットの作成に必要な運転免許試験正誤問題の問題文・解答ペアは、文献 [5] にて作成されたものを用いた。このデータは Web 上から収集した問題例から作成されたものであり、画像を用いた問題は除外されている。

## 4 分類結果

調査した 170 題のうち、データ化できたのは 91 題であった。本節ではデータ化できたもののできなかったもののそれぞれについて、その分類結果や考察を述べる。

### 4.1 データ化できた問題

運転免許試験問題のうち、交通教則などに存在する知識をそのまま問うような問題について、

\*1 <https://www.npa.go.jp/koutsuu/kikaku/kyousoku/>  
最終アクセス日時 2017/01/13 14:56

\*2 <http://protege.stanford.edu/>

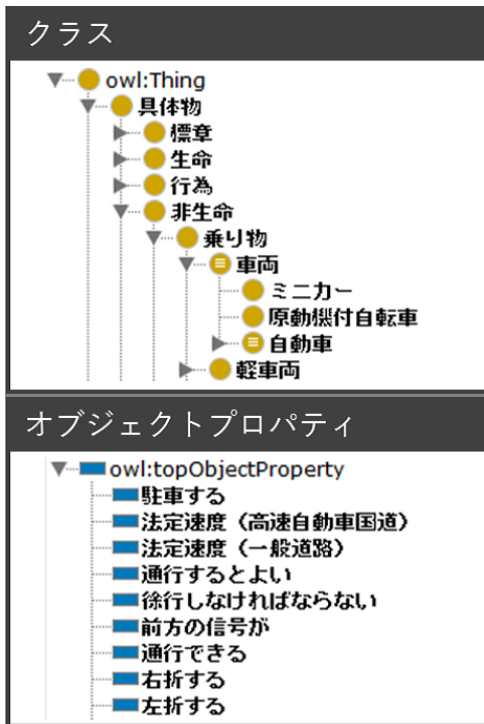


図5 交通オントロジーの語彙の一部 (Protégéの編集画面)

データ化できた。データの一例を図6に示す。図中“SPARQL :”以下にあるように、問題文に対応する SPARQL クエリはすべて ASK 文で作成した。ASK 文は指定した条件にマッチするデータがあれば True, なければ False を出力するので、正誤問題を表すのに適している。作成したデータは SPARQL クエリや論理式の構造の形をもとに分類した。表1にその分類結果を示す。以下にそれぞれの分類の概要と、その分類に属する問題例を記す。なお、問題例の中では下記のような形式の述語論理式を使用している。

述語 (主語, 目的語)

ここで用いる主語や目的語は交通オントロジー上のクラスに該当し、述語はオブジェクトプロパティに該当するが、述語については例外的に上位下位を示す語彙である *subClassOf* (下位クラス, 上位クラス) を用いることがある。また、主語や目的語が問題文において言及されていない場合、変数を用いる。例えば「前方の信号が黄の点滅のときは、徐行しなければならない。」では、行動の主体に言及されていない他、「徐行する」の目的語も抜けているため、これらを変数として扱う。

上位下位型 ものごとの上位・下位関係を問う問

```

- ID : 16
SENTENCE : 排気量50ccのミニカーは、原付免許で運転することができる
ANSWER : x
LOGIC : ミニカー (X0) ^ 原付免許 (X1) ^ 運転できる車種 (X1, X0)
CONDITION : True
SPARQL : ASK WHERE {
  Ontology:mopedLicense owl:equivalentClass [
    owl:onProperty Drive:licenseToDrive;
    owl:someValuesFrom Ontology:Minicar;
  ]
}

```

図6 データセットの例

題。頻出表現として「…には…がある」「…は…である」などが挙げられる。

例:「停止表示器材には停止表示板と停止表示灯がある」

$$P \wedge Q$$

$P$ : *subClassOf*(停止表示板, 停止表示器材)

$Q$ : *subClassOf*(停止表示灯, 停止表示器材)

述語型 上位下位で記述できない問題。

例:「排気量 50cc のミニカーは、原付免許で運転することができる」

$$P$$

$P$ : 運転できる車種 (原付免許, ミニカー)

すべての問題は以上の2つのどちらかあるいは両方に分類できたため、これを基本分類とした。また、特殊なクエリを用いる必要がある問題を、特殊型として以下の4つに分類した。

論理和型 問題文が2つ以上の命題に分けられ、そのいずれかが真であれば解答が決定するタイプの問題。例「徐行や停止をするときは、合図をしなくてもよい」

$$\neg P \vee \neg Q$$

$P$ :  $\lambda x$ .合図をしなければならない ( $x$ , 徐行)

$Q$ :  $\lambda x$ .合図をしなければならない ( $x$ , 停止)

論理包含型 知識が論理包含「 $P \rightarrow Q \equiv \neg P \vee Q$ 」の形で定義されているタイプの問題。

例「前方の信号が黄の点滅のときは、徐行しなければならない。」

$$\neg P \vee Q$$

$P$ :  $\lambda x$ .前方の信号 ( $x$ , 黄色の灯火の点滅)

表1 データの分類結果		
基本分類	上位下位型	23
	述語型	70
特殊型	論理和型	4
	論理包含型	23
	マッチングエラー型	9
	逆説型	5
	総データ数	91

$Q: \lambda x. \lambda y. \text{徐行しなければならない}(x, y)$

マッチングエラー型 オントロジー上に該当する知識がないために SPARQL クエリを生成できない問題。オントロジーに交通法規などに関するすべての知識が存在するという仮定のもとであれば、このような問題に対する解答は×であるとしてよい。

例:「道路標示には、規制標示と案内標示の2種類がある」

$P \wedge Q$

$P: \text{subClassOf}(\text{規制標示}, \text{道路標示})$

$Q: \text{subClassOf}(\text{案内標示}, \text{道路標示})$

「案内標示」に該当するクラスがオントロジー上に存在しない。

逆説型 問題文が2つ以上の命題に分けられ、それぞれの命題の述語が共通で、かつどれかが否定で表される問題。SPARQL の仕様上このような問題を一つのクエリで表すことはできないため、複数のクエリを用いる必要がある。

例:「普通二輪免許では、原動機付自転車は運転できるが小型特殊自動車を運転することはできない。」

$P \wedge \neg Q$

$P: \text{運転できる車種}(\text{普通二輪免許}, \text{原動機付自転車})$

$Q: \text{運転できる車種}(\text{普通二輪免許}, \text{小型特殊自動車})$

#### 4.2 データ化できなかった問題

運転免許試験問題には、知識をそのまま問う問題のほか、推論や計算が必要な問題があった。例えば、「自宅前の歩道に限り、植木や盆栽を置いてもよい」という問題では、オントロジー上には「歩道に障害物を置いてはならない」といった知識があるのに対し、

$\{\text{植木}, \text{盆栽}, \dots\} \in \text{障害物}$

であることを推論する必要がある。ここで、植木や盆栽はどのような状況においても障害物となるわけではなく、また植木と盆栽だけが障害物になり得るわけでもないため、オントロジー上で「障害物」の下位クラスに「盆栽」や「植木」を設けるというアプローチは適切ではないと考えられる。

また、「乗車定員5人の普通乗用車に、運転者の

ほか、大人が2人と13才の子供1人と11才の子供3人を乗せると定員オーバーになり違反である」という問題では、「12歳未満の子供は3人を2人として計算する」という知識を基に、「11才の子供3人」は2人として計算すること、「13才の子供1人」は大人と同じ扱いであることを導いた上で、すべてを足し合わせた数と乗車定員と比較するといった計算が必要になる。

このような問題は解答のプロセスを SPARQL による知識の問い合わせのみでは表現できなかったため、データ化することができなかった。

## 5 おわりに

交通オントロジーを用いた質問応答システムの構築のために、オントロジーと91件の質問文・SPARQL クエリを含むデータセットを作成し、SPARQL クエリや論理式の構造に基づいてデータの分類を行った。知識をそのまま問うタイプの問題についてはデータ化することができたが、知識を利用した推論や計算などといった処理をする必要があるタイプの問題はデータ化できなかった。今後は本提案手法で対応できなかった問題や、研究対象外とした画像とテキストからなる問題についても、解答プロセスを表現できるように手法を改良する必要がある。

## 参考文献

- [1] 河辺一仁ら. 交通オントロジーの半自動拡張のための交通用語認識. *In Proc of. NLP2015*, pages 135–138, 2015.
- [2] Tom Kwiatkowski et al. Scaling Semantic Parsers with On-the-fly Ontology Matching. *In Proc of. EMNLP*, (October):1545–1556, 2013.
- [3] 鈴木遼司ら. 交通オントロジーを対象とした質問文の SPARQL クエリ変換. *In Proc of. NLP2015*, pages 171–174, 2015.
- [4] 平博順ら. 自動車運転免許試験 rte コーパスの構築. *In proc of. 人工知能学会全国大会論文集 Vol.28, 3I4-5*, 2014.
- [5] 杉村皓太ら. 交通規則問題のための解答システムの構築. *In Proc of. NLP2013*, pages 790–793, 2013.