

運転免許試験で使用される語彙と省略語句の分析

Analysis of vocabulary and omitted words in car license tests

的場 成紀 *¹
Seiki Matoba

古賀 雅樹 *¹
Masaki Koga

大塚 基広 *¹
Motohiro Otsuka

小林 一郎 *²
Ichirou Kobayashi

平 博順 *¹
Hirotoshi Taira

*¹大阪工業大学大学院 情報科学研究科

Faculty of Information Science and Technology, Osaka Institute of Technology

*²お茶の水女子大学大学院 人間文化創成科学研究科

Graduate School of Humanities and Sciences, Ochanomizu University

We develop a solver for Japanese car license tests. The test consists of about a hundred of true/false questions about traffic rules, driving manners, architectures of cars and the laws of physics related to cars. While the passing score is 90%, The best score in the previous approaches is about 65%. The approach is based on the sentence similarity between the test sentence and most similar sentence with the gold-standard answer in the database in the solver. Toward the system to pass the test, we analyzed the vocabulary and writing styles of the tests. The results of the analysis showed that the vocabulary is relatively small, which is about 300 words for 100 problems, and the sentences contain a lot of zero pronouns and they cause the low accuracy of the solver. Furthermore, we tried to resolve the antecedents using a previous anaphora resolution system. The results showed that the system cannot resolve the anaphora in the tests, because each problem consists of only one sentence and the clue to resolve the pronoun is very few, and they are more difficult to resolve than ones in standard articles. The analysis has revealed that high-performance systems require the anaphora resolution which is more based on domain specific knowledge.

1. はじめに

自動車免許試験に対して自動解答するシステムの検討を行っている [平 14]. 本研究は自動車免許試験を自動解答するソルバを開発するために免許試験のコーパスの分析を行う. これまでの研究では, 問題文とソルバが持っている問題データベース中の問題文との間の単語類似度を利用して正誤判定を行う手法が提案されている [杉村 13]. この手法では, 図を使用しない自然文のみからなる模擬問題に対して, 約 6 割の正解率が得られている. また, 従来の自動車免許試験についての問題分析では, 問題のトピック, 問題の言い回し, 問題を解くために必要な技術や知識に関する検討が行われている [平 15]. 本研究では, 免許試験問題に対して自動解答する上で必要な語彙について 5W1H の観点から分析するとともに, 問題文中で頻出する省略語句について分析を行った.

2. 普通自動車免許学科試験問題の概要

普通自動車免許を取得するためには, 実際に自動車を運転し, 運転技術について評価する実技試験と交通規則やマナー等, 運転する上で必要となる知識を問う学科試験の 2 つに合格する必要がある. 後者の学科試験は, 制限時間 50 分の筆記試験である. 出題形式は正誤判定問題であり, 交通規則やマナー, 運転知識などについて述べられた問題文に対し, そこで述べられていることが正しいか誤っているかを判定する問題である. 問題は 95 問出題され, 先頭 90 問は, 問題文各 1 文を読み正誤判定を行う問題である. 自然文の問題文だけを読み解答する問題が多いが, 交通標識などのイラストの絵を参照して解答する問題も存在している. それに対し, 末尾の 5 問は, 1 問につき 1 枚の運転席から見た外の様子などのイラストが示され, 3

つの枝問の正誤判定問題を解く問題である. 先頭 90 問は各 1 点, 末尾 5 問は完答で各 2 点で, 合計 100 点満点である. 学科試験の合格基準は 90 点以上である.

3. 学科試験問題で使用されている語彙の分析

3.1 問題文中の語彙数

今回, 分析の対象とした学科試験の模擬問題は「試験によく出る普通免許 1000 題」(倉 宣昭著, 高橋書店)の第 10 回の問題の問題文 100 文とした. まず, 問題文中に含まれる内容語の語彙数について調査したところ, 319 語であった. ただし, 「原動機付自転車」などの語は「原動機」「付」「自転車」と 3 つには分けて 1 語として扱った.

3.2 問題のタイプ

3.2.1 自動解答処理の観点から見た問題分類

試験問題には異なったタイプのものが存在する. そのため, 解答する際にはどこに注目すべきなのかを理解する必要がある. まず, 試験問題は次の通り大きく 3 種に分類する事が出来ると考えた. まず 1 つ目はイラストを見て解答する問題である (以下「イラスト問題」と略す). 分析対象の 100 問中にはイラスト問題は 13 問存在した.

2 つ目は問題文で示された物理現象について, 法則に基づいて計算が必要な問題である. この問題では, 問題文が問われている法則の種類, 計算に必要な情報を特定した上で, 計算を行った上で正誤判定を行う必要がある. これは 100 問中, 2 問存在した.

3 つ目はそれ以外の問題である. この問題は, イラストはなく, 問題の自然文で記されている交通ルールやマナーなどについての記述が正しいかどうかを判定する問題である. この問題は, 100 問中, 85 問存在した.

連絡先: 的場成紀, 大阪工業大学情報科学部, 〒 573-0171 大阪府枚方市北山 1-79-1, e1b15097@st.oit.ac.jp

3.2.2 詳細な分類

以下の異なる 3 つの観点で、問題をより詳細に分類した。

- 1. 5W1H による分類
- 2. 問われ方による分類
- 3. 問題の内容

3.2.3 5W1H による分析

問題文中に出現する内容語の語彙について 5W1H の観点から分析を行った。ここで 5W1H は、時間 (when)、場所 (where)、ガ格 (who)、ヲ格 (what)、述語 (how) とした。

- (1) when: 季節や天候、時間帯、災害時なども含めた時期や時間
- (2) where: 場所
- (3) who: 文中のガ格に相当する内容語
- (4) what: 文中のヲ格に相当する内容語

5W1H の観点から内容語を分類したものを表 1 に示す。
まず、「時間」の中で最も多く出現した単語は「夜間」であった。これは、視界や前照灯に関する問題において、「夜間」という単語が多く用いられているためである。
「場所」に関しては単語「道路」が最も多く出現していた。また、「道路」という単語は、単独で使用されることは少なく、「一方通行の」や「車両通行帯のある」などの修飾語句を伴って使用されることが多かった。
「ガ格」については、「運転手」「自動車」などが多く出現していたが、「ブレーキ」「後車輪」などの自動車の部品に関する単語も多く見られた。また、ガ格である単語が直接的な問題の対象物ではない問題も多く見られた。例えば、

- ファン・ベルトの中央を指で押してみたら、50 ミリメートルぐらいの緩みがあった

の問題では、述語「あった」に対するガ格は「緩み」であるが、問題で問われている知識は、「ファン・ベルトの緩みの大きさ」についての知識であり、問題が問うている対象の解析を難しくしていることが分かった。

3.2.4 問題の問われ方による分析

次に、問題の問われ方について分析を行った。自動車免許試験の正誤判定問題の言い回しは、ある種独特のものであり、「～は～である」といった「事実について述べている文」と「～は～した」といった運転手や車などの「行動について述べている文」の 2 種類が存在し、それらの割合について分析を行った。
分析の結果、表 2 で示す通り、「事実について述べている文」の方が多く、出現数は 84 であった。それに対して、「行動について述べている文」の出現数は 16 であった。さらに、それらの文についての文末表現について分析したところ、「行動について述べている文」に関しては最後が「～した」で終わることが多く、「事実について述べている文」には「～である」「～できる」「～になる」「～よい」など、表現に多様性が見られた。

3.2.5 問題の内容による分析

問題の内容による分析を行う。
大きく分ける種類として 4 種類あり、それぞれ「運転」「準備」「知識」「心がけ」である。運転は運転に関することに対して問われている問題である。例えば「追い越し」などである。

表 1: 5W1H の観点から分類した内容語と頻度

5W1H	単語 (頻度)
when (時間)	夜間 (6), 雨 (3) 冬 (1), 雪 (1) 災害 (1), 光化学スモッグ (1)
where (場所)	道路 (12), 交差点 (6) 高速道路/高速/高速自動車国道 (4), 坂 (4) 踏切 (3), 路側帯 (2), 踏切 (2) 曲がり角 (1), 市街地 (1), トンネル (1)
who (ガ格)	運転手 (4), 速度 (3) 自動車/車/普通自動車 (3), 者 (3) エンジン・オイル (2), 空気圧 (2) 原動機付き自転車/原動機付自転車 (2), 追い越し (1) 警察官 (1), 交通 (1) 光化学スモッグ (1), 交通公害 (1) 後車輪 (1), 視線 (1) 灯火 (1), 燃料 (1) (オイルの) 循環状態 (1), (燃料の) 消費量 (1) タイヤ・チェーン (1), トンネル (1) (排気の) 色 (1), ハンド・ブレーキ (1) 普通二輪車 (1), ブレーキ (1) 前車輪 (1), (タイヤの) 溝 (1) ミニカー (1), 路面電車 (1)
what (ヲ格)	自動車 (8), オートマチック車 (8), 車 (8) ブレーキ (3), 急ブレーキ (3) ペダル (3), ブレーキ・ペダル (3) 原動機付き自転車/原動機付自転車 (2) 人 (2), 2 人 (2), 大型特殊自動車 (1) 違反 (1), 後車輪 (1) 燃料 (1), エンジン (1) ファン・ベルト (1), 速度 (1) 幼児 (1), 前照灯 (1) 急ハンドル (1), 視力 (1) チェンジ・レバー (1), 進路 (1) 通行帯 (1), 路側帯 (1) 危険 (1), 全引きしろ (1) ハイドロ・プレーニング現象 (1) 車間距離 (1) 速度超過 (1), 余地 (1) 積載超過 (1), 間隔 (1) 場所 (1), マイクロバス (1) 許可 (1), 車線 (1) 故障車 (1), 番号標 (1) 荷物 (1)

表 2: 問題の問われ方による分類の結果

問題の問われ方	頻度
事実について述べている文	84
行動について述べている文	16

表 3: 問題の内容による分類の結果

問題の内容	頻度
運転の技術	77
準備	17
知識	4
心がけ	2

準備は運転をする前に行う前の段階である。知識は運転に関する現象である。心がけは道徳的な話が含まれている。

表 3 にこれらの分類の頻度数を示す。一番多く出題されているのは「運転の技術」についてのものである。

次に多く出題されていたのは「準備」であった。「準備」に関する問題は、主に車の部品に関することや免許の種類に関する内容である。

「知識」に関する問題は、運転中におこる現象に関する内容や数式に分類される問題である。

「心がけ」は出題される問題が一番少なかった。

3.3 問題を解く上での必要な知識の分析

問題を解くために必要な知識を考察する。先の問題の分類を踏まえると

1. 運転に関する知識
2. 常識的な知識

の 2 種類が必要であると考えられる。

運転に関する知識とは「交通の方法に関する教則（昭和 53 年 10 月 30 日 国家公安委員会告示第 3 号）」に掲載されている知識について問われている。例えば

- 車の速度が 2 倍になると、制動距離はおおよそ 4 倍になる。

上記のような問題の場合、「制動距離」という単語の意味を知らなければならない。加えて、制動距離の算出の仕方も知っている必要がある。

「常識的な知識」とは運転に関する知識とは違い、明記されていない情報を想像するために必要な知識である。例を挙げると

- オートマチック車で坂を下るときは、チェンジ・レバーを 2 か 1 に入れ、エンジン・ブレーキを活用する。

上記の問題では動作主格となる単語が文中にない。しかし、人間の受験者は、問題文を見たときに「動作主格が運転手である」と容易に想像することができる。これは人が今までに培ってきた経験から推測できるためである。

3.4 問題文の省略解析

運転免許問題には省略語句が多く含まれる。問題文に対して、省略解析を行って分析を試みた。省略解析器として KNP を使用した。

表 4 に KNP による解析結果を示す。100 問の問題文中で、省略格指定が 17 個、省略解析対象指示詞が 11 個、合計 28 個の省略があった。省略格指定では主に「見える」の単語が対象となった。例えば

- 暗いトンネルから明るい場所へ出たときは、視力が急激に低下して、見えなくなることがある。

表 4: 省略解析の結果

ゼロ照応解析の結果	頻度
省略格指定	17
省略解析対象指示詞	11
合計	28

上記のような問題であった場合、「見えなくなる」の部分が省略格指定となる。また、「不特定人が」、「トンネルから」、「場所」が直接係り受け解析の格の対象となる。不特定人とは問題文の中に明記はされていないが主格の対象となるのは人であるときの対象となる。また、「とき」という単語が明記されていないが「見えなくなる」に係り受け解析の結果となっており、「見える」が何に対して指示対象としているのかを表す単語は「視力」となっている。

上記以外にも「～という」「～といえる」の単語が省略格指定の対象となった。例えば

- 運転者は、「酒を飲んだら運転しない」「乗るなら飲まない」という習慣を身につけることが大切である。
- 交通法令を守らなくても、臨機応変に運転して交通事故を起こさなければ、「安全な運転」といえる。

である。「～という」の場合、主に主格となる部分は不特定人となった。また、「～といえる」の場合は主格が不特定人となるが、上記の例では「運転と」が省略されていると特定された。

上記の KNP で省略解析を行った結果と予測していた結果と異なっていた。その原因の一つとして、免許問題が 1 文で構成されることが挙げられる。本来、KNP で省略解析を行う場合は 2 文以上あることを想定している。しかし、今回のように問題文が 1 文しかないために予想とは違った結果となったと考えられる。

4. おわりに

本稿では問題文の分析を 5W1H の観点からの使用語彙と省略語句について分析を行った。5W1H の観点で使用語彙を見た場合、「時間」は「夜間」、「場所」は「道路」、「ガ格」は「運転手」、「ヲ格」は「自動車」が一番多く頻出した。さらに省略語句について省略解析器を用いて問題文の解析を行い分析を行った。今後は、上記の結果を踏まえて自動解答システムを開発する予定である。

謝辞

本研究は JSPS 科研費 18K11452 の助成を受けたものである。

参考文献

- [杉村 13] 杉村 皓太, 佐々木 裕: 交通規則問題のための解答システムの構築, 言語処理学会第 19 回年次大会 発表論文集, pp. 790–793 (2013)
- [平 14] 平 博順, 田中 貴秋, 永田 昌明: 自動車運転免許試験 RTE コーパスの構築, 第 28 回人工知能学会全国大会予稿集, 3I4-5 (2014)
- [平 15] 平 博順: 自動車免許試験自動解答に向けた問題分析, 第 29 回人工知能学会全国大会予稿集, 1K2-2 (2015)