

心理表現を含む運転関連表現辞書の構築

岩井 律子^{1,2}・熊田 孝恒^{1,2}・高橋 憲生¹・河原 大輔¹・黒橋 禎夫¹

¹京都大学大学院情報学研究科

²理研 BSI-トヨタ連携センター

iwai.ritsuko.8n@kyoto-u.ac.jp; {t.kumada, dk, kuro}@i.kyoto-u.ac.jp; takahashi.norio.24x@st.kyoto-u.ac.jp

1. はじめに

人間は、相手の意図を推測し行動を予測するための知識を発達過程で獲得し、その知識を用いて他者と自然なコミュニケーションや共同作業を行っている (Tomasello, 1999)。つまり、知識を基に、他者の行動からその意図や感情等の心理状態、次の行動を予測し、自己を内省し、自分の言動を調整しながら他者と接している。また、相手も同様のことを行い、互いに自分の言動を調整し合うことによって自然なやりとりが成立している。

近年、対話システムやロボットの開発が活発化している。しかし、現在の対話システムや人工知能は上記の知識を備えないため、ユーザーは対話システムやロボットとのやりとりに対して対人同様の自然さを実感しづらい。

そこで、我々はテキストに表れる行動と主観の共起関係から人間の自然なコミュニケーションを支える知識の収集を試みている。主観とは、他者や自己の行動に関する評価や心の動き (感情、思考など) といった、他者が観察できない個々人の心理状態を指す。行動は多岐にわたるため、本研究では比較的知識が限定される自動車運転に着目した。煽り運転・飲酒運転・高齢者による危険運転等が社会的な問題になっており、運転行動と主観に関する知識の収集は重要だと考える。

そのために、運転行動とその行動に対する主観の記述を多く含むと思われるコーパスが必要となる。例えば、「車が進路を妨害する」という運転行動を見て、あの運転手は自分勝手だと主観を持ち、この先も自分勝手な運転をするかもしれないと推測し対策を人間は考える。

しかし、そのようなコーパスは存在しておらず、また構築するのは容易ではない。「車・自動車」でウェブ検索すると、事故ニュース記事の引用、新車のレビューサイト、自動車販売店の宣伝サイト、また最近は運転マナーがなっていないという一般論等も含まれる。そのためには、実際の運転状況を記述しているだろうと判断できる表現が必要である。例えば、「車」という語だけでは、「この車は加速度が優れている」という車の性能を記述したテキストを抽出してしまうが、「車を運転する」や「対向車とすれ違う」という表現を用いれば、実際の運転行動を記述したテキストを抽出できる。また、運転行動に限定せず、主観に関わる表現 (心理表現) も必要である。例えば、ある人が「このパンがとても美味しい」と思うのは主観だが、運転行動に関わるとは想定しづらい。しかし、「渋滞で苛々する」や「運転するのは不安だ」という心理表現は、テキスト中で著者の、ある運転行動に対する主観として出現する可能性が高い。そのため、まず人間の心理表現を含む運転関連表現辞書を構築した。また、構築した辞書を活用して、運転行動・主観コーパスの自動構築を試みた。

2. 関連研究

自動車運転に関わる自然言語処理に関連した先行研究はいくつか存在する。平ら (2014) は、自動車運転免許試験への回答が交通法令や運転のマナー、運転操作など幅広い知識を要求することに着目し、学科試験問題集の問題文について交通法令等を関連付けてテキスト含意認識に関わる評価用コーパスを作成した。河辺ら (2015) は、運転に必要な交通法規や交通マナーなど交通に関する知識のオントロジー構築を、交通教則本を用いて半自動拡張して行った。鈴木ら (2015) は、

交通オントロロジーに対する問い合わせに機械が回答できるような質問文を SQL クエリに変換している。また、高山ら(2017)は、交通オントロロジーを用いた質問応答システム構築を目的に、質問応答データセットの構築を行っている。いずれも、交通法規や運転操作、マナーなどが中心である。

また、中川(2017)は人手で収集した運転シードコーパス(62記事)の単語出現率を元に1,000万ブログ記事から運転行動コーパス(16,011記事)を抽出した。しかし、本コーパスは自転車やバイクのみの場面が多く含まれており本研究には適していなかった。

3. 運転関連表現辞書の収集・構築

運転操作・行動や運転環境に限らず、運転行動に伴う心理的表現を含めた言語資源を収集するために、次の4段階により構築した。

3.1. 運転行動語の獲得

まず、運転行動に深く関連すると思われる単語を収集すべく、(1)中川(2017)が作成した運転関連語(426語)、(2)JUMAN交通関連語(426語)、(3)独自に収集したヒヤリ・苦手回答から抽出した単語(1,292語)を用いた。(2)は、形態素解析器JUMAN¹の基本語辞書からドメインが交通の単語を抽出したものである。

ヒヤリ・苦手回答は、2012年8月から2015年10月にかけて産業総合研究所及び理化学研究所で収集された、639名の研究協力者($M_{age}=42.2$, $SD=22.6$, 女性=260)が手書きで記述した運転体験・操作に関するテキストである。運転歴の浅い若齢ドライバーから40年を超える高齢ドライバーと幅広い運転歴を持つ協力者(平均免許保有年数=20.0年, $SD=19.8$)が、運転中ヒヤリとした体験と運転で苦手な操作(苦手な運転操作・理由・対策を最大10件まで記述)について回答した。

ヒヤリの回答例としては、「長時間のドライブで疲れていた時に、右車線が空いていたので、思わずハンドルを右に切ったところ、右すぐ後方にトラックが来ていた。私の車は左ハンドルだったので、右後方を振り返らずに、右ハンドルを切った」、「信号の無い十字路

で、あらゆる方向のミラーを確認した後、前進しようとした時、右側から原チャが直進してきて、接触しそうだった」などが含まれる。また、苦手操作についての回答例は、対向車とのすれ違いを苦手な運転操作とし、「ぶつかるような気がする。」という苦手理由と「必ず、白線なり、分離帯の位置を確かめる。」という自分の対処法というなどがあった。これらのデータに対し、形態素解析を行い、動詞・名詞(但し、固有名詞・形式名詞を除く)・形容詞を収集した。

以上の3つをすべて統合(1,976語)し、港湾・出国等無関連なものを除き1,173語(59.3%)、更に、動詞・サ変名詞を抽出(836語)し運転行動語をリスト化した。運転、右折、左折、横断、当てる、という運転についてだけではなく、落ち着く、不足、注意、確認等の心理表現単語が含まれるのが特徴である。

3.2. 運転関連述語項構造の取得

前節の運転行動語を元に、以下の3つのテキストに対して述語項解析を行い、1つ以上の項を持つ述語項構造を取得した。

1. ヒヤリ・苦手回答(3.1節(3)と同じ):全記述に対して、出現頻度にかかわらず述語項構造(2,106表現)を収集した。
2. 運転教本(学科・運転):トヨタ名古屋教育センター(1998, 2004)が制作・販売する運転教本を、本研究目的により提供を受けた。出現頻度にかかわらず述語項構造(2,722表現)を収集した。
3. 運転ブログ:運転行動コーパス(中川, 2017)に対し述語項解析を行った。うち、運転行動語を述語、あるいは項に含む述語項構造で、出現頻度上位10,000表現を収集した。

ヒヤリ・苦手回答と運転教本については、運転に焦点化されているため、1つ以上の項を持つすべての述語項構造を収集することにした。「道が狭い」などは、運転行動語を含まないが、運転中に出てくる描写の可能性が高い。このような運転の周辺環境の描写などにも対応するためである。

従って、表現には、運転行動語を述語とする述語項

¹ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

構造 (Type 1)、運転行動語を項に含む述語項構造 (Type 2)、運転行動語を含まない述語項構造 (Type 3) の 3 種類が存在する。

3.3. 日本語・ブログ関連フィルタリング

得られた述語項構造全体のうち、まず日本語として成立しない表現を削除した。また、運転行動コーパスから取得した述語項構造についてはブログ特有表現 (例: 続きを読む、広告がブログに表示、読者になる) を排除した。表 1 に、フィルタリング後の各データの数、並びにタイプの内訳を示す。これらの表現を運転関連表現と呼ぶ。

表 1: 運転関連表現内訳

データ元	Type 1	Type 2	Type 3	合計
手書き	744	110	673	1,527
教本	973	211	564	1,748
ブログ	4,907	857	----	5,764

3.4. カテゴリ付与

前項の運転関連表現に対し、直接的に運転に関わるものから周辺環境、状況や心理関連と内容が多岐にわたるため、以下、4つのカテゴリを付与した。

- A) 運転行動表現: 直接的な運転行動。
- B) 運転文脈表現: 運転中であれば、運転行動として使用可能。
- C) 心理表現: 運転時の意識・心がけに関する表現 (知覚・認知・心理状態等)。例えば、対向車に気をつける、方向に気を配る、集中力を失う。
- D) 運転周辺表現: 運転中に遭遇する、周囲の状況や出来事。例えば、老人が飛び出す、自転車が飛び出す、カーブが続く、路面が凍結等。

運転行動表現とは、このカテゴリに属するいずれかの述語項構造が現れれば、車を運転している記述だと推測できそうな表現である。例えば、「車を運転する」、「バックミラーを確認する」、「ハンドルを切る」、「アクセルを踏む」などが本カテゴリに属する。運転文脈表現とは、文脈から車の運転場面だと明らかであれば、運転行動を表現すると推測できる表現を指す。本

カテゴリには、「距離を保つ」、「速度を出す」が該当する。「距離を保つ」は、人間関係において「距離を保つ」など用いられるが、運転場面で先行車との距離を保つことは運転行動に該当する。また、「速度を出す」はマラソン、歩行などの場面も想定できるが、運転場面では車の速度を出す運転行動が想定できる。

上記の内容を定義書として作成し、3.3 節の表現リストに対し、評価者 (女性 6 名) がカテゴリを付与した。カテゴリは、1 表現に 1 つではなく複数付与されているものも存在する。但し、人的リソースの都合上評価者の運転知識は限定的であり (無免許/ペーパードライバー)、1 名の評価による結果である。

表 2 に各カテゴリの内訳を、また、表 3 に各カテゴリの表現例を示す。

表 2: 各カテゴリの内訳

A	B	C	D	該当無
1,491	2,338	1,825	2,694	3,633

表 3: 各カテゴリの表現例

カテゴリ	例
A	アクセルを踏む、ニュートラルにギア、ハンドルを切る、バックで駐車
B	相手を先に行かせる、スピードが速い、方向に回避、トランクを開く
C	ブレーキを確認、目視をシッカリ、目視が甘い、飛び出しに気を付ける
D	歩行者が飛び出す、信号が黄色に変わる、カーブが続く、渋滞がある

4. 運転行動・主観コーパス構築

運転関連表現辞書を用いて、運転行動とその行動に対する主観が多く含まれるコーパスを構築した。これを運転行動・主観コーパスと呼ぶ。

まず、ブログサイトで公開されている日本語のブログ記事をクローリングし、約 7,000 万記事のブログコーパスを構築した。次に、以下に示す方法でブログコーパスから、運転行動・主観コーパスを構築した。

- (1) 3節の辞書から特徴的な名詞・動詞・形容詞を抽出し、運転特徴語(384単語)を得た。例えば、表3のAの表現例からは、運転特徴語「アクセル、ニュートラル、ギア、ハンドル、バック、駐車」を抽出した。
- (2) 各記事に対して運転特徴語出現率を計算し、上位1,000万件の記事を抽出した。出現率は「運転特徴語数/記事中の文の数」で計算した。
- (3) (2)のブログ記事に対して運転関連表現出現率(以降、表現出現率と呼ぶ)を同様に計算し、上位100万件の記事を抽出した。
- (4) (3)のブログ記事から、「運転行動及び主観」の記述を含むブログ記事の正例296記事、負例569記事、合計865記事を人手で抽出した。
- (5) 目的変数を「運転行動及び主観」の有無、説明変数を記事中の自立語として、RBFカーネルを用いたSVMで学習を行い、分類器を作成した。適合率は0.776、再現率は0.507であった。
- (6) 分類器で、(3)のブログ記事を分類し、23,121件を抽出し、これを運転行動・主観コーパスとした。

運転行動・主観コーパスをクラウドソーシングで評価すべく、2,000記事をランダムに抽出した。1記事当たり5人のワーカーが「運転行動及び主観」の有無を評価した。5人中3人以上が「有」と評価した記事は511件、「無」と評価した記事は1,123件、適合率は0.313であり、(5)より低かった。これは、学習用記事の抽出が表現出現率上位の記事に偏っており、(3)の100万記事全てに適用した場合に表現出現率下位の記事を精度良く分類できなかつたためと考えられる。

また、クラウドソーシングで「運転行動及び主観」が「無」と判定されたブログ記事を分析すると、運転行動を含む記事は多かったが、主観を含む記事は少なかった。これは、学習に使用した正例296記事中に主観に関する表現をカバーすることが難しく、SVMで主観表現の有無を精度良く分類できなかつたと考えられる。

クラウドソーシングの判定結果は今後学習データに加え、運転行動・主観コーパスを改良する予定である。

5. 終わりに

本研究では、運転に関連が高いと思われる心理表現を含む運転関連表現辞書を整備し、運転行動・主観コーパスの自動構築を試みた。

Iwai et al. (2018)は、人手でタグを付与した運転体験コーパスを構築し、本研究の運転行動語と運転関連表現辞書を素性としてCRFによる運転行動(F値0.698)と主観(F値0.749)の自動抽出を試みた。今後は、当該CRFモデルを運転行動・主観コーパスに適用し運転行動と主観の自動抽出を行う予定である。

本研究で作成した運転行動語ならびに運転関連表現辞書は、研究利用に限り公開予定である。

謝辞

本研究にあたりデータ利用を快く許諾してくださった(株)トヨタ名古屋教育センターに感謝いたします。

参考文献

- Iwai, R., Kawahara, D., Kumada, T., & Kurohashi, S. (2018). Annotating a driving experience corpus with behavior and subjectivity. *In Proceedings of PACLIC 32*.
- 河辺一仁ら (2015). 交通オントロジーの半自動拡張のための交通用語認識. *言語処理学会第21回年次大会発表論文集*, 135-138.
- 中川裕貴. (2017). コーパスから獲得した運転関連語に基づく運転行動コーパスの自動構築. *卒業論文*. 京都大学.
- 鈴木遼司ら (2015). 交通オントロジーを対象とした質問文のSPARQLクエリ変換. *言語処理学会第21回年次大会発表論文集*, 171-174.
- 平博順ら (2014). 自動車運転免許試験RTEコーパスの構築. *人工知能学会全国大会論文集*, Vol.28, 314-315.
- 高山隼矢ら. (2017). 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. *言語処理学会第23回年次大会発表論文集*, 691-694.
- Tomosello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press.
- トヨタ名古屋教育センター. (1998). *学科教本*.
- トヨタ名古屋教育センター. (2004). *運転教本*.