

シソーラスを用いた交通経路の認知ルールの抽出

鷹尾 和享 朝倉 康夫

神戸大学大学院

E-mail: 003d912n@y02.kobe-u.ac.jp, asakura@kobe-u.ac.jp

1 はじめに

従来の国土交通計画は、主に経済効果や需要予測という視点から行われてきている。したがって、交通行動研究においても、多くの場合は行動結果に関心が向けられており、心理学的な側面に目が向けられることは少なかった。特に経路選択行動の心理的な要素に目を向けたフレームワークはあまりないのが現状である。

交通工学では交通移動の単位は起点 (Origin) から終点 (Destination) への移動としてとらえられ、「トリップ」 (trip) と呼ばれる。また、起終点のペアは「OD ペア」と呼ばれる。経路選択行動はそれぞれのトリップにおいて、候補となる経路の集合から1つを選択することである。

したがって、各経路がどのように認知されているかを分析することにより、経路選択行動における心理状態をとらえることができる。多くの研究例が物理的に観測可能なデータを用いて経済学的な分析を行っているのに対し、筆者らはことばを用いて心理状態をとらえるを試みる。

その方法として、筆者らは、交通経路を構成する空間の認知と経路選択行動をことばによってモデル化することを試みている。これは、交通空間と認知結果の間を認知ルールの形で表現し、自由記述文で記述された各経路の特徴から認知ルールを抽出するという手法である。本稿ではシソーラスを用いて認知ルールを抽出した結果について報告する。筆者らの最終目標は、ことばを用いて経路選択行動の心理状態を説明するモデルを構築することであり、そのための言語処理の技術の応用という観点から論述する。

2 経路選択行動のモデル

2.1 経路選択のモデルの全体像

筆者らは、交通経路選択のプロセスの全体像を次のように表現する (図1)。詳しくは[1]を参照されたい。

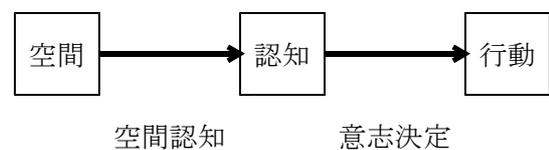


図1：経路選択のプロセス

第1段階は、物理的な交通空間内の交通経路から、各経路の認知結果に至るまでの空間認知プロセスである。この段階は、各経路の印象や特徴が記述された自由記述文を解析することによって、認知ルールの形で抽出する。

第2段階は、それぞれの認知結果を評価して候補となる経路を取捨選択し、最終的な経路を選択する意志決定プロセスである。この段階は Tversky (1972) の Elimination-By-Aspects[2] のような意志決定モデルで表現することができる。

「アスペクト」とは、ある状況を表す特徴、たとえば「明るい」「快適」のような、選択候補のいくつかの経路に共通して現れる認知結果を意味する。Elimination-By-Aspects では、意志決定は、着目しているアスペクトを各経路が持っているか否かによって候補を順に排除していくことで行われる。たとえば、着目するアスペクトが「明るい」の場合、明るくない経路は選択候補から取り除かれる。選択候補が1つに絞られなかった場合は次のアスペクトに着目する。これを繰り返すことによって最終的に1つに絞られた経路が選択されることになる。

本稿では第1段階である空間認知プロセスにつ

いて述べる。ここでは、後の段階の意志決定プロセスに必要な各経路のアスペクトを、自由記述文から如何に抽出するかがテーマとなる。

2.2 空間認知モデル

本稿では、ことばを以下の3つのタイプに分類することにより、認知ルールという形で空間認知プロセスを表現する。

経路の属性 & 空間条件 = 認知結果

「経路の属性」とは、それぞれの経路の静的な属性を知覚または観察した結果を直接、またはそのまま表現したことばである。「空間条件」とは、季節・天気・時刻のように、状況によって動的に変化するような交通空間の条件を表すことばである。なお、急いでいる等の個人の条件も動的に変化するという意味で空間条件に含めて考える。「認知結果」とは、結果として生じる感覚や印象等を表すことばである。たとえば、ある経路に照明灯が存在し、時刻が夜の場合は「明るい」と感じる。これを、

照明灯 & 夜 = 明るい

と表す。

3 データ収集

筆者らは、トリップ目的が同じで、OD ペアが同じである数名の被験者に対し、選択する可能性のある経路を挙げてもらい、それぞれの経路についての印象や思うことを E メールによって自由記述文に書いてもらった。交通手段の違う場合もそれぞれを別経路として扱った。

経路の属性、空間条件、認知結果の3つのタイプのことばを明確化するために、必要ならば再質問を行った。なぜならば、日常会話では、常識や暗黙の了解事項が存在するため、認知結果はしばしば省略されて表面化しないからである。たとえば、「階段がある」で文が終わってしまう場合は、「しんどい」のか「適度な運動は健康に良い」のかを判別できない。日常会話では相手の気持ちをくみ取って「しんどい」ということを言おうとしているのだと推察するが、本稿の場合は明記させ

る必要がある。一連の再質問のやりとりは最初の回答文に含めた。すなわち、再質問の回答文は最初の回答文と同じエントリとして扱った。また、被験者の回答能力を考慮してこれ以上の再質問は無理と判断した場合はその時点でやりとりを打ち切った。

収集したエントリ数は計 66 エントリである。このサイズは通常の言語処理で用いられるコーパスのサイズと比べて小さい。図1の一連の経路選択プロセスをモデル化するため、認知結果の多様化を避ける意味で OD ペアを絞る必要があることと、再質問が必要なので大量のデータを収集するのは現実的に困難なことの2点がある理由である。

4 認知ルールの抽出

本稿ではシソーラスを用いて文中の単語を3つのタイプに判定し、ルールとして抽出した。いったんシソーラスのカテゴリ（意味コード）と3つのタイプの対応関係を求めておけば、シソーラスと対応関係をテンプレートとして用いることにより、自由記述文中の各単語から3つのタイプの単語を抽出し、認知ルールとすることができる（図2）。

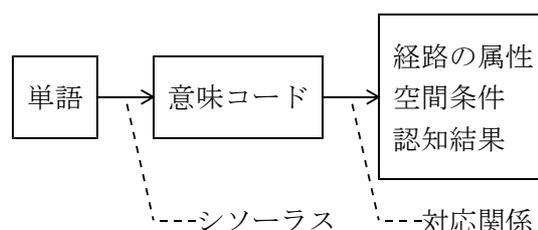


図2：認知ルールの抽出

この方法は、決定リスト法のルールのリストの抽出、つまり、ある入力に対する出力が素性によって変化するルールを抽出する手法と類似の方法である。本稿の場合は入力が経路の属性、出力が認知結果、素性が空間条件となる。しかし、決定リスト法が信頼性の高いルールもそうでないルールもいったんすべて抽出し、そのあとでルールの信

頼性を統計的に求めるのに対し、本稿の場合は、良いルールを抽出するためにできるだけ良いテンプレートにすることに焦点を当てる点が異なっている。テンプレートを作成する具体的な手順は次の通りである。なお、形態素解析は ChaSen[3]を用い、シソーラスは日本語語彙大系[4]を用いた。

- (1) 訓練文を形態素解析する。
- (2) 各単語に意味コードを付与する。シソーラスに載っていない場合は手作業で付与した。
- (3) 各単語を3タイプに判定する。助詞・助動詞や意味の薄い語は無視する。
- (4) 意味コードと3タイプとの関係を求める。

ルール抽出の手順は以下の通りである。

- (1) 対象の文を形態素解析する。
- (2) 各単語に意味コードを付与する。
- (3) 対応関係から3タイプを判定する。3タイプに関係のない意味コードの単語は対応関係がないので無視される。

このルール抽出の手順は初見の文に対して用いることが可能である。しかし、本稿では、手法としての根本的な問題点の有無を見極めるため予期しないエラーの混入を避ける必要性と、データ量の少なさのため、訓練文を用いた。

5 抽出結果とその考察

5.1 評価方法

空間条件と認知結果が正しく抽出できているかによって認知ルールの抽出を評価した。経路の属性については別の機会にゆずる。評価は、「ゴミが混入」等の問題点を列挙し、各エントリにそれぞれの問題点が該当するか否かを評価者が評価することで行った。結果は表1のとおりである。問題点が1つも該当しない場合は「うまく抽出できた」に分類される。

5.2 考察

66 エントリのうち、22 エントリは問題なく抽出できた。

38 エントリは何らかのエラーを含んでいた。表1の(B-1)～(B-4)の細分類はこのべ数であり、1エントリに複数の細分類が該当する場合がある。

(B-1)のように、19 エントリが理解の困難なルールを抽出した。理由の1つは、認知結果を形態素1つで表すのは困難な場合があるためである。たとえば、例のように、「悪い」は何が悪いのか理解困難であり、「時間」がどうなのか不明である。「乗り継ぎが悪い」「時間に余裕」といった形であれば理解できるので、複数の形態素から成る語をシソーラスに追加するか、または、辞書に載っているような1形態素の語に換言する必要がある。別の理由は、本稿で用いたのが1段階の認知ルールのテンプレートなので、複数段階の認知ルールをうまく扱えていないためである。たとえば、(B-1)の2番目の例では「自転車→労力」「労力→暑い」という2段階になっているが、このような場合にはテンプレートがうまく機能しない。

(B-2)のように誤ったルールを抽出したものは5 エントリあった。原因の1つは、1つの文に複数の内容が記述されている場合があるためである。別の原因は、シソーラスの分類体系が荒すぎて、同じ意味コードに空間条件を表す語と認知結果を表す語の両方が含まれている場合があり、対応関係の判定を誤っている場合があるためである。

(B-3)のように何らかのゴミを含んでいるものが22あった。原因の1つは、最初の例の「急に」のように修飾文節の語を拾っている場合である。このようなゴミを除去するためには係り受け解析によって修飾語の判定をする必要があると思われる。また、別の原因は2番目の例のように「万が一…でも」のような言い回しの中の語を拾っている場合である。3番目の例は「気持ち」という意味の薄い語が混入している例である。ただし、意味の薄い語は、後の意志決定の段階でアスペクトとして挙がってこないと思われる。したがって、意味の薄い語は無視されるだけであり、混入していても大きな問題ではないと思われる。これを「何らかのエラー」として扱うのは厳しすぎるかもしれないが、本稿では、テンプレートをできるだけ良くするための指針を見いだす意味で、厳しく扱った。なお、意味の薄い語の混入と、誤ったアス

ペクトをもたらず語の混入との区別が必要になると思われる。

(B-4)の抽出不足は、「行き着ける」のような可能動詞を標準形に直して訓練したため、認知結果として取り出すことができなかつたのが原因である。この場合の標準形の動詞「行き着く」を認知結果として扱った場合、多くのゴミが混入してしまうので、有効な解決策ではない。したがって、可能動詞を別扱いする等の工夫が必要である。

(C)は元の文に認知結果が含まれていない場合である。データの解析手法だけではなく、データの効率的な収集方法についても工夫することが重要であることを示している。

6 まとめ

本稿では、自由記述文から交通経路選択行動の

認知ルールを抽出する方法として、シソーラスをテンプレートとして用いた。今後は、抽出精度をさらに改良するとともに、後の段階の意志決定までを扱う予定である。

参考文献

- [1] 鷹尾和享,朝倉康夫: ことばによる空間認知と交通行動のモデル, 土木計画学研究・講演集 Vol.28, ¥pdf¥274.pdf (2003)
- [2] Tversky,A.: Elimination by Aspects: A Theory of Choice, Psychological Review Vol.79, No.4, pp.281-299 (1972)
- [3] ChaSen, 奈良先端科学技術大学院大学, <http://chasen.aist-nara.ac.jp/>
- [4] NTT コミュニケーション科学基礎研究所監修, 池原他編: 日本語語彙大系 CD-ROM 版, 岩波書店 (1999)

表 1 : 抽出結果

評価分類	数	例文	空間条件	認知結果
(A) うまく抽出できたもの	22	雨の時、傘をさしながらは危険。	雨	危険
(B) 何らかの誤りを含む	38			
(B-1) 抽出されたルールが理解困難	のべ 19	地下鉄烏丸線と東西線の乗り継ぎが悪く、待ち時間があるから時間に余裕がある時でないとは利用しない。 暑い。自転車をこぐという労力を使うので暑い。汗をかく。		悪い,待ち時間,時間,余裕 暑い,労力,暑い,汗
(B-2) 誤ったルールが抽出されたもの	のべ 5	夏は涼しいし、冬はあったかい。	夏,冬	涼しい,あったかい
(B-3) ゴミが混入	のべ 20	観光客が急に立ち止まり危険。 万が一座れても道が混んでるといつ着くかわからないので不安になる。 結構楽チンなのとあともう少し、と思う惜しい気持ちが混じってる。		急,危険 座れる,混む,いつ,不安 楽チン,もう少し,惜しい,気持ち
(B-4) 抽出不足	のべ 1	雨以外は何とか行き着けるので。	(not) 雨	
(C) 被験者の回答能力不足	6	自転車で持ちきれないような荷物の時。		