

スケジュール登録のための発話文解析

藤井良介
鳥取大学工学部
jcyhbb@gmail.com

徳久雅人
鳥取大学工学部
tokuhisa@tottori-u.ac.jp

木村周平
鳥取大学工学部
kimura@tottori-u.ac.jp

1 はじめに

自動車の中などキーボードが使用しにくい状況において、スケジュールのメモをとるためには音声による入力の一つの手段である。運転の直前に追加された予定は運転に集中していると忘れやすいため、メモをとることができていれば行き忘れることを防止できる。

スケジュールという情報は、「日時（開始、終了）、場所（出発、到着）、用件」という5つの項目の組の系列である。スケジュールの登録のための発話文の解析においては、日時や場所の表現抽出、および、開始／終了等の下位分類の判定が必要となる。本稿では、これらの解析を行う手法を示す。

2 スケジュール登録の発話

2.1 例題

スケジュールを登録する発話の例を見ながら、処理目標を解説する。

例)

設定日時：2020年5月13日 8:12
設定場所：自宅駐車場
入力文1：10時半から大学でゼミがあります
入力文2：16時に自宅に帰ります

スケジュール表：

#	日時	場所	用件
1	2020年5月13日 10:30 ～***	大学 大学	ゼミが あります
2	2020年5月13日 *** ～16:00	*** 自宅	帰ります

※日時列上行 = 開始, 日時列下行 = 終了
場所列上行 = 出発, 場所列下行 = 到着
はレコード番号

まず、発話時の設定として、発話の日時および場所が与えられるものとする。次に、発話文が2つ入力される。文1の解析の際、日付は設定日時と同一と仮定する。時刻を文中から取得する。「～から」なので開始の日時とみなす。場所は「大学」である。用件は「ゼミ」である。ゼミの間、場所の変化はないので、「出発、到着」の区別がない。スケジュール表のレコード#1が埋められる。このレコード#1を基準として、文2の解析を行う。次に、文2の開始日時は不明である。終了日時は「16時」である。文2から直接的に得られる場所は「自宅」であり、到着場所と解釈する。こうしてレコード#2が生成される。

2.2 分析用コーパス

発話文とスケジュールを収録したコーパスは構築済みである [1]。発話文には、日時タグ<tmg>、場所タグ<loc>、および助詞タグ<p>が付与されている。

3 解析手法

3.1 概要

本稿では以下の手順で発話文の解析を行う。

- 手順1. 意図解析
- 手順2. スロットフィリング
- 手順3. 開始／終了, 出発／到着の判定

手順1および手順2を [2] の手法で解析する。以下に実行例を示す。

入力文：

10時半から大学でゼミがあります

意図：

```
serviceScheduler.comInformEvent(loc, tmg)
```

スロット：

```
loc="大学"=[6,7]
```

```
tmg="10時半"=[0,1,2,3]
```

ここで、意図は事象の伝達という意味である。スケジュールサービス (serviceScheduler) はドメイン

である。次に loc および tmg がスロット名である。数値は文字番目である。

手順3についてはルールベースによる解析を行う（詳細は次節で述べる）。

「loc で＋あり」というパターンの適用により、loc の示す場所が「出発, 到着」の場所と判定される。「tmg から＋あり」というパターンの適用により「開始」の日時として解析される。

3.2 開始／終了，出発／到着の判定

助詞の意味は係先の動詞に依存することに注意して、「開始／終了, 出発／到着」という日時や場所の下位分類の判定を行う。しかし、日本語の基本用言が約 6,000 語存在するため「助詞と動詞のペア」を準備することは効率が悪い。助詞だけで判断が可能な場合は「助詞と動詞のペア」での判定を避けておきたい。そのため本稿では次の 2 種類のルールセットを作成する。

- 助詞＋動詞によるルールセット
- 助詞によるルールセット

ここで、処理の起点が、スロット tmg の下位分類を求める、もしくは、スロット loc の下位分類を求める、ということであるため、助詞の意味が時間的か場所的かという判断は不要である。

分析用コーパスにおいて出現した表現について分析を行った結果を次節以降で説明する。

3.2.1 助詞＋動詞によるルールセットの作成

時間的表現について動詞の考慮が必要であった助詞は「に」であった。例を示す。

例文 1：10 時に出発します（開始）

例文 2：12 時に到着します（終了）

出発や到着という動詞は、移動の事象における端点を指す。その端点の時間を「に」で示すため、事象の時間帯の「開始」と「終了」という解釈がされやすいと、本稿ではまとめた。

分析用コーパスから対応する表現を列挙したものを表 1 に示す。なお、次節で示すが、「に」の時間的下位分類を「開始」とするので、運用上は終了のパターンのみを使用する。

場所について動詞の考慮が必要な助詞は存在しなかった。

表 1 助詞＋動詞型の時間的下位分類ルールセットの作成

下位分類	パターン
開始	に＋食べる, に＋起きる, に＋もらう に＋買う, に＋乗る, に＋倒れる に＋受ける, に＋引越す, に＋ある に＋飲む, に＋出発, に＋始まる
終了	に＋到着, に＋終わる, に＋行く に＋向かう, に＋来る, に＋帰る に＋する

3.2.2 助詞によるルールセットの作成

時間的表現について、分析用コーパスでは次の助詞が用いられていた。

「は, の, に, まで, から, には」

各助詞の下位分類を定める基準を次のとおりとする。

- 助詞だけを見て開始／終了が明確であればその通りとする。
- 不明確であれば明確な助詞に置き換えて例文を参照し判断する。

上記の基準に従うと、まず、時間の流れを表す助詞として「から」「まで」に注目する。「から」を「開始」「まで」を「終了」とする。次に、不明確であった助詞のうち、例えば「は」については次の文を参照する。

例文：あさっては東京に出張です。

例文の「あさっては」を「あさってから」に置き換えてもスケジュールの内容は変化しない。「あさってまで」に置き換えると異なる内容になる。ゆえに「は」の時間的下位分類を「開始」と判断した。

時間的下位分類を判定する助詞の分類を表 2 にまとめる。

表 2 助詞型の時間的下位分類ルールセット

下位分類	助詞
開始	は, に, から, には
終了	の, まで

場所について、分析用コーパスでは次の助詞が用いられていた。

「から, を, に, まで, で, へ」

各助詞の下位分類を定める基準は、上記の時間的表現の場合と同様である。すなわち、まず、「から」と「まで」はそれぞれ出発と到着を明確に表す助詞と判断する。次に、不明確であった助詞を例文を見ながら判断する。

例文：あした東京を出発します

例文の「東京を」を「東京から」に置き換えてもスケジュールの内容は変化しない。「東京まで」に置き換えると異なる内容になる。ゆえに「を」の場所的下位分類は「出発」と判断した。

次に「で」について例文を見ながら判断する。

例文1：東京で用事があります

例文2：東京で観光します

例文1の「用事」、例文2の「観光」の間、場所の変化がない、または、一定の範囲内である。「出発」と「到着」のどちらにも属するものであると判断した。

例文3：新宿で高速バスに乗ります

例文3ではバスの移動をイメージすると「新宿」が出発点と解釈することになるが、「乗る」という瞬間的な事象に対する解釈を行うので、「出発」と「到着」のどちらにも属するものであると判断した。

場所的下位分類を判定する助詞の分類を表3にまとめる。

表3 助詞型の場所的下位分類ルールセット

下位分類	助詞
出発	から, を
到着	まで, へ, に
出発, 到着	で

4 前処理

発話文解析の背景にある処理および発話文へのクリーニング処理について述べる。

4.1 場所名

発話文解析において場所を表す表現の抽出が一つの問題となる。本稿においては、場所名（店名など）が辞書に登録済みの状態で議論を進めた。

場所名が既知であることを前提にできる根拠は、本稿の解析手法がユーザ適応型の車載器での利用を想定しているためである。車載器には、駐車地点の名称を収集する機能が備わる。ユーザが話題にする場所名は蓄積される。また、スケジュール登録時に新規の場所を話題にする際に場所名の登録を行う対話を開始することが可能である。

4.2 発話文のクリーニング

入力される発話文に対して次の処理を行う。

- 漢数字や全角数字を半角アラビア数字に変換
- 時間表現「hh:mm」を「hh時 mm分」に変換

- 句読点の除去

5 実験

5.1 実験方法

分析用コーパスからタグを外したものをを用いた実験、すなわちクロードテストを行う。入力文は331文である。1文につき1つの5つ組を目標出力とする。3.1節の3つの手順で処理した結果について正答率を確認する。

5.2 実験結果

正答率は79%（正解文数263 / 入力文数331）となった。

解析例を示す。文1において「ご飯に行きます」とある。もし「ご飯」（抽象）を場所として抽出すると誤りであるが、本処理では正しく処理されていた。文2において、「明日」という表現から日付の算出ができており、「映画館」を到着場所として解析できていた。

例)

設定日時：2020年5月13日8:12

設定場所：自宅駐車場

入力文1：午後7時から友達とご飯に行きます

入力文2：明日の午前10時から映画館に行きます

す

スケジュール表：

#	日時	場所	用件
1	2020年5月13日19:00 ～***	*** ***	友達とご飯に行きます
2	2020年5月14日10:00 ～***	*** 映画館	行きます

5.3 誤り分析

誤り事例を分析した結果を表4に示す。各誤りパターンについて説明する。

誤りe1について、相対的な時間的表現とは例えば「部会后」というものがある。例を以下に示す。

例)

設定日時：2020年5月13日9:00

設定場所：自宅

入力文1：今日の夕方6時から部会があります

入力文2：部会后に先輩の家で遊びます

スケジュール表 (出力)

#	日時	場所	用件
1	2020年5月13日 18:00 ～ ***	*** ***	部会が あります
2	2020年5月13日 18:00 ～ ***	先輩の家 先輩の家	遊びます

スケジュール表 (目標)

#	日時	場所	用件
1	2020年5月13日 18:00 ～ ***	*** ***	部会が あります
2	2020年5月13日 19:00 ～ ***	先輩の家 先輩の家	遊びます

スケジュール表の出力と目標を比較すると、レコード#2の開始時刻が誤っていた。部会にかかる時間を推定する必要があるが、文外情報であるので解くことができなかった。

誤り e2 について幅のある時間的表現とは例えば「昼に」というものがある。例を以下に示す。
例)

設定日時：2020年5月13日 10:00

設定場所：自宅

入力文1：昼に友達と遊びます

スケジュール表 (出力)

#	日時	場所	用件
1	2020年5月13日 10:00 ～ ***	*** ***	友達と 遊びます

スケジュール表 (目標)

#	日時	場所	用件
1	2020年5月13日 12:00 ～ ***	*** ***	友達と 遊びます

スケジュール表の出力と目標を比較すると、開始時刻が誤っていた。「昼の1時」に対して「13時」という変換は可能だが「(単に) 昼」に対して「12時」をデフォルト値として変換することができていなかった。

誤り e3 について場所のロットフィリングに失敗が見られた。例を以下に示す。
例)

入力文1：

29時に別のバイト先に出勤します

意図 (出力)：

`serviceScheduler.comInformEvent(tmg)`

スロット (出力)：

`tmg="29時"`

意図 (目標)：

`serviceScheduler.comInformEvent(loc, tmg)`

スロット (目標)：

`tmg="29時"`

`loc="別のバイト先"`

「バイト先」は日本語語彙大系 [3] における意味属性で解析すると「抽象+場所」である。抽象を含む名詞句を「場所」として解析できていなかったため、スロットフィリングに失敗していた。

以上の誤り分析の結果、時間的表現から日時を特定する処理の不足、および、スロットフィリングに用いる意味属性のカバー範囲の調整不足という問題が確認された。しかし、時間的/場所的下位分類の判定における誤りはみられなかった。

表4 誤り事例の分類結果

#	誤りパターン	件数
e1	相対的な時間的表現	17
e2	幅のある時間的表現	23
e3	スロットフィリングの失敗	4

6 おわりに

本稿では、スケジュール登録のための発話文解析において、スケジュールを構成する情報の抽出を試みた。クローズドテストにおいて正答率79%で解析された。誤り分析の結果、日時変換、および、場所のスロットフィリングに改良すべき点がみられた。時間的/場所的下位分類の判定においては、ルールセットが良好であることが確認できた。

謝辞

本研究 JSPS 科研費 JP19K12548 の助成を受けたものです。

参考文献

- [1] 石川琢己, 徳久雅人, 木村周平: スケジュール登録のための発話文コーパスの設計, 言語処理学会年次大会講演論文集, pp.213-216, 2020.
- [2] 徳久雅人, 木村周平: 小型計算機におけるサービス指向発話文解析, 自然言語処理, 26(3), pp.545-578, 2019.
- [3] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙大系, 岩波書店, 1997.