

# スケジュール登録のための発話文コーパスの設計

石川琢己 徳久雅人 木村周平 (鳥取大学)

## 1. はじめに

本稿では、発話によりスケジュールの登録や変更を行うことを目指している。例えば、自動車の運転中に発話による用件の追加や変更を行うことで、巡回漏れの防止が可能になるだろう。スケジュールは用件の並びである。用件は日時・場所・行動で構成する。発話文からこれらの構成要素を抽出し、既存の用件の並びの中に新規の用件を追加する操作について、言語処理的な手法が必要となる。

筆者らは既にスケジュール登録のための発話文解析を作成し、応用として車載器上で使用した[1]。ここでの手法は日時や相対的な時間的表現をルールベースで解析するものである。問題点は特定の人物の言い回しに特化したことである。人々から言い回しを集計するためにコーパスの形式を定める必要がある。

コーパス構築に関する関連研究は多様に存在する。中でも旅行対話コーパス[2]では、予約タスクにおいて日時の表現が見られる。しかし、対話コーパスでは受け答えや推薦などその他の目的での発話が多く含まれている。また、コーパス検索サービス[3]において「予定です」をクエリにするとスケジュールに言及する状況の表現を収集できる。しかし、いずれもスケジュールの登録を行う連続した発話ではない。

そこで本稿では人間が計算機にスケジュールを登録するための発話を収録するコーパスの設計を行うことを目的とする。

## 2. スケジュール登録

スケジュール  $S$  は用件  $e_i$  を時系列に並べたものであり、 $S = (e_1, e_2, \dots, e_n)$  となる。用件は日時  $t_i$ 、場所  $p_i$ 、行動  $a_i$  を要素とし、 $e_i = (t_i, p_i, a_i)$  となる。

スケジュール登録のための発話文には、時間的な表現および場所を表す表現が含まれる。時間的表現は年月日時分だけでなく、発話日時からの相対的表現(例「今日のお昼に」)や話題からの相対的表現(例「その前に」)もある。相対的表現は、正確な日時  $t_i$  に換算されて、時系列による用件の列挙に用いられる。

## 3. コーパスの仕様

### 3.1 コーパスの種類とねらい

既存の用件の並びに新しく用件を追加するための発話を登録発話と呼んでいる。既存の用件について、用件の削除、用件内容の変更、および、用件の並び替えという操作があり、そのための発話を変更発話と呼んでいる。

そこで、次の2種類のコーパスを作成する：

- 登録発話文コーパス
- 変更発話文コーパス

前者は発話文における日時・場所・行動に関する表現および処理を分析するために作成する。後者は用件変更を意図する表現を分析するために作成する。

### 3.2 構成

コーパスは、3つのフィールド群および「補足」フィールドで構成する。

- メタ情報フィールド群
- 発話フィールド群
- 抽出情報フィールド群

メタ情報フィールド群には、「入力方式、発話者、発話日時、発話場所」のフィールドがある。発話フィールド群には、「原文、タグ付き発話文、修正フラグ」のフィールドがある。抽出情報フィールド群には、「日時、場所、行動略称」のフィールドがある。

「入力方式」には、「キーボード発話」および「音声発話」がある。個人適応のために「発話者」を記入する。時間的表現から日時を算出するために「発話日時」を記入する。地域の依存性(地名つき店舗名など)に対処するために「発話場所」を記入する。

入力されたままの文を「原文」に記入する。音声認識誤りや誤入力を修正した発話文にタグを付与したものを「タグ付き発話文」に記入する。修正した際に「修正フラグ」をたてる。なお、文の単位を明確にするために、発話文を「ですます調」に制限する。

一つの用件を表すために複数の発話文を要する際、発話文をセミコロンで接続して記入する。逆に、

複数の用件が一つの発話文で表される際は、用件の数だけレコードを列挙し、第一レコードに発話文および一つ目の用件を記入し、残りのレコードに残りの用件を記入する。ただし、動詞連続(シテ型, シニ型を含む)で行動が述べられ、かつ、日時場所が一意である場合には一つの用件にまとめることができる。

例 1. 家に帰って寝ます(1用件)

例 2. 8時に家に帰って11時に寝ます(2用件)

例 3. 友達と御飯を食べに飯屋に行きます(1用件)

幾つかの発話連続を通じて複数の用件を連続で登録する。そのための発話文の列をまとめて一つの談話とする。談話の開始・終了を表すマーカーとして発話文のフィールドに **START- $k$** ・**END- $k$**  をそれぞれ記入する。ここで登録発話の談話では  $k = S$ 、変更発話の談話では  $k = M$  とする。

「日時」フィールドには年月日時分を記入する。時間が不明確である場合、発話者が意図した日時を記入する。行動の開始と終了の両方の日時を記入する際は“**BGN:日時, END:日時**”とする。

「場所」フィールドに文字通りの場所や補完した正確な場所名を記入できる<sup>\*1</sup>。

例 4. 8時半に大学に行きます(場所=鳥取大学)

移動行動に関して出発と到着の両方の場所を記入する際は“**DEP:名称, ARR:名称**”とする。用件を2つに分割できない場合に用いる。

「行動略称」には時間と場所以外の事柄について用件の内容を記入する。自動的に略称を生成するための正解データとなる。応用システムにおいて、ユーザがスケジュールを瞬間的に可読であるために、略称は有効と予想している。

### 3.3 タグ

発話文に付与するタグは次のとおりである。

- **tmg**: 時間的表現
- **loc**: 場所表現
- **p**: 助詞, 助動詞あるいは相当語句
- **e**: イベントを表す語句
- **i**: 既存の用件を指し示す名詞

時間的表現は、年月日時分、「今日」、「その後、」など用件のタイミングを示す表現である。時間的区間を表す付属語を含む。**tmg** タグ内の付属語に **p** タグを付与する。例を以下に示す。

例 5. <tmg>今日<p>の</p>夕方<p>までは</p></tmg>ゼミがあります

場所表現についても同様に **loc** タグを用いる中で **p** タグを用いる。例を以下に示す。

例 6. <loc>4805 室<p>で</p></loc>ゼミがあります

イベントを表す語句は、他の用件との相対性を表すことがある。時間的表現の内部で **e** タグを付与する。例を以下に示す。

例 7. <tmg><e>晩飯</e><p>の</p>後</tmg>ボーリングをします

なお、イベントを明示しない言い回しには「その後」などがある。指示語に **e** タグを付与しない。

既存の用件を指し示す名詞に **i** タグを付与する。イベントタグの内部に付与することがある。

例 8. <tmg><e><i>ゼミ</i>に行く</e>前<p>に</p></tmg><loc>部室<p>に</p></loc>寄ります

※この発話以前に「ゼミ」が話題になっていた

### 3.4 コーパス種別による構成の違い

登録発話文コーパスでは、発話文に対して用件を検出し、抽出情報フィールド群への記入を行う。発話の順でレコードを並べる。ゆえに、用件の時間的順序が乱れることがある。

変更発話文コーパスでは、用件列に発話文を1つ作用させ、更新した用件列を記入する。これを発話文数だけ繰り返す。用件を時間的順序で整列する。

## 4. 作成結果と分析

登録発話文コーパスにおいて、発話者は10人である。発話文数は335文、談話数は137件となった。用件数は347件となった。付与したタグ数を表1にまとめる。変更発話文コーパスにおいて、発話者は

\*1 応用システムに応じて場所名の精度が異なる。ユーザに予定を表示するシステムならば場所の略称で十分である。ユーザ自身が略称から場所の実体が理解可能だからである。一方、ユーザの都合に配慮して情報推薦を行うシステムならば略称では不十分である。場所の実体を推定し、立ち寄り支援などに繋ぐことが課題であるためである。

6人である。発話文数は33文、談話数は18件となった。用件数はのべ72件となった。コーパスの一部を付録に示す。

表 1. 登録発話文コーパスにおけるタグ数

タグ種別	付与数	パターン数
tmg	317	142
loc	166	100
p	325	-
e	5	-
i	2	-

表 2. 高頻度パターン(上位 5 位まで)

時間的表現	頻度	場所的表現	頻度
夜の#時から	26	自宅で	10
#時に	23	研究室に	10
#時から	18	友達の家で	7
夜の#時に	16	研究室で	6
#:#に	11	マーレに	5

#### 4.1 時間的表現

tmg タグ内の表現について、数値を#に置き換えてパターン化を行った。パターン数は142となった。上位 5 件を表 2 (左) に示す。コロンの用いた時分の表現はキーボード入力の場合のみにみられた。なお、頻度 1 のパターンは 100 件となった。得られた時間的表現は、多くはルールベースにより年月日時分に変換できるが、対処が難しいものがあった。

一つは用件の反復を意味する表現である。下記の例 9 では用件「バイト」が 2 回生じる。ただし、期間で解釈する場合(例 10)もあり、行動の考慮を要する。

例 9. <tmg>土日<p>は</p></tmg>バイトです

例 10. <tmg>土日<p>は</p></tmg>出張です

もう一つは曖昧性である。下記の例は、終日の意味なのだが一語だけを見ると日付と解釈可能である。

例 11. <tmg>一日</tmg>バイトでつぶれます

#### 4.2 場所表現

loc タグ内の表現について字面パターンを集計すると 100 件となった。上位 5 件を表 2 (右) に示す。

場所表現の判定に間違いやすい部分があった。

例 12\*. <loc>カラオケ<p>に</p></loc>行きます

「カラオケ」を場所と判定したことが誤りである。「カラオケ屋さん」であれば場所であるが「カラオケ」はイベントである。正しく判定する方法は「をする」を後続させて自然さを見ることである。自然であればイベント、不自然であれば場所である。

例 13. カラオケをする ※ 自然なのでイベント

例 14. <loc>レストラン<p>に</p></loc>に行きます

例 15\*. レストランをします ※ 不自然なので場所  
また、悩ましいが場所扱いとなる例を以下に示す。

例 16. <loc>歯医者<p>に</p></loc>行きます

#### 4.3 行動表現

行動を略称で表現するパターンは以下があった：

- $N$  をする,  $VRY$  をする,  $N$  する,  $N$  がある,  $N$  で行く,  $N$  をしに行く,  $N$  しに行く,  $\sim$  に行って  $N$  する  
ここで,  $N$  は名詞,  $VRY$  は動詞連用形である。  $N$  や  $VRY$  の部分, および, 掛かる格要素のまとまりを行動略称とする。表 3 に具体例を示す。動詞連用形のままで自然な場合(転成名詞として存在する場合)や基本形に戻すべき場合がある。また, 基本形に戻すと不自然になる場合もある。

表 3. 行動略称の作成例

フレーズ	行動略称	変形
自宅で宿題をする	自宅で宿題	不要
友達と温泉巡りをする	友達と温泉巡り	不可
街で食べ歩きをする	街で食べ歩き	可
後輩とご飯を食べに行く	後輩とご飯を食べる	必要

#### 4.4 変更操作表現

変更操作に関して以下の種類とその表現を得た。

- 削除
- 時間, 場所, 行動の変更
- 順序変更

時間の変更は, 直接的に時分を新たに指定する言い方である。その結果, 用件の時間的順序関係に変化が生じることがある。順序変更は 2 つの用件の間で順序を変えることを主とした言い方である。抽出すべき時分が不明確になりやすい。付録 2 の例では「ゼミより前」の時分が不明である。

変更操作の最中(談話の途中)では, 用件間の矛盾が生じやすい。付録 2 の例では, 「ゼミに行く」用件のみに時間を変更した。学生の立場では, ゼミに遅刻することを予定に入れたものとなる。逆に教員の立場であれば「ゼミを行う」用件の時間も同時に変更されるべきだろう。応用システムで発話を受けて変更処理を行う際は, 最小の変更を行い, ユーザからの追加命令を介して漸進的に変更を行うべきだろう。

## 4.5 その他の表現

行動の理由が付記されることがある。用件を選択する、あるいは避ける理由は、応用システムが立ち寄り先の推薦を行う上での手がかりとなる。

例 17. 日曜日はポイント 10 倍なのでエスマートに行きます

例 18. ■(某飲食店名伏せ字) ■は好きじゃなかった  
ので、ビアドパパに行きます

## 4.6 時間の抽出

抽出フィールド群における「時間」フィールドへの記入において、時間的表現および行動内容の両方に注目する必要があった。

例 19. お昼に友達とご飯を食べます(12時)

例 20. お昼からゼミがあります(13時)

時間的表現が「お昼に／から」であるが、行動が「ご飯を食べる」ならば昼休みの 12 時を、行動が「ゼミ」ならば昼休み後の 13 時をそれぞれ用件の時間として解釈する必要がある。個人について蓄積したスケジュールに基づき、個人に適応した時間の推論を行う必要がある。

## 5. おわりに

本稿では、スケジュールの登録および変更を行うための発話文を収録するコーパスについて設計を行った。応用システムの開発に向けて準備すべきフィールド・レコード・タグ・抽出情報を定義した。コーパスの試作を通じてシステム開発に有益な分析が可能であることを確認した。

### 謝辞

本研究は JSPS 科研費 JP19K12548 の助成を受けたものです。

### 参考文献

[1] 徳久雅人, 木村周平: 車載器に旅行計画を登録するための発話文解析, 言語処理学会第 24 回年次大会 発表論文集, pp.532-535, 2018.

[2] ATR 音声言語データベース

<https://www.atr-p.com/products/sdb.html>

[3] 国立国語研究所コーパス開発センター

[https://pj.ninjal.ac.jp/corpus\\_center/](https://pj.ninjal.ac.jp/corpus_center/)

## 付録 1 スケジュール登録発話文コーパスの一部

メタ情報フィールド群

入力方式	発話者	発話日時	発話場所
キーボード	A者	2019/7/29 23:43	自宅

発話フィールド(一部)および抽出情報フィールド群

タグ付き発話文	日時	場所	行動略称
START-S			
<tmg>明日<p>の</P> 朝10時<p>から</p></ tmg><loc>研究室<p>で </p></loc>ゼミがあり ます	2019/7/30 10:00	研究室	ゼミ
<tmg>その後<p>、</p> </tmg>友達とお昼ごは んに行きます	2019/7/30 12:00	マーレ	友達とお 昼ごはん
END-S			

## 付録 2 スケジュール変更発話文コーパスの一部

メタ情報フィールド群

入力方式	発話者	発話日時	発話場所
キーボード	A者	2019/7/9 8:43	自宅

発話フィールド(一部)および抽出情報フィールド群

タグ付き発話文	日時	場所	行動略称
START-S			
<tmg>今日<p>の</p> 10時半<p>に</p></ tmg>ゼミがあるので< loc>研究室<p>に</p> </loc>行きます	2019/7/9 10:30	研究室	行く
	2019/7/9 10:30	研究室	ゼミ
<tmg>そしてお昼<p>に< /p>サークルに行きます	2019/7/9 14:00	部室	サークル
END-S			
START-M			
<tmg><e><i>ゼミ</i> </e><p>より</p>前<p> >に</p></tmg><i>サー クル</i>に行きます			
	2019/7/9 9:00	部室	サークル
	2019/7/9 10:30	研究室	行く
	2019/7/9 10:30	研究室	ゼミ
<tmg>10時半</tmg>で はなく<tmg>11時<p>に </p></tmg><i>ゼミ</ i>に行きます			
	2019/7/ 9:00	部室	サークル
	2019/7/9 10:30	研究室	ゼミ
	2019/7/9 11:00	研究室	行く
END-M			