

簡略表現を認識し応答に利用できる音声対話システム

秋田谷 樹[†]駒谷 和範[†]佐藤 理史[†]中野 幹生[‡]
[†]名古屋大学大学院 工学研究科 [‡]ホンダ・リサーチ・インスティテュート・ジャパン (HRI-JP)

1. はじめに

人間同士の対話において、長い名称について話すとき、話者は名称の一部を省略した表現（簡略表現）をしばしば使用する。また、簡略表現が使用された対話では、話者間の表現の共有が起きる。このような話者間の表現の共有化は語彙の引き込みと呼ばれる [1]。我々は、この簡略表現における引き込みに注目する。このような対話は、あるトピックに対して複数の表現を使い分ける対話の一例であり、意思疎通の齟齬の一因となる。

我々は簡略表現を認識する機能や応答に使用する機能を持った対話システムを構築する。本研究で実現する簡略表現を用いた対話の例を表1に示す。このシステムは、名古屋市のコンビニ・レストランに対するデータベース (DB) を検索する。このような対話の実現には、簡略表現の認識と、応答生成時の表現の選択が必要となる。そこで、簡略表現の認識においては、文献 [2] で得られた知見をもとに、生成規則に基づく簡略表現の候補生成を行い、この結果を用いて音声認識用言語モデルを拡張した。また、応答生成においては、ユーザが使用した簡略表現をシステムが使用することで、システム側の引き込みを実現する機能や、システムが簡略表現を用いることで、ユーザに引きこみを促す機能を実現する。

本稿では、まず対話システムにおける簡略表現の取り扱いについて述べる。その後、対話システムの概要とそれを構成する各要素について順次説明する。

2. 簡略表現の理解と使用

対話の中で簡略表現をどのように利用するかを述べる。表1の対話は名古屋市内のレストラン検索を想定した対話であり、ユーザが金山駅周辺のマクドナルドを探して、席の数を聞くものである。システムは、U2で発話された”マック”という単語が、S2の後半部の”マクドナルド金山店”を指していると理解して検索を実行する。また、応答時にはユーザが使用した”マック”という単語を用いて応答している。

2.1 簡略表現の理解

システムが簡略表現を正しく理解するには、正しく音声認識を行い、簡略表現の元となる語（元の表現）を同定する必要がある。そこで、簡略表現の生成規則による自動生成結果を利用する。自動生成には文献 [2] において、実例の調査から得た簡略表現の生成規則とその出現頻度を使用する。図1に”セブンイレブン”から簡略表現を生成する例を示す。生成規則の”1形態素抽出”は、”セブン”のように複数形態素語から1形態素を抽出する規則、”2モーラずつ抽出”は”セブイレ”のように複数形態素語の2つの形態素から、語頭の2モーラずつを抽出する規則である。スコアは、規則の出現頻度を付与したもので、その規則の適用されやすさを表す。

表 1: 対話例 (S: システム, U: ユーザ)

ID	発話
S1	検索条件を教えてください
U1	金山駅周辺のファーストフードを教えてください
S2	4件見つかりました マクドナルド金山店, ロッテリア名鉄金山駅店, ...; です
U2	マックについて詳しく知りたい
S3	マックについてどのような情報を知りたいですか
U3	席の数を教えてください
S4	マックの席の数は110席です

元の表現(店名)	簡略表現の候補	生成規則	スコア
セブンイレブン	セブン	1形態素抽出	0.33
	イレブン	1形態素抽出	0.33
	セブイレ	2モーラずつ抽出	0.54
	セブンイレブン	原形	1.00

図 1: 簡略表現の生成例



図 2: 応答生成チャート

簡略表現の音声認識には、生成された簡略表現の候補を使用する。簡略表現の候補は、1つの単語に対して複数出力されるため、語彙の増大による音声認識精度の劣化が懸念される。本システムでは語彙の増加を抑制するため、候補の中から使用されやすい表現のみを手で選択し、辞書に追加することで認識精度を向上させる。これに加え、生成規則では得られないものの、一般によく使われる簡略表現を手で追加した。

元の表現の同定には簡略表現と元の表現の対応やスコアを使用する。例えば”セブン”という単語が入力された場合、元の表現の候補にはコンビニの”セブンイレブン”以外にも、名古屋市内に実在する喫茶店の”セブン”がある。このように元の表現の選択肢が複数得られる場合は、スコアが高い喫茶店の”セブン”を選択する。

2.2 簡略表現を用いた応答

応答の生成は、図2に従って遷移する。初期状態では、元の表現をシステムが持つ簡略表現に言い換える。ユーザから店名が入力された場合、元の表現あるいは簡略表現のどちらであったかに関わらずユーザが使用した表現を使用するものとした。前者の機能は、生成した簡略表現から行い、ユーザに引きこみを促す。後者の機能は、システム側の引き込みを実現する。

Spoken Dialogue System that Recognizes and Responds with Abbreviated Expressions: Tatsuki Akitaya, Kazunori Komatani, Satoshi Sato (Nagoya Univ.) and Mikio Nakano (Honda Research Institute Japan Co., Ltd.)

3. システムの実装

名古屋市内のコンビニ・レストランを検索する対話システムを作成した。本システムはまず、ユーザからの検索条件を入力として受け取り、候補を絞り込む。このときシステムは検索結果の件数や店名を出力する。次に、この出力された店舗の中から、ユーザは任意の店を選択し、具体的な情報を引き出したり情報を比較したりできる。システムの概略図を図3に示す。以降ではシステムを構成する要素について述べる。

3.1 検索対象 DB

DBには、簡略表現が出現しやすいドメインとして、コンビニ、ファーストフード、ファミレス、カフェに該当する店舗情報を人手で収集した。収集可能な店舗情報の違いから、コンビニとそれ以外のカテゴリを、別々のDBとした。

コンビニのDBには、住所、電話番号、ATMや駐車場の有無などを記述した19フィールドを作成した。このDBは519件のレコードを持つ。コンビニ以外のジャンルのDBには、モーニングの有無や座席数の情報などを記述した24フィールドを作成した。このDBはいずれかのジャンルに該当する2549件のレコードを持つ。DB管理ソフトウェアにはH2 Databaseを使用し、PostgreSQLによってDBの検索結果を取得できる。

3.2 音声認識・言語理解

音声認識にはJuliusを使用する。言語モデルはクラスN-gramモデルを使用し、Yahoo知恵袋の料理・グルメ・レシピカテゴリから取得したテキスト(441,872文)と想定発話言語モデルから生成したテキスト(13,240文)から作成した。このモデルでは検索用のDBと対応した名称(shop)、支店名(branch)、最寄り駅(station)、ジャンル(genre)と食べ物(food)をクラス単語として認識する。

DB中の店名に対して、ブランドと支店名をそれぞれshopクラス、branchクラスとして分離しそれぞれのクラス内で簡略表現を取り扱う。店名には”ファミリーマート名古屋大学店”のように、ブランドと支店名がセットとなっているものが多い。このような店名は前後のどちらか一方が省略される場合が多い。ここで、”名古屋大学店”の元の表現が”ファミリーマート名古屋大学店”であるとした場合、同一の支店名を持つ店舗が複数存在すると考えられるため、元の表現を一意に同定すべきではない。クラスを分割することにより、”ファミマ名古屋大学店”のように片方のみ省略する表現も認識可能となるようにした。

音声認識辞書のshopクラスには、自動生成された例の中から選択した単語に、”マクドナルド”の簡略表現である”マック”など、頻繁に使用されるものの自動生成できない例を763語追加した。また、branchクラスには”名古屋大学店”の簡略表現である”名大店”を始めとした438語を追加した。拡張後の辞書のサイズは、shopクラスが3828語、branchクラスは1358語となった。

Juliusは、辞書中の各単語に対して、音声認識に使用しない文字列を記述することができる。そこで認識辞書内に、元の表現の同定に使用する元の表現とスコアの情報を記述し、言語理解部では、音声認識結果から検索要求作成に必要な内容語を抽出する。ここでshop、branchクラスは同一のフィールドに対応する内容語として取り扱う。ユーザが使用した表現および元の表現がこれに対応する。

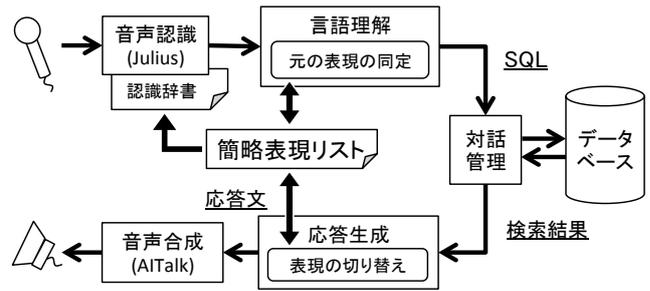


図3: 対話システムの概略図

3.3 対話管理

対話管理部では言語理解部で取得した内容語から対話状態の決定を行なう。システムの状態は大きく検索条件の追加 (search) と情報の提示要求 (info) の2つに分かれる [3]。searchは必須であるgenreとstationフィールドと取得するための状態 (getGenre)、(getStation) とその他条件の追加 (getCondition) の3状態に細分化される。また、infoには1件の情報を提示する状態 (tellInfo) と2件の店舗を比較する状態 (compareInfo) が存在する。

システムの初期状態はsearchであり、検索が4件以下になるまでこの状態を継続する。検索条件が追加され、検索結果が4件以下になると、システムは店名の候補を読み上げる。ここで店名に該当する内容語が入力されたとき、infoに移行する。このとき発話された店名が1件ならtellInfo、2件ならcompareInfoへ移行する。

DBからの情報検索は、SQLによって行われる。SQLは言語理解部で取得した内容語から生成され、条件に合致する店舗情報を取得できる。金山駅周辺のコンビニを検索する場合の生成例を以下に示す。

```
SELECT * FROM database WHERE station LIKE '金山' AND genre LIKE 'コンビニ'
```

3.4 応答生成

生成内容に店名が含まれる場合、簡略表現への言い換えを行うかどうかの決定に基づき応答を生成する。簡略表現への言い換えは、”マクドナルド金山店”の場合、同じブランドを持つ候補が複数なければ”マクドナルド”を使用し、ブランドが重複する場合は”金山店”のように支店名を使用するものとした。

4. 今後の展開

作成したシステムを使用した被験者実験を行い、被験者の簡略表現の引き込みの有無や、簡略表現を用いた対話の有効性について検証する。具体的には、システムが使用した表現を、ユーザが実際に使用したかどうかや、システムが同じ簡略表現を使った場合に、ユーザが好感をもったかを調査する。

参考文献

- [1] Brennan, S. E.: Lexical entrainment in spontaneous dialog, *Proc.ISSD-96* (1996).
- [2] 秋田谷樹, 駒谷和範, 佐藤理史, 中野幹生: 音声対話システムのための簡略表現の生成規則と出現頻度の調査, *情報処理学会全国大会講演論文集*, Vol. 75,6T-6,(2013).
- [3] 西村良太, 駒谷和範: データベース検索音声対話システムにおける対話状態の推定, *情報処理学会研究報告*, Vol. 2012-SLP-90, No. 20 (2012).