

## 入力文の意味理解と連想に基づく自然会話応答

田中亜紀子<sup>†</sup> 吉村枝里子<sup>†</sup> 渡部広一<sup>†</sup> 河岡司<sup>†</sup>  
<sup>†</sup>同志社大学大学院 工学研究科

### 1. はじめに

従来の会話処理では、人間のように会話の中に連想を取り入れた応答はあまり行われていない。本稿では、知的オウム返しなどの従来の手法に加え、連想を取り入れた、新たな応答処理手法として「連想応答」を提案する。具体的には、会話入力文に使われている語に対する種々の連想や入力文の意味理解に基づき、会話の発展性を考慮した応答文を生成する「動詞場所連想」方式と「応答文連想選択」方式の提案を行う。前者は、「『映画館』へ「行く」」など「場所名詞」と「場所動詞」を用いた場所に関する連想から新たに得られる動詞「観る」を用いた応答文を生成する。後者は、目的語と動詞で捉えた入力文の意味とデータベースに準備した代表応答文の意味的関連の強さを連想により評価し、最も適切な応答文の抽出を行う。

### 2. 連想メカニズムと常識判断メカニズム

本研究では、連想メカニズムとして、概念ベース、関連度計算、概念ベース IDF、大規模格フレームを利用している。概念ベースとは語(概念)の特徴を表す語(属性)を集めたものであり、属性には重要度(重み)が定義されている。概念ベースには概念が約 9 万語格納されている。また、関連度計算<sup>[1]</sup>とは、2 つの概念間の関連の強さを定量的に評価するものであり、各概念を二次属性まで展開し、重みも考慮した属性集合の一致度を計算している。概念ベース IDF は、概念ベース内の全概念を検索対象とし、二次属性まで展開した属性全てを母集団としている。大規模格フレーム<sup>[2]</sup>は、Web 上の約 5 億言の日本語テキストから自動的に構築されており、約 5 万用言からなるものである。また、常識判断メカニズムとして、場所、感覚、意味理解の判断システム<sup>[3][4][5]</sup>を用いている。

### 3. 従来の応答文生成法

肯定の情報文から質問応答文を生成する手法として 3 種類挙げられる。6W1H 法は、情報フレーム内の空白部分に対して質問文を生成する方法である。また、具体化・詳細化法は、入力文に対し具体的に問う必要がある場合に適用する方法で、目的語がシソーラスにおいて「名詞・具体」以下の場合に用いる。知的オウム返し法は、連想メカニズムと常識判断メカニズムを利用したオウム返しである。

### 4. 「連想応答」による応答文生成法

「6W1H 法」、「具体化・詳細化法」は、入力文中の語を直接用いた応答であり、「知的オウム返し」は、新たな語を連想しているが、応答の文型がパターン化されている。バリエーション豊富な応答ができるよう、入力文から連想を行うこと(連想応答)が必要となる。「動詞場所連想」と「応答文連想選択」の 2 手法を提案する。

#### 4.1 動詞場所連想

動詞場所連想は、「『映画館』へ「行く」」など「場所に関する名詞」と「場所動詞」の 2 語を用いて、場所に関する 2 語による連想を行い、新たな動詞「観る」を連想し、テンプレートに入れ応答文を生成するという連想応答である。

#### 4.1.1 場所動詞データベース

入力文中に場所に関する名詞を目的語にとる動詞を含むかの判別の際に用いるため、場所動詞 DB を作成した(表 1)。ここで、場所に関する名詞とは、「遊園地」、「郵便局」のようにその名詞単独で、そこには何が存在し、何をやる場所かが連想可能となる語と定義する。つまり、シソーラスに存在する上記の条件を満たすノード「場所」・「家屋」以下に属する全ての語とする。

二格とヲ格の目的格をとり、かつ、動詞の目的語に「場所に関する名詞」を入れた際、その文が適切な文となる場合、その動詞を場所動詞 DB に含めるとする。例えば、「行く」「訪れる」は DB に含めるが、「食べる」「読む」は含めない。

表 1. 場所動詞データベース

「行く」に関する動詞	行く, 訪れる, 訪問する, 出発する, 出かける, 連れて行く, 飛び立つ
「帰る」に関する動詞	帰る, 帰宅する, 戻る
その他	近づく, 入る, 来る

#### 4.1.2 動詞場所連想の手法

入力文を形態素解析して自立語に分け、品詞が「名詞」である語を場所判断システムにかけ「場所に関する名詞」であるかを判別。品詞が「動詞」である語を基本形にし、その語が場所動詞データベースに格納されている語であるかを判別。共に条件を満たす場合のみ を実行し、それ以外は「動詞場所連想不可能」とする。

取得した「場所に関する名詞」を場所判断システムに渡し、場所の目的を表す動詞及び名詞を取得する。

入力文にない新たな動詞を取得するため、得られた語から、「動詞」と「サ変接続の名詞」の語を選択する。

選択した語の概念ベース idf 値を調べ、頻繁に使用される語を得るため、概念ベース idf 値の最も低い語を選択。

選択した語を語尾変換し条件に合うテンプレートに入れる。

### 4.2 「動詞場所連想」の評価実験

#### 4.2.1 動詞場所連想の連想語の精度評価

選択した場所の目的を表す動詞及び名詞が、入力文に対する応答文に用いる語として適切であるかを目視によって評価した。評価データは、中学英語テキストより、「場所に関する名詞」と「場所動詞 DB に存在する場所動詞」を共に含む文 100 文を選択した。3 段階評価( , , ×)を行い、常識度( と の割合)を求めた結果、常識度が 74%となった。

#### 4.2.2 動詞場所連想の応答文の精度評価

選択した場所の目的を表す動詞及び名詞を応答テンプレートに入れて応答文を出力した際、入力文に対して、出力した応答文が適切であるかを評価した。評価データは 4.2.1 節と同じものを用いた。この結果、常識度が 60%となった。

#### 4.2.3 「動詞場所連想」の考察

「連想語の精度評価」について述べる。「私は昨日歯科に行った」という入力文に対し、「治療」という連想語が出力された場合「治療してきたのですか?」のような適切な応答文が考えられる(成功例)。これに対し、「彼女は京都の名所を訪れた」という入力文に対し「住む」が出力された場合、「住む

のですか?」のような不適切な応答文が生成される。このように、場所に関する名詞のみを場所連想するだけでは、柔軟な結果が得られない場合がある。入力文中の場所に関する名詞を一次属性、二次属性にまで拡張し、その後場所連想を行うことで、さらに柔軟な連想が可能となると考えられる。

### 4.3 応答文連想選択

入力文中の「場所動詞以外の動詞」と「目的語」の2語で捉えた入力文の意味と、DB に準備した定型応答文の意味的関連の強さを連想により評価し、最も適切な応答文を抽出する連想応答である。

#### 4.3.1 作成データベース

「代表動詞 DB」に格納する動詞は、場所動詞以外の動詞全てを対象とする。代表動詞 DB には目的語と動詞、情報フレーム<sup>5)</sup>の種類の設定でパターン化して格納する。

目的語に関係なく常に一意の応答文を返す動詞については、目的語と情報フレームの種類を「\*」としておく。以降、目的語の代表的なシソーラスノードを「代表語」と呼ぶ。格納例を表2に示す。表2の代表語は、基本的にシソーラスノードを指標として格納している。格納方法としては、「大規模格フレーム」を用い、各動詞の各フレームの検索結果の上位50件を用いて、シソーラスノードを定め、表2の「代表語」としている。「定型応答文 DB」は各動詞と応答文を対応させたものであり、格納例を表3に示す。

表2. 代表動詞 DB の格納例

ID	フレームの種類	代表語	動詞	応答
1	Who	推量	当たる	7, 8
2	What, Whom	役割, 仕事, 調査	当たる	3
3	What	くじ, 贈り物	当たる	1, 6
4	What	食料	あたる	4
5	*	*	遊ぶ	10
...	...	...	...	...

表3. 定型応答文 DB の格納例

ID	応答文
1	おめでとう!
2	痛かったですか?
3	頑張ってください。
4	大丈夫ですか?
...	...

#### 4.3.2 応答文連想選択の手法

##### 【動詞検索処理】のアルゴリズム

入力文を意味理解システムにかけ、動詞を得る。用言フレームの動詞が「代表動詞 DB」に存在するかを判別し、存在する場合「応答文生成処理」を行う。

##### 【応答文生成処理】のアルゴリズム

代表動詞 DB の中で、取り出した動詞と対応する「フレームの種類」が「\*」であるか判別。「\*」の場合、代表動詞 DB により、動詞に対応する ID を取り出し、定型応答文 DB において、取り出した ID の応答文を出力。「\*」でない場合は の処理を行う。

で判別した「フレームの種類」が、「動詞検索処理」で意味理解システムにかけて語の格納された情報フレームの種類と一致するかを判別する。一致しない場合「応答不可能」とし、一致する場合は の処理へ進む。

で一致した情報フレームに格納された語(A)を取り出す。(A)の「シソーラスの全親」中に代表動詞 DB 中の代表語(B)が存在するか判別。存在する場合、(B)に対応する ID を定型応答文 DB で検索し、応答文を出力。存在しない場合は、 の処理へ。

代表動詞 DB 中の(B)と(A)との関連度計算を行い、閾値 0.202 以上かを判別する。閾値以上なら、代表動詞 DB 中の(B)に対応する ID を定型応答文 DB で検索し、応答文を出力する。閾値以上でない場合「応答不可能」とする。

### 4.4 「応答文連想選択」の評価実験

#### (1) 評価方法

人間が定型応答文で応答可能と考える文を入力し、代表動詞 DB 中の正しいレコードに合致したかを目視によって評価した。代表動詞 DB の正しいレコードに合致し適切な応答文が出力された入力文は とし常識度( の割合)を求めた。データは、WEB と中学英語テキストより、作成 DB 内の動詞をもつ単文 230 文を用いた。

#### (2) 評価結果

精度については、常識度が 72% となった。

#### (3) 考察

「学級の委員長に当たった」という入力文に対し、「頑張ってください」という適切な応答文が出力された(成功例)。これに対し、「葬式が始まった」という入力文に対しては、「楽しんでください!」という不適切な応答文が出力された。これは、「葬式」のシソーラスの全親の「行事」により、「コンサートが始まった」に対する応答文と同じ文が出力される結果となった。同じ「シソーラスの親」をもつ語であっても、応答文が異なる場合の対処法が必要になる。具体的には、「行事」の複数の具体リーフ(コンサート、祭り、葬式、結婚式など)と「葬式」との関連度計算により、入力文の語に最も関連のある語を検索することで解決可能と考えられる。

### 5. おわりに

本研究では、入力文に対し意味理解と連想を伴った「連想応答」の方式を提案した。従来の応答文生成法に「連想応答」を加え、今後、拡張された応答候補文から適切な文の選択を行うことで、より自然な応答文生成の実現につながっていくことが期待される。

#### 謝辞

本研究は、文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクトにおける研究の一環として行ったものである。

#### 参考文献

- [1] 渡部 広一, 河岡 司, “ 常識的判断のための概念間の関連度評価モデル”, 自然言語処理, Vol. 8, No. 2, pp. 39-54, 2001.
- [2] 河原大輔, 黒橋禎夫, “ 高性能計算環境を用いた Web からの大規模格フレーム構築”, 情報処理学会 自然言語処理研究会 171-12, pp.67-73, 2006.
- [3] 杉本二郎, 渡部 広一, 河岡 司, “ 概念ベースを用いた常識場所判断システムの構築”, 情報処理学会自然言語処理研究会資料, 2003-NL-153, pp.81-88, 2003.
- [4] 渡部 広一, 堀口 敦史, 河岡 司, “ 常識的感覚判断システムにおける名詞からの感覚想起手法”, 人工知能学会論文誌, Vol. 19, No. 2, pp.73-82, 2004.
- [5] 篠原直道, 渡部 広一, 河岡 司, “ 常識判断に基づく会話意味理解方式”, 情報処理学会自然言語処理研究会資料, 2003-NL-153, pp.89-96, 2003.