

常識判断に基づく会話意味理解方式

篠原 宜道 渡部 広一 河岡 司

同志社大学大学院 工学研究科

dtb0730@mail4.doshisha.ac.jp, {watabe, kawaoka}@indy.doshisha.ac.jp

1. はじめに

コンピュータを人間にとってより使いやすいものにするには、人間同士が日常行っているコミュニケーション（意思疎通）の仕組みをモデル化し、コンピュータと人間とのインタフェースに、これを取り入れることが望まれる。その基本要素技術として、情報の意味を理解するメカニズムの研究が近年盛んに行われている。

本稿では、コンピュータにおける理解を、推測（連想）することと捉えている。これは、人間における理解とは必ずしも一致していない。人間は子供の頃から眺めたり触れたりという五感を通じた経験によって理解している。しかし、五感を備えた身体をもたないコンピュータが、このように理解することは不可能である。そこで、概念ベースを用いた連想処理を行うことにより、人間の理解に近い知的処理を行う必要がある。

また、コンピュータにおける会話意味理解とは、話しかけられていることの内容を把握し、適切な応答をかえすことをいう。そのためには、会話文の内容を整理し連想する必要がある。

そこで本稿では、6W1H (Who, What, When, Where, Why, Whom, How) + 用言フレームを用いて、会話文の内容を整理し、WhenとWhereに関しては、連想処理を行うことにより、コンピュータに理解させる手法を提案する。

2. 会話意味理解メカニズム

2.1 意味理解

本稿で述べる意味理解とは、置換処理や連想処理を行うことによって会話文の内容を図1のようなフレーム（以下個別知識ベースと記す）に格納し、情報を整理することをいう。但し、本稿で扱う会話文は、複文や重文を含まない文とする。

(ex)昨日店で母が妹にセーターを買った			
情報フレーム		連想フレーム	
Who	母	When	冬
What	セーター	Where	存在：店長,客 目的：買う,売る
When	昨日		
Where	店		
Why			
How			
Whom	妹		
用言	買う		

図1. 個別知識ベース

2.2 フレームの必要性

会話文の内容を整理するのは、適切な応答をかえすためである。よって、問い掛けに対して、返答しやすい形で情報を整理する必要がある。そこで、質問に対する応答が可能になるように疑問詞5W1Hでフレームを構成した。また、目的格（誰に）にも対応できるよう「Whom」をフレーム項目に追加した。これにより、「誰に何をしたか」という質問に対して、応答が可能になる（図2）。

(質問) 誰にセーターを買ったのですか？			
(応答) 妹に買いました			
情報フレーム		連想フレーム	
Who	母	When	冬
What	セーター	Where	存在：店長,客 目的：買う,売る
When	昨日		
Where	店		
Why			
How			
Whom	妹		
用言	買う		

図2. Whom フレームを用いた応答

2.3 各種判断システム

提案する手法では、場所判断システム^[2]・時間判断システム^[3]を利用する。以下ではこれらについて述べる。

2.3.1 場所判断システム

場所判断システムとは、会話で使われた言葉が場所に関係する語かどうかを判断し、場所に関する語と判断された場合は、そこに何が存在し、何をやる場所なのかを連想するシステムのことである。具体的には、事前に人手で用意した小規模の場所に関する知識および国語辞書等から自動的に構築した大規模概念ベースとシソーラス^[4]を用いて、知識にない言葉に関しても場所に関する情報を連想可能としている（図3）。

なお、シソーラスとは一般名詞の意味的用法を表す2700の意味属性（ノード）の上位下位関係、全体部分関係が木構造で示されたものであり、約13万語（リーフ）が登録されている。また、概念ベースとは語（概念）を関連する語の集合で表現したもので、約9万の語（概念）が格納されている（表1）。

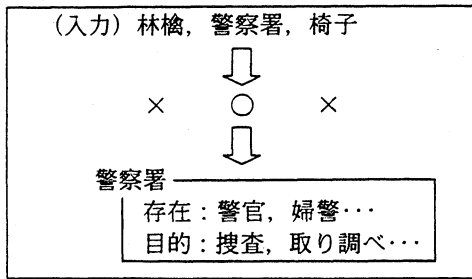


図3. 場所判断システムの例

表1. 概念ベースの例

概念	属性
警察署	役所, 警察官, 安全, 犯罪, …
林檎	果物, 実る, 丸い, 赤い, …

2.3.2 時間判断システム

時間判断システムとは、日常的な表現から時間／季節を得るシステムである。特に、用言の知識ベースを持たずに、名詞と用言の組合せから時間／季節を連想するという特徴をもっている。

具体的には、事前に人手で用意した小規模の時間に関する知識および概念ベースとシソーラスを用いて、知識にない言葉に関しても時間に関する情報を連想可能としている(図4・図5)。

また、このシステムにおいては時間に関する語を以下のように大きく2つに分類して扱っている。

- ①「クリスマス」や「夏」のように日付や季節等を指す言葉
- ②「サンタクロース」や「海水浴」のように時間を指す言葉ではないが、暗黙に時間を指す言葉以下、①を明示的時語、②を暗示的時語と呼ぶ。

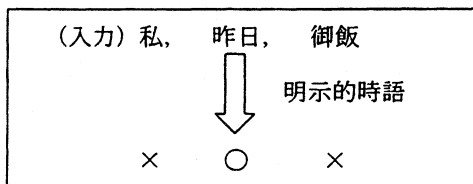


図4. 時間判断システムの例1

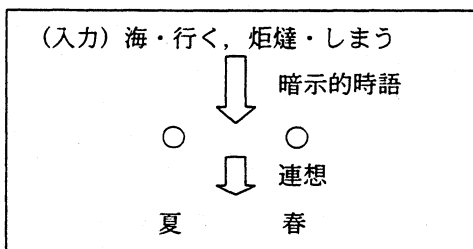


図5. 時間判断システムの例2

2.4 知識ベース

2.4.1 方向語知識ベース

方向と関係している語の中で代表となる語(14語)を集めたものである(表2)。

2.4.2 時間知識ベース

「1999年」や「2月」のように数字と組み合わせられる時間の単位(20語)を集めたものである(表3)。

表2. 方向語知識ベースの一部

代表方向語ID	代表方向語
4	真右
6	真左
7	右斜め前
8	右斜め後ろ

表3. 時間知識ベースの一部

ID	時間軸
1	世紀
2	年
3	月
4	年月

2.5 アルゴリズム

図1のような個別知識ベースを作成するために、形態素解析・日本語国文法・助詞・判断システム(場所・時間判断システム)・シソーラスを用いて、フレームごとの格納ルールを作成した。以下では、このルールについて述べる。なお、ルールの適用順序を図8に示す。

2.5.1 情報フレーム

(1) Who フレーム

このフレームには、文章中で主体語になるものを格納する。以下に処理される順通りルールを示すが、「Who」フレームに主体語が格納された時点で残りのルールは適用しないものとする。

- ① 文章中において、助詞「が・は・も」をとともなう名詞を「Who」フレーム候補として取り出す。但し「には・とは・では」の場合は取り出さない。
- ② 「Who」フレーム候補が1個しかない場合は、それを「Who」フレームに格納する。但し、倒置への対応として、文章が「の」+「が・は・も」の場合は、「が・は・も」の後に続く名詞を格納する(図6)。

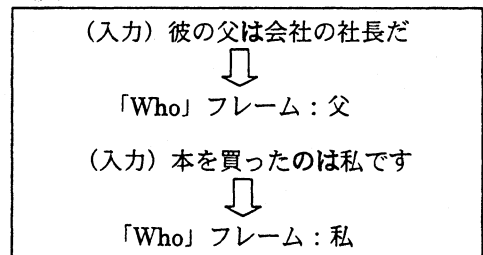


図6. ルール①②の例

- ③ 形態素解析による語の品詞情報に優先順位を設け、最も優先された語を格納する(図7)。

代名詞 > 固有名詞 > 一般 > サ変接続

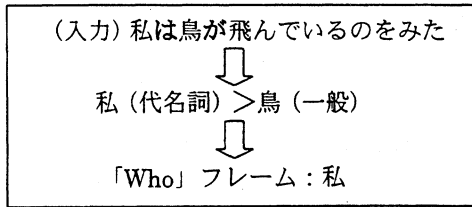


図7. ルール③の例

④ 優先順序が同じ場合は文章中で先にある方を格納する。

(2) Where フレーム

このフレームには、文章中で場所を示すものを格納する。以下にルールを示すが、この場合も「Where」フレームに場所が格納された時点で残りのルールは適用しないものとする。

① 文章中において、助詞「を」「に」「へ」「で」と副助詞「まで」「から」をともなう名詞を「Where」フレーム候補として取り出す。

② 取り出された「Where」フレーム候補が場所判断システムで場所と判断されるか調べ、場所と判断されたものを「Where」フレームに格納する。

③ 取り出された「Where」フレーム候補が方向語知識ベースに登録されているか調べ、登録されていたものを「Where」フレームに格納する。

④ 取り出された「Where」フレーム候補がシソーラスのノード「位置」「方角」「遠近」「地域」のリーフに属するかどうか調べ、属していたものを「Where」フレームに格納する。

(3) When フレーム

このフレームには、明示的時語を格納する。以下にルールを示すが、この場合も「When」フレームに明示的時語が格納された時点で残りのルールは適用しないものとする。

① 文章中において、助詞「を」「に」「で」をともなう名詞を「When」フレーム候補として取り出す。

② 取り出された「When」フレーム候補が時間判断システムで明示的時語と判断されるか調べ、明示的時語と判断されたものを「When」フレームに格納する。

③ 取り出された「When」フレーム候補が時間知識ベースに登録されているか調べ、登録されていたものを「When」フレームに格納する。

④ ①から③の処理を行って、「When」フレームに格納できなかった場合、文章中の全ての名詞を「When」フレーム候補とし、もう一度上記の処理を行う。

(4) How フレーム

このフレームには、文章中で手段を示すものを

格納する。以下にルールを示すが、この場合も「How」フレームに手段が格納された時点で残りのルールは適用しないものとする。

① 文章中において、助詞「で」をともなう名詞がシソーラスのノード「乗り物 (本体)」のリーフに属していれば「How」フレームに格納する。

② 文章中において、助詞「で」をともなう名詞が場所・時間フレームに格納されておらず、なお且つ、シソーラスのノード「具体」のリーフに属していれば、「How」フレームに格納する。

(5) Why フレーム

このフレームには、文章中で理由を示すものを格納する。以下にルールを示す。

① 文章中において、助詞「で」をともなう名詞がWhere・When フレームに格納されていない場合、「Why」フレームに格納する。

(6) Whom フレーム

このフレームには、目的格 (誰に) を示すものを格納する。以下にルールを示す。

① 文章中において、助詞「に」をともなう名詞がシソーラスのノード「人間」のリーフに属していれば、「Whom」フレームに格納する。

(7) What フレーム

このフレームには、文章中で「何」を示すものを格納する。以下にルールを示すが、この場合も「What」フレームに何が格納された時点で残りのルールは適用しないものとする。

① Who・Where・When・How・Why・Whom フレームに格納されていない名詞が文章中において、助詞「を」「に」「が」「は」をともなっている場合、「What」フレームに格納する。

② Who・Where・When・How・Why・Whom フレームに格納されていない名詞がシソーラスのノード「乗り物 (本体)」に属していれば、「What」フレームに格納する。

(8) 用言フレーム

このフレームには、文章中で動作・状態を表す語を格納する。以下にルールを示すが、この場合も「用言」フレームに動作・状態が格納された時点で残りのルールは適用しないものとする。

① 形態素解析結果より、動詞と判断されたものを「用言」フレームに格納する。

② 形態素解析結果より、文末が形容詞もしくは形容動詞と判断された場合、それを「用言」フレームに格納する。

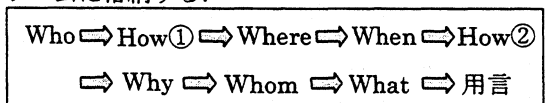


図8. 各ルールの適用順序

2.5.2 連想フレーム

次に、場所・時間判断システムを用いた連想フレームへの格納方法について述べる。

(1) 場所連想

場所連想とは、「Where」フレームに格納された語に、何が存在し、何をすることを目的とした場所なのかを連想することをいう。格納ルールは以下の通りである。

①「Where」フレームに格納された語を場所判断システムに渡し、連想結果の存在・目的を連想フレーム「Where」に格納する。

(2) 時間連想

時間連想とは、名詞と用言の組合せから時間／季節を連想することをいう。格納ルールは以下の通りである。

①「Who」と「用言」、「What」と「用言」、「Where」と「用言」に組み合わせて、時間判断システムに渡し、連想した時間を連想フレーム「When」に格納する。

3. 評価と考察

3.1 情報フレーム

テストデータとして、英会話の参考書⁵⁾から会話文200文を収集し、そのテストデータに対して、人手で作成した個別知識ベースとの一致度を調べた(図9)。評価は下記の3段階とする。

- ・ RankA：完全一致
- ・ RankB：部分一致
- ・ RankC：不一致

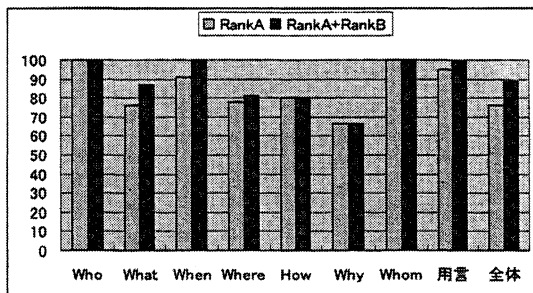


図9. 評価結果

全体では、RankA・RankBを合わせて88.8%の語が正しくフレームに格納できた。また、Who・When・Whom・用言フレームに関しては全ての語を正しく格納できた。

しかし、今回は必ず主語が含まれる文をテストデータとしており、上記で示したルールでは、会話上でよく行われる主語省略に対応することはできない。また、「象は鼻が長い」といった二重主格構文にも対応できていない。よって、今後は対象範囲を拡張し、新たなルールを作成していく

必要があるといえる。

3.2 連想フレーム

3.2.1 場所連想

テストデータ200文中、連想フレーム「Where」に格納されたものは30文である。これに対して、調査を行った。評価は存在語・目的語に関して、正しく連想(O)、矛盾しない連想(Δ)、正しくない連想(X)の3段階で行った(図10, 図11)。

3.2.2 時間連想

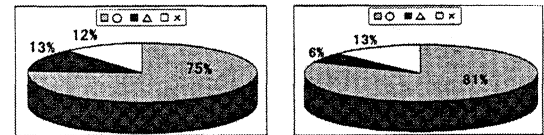


図10. 存在語の評価

図11. 目的語の評価

テストデータ200文中、連想フレーム「When」に格納されたものは48文である。これに対して、調査を行った。評価は正しく連想(O)、矛盾しない連想(Δ)、正しくない連想(X)の3段階で行った(図12)。

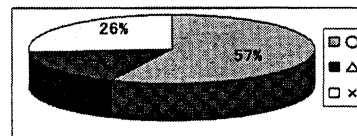


図12. 時間連想の評価

4. おわりに

本稿では、独自のルールを用いて、会話文の内容を整理・連想する手法を提案した。これにより、話者に適切な応答をかせうことができ、会話が円滑に進むことが期待できる。

今後、より日常会話に近い文にも対応できるように対象範囲を広げ、知的な応答が可能なメカニズムを目指す。

なお、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。

参考文献

- [1] 眞鍋康人, 小島一秀, 渡部広一, 河岡司: “概念間の関連度やシソーラスを用いた概念ベースの自動精練法”, 同志社大学理工学研究報告, Vol.42, No.1, pp.9-20, 2001.
- [2] 杉本二郎, 渡部広一, 河岡司: “概念ベースを用いた場所に関する連想メカニズムの構築”, 情報処理学会, 第64回全国大会, 2002 (印刷中).
- [3] 小畑陽一, 渡部広一, 河岡司: “単文の名詞と動詞から時間/季節を判断するメカニズム”, 電子情報通信学会技術研究報告, 人工知能と知識処理, Vol.100, No.530, pp.1-6, 2001.
- [4] NTTコミュニケーション科学研究所監修, 「日本語語彙体系」, (岩波書店, 東京, 1997).
- [5] 向井京子: “6ルールでこんなに話せる! 英会話入門”, 池田書店, 2001