

## 会話モデルにおける言い換え方式：知的おうむ返し

岩下 直人 渡部 広一 河岡 司

同志社大学大学院工学研究科

[dtb0712@mail4.doshisha.ac.jp](mailto:dtb0712@mail4.doshisha.ac.jp), [watabe,kawaoka@indy.doshisha.ac.jp](mailto:watabe,kawaoka@indy.doshisha.ac.jp)

### 1 はじめに

本研究では、ロボットが人間のように自然に会話できることを目標としている。

ロボットが会話できるようになれば、どんなに良いことであろうか。例えば、老人の話し相手になったり、ネット初心者には敷居の高い web の検索(論理演算の使い方)などがより簡単になる。

会話という研究はさまざまな研究機関で行われているが、その原理を解き明かしている研究はまだないと思われる。そこで、研究の第一歩を進めるために、我々は以下の調査を行った。

まず、人間がどのような会話を行っているかを調べるために、例文 100 対について調査した。本稿では(例文1)のように、問題対象としては一対の会話のみを考え、複数の文から成り立つ会話は考えないこととする。

A: 昨日、電柱にぶつかった B1: それは災難でしたね B2: 何を考えていたのですか B3: 君はいつもぶつかるな B4: 電柱って何ですか	… (例文1)
--	---------

この例を見ても分かるように、我々はさまざまな技術を用いて会話を行っていると考えられる。我々はこのような例から6つのルール(知的おうむ返し、話題展開、5W1H、つなぎの会話、未定義語、情報追加)を抽出した。本稿では、その中でも「知的おうむ返し」について述べる。

### 2 知的おうむ返し

「知的おうむ返し」とは単語と短文間の意味を考慮した言い換え方法のことを言う。以下に例を挙げて説明する。

人間は、会話をするときと同じことを言う場合でも違う表現を用いる。応答文に単純に、同じ物を用いるのは人間らしくない。会話を豊かにする

ために「知的おうむ返し」という方法は必要と考えられる(例文2)。

A: 彼女は言いかけていた B: 話そうとしていたの	… (例文2)
-------------------------------	---------

また、次のような会話で「知的おうむ返し」は威力を発揮する。

A: のどが渴いた B: ジュースはいかがですか	… (例文3)
-----------------------------	---------

(例文3)のような会話を出力することは現状では困難である。もちろん、一つの解決方法として、スクリプトを定義しておき、「のどが渴いた」→「ジュース」を連想することは可能である。しかし、これでは「お腹がすいた」→「ご飯」を連想できない。つまり、定義しておかなければならないことが多すぎて、例外が多発する。しかし、本稿で述べる技術を用いれば、(例文4)のような会話が可能になると考える。

A: のどが渴いた B1: アイスティー B2: アイスティーを飲む B3: アイスティーを飲みますか B4: アイスティーはいかがですか B5: 冷たい飲み物はいかがですか	… (例文4)
--	---------

B1からB3までの応答は、短文から短文の言い換えを使用することによって可能になる。また、B4とB5については、本研究の技術(意図理解や感覚判断)を用いることによって可能になる。

また、本研究の利点として、単語→短文といったより大きい単位の意味に機械的に言い換えられることが挙げられる。現在のweb検索では「のどが渴いた」と入力しても、表記の一致しか見ないが、「知的おうむ返し」を利用すると、それに関連する「アイスティー、飲む」に関連する文書をも検索できる。

### 3 用いる辞書等について

#### 3.1 概念ベース

概念ベースとは、概念とそれを表す単語集合（1次属性）からなる知識ベースのことである。なお、概念の属性の属性のことを2次属性と呼んでいる。例えば、概念「鉛筆」では、それが「鉛筆」、「筆記」、「木」、「芯」、「軸」…などの単語で表せると考え、それらに対にして格納している（表1）。

表1 概念ベースにおける概念の例

概念	属性
鉛筆	鉛筆
	筆記
	木
	芯
	⋮

#### 3.2 関連度計算

関連度とは、概念と概念の関連の強さを定量的に評価するものである。例えば「大学」と「講義」の関連性と「大学」と「水」の関連性では人間ならば前者の方が関連性が強いと判断する。

$$Assoc(\text{大学}, \text{学校})=0.20$$

$$Assoc(\text{大学}, \text{講義})=0.18$$

$$Assoc(\text{大学}, \text{水})=0.02$$

(ただし,  $0 \leq Assoc(A, B) \leq 1$ )

図1 関連度の例

計算方法としては、概念連鎖により概念を2次属性まで展開したところで、最も対応の良い1次属性同士を対応付け、それらの一致する属性個数を評価することにより算出する。

概念AとBの関連度  $Assoc(A, B)$  は以下のアルゴリズムにより計算する[渡部 01]。

① まず、2つの概念A, Bを1次属性  $a_i, b_j$  を用いて、

$$A = \{a_i \mid i = 1 \sim L\}$$

$$B = \{b_j \mid j = 1 \sim M\}$$

と定義する。ここで、属性個数は重みの大きいものから30個を上限（実験的に検証された）[渡部 01]として展開するものとする。

② 1次属性数の少ない方の概念を概念Aとし ( $L \leq M$ )、概念Aの1次属性の並びを固定する。

$$A = (a_1, a_2, \dots, a_L)$$

③ 概念Bの各1次属性を対応する概念Aの各1次属性との一致度 ( $Match$ ) の合計が最大になるように並べ替える。ただし、対応にあふれた概念Bの1次属性 ( $b_{x_j}, j = L+1, \dots, b_{x_L}$ ) は無視する。

$$B_x = (b_{x_1}, b_{x_2}, \dots, b_{x_L})$$

④ 概念Aと概念Bとの関連度  $Assoc(A, B)$  は、

$$Assoc(A, B) = (s/L + s/M)/2$$

$$s = \sum_{i=1}^L Match(a_i, b_{x_i})$$

とする。

また、概念Aと概念Bの一致度  $Match(A, B)$  は、一致する1次属性の個数（すなわち、 $a_i = b_j$  なる  $a_i$  の個数）を  $s$  個とすると、次式で定義する。

$$Match(A, B) = (s/L + s/M)/2$$

この式は、概念Aと概念Bの一致割合を評価する一つの方式として、概念Aから見たときの属性の一致割合  $s/L$  と概念Bから見たときの一致割合  $s/M$  の平均を採用している。

#### 3.3 シソーラス

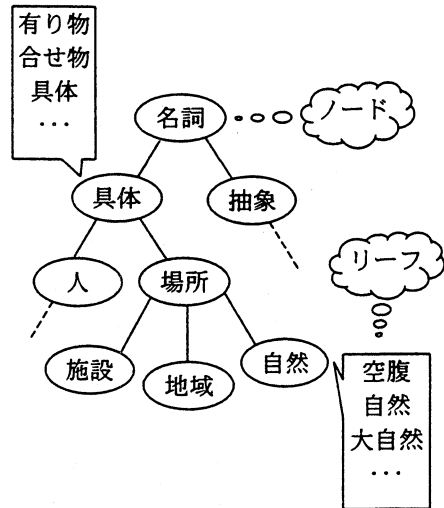


図2 シソーラス

シソーラスとは、単語間の関係を木構造を用いて表したものである。ノードとリーフからなる。ノードに属している単語のことをリーフと呼ぶ。

### 3.4 同義語辞書

ある単語とその同義語が対になった辞書を用いている。(表2)

表2 同義語辞書

絵	図絵
絵	図画
学校	学館
学校	学園
:	:

## 4 言い換えアルゴリズム

### 4.1 単語→単語

単語から単語の言い換えが一番単純である。本稿では、一番単純な方法として同義語辞書を利用する。品詞ごとに同義語が存在するが、今回実験を行ったのは、名詞についてのみである。

### 4.2 短文→単語

短文から単語の言い換えには、概念ベースを利用する。本稿では実験の対象として、修飾単語+被修飾単語に限定した。限定した理由として、長い文にこのアルゴリズムを適用すると、色々な語の雑音のため、一つの単語に限定できない可能性があると考えたからである。また、何語までこのアルゴリズムが適用できるか調査するのも興味深い。

アルゴリズムは以下のとおりである。

- ① まず概念ベースを用いて修飾単語の属性と被修飾単語の属性を取得し、これらを置換概念候補とする。
- ② その置換概念候補それぞれと修飾単語、被修飾単語の関連度を計算する。
- ③ それらの平均関連度が最も高かった単語を置換概念とする。

また、置換概念候補として、(修飾語の属性 or 被修飾語の属性) と (修飾語の属性 and 被修飾語の属性) の二つの場合を検証した。

(修飾語の属性 or 被修飾語の属性) の性質として、ヒットする数は多くなるが言い換えられた単語の意味的距離が大きくなり、反対に(修飾語の属性 and 被修飾語の属性) の性質はヒットする数は少ないが言い換えられた単語の意味的距離が小さくなる傾向がある。

### 4.3 単語→短文

以下のアルゴリズムを用いて、単語から短文への言い換えを行う。

- ① 指定された単語の属性の中から動詞をピックアップする
- ② シソーラスの文情報を用いて、その動詞の構文を探す
- ③ シソーラスの文情報を利用して、主語及び目的語となる名詞をピックアップする
- ④ 主語及び目的語となる名詞が指定された概念の属性の中にあれば、それを使用して短文を作成する

### 4.4 短文→短文

応用として、短文から短文の言い換えも可能になると考えられる。アルゴリズムとしては、今まで説明してきたものを用いる。全体的な流れは、短文→単語→短文である。

短文から単語の言い換えは 4.2 で説明したアルゴリズムを用いる。次に単語から短文の言い換えは 4.3 で説明したアルゴリズムを用いる。

## 5 実験

### 5.1 短文→単語

修飾語+被修飾語を 50 組作成し、評価を行った。人が見て、コンピュータが出力した結果が意味的に近いものであれば正解と判断した。

評価基準は、1. 正解、2. 不正解、3. どちらともいえない、4. 出力なし (null)、の4つを用いた。

#### 5.1.1 and 条件の場合

and 条件の場合の評価結果を表3に示す。

表3 評価結果 (and)

正解/全体	0.52
(正解+どちらともいえない)/全体	0.7
正解/(全体-null)	0.703
(正解+どちらともいえない)/(全体-null)	0.946

答えを出力しないこともあるが、出力されれば正解率が高い。(正解+どちらともいえない)/(全体-null)の正解率が94.6%と高く出ることから、この技術は有用であると考えられる。

表4に成功例と失敗例を示す。

表4 実験例 (and)

成功例	失敗例
赤い+食べ物=さくらんぼ	青い+空=空虚
大きい+動物=熊	美味しい+水=伏流
味+感覚=味覚	厚い+本=null

### 5.1.2 or 条件の場合

or 条件の場合の評価結果を表5に示す。

表5 評価結果 (or)

正解/全体	0.32
(正解+どちらともいえない)/全体	0.76
正解/(全体-null)	0.32
(正解+どちらともいえない)/(全体-null)	0.76

例題すべての場合で答えを出力するが、「どちらともいえない」の割合が and の場合と比べて多い。

表6に成功例と失敗例を示す。

表6 実験例 (or)

成功例	失敗例
暖かい+服=スキーウェア	触る+感覚=くぐれる
甘い+お菓子=綿菓子	賢い+人=凡人
高い+建物=東京タワー	臭う+感覚=匂う

### 5.2 単語→短文

任意の単語 50 個を選び、プログラムにこの 50 個をかけて、出力する結果が正しいかどうか人間が判断した。いくつか答えが出力されるが、1 つでも正しいものが出力されれば正解とした。正解率は 74% という結果になった。

表7に成功例と失敗例を示す。

表7 実験例 (単語→短文)

成功例	失敗例
学校=学生+入学する	秋=null
出産=赤ちゃん+生まれる	音楽=null
足し算=数+加える	磁石=null

## 6 考察

本稿では、短文→単語と単語→短文の評価を行った。短文→単語 (and) の正解率が (正解+どちらともいえない)/(全体-null) で 94.6%、単語→短文では、76% という結果になった。会話のアルゴリズムの一つということを考慮すれば、この結果は実用に耐え得る物と考える。

短文→単語の悪くなった例として、「青い+空

=空虚」というものがあるが、これも答えを複数許せば「青空」という答えが返ってきているので、システムを答えを複数許すようにすればより良い結果が得られそうである。また、同じような意味の単語でもシステムの答えが違う場合がある。例えば、「青い+空」の尋ね方を変更して「青+空」とすると「真っ青」という正しい答えを返した。この問題を解決するためには、質問単語を同義語情報を用いて膨らます方法が考えられる。

次に、単語→短文のほうであるが、実験の範囲では、答えを返す率が低かった。この原因として、概念ベースの動詞が少ないことが挙げられる。これは、概念ベースの質の向上を目指す処理を重ねるごとに名詞が多くなっているからだと思われる。また、概念ベースに日常的には使用しない単語が入っていることも問題である。例えば、概念「音楽」の属性に「奏する」と言う動詞が入っているが、これは、日常使用しない単語である。そのためシソーラスの文情報にもヒットしない。

## 7 おわりに

今後の課題としては、「知的おうむ返し」の精度向上のために、概念ベースやシソーラスの文情報を良くしていく方法を考案することが挙げられる。また、「知的おうむ返し」以外の会話モデルを提案する必要がある。

### 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「知能情報科学とその応用」における研究の一環として行った。

### 参考文献

- [渡部 01] 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  
 [笠原 97] 笠原 要, 松澤 和光, 石川 勉: 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌, vol.38, No.7, pp.1272-1283, 1997