

日本語対話処理のためのユーザ入力支援

7E-1

松尾 比呂志, 大山 芳史, 中川 優

N T T 情報通信処理研究所

1. はじめに

自然語による対話処理システムでは、日常の言葉による計算機とのコミュニケーションを目指している。しかし、人間が用いる多様な表現を正しく理解するには、人間と同程度の言語理解能力が必要となる。このような能力を計算機に持たせることは現状では不可能であるため、システムが受理可能な文が制限されている。しかし、ユーザにはどのような文が入力できるかわからないため、受理できない文を入力することが少なくなく、計算機との円滑な対話の障害となっていた。本稿では、入力文を基に受理可能な文を例示することにより、ユーザが容易に受理可能な文を入力できる支援処理を提案する。

2. 入力支援の基本的考え方

ここで、正しく意味を理解できず確かな応答をユーザに返すことが保証されている文を正規文と呼ぶ。入力文が正規文であって、システムが持つ知識ベースと対応付けができない文の場合には、“なぜユーザが期待する応答ができないか”をユーザに提示することが可能である。しかし、ユーザが入力した文が正規文でない文の場合には、入力支援処理がなければ、誤った解釈のまま応答を返したり、理解できないことを示すことしかできないため、円滑な対話の障害となる。入力支援処理では、正規文と異なる文の場合には、あらかじめ登録した正規文と照合を行い、入力文に類似した文を抽出し、システムが受理可能な文をユーザに提示して選択させる。このため、システムには受理可能な文のみが入力される。図1に入力支援処理の例を示している。この例では、正規文でない文「ハーフトーンのFAXがあればよい。」が入力され、3つの例文が提示された場合を示している。ユーザはこれらの3つの文の中から自分の意図に合った文を選択すればよい。

3. 文の照合処理による入力支援

ユーザが入力した文に近い正規文を抽出するために、形態素解析処理後、以下の評価規準に基づいた類似性の評価値を算出することによって、類似性の高い正規文を抽出している。

- 1) 入力文と正規文で意味属性が一致した自立語が何組存在するか。(意味属性には、汎用的な意味属性¹⁾の他に、「kinou_mei」や「syubetu」のよう

なデータベースの項目に対応した分野固有の意味属性も設けている。)

- 2) 意味属性が一致した自立語に対して、各自立語に接続される語の文法的属性(助詞の場合は表層格、自立語の場合は品詞)が一致しているか。

例えば、図2は2つの文間の類似性の評価値算出方法を示している。この例では、意味属性が一致した自立語に得点 α 、さらに後続語の文法的属性の一致に対して得点 β を与えた場合を示している($\alpha > \beta$)。この2つの文では、「kinou_mei」「syubetu」「sonzai」「kibou」の4つの意味属性で一致して得点 4α が得られ、「syubetu」の表層格が“が”格で一致しているためさらに得点 β が加えられ、 $4\alpha + \beta$ の評価値が得られる。2)の評価規準は、「Aが欲しい。」と「Aも欲しい。」のように1)の評価規準だけでは評価値に差が生じない正規文に差を付けるためであり、1)の評価規準に比べ低い得点を与えている。

照合処理のアルゴリズムを図3に示す。正規文の各々を順次比較する総当たり法では、実用的な時間で処理できないため、あらかじめ生成した以下のインデックステーブルにより、高速な照合処理²⁾³⁾を行う。

- 1) 正規文の形態素解析で得られた各単語から正規文へのインデックステーブル。
 - 2) 各意味属性から単語へのインデックステーブル。
- 照合処理のフローを以下に示す。
- 1) 入力文の形態素解析を行い、各単語の意味属性と文法的属性を抽出する。
 - 2) 各意味属性からインデックステーブルを用いて、単語を抽出し、各単語から正規文を取り出して、正規文の評価値を加算していく(評価規準1)。
 - 3) 評価値が高い正規文に対して、評価規準2のための評価値の補正を行う。
 - 4) 評価値の高い正規文を抽出する。

例えば、入力文「ハーフトーンのFAXがあればよい。」に対して、「ハーフトーン」から意味属性「kinou」が抽出され、意味属性「kinou」を持つ単語を含む正規文の一つとして「複写機能があるFAXが欲しい。」が抽出される。

表示文の生成処理では、「複写」が「ハーフトーン」に置き換えられてユーザに表示される。

4. システム評価

(1) 文の照合処理の高速性

登録した正規文約180文において、総当たり照合に比べ1/100程度の高速化が図られた。インデックステーブルを用いているため、正規文を増やしたとしても処理量の増加はnオーダー以下である。このため、大量の正規文に対しても実用に耐える時間での処理が期待できる。

(2) 入力支援の有効性

FAX製品の情報案内の対話例で実験を行った(約180文登録)。初めて使用する被験者(1)において、40文中約60%の文に対して、学習した被験者(2)においては、約80%の文に対して、4位以内にユーザの意図に合った文が表示された。成功した入力文には、「コピー付きのもの。」のように、「～が欲しい。」というようなユーザの意図を表す言葉が省略されたものも多くあった。このように、完全な文でなくてもユーザ入力支援により自分の意図を入力できることができ、入力支援の有効性が確かめられた。入力支援が失敗した例としては、被験者が専門用語を知らないために、その用語を用いずにユーザ独自の表現を用いた入力文が、被験者1で失敗例中の1/3を占めた。例えば、「同報通信」を表現するのに「いっぺんに送れるもの」というような表現を用いている。被験者2が被験者1に比べ高い成功率を示したのは、このような表現が少ないためである。このような表現に対しては、想定されるユーザ独自の表現を登録して専門用語と結び付けることにより、入力支援の成功率を高めることが可能である。

5. おわりに

ユーザが入力する多様な表現から、システムが理解可能な文を提示することによって、ユーザの意図を的確にシステムへ入力できるユーザ入力支援処理を提案した。提案した文の照合処理は、高速に意味が類似した文を抽出でき、テキストベースの知識の検索⁴⁾にも使用できる。今後は、文脈情報を使った処理や句レベルの照合とを組み合わせた処理により、より多様な表現を許容できるよう拡張する。

[参考文献]

- (1) 池原他「言語における話者の認識と多段階翻訳方式」情処論 28, 12, pp.1269-1279
- (2) 松尾他「特徴要素別にカテゴリ選択を行う高速パターン照合法」信学論 J70-d, 12, pp.2503-2509 1987
- (3) 松尾他「連想統合型照合による単語あいまい検索法」情処34回全国大会pp.1845-pp.1846 1987
- (4) 西山他「検索キーの文章定義による日本語対話処理」情処38回全国大会投稿予定

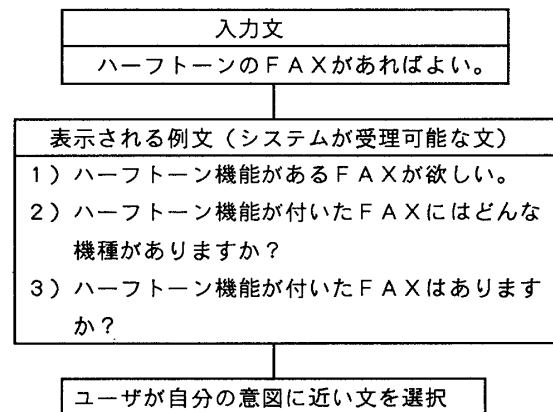


図1 入力支援処理の例

入力文 「ハーフトーンのFAXがあればよい。」
正規文 「複写機能があるFAXが欲しい。」

| 入力文 | | 正規文 | | 得点 |
|--------|-----------|-----|--|----|
| 自立語 | 意味属性 | 自立語 | | |
| ハーフトーン | kinou_mei | 複写 | | a |
| - | kinou | 機能 | | 0 |
| FAX | syubetu | FAX | | a |
| あれ | sonzai | ある | | a |
| よい | kibou | 欲しい | | a |

$$\begin{aligned}
 \text{評価値} &= 4a \quad \dots \text{評価規準(1)} \\
 &+ \beta (\text{「syubetu」が}) \dots \text{評価規準(2)} \\
 &= 43 \quad (\alpha = 10, \beta = 3 \text{の場合})
 \end{aligned}$$

図2 評価値算出法

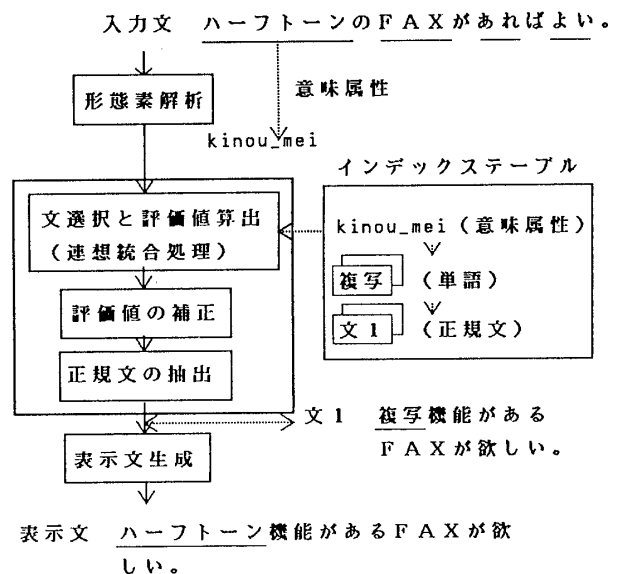


図3 文の照合処理