6F－4

対話データベースを用いた各種言語現象の検索

橋本 一男　小倉 健太郎　江原暉将　森元 逞
ATR自動翻訳電話研究所

【1】はじめに

ATRでは自動翻訳電話の実現に向け、対話文を対象とする大規模言語データベースADD(ATR Dialogue Database)を構築している[1]。話し言葉の言語現象を把握するために、ADDでは対話テキストを単語、文などの言語単位ごとに要素分割し、各要素に属性データや要素間の関連データを付与してある。本稿では、まずADDの構成と検索システムの概要を述べた後、属性や関連のデータを用いた言語現象の検索例を示しADDの有用性を明らかにする。なお、ADDの内容については別稿[2]で述べる。

【2】ADDの構成

これまでにも言語データベースは数多く構築されてきているが、実際の話し言葉を対象にした本格的なものとなると国立国語研究所のものが唯一であり、その内容も単語レベルに限られる[3]。自動翻訳技術を研究するためには、単語個々のテキスト表記や属性だけでなく、さらに単語の連鎖/共起や係受け、英語との対応などが必要になる。ADDでは、言語要素に次のような関連情報を付与してあり、これらを用いて検索を行うことができる。

1) 包含関係: 文-文節、文節-単語といった2レベル間の全体-部分関係(part-of、composed-of)を示す。
2) 順序関係: 文節同士、単語同士などの同じレベルでの前後関係(prev-is、next-is)を示す。
3) 日英対応関係: 文、文節、単語などの各レベルで、対応する英語との関係を示す。
4) 係受け関係: 文節間の係受け関係を示す。ADDでは便宜上、文節を代表する単語間に係受けをふっている[4]。

【3】検索システム

ADD検索の大きな特徴となる多様性に応えるため専用検索システムが用意されている。本システムは、利用者が検索条件と出力形式を記述すると、データ間の関連を自動的にたどって検索処理を行う機能を備えている[5]。図1に「NのNのN(Nは名詞句)」というパターンを含む文を検索する例を示す。ldbshコマンドは検索システムの起動、target、modeコマンドはそれぞれ検索範囲の指定、出力形式の指定に用いる。wsearchコマンドは、単語列検索用のコマンドで

```
wsearch の + * + の      (* はワイルドカード)
```

と入力すれば、精度(検索用例中にどれだけ目的に合致するものがあるか)は多少低いが、十分な再現性(必要な用例がどれだけADDから検索できるか)が得られる。精度を向上させるには、

```
wsearch の + * + 品詞:名詞 + * + の
```

と書けばよい。さらに複雑な検索を記述するためには、表記入形式の記述支援ツールを用いる。図2は、「教える」という単語と係受け関係にある文節の検索記述例である。

```
% ldbsh
ldbsh: target
(範囲指定画面で範囲指定を行う)
対象会話:3045 3046 3048 3049 3050
ldbsh: mode
(出力形式指定画面で形式指定を行う)
ldbsh: wsearch の+*+の> kekka &
(上の&はBG処理の指定。この間、別の操作が可能)
検索(kekka)が終了しました
ldbsh:!cat kekka
1. ３カ月後のそちらからの次のサーキ
2. ３カ月後のそちらからの次のサーキ
   .........
8. 一応今のところは各分科会のテーマ
9. この会議の焦点はＡＩにあるのです
```

図1　検索システムの操作例

```
パターン名:   Q01
RETURN    :   #<文節1>,#<文節2>,#Z,#x,#y
-----------------------------------------
係受け      :#<係受け1>
   HEAD          :#<単語1>
   MODIFIER      :#<単語2>
   意味関係名    :#x
   構文関係名    :#y
-----------------------------------------
単語        :#<単語1>
   PART-OF       :#<文節1>
   表記          :
   読み          :
   標準表現      :教える
   品詞          :
   . . . . . .
```

図2　記述支援ツールの画面例

Retrieving Various Linguistic Phenomena from A Dialogue Database
Kazuo HASHIMOTO, Kentaro OGURA,.Terumasa EHARA, Tsuyoshi MORIMOTO
ATR Interpreting Telephony Research Laboratories

【4】検索例

ATRにおける利用例を3つ示す。

4.1 単純な検索(音声認識のための検索例)

音声認識では、短単語に関して特に詳細な文法が要求される[6]。代名詞は普通名詞、固有名詞に比べて短く、また後続の助詞も概して短いため、文法を書く上で図3のような「代名詞」を含む文節の頻度付き用例は重要な情報となる。

4.2 順序関係、日英対応関係を用いた検索(日英翻訳のための検索例1)

翻訳技術に関する問題のひとつに訳し分けがある。用例に基づいた翻訳方式では、実際の日英対応用例をデータベースとして用いる[7]。図4は「の」による修飾複合という連鎖パターンを持つ文の日英対応用例である。

4.3 係受け関係、日英対応関係を用いた検索(日英翻訳のための検索例2)

また、多義語訳し分けのために、多義語と係受け関係にある語に着目して、その対訳を調べるというアプローチがある[8]。図5は「教える」についての参考用例である。

【5】おわりに

現在、ADDは約20万語が利用可能である。本稿では用例検索について述べたが、統計情報抽出のツールも整備されている。これについては別の機会に報告したい。

検索システムはVT282を端末として想定しているが、エミュレーションにより他機種(たとえばPC-9801)でも利用可能であり、ATR内外の研究者が手軽に活用できる。

謝辞

日頃、ご指導いただく本研究所樽松明社長に感謝いたします。保坂順子、隅田英一郎、工藤育夫研究員をはじめとするデータ処理研究室、言語処理研究室諸氏には、ADD利用について数多くの議論をしていただきました。ここに感謝の意を表します。

文献

[1] 森元逞ほか:自動翻訳電話研究用言語データベースの収集について、第36回情処全国大会4U-5、1988
[2] 江原暉将ほか:電話対話データベースの構築、第40回情処全国大会、1990
[3] 中野洋:話し言葉の語彙調査、情処NL30-1、1982
[4] 井ノ上直己ほか:言語データベース用単語間の関係データ、第37回情処全国大会5B-7、1988
[5] 橋本一男ほか:フレーム表現による検索機能を有する言語データベース管理システム、情処アドバンストデータベースシンポジウム、1989
[6] 保坂順子ほか:音声認識のための文法構築(仮題)、第40回情処全国大会、1990
[7] 隅田英一郎ほか:用例に基づいた翻訳、第40回情処全国大会、1990
[8] Ikuo Kudo et al.:Lexical-Functional Transfer、Coling'86

1.	41:	私の
2.	39:	何か
3.	39:	こちら
4.	31:	私、
5.	28:	こちらの
6.	27:	私が
7.	26:	私は
.....		

図3 「代名詞」を含む文節の用例
中央の番号は頻度を示している

.....

27. これからの会議日程、発表論文の依頼の連絡を取らせて頂きたいと思います。
27. I am looking forward to your visit to Japan , and I will be talking to you soon about the Conference schedule and your paper .
28. これからの会議日程、発表論文の依頼の連絡を取らせて頂きたいと思います。
28. I am looking forward to your visit to Japan , and I will be talking to you soon about the Conference schedule and your paper .

.....

42. 先生の発表は会議3日目の29日第一分科会の10時から11時までです。
42. You are supposed to give a speech in the first session , from 10 o'clock a.m. to 11 o'clock on 29th , the third day of the Conference .

.....

図4 「の」による複合修飾の用例

私このような会議に参加するのは初めてですので、チェックインの方法を教えて頂けますか?
I've never attended a conference such as this before . Could you tell me the check-in procedure please ?
教えて頂けますか?|方法|普通名詞|対象|連用格
Could you tell me

この事について、もう少し詳しく教えていたいただけますか?
However , the secretary of our lab tells me that you now require some additional material - could you give me some more details on that please ?
教えていたいただけますか?|詳しい|形容詞|方式、様態|述語連用修飾
教えていたいただけますか?|事|普通名詞|範囲規定、関係|連用格
could you give me please

図5 「教える」の係受けと対訳の用例