

言語データベース統合管理システムの 2C-4 マンマシンインターフェース

橋本 一男 小倉 健太郎 森元 還

ATR自動翻訳電話研究所

【1】はじめに

言語データベースは、会話文の自動翻訳技術を研究するための基礎データとなる[1]。そこには、大量の会話用例に加え、それらに付随する多種多様な構造情報（品詞、意味関係、構文関係、多言語間の対応等）が蓄積される[2]。この複雑な様相を持つ言語データベースを有効に利用するために、統合的な管理システムが必要である[3]。我々はまず、簡単で効率のよいデータ操作実現のため、オブジェクト指向を利用した視覚的なマンマシンインターフェース部の試作を行った。本稿では、本システムの概要について述べた後、その中心機能である「検索」について述べる。その他の機能については参考文献[3]を参照されたい。

【2】システムの概要**2.1 基本要件**

言語データベースは、主として会話文の言語現象を解明するためのものであり、その管理システムは次の要件を満たさねばならない。

- (1) 言語データの複雑な様相を表現できること。
さらに言語現象分析のために多様な角度からのデータ操作ができること。
- (2) ユーザとして言語分析者、特に機械翻訳に関わる研究者を対象にしているため、人によりコンピュータ操作の習熟度に大きな差がある。
したがって、初心者から熟練者まで、ユーザのレベルに応じた操作環境を提供できること。
このためオブジェクト指向のデータ構造を採用するとともに、定型的な検索パターンの登録／編集機能の実現、マルチウィンドウ、マウス等の利用を行うことにより、上記要件に応えた。

2.2 データ構造

言語データの情報構造はクラス／インスタンス、部分／全体および多面的なis_a階層等から成る。
このため、これらを十分に表現できるよう、オブジェクト指向のデータ構造を採用している[3]。

2.3 高機能ワークステーション

本システムは、(1)ユーザインターフェースにキーボード、マウス、マルチウィンドウ、アイコンが使えること、(2)Common Lispを使ったシステム記述に良好な開発環境であること、(3)オブジェクト指向の表現にFlavorが使えること等を理由に、

Symbolics 3600上で開発／実装した。

さらに、このように高機能ワークステーションとオブジェクト指向を用いることで、ディスプレイ上のほとんどのオブジェクト（データ、ウィンドウ等）をマウスセンシティブにでき、操作は一部のキーボード入力を除いて、ほとんどマウスで行うことができる。

【3】検索機能**3.1 検索フレームワーク**

本システムは、初心者用に使用頻度の高い検索パターンをいくつか用意している。しかし、言語データに対する検索要求は多種多様であり、それら1つ1つにシステム組込みの検索パターンを用意するのは困難である。そこで検索フレームワーク(SFW)を用意し、それを用いることでユーザーが自由に検索パターンを定義できるようにした。たとえば、『動詞+助詞「か」のパターンを持つ文を求める』場合の検索パターンは、図1のように定義される。このようにSFWは、

- (1)検索パターン宣言 (defsearch)
 - (2)検索パターン名（任意、図1ではK10）
 - (3)コメント（”で囲まれた部分）
 - (4)検索パターン記述部 (object-pattern)
 - (5)リターンフォーム記述部 (return-form)
- の5つの部分からなる。このうち(4)では、検索パターンを指定する。また(5)では、検索結果の表示単位（文、単語等）を指定する。

```
(DEFSEARCH K10 "動詞 + * + 助詞か"
(OBJECT-PATTERN (VAR 文節)
  ((INSTANCE-OF 文節)
   (COMPOSED-OF * (VAR 単語1) *
                 (VAR 単語2) *)))
(OBJECT-PATTERN (VAR 単語1)
  ((INSTANCE-OF 単語)
   (PART-OF (VAR 文節1))
   (品詞 動詞)))
(OBJECT-PATTERN (VAR 単語2)
  ((INSTANCE-OF 単語)
   (PART-OF (VAR 文節1))
   (出現単語 か)
   (品詞 助詞)))
(RETURNS-FORM 文 0 0))
```

図1 検索パターンの定義例

3.2 パターンエディタ

SFWはLisp形であり、そのままでは記述に若干の煩わしさが残る。そこでユーザが、検索パターンを対話的に容易に定義できるツール（パー

エディタ：P E）を用意した。P Eはユーザが言語データの部分／全体関係、順序関係、スロット条件等を対話的に記入すると（図2）、それを図1のような検索パターンに変換する。

3.3 言語データの表示

検索された言語データはまず、日本語テキストの形で表示する（図3右上）。特に反転表示によりマッチングに成功した部分（単語）を強調する。

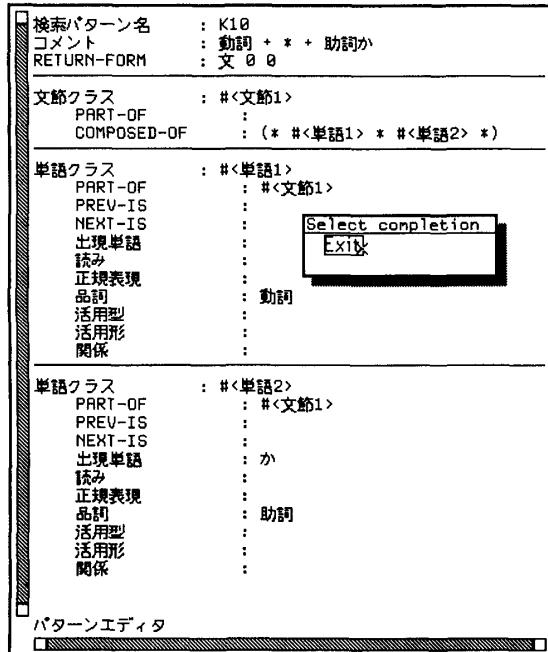


図2 パターンエディタ（P E）

各日本語テキストは、それぞれに関連情報に対するポインターを持たせ、さらにマウスセンシティブにすることにより、

(1)日英間の文レベルの対比（図3右下）

(2)形態素の詳細データ（図3左上）

(3)言語データの部分／全体関係（図3左下）等をマウス操作だけで連続的に検索できる。また各種検索結果は、ウィンドウに表示されたイメージでファイルに出力もできる。

【4】おわりに

オブジェクト指向プログラミングを用いて、複雑な様相を持つ言語データに対して、簡単で効率的なデータ操作のできるマンマシンインタフェースの試作を行った。現在、テストデータによる性能評価を行っている。今後は、関係データベースを用いた言語データ格納機能を開発し、本システムとの接続を行う予定である。

【謝辞】

本発表の機会を与えて下さいました当研究所構松明社長、日頃熱心に討論して下さいましたデータ処理研究室および応用技術諸氏に感謝します。

＜参考文献＞

- [1]森元、小倉、飯田：自動翻訳電話研究用言語データベースの収集について、情報処理36回全国大会4U-5、1988
- [2]小倉、篠崎、森元：言語データベース収集支援システム、情報処理36回全国大会4U-4、1988
- [3]小倉、橋本、森元：言語データベース統合管理システム、情報処理37回全国大会、1988

図3 検索結果の各種表示