# コンピュータと言語処理

NTTヒューマンインタフェース研究所　小　橋　史　彦

自然言語処理技術は，基本技術から応用技術まで幅広く研究開発が進められている．応用面に絞って見てみると，それらの技術は大きく3つの分野に分類できる．第1番目はメディア変換処理に関するもので，日本語の入出力を各種のメディア（文字，図形，画像，音声）を介して行なう場合に，自然言語処理技術を用いて言語との相互変換を高精度に行なう技術である．かな漢字変換技術，音声合成用言語処理技術が先行し，音声認識用言語処理技術，文字認識用言語処理技術の開発が実用化に向けて進められている．また，画像と言語との相互変換技術も基礎的研究が開始されている．第2番目は計算機対話処理に関するもので，データベース検索や質問応答などにおける計算機との対話を自然言語を用いて行なう技術である．キーボードを介した自然言語対話システムとして，ワークステーションなどの端末を利用したパーソナル用インタフェース，およびホストコンピュータを利用したパブリック用インタフェースの双方から研究が進められている．また，音声を用いた対話処理の研究も開始されている．第3番目は文章処理に関するもので，計算機に入力されたテキストを自動加工して，人間の情報活動を支援する技術である．機械翻訳，文章推敲，テキスト検索，分類などの研究が進められ，一部実用化もなされている．また，文章生成，物語理解などの研究も始まっている．

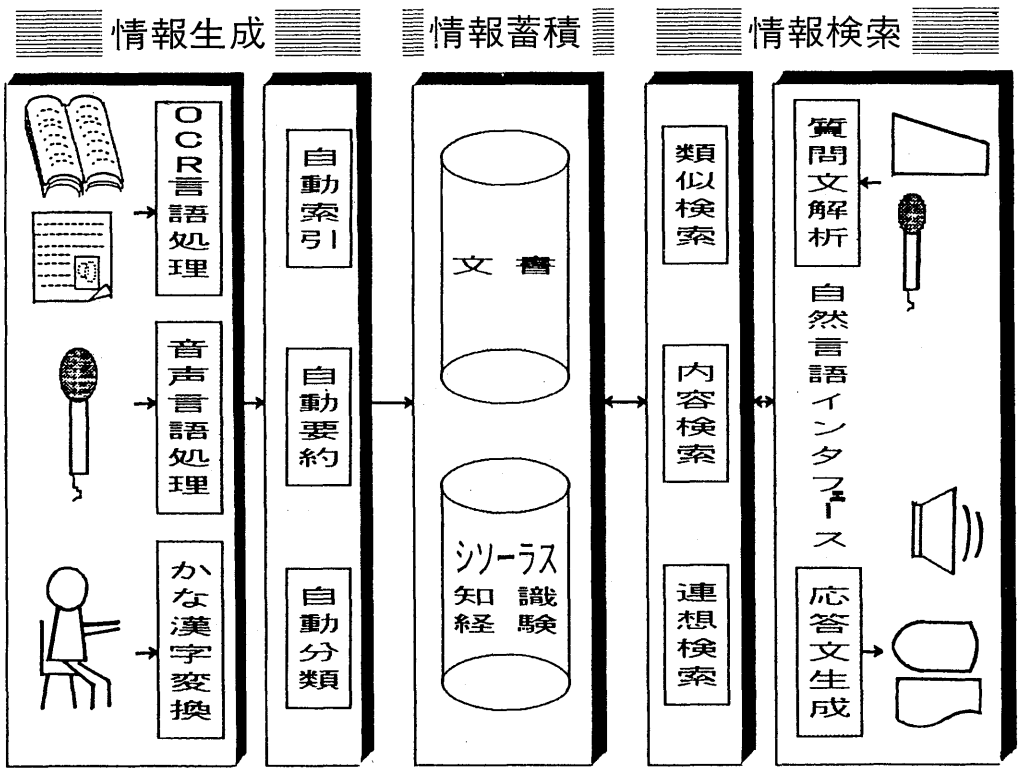


図1．自然言語ファイリング．

現在，我々はこれらの自然言語処理技術でもって情報の生成，蓄積，検索の一連の過程を支援する自然言語ファイリングシステム（図1）の研究を進めている．ここでは，その中からいくつかの研究例を紹介し，技術の現状と今後の展望にふれる．

情報入力の省力化の観点からは，文字認識技術の研究を進めている．単一フォントの印刷文字および丁寧に筆記された手書き文字の認識技術は既に実用化しており，マルチフォントの印刷文字および通常筆記の手書き文字を文脈情報も解析して高精度に読み取る方法について研究を進めている．

期待の大きい入力技術として，音声認識入力技術がある．既に単音節単位の入力技術は確立しており，より自然な入力法としての文節単位の連続発声による入力法について，音声情報と言語情報の融合処理で正しい日本語文（漢字かな混じり文）に変換する研究を進めている．

情報を自動抽出する観点からは，自動索引抽出技術の研究を進めている．あらかじめ定められた統制語に対する索引抽出技術は実現されており，さらにフルテキストにおける自由語に対しても，文章の構造を解析して文章中における重要度を反映した構造化キーワードを抽出する研究を進めている．

自然言語を用いた計算機対話技術は，入力できる文型を固定するなどの使用上の制約のもとで実現されているが，より柔軟性に富み，類義な表現でも許容する，タフな対話法の実現を目指し，語順の変化や省略表現に対する手法や，言葉の同義関係，類義関係，対義関係をシソーラスの形で表現する手法などについて研究を進めている．

## 日蓮の文体について

群馬大学 教育学部 古瀬順一

1988年3月に，愛知教育大学国語科を卒業した大野秀幸君が，「日蓮遺文における文体研究――統計的分析を通して――」と題する卒業論文を，私のもとで書いた．ここでの報告は，その内容紹介である．

この卒業論文で，大野君が対象にしたのは，次の9作品である．

A 「念仏無間地獄抄」　B 「一生成仏抄」　C 「主師親御書」
D 「四恩抄」　E 「開目抄(上)」　F 「さじき女房御返事」
G 「南条殿御返事」　H 「種種御振舞御書」　I 「妙法比丘尼御返事」

この作品選定には，成立年と文章内容，それに受け手の性別等を考慮した基準が用いられている．分析には，樺島（1979）による方法が採られている．

各作品から，乱数表を使って，分析対象となる50文ずつ(50文に満たない作品にあっては全文）を無作為抽出する．そして，文節に区切った後，品詞分解をし，以下の8項目について数値を求めていく．

(1) 名詞比率　(2) MVR　(3) 指示詞比率　(4) 字音語比率
(5) 文長　(6) 引用文比率　(7) 接続文比率　(8) 現在止め文比率

分析結果から，次のようなことが指摘できる．(1)：作品Fがもっとも高く，したがって，これが要約的表現を一番高い割合で含んでいる．(2)：この数値の大きいHなどは，描写的な文章であり，逆にAは描写よりも動きをとらえて描く文章である．(3)：分脈への依存度が一番