

大量の情報を整理してくれる、英語もわかる電子秘書

人がふだん使う書きことばや話しことばを自然言語と言い、それをコンピュータで理解したり翻訳したりする技術を自然言語処理技術と呼びます。人どうしのコミュニケーションなど、様々な場面で人は自然言語を使います。他の人に自分の意図を伝えるために絵やジェスチャーなども使いますが、自然言語がもっとも中心的な役割を果たしているのは言うまでもありません。

当社は、日本語ワープロを日本で初めて製品化し(1978年)、日本の文字文化に大きな影響を与えました。自然言語処理技術の一つである“かな漢字変換技術”の実用化がキーとなりました。そして現在、様々な応用へと自然言語処理の技術が適用されるようになってきています。

単語を認識するための形態素解析

英語では、単語と単語の間に空白がありますが、日本語にはありません。そこで、コンピュータが日本語を理解するためには、まず単語の切れ目を認識する必要があります。当社は、通常の単語の品詞分類(名詞、形容詞、など)と比べて、より詳細な品詞分類に基づく形態素解析エンジンを開発しました。詳細な品詞分類に基づき単語間の接続性をチェックし、日本語として成立しない並びを排除することにより、精度の良い解析を実現しています。

単語間の関係を解析する構文意味解析

自然言語文における単語と単語の関係を表現した構造を、構文意味構造と呼びます。構文意味解析は、形態素解析が出力する単語の並びから、この構造を導き出す処理のことです。また、例えば、

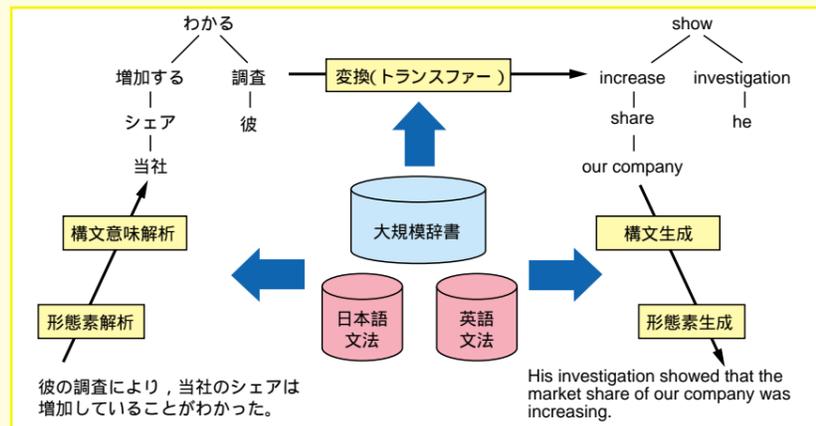


図1. 機械翻訳システムの概要 大規模な辞書・文法に基づき、ソース言語の文(原文)を解析し、目的言語(訳文)に翻訳します。機械翻訳システムは、形態素解析、構文意味解析を応用したシステムです。

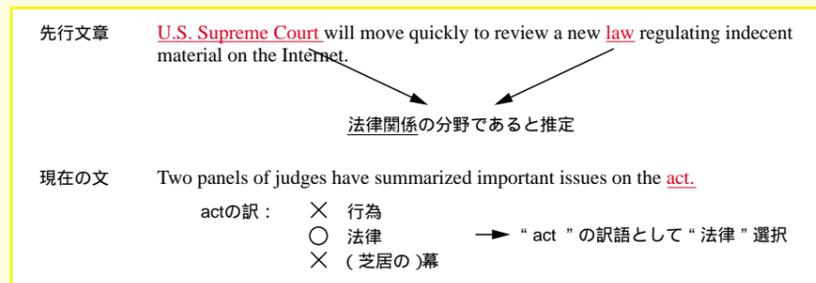


図2. 分野推定に基づく訳語選択 従来の方式では、訳語を決定する手がかりとなる情報がない場合、“行為”という訳を一律に選択していました。先行する文章から、話題や専門分野を推定し、それを元に、特定分野で使用頻度の高い訳語を選択するための方式を開発し、より適切な訳語を生成できるようになりました。

“... take a bath”という文では“take”は“入る”という意味になりますが、“... take a taxi”という文では“乗る”という意味になります。意味を特定できなければ、正しく文を理解したことになりません。構文意味解析では、このような個別の単語に関する意味の同定も必要となります。当社は、これまで個々の単語レベルにまで及ぶ大規模な辞書と文法を開発してきました。これらに基づいて高精度な解析を行います。

自然言語処理の応用

自然言語処理技術は、応用システムを

通じて人の知的活動を支援するコア技術と言えます。

機械翻訳

例えば、“The翻訳™”は、英語を日本語に、日本語を英語に翻訳するパーソナルユースのパッケージソフトウェアです。また、当社のパソコン(PC)“Dyna-Book”にもプリインストールされています。

機械翻訳では、形態素解析、構文意味解析によって入力文の構文意味構造を導き出した後、対象となる言語の構文意味構造に変換(トランスファー)し、構文生成、形態素生成を経て、訳文を生成しま

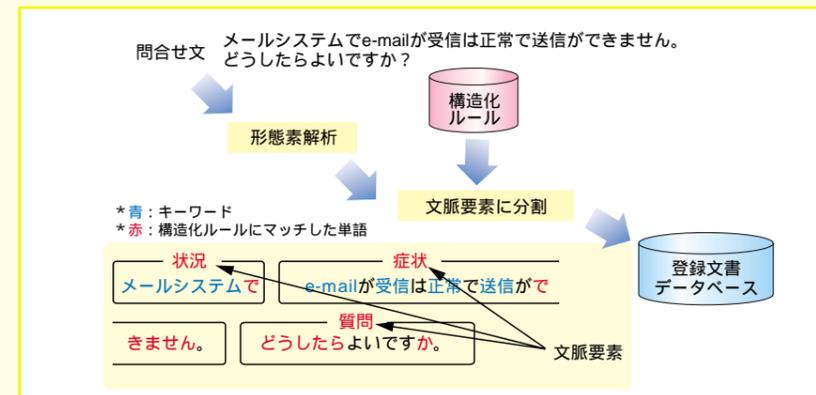


図3. 知識情報構造化の例 コールセンターにおける問合せ事例の場合、状況、症状、質問などが文脈要素になります。登録文書と検索のための問合せ文に対してこのような構造化を行うことにより、文脈を考慮した検索を実現しました。

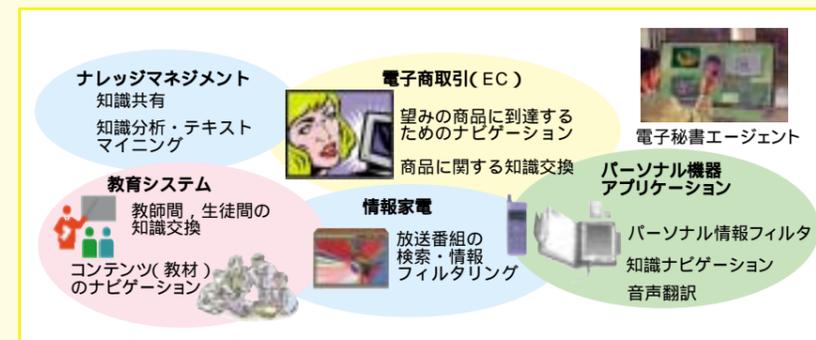


図4. 自然言語処理の応用 自然言語は、人にとって知識を表現する基本メディアです。このため、自然言語処理技術は、様々な応用領域において、人と人とのコミュニケーション、コンテンツのナビゲーションなどを知的に行うコア技術と位置づけられます。

す(図1)。次のような特長があり、雑誌などの評価では、トップクラスの翻訳精度であるとの評価を得ています。

- 1) 20年にわたって開発してきた大規模な翻訳辞書・文法に基づき高精度に翻訳を行います。
- 2) 分野推定に基づく訳語選択機能により、1文内の情報だけでなく、文脈情報に基づく訳語の選択を行います(図2)。

自然言語検索

“KnowledgeMeister™”はナレッジマネジメントを志向したサーバタイプのパッケージソフトウェアです。特別なキ

ワードや検索条件を入力する代わりに、日常使う質問文の形で入力すれば、蓄積されている文書の中から関連する文書を検索することができます。

このような、よりユーザーの意図に沿った検索を実現するため、知識情報構造化という技術を開発しました(図3)。これは、入力された問合せ文の形態素解析結果に対し、句読点や接続助詞などを手がかりにし、単位文に分割します。その後、構造化ルールと照合して文脈要素を特定します。そして、文書中の単語を文脈要素と組合せてデータベースに登録します。検索時には、単語の統計的な重み

付けと、これらの文脈要素を組み合わせ、文書を検索します。これにより、日常使われる質問文による文書の検索が可能となり、ナレッジマネジメントにおける知識共有に効果を発揮します。

その他の応用

上記応用以外にも、自然言語処理を適用した様々な応用があります。

- 1) テキストマイニング 日報やアンケートなど、大量のテキストデータから意味のある知識に集約します。そして、集約した知識をマーケティングや経営判断に生かします。ナレッジマネジメントにおける知識創造につながる応用です。
- 2) 情報フィルタリング 日々発生する新聞記事などを対象に、ユーザーの興味のある内容の記事だけを高精度に選び出します。情報洪水が顕著になってきた今、様々な情報源に対しての適応が期待されます。

このほか、音声認識・音声合成技術などの組合せにより、更に様々な応用へと発展させることが可能です(図4)。

インターネットやPCの普及に伴い、情報洪水とも言えるほど情報が氾濫(はんらん)しつつあります。同様に、メールやチャット(PC間でのメッセージのやり取り)などネットワークを介してコミュニケーションを行う機会も増えてきています。将来は、自然言語による命令を解して、ユーザーの知的活動を支援してくれる電子的な秘書エージェントの実用化も夢ではありません。

研究開発センター
知識メディアラボラトリー 室長

住田 一男