

解説

文理解に基づく機械翻訳システム†

—ATLAS/II—



内田 裕士††

1. はじめに

ATLAS/II は多言語翻訳を目標とした機械翻訳システムである。この ATLAS/II に基づいて作成された日本語から英語への翻訳を行う日英自動翻訳システム ATLAS/II は 1985 年 6 月に発売され、現在数十のユーザで使用されている。

ATLAS/II は、FACOM M シリーズの OSIV/F 4 MSP 上で動作する。日本語基本辞書約 5 万語、英語基本辞書約 5 万語、科学技術専門用語辞書を各々 25 万語もち、M 380 上で動かす場合に、1 時間に 6 万語の速度で翻訳することができる。

2. 設計思想

ATLAS/II の目標は、ある言語で書かれた文章を理解し、理解した内容を他の言語で表現し直す人間の翻訳過程をモデル化することにある。これは自然言語に対する次のような考え方に基づいている。

自然言語は人間同志がコミュニケーションを行うために作り上げてきた言語である。したがって、その構文的な規約などは人間ならば誰もがもっている意味的な理解や文脈を踏まえた理解能力を十分に駆使し得るとの前提で決められたはずである。こういった言語をコンピュータで解析したり生成したりするためには、コンピュータも人間と同様の能力を備えていなければならないのは当然である。

人間がいわゆる意味解析や文脈解析を簡単になし得るのは、その言語に関する知識はもちろんのこと、言葉の表し得る概念や、それらの概念がどのように結びついてもっと複雑な概念を構成しているのか、など常識や因果関係、人間の行動原理までを含む膨大な知識をもっており、それらの知識を適切に使用することが

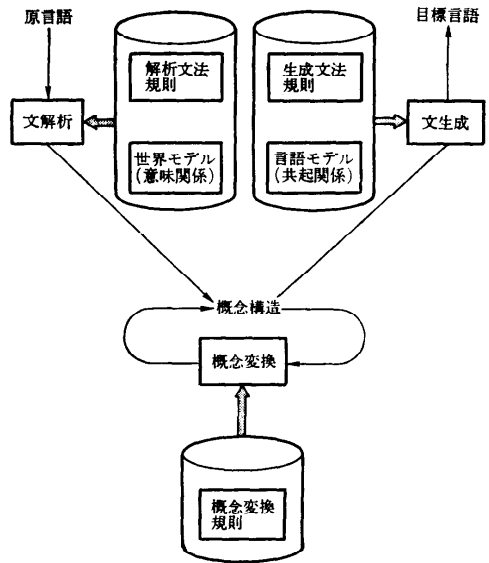


図-1 ATLAS/II の処理概要

できるからである。また、人間が考えたことを自由に表現できるのは、上記の知識に加えて、言葉の用法や言い回しなど、言葉の使い方に関する知識も持っているからである。

ATLAS/II の処理概要を 図-1 に示す。ここでは、人間のもっている言語的な知識を解析文法規則・生成文法規則・辞書の形で記憶し、人間のもっている常識などの知識は概念と概念の間に起こり得る関係を意味関係として表現した世界モデルの形で記憶している。また、言葉の使い方に関する知識は言葉と言葉の間に起こり得る関係を共起関係として表現した言語モデルの形で記憶している。

翻訳処理は 図-1 に示しているように、まず解析文法規則と世界モデルを使用して原言語の文を解析し、中間表現である概念構造に変換する。次にこの概念構造から生成文法規則と言語モデルを使用して目標言語文（翻訳文）を生成する。原言語での言い回しと目標言語での言い回しが異なる場合は、文解析によって得

† A Machine Translation System by Sentence Understanding —ATLAS/II— by Hiroshi UCHIDA (Natural Language Processing Section, Software Laboratory, Fujitsu Laboratories LTD, Japan).

†† (株)富士通研究所ソフトウェア研究部第一研究室

られた概念構造を目標言語になじみやすい概念構造に変換した後で文を生成する。

概念構造は意味ネットワーク形式による入力文の意

味表現である。図-2に「太郎は新しい車を買った。」という文に対応する概念構造を示す。

ATLAS/IIは概念構造という表現形式で入力文を太郎は新しい車を買った。

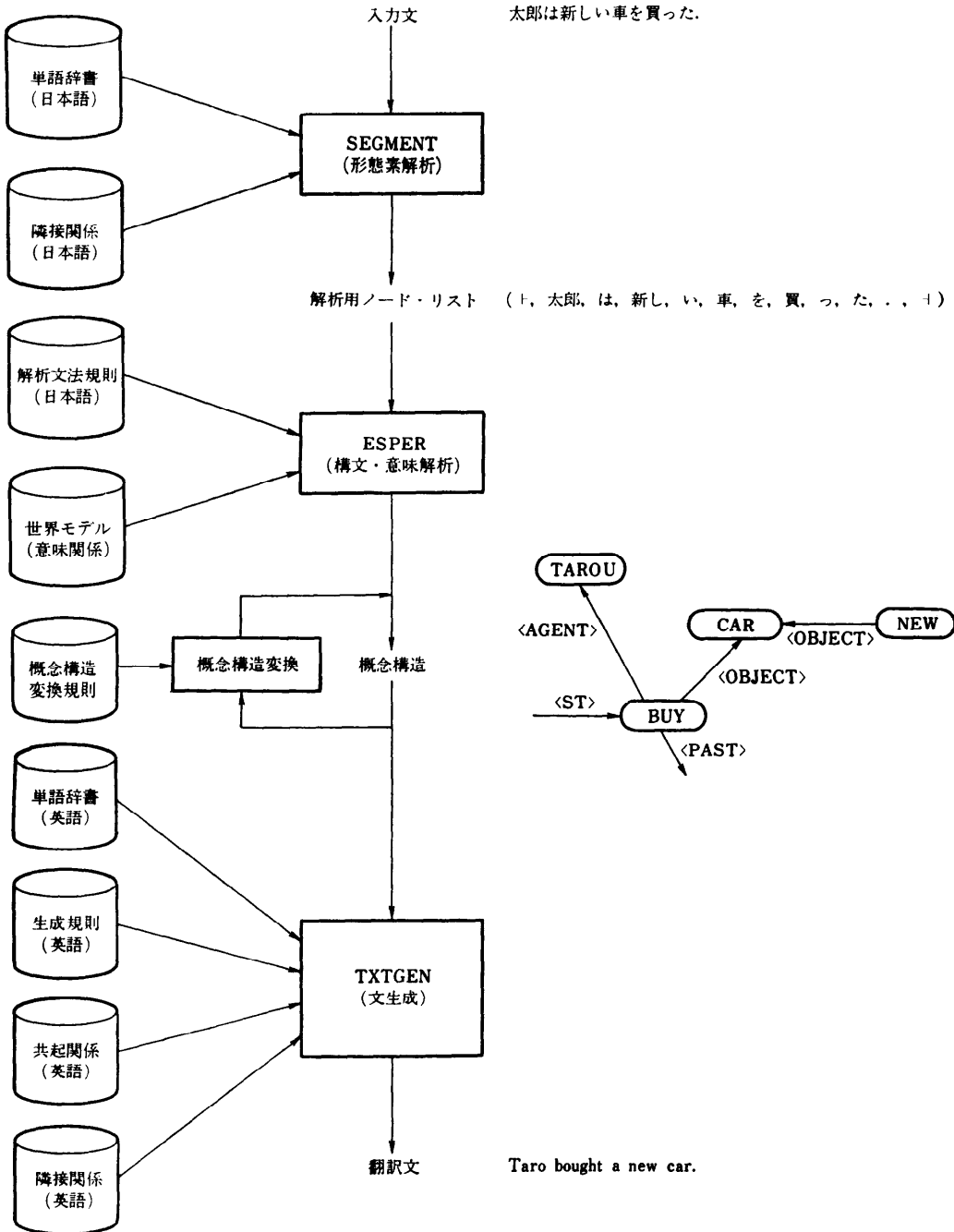


図-2 翻訳処理の流れ

理解する。人間が文を理解するときは、自分のもっている知識に照らし合わせて理解する。ATLAS/IIの場合も人間と同様に、世界モデルとして表現されている知識に照らし合わせることによって文を理解する。世界モデルには、概念間の関係として成り立ち得るすべての関係が概念構造と同様の形式で定義してある。すなわち、あらゆる入力文の解釈として成り立ち得る概念構造のプロトタイプがすべて世界モデルとして定義されている。システムは入力文を解析して得た概念構造が世界モデルの中に含まれるかどうかを調べ、含まれている場合には、その形で理解する。もし含まれていなければ解析結果が正しくないものと判断し、別の解析結果を調べる。

ATLAS/II では以上のような方式で人間の翻訳過程をモデル化している。

### 3. 翻訳処理

ATLAS/II における翻訳処理の流れを図-2に示す。

SEGMENT は、単語（形態素）辞書を参照しながら入力文を形態素に分割する。一般に形態素解析はもともと言語に依存する部分である。しかし本システムでは多言語間の翻訳を目標としているので、言語に依存しない辞書形式と解析方式を採用している。採用した形態素解析法は最尤評価法といい、すべての形態素を含む単語辞書と形態素間の接続可能性を定義した隣接関係行列を用いる<sup>5)</sup>。

形態素解析で抽出された形態素はリスト構造の形で出力される。これが解析用ノードリストである。ESPER は、この解析用ノードリストに対し、解析文法規則を適用し、解析木を生成すると同時に概念構造も作り上げる。すなわち、ESPER は、文脈自由文法を基本とし、構文操作と意味操作を関係づけた解析文法規則を使って入力文をボトムアップに解析する<sup>2)</sup>。構文操作の結果が解析木として表され、意味操作の結果が概念構造となる。

概念構造において、概念間の関係子はかなりの程度まで普遍的なものに定めることができる。しかし概念そのものは、文化的な背景がある以上、すべてを普遍的な概念で構成するわけにはいかない。したがって入力言語に依存した概念を含む概念構造は、意識に相当する過程を経て、目標言語向きの概念構造に変換しなければならない。これを行うのが概念構造変換である。

概念構造変換はネットワークを変換する概念構造変換規則を概念構造に適用することによって行われる<sup>4)</sup>。

訳文の生成は、意味ネットワーク形式の概念構造を一次元の単語列に直接変換する方式をとっている<sup>3)</sup>。このようにすると、疑問や受け身など言語に依存した構造の変形や形態素の変形などを行わなくて済み、生成メカニズムだけでなく、生成文法規則に関しても言語への独立性が高められるという利点がある。この方式のもう一つの特徴は、構文生成と形態素生成を、同一のメカニズム、同一の形式の規則で同時に行うことである。このため、文生成メカニズムは簡潔になる。訳文の生成は概念構造を構成している概念を表す目標言語の単語を適切な順序で出力していくことによって行われる。ただ、一般にある概念を表し得る単語は複数個あるので、その選択を行うために共起関係と隣接関係が使用される。共起関係は、単語と単語の語用論的な結び付きに関する束縛条件を与え、隣接関係は単語の並び方に関する束縛条件を与えるものである。

### 4. おわりに

現在の機械翻訳システムは、翻訳する際に使用する情報量を人間が翻訳する際に使用する情報量と比べると機械翻訳システムの方が圧倒的に少なく、まだまだ不完全なものである。したがって現時点では、前編集、後編集、辞書編集など機械翻訳を支援する機能が重要な役割を果たす。ここ当面は、機械翻訳技術を向上させる努力と並行して機械翻訳を使いこなすための技術の開発にも力を注いでいく必要があると思っている。

### 参考文献

- 1) 内田, 小部, 西野, 増山, 松井: 日英機械翻訳システム ATLAS/U, 情報処理学会, 自然言語処理研究会資料, No. 29-3 (Jan. 1982).
- 2) 内田, 増山, 小部: 意味解析向きの自然言語パーサ, 情報処理学会, 自然言語処理研究会資料, No. 34-3 (Dec. 1982).
- 3) 内田, 西野: 概念構造からの自然言語文生成, 情報処理学会, 自然言語処理技術シンポジウム資料 pp. 57-65 (Jun. 1983).
- 4) 内田, 増山: 機械翻訳における概念変換について, 情報処理学会, 自然言語処理研究会資料, No. 39-3 (Sep. 1983).
- 5) 内田, 杉山: 自由入力形式のカナ漢字変換, 情報処理学会, 自然言語処理研究会資料, No. 27-3 (Sep. 1981).

(昭和60年9月11日受付)