

英日機械翻訳システムにおける 語彙的意味の記述について

三池誠司 平川秀樹 天野真家

(株)東芝総合研究所

我々は、科学技術文献等を対象にした英日機械翻訳システムを開発した。実用英文の翻訳においては、広範な言語現象の調査とそれらに対応できる柔軟性の高い機械文法の枠組が必要である。我々は、機械文法の枠組として、語彙遷移ネットワーク文法を用いている。語彙遷移ネットワーク文法の特徴の1つは、意味処理が語彙的意味に基づいて行われることである。

本稿では、英日翻訳において意味的处理を必要とする言語現象を列挙し、それに関する語彙的意味の記述とその処理について述べる。語彙的意味の記述は次のように分類される。

- (1) 英語概念構造中の係り受けに関する記述
- (2) 英語概念構造中の意味解釈に関する記述
- (3) 英語概念構造から日本語概念構造への変換に関する記述

これらの語彙的意味記述を用いることにより高品質翻訳が可能である。

On Lexical Meanings in Semantic Analysis in an English-Japanese Machine Translation System

Seiji Miike Hideki Hirakawa Shin-ya Amano

Toshiba Research and Development Center
1, Komukai Toshibacho, Saiwai-ku, Kawasaki-City, Kanagawa, 210 Japan

We have developed an English-Japanese machine translation system for scientific and technical documents. In order to treat realistic documents, the investigation of many linguistic phenomena and a flexible framework for writing the system's grammars are both required. the system uses the Lexical Transition Network Grammar (LTNG) proposed by the authors. One of the characteristics of LTNG is that the semantic analysis is based on lexical meanings.

This paper describes the linguistic phenomena which require semantic analysis, and discusses the descriptions of lexical meanings. The descriptions of lexical meanings are classified as follows:

- (1) descriptions of dependency in English structure
- (2) descriptions of the interpretation of English structure
- (3) descriptions of the transfer of English structure to Japanese structure

The incorporation of the lexical meanings greatly improves the quality of translation.

1. はじめに

我々は、科学技術文献、マニュアル等を対象とした英日機械翻訳システムを開発した。実用英文の翻訳を行うためには、広範囲の言語現象を処理するための柔軟性の高い文法の枠組みと、広範囲のテキストを処理するための大規模な辞書（数万語以上）が必要である。我々は、機械文法の枠組みとして語彙遷移ネットワーク文法[1]を用いており、その特徴の1つは、意味処理が語彙に基づいて行われることである。これは言い換えれば、意味処理情報が語彙項目（辞書）に記述されることであり、翻訳処理における辞書情報は非常に大きな役割を担っている。本稿では、この語彙の意味処理に焦点をあて、意味的な規則に基づく言語現象の例とその意味記述と処理過程について述べる。統語的な規則に基づく言語現象とその解析方法については文献[2]を参照されたい。

第2章では、語彙遷移ネットワーク文法について述べる。第3章と4章では、意味的処理を要する言語現象と語彙の意味について述べ、第5章では、翻訳処理過程について述べる。第6章で今後の課題を述べる。

2. 語彙遷移ネットワーク文法について

我々は、次の意味モデルを提案した。

文の意味は統語構造の集合へ語彙の意味の集合を作用させた結果である。

この意味モデルに基づき統語-意味分析システムとして語彙遷移ネットワーク文法を実現した。語彙遷移ネットワーク文法は、次の特徴をもつ。

- (1) 一般に構文解析は長時間を要するが、本文法では、文法規則に意味規則をもちこまずに、品詞により純粹に統語的な解析のみを行うことによって構文解析の負担を軽くする。構文解析で用いる文法を一般文法とよぶ。さらに、意味処理は計算の複雑さが線形である部分で行うことによって全体の計算のコストダウンを実現する。
- (2) 意味は基本的に語彙的であるという立場をとり、語彙の意味に基づいて意味処理を行う。この意味規

則を語彙文法とよび、語彙的であることから、各語彙に記述する。語彙文法による意味処理では、次の2つを目的としている。

- (1) 一般文法で解消されない構文的曖昧さを解消する。
語彙遷移ネットワーク文法の特徴にあるように、構文的曖昧さを含む文を解析する場合、構文解析ではこの曖昧さを内在する唯一の解析結果だけを構築し、後の意味解析で曖昧さを解消する。この唯一の解析結果を基準概念構造とよぶ。
- (2) 語および句の意味を決定し、構造を変換する。
英語における意味解釈と英日変換において、語彙文法により処理を行う。

3. 英語における言語現象

英語における意味的処理を要する現象には、構文的曖昧さに関する言語現象と、意味解釈に関する言語現象とがある。本章では、それらの言語現象の例をあげる。

3-1 係り受けに関する言語現象

品詞という縮退した世界において、その品詞の並びにより英語統語構造中に複数の可能な係り受け関係が存在しえる。ここで可能な係り受け関係とは、係り受けの非交差条件を満足しているという意味である。この複数の可能な係り受け関係の中から、基準概念構造の中の語そのものがもつ語彙の意味を用いることによって係り受け関係を決定することができる。このように語彙の意味により、係り受け関係に関して基準概念構造を概念構造へ写像することができる。

係り受け関係の曖昧さに関する言語現象の中で、ここでは、前置詞句における係り受けについて考察する。統語的には前置詞句は名詞と動詞に係ることが可能である。このため、前置詞句は複数個の名詞および複数個の動詞にかかる可能性をもつ場合がある。このような場合、次のような語彙の意味により、前置詞句の係り先を決定できる場合がある。

(1) 格フレーム

動詞やある種の名詞などの格フレームは、それらと前置詞との係り受けを決定する1つの語彙的意味である。次にその例をあげる。

① transfer; 動詞

「transfer A from B to C」において、基準概念構造を図1(a)に示し、transferの格フレーム情報により、意味的に解釈された結果生じる概念構造を(b)に示す。ここで、名詞A, B, Cは物、場所などの意味をもつ名詞に限られる。

(例) That transfers the instruction from the program memory to the microprocessor.

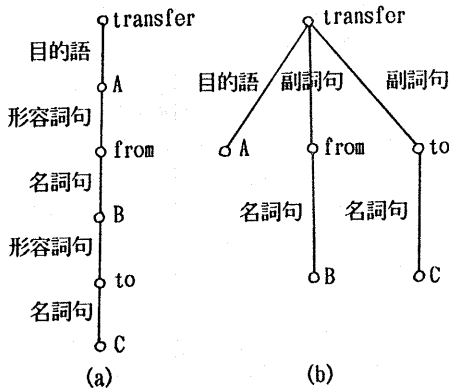


図1 transferの格フレームによる構文的曖昧さの解消

② ratio; 名詞

いわゆる動作性名詞を含むある種の名詞は動詞と同様の格フレームをもち、動詞と同様の処理が行われる。

「ratio of A to B」の構造において、基準概念構造を図2(a)に示し、ratioの格フレーム情報により、意味的に解釈された結果生じる概念構造を(b)に示す。

(例) The ratio of each peak value to the average value is calculated.

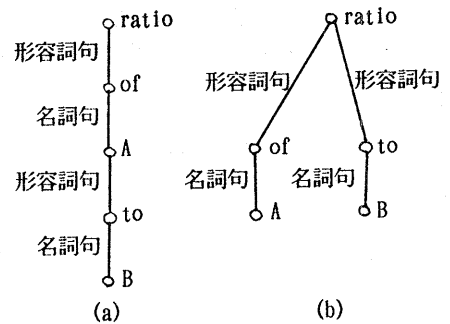


図2 ratioの格フレームによる構文的曖昧さの解消

(2) 前置詞が支配する名詞の意味

動詞の任意的な格要素となる前置詞句に関しても、前置詞句が支配する名詞の意味により、係り受けの可否を決定することができる。

① at; 前置詞

「A(動詞) ~ at B(名詞)」の構造において、前置詞atは、割合・程度の意味をもつ名詞を支配し、その前置詞句は動詞を修飾する。

(例) The fuel is controlled to provide an equal flow at rate consistent with the load requirements of the turbine.

3-2 意味解釈に関する言語現象

品詞という縮退した世界において、その品詞の並びにより生成される統語構造は常に同じ統語構造である。しかしながら、統語構造中の語そのものがもつ語彙的意味を用いることによって統語構造を概念構造へ変換することができる場合がある。ここでは、英語単独の言語現象で、語彙的意味による統語構造から概念構造への変換を意味解釈とよぶ。本節では、意味解釈に関する言語現象の例をあげる。

(1) 現在分詞の目的語の解析

例文①において、下線部は現在分詞と副詞句の並びである。ここで、副詞句中のofの語彙的意味に従えば、ofが支配する名詞は現在分詞の目的語である。

図3において、(a)は統語構造の部分構造を示し、(b)はofの語彙的意味により意味解釈された結果生じる概念構造を示す。

(例) - introduction of predicting programs
recording of the condition of the equipment

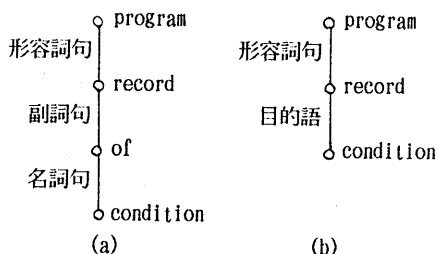


図3 現在分詞の目的語の解析

(2) 不定冠詞による名詞の意味の決定

名詞の意味の意味を決定する手掛かりに、可算と不可算という範疇が存在するが、名詞の中には、可算の意味と不可算の意味の両方をもつものがある。このような名詞に不定冠詞が依存する統語構造において、不定冠詞がその名詞の意味が可算の意味であることを決定する。すなわち、英語概念構造の中で、その名詞の意味が絞られる。その結果、ある言語対において、絞り込まれた英語の意味に対して目的言語の訳語を得ることができる。例えば、例文①では1枚の「線画」という訳語を得ることになる。

(例) This is a drawing of the plant.

(3) 強意形容詞の機能の解析

形容詞の中には、形容詞が依存する語に対して意味を強めたり弱めたりする効果をもつものがある。

例えば、「particular」という語は「指示形容詞に添えて意味を強める」という言語現象をとる。例文①は「特にこの定義に〜」という意味をもっている。

(例) Since its use is to be restricted to this particular definition of URGENT-MESSAGE, it itself is not generic.

4. 英日変換における言語現象

本章では、語が対応する「概念」に関して語のレベルでの変換という観点から、英日変換において現れる言語現象の分類と例を示す。その中には、意味の決定と構造変換に関する言語現象が含まれる。

4-1 単純語彙変換の場合

単純語彙変換とは、語対語の翻訳を意味する。ここで、語とは一般的に英語では単語・熟語を、日本語では自立語を意味する。

(1) 英語の語の意味を決定できる場合

(i) 英語の語の意味が一意である場合

英語と日本語で、それぞれの語が一意に対応する概念が共通である。すなわち、ある概念を媒介として一対一に対応する語が英語と日本語に存在する場合である。このような単語には、deskやbookのような一般用語とwheel やgearのように分野によって意味を一意に決定できる専門用語とがある。一般には、対応する日本語は表記上の違いなどで一つとは限らない。

(ii) 文脈等により英語の語の意味が決定できる場合

英語概念構造中の係り受け構造や語のもつ素性により、語の意味を決定して、あるいは絞って日本語の語に変換することができる。例えば、「形容詞において、叙述的用法と限定的用法で意味が異なる」、「名詞において、単数形と複数形で意味が異なる」というような言語現象である。

次に例を示す。例において、まず英単語と品詞を並べて示す。C1, C1-1のようにCで始まる記号の項は、その概念を表わす日本語を示す。E1, E-1のようにEで始まる記号の項は、その概念における英語の言語現象を示す。J1, J-1のようにJで始まる記号の項は、それに対応する日本語の言語現象または訳語を示す。さらに、その言語現象を含む英文と一部について下線部分の日本語とを示す。

① free; 形容詞

C1; 制約がない、制限がない

E1: 「free from」の構造

J1: 「～がない」

(例) - a method so adapted as to obtain
correct value of the signal free from noise
ノイズがない

c1: 自由な

E2: 限定的用法

J2: 単語の置換え

(例) A program unit must be written in
either fixed form or free form.
自由形式

② potential: 形容詞

C1: 潜在的な、可能性のある

E1: 「potential to-inf」あるいは
「potential for -ing」の構造

J1: 「～できる」

(例) - a design potetial for increasing
the unit power of the reactor
増加させることができる

(2) 英語の語の意味を決定することが困難である場合

英語の語の意味が一意でなく、文脈等の情報を利用できない場合である。しかし、この場合でも、翻訳の対象としている分野を考慮することにより、訳語の候補あるいはその順序に選好性（訳語の好ましき）の要素をとり入れて訳語を決定することができる。

例えば科学技術文の翻訳では、語源的な意味、派生的な意味や比喩的な意味の語の選好性は低いと考えられる。例①では、absence に対応する概念に対して、日本語では、その概念の部分集合の概念の各々に対応する日本語が存在している。従って、それらの日本語の中から訳語を選択するためには修飾する名詞の意味素性等を参照する必要がある。しかし、文脈によっては、必ずしも常に適切な日本語を決定できるとは限らない。そこで、翻訳の対象を科学技術文献とする場合、訳語の選好性の点から 4. の意味を第一義とすることができる。

① absence: 名詞

C1: 不在、留守

C2: 不在期間

C3: 欠席、欠勤

C4: 欠如、不足

C5: 放心

(例) The distinguishing features is the
absence of a pressurized body.

4-2 構造的語彙変換の場合

構造的語彙変換とは、英語概念構造の変形が必要であることを意味する。

(1) 普遍的概念に対する表現の違いに基づく概念構造の変換の場合

例えば「時刻」のような概念に対しては、英語と日本語にそれぞれその概念を表す表現があるが、その表現形態が異なる。このような場合である。

(2) 語に基づく概念構造の変換の場合

(i) 英語では語であるが、日本語では語でない場合

例えば、「英語では単語であるが、日本語ではモダリティなどの述語表現に相当する」のような言語現象である。

例①では、英語概念構造を変換すると同時に、日本語概念構造に、モダリティと否定を用いている。

① likely: 形容詞

C1: ～しそうである、～しそうな

E1: 「likely to-inf」の構造

J1: モダリティ（様態）で表す。

(例) It would be efficient to select that
leaf node most likely to change the
estimate of the best solution graph.
最も変えそうな

(ii) 英語の語と日本語の語が 1 対 1 に対応せず、概念構造の変換が必要である場合

この場合には [1] ～ [3] の 3 通りがある。その中で [1] と [2] は、一方の言語では一語で表現できる概念が、他方では一語で表現できないという言語現象を意味する。

[1] 1対多の対応

英語では1つの語で表現する概念を、日本語では複数の語で表現する。

[2] 多対1の対応

英語では複数の語で表現する概念を、日本語では1つの語で表現する。

例えば、日本語の「前半」は1語であり、英語では2語の単語で表現する。

① half; 名詞

C1; 半分、2分の1

E1-1; 「the (first, former) half」

J1-1; 「前半」

(例) During the first half of the cycle, the address is clocked into the register

E1-2; 「the (second, latter) half」

J1-2; 「後半」

[3] 多対多の対応

語の並びすなわち句のレベルで1対1に対応する言語現象である。例えば、「英語の2つの単語の並びが、日本語では接頭語と自立語に相当する」のような言語現象である。

例①において、decadeは1語で「10年間の長さ、あるいは10の単位」という概念に対応する。日本語の1語でこの概念に対応するものは見当たらない。この場合には、「数詞+decade」の構造に対して、数値の処理を行う必要がある。

① decade; 名詞

C1; 10年間、10の単位

E1; 「数詞 decade」の構造

J1; 「x年間、x単位」

ここでxは、『数詞×10』の値である。

(例) The silicon detector is used to measure linearly up to six decades of light sensitivity.

60単位

5. 翻訳処理過程について

本章では、まず語彙遷移ネットワーク文法に基づく英日翻訳システムにおける翻訳処理過程について述べる。さらに例文を用い、各翻訳処理過程で用いられる語彙的意味を示す。

5-1 翻訳処理過程の概要

英日機械翻訳システムにおける処理過程を次に示す。

1 英語解析過程

(1) 英語形態素解析過程

(2) 英語構文解析過程

(3) 英語意味解析過程

(i) 係り受け調整

(ii) 意味解釈

2 英日変換過程

(1) 英日語彙変換過程

(2) 英日一般構造変換過程

3 日本語生成過程

(1) 日本語構文生成過程

(2) 日本語形態素生成過程

図4 英日機械翻訳システムにおける処理過程

図4の1において、構文解析過程は品詞に基づいて入力英文の統語構造を解析する。この意味解析過程の係り受け調整では、3-1節で述べた言語現象を対象とした意味処理を行い、意味解釈では、3-2節で述べた言語現象を対象とした意味処理を行う。

図4の2において、語彙変換過程は、語彙的意味に基づいて英語概念構造を日本語概念構造に変換する。一般構造変換過程は、語彙に依存しない英日の構文上の違いに基づいて、英語概念構造を日本語概念構造に変換する。語彙変換過程において、第4章で述べた言語現象を対象とした意味処理を行う。

5-2 翻訳処理過程と語彙的意味

例文を次に示す。各例文における部分構造の変化を図5と図6に示す。各図において、(a)は英語構文解析過程の出力である英語統語構造を示し、(c)は英日変換過

程の出力である日本語概念構造を示す。図5(b)は英語意味解析過程(係り受け調整)の出力である英語概念構造を示し、図6(b)は英日語彙変換過程の出力である英語概念構造を示す。

- (例1) Users can use the command with no change.
 (例2) These data frames include only the data words and no parity bits.

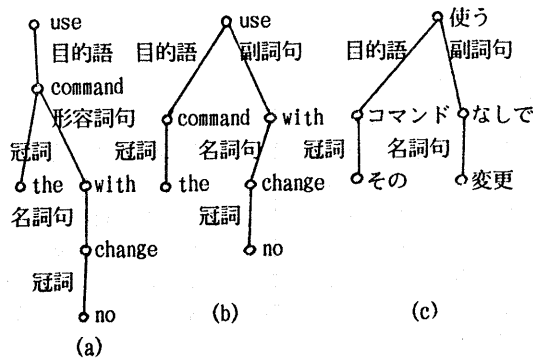


図5 例1における語彙の意味処理

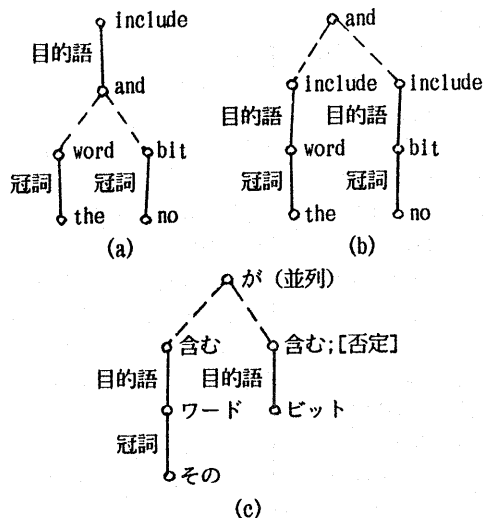


図6 例2における語彙の意味処理

図5では、『「with 名詞(動作性)」の構造は動詞を修飾する』、および『英語の「with no 名詞」の構造は日本語で「名詞なしで」の構造に対応する』という

「with」の語彙的意味に基づいている。これらの語彙的意味は、語彙項目に記述されており、図5の各処理過程において、それぞれの語彙的意味に基づいて処理を行う。

「no」は、『noが動詞の必須格要素を修飾するならば動詞に属性として否定を与える』という語彙的意味をもつが、図5の英語統語構造はこれに当てはまらない。さらに、図6の場合には、「no」は並列名詞句の一部の名詞のみを修飾している。この場合には、日本語では一般に否定は用言で表現するという表現の違いから、これに基づいて英語概念構造を日本語概念構造に変換している。

また、第4章では、語が対応する「概念」に関して語のレベルでの変換について述べたが、英日変換過程において、それらの変換と同時に、英語を自然な日本語にするための変換も行っている。この処理規則もこれまで述べた語彙的意味と同様に語彙項目に記述している。その処理の例を次に示す。例文と例文における部分構造の変化を示す。図7(a)は英語意味解析過程(係り受け調整)の出力である英語概念構造を示し、(b)は意味解釈の出力である英語概念構造を示し、(c)は英日語彙変換過程の出力である日本語概念構造を示す。

- (例3) The procedure explicitly required the level to be lowered to 12.4 meters.

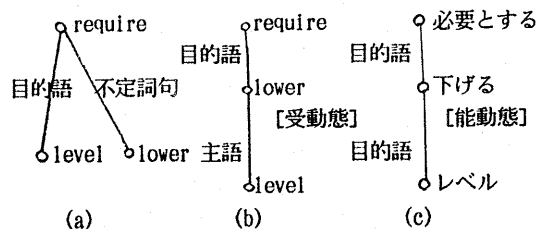


図7 例3における語彙の意味処理

図7では、英日語彙変換過程において、require の語彙文法により、lower の態を受動態から能動態に変換し、level をlower の目的語に変換している。

次の例4では、require の語彙文法により、英日語彙変換過程において、「AはBが～することを可能にする」という表現ではなく、「AによってBは～できる」という表現にする。英日語彙変換過程における英語概念構造から日本語概念構造への部分構造の変化を図8に示す。

(例4) The command language enables users to extend their command repertoire easily.

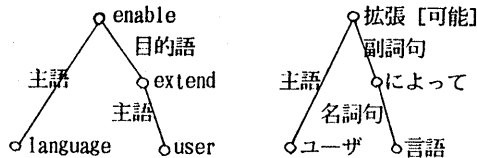


図8 例4における語彙の意味処理

6. 今後の課題

本稿では、語彙の意味に基づく処理について述べてきた。第4章で英日変換における言語現象について考察したが、第5章で触れたように、より自然な日本語へ変換することが今後の目標であり、そのためにはさらに自然言語に対する研究が必要である。今後の課題のいくつかを次にあげる。

例えば、potential という語に対して、「潜在的な」のような日本語で置換えればよい場合も多いが、次の英文では、potential は「障害になる可能性のある物体」あるいは「障害になりえる物体」という意味を構成する。

(例1) The sensor detects a potential enemy object.

現在のシステムのレベルでも例1をこのように翻訳することは可能であるが、potential という語が用いられているあらゆる文に対して、常に適切な日本語を生成するには、大量の実験データに基づく語彙の意味の調査と、これらの語彙の意味を適切に適用するシステムの機能の研究が課題である。

次は、文脈の情報を必要とする例である。例2は3文

(例2) At your work station, keep cables of your micro out of the way so that people can not trip over them. Guide them down behind the desk and to a distant socket through cable bridge. Wind an excessive length into loops and secure it.

からなる。3文目の下線部分の翻訳には、「長いケーブルは」、あるいは「もしケーブルが長ければ」という日本語が好ましい。「ケーブル」という語を生成するには文脈の情報が必要である。しかし、この場合においても、「不定冠詞+形容詞+名詞」の構造が1つの条件節に相当するという英語の言語現象をみることができる。この構造のみでそのように変換することはできないが、機械処理のための重要な情報の1つである。このような文脈処理は今後の課題である。

7. おわりに

語彙遷移ネットワーク文法では、時間を要する統語解析では統語構造の解析のみを行い、次に語彙の意味を用いて概念構造を生成する。このことにより、処理速度を損なわずに、かつ高品質な翻訳を実現することができる。語彙の意味を、言語現象に応じてそれぞれのレベルで各語彙に記述し、システムは段階的にそれらのルールを適用する。

【参考文献】

- [1] 天野、野上、三池：語彙遷移ネットワーク文法について、情報処理学会自然言語処理研究、49-4、1985
- [2] 野上、杉浦、平川、天野：英日機械翻訳システムにおける英文の解析方法について、情報処理学会自然言語処理研究、61-4、1987
- [3] 中里、堤、平川、天野：日英機械翻訳システムにおける日本語解析について、情報処理学会第32全大、1986
- [4] 熊野、小山、平川、天野：日英機械翻訳システムにおける構造変換について、情報処理学会第32全大、1986
- [5] 久米、辻井、長尾：Muプロジェクトプロジェクトにおける英日翻訳システムの構成と翻訳結果、情報処理学会自然言語処理研究、59-6、1987
- [6] Nagao, M.: Role of structural transformation in a machine translation system: Machine translation, Cambridge University Press, 1987
- [7] 堤、西嶋、寛：英日機械翻訳システムSHALTにおける英日トランスファー、情報処理学会自然言語処理研究、53-4、1986
- [8] Brouhard, 水上：技術英語基礎講座 JSTC, 1987
- [9] 國廣：意味と語彙 大修館書店, 1982