

機械翻訳における多義性の諸問題

— 機械翻訳の射程 —

成田 一

大阪大学言語文化部

narita@lisa.lang.osaka-u.ac.jp

言語形式によっては複数の構造へ解析可能だが、
どれかに絞り込まないと翻訳ができない。実際には
文脈などを考慮しないと構造が分からないもの
も多く、翻訳の壁となっている。この壁を破るシ
ステム化可能な言語的方法を考察する。

機械翻訳の射程

構造解析の曖昧性は、機械だけではなく人間にとっても、重要な問題である。認知言語心理学では、(1)文は言語レベルでいくつかの可能な構造に解析され、(2)その中から文脈に適合する構造が統合レベルにおいて選択される、とする仮説が有力になっている。人間の脳中において2段階の処理が仮定されるならば、「言語構造の解析処理を機械に委せ、文脈情報や知識情報を動員する構造選択は人間が行なう」という、機械翻訳が暗黙の了解として採ってきた枠組は、実は認知的な合理性を持っていたことになるだろう。

これは文脈や知識を扱うシステムが構築できない実情を踏まえた、やむを得ない枠組であったが、人間の言語処理過程にも対応するというのであれば、ネガティブに捉える必要はない。ただし、一般的には構造解析が曖昧であるという場合、全て文脈、知識といった高次処理の問題として放置され、曖昧性を解消する方策をまともに研究することが少なかった。しかし、実際には言語的レベルで解消可能な曖昧性も多い。

ここでは、現段階での翻訳ソフトの構造処理能力を基礎として、解析が困難な言語形式をいくつか選び、解析上の問題点を検討の上、言語レベルの手がかりだけで構造を絞り込む方策を考える。なお、翻訳例は英語文法データを市販の翻訳ソフトで処理したものの一部である(脚注参照)。

検証ソフト ログヴィスタ「LogoVista E to J」カテナ「コリヤ英和!」ノヴァ「PC-Transer ej」三洋「トランス・サポーター EJ」(ザ・国際人) 沖「パンセ(PENSEE)」富士通「アトラス(ATLAS Win)」高電社「j・London/EJ」(NECライセンス)

曖昧な構造の解析

翻訳ソフトの解析例をみると、潜在的に複数の解析を許す言語形式は、ほとんど全ての解析がみられることが分かる。ここでは、準動詞句、名詞修飾節の中から代表例を取り上げ、適切な解析をするための言語的条件について検討する。

準動詞句

[V...to V^{inf}...]の機能

これは副詞用法の不定詞句の機能をいかに決定するかという問題である。不定詞句は動詞を修飾すると解析された場合、「目的」と「結果」のいずれの機能を担うかを決定しなければならないが、翻訳ソフトでは普通「目的の副詞」として扱われる。I woke up [to find myself on the sofa]. は「結果」を表すのが正解だが、だいたい「私は、ソファ(の上)に/で私自身を見つけるために目を覚ました/起きた」(ノヴ/富士/高電)のように訳されている。一般に、こうした機能の判断は「意味関係の理解がなければ無理」と諦められがちだが、言語的情報だけでも十分対応できる。

(A)主文の動詞 **or**

(B)不定詞句の動詞

が「意志で制御できない」場合、

→ 不定詞句を結果として訳す。

このように設定すれば不定詞句が適切に翻訳できる。システム的には動詞の素性情報に依拠して動作させるのが望ましいが、簡単な措置としては、辞書で所定の動詞に不定詞を「結果」として訳す

九州松下「翻訳工房」東芝「エーエストランザック(ASTRANSAC)」シャープ「デュエット(DUET)」日立「ハイキャッツ(HICATS)」

★英語文法データ

[基本390例; 総合2478例; 補助440例]

ように指定しても良い [「私は、目覚めて気がつく」とソファーにいる」(シャ)]。いずれにせよ、知識処理機構にアクセスするには及ばない。

[V_{ing} ...]の解析

[V_{ing}] という形態の語は、目的語などを伴わずに単独で使われると、単に ing で終る(A)名詞として解析されるか、(B)動名詞ないし(C)分詞 [進行形] など動詞の文法形式として解析される可能性がある。実際の翻訳ソフトでは名詞解析が優勢だが、この形態の語として独立に辞書登録されていれば、それが妥当な解析ということになる。[Jogging] is very good for our health. は「ジョギングは私達/我々の健康に(は)非常に良い」(カテ/富士/高電/シャ/日立)のほか「ジョギングすることは我々の健康のために非常に良い」(ロゴ)と動名詞に訳されている。

しかし、[V_{ing} ...]という構造がBE動詞に続く場合、その機能を適切に判断できるソフトはない。「連結辞の 'BE'」の後では V_{ing} が動名詞であっても、もっぱら進行形として解析され、His work in the morning is [reading a newspaper]. は「彼の、朝の仕事は新聞を読んでいる」となる。しかし、これは動詞 read が [人間] を主語に取るという基本的情報を辞書に記載し、それぞれの名詞にもそうした意味特性を記載していれば、問題なく解析できる構造である。

注 実際、the [sleeping] car. を「寝台車」、the [sleeping] girl. を「眠っている少女」と訳し分けるソフトもある。この機能の決定は、本質的には、名詞との意味的依存関係 [選択制限] に基づいて行うのが適当だが、もっと簡便に対応できる解析手続きもある。sleeping car/bag など、(1)少数の語結合を複合語として辞書登録し、(2)これ以外の語結合は全て名詞修飾の分詞に解析するように設定すれば、意味的依存関係に依らなくても正しい構造解析が得られる。

[動詞 名詞 V_{ing} ...]の解析

主文動詞の目的語が動名詞の「意味上の主語」となるのは珍しいことではない。しかし、「目的語が代名詞」の場合は問題が起る。たとえば、My father insisted on [my studying] at home. は「私の父親は家において私の勉強を強く要求し

た」のように動名詞が名詞として解析される。また、口語体によく見られるように、my が me になると動名詞句が分詞構文に解析され、「私の父親は私の部屋で勉強して私を強く要求した」という訳も見られる。「me が前置詞の目的語に判断された所で主文が終了する」と解析されている。

こうした解析の曖昧性には「動詞や名詞の種類ないし意味特性」などの情報を利用することによって解消できるものもある。「主張などを含む動詞」の場合、目的語の位置には「行為名詞」ないし「補文構造」がなければならない。このため、制約に違反しない解析に絞られる。また、主文が知覚・使役動詞の場合にも同じ制約が適用できる。知覚構文の She watched her son sleeping in the bed. は「彼女は [彼女の息子がベッドで眠っている] のを見た」、使役構文の John got his car stolen. は「ジョンは [彼の自動車が盗まれる] ようにした」となる。

[名詞 V_{ing} 名詞]の解析

前置型準動詞句 [[N₂ V_{ing} N₁]]

「目的語に先導される V_{ing} が主名詞を修飾する」この言語形式の場合、N₂ は「特定対象を指示できない」という制約が課せられる。「犬」や「猫」を一般的に表すことはできても、「その猫」とか「隣の犬」は指せない。英語では、冠詞が付いたり複数形に活用することがなく、N₂ は名詞だけの構成 ([[oil producing]well]「油井」)になる。なお、これは複合語に関する一般制約で、日本語でも「[野犬]狩り」は言えても「[この野犬]狩り」は許されない。

後置型準動詞句 [NP₁[V_{ing} NP₂]]

類似構造を示すものに「主部の後に V_{ing} が続く」言語形式がある。目的語を従える分詞句が主部の名詞句を逆行的に修飾する構造だ。この場合、主部にも目的語にも特に制約はない。冠詞が付いて複数にもなる。どちらも [限定詞+名詞] の構成 ([[the program][controlling the process])になる。したがって、準動詞の後の名詞に限定詞が付けば「後置型」に解析する設定で良い。

前置型か後置型か?

目的語が固有名詞や複数名詞の場合、NP₂ に限定詞がないこともある。仮に主部の NP₁ も限定詞を取らないとすれば、形式的には[N V_{ing} N] ([[techniques][underlying Ethernet]])とな

り、「前置型」と区別できない。ただし、最初の名詞が活用も欠いているならば、これは準動詞の目的語であるという判定が下せる。目的語をとる動詞を準動詞とするこの2つの言語形式は、表面的には似かよっているが、基本的には形態および構造的な手がかりによって解析できるのだ。

名詞修飾節

内容節

内容節は主部の内容を具体的に表すが、節中には欠如している必須格要素がない。英語の補文化辞は **that** が標準で、内容節を直接主部につなぐ翻訳で済む場合もあるが、補文導入要素「という」がなければ不自然になることが少なくない。そこで、これを介在させる設定にする。例えば、His claim that he is not guilty is groundless. は「[[彼が有罪でない]という]彼の主張は根拠がない」といった訳になる。主部の名詞によっては **whether** が内容節を率いるが、この場合、「かどうか/か否か」を「という」の前に補充する設定にすると自然な訳文が得られる(次例参照)。

We must answer the question whether we should resist the pressure.

ロゴ「我々は我々が圧力に反抗するべきであるかどうかの質問に答えなくてはならない」(○)

ノヴ「我々がその圧力に抵抗しなければならないかどうかに関係なく、我々はその問題に答えなければならない」[副詞節]

カテ「我々は我々が圧力に抵抗するべきであるという、問題に...」(△)[「である」は不適切]

沖「我々が圧力に抵抗する必要があるかどうか、我々は質問に...」(○)

富士「我々は、我々が圧力に耐えるべきであるかどうかという質問に...」(○)

三洋「我々は、圧力に抵抗するべきかどうか質問に...」(○)[「べきだ-か」は不適切]

高電「私達は、私達がプレッシャーに抵抗するべきであるかどうかという問題に...」(○)

シャ「我々がプレッシャーに抵抗するべきであるかどうかにかかわらず、我々は、質問に...」[副詞節]

東芝「我々は圧力に抵抗するべきであるかどうかという質問に...」(○)

日立「我々は、我々が圧力に抵抗するべきであるかどうかという質問に...」(○)

この **whether** という語には副詞節用法もあるため、「かどうかに関係なく/か否かに拘らず」とする翻訳もみられるが、下記の設定にすれば、適切な表記に従うものは正しく判別される。

(A)カンマが先行する場合に副詞節に分析する

(B)名詞に続く場合に内容節に分析する

しかし、現実には全ての文書においてこの表記が守られていることを期待することは難しい。ここで「補文標識として **whether** を取るような選択を迫る意味特性」を持つ名詞を疑問名詞と規定しよう。そうすると、次のような言語的な依存関係に基づいた解析手続きの設定ができる。

疑問名詞が先行する場合は内容節に分析する

複文関係節

主部と同じ名詞が修飾節の中にある場合には関係節が形成される。英語ではこの名詞に代る関係詞が節頭に現れるのに対し、日本語では同一名詞が削除され何も現れない。このため、英文の(1)節内部に存在する要素を翻訳し、そのまま(2)主部を訳出する、というように浅い解析だけでも訳文が得られる。アルゴリズム的には「内容節と同じ翻訳プロセス」で十分対応できるのだ。もちろん、「述語の格フレーム情報に基づいて、節内の欠如要素を求める」という本格的な深い解析をするソフトも少なくない。

The girl whom you know that I met yesterday is beautiful.

ロゴ「あなた(達)が私が昨日会ったことを知っている少女は美しい」(○)

ノヴ「私が昨日会ったということを、あなたが知っている少女は、美しい」(○)

カテ「私が会ったとあなたが知っている女の子は、昨日美しい」(○)[「昨日」の帰属は誤り]

沖「私が昨日会ったこと、あなたが知っている少女は、美しい」(×)

富士「あなたが知っている少女...私が昨日会った...美しい」(×)

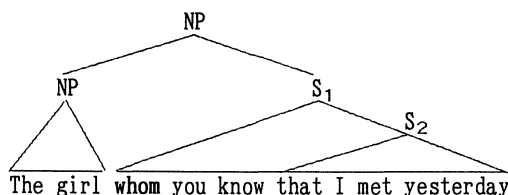
三洋「私が会ったあなたがきのう知っている少女は美しい」[異分析:「きのう」の帰属は誤り]

高電「昨日私が会ったあなたが知っている少女は美しい...」(○)

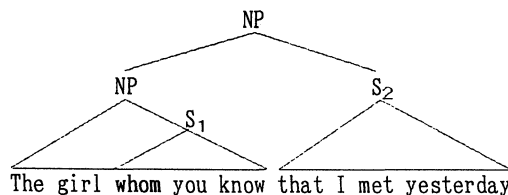
複文関係節の解析

関係節 S_1 が別の節 S_2 を目的語として埋め込む複文の場合、(A)の構造が正しいが、そうした英文の構造的な可能性としては(B)もある。

(A) $[NP \times [S_1 \dots [S_2 \dots]]]$



(B) $[NP [NP \times [S_1 \dots]] [S_2 \dots]]$



(B)は、関係節の主節が先行詞を修飾してきた名詞句 $[NP \times [S_1 \dots]]$ を、さらに、*that*の率いる従節 $[S_2 \dots]$ が修飾する構造である。実際、ソフトの翻訳例には、(B)の構造への誤分析と推定されるケース(下記参照)が認められる。そうした二重関係節自体は実在する構造だが、上の英文の訳としては不適当だろう。[従文 S_2 が直接「あなた」を修飾する構造も原文からは考え難い。]

「 $[S_2$ 私が会った $[NP [S_1$ あなたが[*きのう]知っている]少女] $]$ は美しい」(三洋)

「 $[S_2$ 昨日私が会った $[NP [S_1$ あなたが知っている]少女] $]$ は美しい」(高電)

また、解析としては失敗ではあるが、ソフトの翻訳例には、従節を帰属のない独立の節に解析するものもある。

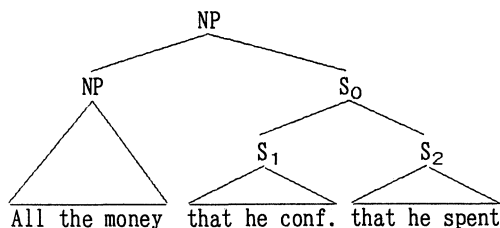
「 $[S_2$ 私が昨日会ったこと]、 $[NP [S_1$ あなたが知っている]少女] $]$ は、美しい」(沖)

このほか、「関係節の主節と従節が先行詞を並列的に修飾する」並列関係節[(C)の構造]と解析する可能性も残る。上記の英文の例ではそうし

た翻訳はみられないが、訳文に「連用中止」という形態があれば、並列関係節への解析を行っているとも認めるのに十分だろう。英文の解析は失敗しているが、下記の和訳例がこれに相当する。

$[NP$ All the money $[S_1$ that he confessed $[S_2$ that he spent]] is his father's.
「 $[NP$ *すべて $[S_2$ 消費し] $[S_1$ 彼が告白した]お金は彼の父のものである」(三洋)

(C) $[NP \times [S_0 [S_1 \dots] [S_2 \dots]]]$



以上のように、複文関係節構造を持つ英文については、論理的可能性が全て種々のソフトの解析に現れる。現状では、この構造の解析は安定していないのだ。言語的に有効な解析は(A)か(B)であるが、現実には(B)の構造が使われることはめったにない。したがって、(A)の構造への解析を優先させるように設定するのが、翻訳率の向上には直接的に寄与すると考えられる。

系 言 語

曖昧な構造に限らず、機械による解析の絞り込みは普遍的な言語処理プロセスに従うものと仮定されるが、言語的操作を組み込んで機械処理モデルを高度化し適切に動作するように設計することは、結果的に、人間の情報処理機構の解明にも具体的な形で貢献するものと期待される。

参 考 文 献

- 成田一(1994)編著『こうすれば使える機械翻訳』バベルプレス
成田一(1994-96)「翻訳ソフトこんなに使える！」『Professional English』バベルプレス連載
佐々木美樹・坂本仁(1995)「文章一括処理による係り受け関係の解析」言語処理学会第1回年次大会発表論文集