

言語類型と機械翻訳

成田一

大阪大学言語文化部

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

機械翻訳における言語処理は、翻訳対象となる言語間の構造的な距離に依存する。言語類型が同じ言語間の翻訳では、構造解析が浅いレベルのものでも、生成規則がしっかりしていれば、相当高い翻訳品質が得られる。これに対し、言語類型が異なる言語間では、構造解析が充実していることが不可欠だけでなく、文法・意味情報を広範囲に扱うことが、適切な翻訳には欠かせない。しかし、そうした深いレベルの言語情報に対応する翻訳システムであっても、同じ言語類型の言語間の浅い翻訳と較べると、翻訳品質は遥かに低いのである。本論では、英語と欧州語間の翻訳、日韓翻訳などの同型言語間の翻訳と、英中翻訳と日英翻訳など異型言語間の翻訳について、文法データの翻訳結果を検査し、これを踏まえて言語類型と言語処理の関係を考察する。

LANGUAGE TYPOLOGY AND MACHINE TRANSLATION

Hajime NARITA

Faculty of Language and Culture

OSAKA UNIVERSITY

1-8 MACHIKANEYAMA-CHO, TOYONAKA-SHI, OSAKA, 560 JAPAN

Language processing in machine translation is closely related to the distance of the structures of the target languages. Translation of the languages of the same type can produce results of high quality by structural analysis on the shallow level if the generative rules are well motivated. On the contrary, translation of the languages of different types requires a system which deals with rich semantic information, supported by a reliable parser. Such an integrated system, however, generates translation of a quality much poorer than an MT system designed for languages of the same type. This paper will clarify, for the purpose of designing optimal translation systems, how language types and language processing are co-related, on the basis of the translation results of the MT systems for English-European languages, Japanese-Korean, English-Chinese, and Japanese-English.

検証ソフト	Globalink「Power Translator™」	シャープ「デュエット(DUET)」日立「ハイキャ
[英→仏/西/独/中]	高電社「日韓翻訳 j・Seoul」	ッツ(HICATS)」NTT「ALT」
日英翻訳: ノヴァ「PC-Transer je」	エクシング	謝辞 英中データ[338例]は大阪大学言語文化研
「TransLand」	パシフィックアイ「ブラビスJ/E」	究科の黄愛鈴、日韓データ[335例]は大阪女学院
三洋「ザ・国際人JE」	高電社「j・London/JE」	短大の金菊子(兼任)講師に、英欧データ[338例]
(NEC「ピボット」)のライセンス供与)	富士通	は、大阪大学言語文化部の三藤博講師(仏語)、
「アトラス(ATLAS Win)」	東芝「エーエストラ	William Karkavelas教官(独語)、院生の幡野光
ザック(ASTRANSAC)」	沖「バンセ(PENSEE)」	メリリアン(西語)に判断を仰いだ。

機械翻訳においては、翻訳対象となる言語の構造解析と意味解析が適切に遂行され、さらに多義性の解消も行われることが望ましい。しかし、これは言語類型が大きく違う言語間の翻訳に関して必要となる言語処理で、言語類型が同じないし同族言語であるという場合には不可欠なものではない。語彙の同定、形態素の解析は必須条件になるが、構造解析は比較的小さな句〔日本語で言えば「文節」に相当するような〕レベルの構造が解析できれば、修飾関係など機能的な情報は把握しなくても構わない。ましてや意味解析、多義性の解消ということは、基本的に不必要である。

言語類型が同じ(特に同族言語)場合、生成部の規則が適切である限り、深いレベルの解析をしなくても「原文の構造をそのまま訳文に移す」ことが可能だ。I saw a woman with a telescope. は J'ai vu une femme avec un telescope. で訳文として成立する。修飾関係に曖昧性があつたとしても曖昧のまま訳文を生成し、これを読む人間が文の意味ないし文脈から判断すればよい。これは原文を読む人間が通常行う作業と同じことで、機械が負担するには及ばない。そうした意味で、同型言語の翻訳システムは、浅い解析部と形態規則レベルの生成部があれば基本的には十分であり、翻訳品質の面でも実用性が高い。深い解析を行うと機能を誤って判断することもあるが、そうしたミスが実際にはかなり起こるので、浅い解析の方が翻訳品質が高くなってしまふ。

一方、言語類型が異なる言語間の場合には、「原文の構造をそのまま訳文に移す」ことはできない。修飾関係など機能的な情報が分からないと、どういう構造を生成していいのか決定できないのである。先の例文の場合、with a telescope が副詞句ならば「望遠鏡で」と訳し、形容詞句ならば「望遠鏡を持った」と訳す。これは機能関係が分かって初めてできることだ。そうした意味で、翻訳システムの設計にあたっては、翻訳対象となる言語の類型に対応する最適なモデルを見極めることが肝要だろう。

同型言語の翻訳

同型の言語の翻訳として、日本語から韓国語への翻訳と英語から欧州語への翻訳を取り上げる。

日韓翻訳

日本語と韓国語は極めてよく似た言語である。漢語を背景とする語彙はもちろんだが、助詞などの文法機能を担う語彙についても、用法の違いはあるもののかなり対応が認められる。また、構造的な面でも共通性が高く、日本では一般に日本語固有と信じられてきた副次的な諸構造もそのほとんどが韓国語にも存在する。したがって、構造レベルの機能の解析は問題とならない。

ただし、日韓語の翻訳では、文中の語連鎖の「形態素解析における誤り」と「多義を解消できないこと」が誤訳の2大原因¹だが、生成面でも「助詞や時制を選択する情報の欠如」とか「形態接続規則の不備」が訳文の不具合を生んでいる。

注1 日本語は、漢字、ひらがな、カタカナといった「表記による語彙境界の識別」や漢字の「表意性による多義の判別」ができるが、韓国語のハングルは表音文字でこの機能を欠くため、韓日翻訳は日韓翻訳より翻訳率がかかなり落ちる。助詞などは韓国語の方が日本語より複雑だが、一般に「複雑系から単純系への投射」は固定的な設定で済むので、こうした統語現象については韓日翻訳の方が有利である。

日韓翻訳の実力 「これに抗議した学生が多い」「これを支持した学生はない/少ない」などの述語否定文、「[三人の学生]が来た」「[学生][三人]が来た」「[学生]が[三人]来た」といった類別詞基本・浮遊形式はもちろん、「僕は原稿を書いて-いる/しまった/(い)ない」「僕は原稿を書かない(で済ました)/き始めた」など種々のアスペクト形式や、敬語との複合形式「先生は原稿を始められた」にも適切に対応している。日本語固有とみられている表現構造も実は韓国語と共通なために、二重主語文「象は鼻が長い」だけではなく被害の受身「僕の女房はその頃父親に死なれた」なども完全な訳が得られる。通常の関係節はもちろん、その派生形式として捉えられる疑似/場所関係節「僕は母親が魚を焼く]臭いがとても好きだった」「[学生が歩いている]傍を警官が通った」や「の節」「[子供が走り寄ってきた]のを抱き上げた」も問題ない。韓国語では「の」の代りに「こと」に相当する語を使わなければならないが、構文的には日本語と同じ機能を果たす。ただし、「ところ節」「[子供が走り寄ってきた]ところを抱き上げた」の場合は、「ところ」が「場所」の意味でしか解釈できない。推量文「彼はその事務

員を教授と間違えたようだ」や分裂文「[その教科書を注文した]のは先生だ」のほか、条件文「君が反対するなら、僕はそこに行かない」なども対応している。助詞や活用など些細な誤りを除けば、翻訳率は9割を悠に超える。検証したソフトは市販されるパソコン用翻訳ソフトの中では最高水準 [ハングルカナ(チャンシンコンピュータ) HICOM/MT(日立情報ネットワーク)] という評価 [96年言語処理学会：金泰完氏講演] である。

英欧言語記

同じ印欧語族に属する欧州語間の翻訳の場合、時制系を除けば、言語構造の諸局面や表現形式が極めて類似している。このため、形態解析がうまくできれば語彙置換レベルの直訳でも概してまともな翻訳になる。もちろん、「構文の違い」はあるし、語源的に対応する語彙でも「意味、用法のずれ」などから、直訳では不適切なものもある。たとえば、スペイン(西)語では「属性」か「状態や存在」か意味機能によって動詞 *ser* と *estar* を使い分けるので、英語の *be* を常に *ser* に訳す固定した現状の設定ではいけない。

時制系の機構に関しては、ロマンス系言語はゲルマン系言語とかなり異なる。フランス(仏)語では、直接法、条件法、接続法、命令法があり、直接法の過去時制だけでも5種類で、それぞれ6通りの人称変化をする。英語の動詞活用の情報だけでは、仏西語などの複雑な時制形式に適切に投射するシステムはなかなか構築できない。英語の過去形は現代仏語では複合過去という完了形[avoir+過去分詞]が対応するが、現状では、そのまま過去形(単純過去)に訳す設定になっている。

呼応機構に関しては、名詞に「男/女/中性」と「4つの格」があるドイツ(独)語が典型だが、欧州語では「名詞の性、数、格に冠詞類と形容詞が一致」する。格関係の分析に若干不備が残るものの、システム的には対応できている。

英欧翻訳の実力 ここでは英欧語間の翻訳能力を概観する。言語的類似性からいうと、英語と独語は同じゲルマン語族だがいとも関係ほどに違いが出ている。英語と仏語は、300年余のノルマン王朝支配による語彙を中心とした仏語化の結果、

姉妹ほどの関係になっているが、現代はフランスが英語の影響を懸念し、流入を阻止する姿勢を示しており、仲が良いとは言えない。これに対し、仏語と西語は同じロマンス語族として極めて良く似た姉妹だが、西語とイタリア(伊)語は一卵性の双子が異なる地域、環境に育ったほどしか違わない。せいぜい方言くらいの差なので、それぞれの話者が母語で話しても8割5分ないし9割は理解できる。このため、英独翻訳は8割、英仏/英西翻訳は8割5分ほどの正訳である。製品化するほどの需要はないが、逐語処理的なシステムでも形態変化規則さえ充実していれば、仏西翻訳は9割以上、西伊翻訳は9割5分ほどが、時制形式を含め、文法的にも整った訳となることが望める。構文の違い He gave the book [to Mary]. は [SV0+PP] という構文だが、これは仏語や西語にも共通なので Il donnait le livre [a Mary]. や El dio el libro [a Mary]. に直訳できる。しかし与格を使う He gave Mary a book.[SV00] という文型は仏、西語にないので *Il donnait Mary le livre. や *El dio Mary el libro. と直訳するのは不適切である。逆に、独語には Er gab Mary das Buch. という与格構文しかない。

形容詞補部 John is eager [to see you]. の場合も、Jean est desireux [a vous voir].(仏)、John es avido [para ver usted].(西)、John ist eifrig, [Sie zu sehen].(独)に訳される。いわゆる難易文もそのまま訳され、John is easy [to cheat]. は Jean est facile [a [<de] tricher[<tromper]].(仏)、John es facil [de enganar].(西)、John ist leicht [zu betrogen].(独)になる。一部に α [< β] で示すように、語彙選択に改良の余地がある。

受動構文 My sister wrote this article.(SV0) は Ma soeur ecrivait cet article.(仏)、Mi hermana escribio este articulo.(西)、Meine Schwester schrieb diesen Artikel.(独)に訳され、受動態は Cet article etait ecrit [par ma soeur].(仏)、Este articulo era escrito [por mi hermana].(西)、Dieser Artikel wurde [von meiner Schwester] geschrieben.(独)になる。

翻訳率 英文法データ338例の仏訳の分析結果では、完訳やそれに準ずる訳が95例あり、時制、冠詞、語彙選択など比較的軽微な誤訳が195例、構文の表現形式などを改善すべきものが70例ほどで、意味が一応通る翻訳の比率は9割を超える。英西

翻訳もこれに準じる。英独翻訳では副詞などの語順や語彙の意味、用法のずれが目立つが、意味の通る翻訳が8割5分は超える。

N.B. 欧州諸語の文字に重なる補助記号=アクセサン[[^]...]については出力できないので省略する。

目的補語をとる They found the house empty. (SVOC) は Ils trouvaient la maison vide.(仏)、Ellos encontraron la casa vaciar.(西)、Sie fanden das Haus leer.(独)に翻訳され、受動態は La maison etait trouvee vide.(仏)、La casa se encontro vacia.(西)、Das Haus war leer gefunden.(独)になる。

不定主語構文 知覚構文の They saw the man [eating an apple]. は Ils voyaient l'homme [mangeant une pomme].(仏)に訳される。これを受動化すると *L'homme etait vu [mangeant une pomme].(仏)になるが、不定主語を立てた能動文 On a vu l'homme~. が普通である。独語訳 Sie sahen den Mann, [der einen Apfel isst]. は知覚構文ではなく関係節になっている。

強調構文 It is that house [where we have lived before]. は、仏訳では Il est cette maison [ou nous avons vecu avant].のように、英語の It を Il に直訳せずに、C' [=Ce]に改めれば正しい文になる。独語では Es ist das Haus, [wo wir vorher gelebt haben]. のように、「助動詞が従属節の末尾に」置かれる。

分詞構文 He went out the room, [slamming the door]. は Il allait dehors [<a quitte : 時制と語彙選択] la piece, [claquant la porte]. (仏)、El salio [de] la sala, [golpeando la puerta].(西)に翻訳される。分詞構文は元来ゲルマン語にはなく、英語の場合仏語から借用された文法形式であるため、独語には訳出できない。

比較 She is as tall as her sister. は Elle est aussi grande que sa soeur.(仏)になるが、「異なる性質の程度比較」 He is cleverer than Mary is pretty. も Il est plus astucieux que Mary est joli.(仏)に直訳できる。

語法 He said to me, "I will be absent from school tomorrow." は Il disait a moi, "je serai absent d'ecole demain."(仏)に、He told me [that he would be absent from school today]. は Il me disait [qu'il serait absent d'ecole aujourd'hui].(仏)に翻訳される。いずれも主文の述部は m'a dit/demande のように複合過去形の使用と人称代名詞の縮約が必要である。

内容節 We admit his decision [that he should resign]. は Nous admettons sa decision [qu'il devrait demissionner].(仏)、Nosotros admitimos su decision [que el deberia re-

nunciar].(西)、Wir lassen seine Entscheidung zu, [die er zurucktreten soll].(独)に訳される。関係節 He is the only boy [that I don't forget]. は Il est le seul garçon [que je n'oublie pas].(仏)、El es el muchacho unico [que Yo no olvido].(西)、Er ist der einzige Junge, [den ich nicht vergesse].(独)に翻訳される。また、複合関係副詞節の [Wherever] you go, I always think of you. も [N'importe ou] vous allez, je pense toujours de vous.(仏)、[Adondequiera que] usted vaya, Yo siempre pienso de usted.(西)になる。複文関係節 This is the book [which you insist [that I stole yesterday]].にしても、C'est le livre [que vous insistez [que je volais hier]].(仏)、Este es el libro [que usted insiste [que Yo robe ayer]].(西)のように直訳で構造的に成立する。ただし、独語では Dieses ist das Buch, *der [<das] Sie beharren, [dass Ich gestern stahl]]. のように関係詞の格が誤っている。

英欧翻訳は、時制系はともかく、形態解析部と名詞を中心とする呼応系を管理する生成部が充実していれば、かなり高い実用レベルの翻訳が達成でき、英日間翻訳のような厚い壁はないのである。また、日英間翻訳と違い、英欧語間では(日韓語間と同じく)訳文に原文と同じ曖昧性を持たせたままでも翻訳が成立するという点、知識処理を考慮しなくても済むのでかなり優利である。

欧英番訳

英語から欧州諸語への翻訳については、仏語、独語、西語への翻訳を検証したわけだが、現在、イタリア語、ロシア語への翻訳も準備中である。これと並行して、こうした欧州諸語から英語への翻訳能力についても検証の準備を進めており、仏語については既に結果が得られている。

英語対応データの作成 英欧翻訳と欧英翻訳の翻訳能力の違いを正確に検証するためには、英語データと欧語データが対照可能なものでなければならない。このため、欧語データの作成に当っては、英語データを英欧翻訳ソフトによって処理した欧訳データ [誤訳、非文が含まれる] を基礎に、これに最小限の調整を加えて適格な欧語データを作成するという方法が適当であると考えられる。これを欧英翻訳ソフトによって処理すれば、英欧翻訳の英語データに最も直接的に対比できる英訳データが得られると仮定される。

比較対照可能な翻訳データの作成

英語データ → MT → 欧訳データ
英訳データ ← MT ← 欧語データ

仏英番訳

仏語から英語に翻訳した場合、仏語の表現形式がそのまま英語に移されることが珍しくない。

語彙レベル 英語の [two years] ago は仏語では il y a [deux ans] という慣用表現で表されるが、これは文字通りには there are [two years] という英語に相当する。仏英翻訳ソフトでは正にこの形が訳文として出力するが、これでは英語の two years ago には了解されない。また、仏語では前置詞が英語ほど細分化されていないため、一つの前置詞がいくつかの英語の前置詞に訳し分けられなければならないのだが、現状では固定した訳になっている。中には前置詞 a ないしその冠詞 (le) との結合形 au のように前後の構造環境に関わりなく常に英語の has という語に訳されるなど、極めてお粗末な処理もみられる。代名詞も男性、女性の人称代名詞が中性の it にしか訳されないのも期待はずれであった。しかし、こうした誤訳は全く形態レベルのもので構造には関係しないため、自動ないし逐語的な置換操作によって正訳に変える対応ができる。

文法レベル 疑問文を形成する Est-ce que は文字通り Is-this that と逐語訳しても、英語では意味をなさない。これに続く埋め込み文が疑問文であることを表す標識なのである。埋め込み文を平叙文の形から疑問文の形に変換する一連の文法操作をしないと翻訳にならない。たとえば、Est-ce que [s vous avez obtenu toute information sur l'administration]. ならば、Is-this that [s you have obtained all information on the administration]. という訳になるが、Is-this that を除いて、埋め込み文を Have you ...? という形式に改める操作が施される設定でないといけない。仏文を構成する成分を語彙レベルで置換し活用などの形態を整えるだけで翻訳が成立することが多いので、代名詞など固定的な訳しかでないものもあるが、構造的にはかなり複雑なものでもほぼまともな訳が得られるのである。なお、時制については、英語の過去形は仏語では複合過去という完了形が対応するが、英仏翻訳で仏語も過去形 (単純過去) のままで訳されたように、仏英

翻訳においても仏語の完了形 [avoir+過去分詞] がそのまま [have+過去分詞] 英訳されている。

受動構文 Cet article a ete ecrit par ma soeur. は This article has been written by my sister. となり、目的補語をとる La maison a ete trouvee vide. は、The house has been found void. に訳される。

強調構文 C'est cette maison [ou nous avons vecu avant]. は、It is this house [or we have lived before]. になるが、ou (=where) がいつも or に訳されており、これは構造ではなく語彙レベルの誤りになる。

分詞構文 Il a quitte la piece, [claquant la porte]. は It has left the coin, [smacking the door]. のように構文的には正しいが、代名詞と名詞の選択 [coin は piece の一義だが room が正解] が間違っている。

内容節 Nous admettons sa decision [qu'il devrait demissionner]. は We admit its decision [that it would have to resign]. となる。

関係節 Il est le seul garcon [que je n'oublie pas]. は It is the alone boy [that I don't forget]. に翻訳される。複文関係節 C'est le livre [que vous insistez [que j'ai vole hier]]. にしても、It is the book [that you insist [that I have flown yesterday]]. となり、構造的には文法的な訳文が得られる。

欧州における品質要求 日英間翻訳などの水準から見たら遥かに高い翻訳結果が得られるにも関わらず、欧州における機械翻訳の開発意欲と利用者の満足度は決して日本より高いものにはなっていない。構造的な処理は極めて高いという結果になっているが、これは基本的な解析で間に合うだけ対象言語が構造的にパラレルであることに起因する。しかし、構造的に近く語彙的にも共通性がみられるというのは、欧州には母語とほかの欧州言語を容易に翻訳できる人が溢れているということで、機械翻訳での粗訳など必要としないのだ。始めから人間並の水準を期待してしまう。また、時制、活用などはどの文にも現れるが、こうした面で誤りが多いと、それだけ後編集作業に手間と時間がかかる。これでは機械翻訳を利用するメリットが得られない。欧州においてプロの翻訳者から機械翻訳の品質と効用に対する不満が出るのは、そうした言語的背景と翻訳文化ないし風土が関係していることを忘れてはならない。

異型言語の翻訳

類型の異なる言語間の翻訳として、英語から中国語への翻訳と日本語から英語への翻訳を見る。

英中番訳

中国語の特徴 中国語は、文成分の配列が [SVO] であり助詞が前置されるなどの特徴から、英語に近いと信じ込まれているが、これは一面にすぎない。主語、目的語を構造の支えとする英語とは違い、省略などで [VO] ないし [V] が一文を構成することが多い。複合語でも「売券」など [VO] 構造が基本だが、名詞句の内部構造は日本語に近似しており、連体修飾節は補文化辞「的」を介して主部の前に置かれる。文はそのままでも主語、目的語として埋め込めるが、「文を名詞化する」補文化辞「得」を付加することもできる。この場合、[[[SVO]V]得] のように「動詞を重複し元の動詞を任意に省略する」ので、補文は [[[SO]V]得] のように日本語と同じ配列 [SOV] になる。特に、人称、数の一致、時制がないので、語彙が活用しないため、形態的に文法関係を認定することは困難で、意味情報などが必要になる。

英中翻訳の実力 基本 5 文型はほぼまともな訳になっているが、その受動態は半分も訳せない。ただし、知覚文とその受身の The man was seen eating an apple. は適訳となる。名詞補部表現 We had [a discussion][on the serious issues]. や We admit [his decision][that he should resign]. のような内容節は意外に適訳が得られる。しかし、修飾表現全体としては半分程度しか訳せない。特に、He painted the wall red. のような状態補語は意味不明の訳になる。それでも、I didn't marry her because I loved her. のような副詞節を含む文は適訳になっている。Here comes the bus! など倒置は苦手だが、語順による強調 There the dog runs! なら訳せる。しかし、It is that house [where we have lived before]. のような構文による強調は全く手が出ない。It was very kind of you [to introduce

the president]. は正訳になるが、仮主語を含む構造も半分ほどしか訳せない。不定詞は名詞用法が半分訳せる程度、動名詞は I don't mind your smoking. などほぼ適訳になる。分詞はせいぜい半分程度で分詞構文もかなり苦手だが、準動詞は名詞用法が得意だ。比較はまあまあだが、仮定法、話法は半分止まり。関係節は 2~3 割しか対応できず、複文関係節は全く歯が立たない。これは中国語が [SVO] と [SOV] の混合型であるために、言語処理に強い制約がかかるためであると仮定される。そうした言語的特質のためか、一般に文の長さが比較的短くなる傾向にある。

言語類型のかなり違う英語から中国語への翻訳では、英文の解析も深くなければならぬし生成にも工夫が必要だが、浅い処理に終わっている。文法的に若干修正すべき箇所が含まれる訳例を認めても翻訳率が 5 割台に達しない。

日英番訳

日英翻訳の難しさ 日本語と英語は典型的に類型が異なるため翻訳が困難だが、翻訳方向によっても訳質は大きく違う。日英翻訳は英日翻訳よりも品質がかなり低いという状況だが、これには原因として、(1)英語では原文が簡明に書かれていることのほか、(2)日本語の言語的特徴が挙げられる。特に、英語では代名詞になるところが日本語では削除されているので、日英翻訳にあたってはどの要素が削除されているかを推定しなければならない。「戸を開けると、椅子に男が座っていた」では「戸を開けた」のは「僕」か「その娘」かが分からないと英文が構成できない。これには文脈情報などが必要だが、文脈処理システムはまだ「おもちゃのモデル」といった段階にあるから、この種の言語的課題を含む局面は今後も翻訳のハードルとなる。[ただし、「大阪へいらっしやる時は、是非ご一報下さい」では従文、主文の省略成分が何かを言語情報だけで決定できる。]

日英翻訳の実力 「あの女性は看護婦だ」は That woman is a nurse. になるが、推量の助動詞の入る「あの女性は看護婦らしい」でも That woman

日英翻訳の評価法 国文法教育は活用が中心になっており、構造的な観点からの説明はほとんどない。このため日本語の構造については一般に極めて不十分な理解しかできていないが、英文法教育では全体的に英語の構造を扱うので、ほぼ共通の認識が期待できる。したがって、英日翻訳データに対照できるデータにより評価した。この評価法

であれば、(1)英日翻訳との比較が容易で、(2)一般の日本人だけでなく日本語を良く知らない外国人も構造処理能力がほぼ揃える。なお、日本語固有の構造については英語など他言語と共通の中核的「一次構造」と区別し、派生的な「二次構造」(文献 2 参照)として扱うので、翻訳能力を二つのレベルで捉えることができる。

seems to be a nurse. と訳す。「僕の女房は魚が嫌いだ」のように目的語の助詞が「を」から「が」に交替する場合には、My wife dislikes a fish. のように正しく解析するソフトと A fish is hateful to my wife. のように主語に解析してしまうソフトに分かれる。これをクリアできても、「課長にコンピューターが分からない」という「が」から「に」に変る助詞パターンになると、A computer doesn't know it to the sectional chief. のように関係を誤認する。「その男が私に車を買わせた」は The man let/made me buy a car. のように使役動詞表現になる。「僕はその山が爆発するのを見た」が I saw the mountain explode. と不定詞構文になるソフトでも、「その男は遠くで犬が吠えるのを聞いた」は、The man heard that a dog barked in the distance. のように節構造に翻訳する。これは「伝聞」の意味にしかならずリアルタイムな現象知覚には使えない。これからも複文が解析できることが分かるが、補文標識が「の」ではなく「と」「こと」などの場合も問題ない。「私はこの問題は難しいと思う」は I think that this problem is difficult. のように適切に翻訳される。一種の強調構文 [分裂文] の「その教科書を注文したのは先生だ」は「の」を形式名詞に解析し What ordered the textbook is a teacher. とする訳もあるが、It was a teacher that ordered the textbook. という時制の一致にも配慮した翻訳もみられる。日本語固有の表現 この中には適切な英語への翻訳が困難なものもあるが、媒介的な言語処理を設定して対応できることも少なくない。また、英語では通常表現されることのない言語形式は、日本語の解析工程で検出しても訳文に反映しない。敢えて訳出すると過剰ないし歪んだ翻訳になってしまう。日英翻訳については、その合理的目標と人間の役割を考えなければならない。

被害の受身 この構造への全体的対応は限られるが、「彼女は雨に降られた」を She was caught in the rain.、「私は母に死なれた」を Mother died on me.、「私は泥棒に財布を盗まれた」を I had a wallet stolen by a thief. に訳すソフトもあり、決して克服できない構文ではない。

類別詞 日本語では数詞とともに類別詞が使われるが、現状では代表的なものにしか対応していない。類別詞が名詞句の内部だけではなく外部にまで浮遊する現象は対応できないソフトも目立つ。

「[三人の学生]が来た」は Three students came. と訳せても「[学生][三人]が来た」は Student 3 men came. 「[学生]が[三人]来た」は A student came 3. などになるのである。ただし、全て正解のソフトも複数ある。

アスペクト助動詞 時間相を表す助動詞は、対応する形式が英語にない場合、無訳の設定にする。て-「しまう」は完了状態を表すので英語は完了形に訳す。「花瓶を壊してしまった」は 8 社が A vase has been broken. のようになり、「火が消えてしまった」は Fire has disappeared.(プ/三/日)、Fire has gone out.(富/東/沖)などのほか、これらの過去完了形(ノ/高)や、Fire went out.(エ/シ/NT)などの過去形もみられる。

て-「ある/おく」も動作完了形に翻訳できるが、(be動詞による)状態完了形が適切なことも多いようである。「もう部屋は掃除してある」は 6 社が The room has already been cleaned.、4 社が The room is already cleaned. など正しく訳すが [完了形は「されてしまった」という含み]、副詞 already は文頭、文末にも配置される。

「冷蔵庫にビールを冷やしておきました」は (The) beer was cooled in the refrigerator. (エ/シ/東/三[to]/沖[on]/日[is~to])のほか、<S>/I cooled beer in/to a refrigerator. (ノ/NT)など主語を補う訳もみられる。

授受動詞 日本語では、授受行為における行為者と受益者の関係を話者の視点から捉えて、「やる」「くれる」「あげる」「もらう」などの授受動詞を選択することが迫られるが、英語への翻訳にあたっては全て give でこと足りる。「妹は花に水をやった」は A younger sister gave water to a flower.(ノ/沖/東/日[sent]/[did](プ/三/富))、A younger sister gave a flower water.(エ/高)、The (younger) sister watered a flower.(シ/NT [My])などに訳される。「友達に妹に飴をくれた」は A friend gave a younger sister (a) candy. (ノ/エ/高/東)以外は A friend gave a candy to a (younger) sister.(NT[their sister])となる。「弟は叔父にお金を貰った」は大体 A younger brother got money from (an) uncle.(NT[My]/高 [gold]/[received...to(プ/三/沖)])に翻訳される。ただし、授受動詞の翻訳が本当に問題となるのは、訳し分けの条件を必要とする英日翻訳である。うなぎ文 うなぎ文(「僕は鰻だ」)は情報を極めてコンパクトに収めた表現である。構造的には

[Xは[Yだ]]となるが、意味的に(「僕は学生だ」のような)X=Yという同定関係が成立しない場合、指定文(うなぎ文)として解釈する。これは[X(が[Vした/する]の)は[Yだ]]という構造(疑似分裂文)の縮約形式と考えることができる(V=動詞/形容動詞)。たとえば、「[僕が注文した]の[の]は[コーヒー]だ」>「僕はコーヒーだ」という関係があると仮定するのである。機械翻訳では、I am coffee. のような直訳で構わない。I orderd coffee. という解釈を生む理解プロセスは人間が分担すれば良い。同定関係が認められなければ、指定文として文脈に沿った解釈をするのである。これには「文脈、状況を的確に把握して、意図された関係を汲み取る」知識処理プロセスが要求されるが、機械による英訳を利用する際にも同じ知的作業を負担をするだけのことである。象鼻文 日本語には[AはBがCだ]という言語形式があるが、Bが身体の一部を表す「身体名詞」ならば、これは「Aの身体的特徴を記述する」表現になる。この表現を構造レベルで機械翻訳する場合、[Aは[_S BがCだ]]というように、Aをトピックとして扱い、[_S BがCだ]をそのまま翻訳する。「象は鼻が長い」は *As for an elephant a nose is long.* と訳す。しかし、名詞の「意味特性を扱えるシステム」の場合、もう少し意味関係を反映した翻訳が得られる。Bが身体名詞であることが検知されれば、Aをその所有者とみて、[The nose of the elephant] is long. ないしは [Elephant's trunk] is long. と翻訳するのである。さらに、「英語では身体記述に所有文を好んで使う」ことに対応するシステムは、*Elephants have long trunks.* という自然な英文に構造展開する。[「鼻」に "trunk"(幹状の鼻)を当てるのは「象」との緊密な関係が辞書情報に記載されているためである。]「太郎は父が医者だ」のようにBが親族名詞などでも「[太郎の父]は医者だ」に対応する翻訳が可能である。

日本語固有の節構造 こうした節構造は他言語と共通の中核的な「一次構造」と関連するが、これと区別される意味的・統語的特徴を示すため、派生的な「二次構造」(文献2参照)として扱う。疑似関係節 節の記述する活動に関連する現象、産物などを表す関連内容語が主部になる日本語固有の構造が疑似関係節である。内容節に近いものの、論理的な関係を表すには、通常の関係節に変換する処理を施す。「[娘が琴を弾く]音」は The

sound [that a daughter plays the Japanese harp] に翻訳されるが、これは補文に主部が付加されただけの操作である。通常の関係節を含む構造へ変換すると、「[[娘が弾く]琴の]音」が The sound of the Japanese harp [which a daughter plays] と正しい関係に訳される。

の/ところ節 日本語には主部に形式名詞「の」「ところ」が現れる構造がある。論理的な意味は関係節と同じだが、状況を描出する機能が顕著だ。「[父がボールを投げた]の[の]を打ち返した」は逸脱文 The thing [that father threw a ball] was struck back. と訳出されるが、関係節を含む構造「[父が投げた]ボールを打ち返した」に変える操作を介在させれば The ball [which father threw] was struck back. となる。「[泥棒が出てきた]ところを警察が捕まえた」は The police caught the place [where a thief came out]. という場所そのものを表す訳がほとんどで、例外的に The police captured thief's coming out. (富士)があるにすぎない。「[出てきた]泥棒を」に変えないと適訳にならない。

このように、日本語固有の節構造については、そのままではどのソフトも対応できないが、通常の関係節に変換する構造還元操作を介在させることにより、翻訳できるようになる。この操作は関連内容語を中心とした知識処理機構と組み合わせれば自動化することが可能である。

糸吉 語

本稿では、各言語の特徴を踏まえ、言語類型が近い言語間の翻訳ほど浅い言語処理の方が高い品質となり、遠い言語間の翻訳ほど、意味、知識にまで及ぶ深い処理が必要になる、という言語とシステムの関係をソフトの検証を通して確認した。

主要参考文献

- 1) 成田一編著『こうすれば使える機械翻訳』バベルプレス
- 2) 成田一著「連体修飾節の構造特性と言語処理」(田窪行則編『日本語の名詞修飾表現』くろしお出版に収録)
- 3) 成田一著「翻訳ソフトこんなに使える！」『PROFESSIONAL ENGLISH』バベルプレス (1994年11月号～1996年7月号)
- 4) 成田一著「機械翻訳における多義性の諸問題」言語処理学会第2回年次大会発表論文集