

## 翻訳ソフトの実力評価

成田一

大阪大学言語文化部

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

あらまし 日・英語間の翻訳ソフトの翻訳能力をジャンルごとにテキストデータを通して検証し、訳質の改善状況についても言及する。なお、テキスト翻訳の評価に当たってはその評価方法を改めて検討した。また、欧英翻訳の結果を各種英日翻訳ソフトで処理し、欧日翻訳ソフトと実力を比較する。これによって、英語を中間言語とするブリッジ方式の有効性とその条件を示した。

キーワード 機械翻訳、評価法、ブリッジ方式、多言語翻訳

## Evaluation of MT Softwares

Hajime NARITA

Osaka University

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

**Abstract** This paper proposes measures to evaluate machine translation programs by processing a set of text data. Efforts for the improvement of translation capacity will be surveyed. The data processed by MT programs for English-Japanese, English-European and Japanese-European languages will be compared to verify the validity and conditions of the bridge method in which English serves as an interlingua.

**Key words** machine translation, evaluation measures, bridge method, multilingual translation

### 0 はじめに

「言語類型と機械翻訳」(成田 1996)においては、文法データによる構造処理能力の検証を通して、英・欧語、日・韓語など同型の言語間と英・中語、日・英語など異型の言語間の翻訳の解析・生成レベルが大きく違うことを示した。今回は、日・英語間の翻訳ソフトの翻訳能力をテキストデータを通して検証し、訳質の改善状況についても言及する。なお、テキスト翻訳の評価に当たって、まずその評価方法を改めて検討する。また、欧英翻訳の結果を各種英日翻訳ソフトで処理し、欧日翻訳ソフトと実力を比較する。これによって、ブリッジ方式の有効性とその条件を示す。

### 1 翻訳品質の評価法再考

米国の「ALPAC 報告」(1966)において採用された評価法を基本的に踏襲したのが京大を中心とする「Mu プロジェクト」における評価法(1985)だ。これは(1)理解度、(2)忠実度という尺度によって訳文を5ないし6段階で評価する。

理解度の評価にあたっては、原文と対照しないで、英米人が「訳文をどれだけ理解できるか」をみることになっている。このため、訳文の句、節の係り受けや指示対象、時制が原文と違っていたり、翻訳できなかった部分が脱落していても、訳文自体は理解度が高いことが少なくない。例えば、「John was elected mayor.」を「ジョンは選ばれた市長だった」と誤訳し

でも理解度は高いと判断されるだろう。

忠実度の評価では「どれだけ原文が忠実に訳文に反映されているか」を見る。が、作業的には(語、句など)原文要素の脱落、係り受けの誤りといった大雑把な項目にチェックするにすぎない。「I saw Mary with the bag.」を「私はバッグでメアリを見た」とする誤訳に対しても「係り受け」の正否が記されるだけだ。「どういふ修飾関係がどう誤って翻訳されるか」という具体的な情報が抜けている。

### 1.1 評価尺度のコンセプト

このように、理解度、忠実度という評価尺度は、いろいろ不備な点や欠陥がある。が、有効性がないわけではない。そこで、こうした尺度の持つ本質を検討し、そのコンセプトを改めた。

忠実度は「原文の情報がどれだけ正確に伝えられるか」という趣旨のものだが、(意味を支える)構造関係が正しく捉えられていれば情報は忠実に伝わる。つまり、一義的には「原文(の構造)が正しく解析されているか」を検証するというので、併せて「文法的な訳文になっているか」についても見ることになる。したがって、評価にあたっては原文と訳文を対照することが不可欠だ。構造面を分析、照合するから、制御された(網羅的)文法データが評価対象になる。

理解度は、構造関係が正しく捉えられていれば、必然的に高くなる。つまり忠実度に比例する評価結果になる。ただし、評価内容は「原文と変わらない意味で理解できるか」ということだ。やはり、原文と対照しなければならぬ。が、構造処理を個別に見ると違い、「雑多な成分を含む文が全体として理解できるかを判断する」のが特徴だから、生の文章(テキスト)の翻訳能力を見るのに適している。実務で使う文書を訳してみれば、現場の用途にどの程度対応できるか、などおおよその印象は掴める。

では、理解度と忠実度はどう違うのか。文章中の文が評価対象であっても、語や句など比較的小さい構造ごとに検証する限り、実質的に構造の解析を個別に調べることになる。ただし、節を幾つも含む文や文章に領域を広げると、全体的な理解度が見えてくる。「構造の個別評価が忠実度」で、「文(章)の総体的評価が理解度」であると捉えても構わない。評価尺度は

評価対象と依存的な関係にある。いずれにせよ、構造解析のよし悪しが理解のしやすさに繋がる。

### 1.2 評価段階

「忠実度による評価」では、(解析・生成を含め)文法的に次の3段階評価が適当だ。

- (1)ほぼ正確＝「良」(○)
- (2)どうにか許せる＝「可」(△)
- (3)ほとんど失敗＝「不可」(×)

「理解度による評価」でも、原文と対照するとそれなりに信頼のおける評価が可能だ。若干判断の揺れはあるが、以下の3段階で評価するのが適当だろう。(あまり多段階の評価だと、段階の境界で判断が揺れ、安定的な評価が困難。)なお、「ほぼ理解できる」で、かつ「文法的に正しい」ならば、「優」(◎)という評価を与えても良い。これは忠実度の評価事項を持ち込んだ折衷案になるが、これで実力がより正確に掴める。

- (0)文法的にも正しい＝「優」(◎)
- (1)ほぼ理解できる＝「良」(○)
- (2)どうにか理解できる＝「可」(△)
- (3)ほとんど理解できない＝「不可」(×)

### 1.3 テキストによる検証

実務文書中の文は、単文のこともあれば重文、複文のこともある。また、長い文ほど処理すべき構造の種類や数が多くなり、解析が失敗する可能性も増え、全体の評価が短い文よりも低くなる。これは修飾の曖昧性など、知識処理が必要な関係が増えるためだが、40字を超えない文を処理対象とするのが適当だ。

もちろん、文ではなく、句や単語などを選んで、その訳を見るということであっても良いし、段落の文章全体であっても構わない。句や単語だと忠実度、文章レベルだと理解度を見ることになる。

ただし、テキストの場合、(1)文法操作や構造が全て含まれているわけではない、(2)単独では処理できる構造でも文中の他の構造や語彙が原因で解析に失敗する、などの理由から、どういった尺度で見ても、文法構造全般を漏れなく検証することはできない。構造処理上不備な点についてどう対処するかも分らない。

自然性の評価作業 日英翻訳ならば英米人に「自然で理解

しやすい英文]になっているかを評価してもらえる。ただし、原文との意味の照合作業まで英米人に望むことはできない。このため、日本人が原文を英文に翻訳し、英米人がその英文と機械の英文の意味を照合した上で評価する。

## 2 ジャンル別翻訳能力

ここではジャンル別にソフトの翻訳能力を検証する。長い文ほど処理すべき構造の種類や数が多くなり、解析が失敗する可能性があるため、2箇所ほどに分けて評価することもある。それでも、文章の場合は、制御されたデータによって個別の文法構造の処理能力を検証するような客観的な評価にはならない。

句読点と構造解析 文脈情報を使えない機械翻訳では、句読点の使い方が構造解析において極めて重要な手がかりを与えることが少なくない。誤った個所に読点を打てば解析も誤る。長文の場合、この程度の文章チェックを行なうのは、機械に処理を委ねるための最低条件である。句読点の使い方が適切ではなく、あるべき箇所に読点がないために、構造解析に失敗するケースもみられる。原文に読点を補ってデータ文の補正を行ない、ソフトの翻訳能力を改めて評価する。

### 2.1 英日翻訳

英日翻訳ソフトはかなり実力に違いがあるが、優秀なものは実用レベルに達している。ほんの少し手を加えればちゃんとした翻訳になる。ここでは、2つのジャンルの文章について翻訳ソフトの実力を検証する。

#### 取扱説明書(ゲームソフト MANAGING A COLONY)

If you do not want a new colony in this system, you can close the System window instead. This instructs a Colony Ship to await further instructions, but destroys a Colony Base. When you're prompted to select a world for a Colony Base, you must do so or else lose the Colony Base (and all the production that went toward building it).

「もしこのシステムに新しい植民地が要らないなら、代りにシステム・ウィンドウを閉じることができます。これにより植民地船が更なる指示を待つように命じられるが、植民地は破壊されます。植民地に新しい世界を選ぶように促された場合、そうしなければならない。さもないと植民地(とこれを建設することに向けた全ての生産)を失うことになります。」(試訳)

第1文の If you do ~ window instead. は「システムウィンドウ」のほか「システム窓」と訳したり System がそのまま訳に出る例があるが、構造的にはいずれも正しく訳されている。

This instructs a Colony Ship to await further instructions, but destroys a Colony Base.

NV◎これは、植民地船に更なる指示を待つように命令するが、植民地ベースを破壊します。

LV◎これは Colony 船にそれ以上の指示を待ち受けるよう指示します、しかし Colony ベースを破壊します。

TSOこれは、一層の指示を待つように植民地船(Colony Ship)に命じるが、植民地基礎(Colony Base)を破壊する。

SHOこれは、更なる命令を待つように Colony Ship に指示する、しかし、Colony Base を破壊する。

HI△これは、それ以上のインストラクションを待つ Colony の Ship を教えるが、Colony ベースを壊す。

FU○これは一層の指示を待ち受けるように Colony Ship に命令するが、Colony Base を破壊する。

IBMOこれは、植民地船にそれ以上のインストラクションを待つように指示します、しかし、植民地基盤を破壊します。

AIL△これは植民地船がもっと遠い命令に用意されるように指示するが、植民地基礎を破壊する。

第2文はどれも無生物主語をそのまま訳しているので直訳調だが、論理的な意味は一部を除いて適切に捉えられている。日立では不定詞句 to ~ instructions が前の名詞を修飾する解釈になっているが、これは動詞 instruct の構文への解析を優先させることで避けられる。Colony Base が訳出されていないのは固有名詞の扱いのためだ。

When you're prompted to select a world for a Colony Base, you must do so or else lose the Colony Base (and all the production that went toward building it).

NV◎あなたが植民地ベースのために世界を選ぶことを促されるとき、あなたはそうしなければならないか、他に植民地ベース(そして、それを造ることに向かった全ての生産)を失わなければならない。

LVOあなたが Colony ベースのために世界を選択するよう促される時、あなたはそうするか、あるいはほかに Colony ベース(そしてそれを建てることに行つたすべての生産)を失わなくてはなりません。

TS◎あなたは植民地基礎(Colony Base)のために世界を選ぶように促される場合、そのように行わなければならないかあ

るいはさもないと植民地基礎(Colony Base)(およびそれを構築するへの行ったすべての生産)を失わなければならない。SHOあなたが Colony Base のために世界を選択するように促されるとき、あなたは、そうする、或いは、Colony Base (そして、それを造ることの方へ向かった生産全て)を失わなければならない。

HI△あなたが Colony ベースに対して世界を選択させられる時、あなたは、そのように行なうか、または、他に Colony ベース(そしてそれを造ることに使われたすべての生産)を失う必要がある。

FU△Colony Base のために世界を選択するようにながされるとき、そのようにまたは失いなさいくらい、あなたは Colony Base(そしてそれを建てることに向かって行ったすべての生産)をしなければならない。

IBM△植民地基礎のための世界を選ぶために促進されるときに、あなたは、して、したがってさもないと、植民地基礎(およびそれを構築することへ行ったすべての生産)を失わなくてはなりません。

AIL△あなたが世界を植民地基礎のために選ばせられるとき、そのようにあなたはしなければならず、か植民地基礎を他に失う、(そしてそれを建てることの方へ、行ったすべての製造)。

第3文は比較的長文だが、大体意味が通る訳が得られる。ただ、or else については、「ほかに」などがほとんどで東芝とIBMだけが「さもない」と正しく訳す。

調整後の訳 第3文に you must do so, or else lose のように読点を補うと、いくつかのソフトで訳に変化がみられる。シャープでは「あなたは、そうする、さもないと、」となり、日立は「あなたは、そのように行なう必要がある、でなければ、Colony ベース(...)を失う」に変わる。富士通は「あなたはそうしなければならないか、または Colony Base(...)に負けなければならない」になる。が、IBM は「あなたは、そして、あるいはそのほかに、植民地基礎(...)を失わなくてはなりません」のように or else の訳については劣化している。

#### CNNニュース(ブラジル警察のストライキの記事)

A tidal wave of police strikes has led to a widespread breakdown in law and order in Brazil and brought the country to the brink of chaos. ... A protest is scheduled Friday in Sao Paulo, Brazil's richest and most populous state, and another

was held Thursday in Rio de Janeiro.

「津波のような警察のストライキはブラジルの法と秩序に広範な機能不全をもたらし、国家を混沌の淵に投げ込んだ。... ブラジルで最も裕福で人口の多いサンパウロでは金曜日に抗議行動が予定されており、リオデジャネイロでは木曜日に行なわれている。」(試訳)

A tidal wave of police strikes has led to a widespread breakdown in law and order in Brazil and brought the country to the brink of chaos.

NV◎警察ストライキの津波は、ブラジルの法と秩序の中の普及した故障に至って、カオスのふちに国を持ってきました。

LV◎警察ストライキの津波がブラジルで法と秩序で広範囲にわたる故障に導いて、そしてカントリーをカオスの瀬戸際にもたらしました。

TSO警察ストライキの A 津波は、ブラジルで治安強化中の広範囲のブレイクダウンに通じており、混乱の縁に国を連れて来た。

SHO警察ストライキの A 津波は、ブラジルの法と秩序における広範囲にわたるブレイクダウンという結果を生み、そして、国民を無秩序の縁に連れて来た。

HI×警官の当たりの大変動は、ブラジルに治安を広く行きわたった故障へ通じて、国民は混乱状態の縁まで来た。

FU△警察の打撃の A 津波はブラジルの法と秩序で広範囲の故障に通じて、国を混沌縁に持って来た。

IBMO警察ストライキの高波は、ブラジルでの法と秩序における広げている突然の故障に至って、そして国をカオスの縁に運びました。

AIL×警察の打つことの潮の波はブラジルで法および順序での広範囲の故障に通じ、国を無秩序の縁にもたす。

第1文では A tidal wave が「津波」のほかに「高波」「潮の波」「大変動」などに訳されているが、A が冠詞として解析されていない例が3つもある。また breakdown が「故障」でも意味は分かるが、カタカナでは日本語にならない。

A protest is scheduled Friday in Sao Paulo, Brazil's richest and most populous state, and another was held Thursday in Rio de Janeiro.

NV◎抗議は金曜日にサンパウロ(ブラジルの最も豊富で最も人口の多い状態)で予定されます、そして、他のものは木曜日にリオデジャネイロで持たれました。

LV◎抗議がサンパウロ、ブラジルの最も金持ちの、そして最

も人口ちゆう密な国家で金曜日に予定されます、そしてもう1 (つ・人)はリオデジャネイロで木曜日に持たれました。

TS△抗議は金曜日に、サンパウロ、最も豊富であり(ブラジルの)、最も人口稠密な状態で予定される。また、別のものは木曜日にリオデジャネイロの中に保持された。

SH○抗議は、サンパウロ、ブラジルの最も豊かな、そして、最も人口の多い状態で金曜日にスケジュールされ、そして、別のものは、リオデジャネイロで木曜日に保持された。

HI×★★---抗議は、Sao の Paulo で予定される---。---金曜日、ブラジルの最も豊富で、最も人口が多い国家、およびもう一つは、木曜日、リオデジャネイロで持たれた。

FU×抗議が金曜に Sao Paulo で予定される…ブラジルのもの最も豊かな最も人口の多い状態…別のものは木曜に Rio 反-Janeiro に保持された。

IBMO抗議は、サン・パウロ、ブラジルの最もリッチなおよび最も人口が多い状態で金曜日に予定されます、そして、もう一つのもの、リオデジャネイロで木曜日に開催されました。

AIL×リオデジャネイロで、木曜日、抗議は金曜日 Sao Paulo の、ブラジルの最も金持ちの、最も人口の多い状態での予定に組み入れ、他の人は所持された。

第2文では Sao Paulo を説明する Brazil's ~ state が英文と同じ順序で訳されているが、できれば「~州、サンパウロ」のように後ろから訳した方が分かりやすいだろう。

## 2.2 日英翻訳

日英翻訳ソフトはまだ実用段階に達したとは言えない。このために、辞書引きのほか、例文参照など、利用者の判断による校正作業を前提とした編集機能を備えているものが少なくない。利用者が原文と訳文を対照して誤訳や不備の有無を確認しながら問題のない部分を適当に調整して使うということになる。しかし、専門的な用語や表現に対応しているソフトの場合、かなり参考になる訳文が得られることもある。

### 原子炉(原子炉内部の構造説明)

原子炉は巨大な圧力容器(直径約70フィートの球形容器)の中に納められている。その圧力ドラムの内部には何千もの黒鉛ブロックがあって、それらのブロックとブロックの間のチャンネルの中にはマグネシウムの合金で被覆されたウランウム金属の棒が立ち並んでいる。

A nuclear reactor is placed in a huge pressure container (a spherical container with a diameter of about 70 feet). There are thousands of graphite blocks in the pressure drum, and rods of uranium metal, coated with magnesium alloy, are lined up in the channels between these blocks. (試訳)

原子炉は巨大な圧力容器(直径約70フィートの球形容器)の中に納められている。

NV◎ A nuclear reactor is put in huge pressure container (○ a global container of about a diameter of 70 feet).

FU○The nuclear reactor is supplied to a huge pressure container (○ globular container of about 70 feet in the diameter).

BR◎ A reactor is put in the huge pressure container (△ about diameter 70-foot globular containers).

NECO A nuclear reactor is supplied into the huge pressure container (○ the globular container with about 70-foot diameter).

第1文の「納められている」は「収納されている」という状態を表しているが、is supplied to/into などの訳では「供給する」という意味の「納める」行為を表すので、原文とは多少ずれている。put into でも厳密には行為を表すが、put in は状態を表すことを前提とした翻訳であり許容範囲に入るだろう。丸括弧部分は富士通、ノヴァ、NEC の順でほぼ容認できるが、ブラザーは改善が必要だ。「球形容器」の「球形」は spherical でなく global や globular でも構わない。ただ「直径約70フィートの」のような「数量表現」は技術的な文章には頻繁に表れるものであり、改善が欲しいところだ。富士通は globular の前に a 欠けていて、diameter の前に the が余分にあるなど冠詞の不備を別にすれば完璧で、ノヴァは about を 70 feet の前に移し、diameter の前の of を with に変え、NEC は冠詞を a に改め with を of に変えればほぼ正しい表現になる。

その圧力ドラムの内部には何千もの黒鉛ブロックがあって、それらのブロックとブロックの間のチャンネルの中にはマグネシウムの合金で被覆されたウランウム金属の棒が立ち並んでいる。

NVO There is thousands of graphite block in the inside of the pressure drum //△ and lines with a stick of uranium metal coated by an alloy of magnesium in a channel between things of those blocks and a block.

FU○ There is thousands of one black lead block in the pressure drum, //△ and (\*S) lines in the channel during those blocks and blocks with the stick of the uranium metal covered with the alloy of magnesium.

BR◎ Thousands of black lead blocks are inside the pressure drum, //○ and the stick of the uranium metal that it is covered with the alloy of マグネシウム in the channel between those blocks and the block lines up.

NEC◎ There are thousands of graphite blocks inside the pressure drum //◎ and the stick of uranium metal which was matted with the alloy of magnesium ranges in the channel between those blocks and the block.

第2文は前半の「その圧力～があって」と後半の「それらの～立ち並んでいる」に分けて評価できる。前半部は NEC が完璧で、ほかは動詞や名詞の数に不備がある。「黒鉛」が black lead でなく graphite に訳されるソフトは専門語に強いのが分かる。「の内部」は in ないし inside で良く、ノヴァは過度に分析的な表現になっている。

後半部も NEC がかなり読めるものになっており、stick を rods に変え、「それらのブロックとブロックの間」を between these blocks のように余剰性のない表現に改めれば完璧に近づく。ただ、知識処理機構を含まない現在の機械翻訳システムでは、そこまで要求することは酷な気もする。通常「の間」は2者を「と」で結ぶのが普通で、言語的には正しく訳されていると言える。その2者が同じものなのに繰り返されるのがごちないのだが、むしろ、原文が余剰性のない表現になっていないことが問題かもしれない。

ブラザーは「マグネシウム」がそのまま出るなど専門語に対応できていない。「並んでいる」の lines up が文末に置かれるなどごちなさはあるが、構造的な処理はほぼできている。ノヴァ、富士通は「立ち並んでいる」が主語無しで lines with と訳されており、with の後の名詞句が意味的には主語の働きをしているが、この関係を推定するのは困難だ。

### 3 ブリッジ翻訳

欧英翻訳は言語類型的にも近いために、翻訳に必要な解析・生成が比較的簡単に構成でき、翻訳精度も高い、ということ、96年の自然言語処理研究会において発表（「言語類型と機械翻訳」）した。この考察を受けて、『パソコン翻訳の世界』（講談社）の中では、

(1)欧州の言語を英語に翻訳し、(2)これを日本語に翻訳する、というブリッジ翻訳によって、英語圏以外の地域の情報を得ることを推奨した。これは欧英翻訳によって得られる英文が高い訳質になっていることが前提であるが、それだけではなく、英日翻訳も標準レベルをクリアすることが条件になる。

ところが、欧日翻訳ソフト「AllLogic Translator 五カ国語」(AllLogic)は欧英翻訳部に優秀な Globalink (GL)社の翻訳エンジンを採用しているものの、英日翻訳部には(雑誌でも)業界水準より遥かに劣ると評価される自社の英日ソフト「E-JBANK」を組み込んでいる。ブリッジ方式だが英日翻訳が条件を満たしていない。

そこで、インターネットなどの欧文を GL 社の翻訳ソフトで処理し、その訳文をいくつかの英日翻訳ソフトで処理した結果を問題の欧日翻訳ソフトと比較してみた。欧文はフランス語、ドイツ語、スペイン語を対象としたが、紙面の関係上、フランス語の文章を取り上げた。

#### 原文(フランス語)

Vous êtes sur le site de la filière Communication de l'Université de Nantes (département de Lettres Modernes). Les pages que vous pouvez consulter sont l'ouvrage des étudiants, des techniciens et des enseignants de la filière. Celle-ci offre un enseignement de niveau licence et maîtrise. Son objectif est de former les futurs spécialistes de l'édition multimédia.

あなたはナント大学(現代文学部)のコミュニケーションコースのサイトにいます。ご覧になれるページはコースの学生、専門家、先生の制作です。このコースは学士、修士レベルの教育を提供します。その目標は、マルチメディア編集の将来のスペシャリストを育てることです。(試訳)

#### 仏日翻訳文(「AllLogic Translator 五カ国語」)

あなた? Universit 通信? Nantes(現代手紙 d? ページ、それ、いる、あなたは仕事に意見を聞くことができるか? このある申し出の水平な免許の教えること、そしてママ、? 彼に/彼女に/その客観的さは未来の sp を作ることであるか?

#### 仏英翻訳文(「WebTranslator」Globalink)

You are on the site of the Communication path of the University of Secure (department Advanced Letters). Pages that you can consult are the work of students, technicians

and teachers of the path. The former offers a teaching of license level and mastery. Its objective is to shape the future specialists of the multimedia editing.

この仏英訳では Nantes が何故か Secure になっている。Modernes が Advanced に変わり、先行する de が脱落しているほか、la filière 「コース」を指す Celle-ci 「こちら」が The former となっている。また、un enseignement de niveau licence et maîtrise 「学士、修士レベルの教育」は de 以下の修飾部を目的語の関係で捉えている。が、これを除けば、構造解析はほぼ正しい。

#### 英日翻訳文

仏語の英訳で不備がある部分は、英日翻訳においてもこれを原文とした翻訳にしかならない。仏語の原文とは違った訳になるが、この英訳を基礎に評価すると、英訳通りの下線部を含め、かなり良い訳になっている。ただ、of the path の修飾関係を誤るソフトもある。文脈処理はできないから path が邦訳で「パス」となるのは仕方がない。

SH△あなたは、Secure(部門 Advanced Letters)大学の Communication パスのサイトにいる。○あなたが相談し得るページは、パスの学生、専門家、及び、先生のワークである。前者は、ライセンスレベル、及び、支配の教示を提示する。◎その目的は、マルチメディア編集の将来のスペシャリストを形造ることである。

TS○大学のコミュニケーションパスのサイト上にいる……の……安全(部は文字を進めた)。◎意見を求めることができるページはパスの、学生、専門家および教師の仕事である。前者はライセンスレベルおよび制御の教えることを提示する。◎その目的はマルチメディア編集の将来の専門家を形作ることである。

LVOあなたは安全な(課 Advanced 手紙)の大学の通信パスのサイトの上にあります。△あなたが調べることができるページは学生、パスの技術者と教師の仕事です。前者は許可証レベルと優越の教授を提供します。◎その目的はマルチメディアのエディットの将来の専門家を形づくることです。

NV△あなたは、Secure(部 Advanced Letters)大学の Communication パスのサイトの上にいる。△あなたが意見を聞くことができるページは、学生、専門家とそのパスの先生の仕事である。前者は、許可レベルと支配の教授を提供する。○その目的は、マルチメディアのその編集する将来の専門家を形づくることになっている。

FU△あなたは Secure(部の Advanced Letters)の大学の Communication 経路の現場にいる。◎あなたが相談することができるページは経路の学生、技術者および教師の仕事である。前者は免許レベルと主人の教育を提供する。◎その目的はマルチメディア編集の今後の専門家に形成することである。IBM○あなたは、確保する(部門が、レターを進めました)大学のコミュニケーションパスのサイトにいます。◎あなたが、調べることができるページは、パスの学生、専門家、および教師の仕事です。前の提案ライセンスレベル、および熟達の教育。◎その目的は、エディットするマルチメディアの将来のスペシャリストを形づくるはず(こと、予定)です。

欧日翻訳ソフトの翻訳結果はほとんど意味をなさない。構文解析も劣悪だ。では、何が問題なのか。このソフトの内部処理は欧英翻訳と英日翻訳に分かれるが、欧英翻訳は GL 社の「WebTranslator」と同じエンジンなので、その翻訳結果と同じである。この英訳は実用レベルだ。すると悲惨な翻訳結果は「E-JBANK」と同じエンジンの英日翻訳の責任であることになる。このことは業界水準を達成している翻訳ソフトによる翻訳結果を見ても明らかだろう。

対象となる言語や文書の難易度にもよるが、欧英翻訳でも平均すると10%ないし20%程度は誤訳だけではなく未処理の原語や表現を含む不適切な訳文になる。標準的な英日翻訳ソフトは70%前後の翻訳率になるだろう。そうすると、欧日翻訳にあたっては英日翻訳の原文となる英文自体がまともではない部分がある程度含んでいることになる。

翻訳ソフトは、原文が文法的な文であることを前提に、解析・生成文法が書かれている。特に、ドイツ語などの長い複合語が未処理のまま英文に含まれる場合は、未知語として処理するか翻訳処理が動作しないこともある。したがって、欧文が原文の場合、日本語になる段階では最高レベルの英日翻訳ソフトを使っても50%ないし70%ほどが翻訳されるにすぎない。

できれば英語に翻訳された段階で、英文を最小限の調整をしてほぼまともな形に改める作業を施したい。その後で日本語に翻訳すれば翻訳ソフトのメリットが十分に生かせる。欧日翻訳を内部的に全て行い英訳に触れない処理工程だと、情報はほぼ取れても満足

な邦訳にはならないことも多い。その点をしっかり認識して、開発者にはシステムを組んで欲しいし、ユーザーにも賢い使い方をして欲しい。

#### 4 翻訳品質の改善

90年頃までに大手の電気通信系メーカーはほとんどが英日ないし日英システムを構築していたが、バブル崩壊後には機械翻訳部門を閉鎖した企業が多い。その後、パソコン版のソフトとして移植しているが、販売部門などがサポートの一環としてインターネット関連の語彙などを増強している。この体制では翻訳の根幹をなす解析・生成のための文法は変更できない。現在も開発を続ける企業はソフトハウス系が中心だが、言語系開発員主導で言語処理の問題に地道に取り組み、言語現象のかなり細部に及んで具体的な答えを着実に出している。ここでは旧版で対応できなかった言語現象が新版でどのように改善できたかを見る。

『NetSurfer/ej』(ノヴァ): 通常、動詞 see は「見る/見える」とされるが、目的語が[人間]という意味特性を持たば、「会う」の方が適切なケースが多いので、この訳語を選ぶ設定にする。I am glad to see you here. は旧版では「私は、あなたをここで見てうれしい」だったが、新版では「会ってうれしい」になっている。動詞 boil は、水などの[液体]が目的語の時は「沸騰させる」だが、野菜や卵など[食物]が目的語ならば「ゆでる」が適切だ。Mary likes eggs boiled by her mother. は、旧版の「メアリーは、彼女の母によって沸騰させられた卵が好きである」から新版では「ゆでられた卵が好きである」に変わっている。目や鼻や髪など[身体の部分]を意味する語がその目的語で、前置詞句全体が前の名詞を修飾する場合、前置詞 with は「持つ/持った」ではなく「の」が自然だ。My father is the man with silver hair. は「私の父は、銀の髪をもつ男性である」が「銀の髪の男性である」になっている。

『PowerE/J』(シャープ): 企業への販売実績を反映して、ビジネスライターなどにおける一連の定型表現にも特に配慮している。Let me introduce myself. は「私に自己紹介させなさい」が「自己紹介させて頂きたい」になっている。Enclosed are the R&D reports. のように過去分詞が主語の位置に立つ表現は文法的には倒置形だ。旧版では「囲まれている#研究開発レポートである」のように解析に失敗しているが、新版では「同封されているのは、研究開発レポートである」に改善さ

れている。経済記事についても対応が図られ、Gold opened at \$391 a troy ounce, up from \$382 late Monday. が「月曜日遅くの \$382 からアップして、金は、1 トロイオンス当たり \$391 で寄り付いた」に翻訳される。

『LogoVista』(ロゴヴィスタ): 法律文書に特有の表現についても改善がみられる。A method [whereby this problem can be resolved] is discussed in Section 5. は「A それによって方法#この問題は解決されることが出来る#セクション5で論じられる」から「それによってこの問題が解決されることが出来る方法がセクション5で論じられる」に改められた。A last-minute computer failure delayed NASA's launch early today of A MARS SPACECRAFT carrying the first-ever interplanetary rover. のように、NASA's launch とその目的語の機能を担う修飾語句 of A MARS SPACECRAFT が離れているときでも、適切にその関係が解析できるようにしている。文中の語句を手がかりに文書の分野を自動判断する機能もある。He knows many PASSAGE from Shakespeare by heart. は「彼はシェークスピア(Shakespeare)から多くの節を暗記して知っている」に訳し、He worked for PASSAGE and signed NAFTA. は「彼は議会通過のために努力して、そして NAFTA に署名した」に訳し分けている。

検証ソフト: 英日翻訳(「Power E/J」シャープ[SH]: 「NetSurfer/ej」ノヴァ[NV]: 「ASTRANSAC」東芝[TS]: 「LogoVista E to J Internet Plus」LogoVista[LV]: 「ATLAS」富士通[FU]: 「たちまち翻訳」日立[HI]: 「翻訳アダプタII」NEC: 「翻訳の王様」日本 IBM: 「Transpad」AILogic) 日英翻訳(「TransLand」ブラザー[BR]: 「PC-Transer je」ノヴァ: 「翻訳アダプタII」NEC) 欧英翻訳(「WebTranslator」Globalink) 欧日翻訳(「AILogic Translator 五カ国語」)

#### 主要参考文献

- 1) 長尾真(1985): 「Mu プロジェクトにおける日英翻訳結果の評価」情報処理学会報告. NL 47-11
- 2) 成田一(1988): 「機械翻訳における構造処理能力の評価」情報処理学会報告. NL 69-1
- 3) 成田一(1994. 11~1996. 7) 「翻訳ソフトこんなに使える!」『PROFESSIONAL ENGLISH』バベルプレス
- 4) 成田一(1996): 「言語類型と機械翻訳」情報処理学会報告. NL 114-21