

J E I D A 機械翻訳システム評価基準 (開発者編)

— 開発者のための評価項目の作成と検討 —

中岩浩巳 森本康嗣 松平正樹 成田真澄 野村浩郷
NTT (株) (株) 日立製作所 沖電気工業 (株) (株) リコー 九州工業大学

(社) 日本電子工業振興協会の機械翻訳システム調査専門委員会技術動向調査ワーキンググループでは、機械翻訳評価基準の検討を (1) ユーザーによる評価、(2) 開発者からの評価、(3) 翻訳品質の評価の3種類の観点から進めてきた。本稿では、このうち、機械翻訳システムの開発者が自らのシステムを評価する際にチェックすべき技術的項目及びそれを用いた評価法について報告する。この開発者からの評価基準は、機械翻訳システムが扱う言語対に依存する技術項目と、依存しない技術項目に大別される。それぞれの設問には、評価の観点を示す6つの軸と技術レベルの評価軸のポイントが付与されており、それを用いることによって機械翻訳システムの特徴や弱点が客観的に評価できる。

JEIDA's Proposed Method for Evaluating Machine Translation
(Developer's Guidelines)

— Technical Evaluation by Developers —

Hiromi Nakaiwa Yasutsugu Morimoto Masaki Matsudaira Masumi Narita Hirosato Nomura
NTT Corp. Hitachi Ltd. Oki Electric Industry Ricoh Co., Ltd. Kyushu Institute
Co., Ltd. of Technology

This paper describes the guidelines for making an in-house evaluation of the technical level of a system from the developer's point of view. Technical issues have been classified into two groups --- those that depend on a specific language pair (Japanese as a source language and English as a target language) and those that are language independent. The system's strengths and weaknesses are identified by giving points for six fields corresponding to different evaluation criteria and for technical fields that evaluate the translation process.

1. はじめに

(社)日本電子工業振興協会の機械翻訳システム調査専門委員会技術動向調査ワーキンググループ¹⁾では、数年来、機械翻訳評価基準の検討を進めてきた。その成果は昨年に第1版と言う形で報告書にまとめ¹⁾、その概要は本研究会でも報告した²⁾。この評価基準は、昨年度の活動を通してさらに詳細化し、新たな観点からの評価基準も追加した。これは、評価基準第2版として平成4年度の電子協報告書の形でまとめた³⁾。報告書では3種類の観点(ユーザ側からの評価、開発者からの評価、品質面からの評価)からの機械翻訳システムの評価基準を提案した。我々のサブグループが担当した開発者からの評価基準は、機械翻訳システムの研究開発に従事している研究者、開発者、プログラマ、及び研究開発プロジェクトを管理している管理者自身が、自らが開発しているシステムを評価する際にチェックすべき評価項目をまとめた「開発者側からの評価基準」である。これは、そのシステムが技術的にどのレベルまで達成しているか、開発目的に合致したシステムとなっているかに関して、内部的に評価することを目標に作成したものである。機械翻訳システムを開発者側の立場から評価する場合、評価項目は多岐にわたり、それらを個別に評価することは困難であるため、本評価基準では評価の観点を示す6つの軸と技術レベルの評価軸を設け、各々の軸の観点から評価し、システムの特徴や弱点を明確にすることを試みた。

第2版の評価基準と第1版の評価基準との主な差分は、1) 翻訳言語対固有の翻訳技術を評価するための項目(平成4年度報告書には、日英翻訳技術に関するもののみ)を追加した、2) 設問項目に対する説明を追加した、3) 設問や選択肢を若干修正した、点である。

本稿では、「開発者側からの評価基準」第2版の概要について述べる。まず第2章では、本稿で提案する技術評価項目を用いてシステムを評価するための方法について述べる。次に3章では、第1版の評価基準で提案した言語対に依存しない汎用的項目について、昨年のN L研の報告で詳細に説明できなかった点を中心に概説する。次に4章では、第2版で新たに追加した項目である、日英翻訳技術に関する翻訳言語対依存の翻訳技術の評価項目について述べる。

2. 評価方法

評価基準では、システム全体をさまざまな角度から評価するために以下のような評価の観点を示す6種類の軸を設けて、関連する設問、回答にその軸を示す記号を付与している。

- ・汎用性/分野依存性：システムが汎用的なものか分野に特化したものかを示す

- ・機能網羅度：システムが基本的な機能を網羅しているかどうかを示す
- ・緻密度：システムが細かなところまで考慮して作成されているかどうかを示す
- ・オリジナリティ度：システムに用いられている要素技術にオリジナル性があるかどうかを示す
- ・システム解放度：システムがユーザに解放されているかどうかを示す
- ・使い勝手度：システムが使いやすいかどうかを示す

同時に、各設問、回答に技術的な難易度を示す記号を付与している。

- ・難易度大(A) = 3点
- ・難易度中(B) = 2点
- ・難易度小(C) = 1点

これは、以下のように分類したシステム内部の各処理を技術的に評価するためのものである。

- ・辞書
- ・解析技術
- ・中間表現と加工
- ・生成
- ・非文法的な現象の処理
- ・カスタマイズ、学習機能
- ・環境、操作
- ・言語対依存解析技術(日英翻訳技術)

本評価基準での評価方法は、

- 1) 機械翻訳システムの開発者などが、そのシステムに関する評価基準の設問に回答する。
- 2) 各評価軸(汎用性/分野依存性、…)ごとに、軸の記号が回答に付与されている場合に1点として合計点数を求める。また、各処理(辞書、解析技術、…)ごとにA=3点、B=2点、C=1点として技術的な難易度の合計点数を求める。
- 3) 各評価軸・処理ごとに満点を100に換算して評点を求める。
- 4) 各評価軸の点数をもとに目標としているシステムとのギャップを認識する。また、各処理の点数から技術的に進んでいる部分および遅れている部分を認識する。

図1に、評価基準の設問の例を示す。ここで、設問の選択肢に付与されている記号FとDは、それぞれ、機能網羅度、緻密度の評価に影響を及ぼすことを意味し、記号Cは、その選択肢の技術的な難易度が低いことを示す。

- (f) 連語抽出処理を行なっていますか。
 いいえ
 はい (F)
- (f-1) 変化形に対応できますか。
 はい (C)
 いいえ
- (f-2) 条件によって連語にしないものもありますか。
 いいえ
 はい (C, D)
- (f-2-1) 連語にしない場合の条件は何ですか。
 前後の単語の品詞
 その他 ()

¹⁾ 平成5年度から自然言語処理応用システム専門委員会機械翻訳分科会

図1 評価基準の設問例

3. 汎用的評価項目

機械翻訳システムには、翻訳言語対に依存しない共通的な技術が多数存在する。ここでは、このような言語対に依存しない汎用的な技術の評価する項目について述べる。ただし、紙面の都合で実際の設問は割愛させていただく。

3.1 システムの特徴

(1) システムの概要

ここでは、システム名、開発者名、動作するコンピュータなどシステムの概略を知るための質問を行なう。これらは開発者自身がシステムを評価する際には分かり切ったことかも知れないが、第三者にシステムを説明する際には重要な項目である。この項目によってシステムがどういったコンセプトで作られているかがはっきりする。なお、システムの概要では評価軸による採点はない。

(2) 言語対

ここでは、翻訳対象である原言語（ソース言語）と翻訳結果である目標言語（ターゲット言語）の種類や、1つのシステムで言語対の双方向に翻訳か、ソース言語やターゲット言語を複数個持つ（多方向）かに関する項目を設定した。

(3) システムの特徴

システム開発者が想定している用途について質問する。システムが対象言語や分野に依存性を持つ場合には、その対象言語や分野に関して高い翻訳性能を確保できる可能性があるが、汎用性はなくなる。また、翻訳結果の利用方法によって翻訳の要求品質や機能、速度性能等に制約が生じる。ここでは、対象言語の特性・対象分野への依存性、要求される訳文品質、想定ユーザ、利用形態、速度に関する項目を設定した。

3.2 辞書

機械翻訳においては、辞書にどのような情報が記述されているかが翻訳性能に直接的に影響する。以下の項目では、辞書の各側面を明らかにするための質問を用意した。

(1) 辞書の種類について

ここでは、どのような種類の辞書がシステムに用意されているかを質問する。一般的には、基本辞書、専門用語辞書、ユーザ辞書、学習辞書などが用意されている。通常、システムの持つ辞書の種類や登録語数が多ければ翻訳性能は高い。しかし、実際には、システムが適切な訳語を選択して翻訳結果として出力するには、該当する語の辞書登録は必要条件とはなるが十分条件とはならない。ここでは、はっきりと答えられる辞書の種類と登録語数のみの項目を設定した。

(2) 複数の辞書の使用

基本辞書に加えて同時に使用できる専門用語辞書やユーザ辞書の数は、きめ細かな訳文の生成に影響する。ここでは、これに関して質問する。

(3) 基本辞書の記述内容について

ここでは、基本辞書の記述能力について質問する。記述能力の高い辞書を持つシステムの方が翻訳の対象文書に適した訳し分けが行なえる。特に、共起関係（ある単語がある特定の単語あるいは特定の意味素性を持つ単語と同一文中に出現することによってその語意が決定されること。例えば、"take a shower"の訳として「シャワーを浴びる」という訳を得るためには、動詞"take"が目的語に"shower"をとれば「浴びる」という訳語を持つという情報が辞書に記述されていればよいことになる。）の記述は重要であり、この記述に用いられる意味素性や概念辞書（語の概念的な関係を記述した辞書）の有無も大きく影響する。

(4) 辞書の拡張

翻訳性能を高めて、所望の訳語を得るためには辞書の更新を行なう機能が重要である。しかし、辞書を不用意に編集することによって思わぬ副作用が出てしまうことがあるため、システムによっては編集できる項目に制限を設けている場合がある。ここでは質問項目では、辞書の作成・登録・編集機能について取り上げる。

(5) 電子辞書

機械翻訳処理用の辞書を人間が辞書引きのための電子辞書として使用することができるかに関して質問する。また、ポストエディット等に活用できるかについて質問する。

(6) コーパス

翻訳システムがどのようなコーパスを持っているかは良い機械翻訳システムを構築するためには、非常に重要である。また、単に生の文章を収集するだけでなく、対訳や品詞などの情報が付与されていれば利用価値が上がる。ここでは、翻訳システムにコーパスが利用できるかの観点で項目を設けた。

3.3 解析技術

ここでいう解析技術とは入力方式、形態素解析、構文解析、意味解析、文脈解析の5つである。システムによっては構成としてこの5つに分かれているとは限らず、後の2つは構文解析の部分として扱っているシステムが多いと思われるが、ここでは、機能としてこれらの処理を行なっているかを質問している。これらの処理がどれぐらい深く行なわれているかが一般的には翻訳性能に大きく関わる。

(1) 文字の入出力方式

入出力は、最初からワープロで作成された場合のように、電子化されたテキストの場合には問題にならないが、紙に書かれた文章を対象とする場合や、音声入力の場合には、文字認識、言語解析と音声認識処理の融合などを考慮して、システムを構築する必要がある。ここではその点を中心に質問する。

(2) 形態素解析

形態素解析は原言語の文書から単語・文章単位を認識し、単語の活用形、品詞などの辞書情報を後段の構文解析に渡す役目をする部分である。形態素解析に誤りがあると絶対に正しい解析結果が得られないため非常に高い解析性能が要求される。また、原言語の種類によって要求される機能・処理が大きく変わる。ここでは、一部言語固有の処理も含め項目を設定した。

(3) 構文解析

構文解析は形態素解析で得られた原言語の単語の単位、活用形、品詞などの情報をもとに、単語間の修飾関係を認定し、構文構造を決定していく部分である。構文解析は、システムによってただ単に構文構造を決定していくものから、意味処理や文脈処理も同時に行なって全体の文構造を決定していくものまで様々な手法を取り入れたものがある。修飾関係の認定では解析多義が発生することが多いため、その多義の取り扱い(格納、絞り込み)も重要な検討課題となる。また、構築する構文構造は、原言語の種類によって異なる場合がある。本処理もやはり、正しい解析結果が得られないと正しい翻訳結果が得られないので、高い解析性能が要求される。ここでは、下記の4項目に関して設問を設定した。

- ・文法—構文解析に利用する文法の種類
- ・解析アルゴリズム—bottom upかtop downか、パラレル型かシリアル型か
- ・個別解析技術—省略要素認定、テンス・アスペクト・モーダル解析、疑問文解析、倒置文解析、並列表現解析、感嘆文解析、命令文解析、付加疑問文解析、同格表現解析、複文解析、不定詞の用法解析、現在分詞解析、過去分詞解析、時制の一致
- ・構文構造解析—共起・相関語句の扱い、複文の曖昧性、自由格/adjunctの認定能力、依存関係解析

(4) 意味解析

意味解析は、単語間の意味的關係の認定や、単語の語義の決定等を行ない、より適切な訳語を得るための処理である。ここでは、意味解析の種類、意味的構造の曖昧性の取り扱い、意味表現形式等が評価項目となる。

(5) 文脈解析

ここでは、意味解析までで解消できなかった曖昧性の解消や省略要素の補完などに必要な要素技術である文脈解析についてどの程度行なわれているかを質問する。また文脈解析を実際に行なうためにはかなり深い知識が必要となるので、ドメインを限っているかも評価項目として挙げている。

3. 4 中間表現と加工

ここでは、どのような翻訳方式を採用しているかについて質問し、さらに各方式毎に設問を設定

した。

3. 5 生成

ここでは、訳語選択、自然な訳語を生成するための処理などについて質問する。

(1) 訳語選択

訳語選択処理がどの程度行なわれているかに関する技術項目を設定した。

(2) 非文のチェック

訳文が非文かどうかのチェックを行なっているかに関する技術項目を設定した。

(3) 生成処理

生成処理について、自然な目標言語を生成するための特別処理、形容詞の語順、副詞相当句の生成位置に関する技術項目を設定した。

3. 6 非文法的な現象の処理

ここでは、一般的な文法現象とは異なる言語現象を含むテキストの処理について質問する。これらの処理は、言語学などで扱われることはないが、システムの性能、ユーザの使い勝手の両面から非常に重要である。箇条書、タイトル文、引用句・挿入句、図・表、数式、手紙などの定型、日付、数量に関する表現、記号、文法範疇外の文等に関する技術項目を設定した。

3. 7 カスタマイズ・学習機能

ここでは、システムをユーザ毎にカスタマイズする機能に関する項目を設定する。カスタマイズには、2種類あって、1つはユーザ固有の情報などで一般的でないものを追加することである。もう1つは、既にシステムが持っている情報に対して、必要なもののみを取り出したり、優先順位を変更することである。

(1) 文種・分野

文種・分野によるカスタマイズは、システムが持っている情報の変更には属する。例えば辞書において、1つの見出し語に対して、多くの訳語が存在するが、特定の分野では、1つの訳語しか用いられないことがよくある。このような現象を利用することにより、システムをカスタマイズすることができる。ここでは、文種・分野の種類、フィルタプログラムに関する技術項目を設定した。

(2) 形態素解析

大文字小文字の区別に関してカスタマイズ出来るかの技術項目を設定した。

(3) 解析

プリファレンス、解析の時間制限、生成する解の個数に関する技術項目を設定した。

(4) 生成

生成過程でのカスタマイズとして、文体、態に関する技術項目を設定した。

(5) 辞書

辞書は、最も一般的なカスタマイズ手段である。

辞書のカスタマイズにおいては、新しい情報の追加と既に登録済みの情報の変更の両方が行なわれる。すなわち、登録されていない見出し語や訳語に対しては、新たにユーザ辞書に登録することが行われる。また、既に登録済みの訳語に対して、優先順位を変更したり、必要な訳語のみをユーザ辞書に登録することが行なわれる。ここでは、ユーザ辞書、辞書の学習機能に関する技術項目を設定した。

(6) 文法

文法のカスタマイズは、辞書のそれに比べてかなり困難である。翻訳精度を向上させ、文法変更に伴う副作用を最小限に抑えるには言語学的知識やシステムで使用している文法に関する知識が必要となる。ここでは、ユーザへの開放度、制限文法、文法の学習機能に関する技術項目を設定した。

3. 8 環境・操作

ここでの質問は、機械翻訳システムの本質的な性能(訳文の質、翻訳速度)とは関係なく、主に使い勝手に関するものである。ただし、このことは、この質問が重要ではないことを意味するわけではない。現在の機械翻訳システムは、完全自動翻訳のレベルには達しておらず、何らかの形で人手の介入を必要とする。そのため、その機械翻訳システムを利用することにより、翻訳コストの削減、翻訳時間の短縮などが実現できるかどうかは、ここで質問されている機能がサポートされているかどうかによって強く影響される。ここでは、入力、前編集、インターエディット、後編集、開発環境に関する質問を設けている。

4. 言語対依存評価項目

機械翻訳において評価すべき項目は、それがある自然言語を別の自然言語に変換するものである。それら言語対に依存するものが多く存在する。ここでは、翻訳言語対に固有な技術項目について説明する。本稿では、日英翻訳技術のみについて列挙する。

4. 1 述部

ここでは、日本語の述部に関して日英機械翻訳で考慮すべき技術項目について述べる。原言語である日本語と目標言語である英語は言語構造的にかなり異なる。この構造の違いをうまく変換する枠組みが必要となる。述部の翻訳の枠組みでこれらの言語間の構造的な違いを取り扱おうとするシステムも多い。

(1) 述部の訳し分け

述部、とくに動詞や形容動詞は、ほとんどの場合訳し分けが必要となってくる。訳し分けのために用いる情報としては、助詞、名詞の意味属性、副詞との共起等が一般的であるが、文脈を用いる方法なども近年提案されてきている。ここでは、訳し分けの可能性とその手法についての項目を設

定した。

(2) 断定文

断定文とは、「AはBだ」のタイプで、Bが名詞である文のことである。一般的にこのタイプの文は、「A is B.」と訳すケースが多いが、Bを用言化したり、構造自体を大幅に変更したりする必要がある場合も多い。ここでは、訳し分けの可能性、解析手法に関する項目を設定した。

(3) 体言述語

「同社では商用サービスに移る予定」の「予定」は名詞述語というよりも、機能的には用言である。このような体言述語が正しく動詞や形容詞として認識できるかという点が機械翻訳システムでは、正しい訳を生成するためには必須である。ここでは、このような体言述語の処理可能性と、その手法に関する項目を設定した。

(4) 複合述部

「牛乳は栄養豊富である」の「栄養豊富」のように、述部が連続複合語となっている場合、複合語の単語間の関係を正しく認識することが必要となる。例えば、この例の場合は、「栄養」が「豊富」であるという意味に解析する必要がある。このような技術は、機械翻訳システムでは、処理対象分野によっては出現頻度が高いので、必須技術となる。ここでは、処理可能性と、複合述部内の要素で認定できる関係(並列用言、格関係等)の種類に関する項目を設定した。

(5) 並列用言だが訳は1用言となる述語表現

「犬が子供に飛び掛かる」の「飛び掛かる」は用言「飛ぶ」と「掛かる」の並列表現であるが、英語で訳す場合には、「jump」の1用言となる。このような表現を翻訳する場合には、並列用言として訳したほうが良い一般的な表現と区別する処理が必要となってくる。ここでは、処理可能性等に関する項目を設定した。

(6) 補助動詞

補助動詞はmodal等の情報を抽出するために必要となり、本動詞との区別も必要となる。ここでは、みる、いく、いる、ある等の補助動詞について本動詞との区別可能性等に関する項目を設定した。

(7) 基本動詞の言い換え

基本述語(ある、ない、おいしい、すくない、できるなど)が原文中にはない訳語に翻訳される場合がある。(例、「温度は一定とする。」→"It is assumed that the temperature is constant.")ここでは、このような処理が出来るかに関する項目を設定した。

4. 2 名詞

ここでは、日本語名詞を英語に翻訳する際に考慮すべき点について質問する。日本語の名詞体系は、英語の名詞体系とはかなり異なるので、訳し分けが必要であったり、品詞を変換する必要があったりする。また、固有名詞のような、バリエーショ

ンの大きい表現も頻繁に出現する。名詞の翻訳は、文章内容を正確に伝えるうえで、特に正確に翻訳する必要があるため、極めて重要な技術と言える。

(1) 訳し分け

ここでは、訳し分けの可能性と手法に関する項目を設定した。

(2) 複合名詞

複数の単語から構成され慣用的に1語として扱われるものを複合名詞と呼ぶことにする。ここでは、複合名詞の範囲認定、中心名詞の認定、複合名詞内の係り受けに関する項目を設定した。

(3) 「〈名詞1〉の〈名詞2〉」型表現

名詞1が名詞2を意味的に限定する関係で結びつく最も一般的な形は、「〈名詞1〉の〈名詞2〉」という形である。助詞「の」により結びつけられた2つの名詞の意味関係は多様であり、その関係を正しく認定することは重要な課題である。ここでは、認定できる修飾関係、認定方法等に関する項目を設定した。また、「〈名詞1〉の〈名詞2〉の〈名詞3〉」型表現も同様に項目を設定した。

(4) 並列構造を持つ名詞句

2つ以上の名詞が助詞「と」や助詞「や」などを介して並立的に結合して、全体として名詞句を形成することがある。たとえば、「私は学校と公園に行った」という例においては、「学校」と「公園」という2つの名詞が助詞「と」によって連結され、「学校と公園」というまとまりで1つの名詞句を形成している。しかし、一見同じような構造を持つように見える「私は彼と公園に行った」という表現では、「彼」と「公園」は並列関係にはない。これは、いくつかの名詞をつないで1つの名詞句を形成するのは、個々の名詞が或る集合のメンバーとして何か共通の特徴を持つという見方から生まれているものだからである。並列構造を持つ名詞句の出現頻度は高いので、並列構成要素を正しく認定することが必要である。ここでは、並列構造の認定方法、英文構造への反映方法に関する項目を設定した。

(5) 固有名詞表現

固有名詞表現は、一般的な技術文章、新聞等の分野を問わず頻繁に出現するが、そのバリエーションが余りにも多いため、全ての語を登録することは困難である。よって、固有名詞に対しては何らかの対処法を検討する必要がある。ここでは、固有名詞の処理方法、変換方法、処理可能な固有名詞表現に関する項目を設定した。

(6) 用言性名詞(サ変名詞)

用言性名詞とは「する」が付くと動詞化する名詞のことを言う。この名詞は、動詞的な振る舞いもするので、一般的な名詞と区別して取り扱う必要がある。翻訳する際には、品詞を用言に変更したり、動名詞化したりすることが必要となる場合がある。ここでは、用言性名詞を区別するか、翻

訳スタイル、用言性名詞を含む複合語の処理に関する項目を設定した。

(7) 形式名詞

形式名詞(「こと」、「もの」など)は、用言に後接して名詞句を形成する。"thing"のように訳す場合もあれば、to不定詞やthat節として訳す場合もある。ここでは、形式名詞の処理方法に関する項目を設定した。

(8) 関係を示す名詞

ここでは、ものや概念に関する関係を示す名詞についての取り扱いに関して質問する。関係を示す語は、日本語と英語で、語と実際の関係の間の対応関係(範囲)がかなり異なるため、特に注意して取り扱う必要がある。例えば、「上」という語は、「食器はテーブルの上にある」の「上」のようにある物がある物に直接接触した上にある場合には"on"と訳す必要があるが、「雪が車の上を覆っている」の「上」の様にある物が有るものの上を覆っている場合には"over"と訳す必要がある。このように、厳密にこれらの選択をするためには、文章中のもの(概念)の間の実際の関係(形態)を具体的に抽出(解析)して、訳語を選択する必要がある。ここでは、「上下」、「前後」、「左右」、「中・内・外」、「間」、「横・そば・内」の関係を示す名詞の処理に関する項目を設定した。

4. 3 副詞

ここでは、日本語の副詞に関して日英機械翻訳で考慮すべき技術項目について述べる。日本語副詞は英語でも副詞に翻訳出来るとは限らず、品詞が変わったり、大幅に構造が異なる場合もある。

(1) 副詞

副詞には、「昨日買った」の「昨日」のように動詞を修飾する場合、「とても美しい女性」の「とても」のように形容詞を修飾する場合、「非常にゆっくり歩く」の「非常に」のように副詞を修飾する場合、「ちょうど真ん中」の「ちょうど」のように名詞を修飾する場合などがある。ここでは、取り扱える修飾先、訳し分け、訳出出来る品詞、訳出位置、時制・アスペクトの変更、動詞連用型・形容詞連用型副詞の処理に関する項目を設定した。

(2) 副詞句

副詞句は成句的な要素(「一般的に言って」等)と連語要素(「～により」等)等がある。それぞれ、英語に翻訳する際には品詞の変更が必要となる場合が多い。ここでは、これらの認定方法と、訳出出来る品詞に関する項目を設定した。

(3) 副詞節

「彼はバスに乗って学校へ行く」という日本語文を"He goes to school by bus."という英語文にうまく訳出しようとするためには、「バスに乗って」を動詞の並列ではなく、副詞的に扱うための処理

が必要である。ここでは、連用型接続の副詞節の処理、品詞の変更、形式的な名詞を含む副詞節の処理に関する項目を設定した。

4. 4 連体修飾語句

連体修飾語句には、いわゆる限定用法の形容詞のほかに、連体詞、格助詞相当語、埋め込み文修飾要素等がある。ここでは、日英翻訳システムにおいて特に取り扱いに注意すべきこれらの表現についての設問を設けた。

(1) 連体詞

連体詞は、体言を修飾する機能を持つ。よって、日本語では形容詞的な取り扱いをするが、英語に翻訳する場合には形容詞に限らず様々な品詞で翻訳できることが必要となる。ここでは、処理可能な連体詞、訳し分けに関する項目を設定した。

(2) 用言性連体詞

用言性連体詞は、英語に翻訳する際に、形容詞として訳す必要がある、一般的な埋め込み文とは異なった取り扱いをする必要がある。また、解析の際に、用言性連体詞であるか否かの判断をすることによって、解析の際に形容詞的に取り扱うか、埋め込み文修飾している要素として解析した後、形容詞に翻訳するかを判断して処理する必要がある。ここでは、処理可能な用言性連体詞、訳し分け、翻訳スタイル、認定方法に関する項目を設定した。

(3) 格助詞相当句

いくつかの語で構成される句が、全体として格助詞に相当する働きをすることがある。このような句を「格助詞相当句」と呼ぶ。格助詞が多様な用法を持っているのに対して、格助詞相当句は用法が限定されていることが多い。ここでは、認定法、処理可能な格助詞相当句に関する項目を設定した。

(4) 埋め込み文修飾

埋め込み文は、それが修飾する名詞（先行詞ともいう）に対する修飾の仕方の違いにより「補足語修飾節」、「相対名詞修飾節」、「内容節」という3つの用法に大別することができる。これらの埋め込み文は、英語においては関係節や同格節といった構造で訳出されることが多いが、先行詞との修飾関係を正しく解析しておかなければ適切な英語構造を生成することができない。ここでは、用法の決定、認定できる修飾関係、認定した結果の英文への反映、先行詞が構造化された名詞句の場合に関する項目を設定した。

4. 5 助詞

日本語における助詞は、格の機能（深層格）を決定するだけではなく、副詞的な様相を生成したり、文脈構造を規定したり、様々な働きをする。日英翻訳においては、前置詞の決定、副詞の生成等で助詞の解析は極めて重要となる。ここでは、

訳し分け、深層格の認定、認定可能な助詞、副助詞に関する項目を設定した。

4. 6 接辞

接辞は、名詞等に接続して、概念を追加したり（「上寿司」の「上」等）、語全体の機能を決定したり（「客観的」の「的」）する、独立性の少ない語のことを言う。翻訳の際には、語の機能によりかなり異なった取り扱いが必要となり、場合によっては、品詞の変更が必要な場合があるため、取り扱いには注意が必要となる。ここでは、処理できる接辞の種類・規模、解析方法、訳し分け、訳出方法に関する項目を設定した。

4. 7 テンス・アスペクト・モーダル

(1) テンス

テンスを表す最も重要な形式は、述語の基本形とタ形である。一般に、述語の基本形は述語の種類によって「現在」または「未来」の時を表現し、タ形は「過去」の時を表現する。たとえば「花子が部屋を掃除する」においては、述語「掃除する」は動態述語であり、テンスは「現在」か「未来」のどちらかとなる。この場合、「明日」のような未来を表す副詞と共起する（例、「花子は、明日、部屋を掃除する」）と、テンスは「未来」に限定される。このように、実際には、述語と共起している副詞や名詞によってテンスが決定されることが多い。さらに、文脈情報を参照しないと決定できない場合もある。ここでは、解析方法、多義解消、生成処理への反映に関する項目を設定した。

(2) アスペクト

動きの展開の様々な局面を表す文法形式であるアスペクトは、種々の表現を持ちうる。アスペクトの形式には、「動詞のテ形+イル、アル、シマウ、イク、クル」といった形式や「動詞の連用形+ハジメル、ツツケル、オワルのような動きの開始、継続、終結等を表す動詞」という形式がある。そのほか、アスペクトを表す語句に、「～ところだ」、「動詞の意志形+とする、としている」、「～ばかりだ」等がある。ここでは、解析方法、多義解消、「動詞のテ形+イル」形の処理、生成処理への反映に関する項目を設定した。

(3) モーダル

文が客観的事態だけを表すというのは、接続節の一部等非常に限られており、その事態や聞き手に対する話し手の様々な判断や態度が同時に表現されるのが普通である。このような事態や相手に対する話し手の判断・態度を表す文法形式がモーダルであり、述語の活用形、助動詞や終助詞といった形式で文末に表現される。文に含まれる客観的事態とモーダルをそれぞれ正しく認定することは重要な課題である。ここでは、解析方法、処理可能なモーダル、生成処理への反映に関する項目を設定した。

4. 8 特殊構造表現

(1) 慣用表現

慣用表現は、要素である単語の本来の意味が失われ全体としてある意味を持つような表現である。よって、慣用表現を字義通りに機械翻訳すると全く意味が通らない文になってしまう。このため、慣用表現に対しては特別の処理が必要となる。ここでは、固定的な慣用表現、変形可能な慣用表現、呼応表現、慣用表現処理適用の制御に関する項目を設定した。

(2) 天気・気象表現

英文では天候・気象表現において、"rain","snow"のような非人称動詞が良く用いられるため、非人称の"it"を補う必要がある。また、言語間の発想の違いから、"we"のような総称的な人称代名詞を補う必要がある。ここでは、これらに関する項目を設定した。

(3) 「はが」構文

「AはBが～」という構文は、日本語では頻出する表現であるが、「は」は非常に多くの意味で用いられるため、処理が困難である。ここでは、主格・目的格の「はが」、所有関係・部分全体関係の「はが」、副詞句の「は」等に関する項目を設定した。

(4) 比較表現

副詞的表現の処理の一部と考えられるが、特に良く用いられるものであるため項目としてまとめた。数量・程度などを比較するとき用いられる表現である。ここでは、同等比較表現、相対比較表現、最上比較表現等に関する項目を設定した。

(5) 比喩表現

比喩表現は、科学技術文献ではそれほど用いられないため実用システムで処理がサポートされているものはすくないと思われるが、翻訳対象分野によっては、頻出するため大きな問題となる場合がある。ここでは、比喩的慣用句、直喩、陰喩に関する項目を設定した。

5. まとめ

開発者自らが機械翻訳システムを評価するための、評価の観点を示す6つの軸と技術レベルの評価軸による評価を行なう設問を示した。紙面の都合上、具体的な例は示さなかったがそれぞれの項目に関する設問に開発者が答えていくことにより、自らが開発している機械翻訳システムの評価を客観的に行なうことが出来る。

この評価により、

1. 評価軸ごとのシステムの総合評価
2. 各項目ごとのシステムの評価

が得られ、総合的・個別的なシステムの長所、短所が浮かび上がる。たとえば、自社のシステムの使い勝手の評価や、解析系の弱点の断定といった評価が行なえる。

本稿で提案した評価項目、および評価軸、難易度は、日本で一般的と思われる機械翻訳システムを想定し、主観的に決定したものである。したがって、できるだけ客観的な評価基準となることを心掛けて作成しているが、項目によって細かい部分と粗い部分のある可能性がある。また、特殊なタイプのシステム、たとえば、Meteoのような(天気予報の翻訳だけを行なうといった)システムでは全体に低い評価しかえられないため、そういう場合には、システムの実用性と本評価とが対応しなくなる。こういったシステムを評価する目的には、本評価基準は充分ではない。本評価基準は、それぞれの翻訳システムが基本的な機能をどれくらい備えているかを知るものであって、そのための指標である。

今後の課題として、さらなる評価項目の詳細化、評価軸の追加等の見直し、難易度の細分化と見直しなどがあげられる。また、言語対固有の評価項目としては英日翻訳技術に関する評価項目の追加も、日本で開発されている翻訳システムの種類を考えた場合に必要であろう。

本稿で提案した評価基準は、開発者がシステムを評価する目的だけではなく、これから翻訳システムを研究・開発しようとしている研究者開発者にとっても極めて有意義な情報が含まれている。今後は、書式を変更して、機械翻訳技術をまとめた百科辞典的な資料としていくことも、必要であろう。

参考文献

- [1]機械翻訳システムの実用化に関する調査研究, 92-計-1, (社)日本電子工業振興協会, 1992.3
- [2]野村,井佐原:機械翻訳の評価基準について, 情報処理学会自然言語処理研究会,89-9, 1992
- [3]機械翻訳システムの実用化に関する調査研究, 93-計-6, (社)日本電子工業振興協会, 1993.3
- [4]高山,伊藤,八木沢,茂木,野村:JEIDA機械翻訳システム評価基準(利用者編), 情報処理学会自然言語処理研究会,96-9, 1993
- [5]井佐原,新納,山端,森口,野村:JEIDA機械翻訳システム評価基準(品質評価編), 情報処理学会自然言語処理研究会,96-11, 1993
- [6]池原,白井:日英機械翻訳機能試験項目の体系化, 電子情報通信学会言語理解とコミュニケーション研究会,NLC90-43, 1990
- [7]益岡,田窪:基礎日本語文法一改訂版, くろしお出版, 1992
- [8]井上:変形文法と日本語 下, 大修館書店, 1976