

TR-IT-0188

## JEIDA 機械翻訳システム評価基準を用いた TDMT の評価

安藤 真一      隅田 英一郎  
Shinichi ANDO    Eiichiro SUMITA

1996年9月

### 要旨

本稿では、TDMT 評価の一つとして、言語現象への網羅性という観点から行った評価について報告する。ここでは特に JEIDA 日英機械翻訳システム品質評価用テストセットに記載された各種言語現象を対象とし、TDMT で正しく扱えるかどうかについて調査した。その結果、単文に起こり得るほとんどの言語現象については対応可能だが、複文、重文ではやや難が残る場合もあることが分かった。これは、特に部分的な情報だけでは訳が決められない場合にも決定的な処理を行っている点が原因であると考えらる。その解決策を含めて検討を進めたので、これについても報告する。

エイ・ティ・アール音声翻訳通信研究所  
ATR Interpreting Telecommunications Research Laboratories

©(株) エイ・ティ・アール音声翻訳通信研究所  
1996 by ATR Interpreting Telecommunications Research Laboratories

## もくじ

1	はじめに	1
2	評価方法	2
3	評価結果概要	3
4	問題点の整理と対策の検討	5
4.1	ローカル辞書	5
4.2	下位構造の参照	5
4.3	構造変換	6
4.4	ルール間の整合性	7
4.5	意味距離計算の妥当性	7
4.6	並列構造の処理	8
4.7	文脈処理	8
4.8	ルール作成	8
5	評価結果の詳細について	10
5.1	述部	10
5.1.1	述部の訳し分け	10
5.1.2	断定文	10
5.1.3	体言述語	11
5.1.4	複合述部	12
5.1.5	訳が一用言となる並列用言	13
5.1.6	用言の副詞(句)化	13
5.1.7	補助動詞	13
5.1.8	基本動詞の訳し分け	14
5.2	名詞	14
5.2.1	名詞の訳し分け	14
5.2.2	複合名詞	15
5.2.3	「名詞1の名詞2」という構造を持つ名詞句の処理	16
5.2.4	「名詞1の名詞2の名詞3」という構造を持つ名詞句の処理	16
5.2.5	並列構造を持つ名詞句の処理	17
5.2.6	疑問表現の名詞節の処理	18
5.2.7	用言性名詞(サ変名詞)	18
5.2.8	英語における数の扱い	19
5.2.9	固有名詞表現	19
5.2.10	形式名詞	20
5.2.11	関係を示す名詞	21
5.3	副詞	21
5.3.1	副詞のタイプ	21
5.3.2	副詞句	22

5.3.3	擬音語・擬態語	23
5.4	連体修飾語句	23
5.4.1	非活用連体修飾	23
5.4.2	用言性連体詞	24
5.4.3	格助詞相当句	25
5.4.4	埋め込み文修飾	25
5.5	助詞	26
5.6	接辞	27
5.7	テンス、アスペクト、モーダル	28
5.7.1	テンスの処理	28
5.7.2	アスペクトの処理	28
5.7.3	モーダルの処理	29
5.7.4	ボイスの処理	29
5.8	特殊構造表現	30
5.8.1	慣用表現の処理に関して	30
5.8.2	四字熟語	30
5.8.3	呼応表現	31
5.8.4	天候・気象表現	31
5.8.5	無生物主語構文	31
5.8.6	「はが」構文	32
5.8.7	比較表現	32
5.8.8	比喩表現	33
5.8.9	部分否定、二重否定、倒置文	33
5.8.10	敬語	34
5.8.11	引用・伝聞表現	34
5.8.12	例示・列挙表現	34
6	おわりに	35
A	評価対象文一覧	36

## 1 はじめに

ATR 音声翻訳通信研究所の中間報告に際し、TDMT の現状把握と課題抽出を目的として品質評価を行ったので、これについて報告する。

機械翻訳システムの品質評価では標準的な評価方法が確立しておらず、従来から様々な手法が提案されている [1]。

この中で最もよく行われているのが翻訳結果の適切さを人手で主観的に評価する方法である。例えば、アジア太平洋機械翻訳協会 (AAMT) はシステムの出力した文が翻訳文として通用するかどうかを 3 段階の基準で評価しており [2]、また ARPA はシステム出力文の各節に対して理解容易性、忠実度などを各々 5 段階の基準で主観評価している [3]。このような主観評価は、翻訳処理全体の性能を人間や他のシステムのそれと比較できるため、システムの現状把握には重要である。

しかしこの評価法では、評価結果からシステムの問題点を特定することが難しく、システムの改良には結びつきにくい。このため様々な言語現象を網羅的に収集した例文集を作成し、その個々について評価を行う手法 [4, 5] や、モジュール毎に評価を進める手法 [6] も提案されている。

そこで我々は両者の利点を取り込んで総合的に TDMT の品質評価を行うために、下記の 2 つの評価を独立に行った。

1. 理解容易性、忠実度
2. 各種言語現象 (文法事項) への網羅性

これは、1. の主観評価によってシステム全体の性能を評価し、2. で各言語現象に対するシステムの対処法を調べることによってシステム設計上の問題点やルール記述における改良点を明らかにしようとするものである。

本稿では、特に後者の文法的側面に焦点をあてた評価について述べる<sup>1</sup>。まず、その評価方法と各文法事項における評価結果の概略を記し、主な問題点とその解決策についてを述べる。さらに、各々の文法事項について詳細に考察する。

---

<sup>1</sup>前者については TR-IT-189、TR-IT-190、TR-IT-191 を参照

## 2 評価方法

我々は評価用の例文集として、日本電子工業振興協会 (JEIDA) が翻訳システムの対応すべき言語現象と評価文をまとめた JEIDA 日英機械翻訳システム品質評価テストセット [7] を採用した。

上記のテストセットは書き言葉を対象として構成されており、話し言葉は特に意識されていない。しかし、両者のコアとなる部分には共通の現象も多く、ここに列挙された言語現象のほとんどは話し言葉でも現れるものである。このため、TDMT も音声翻訳の要素技術としてこれらを網羅的に扱える必要がある。

そこで今回は上記テストセットに記述された言語現象を対象として、その各々について評価を行った。

ただし、TDMT では旅行会話やビジネスレターを対象ドメインとしており、対象語彙が異なるため、上記テストセットに収められた評価文をそのまま処理することはできない。このため評価文に対する形態素、変換辞書、シソーラスや変換ルールを実際に記述しながら、

1. TDMT 変換ルールによる文法記述力
2. TDMT システムにおける問題点およびその解決策

を中心に検討することにした。

ここでは形態素解析が正しい結果を出力することを前提とし、さらに評価文に出現する単語は全て用例に加えた。このため評価文が翻訳できたとしても、システムがその言語現象に対応しているかどうかは分からない。そこで今回は、特に

- 適切な情報を利用して、適切な構造、訳語を与えているか
- 一般性があるか
- 他のルールとの衝突が起らないか

に注意して検討を進めた。

### 3 評価結果概要

本章では TDMT が各々の文法事項を扱えるかどうかについて検討した結果をまとめる。各々の結果の詳細については 5 章「評価結果の詳細について」で述べる。

以下に評価対象とした文法項目とその各々に対する評価結果を表形式で示す。ここで各項目は、JEIDA 日英機械翻訳システム品質評価テストセットで取り上げられた文法項目をそのまま採用した。また評価結果は、○で「ほぼ扱える」、△で「特定の機能追加などで対応可能」、×で「扱えない、あるいは扱いが不明」を表している。ただし未知語処理については考慮していない。

(1)述部	(1-1) 述部の訳し分け	○
	(1-2) 断定文	○
	(1-3) 体言述語	○
	(1-4) 複合述部	○
	(1-5) 訳が一用言となる並列用言	○
	(1-6) 用言の副詞(句)化	○
	(1-7) 補助動詞	○
	(1-8) 基本動詞の訳し分け	○(一部△)
(2)名詞	(2-1) 名詞の訳し分け	評価文なし
	(2-2) 複合名詞	○
	(2-3) 「名詞1の名詞2」という構造を持つ名詞句の処理	○
	(2-4) 「名詞1の名詞2の名詞3」という構造を持つ名詞句の処理	○
	(2-5) 並列構造を持つ名詞句の処理	△
	(2-6) 疑問表現の名詞節の処理	△
	(2-7) 用言性名詞(サ変名詞)	○
	(2-8) 英語における数の扱い	△(一部×)
	(2-9) 固有名詞表現	○
	(2-10) 形式名詞	△
	(2-11) 関係を示す名詞	△
(3)副詞	(3-1) 副詞のタイプ	○(一部△)
	(3-2) 副詞句	△
	(3-3) 擬音語・擬態語	○
(4)連体修飾語句	(4-1) 非活用連体修飾	○(一部△)
	(4-2) 用言性連体詞	△
	(4-3) 格助詞相当句	○
	(4-4) 埋め込み文修飾	△
(5)助詞		○(一部△)

(6)接辞		○
(7)テンス、アスペクト、 モーダル	(7-1) テンスの処理	△
	(7-2) アスペクトの処理	△
	(7-3) モーダルの処理	△(一部×)
	(7-4) ボイスの処理	○(一部△)
(8)特殊構造表現	(8-1) 慣用表現の処理に関して	△
	(8-2) 四字熟語	○
	(8-3) 呼応表現	△
	(8-4) 天候・気象表現	○
	(8-5) 無生物主語構文	△
	(8-6) 「はが」構文	○
	(8-7) 比較表現	○(一部△)
	(8-8) 比喩表現	○(一部×)
	(8-9) 部分否定、二重否定、倒置文	○(一部△)
	(8-10) 敬語	○
	(8-11) 引用・伝聞表現	△
	(8-12) 例示・列挙表現	△

## 4 問題点の整理と対策の検討

本章では、今回の評価で明らかになった問題点のうち主なものを横断的に整理し、その各々について対策を検討する。

以下では、問題点毎に「関連する言語現象」「問題の説明」「対策案」を記す。この「関連する言語現象」では3章の「評価結果概要」で用いた文法項目と同じ項目を用い、またその詳細について記した章の番号を各々に()で付す。

### 4.1 ローカル辞書

#### [関連する言語現象]

述部の訳し分け(5.1.1)、断定文(5.1.2)、体言述語(5.1.3)、用言の副詞(句)化(5.1.6)、基本動詞の訳し分け(5.1.8)、副詞のタイプ(5.3.1)、擬音語・擬態語(5.3.3)、用言性連体詞(5.4.2)、慣用表現の処理に関して(5.8.1)、無生物主語構文(5.8.5)、「はが」構文(5.8.6)

#### [問題の説明]

TDMTでは、原言語のデフォルト訳は変換辞書に記述されているが、このデフォルト訳は1つだけしか記述できない。このため訳語を変更する場合には、変換ルール内のみで有効なローカル辞書を用いて訳語や品詞を変更する。ここでデフォルト訳と異なる品詞を与えるようなローカル辞書を利用すると、そのルールでの出力パターン(構文)は、ローカル辞書を含まない通常のルールと異なる場合が多い。このため、ローカル辞書を含むルールで、デフォルト訳のみを使う単語列が処理されると、訳文の構文がくずれてしまう可能性が大きい。また、訳語のみを変更する場合でも、その訳語の用法がデフォルト訳の用法と異なると同様のことが起こる。

#### [対策案]

○上記のような危険を避けるために、ローカル辞書を含むルールはその辞書項目を用例に記述するのではなく、パターン内の固定項として記述すべきである。これによって、他の一般ルールで処理されるべき単語列が、ローカル辞書を含む特殊ルールで処理されることを避けることができる。このとき、特に述部を処理するルールは述部に対する格関係を記述したルールとなり、格フレーム付きの辞書として働く。これを考慮すると、訳し分けのヴァリエーションが多い語(例えば、基本動詞)については、通常の辞書内にこのパターンを記述して管理して運用時にルールとして利用するなどして、管理方法を整理した方がよいように思う。

○品詞変換のためにローカル辞書を使用するのであれば、辞書内で派生語間の関係を記述するか、派生語処理を導入するなどして、ルール内からは品詞のみを指定することで品詞が変更できれば、ルールの一般性をより高めることができるだろう。

### 4.2 下位構造の参照

#### [関連する言語現象]

並列構造を持つ名詞句の処理(5.2.5)、疑問表現の名詞節の処理(5.2.6)、形式名詞(5.2.10)、関係を示す名詞(5.2.11)、副詞のタイプ(5.3.1)、副詞句(5.3.2)、非活用連体修飾(5.4.1)、用言性連体詞(5.4.2)、埋め込み文修飾(5.4.4)、(5.5)、テンスの処理(5.7.1)、アスペクトの処理(5.7.2)、モーダルの処理(5.7.3)、慣用表現の処理に関して(5.8.1)、呼応表現(5.8.3)、「はが」構文(5.8.6)、比較表現(5.8.7)、比喩表現(5.8.8)、引用・伝聞表現(5.8.11)、例示・列挙表現(5.8.12)



#### [問題の説明]

現行の TDMT ではルール間で受渡しできる情報が非常に限定されている。例えば、「動詞-接助-動詞」のような文と文をつなぐルールからは、各々の動詞の下位構造が参照できず、動詞にどのような助動詞が係っているのかを参照することができない。このため、例えば時制の一致をルール上で表現することは困難になる。また従属節の主語を主節の主語と一致させたい場合にも、これを記述することができない。

また、「名詞-格助-動詞」のような格関係を作るルールにおいても、動詞の下位構造にどのような格が存在するかをチェックすることもできないため、既に下位に存在する格と同じ格して解析してしまう危険性もある。また慣用句や副詞の場合には格要素の存在がチェックポイントになる場合もある。

さらに並列や関係節を扱う場合には、数、主語、時制、格などの情報をルール間で参照する必要がある。また、日本語で明示されていない所有格代名詞を付加するといった場合にも、主語の参照を必要とするものも多い。これらの場合でもお互いの下位構造の情報が参照できないために、処理できない。

#### [対策案]

○今回の評価で参照できないことが問題となったのは、名詞(句)では数の情報がほとんどであった。また、動詞(句)では時制、既に埋まっている格、主語が主なものであった。これらだけでも、まとまった構造全体に対する属性として定義し、ルール間の相互で参照できるようにした方がよいだろう。また、さらに受身や使役なども参照可能な情報として加えれば、2ヘッド構造を使わなくてもルールが記述できるようになると思われる。

○主語の一致や所有格代名詞の人称一致に関しては、参照先を指定する特殊な属性を導入することで、生成側で対処できるかもしれない。

### 4.3 構造変換

#### [関連する言語現象]

基本動詞の訳し分け(5.1.8)、「名詞1の名詞2」という構造を持つ名詞句の処理(5.2.3)、「名詞1の名詞2の名詞3」という構造を持つ名詞句の処理(5.2.4)、埋め込み文修飾(5.4.4)

#### [問題の説明]

TDMT は前から一つずつ形態素を読み込み、その各段階でルールを適用することで訳語を決定するが、ここでできた構造を後から変更することができない。例えば「彼の行動は素早い」という例を考えると、「彼の行動」まで処理した段階で「his action」と訳すことを決めてしまっており、この構造を後から変更できないため、「He acts quickly」と翻訳するルールを記述することは困難である。

#### [対策案]

○文全体を含む長いパターン(例えば“(XのYはZ)”)を書けば現状でも、ある程度は対応できるが、ルールの一般性が低くなってしまふ。また「彼の行動」=「he acts」のような曖昧性を持たせる手法もあるが、下位構造の参照やルール間の整合性が問題となり、また増加する曖昧性をどう解消すればよいのかが明らかではない。このため、日本語とほぼ同じ構造の訳文を作った後、必要な場合には後処理で全体構造を変更するなどの処理を導入した方がよいように思える。

#### 4.4 ルール間の整合性

##### [関連する言語現象]

断定文(5.1.2)、体言述語(5.1.3)、基本動詞の訳し分け(5.1.8)、「名詞1の名詞2の名詞3」という構造を持つ名詞句の処理(5.2.4)、呼応表現(5.8.3)、「はが」構文(5.8.6)、比較表現(5.8.7)

##### [問題の説明]

変数項や固定項を多く含んだ長いパターンを利用すると、ルールの一般性が失われ、様々なヴァリエーションに対応した大量のパターンを用意する必要が出てくる。このため、なるべく短いパターンでルール記述した方がルール記述の点では効率がよい。ただしルールの組み合わせ方が多様になり、誤ったルールセットが適用される可能性が増大するため、ルール間の整合性が問題となる。ルール間で整合性を保つためにはパターンにマッチする条件を厳しくする必要があるが、現行の TDMT では下位構造の参照ができないため、厳しい条件を設定することができない。

これは、構造変換を伴うルールの場合、特に深刻である。例えば、名詞句を動詞句として訳すルールが適用された後は、適用されるルールが通常とは全く異なってくる。これらが混同されると構文におかしな訳文が出力される可能性が高くなり、安定性を保証できない。

##### [対策案]

○変数項の値や品詞を指定することで十分に整合性がとれる場合もあり、これを十分に活用すべきだろう。ただし構造変換を伴うルールを考えると、現言語側の品詞だけでなく、目的言語側の品詞まで条件指定できた方がよいと思われる。

#### 4.5 意味距離計算の妥当性

##### [関連する言語現象]

複合述部(5.1.4)、複合名詞(5.2.2)、並列構造を持つ名詞句の処理(5.2.5)、非活用連体修飾(5.4.1)、アスペクトの処理(5.7.2)、比喩表現(5.8.8)、例示・列挙表現(5.8.12)

##### [問題の説明]

TDMT は用例との意味距離計算に基づいて、深層格や他の要素間関係を認定する。しかし、例えば、以下のような場合にはその妥当性に疑問が残る。

1. 並列や同格の認定では用例内の2つの要素の類似性が問題となる。しかし、現行の TDMT では、意味距離計算は用例全体との意味的近さとして計算されるため、その片方が意味的に近いことで全体の値が小さくなってしまふ。この場合、現行の距離計算法では不適當であろう。
2. 現行の TDMT では、数量表現の選択なども意味距離計算で行っているが、シソーラスにどの程度これらを反映できるかは疑問が残る。

##### [対策案]

○意味距離計算によるルール選択は TDMT 独特の特徴であるため、その関係別により詳細な調査を行って、正当性を確認しておく必要があるだろう。また現在、変数部に現れる要素全てを用例に加えているが、その片方の意味距離だけが効くと思われる場合もあり、この点も考慮して調査を進めるべきであろう。

○数量表現の選択などは意味距離計算を利用するよりも、辞書に記述して生成で対処した方がよいと考える。これらは目的言語によって異なるものであり、シソーラスに反映できないものではないだろうか。

○並列や同格については後述する。

#### 4.6 並列構造の処理

##### [関連する言語現象]

複合述部 (5.1.4)、並列構造を持つ名詞句の処理 (5.2.5)、非活用連体修飾 (5.4.1)、例示・列挙表現 (5.8.12)

[問題の説明] 並列構造の認定では、前述した通り意味距離計算の妥当性に疑問が残る。さらに、構造的な類似性なども考慮に入れる必要があるが、現行の TDMT ではこれらを記述できない。

##### [対策案]

○並列構造はローカルな構造だけを見ても検出しにくいいため、前処理などでよりグローバルな視点から並列性検出する機構 (例えば、黒橋らの並列解析 [8] など) が有効だと考える。

#### 4.7 文脈処理

##### [関連する言語現象]

述部の訳し分け (5.1.1)、名詞の訳し分け (5.2.1)、形式名詞 (5.2.10)、助詞 (5.5)、アスペクトの処理 (5.7.2)、モーダルの処理 (5.7.3)、慣用表現の処理に関して (5.8.1)、比喻表現 (5.8.8)

##### [問題の説明]

冠詞、省略補完、集合名詞の単複などの処理や、一部のモーダルで見られる曖昧性解消では文脈まで考慮しなければ決定できない場合が多いが、現行の TDMT は文脈処理機構を持っていない。

##### [対策案]

○文脈処理機構を持たない TDMT は、デフォルト処理用のルールを用意することで、文内だけは決定できない場合にも出力が可能になっている。ただし、後に文脈処理機構が導入されることを考慮して、デフォルト処理用ルールやそれによる出力が判別できるような工夫をしておく必要はあるだろう。

#### 4.8 ルール作成

##### [関連する言語現象]

述部の訳し分け (5.1.1)、補助動詞 (5.1.7)、用言性連体詞 (5.4.2)、格助詞相当句 (5.4.3)、助詞 (5.5)、接辞 (5.6)

##### [問題の説明]

TDMT は一つのルールでカバーする言語現象が小さく、多くのルールを記述する必要がある。このときルール作成にかかるコストが問題となる。コストを増大させる一つの要因に、同じ言語現象であっても複数の出現形があればその各々に対してパターンルールを記述しなければならないことがある。例えば TDMT

は、特定の品詞列を識別するマーカを利用して高速なパターンマッチを行っているが、パターン内では全てのマーカを固定項として扱っているため、マーカが意味を持たない場合にも、その種類の数だけルールを記述しなければならない。

またコストだけでなく、記述すべきルールを効率よく、網羅的に集めることも考えるべきだろう。現在、ルールの記述作業は学習用の例文に現れる言語現象のみに対して行われている。しかしこの場合、例えば動詞の能動態での利用を想定して訳し分けルールを記述しても、受動態に関しては完全に別のルールであるため、能動態と同じように訳し分けができるとは限らない。

#### [対策案]

○品詞では品詞カテゴリを自由に組み合わせることができ、これを利用することでルール作成のコストダウンを実現している。このような機能の適用範囲を広げることでルール作成コストを削減できると考えられる。マーカに関しても、例えば、動詞類と名詞類の間に入るマーカ全てとマッチするマーカカテゴリをパターン内に記述できれば、用言性連体詞の扱いを一つのルールに集約することができる。またルールの記法として、パターン部を複数記述できるようにして、処理部を一本化するという手法も考えられる。このように処理部分をまとめることによって分散化されている用例もまとまり、ルールの理解性もよくなることが期待できるだろう。

○またルール記述の効率や網羅性をあげるためには、一つの例文からより多くのルールを抽出することが必要であると考えられる。例えば、上記のような能動態ルール記述の際の注意点などを列挙したマニュアル作りなども有用であろう。これらがある程度まとめれば、さらにルール記述ツールとして実現することも可能だろう。

## 5 評価結果の詳細について

本章では文法項目毎の詳細な評価結果を記す。ここでは各々の文法項目について「一般事項」「TDMTでの対応」「考察」に分けて記す。なお実際の評価文は付録1を、その各々に対して記述したルールは宅崎のレポート[9]を参照されたい。

### 5.1 述部

#### 5.1.1 述部の訳し分け

##### [一般事項]

述部を訳し分けるために必要な観点として以下がある。

- (a) 助詞との共起
- (b) 格要素の名詞の意味属性
- (c) 副詞との共起
- (d) 文型
- (e) 文脈情報

##### [TDMTでの対応]

TDMTでは、“(XをY)”のパターンを用いることで、評価文の述部を訳し分けることができた。このパターンは名詞＋助詞＋述部の組合せであり、上記の(a)(b)にあたる。同様にして(c)も記述でき、また長いパターンを使えば(d)を表現することも可能である。

##### [考察]

TDMTでは、原言語1語に対し目的言語1語をデフォルト訳として、変換辞書中に記述している。この訳語を変更するためには、変換ルール内でローカル辞書を用いて訳語の変更を記述しなければならない(図1参照)。しかしローカル辞書を含むルールは、この辞書に記述された単語以外を含むパターンとマッチしても辞書機能が働かないために、間違いを犯す可能性が高い(ローカル辞書問題と呼ぶ)。このため、“(Xを集める)”のようにローカル辞書に記述した単語を固定項を含むパターンを用いて、その単語のための特別ルールとして記述する方が安全であろう。このように訳語変更の対象となる単語をパターン内に含めることにより、このルールは格要素への意味制約が記述された辞書として働くことになり、管理面でも何らかの工夫が必要かも知れない。

また、受動態などでは格の交替が起こるが、TDMTでは完全に別のルールとしてこれらを記述しなければならない(ルール作成の問題)。このため、能動態(受動態)の訳し分けができたからといって、受動態(能動態)も訳し分けられるという保証はない。網羅性を高めるためには、この点も含めて辞書やルールの管理形態や作成環境を考える必要があるだろう。

#### 5.1.2 断定文

##### [一般事項]

断定文(「～が～だ」の表現)の述部を訳し分けるために必要な観点として、以下がある。

- (a) 助詞の種類

```

(define-pattern wo-np-h3 :j-e np
  (?x "を" ?y) (:head 3)(:x group-n)
  =>
  . . . . . 中略 . . . . .
  (((!y !x)(obj {pl+} 2)))
  (
    ("研究者") ("集める")
    ("論文") ("集める")
  )
  (
    (y
      ((本動詞 "集める") (v "gather"))) ← ローカル辞書
    )
  )
  . . . . . 中略 . . . . .
)

```

図 1: ローカル辞書の例

(b) 名詞の種類

(c) 副詞の種類

また、述部(名詞)を動詞や形容詞などに交換しなければならない場合もあり、注意が必要である。

#### [TDMT での対応]

TDMTでは、「AはBだ」に対して“(XはY)”のパターンでYに名詞あるいは形容詞の訳を与えて訳し分け、“(Xだ)”のパターンで全体をまとめる。ここでは名詞+助詞+名詞のパターンを利用しており、上記観点の(a)(b)に対応する。この方法で全ての評価文を正しく処理することができた。またパターンに副詞を含ませることで、(c)にも対応可能だろう。

#### [考察]

訳語を与える際にローカル辞書が必要になる場合があり、このとき、上記5.1.1章で述べたローカル辞書の問題が起こる。ただし、ここでは品詞を交換する目的でローカル辞書を用いている例が多く見られたため、派生語処理を導入することで、より一般性を高めることができるだろう。

また、2つ(以上)に分かれた上記のルールが、文内の単語が用例にない場合にも正常動作するには、パターンで品詞等までチェックする必要があるだろう(ルール間の整合性)。

### 5.1.3 体言述語

#### [一般事項]

名詞句や体言止めのような体言述語を用言として訳すためには、以下の情報を利用する必要がある。

- (a) 用言性の利用
- (b) 助詞の種類
- (c) 格要素の名詞の意味
- (d) 副詞の種類の情報

#### [TDMT での対応]

TDMT の断定文の処理において、断定の助動詞「だ」などは文内にまとめられるだけであり、これらが存在しなくても文として成立することを認めている。このため断定文と同様に、(b)-(d) への対応は可能である。また (a) もパターン内で品詞を指定するなどの方法で記述できる。これにより、サ変動詞語幹、形式名詞、形容動詞語幹のそれぞれが述語となっている評価文を全て翻訳することができた。

#### [考察]

断定文(5.1.2章)と同様に、ローカル辞書問題やルール間の整合性問題に対する注意が必要である。

#### 5.1.4 複合述部

##### [一般事項]

複合述部の処理では、並列用言とするか、格要素・副詞句＋用言とするかを区別する必要がある。

並列用言を認識するために必要な情報としては、一般的に以下がある。

- (1a) 助詞の種類
- (1b) 助詞の種類と名詞の意味属性
- (1c) 用言性単語の認定

また複合述部の要素を格要素として認識するには、以下の方法がある。

- (2a) 複合語要素間の関係を用言と格要素への意味的制約により解析
- (2b) 用言性の部分と他の部分を判断してデフォルト的に格関係を推定

さらに複合述部の要素を副詞句として解釈するには、以下の方法がある。

- (3a) 語の種類により副詞句となりえる要素を複合語より抽出
- (3b) 用言性の部分と副詞句となりえる部分との共起可能性から判断

#### [TDMT での対応]

TDMT では品詞の並びによって挿入するマーカーが異なり、このため品詞列の種類によって適用されるルールも別のものとなる。基本的にはこの機構と用例によって複合述部の種類を認定しており、これは上記の(1a)-(1c)、(2a)、(3b)に相当する。この手法によって評価文全てを翻訳することができた。

#### [考察]

ここは、格助詞などが明示されていなくても語と語の関係が推定できるという TDMT の強みともいえる部分である。ただし意味距離計算の妥当性を検証する意味で、別途、この推定の正解率を評価する必要があるだろう。

### 5.1.5 訳が一用言となる並列用言

#### [一般事項]

日本語では並列用言であるが、訳文では一用言として表現する場合がある。このような用言の翻訳を行うためには、以下の方法がある。

- (a) 並列用言を一用言として辞書に登録
- (b) 二つの用言の種類と格による制約により認定

#### [TDMT での対応]

TDMT では (a) を基本方針としており、辞書登録を行う。また、多言語間翻訳を前提としているため二つの用言として登録される場合もあるが、TDMT はルールによってこれを一つにまとめる合成機能も有している。これにより評価文を正しく処理することができた。

#### [考察]

このようにシステムが知識として持つべき情報は、ある程度網羅的に収集する必要がある。その方法論についても検討すべきだろう。

### 5.1.6 用言の副詞(句)化

#### [一般事項]

一部の用言には、副詞的に訳出すべき場合がある。

#### [TDMT での対応]

評価文として「必要ならば」という成句的な表現があげられており、TDMT では“(X ならば Y)” というパターンに副詞的な訳を記述したローカル辞書を用いた。これにより評価文は処理できた。

#### [考察]

先に示したローカル辞書問題や「Y が副詞化の認定にほとんど寄与しない」ということから、“(必要ならば Y)” という固定項からなるパターンで記述した方が一般性の高いルールとなるだろう。

### 5.1.7 補助動詞

#### [一般事項]

動詞の中には、「みる」や「いく」のように、補助動詞として使われるものがある。本動詞と補助動詞の区別するには、以下の方法がある。

- (a) 漢字か仮名かで判断
- (b) 本動詞と補助動詞の共起情報を利用

#### [TDMT での対応]

TDMT の品詞体系では本動詞と補助動詞を区別しており、各々は本動詞と助動詞に振り分けられる。このため形態素解析で正確に認定できれば、翻訳処理側では通常の助動詞と同じ扱いでこれらを処理することができる。



## [考察]

評価文中で TDMT 品詞体系で補助動詞として認定できないのは「使えなかった」だけであった。その解析結果は「使え(本動詞)なく(助動詞)なっ(本動詞)た(助動詞)」となる。ここで日英翻訳を考えると、否定の助動詞「ない」は本動詞「使える」でなく、補助動詞「なる」に対する否定として解析すべきである。このため「なくなる」を助動詞として合成して、他の補助動詞と同じ扱いにしたほうがよい(ルール作成の問題)。実際、旅行会話システムでは既にこの手法を用いて処理している。

また、例えば「いる」は「ている」(「している」)や「でいる」(「読んでいる」)という2種類の助動詞として解析されるが、その機能は同じである(ルール作成の問題)。そこで、これらを lexical-transformation で正規化することで同じ変換ルールで扱えるようになる。旅行会話システムでは既にこれも取り入れている。

なお、補助動詞の訳出に関してはアスペクトの章(5.7.2章)で述べる。

### 5.1.8 基本動詞の訳し分け

#### [一般事項]

基本述語(する、ある、ない、できる、多い、少ないなど)は、原文中になく訳語として訳さなければならない場合がある。例えば、「温度は一定とする」は、「温度は一定と仮定する」という意味で解釈され、翻訳されることが多い。このように、翻訳の過程で、基本動詞の言い換えを行うには、次の方法がある。

- (a) 省略と考え、解析時に補完
- (b) 言い換え用の訳語を持ち、共起などで訳語を選択

#### [TDMT での対応]

TDMT では特定の格関係ある名詞との共起によって訳語を切り替えており、これは (b) にあたる。ここで言い換え用の訳語はローカル辞書に記述する。

## [考察]

ローカル辞書を使っているため、ローカル辞書問題が生じる。

また、1つの格だけを見て訳語を決定しているため、ルールの競合が起きると入力された文節の順序によって最終的な訳語が変わってしまう危険性がある(ルール間の整合性問題)。これを考えると、途中段階では訳語候補を複数保持し、全体の構造が決まった後に最終決定するような仕組みの必要性を感じる。

さらに、大きな構造変換を伴うルールが存在する場合には特に注意が必要である。例えば「ある」を考えると、「have」などの動詞として訳出するルールと「There be」構文を使うルールが考えられるが、主語の扱いなどの点で適用すべきルールがそれぞれで異なる。これらのルールが互いに干渉しあうと、構文的に不正な解析結果が大量に生まれ、正しい訳文が埋もれてしまう。このようなルールは、訳語の品詞などに関する適用条件を厳しくするなどして、切り分けるべきだろう。同じことは「ない」「多い」「少ない」でも言えるだろう。

## 5.2 名詞

### 5.2.1 名詞の訳し分け

#### [一般事項]

名詞の訳し分けには、以下の情報が使われることが多い。

- (a) 他の名詞との共起
- (b) 意味による制約
- (c) 文字面による制約(規則ベース)
- (d) 用言との共起
- (e) 文脈情報
- (f) 統計情報(ドメイン毎)

#### [TDMT での対応]

TDMT では基本的にパターンを構成する他の要素を含む意味距離によって訳し分けを行う。これは上記の(a) (b) (d)にあたる。評価文がないため、詳細は以下の各章に記す。

#### 5.2.2 複合名詞

##### [一般事項]

複合名詞の処理では、その範囲、ヘッド、係り受けが問題となる。

ここで、複合名詞の範囲を決める方法には以下の方法がある。

- (1a) 文法により処理
- (1b) 頻出する単語列を候補とする

また複合名詞のヘッドを決める方法としては以下があげられる

- (2a) 最後の名詞をヘッドとする
- (2b) ヘッドにならない名詞を登録し、これを省いた最後の名詞をヘッドとする

さらに複合名詞内の係り受けは以下の情報を利用するが多い。

- (3a) 単語間の位置情報
- (3b) 単語間の構造情報(単語の文法的な性質、補助的な品詞情報など)
- (3c) 単語間の意味情報(単語の意味的な性質、2単語間の意味関係など)

#### [TDMT での対応]

TDMT では複合名詞の範囲、ヘッドは用例から計算される変換ルールの優先度によって決められる。これは、上記の中では(1a) (2b)に近い方法と位置付けられるだろう。また係り受けも同様の処理で決定され、これは上記(3c)にあたる。これによって評価文は全て正しく処理できた。

#### [考察]

上記のルールはほとんどの場合、他のルールと競合する。例えば前の名詞が後ろの名詞に係るように処理するルールと、前の名詞を副詞として処理するルールは互いに競合する。今回、評価文中の単語は全て用例として登録したためにこれらのルールが正しく適用されたが、入力文中の単語が用例として登録されていない場合の詳細な動作は不明である。登録された用例が正当かどうかを調べるために、より大規模なオープンテストを行う必要があるかも知れない。

### 5.2.3 「名詞 1 の名詞 2」という構造を持つ名詞句の処理

#### [一般事項]

2つの名詞の修飾関係を認定する方法には以下がある。

- (1a) 「の」によって連結された名詞の表層の共起情報を利用
- (1b) 「の」によって連結された各々の名詞の意味関係に着目
- (1c) 名詞の用言形の格パターンを利用
- (1d) 世界知識を利用

また、認定した修飾関係を英訳に反映させる方法として、以下が上げられる。

- (2a) 名詞連続の複合名詞にする
- (2b) 前置詞 "of" によって名詞を連結する
- (2c) 適切な前置詞を選択して名詞を連結する
- (2d) 一方の名詞を所有格にする
- (2e) 一方の名詞を形容詞化する

#### [TDMT での対応]

TDMTでは(XのY)のパターンで、用例によって(2a)-(2e)を選択している。これは上記の(1b)に当たり、評価文のほとんど全てを正しく翻訳することができた。

#### [考察]

訳出方法としては(2a)-(2e)以外に、原言語の全体構造を変えた方が自然な場合がある。例えば、以下のような評価文、対訳例が見られた。

「彼の行動は素早い」 → 「He acts quickly」

この場合、TDMTでは「((彼, の, 行動), は, 素早い)」の括弧の内側から、その中の情報だけで決定的に訳を与えていく。このため、基本的には日本語の名詞句は対訳中でも名詞句として訳すことになり、「His action is quick」以外の訳を出力することは難しい。これは構造変換に関わる問題であり、全体がまとまった後で構造変換を行うなどの処理が必要だろう。

### 5.2.4 「名詞 1 の名詞 2 の名詞 3」という構造を持つ名詞句の処理

#### [一般事項]

3つの名詞の修飾関係を認定する方法として以下があげられる。

- (1a) 3項関係(3つの名詞間の意味関係)を利用
- (1b) 2項関係の合成として処理
- (1c) 名詞の用言形の格パターンを利用
- (1d) 世界知識を利用

また、認定した名詞の修飾関係を生成する英文構造に反映するには、以下の手法がある。

- (2a) 名詞連続の複合名詞にする

- (2b) 前置詞 "of" によって名詞を連結する
- (2c) 適切な前置詞を選択して名詞を連結する
- (2d) 所有格の名詞を使用する
- (2e) 名詞を形容詞化した表現を使用する

#### [TDMT での対応]

TDMT では、3つの名詞の修飾関係も2つの名詞の修飾関係(5.2.3章参照)の繰り返しによって認定しており、これは(1b)にあたる。ただし3変数のパターンを記述することによって(1a)も記述できる。評価文は全て正しく処理できた。

#### [考察]

前の5.2.3章で述べた問題と同じ問題がある。

また、2つの名詞の修飾関係の処理を組み合わせて処理しているため、これらのルールがどのような順番で係るかが名詞句内の全体の係り受けを決める。これは、各々のルールにどのような用例が入っているかに依存することになるので、注意が必要だろう(ルール間の整合性問題)。

#### 5.2.5 並列構造を持つ名詞句の処理

##### [一般事項]

並列構造の認定方法のヴァリエーションとして、以下の処理がある。

- (a) 並列構成要素を持つ意味属性の類似度に基づく処理
- (b) 格による制約条件に基づく処理
- (c) 表層文字列の一致に基づく処理
- (d) 読点の有無に基づく処理
- (e) 構造上のバランス評価に基づく処理

#### [TDMT での対応]

TDMT は、ほとんどの並列構造を"(X と Y)" というパターンを利用して認定しており、これは(a)のみに対応していることになる。評価文のほとんどは正しく翻訳できた。

#### [考察]

評価文が処理できたのは評価文中の単語を全て用例に登録しているためである。しかし、並列になる組合せ全てを予め登録することは不可能であり、ここで用いたルールに一般性はない。話し言葉を対象とすれば(d)に対応する必要はないが、(b)(c)(e)には対応する必要があるだろう。

並列認定では並列となる2つの要素の関係(類似性)が問題となる。例えば「神戸と大阪」の場合、「神戸」と「大阪」がともに地名であることが重要なのであって、用例との意味距離の総合値として計算する現行の評価法は不適當であるように思える。並列認定での意味距離計算の妥当性については、さらに調査が必要だろう。

また現行の TDMT では、一度まとまった構造の下位構造の参照ができないため、構造的な類似性をチェックすることはできない。

さらに同じ理由で、並列構造として1つの構造にまとまった句を、後から他の構造と区別することができないため、並列構造に係る先の要素を複数として処理することもできない。

## 5.2.6 疑問表現の名詞節の処理

### [一般事項]

疑問表現の名詞句には5W1Hの種類があるが、名詞句に対応する場合と、副詞句に対応する場合とがある。さらにその訳し方にはwh節に訳される場合と、不定詞句に訳した方が自然な場合とがある。

### [TDMTでの対応]

TDMTでは”パターン(XがY)+用例(開発者, 誰)”のようなパターンを用いて処理することで、評価文を翻訳することができた。

### [考察]

上記パターンの部分だけでは埋め込み文として処理するのか、疑問文として処理するのかが用例では決まらないため、両者のルールは衝突することになる。そのどちらを選ぶかは後で決まることであるが、下位構造の参照ができないため、選択すべきときには既に両者の区別がつかなくなってしまう。構造全体に属性を付加するか、下位構造の参照を可能にするかして、この区別を明確にする必要があるだろう。

また、特殊な扱いが必要になる「誰」や「どこ」などの疑問詞を用例に記述すると、他の普通名詞などにもこのルールに係る可能性があるため、これらをパターン内の固定項に含めるなどしてルールを完全に切り分けた方が安全だろう。

## 5.2.7 用言性名詞(サ変名詞)

### [一般事項]

用言性名詞は、英訳にする際には動詞句や不定詞句にした方が自然であることも多い。その訳し方には以下の方法がある。

- (a) 名詞として翻訳
- (b) 用言に変換して構造化し、翻訳
- (c) 条件によって、名詞か用言かを切り分ける
- (d) 条件によって、命令形にする(サ変名詞+「のこと」など)

### [TDMTでの対応]

TDMTでは、

- 1) ”(XをY)”(「調査を行う」など)
- 2) ”(X<名詞連続マーカ(名詞の種類により数種類)>Y)”(「市場開発」など)
- 3) ”(Xのこと)”

などのパターンを用いて、上記の(a)-(d)の全てに対応している。これによって評価文の全てを正しく処理することができた。

### [考察]

上記のルールは全て、用例からの距離計算によって適用の優先度が決まる。しかし、上記1)と3)ではその名詞の用言性が大きく関与していると考えられ、品詞の指定によってルールの適用対象を切り分ける方が適当であろう。

また、2) は普通名詞、サ変名詞、形容名詞などの並びによって訳し分けているが、品詞の並びで訳出パターンが決定できるか、決定できるならどのように訳出すればよいかといったことは、複合名詞(5.2.2章)で述べたような調査で明らかにする必要があるだろう。

#### 5.2.8 英語における数の扱い

##### [一般事項]

日本語を英訳する場合、名詞に対する単数複数を明確に表出させる必要があり、例えば以下の現象に対応しなければならない。

- (a) 集合名詞
- (b) 連体修飾による単数・複数の変化
- (c) 可算名詞
- (d) 不可算名詞
- (e) 絶対複数
- (f) 合成名詞の複数形

##### [TDMT での対応]

TDMT では、各名詞の複数形は生成辞書に記述されている。変換辞書に直接複数形が記述されている絶対複数の名詞のような場合もあるが、上記分類にあたる情報は無い。このため TDMT では単複の処理は全て(c)として扱っており、他の特殊な処理には対応していない。

##### [考察]

上記の種別などの情報がなくては処理できないので、生成辞書をより充実させる必要があるだろう。また実際に数を決めるには文脈処理が必要な場合も多く、これは将来の課題である。

#### 5.2.9 固有名詞表現

##### [一般事項]

固有名詞の全てを辞書に登録できないため、未知語処理などから固有名詞の認定が必要になる。その種類には以下のようなものがあり、各々に特殊な表現形式が存在する。

- (a) 地名・地形名
- (b) 人名
- (c) 役職名
- (d) 組織・企業名
- (e) キャプション番号など

##### [TDMT での対応]

現行の TDMT は全ての単語が登録されていることを前提としており、未知語処理は存在しない。このため、辞書登録されていない固有名詞は処理できない。

##### [考察]

TDMT 品詞に固有名詞という品詞は存在せず、普通名詞として扱われる。このため、変換処理は複合名詞と全く同じになる。しかし住所表現などのための特殊なルールが複合名詞の処理に影響を及ぼす可能性が大きいので、何らかの対策が必要であると考えられる。

#### 5.2.10 形式名詞

##### [一般事項]

実質的・具体的な内容を持たない「こと、ため、とき」などの形式名詞は、そのまま訳出すると不自然な訳文となるので、特殊な処理が必要である。具体的には以下のような方法がある。

- (1a) 単独で使われる普通名詞と全く同様に処理
- (1b) 連体修飾を伴った普通名詞の訳し分けと同様に処理
- (1c) 定型的に処理 (常に that 節にするなど)
- (1d) 個々の形式名詞や用法によって生成を制御

ここで特に、(1d) を制御するための情報として以下があげられる。

- (1d-1) 形式名詞と形式名詞句に修飾される名詞との共起
- (1d-2) 形式名詞と形式名詞を修飾する名詞 (句) との共起
- (1d-3) 形式名詞と形式名詞句に修飾される用言との共起
- (1d-4) 形式名詞と形式名詞を修飾する用言 (動詞句) との共起
- (1d-5) 常識・世界知識
- (1d-6) 文脈 (状況) 情報

また、形式名詞と普通名詞とを区別するための情報としては

- (2a) 語ごとの固有の情報 (辞書などにあらかじめ登録する)
- (2b) 接続している語との共起
- (2c) 係り受け関係

などがある。

##### [TDMT での対応]

ここでは「ため」「こと」「そう・そうな」「点」「ところ」「とき」「の」「もの」「わけ」「はず」が代表的な形式名詞として取り上げられているが、TDMTではこのうち「こと」「ところ」「とき」「もの」「わけ」「はず」だけが形式名詞として定義されており、この他のものは普通名詞として扱われる。しかし共に、これを固定項に含む連体句とのパターン (“(X </kihon-fn> こと)” など) で処理することで、評価文を翻訳することができた。これは上記の (1d) (1d-1)-(1d-4) (2a)-(2c) にあたる。

##### [考察]

上記にもあるように形式名詞は特殊な処理が必要となるため、いわゆる普通名詞の処理と完全に区別できるように、固定項を含むパターンを利用する方がよいだろう。

また大きな問題として、パターン内変数に一致した構造の下位構造が見えないことがある (下位構造の参照)。例えば、「～するため」(目的) と「～したため」(原因) の訳し分けでは、先行する動詞句の時制が訳

し分けのキーとなりうる。しかし、「～＋た」(過去)として動詞句がまとまった後にかかるルールからはXがどのような構造なのかを知ることができず、”(X </kihon-fn> ため)”のルールから過去の情報を使うことができない。これは用例だけでは切り分けられない現象であり、なんらかの方法で下位構造を伝達する仕組みが必要であることを示している。

### 5.2.11 関係を示す名詞

#### [一般事項]

関係を示す名詞は普通名詞として扱うべき場合と形式名詞として扱うべき場合がある。例えば、「上」「下」「前」「後」などがあげられる。

#### [TDMT での対応]

TDMT では 5.2.10 章で示したように、形式名詞として扱う場合にはそれを固定項に含むパターンを利用する。

#### [考察]

文中の単語を用例として与えたため、この手法で評価文の全てを処理できたが、形式名詞の処理(5.2.10章)で述べた問題と同じ問題が起こる。

## 5.3 副詞

ここでは便宜上、格助詞句以外の連用修飾を副詞として扱う。このため、一単語としての副詞の他に、用言の連用形(暖かく、など)や成句、連語要素など、英語での副詞相当の語句や副詞句が含まれる。

### 5.3.1 副詞のタイプ

#### [一般事項]

副詞はそれが修飾する品詞によって分類することができる。ここでは特に、副詞に係り得る品詞として動詞、形容詞、形容動詞、副詞、名詞、数詞、連体詞があげられている。また副詞を訳す際の注意点として、

- (a) 用言との共起を用いた訳し分け
- (b) 訳出位置
- (c) 副詞句以外の修飾要素や格としての認定(例えば「全て」)
- (d) 副詞による時制 / アスペクトの変化

があげられる。

#### [TDMT での対応]

TDMT では、副詞の処理は係り先の単語を用例に記述したパターンに記述されている。また TDMT で採用されている品詞体系では、複数の品詞に対応する単語は副詞以外に割り振る傾向にあるため、これらを副詞として処理するルールがある。これらも通常の副詞と同様に係り先の単語によって制御されている。上記の修飾する品詞による切り分けは行っていないが、評価文はほぼ正しく処理することができた。

#### [考察]



上記 (a) に対する評価文の処理ではローカル辞書を用いているため、副詞部分を固定項に含むパターンで記述した方がよいだろう (ローカル辞書問題)。また、係り先の品詞まで指定した方がより一般的なルールになると思われる。

(b) については文先頭、動詞修飾、形容詞修飾、文最後の指定が可能で、評価文はこれらの指定で十分制御できた。

(c) に関しては、TDMT 品詞体系の特性から通常の処理と変わらない。

(d) については評価文が正しく処理できたものの、以下の問題があると思われる。

その一つは時制をコントロールする属性の不足である。TDMT ルールでは動詞 (句) に対して {past+} や {pp+} という動詞の変化形を指定する属性を振ることで時制をコントロールする。しかし、現在形を出力する属性がないため、助動詞の処理で変化した動詞を副詞の処理で現在形に戻すといった操作ができない。

もう一つは、変換ルールでまとまった後でその下位構造を知ることができないことによる (下位構造の参照)。例えば「今～する」と「今～した」などの例を考えると、副詞による時制、アスペクトの変化は副詞だけでなく、助動詞との対で決まる場合があることが分かる。現行の TDMT では、これら全体を含む長いパターンを使う以外にこれを表現することができず、間違った解析結果を除くことができない。上記の例では“(今 X)”と“(今 X た)”のパターンを使うことになるが、「今～した」はパターン“(今 X)”と過去のパターン“(X た)”によっても解析されてしまう(“(今～する) た)”)など。これを避けるためには、下位構造をチェックが必須となる。

### 5.3.2 副詞句

#### [一般事項]

ここでは副詞句の種類として、動詞連用形、形容詞連用形、成句、連語要素があげられている。各々に対する対策は以下の通りである。

#### 1. 動詞連用形 (「改めて」など)

- (1-a) 副詞の辞書項目を作成
- (1-b) 動詞の辞書項目に記述し、格パターンのチェックにより区別
- (1-c) 動詞の辞書項目に記述し、意味のチェックにより区別

#### 2. 形容詞連用形 (「正しく」など)

- (2-a) 副詞の辞書項目を作成
- (2-b) 形容詞の辞書項目に記述

#### 3. 成句 (「簡単に言うと」など)

- (3-a) 副詞句イディオムを作成
- (3-b) 中心となる単語に記述

#### 4. 連語要素 (「～により」など)

- (4-a) 全体を辞書登録
- (4-b) 「～で行く」のようなパターンを登録
- (4-c) 文法で対処
- (4-d) 副助詞句イディオムを作成
- (4-e) 中心となる単語に記述

#### [TDMT での対応]

今回、TDMT では上記 (1)-(4) の全てを他の副詞と同様に、副詞句となる部分を固定項に含んだパターンによって記述した。これは上記では、(1-a) (2-a) (3-a) (4-d) に近い方法だろう。

#### [考察]

変換ルールに記述することによって、副詞として訳すかどうかの曖昧性を扱うことも可能である。ただし、副詞かどうかの判断の基準となる格チェックや変数部の下位構造のチェックなどが記述できない(下位構造の参照の問題)ため、どこまで正しく記述できるかは不明である。

連語要素の品詞変換については、副詞の処理が動詞句処理内で行われることを考えると、前置詞句、接続詞句に変換できることが分かる。しかし、主語などの格要素に変換するには、動詞に係る他の要素のシフトなどを考慮する必要があるため、副詞句と動詞部分だけの操作では記述できない。格要素への変換を考えるならば、少なくとも変数部の下位構造が操作できる必要があるだろう。

### 5.3.3 擬音語・擬態語

#### [一般事項]

日本語の擬音語・擬態語は、英語に訳出できない場合がある。このような場合、動詞がその擬音語・擬態語の意味を含んでいることが多い。

#### [TDMT での対応]

今回、TDMT では合成処理により「ガーガー鳴く」を一語の動詞として処理することで、評価文を処理することができた。

#### [考察]

「ガーガー」と「鳴く」が文中で離れ得ることを考えると、「ガーガー」を副詞として動詞句を作る際にローカル辞書で処理すべきと考える。

### 5.4 連体修飾語句

#### 5.4.1 非活用連体修飾

#### [一般事項]

ここでは連体詞のように活用なしで体言を修飾する表現を非活用連体修飾と呼んでいる。これには単独型と「の」などで連結する連体修飾があり、数量表現のように、訳し分けのために被修飾要素である体言の情報等を利用しなければならない場合もある。また訳出の仕方には以下のようなヴァリエーションがあり、適切な方法を選ぶ必要がある。

- (a) 同格の名詞
- (b) 形容詞
- (c) 指示形容詞
- (d) 副詞
- (e) 前置詞句
- (f) スタイルを制御(ダブルクォートやフォントなどで表現)

さらに修飾する語句が、被修飾語句が単数なのか複数なのかを区別するキーになることがある。数詞や数量形容詞などがこれにあたり、また連体修飾句の並列性から判断できることもある。

#### [TDMT での対応]

TDMT では、単独型は副詞と同様にこれを固定項を含むパターンで、「～の」型はパターン”(X の Y)”で記述している。これらは被修飾要素である体言を含むパターンであり、その用例を用いて訳し分けている。また、これらは名詞句の処理として行われるため、変換後が名詞句の形をしていれば対応可能であり、上記の (a)-(f) の全てに対応することができる。また、数詞や数量形容詞のルールで被修飾語句の数量を記述することも可能である。

#### [考察]

訳し分けの判断には意味距離計算を用いているが、ここでも意味距離計算の妥当性に問題があると思われる。例えば、数量表現がどの程度訳し分けられるかはシソーラスに依存し、不明である。また同格の判断でも修飾語句と被修飾語句の全体と用例との類似度を計算する意味距離計算を利用しているが、この場合には修飾語句と被修飾語句の間の類似性が重要であり、正当性については疑問が残る。

5.2.5章で述べたように並列性に関してはその認定は難しい。また認定できても連体修飾を処理するルールからは修飾語句が並列かどうかを調べることができない(下位構造の参照)。このため、並列性を利用した被修飾語句の数量認定は記述できない。

#### 5.4.2 用言性連体詞

##### [一般事項]

ここでは、形容詞のように連体修飾表現の主要部が活用可能なものを用言性連体詞と呼んでいる。用言性連体詞を認定する最も簡単な方法は辞書に一単語として登録することであるが、この他にも用言として解析した後、条件によって連体詞と認定したり、変換時に、訳出の必要性により連体詞と認定する方法がある。その種類としては

- (a) 形容詞変化形単独
- (b) 副詞 + 形容詞変化形
- (c) 動詞変化形単独
- (d) 副詞 + 動詞変化形
- (e) 格要素 + 動詞変化形

があり、また訳出可能なスタイルとしては、形容詞、動名詞(現在分詞)、過去分詞、To不定詞、前置詞句、関係節、前置詞句などがある。

#### [TDMT での対応]

TDMT では連体要素をまとめた後に、「パターン”(X Y)”で連体修飾部と被連体修飾部を組み合わせる名詞句を作る。これにより (a)-(e) の評価文のほとんどを正しく翻訳することができた。

#### [考察]

変化形ではなく、派生語を用いて訳出する場合にはローカル辞書を用いることになり、ローカル辞書問題が発生する。

(e)などで関係詞を用いる場合には問題がある。一般的なルールとするためには、関係詞節内をまとめる処理と先行詞と関係詞をまとめる処理を別々に記述した方がよい。しかし、TDMTではその2つの処理の間で情報交換(下位構造の参照)ができないため、例えば主文の処理で先行詞となる名詞が複数だと分かっていても、主格関係節内にこれを反映することができないなどの問題が生じる。

#### 5.4.3 格助詞相当句

##### [一般事項]

ここでは、「に関する」や「を伴う」のように、「格助詞+動詞連体形」全体で格助詞として機能すると見なすことができる表現を、格助詞相当句と呼んでいる。格助詞相当句を認定する方法としては、連語として辞書登録しておき、ひとかたまりとして認定するのが最も単純であるが、個々の単語を構成的に処理すべき場合もあるので注意が必要である。また、訳語の生成に関しては、連語に対する訳語を予め一意に設定しないで、通常の訳し分けと同様、訳し分けを行なう場合もある。このような訳し分けをするには、連語内の述語をキーとした格解析などを行なう必要がある。

##### [TDMTでの対応]

TDMTでは「に関する」などのように一語登録されているものと、「に対応する」などのように二語以上に分かれているものがある。これらは助詞の処理と同様に、その語句を固定項に含むパターン(例えば“(Xに関する Y)”など)を用いて、処理している。これによって評価文全てを正しく処理することができた。

##### [考察]

二語以上に分かれているものは動詞を含んでいるため、</kihon-cn>などのマークが解析対象文に挿入され、ルール中のパターンにもこれを含めなければならない。ここで、例えば</kihon-cn>をパターンに含めた場合、後続する語句は普通名詞でなければならず、サ変名詞などの場合のルールは別途用意する必要がある。ルール記述のコストを考えると、マークにもカテゴリを用意するなどして、同じ機能のルールはまとめて記述した方がよいだろう。

#### 5.4.4 埋め込み文修飾

##### [一般事項]

ここでは、体言を修飾する表現が埋め込み文である連体修飾を埋め込み文修飾と呼んでいる。埋め込み文修飾を適切に翻訳するには、埋め込み文と被修飾句との修飾関係を正しく認定する必要がある。修飾関係としては、主格修飾関係、目的格修飾関係、所有格修飾関係、場所修飾関係、時間修飾関係、原因・理由・方法の修飾関係、相対名詞修飾関係、内容修飾(同格や引用)関係、内容修飾(感覚)関係などがあり、各々を認定するために以下のような解析が必要となる。

- (1a) 埋め込み文中の述語の格パターンを利用
- (1b) 名詞の意味属性や意味ネットワークを利用
- (1c) 先行詞になりうる名詞の制約条件を利用
- (1d) 埋め込み文と先行詞をつなぐ特定の語に着目

また、埋め込み文修飾の処理では以下の点に注意する必要がある。

- (2a) 関係代名詞や関係副詞の用法による使い分け

- (2b) “that” 節、不定詞節、前置詞句を含んだ名詞句などの使い分け
- (2c) 先行詞が構造化された名詞句
- (2d) 被修飾語である体言が他の語に吸収される場合
- (2e) 形容詞が複数ある場合の語順

#### [TDMT での対応]

TDMT では、関係節内をまとめた後に、関係節句、マーカあるいは特定の語句、先行詞からなるパターンを利用することで、埋め込み文を処理する。これらのルールは用例と意味距離計算で制御しており、上記の(1b)(1c)に対応する。またパターン内に特定の語句が記述できるため、上記の(1d)を記述することも可能である。これらによって評価文を全て処理することができた。

#### [考察]

上記(1a)にあたる格パターンのチェック(下位構造の参照)ができないため、修飾関係に曖昧性が生じた場合にこれを十分解消できるかどうかは疑わしい。

また、感覚を表す名詞が先行詞となるとき、この名詞を動詞化して訳出する方が自然な場合があるが、TDMT に全体構造を見ながら修正する機構がないためこれはルールに記述できない(構造変換の問題)。

次に上記一般事項の注意点を考えると、まず(2a)については現行の TDMT でも記述できる。しかし、これは関係節内の動詞に対する訳語が決まった後に判定可能なものである。このため辞書中の訳語の変更によってルールに対しても修正が必要となるため、ルールライターへの負担が大きいと考えられる。格フレームを導入する、生成あるいは他のモジュールで対応するといった対処を考えた方がよいかもしれない。

また、(2d)については構造変換を伴うため記述できないが、関係詞を用いて訳出してもさほど大きな問題は無いと考えられる。

さらに(2e)については、TDMT では日本語と同じ順序でしか訳出できない。しかし英語では形容詞の順番に制約のある場合があり、生成で対処するなどの対応が必要だと考えられる。

### 5.5 助詞

#### [一般事項]

ここでは格助詞と副助詞を扱う。格助詞については深層格の認定が重要であり、副助詞については副詞や形容詞などを用いた表現への変換が重要になる。これらを訳し分けるには、以下の情報を利用する必要がある。

- (a) 動詞との共起情報
- (b) 格となる名詞の意味属性
- (c) 副詞との共起情報
- (d) 動詞のモーダルヤテンス・アスペクト情報
- (e) 文脈情報

#### [TDMT での対応]

深層格の認定については、TDMT では助詞の前後を変数に、助詞を固定項に持つパターン(例えば“(Xは Y)”)を用いている。これらのルールは用例で制御されており、上記の(a)(b)の情報を用いていることになる。(c)も同様に記述可能であり、また動詞部をまとめる際に2ヘッドの構造を持たせていれば、(d)

も記述できる。また副助詞に関しては、格となる名詞に何らかの修飾詞を付加することで表現可能な場合にはその名詞句だけをパターンとして記述している。これにより評価文のほとんどは処理することができた。

#### [考察]

パターン記述の整理が必要であると感じる例が幾つか見られた。例えば、副助詞で表現され格助詞が消失する場合、格となる名詞と副助詞を先に処理した後、名詞と動詞からなるパターン“(X Y)”によって処理している。現在、ここには実際に評価文で出て来た表現だけが用例として付加されているが、同じ関係にある用例は格助詞を含むパターンにも記述されており、これらの用例が共有できれば、より効率良く用例を増やすことができる。ただし名詞と動詞の対だけでどの程度正しい深層格認定ができるのかは調査しておく必要があるだろう。また逆に、同じ機能を持つ助詞が表記が異なるために別れているルールもあり、助詞のカテゴリを作るなどしてルールをまとめた方がルールの作成効率がよくなるだろう。ただしこのカテゴリは深層格そのものにあたり、カテゴリ設定の問題があるかも知れない(ルール作成の問題)。

また、上記の機構だけでは曖昧性が解消しきれない場合のあることが分かっている。例えば「彼は私も呼んだ」「私も彼を呼んだ」では「私」と「呼ぶ」の用例ではその格関係を知ることができない。これを解消するには既に埋まっている格のチェックが必要となるが、TDMTには下位構造の参照機能がないため、これを記述することができない。

さらに呼応の表現などは、その全ての部分を含む長いパターンで記述せざるを得ず、一般性の高いルールを記述することができない。一般性を向上するためには、複数のルールに分けて記述できることが望ましいが、現行のTDMTはルール間で情報交換(下位構造の参照)ができないため、これも不可能である。

## 5.6 接辞

### [一般事項]

接頭辞、接尾辞、接中辞などの処理方法には、以下のような方法がある。

- (1a) 接辞を含んだ長単位で辞書登録
- (1b) 各接辞に対する翻訳パターンを登録
- (1c) 接辞として訳出し、主名詞の訳に繋げる
- (1d) 主名詞との共起関係により、品詞を変えたり訳を選択する

また訳し分けを行うためには、一般的に以下の情報が使われる。

- (2a) 主名詞の意味属性
- (2b) 主名詞の文字面
- (2c) 接辞と接辞以外への意味的制約
- (2d) 接辞と接辞以外への品詞による制約

### [TDMTでの対応]

TDMTでは、ほとんどの接辞は上記の(1a)で処理している。ただし、一部の接辞(「的」「化」など)は合成処理で長単位を生成しているため、(2b)と(2d)を用いて(1b)を行っていることになる。

### [考察]

上記のように接辞の問題はほとんど長単位の辞書登録で対処しているため、辞書がいかに充実しているかが品質に効いてくる。耐性を上げるためには、語彙の網羅的な収集が求められるだろう。

また合成ルールにおいては品詞による制約を用いているが、同じ処理でも普通名詞とサ変名詞でルールを分けている場合がある(ルール作成の問題)。ここでも品詞カテゴリを用いた方がルールの記述効率を上げることができるだろう。

## 5.7 テンス、アスペクト、モーダル

### 5.7.1 テンスの処理

#### [一般事項]

ここではテンスの解析法として、以下の方法が上げられている。

- (a) 述語を分類せず、その変化形によって一意に決定
- (b) 述語を分類し、その変化形によって一意に決定
- (c) 述語を分類し、その変化形によって生じる多義性を、述語と共起している副詞や名詞の情報で解消

また注意すべき事項として、時制の一致や仮定法過去があげられている。

#### [TDMTでの対応]

TDMTでは、述部に対する用例とそれに付加する助動詞の種類によって、時制を決めている。また仮定法や仮定法過去については、“(もし X ば Y)” や“(もし X ば Y のだが)” といったパターンを利用して記述している。

#### [考察]

述部の処理で決まった時制は他のルールから参照できない(下位構造の参照ができない)ため、主節と従属節の間で時制の調整が正確にできず、時制の一致が記述できない。この問題に関しては、直接話法を用いるという手もあるが、話の途中段階で訳すことの多い同時通訳者が、時制の一致をどう処理しているのかを調査した方がいいとの意見が出た。

また仮定法や仮定法過去に関するルールは、長いパターンを利用しているため、一般性は低い。しかし、従属節の真偽が問題となるため、現段階では仕方がないと考えられる。

### 5.7.2 アスペクトの処理

#### [一般事項]

アスペクトの解析方法としては、アスペクト素性を導入し、素性間の演算処理を行ったり、述語と補助動詞の可能な組合せに対して意味解釈規則を用意し、適用する方法があげられている。さらにアスペクト値の多義性を解消するには、述語と共起している副詞や名詞の情報や、さらに文脈情報まで参照する必要がある。

#### [TDMTでの対応]

TDMTではテンスの処理と同様に、述部の用例と助動詞の種類によってアスペクトを処理している。これによって評価文は全て処理することができた。

#### [考察]

ここでは用例として与えられている述部のヘッドが訳し分けのキーとなる。しかし、意味距離でこれらがうまく分類できるかどうかはシソーラスによるため、アスペクト処理のデフォルトは意味分類とは別に考えた方がよいだろう。現行の TDMT では、全ての用言をアスペクト処理ルールの用例に振り分けることで記述は可能である。しかし、これは表現毎に記述しなければならず、ルール作成のコストが高くなるため、辞書で単語に対する属性として振るなどして、これを利用する方がよいだろう。

また訳出に関しては、動詞(句)に対して {past+} や {pp+} で変化形を指定する属性を振るか、「have」「will」などの助動詞を出力パターンに付加することでコントロールしている。しかし、これらが別々のルールで与えられると、お互いの構造情報が参照できない(下位構造の参照ができない)ため、整合性が取れなくなる場合があると考えられる。これらの情報は単語単位よりも、用言単位で一元管理した方がよいように思われる。

### 5.7.3 モーダルの処理

#### [一般事項]

モーダルの認定は、文末表現による判断に加えて文脈情報を参照して判断する必要がある。

#### [TDMT での対応]

TDMT では述部の用例と文末表現によってモーダルを判断している。評価文のほとんどは正しく処理することができた。

#### [考察]

述部と文末表現だけでは曖昧性の解消できない場合が多々ある。例えば助動詞「う」には一人称命令と意志の曖昧性がある。この場合には主語の有無や種類などが曖昧性解消に役立つと思われるが、ルール間の情報交換(下位構造の参照)ができないために、現行の TDMT ではこれらのルールを記述することができない。

また上記の曖昧性には、一般的に文脈処理を行わなければ解消できず、一文内では決定することはできない場合もある。現在の TDMT に文脈処理機構はないが、将来の導入を考慮して決定的な通常のルールとデフォルト処理用ルールが判別できるような工夫をしておく必要はあるだろう。

### 5.7.4 ボイスの処理

#### [一般事項]

受身、使役、受益といったボイスの認定には、文末表現による判断に加えて文脈情報を参照して判断する必要がある。

#### [TDMT での対応]

TDMT では述部の用例と文末表現によってボイスを認定する。これらのルールでは特に格のシフトを考慮し、述部と助動詞の両方をヘッドとする 2 ヘッド構造を用いている。これにより、評価文はうまく処理できた。

#### [考察]



使役の助動詞「せる」の訳出パターンも、TDMTでは用例で訳し分けている。しかし、この訳出パターンは動詞部の英訳が自動詞か他動詞かに依存しており、用例による意味距離ではうまく判別できないと考えられる。このため、何らかの形でこれら訳語の情報も使う必要があると考えられる。

## 5.8 特殊構造表現

### 5.8.1 慣用表現の処理に関して

#### [一般事項]

慣用表現は、文の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り、字義通りに翻訳すると意味の通らない文になってしまう場合がある。このため、慣用表現に対しては特別な処理が必要となる。また、字義通りに訳す方が正しい場合もあるため、一般的な表現なのか、慣用表現なのかを認定する必要がある。その認定の方法には、以下の方法があげられている。

- (a) 語と語の共起を使用
- (b) 格要素により限定
- (c) 文脈情報を使用

#### [TDMTでの対応]

現行のTDMTでは通常の助詞ルールに用例として加え、ローカル辞書による訳語変更で対処している。

#### [考察]

慣用表現では全体が一致して慣用表現であると認定できるので、その要素全てを固定項に含んだパターンで記述した方がよいだろう。例えば、現行では「油を売る」は“(XをY)”のパターンに(油、売る)の用例を加えることで処理しているが、慣用表現でない「油を販売する」でもこのルールが適用され、ローカル辞書が働いてしまう。これを避けるために、合成ルールによって処理する方法も考えられるが、字義通りの解釈ができなくなってしまうため、“(油を売る)”という固定項からなるパターンのルールで慣用表現として処理する方がよいだろう。

また、字義通りに解釈するための格チェックなどができないため、その訳し分け条件が十分に記述できない。ここでも下位構造の参照機能が必要であろう。

### 5.8.2 四字熟語

#### [一般事項]

一般的に四字熟語と呼ばれる語は単純に訳出することができず、その意味を表す英語に翻訳するか、原語のまま(またはローマ字等で)表現する。

#### [TDMTでの対応]

TDMTでは未知語がないことを前提としているため、四字熟語の単位が形態素となり、訳語も振っている。これによって評価文は正しく処理することができた。

#### [考察]

未知語処理のないTDMTでは、このような表現は網羅的に収集する必要があるだろう。ただし旅行会話にはほとんど出現しないと考えられる。

### 5.8.3 呼応表現

#### [一般事項]

呼応表現とは先行するある語に対して、後に特定の語・語形が来るものである。イディオムと同様に2つ以上の単語が共同で1つの意味を表すなど要素合成原理を満たさないことが多いため、特別の処理が必要となる。イディオムと若干異なる点は、キーとなる単語や言い回しが文中でかなり離れて出現することが多い点であり、その分処理が困難となる。

#### [TDMTでの対応]

TDMTでは呼応表現を全て含む長いパターンを利用して処理している。ここで、特に呼応表現にあたる部分はパターン中では固定項として記述している。これによって評価文を処理することはできる。

#### [考察]

多くの固定項を含む長いパターンを記述すると、ルールの一般性が低くなるという問題点がある。しかし、各々を別々のルールとして記述できるものや、片方を副詞として訳出するなどすることでルールが分割できる例も見られ、なるべくパターンを短くするように工夫する余地がみられる。下位構造の参照ができれば、ルール間の整合性を保ったまま、より多くの呼応の表現がより短いより一般的な形でルール化できると思われる。

また、例えば一文中に「also」が2つ以上出現する場合に1つだけを表出させるなど、生成モジュールが対応すべき問題も見られた。

### 5.8.4 天候・気象表現

#### [一般事項]

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よく用いられるため、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場合もある。

#### [TDMTでの対応]

TDMTでは、「it」を補うルールも「we」を補うルールも記述できるが、その訳し分け基準が不明であり、訳し分けまで正当に記述できるかは分からなかった。

#### [考察]

小規模なネイティブチェックを行ったところ、「we」を補うべきとした評価文を「it」で訳出してもほとんど不具合がないとの結果が得られた。このため、訳し分けを考えずに、全て「it」で補っても問題はないと思われる。

ただし、「it」を補うべき非人称動詞の認定を用例からの意味距離で決めて良いかどうかには疑問が残る。

### 5.8.5 無生物主語構文

#### [一般事項]

主格以外の要素が訳文の主語になる場合がある。

#### [TDMTでの対応]

“(XによりY)”や“(XでY)”のルールでYに「分かる」という用例を与えることで、評価文のほとんどは一応処理することができた。

#### [考察]

TDMTでは入力文の部分的な情報から次々と翻訳しているため、入力文の構造が変更できない。このため一般的なルールで、無生物主語構文を出力することは難しい。例えば、上記のルール記述で、Xを主語にするルールを記述することはできる。しかしこのとき、「分かる」にはローカル辞書を用いて字義と全く異なった訳を与える必要があり、ローカル辞書問題が生じる。ここではパターン中の固定項には構造を持つものがマッチしないため、「分かる」を固定項に含めることもできない。

#### 5.8.6 「はが」構文

##### [一般事項]

「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」が非常に多くの意味で用いられるため、処理が難しくなる。一般的に以下のような場合がある。

- (a) 主格・目的格の「は」「が」
- (b) 所有関係・部分全体関係の「は」「が」
- (c) 副詞句としての「は」
- (d) その他(「私の家は学校が近い」など)

#### [TDMTでの対応]

TDMTでは、名詞と動詞の組からその間の深層格や副詞関係を推定することができるため、(a)(c)については一般ルールと同じルールで記述することができる。また(b)については慣用句的に行うべき処理が含まれており(「耳が遠い」「背が高い」など)、これについては慣用句の処理(5.8.1章)と同じ手法を用いている。(d)についても記述でき、評価文をほぼ全て正しく処理することができた。

#### [考察]

(a)(c)については「は」と「が」の各々を別々に評価しているため、これらが同じ格にならないように、下位構造の参照によるチェック機構を導入した方がよりよいだろう。

また(b)については慣用句の処理(5.8.1章)と同じ議論が成り立つ。さらに「名詞1は名詞2が形容詞」に対して、「名詞1 have 形容詞 名詞2」のような長いパターンを用意しておくことで、より一般性を増すことができるだろう。

(d)に関しては特殊な用法であるため、「近い」「遠い」を固定項に含むパターンで記述し、他のルールと区別した方がよいように思える。

#### 5.8.7 比較表現

##### [一般事項]

比較表現には、同格比較、相対比較、最上級比較などがある。

#### [TDMTでの対応]

同格比較に関しては、TDMTでは“(Xと同じくらい Y)”というパターンを利用することで、評価文を処理することができた。また相対比較、最上級比較については、“(Xより Y)”、“(最も X)”という一般的なパターンを利用して、評価文を処理することができた。さらにその他の比較表現として「～すれば～するほど～」があげられており、この場合「～すれば～するほど」の表現から比較表現が現れるため、TDMTではその全てを含む長いパターンで記述した。

#### [考察]

上記の同格比較ルールについては、固定項部分に副詞に係る表現(「～と全く同じくらい～」など)には別ルールを用意せざるを得ない。このため、一般的な副詞のルールで処理できるようにルールを分ける必要があるだろう。ただし、ルール間の整合性を保つ工夫は必要だろう。また、同格比較には「～と同程度～」などのように表現のヴァリエーションも多いため、網羅的な収集も必要かも知れない。さらに英語でY部は、X部と類似した構造を補完すべき場合も多いが、このルールからはX部の構造の参照(下位構造の参照)ができないため、このような補完ルールを記述できないという問題もある。

また上記の相対比較、最上級比較のルールでは、曖昧性が生じた場合に、用例で訳し分けられるかどうかには疑問が残る。

さらに「～すれば～するほど～」のパターンを用いても曖昧性が残る場合があり、第1項目の変数と第2項目の変数が一致するなどの条件が記述できれば、より精度の高いルールになると思われる。

### 5.8.8 比喩表現

#### [一般事項]

比喩表現には直喩と隠喩があり、字義通り訳して良いかどうかなどを判断して処理する必要がある。

#### [TDMTでの対応]

評価文は、TDMTでは“(Xのように Y)”のパターンを用いることで処理することができた。

#### [考察]

上記パターンのX部の単複はその比較対象の単複と一致するが、現行のTDMTでは下位構造の参照が許されていないため、このような操作を記述することができない。またこのパターン中のY部は係り先であって、比較対象はその下位構造に含まれているため、これを追うこともできない。

隠喩を比喩として訳す場合には、文全体の意味解析や文脈処理が必須であり、膨大な背景知識と高次の推論が必要となる。TDMTは字義通りの訳を出力しているが、現在のところこれ以上のことをする必要性は低いと思われる。

### 5.8.9 部分否定、二重否定、倒置文

#### [TDMTでの対応]

部分否定については、“(Xとも Y ない)”や“(Xともは Y ない)”といった長いパターンを利用して評価文に対するルールを記述した。また二重否定の評価文「それは望みがないこともない」については、その内側の否定「望みがない」に「hopeless」という一語の訳を与え、“(X こともない)”というパターンで文全体を否定形にすることで対処した。さらに倒置文に関しても記述可能である。

[考察]

上記は英語上で呼応表現を取っているため、ルールとして分割することは難しいだろう。しかし、対応する日本語表現のヴァリエーションが多いため、同様の表現を網羅的に収集する方法が問題になるだろう。

また二重否定については、内側の否定に対して一語訳が存在するとは限らず、より一般的な形式を模索する必要があると思われる。

倒置文のルールは他のルールと衝突する可能性が大きいですが、その必要性は低いと考えられ、ルール化しない方がよいだろう。

#### 5.8.10 敬語

[TDMT での対応]

TDMT では辞書中に敬語を持っており、通常の処理と同じ処理でこれらを解析、翻訳できる。評価文も全て正しく翻訳することができた。

#### 5.8.11 引用・伝聞表現

[一般事項]

ここでは「～と言った」「～と言われる」などの表現を引用・伝聞表現として取り上げている。

[TDMT での対応]

TDMT では基本的に“(X と Y)” というパターンで格助詞「と」の機能を認識し、動詞句と動詞句をつなぐ。

[考察]

文と文をつなぐことから、時制の一致や主語の一致が記述できないという問題がある。

#### 5.8.12 例示・列挙表現

[一般事項]

ここでは特に例示表現、体言性列挙、反意語並列、同型文並列を扱っている。

[TDMT での対応]

ここにあげられた評価文は並列部分全体を含む長いパターンを用いて記述した。

[考察]

並列構造を持つ名詞句の処理(5.2.5章)で述べたように、TDMT で並列構造を認定することは難しく、上記のルールには一般性がない。並列構造を処理するためには何らかの機構を考える必要があるだろう。

## 6 おわりに

TDMT の評価の 1 つとして、翻訳システムが対応すべき基本的な文法事項の TDMT 上での実装について検討した。その結果、単文の翻訳ルールはほぼ記述可能であるが、複文重文の翻訳ではやや難が残る場合があることが分かった。これは、特に部分的な情報だけでは訳が決められない場合にも決定的な処理を行っている点が原因であると考えられる。これらに対応する機構の検討も行ったので、今後実装に向けて、さらなる検討を進める予定である。また、ルール間の整合性や拡張性も考慮して検討を進めたので、これらは今後のルール作成方針としてまとめる予定である。

## 参考文献

- [1] J.Hutchings, "Evaluation of Machine Translation and Translation Tools",  
<http://www.cse.ogi.edu/CSLU/HLTsurvey/ch13node5.html>.
- [2] "例文評価研究会報告書", アジア太平洋機械翻訳協会 研究成果報告書 (94 年版), 1994.
- [3] J.S.White 他, "The ARPA MT evaluation methodologies: evolution, lessons, and future approaches", *Proc. of the 1st Conference of the AMTA*, p.193-205, 1994.
- [4] 浦谷則好 他, "翻訳機能の評価法", ATR テクニカルレポート TR-IT-0057, 1994.
- [5] H.Isahara 他, "Technical Evaluation of MT Systems from the Developer's Point of View: Exploiting Test-Sets for Quality Evaluation", *Proc. of the 1st Conference of the AMTA*, 1994.
- [6] E.H.Nyberg 他, "Evaluation Metrics for Knowledge-Based Machine Translation", *COLING94 proc.*, Vol.1, p.95-99, 1994.
- [7] 日本電子工業振興協会, "機械翻訳システム評価基準 - 品質評価用テストセット -", 日本電子工業振興協会 95-計-17, 1995.
- [8] 黒橋禎夫 他, "並列構造の検出に基づく長い日本語文の構文解析", 自然言語処理, Vol.1, No.1, pp.35-57, 1994.
- [9] 宅崎成志, "TDMT ルール作成 -JEIDA 評価報告書-", (株) コングレ, 1996.

## A 評価対象文一覧

### 述部

#### 述部の訳し分け

- 「彼は多くの研究者を集めた」
- 「彼は標本を集めている」
- 「彼は論文を集めて本にした」
- 「彼らは会議室に集まった」
- 「学生が教室に集められた」

#### 断定文

- 「この装置はバッテリー駆動だ」
- 「手順は左右同一である」
- 「プッシュボタンは簡易操作に最適である」

#### 体言述語

- 「委員会は彼らの訴えを却下」
- 「部品を検査するため」
- 「部品の検査のため」
- 「部品 A は検査不要」

#### 複合述部

- 「私達は研究開発する」
- 「検査者は部品を修理、計器を点検する」
- 「牛乳は栄養豊富である」
- 「資料は当日配布すること」
- 「渋滞が自然解消する」
- 「住民が自然保護する」

#### 訳が一用言となる並列用言

- 「ナット A を軸に組み付ける」

#### 用言の副詞(句)化

- 「クーラントを点検し、必要ならば補給する」

#### 補助動詞

- 「彼は実験を続けてみる」
- 「彼女が電流計を見ている」
- 「雪が溶けていく」
- 「私がこの装置の増幅度を上げていく」
- 「彼は東京へ行く」
- 「私は本を読んでいる」
- 「軸受けに油が溢れている」
- 「机の下に猫がいる」
- 「私はこの本は読んである」
- 「操作法はマニュアルに全部説明してある」
- 「最大値・最小値の値を示してある」
- 「テーブルの上に蜜柑がある」
- 「この計算機には3つのタイプがある」
- 「中心部にひびが入ってくる」
- 「結果に違いがでてくる」
- 「鶴は秋に日本へ来る」

「地形が交通の障害となる」  
「私は算盤が使えなくなった」  
「すべての物質は原子からなる」  
「パイプにしっかりと栓をしておきなさい」  
「机の上に本をおく」

#### 基本動詞の訳し分け

「温度は一定とする」  
「この工場には最新設備がある」  
「我々の会社は都心にある」  
「基本構成にはフレキシブル・ディスク装置がない」  
「私にはその問題を解決する手掛かりが何もない」  
「あの会社は倉庫に在庫がない」  
「その記述には一貫性がない」  
「我々はその問題を解決できるだろう」  
「明日にはこの報告書ができる」  
「バターは牛乳からできる」  
「ディスプレイの表面がガラスでできている」  
「彼の実験で新しい物質ができた」  
「計画のための委員会ができた」  
「香港に当社の支社ができた」  
「彼女は数学ができる」  
「この報告書には誤りが多い」  
「この機械は故障が多い」  
「新鮮な果物はビタミンが多い」  
「この報告書は誤植が少ない」  
「実験材料は鉄分が少ない」  
「労働者の不満を少なくせよ」  
「この機械が故障することは少ない」

#### 名詞

##### 複合名詞

「オブジェクト指向プログラミング入門」  
「価格算定基準」

##### 「名詞1の名詞2」という構造を持つ名詞句の処理

「私の指は太い」  
「剃刀の刃は危険です」  
「空のボトル」  
「有色の印刷物を複写する」  
「不断の研究がこの成果をもたらした」  
「盲目の少年はジェスチャーで感情を表した」  
「雄の体に紅色の色素が含まれる」  
「エボナイトの棒を毛皮でこする」  
「この国には伝染病の死亡者が多い」  
「私は電話の勧誘は嫌いだ」  
「模型の試験」  
「政府の決定により、試験運転は延期された」  
「重量の節減」



「色素の沈着が強くなる」  
 「気温の変化が激しい」  
 「彼の行動は素早い」  
 「専用工具の使用が望ましい」  
 「その機械は管の空気を振動させる」  
 「A社は世界のコンピュータ市場を独占する」  
 「x方向の速度を変化させる」  
 「半径方向のひずみを測定する」  
 「私は今日の新聞の記事を読んだ」  
 「私は今年目標をすべて達成した」  
 「現代の半導体技術」  
 「彼はレンズの中心を拭く」  
 「金属の硬度」  
 「空気の密度が小さい」  
 「長時間勤務は労働者の能率を低下させる」  
 「医者がその薬の効果を証明する」  
 「研究の見通しが明るい」  
 「彼らは本のページに番号をつける」  
 「彼女は数学の才能を発揮し出した。」  
 「融点の差は大きい」  
 「彼は化学の分野で偉業を成し遂げた」  
 「自動車の事故が多い」  
 「磁気テープ上の情報を文字に変換する」  
 「自動車の事故が多い」  
 「剃刀の刃は危険です」  
 「彼女は数学の才能を発揮し出した」  
 「私の指は太い」  
 「現代の半導体技術」

「名詞1の名詞2の名詞3」という構造を持つ名詞句の処理

「この章の終わりの練習問題をやりなさい」  
 「メーターの読み取りの精度」  
 「彼女は山田さんの長女の由美さんです」

並列構造を持つ名詞句の処理

「私は、神戸と大阪に行った」  
 「並列コンピュータの研究と開発」  
 「システム300とシステム400の価格」  
 「読込部と印字部の処理精度を良くする」  
 「入力部と、入力されたコマンドを処理する処理部からなるシステム」  
 「システムの構成を図3に、特徴を表1に示す」

疑問表現の名詞節の処理

「私はあの装置の開発者が誰か知らない」  
 「ぼくは君がどこに住んでいるか知っている」  
 「その機械はいつ故障するのかわからない」  
 「私は何と答えてよいか途方にくれた」  
 「彼がこの本を書いたかどうか疑わしい」  
 「ぼくは君がどこに住んでいるか知っている」  
 「ぼくはどこに行くべきか分からない」

### 用言性名詞 (サ変名詞)

- 「私は故障の原因の調査をする」
- 「私は故障の原因の調査を行なう」
- 「市場開発」
- 「駐車禁止」
- 「逐次変換」
- 「変化量」
- 「説明書を参照のこと」
- 「濃い溶液の漏れのないこと」

### 英語における数の扱い

- 「私の家族は大家族だ」
- 「私の家族はみな早起きだ」
- 「牛の群れを追っていた狼の群れを狩人が追った」
- 「その大学は全国の高校から学生を集める」
- 「私は各々のコンピュータを調べた」
- 「私は全てのコンピュータを調べた」
- 「私は各々のソフトウェアを調べた」
- 「私は全てのソフトウェアを調べた」
- 「はさみ」
- 「たくさんのはさみ」
- 「女性ドライバー」

### 固有名詞表現

- 「瀬戸内海」
- 「瀬戸内海」
- 「瀬戸内海」
- 「日本海」
- 「東京都港区芝公園 3-5-8」
- 「山田太郎」
- 「山田部長」
- 「内務長官」
- 「情報処理学会」
- 「図 A」

### 形式名詞

- 「プログラムを実行するための作業」
- 「プログラムの実行のための作業」
- 「私はプログラムを実行するためディスクを拡張する」
- 「プログラムを実行したためディスクが満杯になる」
- 「プログラムを実行したためにディスクが満杯になる」
- 「ユーザのためのマニュアル」

### 関係を示す名詞

- 「ディスプレイの上に物を置かないで下さい」
- 「彼は私より年が上だ」
- 「私はボールを上を投げた」
- 「人の上立つのは難しい」
- 「プログラムを実行する上で重要なこと」
- 「プログラムの実行の前にパラメータを設定して下さい」

「プログラムを実行をする前にパラメータを設定して下さい」  
「前に実行したプログラム」  
「ユーザはディスプレイの前に座ります」

## 副詞

### 副詞のタイプ

「ディスプレイを必ず消して下さい」  
「このシステムは徐々に正解を学習する」  
「このコマンドは他のプロセスの実行時間をかなり遅らせる」  
「このアルゴリズムはかなり速い」  
「更に容量が大きいディスクが開発される」  
「この機能は一番早く動く」  
「最も重要なプログラムはオペレーティングシステムである」  
「このプログラムは更に簡潔に書かれている」  
「更にゆっくりカーソルを移動する」  
「最も内側のループを抜ける」  
「かなり多くの記憶領域を保持する」  
「このシステムは僅か1分で計算を終える」  
「最も小さなブロックを抜ける」  
「彼女はよくテニスをする」  
「彼女は黄色い服がよく似合う」  
「昨日私は本を買った」  
「幸い、ファイルの内容はディスクに保存されています」  
「私は冗談を言っただけだが、彼は図らずも完全にそれを信じて、とても喜んだ」  
「全て順調です」  
「彼は作業を全て終えた」  
「管理者がシステムを更新する」  
「随時管理者がシステムを更新する」  
「明日管理者がシステムを更新する」  
「管理者がシステムを更新している」  
「管理者が今システムを更新している」  
「管理者が随時システムを更新している」

### 副詞句

「改めてパラメータを設定して下さい」  
「ディスプレイを正しく設置して下さい」  
「簡単に言うと、カーソルが右に動きます」  
「パラメータの設定により画面の明るさが変化する」  
「LSIが発達したことにより、パーソナルコンピュータが実現された」

### 擬音語・擬態語

「アヒルがガーガー鳴く」

### 連体修飾語句

#### 非活用連体修飾

「手差しトレイには、あらゆる大きさの紙をセットできる」  
「たくさんのコピーをとる場合、ソータを使用して下さい」  
「彼の分析で、その物質に対するたくさん情報が得られた」  
「私たち日本人」

「この現象方式は、いわゆる新世代の技術で可能になった」  
「このキーを押すと F が表示される」  
「点滅は、ほんの数秒で点灯に変わる」  
「このうちの一つを ON にしなければならない」  
「パネル上のスイッチを ON にしなければならない」  
「これがいわゆる金属疲労だ」  
「3本のワイヤ」  
「すべての手法」  
「A と B からの出力信号」

#### 用言性連体詞

「新しい設定にする場合には、リセットボタンを押す」  
「新たな設定にする場合には、リセットボタンを押す」  
「よりよい性能」  
「一番安い車」  
「出力する装置を ON にする」  
「こうした操作は信号レベルの低下を招く」  
「本章は背景とこれに続く問題点から構成される」

#### 格助詞相当句

「減税に関する原案を提出する」  
「雪に対応した装備を身につける」  
「戦争に伴う石油価格の上昇が懸念される」  
「人手不足による倒産が目立つ」  
「教育制度をめぐる議論を繰り広げている」  
「副作用を伴う薬品の製造中止が論議された」  
「この製品は、ソータが付いていないという点を除いてまったく同じものです」  
「それは、車輪を有する乗り物に含まれる」  
「脳死問題と関連する移植について話し合った」  
「毒蜘蛛や蠍といった危険な生物が生息している」

#### 埋め込み文修飾

「この論文を書いた研究者」  
「その研究者が書いた論文」  
「底の丸い部品」  
「アームが収納されるスペース」  
「ファイルが作成された日時」  
「プログラムがエラーメッセージを出した理由」  
「ディスクを交換する前日」  
「経済が回復期に入ったという報告」  
「首相が賄賂をもらったという事実」  
「ものが焼ける匂いがある場合には電源プラグを抜いて下さい」  
「事故につながるとした教授の警告」  
「クレーンを操作する人」

#### 助詞

「彼に聞いた話」  
「私は映画を見に行く」  
「私は大阪に行った」  
「大阪には行った」

「私は大阪にも行った」  
「太陽は西に沈む」  
「ススキの穂が風に揺れている」  
「私は母へ手紙を書いた」  
「彼は窓へ腰掛ける」  
「私は外へ出た」  
「私は父から腕時計をもらった」  
「私は窓から忍び込んだ」  
「風は東から吹いている」  
「四月から新学期が始まる」  
「私はまず玄関の掃除から始めた」  
「パンは小麦粉から作られる」  
「私は不注意から事故を起こした」  
「お前までそんなことを言うのか」  
「当時の人は雑草まで食べた」  
「私は大阪まで行く」  
「子供は5時まで遊んだ」  
「兄は父より背が高い」  
「私はワープロにより文章を作る」  
「天気は西より崩れる」  
「私は3時より仕事を始める」  
「このチームは20人より成る」  
「私も本を買った」  
「私はドイツ語も話せる」  
「彼は1時間も走っていた」  
「彼は新製品だけを売る」  
「この新製品は単体でもソフトと抱き合わせでも入手できる」  
「彼こそ真の英雄だ」  
「私は墓場にさえ行った」  
「その子は4歳にしかならない子供だ」  
「それは誰でも知っている」  
「水泳くらい誰でもできる」  
「彼は学校だけに行った」  
「彼は学校にだけ行った」  
「彼女ばかり泣いていた」  
「これは研究者のみが知る喜びだ」  
「病人は水すら喉を通らなかつた」  
「屋根などに登ってはいけない」  
「私は地震も怖くない」  
「このCPUこそ最も速い」

#### 接辞

「上寿司」  
「クリーム状のスープ」  
「量対質」  
「一般的な事項」  
「解析的な手法」  
「日本の社会がアメリカ化される」

「不況が深刻化する」  
「無節操な行動」  
「日本電気製のコンピュータ」  
「ドイツ製のペン」  
「ジュラルミン製のアタッシュケース」

#### テンス、アスペクト、モーダル

##### テンスの処理

「彼は、誤りを犯したことがわかった」  
「彼は、誤りを犯したことがわかっている」  
「彼は、必要なものをすべて準備した」  
「もし走って行けば、間に合うだろう」  
「地図を持っていれば、貸してあげるのだが」

##### アスペクトの処理

「私はボタンを押している」  
「窓が開いている」  
「ピストンは1分間に1000回往復している」  
「このマニュアルは3回改訂している」

##### モーダルの処理

「一致団結してこの危機に臨もう」  
「ボタンを押しなさい」  
「彼らを電車で学校へ行かせなさい」  
「配電盤には常に人を配置して置くようにすること」  
「受験者は教科書を試験場に持ち込んではいならない」  
「決して彼を1人にするな」  
「ここに駐車してもよろしい」  
「彼は外へ行ってもよかった」  
「その手紙のコピーをいただけますか」  
「道を教えていただけますか」  
「所得税の申告書は正しいものを提出しなければならない」  
「信号で止まらなければならなかった」  
「私は明日の朝、出発しよう」  
「コーヒーでも飲みましょうか」  
「僕はパイロットになりたい」  
「ねじが緩んでいるようだ」  
「ねじが緩んでいるらしい」  
「ねじが緩んでいたのだ」  
「正月と盆が一度に来たようだ」

##### ボイスの処理

「ICカードが挿入されている場合にシステムが動作する」  
「ボタンを押して停止させる」  
「暴走した場合には強制停止ボタンを押して停止させる」  
「私は秘書に今度の会議の原稿のチェックをしてもらった」  
「ログが残っていない場合には、ユーザに使用状況を再現してもらおうと良い」

#### 特殊構造表現

##### 慣用表現の処理

「彼らは木陰で油を売っている」  
「私は彼のしっぽをつかんだ」  
「彼は背が高い」  
「時は金なり」  
「ローマは1日にしてならず」  
「私は首の差でレースに勝った」

#### 四字熟語

「呉越同舟」  
「これは一朝一夕に解決できる問題ではない」

#### 呼応表現

「彼は英語をまるで母国語のように話す」  
「道路の狭いのに加えて、雨が降っていた」  
「あの会社は通信機器販売のほか、ソフトウェア開発も行なっている」  
「雪も前ほど降らなくなった」  
「彼は自分の名前すらかけない」  
「夕方にはきつと雨が降る」  
「この本は子どもでさえ読める」  
「例は少なくとも1つ以上はある」

#### 天候・気象表現

「雨が降っている」  
「去年は雨が多かった」

#### 無生物主語構文

「調査により性能が向上したことがわかった」  
「研究で、摩擦でロープが切れることがわかった」  
「図3を見ると温度が上昇することがわかる」  
「コントロール・パネルによりハードウェアに適切な設定を選択することができる」

#### 「はが」構文

「私は字が読める」  
「彼女は花が好きです」  
「彼は耳が遠い」  
「その建物は外観が良い」  
「冬は空気が乾燥する」  
「私の家は学校が近い」

#### 比較表現

「私は彼と同じくらい背が高い」  
「弟は数学では兄より優れている」  
「その火山は以前より高くなった」  
「経済成長率は日本が最も高かった」  
「ロケットが速くへ行けば行くほど、信号は弱くなる」

#### 比喻表現

「彼女は雲雀のように歌う」  
「発明の母」

#### 部分否定、二重否定、倒置文

「私は2人とも知らない」  
「私は2人ともは知らない」

「それは望みがないこともない」

「私はそんなことを絶対に思っていない」

敬語

「詳細はマニュアルをご覧ください」

「私はこの本を先生から頂きました」

引用・伝聞表現

「彼は又来ると言った」

「聴衆は彼の理論が間違いだと判断した」

例示・列挙表現

「東京の駅、例えば新宿などは混雑が激しい」

「彼は、物理、数学などの科学の分野を研究している」

「G7、先進7カ国の蔵相、中央銀行総裁が会議に出席した」

「その要素は集合Aにも集合Bにも属さない」

「インストール時にはアクセスランプがついたり消えたりします」

「父はタバコも吸わず、酒も飲まない」