

拡張性を重視した日英機械翻訳システム

北村美穂子・甲斐郷子・岡田恵太・永田淳次
(沖電気工業(株))

拡張性を重視した日英機械翻訳システムの文法規則の設計方針と実現方法について述べる。機械翻訳システムの実用化のためには、システムリリース後も文法規則の改良は必要である。我々は、実用化機械翻訳システムの開発プロセスにおける問題点と解決策を検討し、その問題点を解決するための一手法として、文法規則を

1. 基本規則(basic rule)
2. 一般規則(general rule)
3. 経験的規則(heuristic rule)
4. 例外規則(non-general rule)

に階層化し、さらに各規則を基本処理単位ごとにモジュール化するという「階層化」「モジュール化」をとり入れた機械翻訳システムを実現した。

この手法により、文法規則の改良容易性が高められ、開発効率が向上した。

An Extensible Japanese to English Machine Translation System

Mihoko KITAMURA, Kyoko KAI, Keita OKADA, Junji NAGATA
Oki Electric Industry Co., Ltd.
2-27, Shiromi 1-choume, Chuo-ku, Osaka 540, Japan

We describe the rational and implementation techniques in designing grammar rules for an extensible machine translation system.

An important feature of a practical machine translation system is the extensibility of their grammar rules during the entire life time of the system. We analyzed the problem in designing and improving grammar rules and developed a method for designing an extensible grammar rules. The method is based on the following two techniques. The first is to organize grammar rules in the hierarchy of four layers: basic rules, general rules, specific rules, and heuristic rules. The second is to implement code to process each rule as a module.

Our experimentation showed that these two techniques yield a more extensible machine translation system, and they also make the development process of a system more efficient.

1. はじめに

情報流通の国際化が進むとともに、翻訳のニーズは高まる一方である。このニーズに答えるために、近年、機械翻訳システムの実用化が注目されている。しかし、現状の機械翻訳システムの翻訳精度（翻訳の質）は実用に耐えるとは言い難く、利用者が期待する翻訳結果はなかなか得られない。実用化における問題点を解決策するためには、学習機能などの翻訳支援環境の整備や専門辞書の拡充が試みられている[1][2][3]。

我々は、機械翻訳システムの実用化を目指す立場から、機械翻訳システム開発における翻訳精度向上のための問題点を分析し、その解決策を検討した。

本稿では、まず、機械翻訳システムにおける文法開発プロセスについて述べる。そして、そのプロセス中に発生する問題点を解消するための手法として、拡張性を重視した機械翻訳システムの設計方針を提示し、さらに、その実現方法についてを述べる。最後に、今後の課題を述べる。

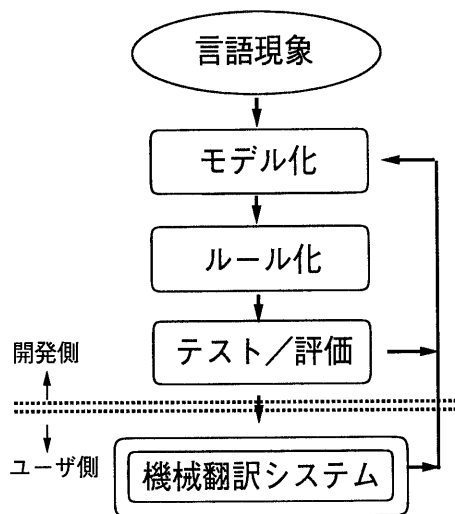


図1 文法開発プロセスの流れ

2. 機械翻訳システムの開発

2.1 文法開発プロセス

機械翻訳システムは、エキスパートシステムの一つであり、翻訳に必要な知識として文法規則やイディオム、単語などの辞書から構成される。翻訳精度の向上には文法規則の改良、イディオム、辞書の追加が必要である。

実用レベルを意識した機械翻訳システムは、約1,000を越す文法規則を持っている[4]。そして、個々の文法規則は、相互に関係し合い翻訳処理を実現している。

機械翻訳システムに対するユーザの要求は拡大する一方である。我々は、そのユーザの要求に答えるため、また、さらに高度な課題を解決するため、文法規則の改良を実施している。

文法規則の改良は図1の流れにしたがって、行なわれている。

(1)モデル化工程

モデル化工程は、一般に設計工程と呼ばれる工程と同じである。

ここでは、ユーザからの要求等の課題をもとに、翻訳の対象となる言語現象から仮説を導き出す。この仮説を「モデル」と呼ぶ。モデルは言語学の研究成果や、コーパス、翻訳者の内観などに基づいて、設計される。この工程の作業は、言語現象に関する知識が豊富な人が行なうのが、一般的である。

モデル化工程は、さらに、次の二段階に分かれる。

(a) 言語現象の解析と形式化

課題をもとに、言語現象の解析を行ない、その結果を仮説として形式化する。

(b) 既存モデルとの整合

(a)で考えられた仮説をもとに既存モデルとの関係を調べ、矛盾がないかを確認し、モデルとする。

開発当初の機械翻訳システムでは既存のモデル数が少ないため、言語現象にのみ注目するだけでモデル化が可能であるが、既に多くのモデ

ルが存在する改良の場合のモデル化では特に次のことを考慮する必要がある。

(a)ある特定の言語現象がどのレベルで解決可能であるかを判断する。

レベルとは言語学における語い論、構文論、意味論に相当する。

この判断の結果、そのモデルがどのような枠組によって解釈できるかが判明しやすくなる。

(b)モデルの一般性または特殊性を判断する。

そのモデルがある分野の文章に特有の言語現象か、一般に通じる言語現象かを見定める。これは、ルール化の際、モデルの優先度を定める時の決め手となる。

(c)モデルの対象を規定する。

モデルを導き出す際、そのモデルの対象外の言語現象も把握しておくことが望ましい。これは、新しいモデルを導き出す時の手助けとなる。

(2) ルール化工程

ルール化工程は、モデル化工程で導き出されたモデルを翻訳処理内部の文法規則（ルール）の形式に変換し、機械翻訳システム内に組み込む工程である。つまり、モデルを実現する工程に当たる。

翻訳処理内部の文法規則は、通常、ある特定の計算機言語により表現されるので、この作業は計算機言語に関する知識が豊富な人に向いている。

この工程では、モデルを実現するための文法規則を追加する作業が主であるが、既存の文法規則に変更を加えたり、不要とされる文法規則を削除する作業もある。

文法規則の追加／削除は、既存の文法規則の振舞いを把握した上で行なう。変更はさらに既存の文法規則と関係のある文法規則を調べた上で行なう。

そのため、以下の点に留意して、ルール化を行なうことが重要である。

(a)優先順位の変更

ある文法規則に適合する一文が、相反する別の文法規則にも適合した場合、どちらを優先的に採用させるかを見定める。

(b)相互作用の把握

文法規則数が多くなり、文法規則間の相互関係が複雑になれば、文法規則間の相互作用も複雑になる。相互作用を把握した上で、ルール化を行なう。

(c)変更仕様の明確化

既存する文法規則と新しく追加する文法規則間の適用順序や相互作用に関する仕様を明確にし、今後の文法規則の変更／追加作業にも耐えるようにする。

(3) テスト／評価工程

(a) テスト

テストとは、追加及び変更した文法規則の改良効果を確認する作業である。この作業は、モデル化の対象となる言語現象の典型的な例文を用いて翻訳を行ない、その結果をもとにして、改良効果を確認することが一般的である。

改良によって翻訳精度が低下していないかどうかを確認する。この翻訳精度の低下を「副作用」と呼ぶこととする。

本工程においては、この副作用の確認が重要となる。

この確認作業は、あらかじめ正解例（既存の翻訳結果を利用することが多い）を用意しておき、正解例と翻訳結果の差分をとることによって行なわれる。しかし、実際には、文法規則の改良は、多くの文に影響を与えるため、多くの差分が出現する。

ゆえに、差分が出現した一文一文について、副作用かどうかを確認しなければならないことになる。テストにおけるこの作業は手間のかかる困難な作業である。それゆえ、モデル化・ルール化の際には、この副作用を考慮に入れることが重要となる。

(b) 評価

評価とは、翻訳結果が文として適切かどうか総合的に判断したり、翻訳結果の欠陥を指摘する作業である[5]。

評価は、一般に、翻訳対象言語と翻訳目的言語両方の知識が豊富な人により行なわれる。

改良は、評価者が抽出した欠陥を分析し、改良項目を決め、モデル化・ルール化のサイクルを繰り返すことにより行なわれる。しかし全ての欠陥が単純な文法規則の追加で解消できるとは限らないため、「改良項目の決定」には慎重さを要する。

ゆえに、欠陥の指摘がすぐに改良に結び付くとは限らない。

2.2 機械翻訳システム開発の特徴

機械翻訳システム開発に関しては、元来、次に示すような特徴がある。

- ・ 言語現象は時代と共に変化する。
- ・ 全ての言語現象をあらかじめ想定して、それを満足する文法規則を構築することは事実上不可能である。
- ・ ユーザ要求のレベルがあらかじめ明示されることはない。
- ・ 一度、機械翻訳システムを試用したユーザーの要求はさらに高度なレベルを要求する。
- ・ 言語学で扱う文法理論の枠組を越える言語現象が存在する。

この特徴は、機械翻訳システムにおける文法規則の開発は、最初から完全な形態の構築を目指すのではなく、まず核となる原型を構築し、その後、前述した文法開発プロセスのサイクル（モデル化・ルール化・テスト／評価）を繰り返すことにより開発せざるを得ないことを示している。

また、個々のモデルごとに多くのサイクルを繰り返すことになる。

3. 設計方針

前述したように、機械翻訳システムの文法規則の仕様は、あらかじめ定義することはできないので、一般に文法規則は単純な集合として構築される。

したがって、新たな言語現象に対する処理の追加は相互作用を無視した文法規則の単純な追加となってしまう。また、ユーザの要求（問題の指摘）に対する改良もその現象を押え込むだけの文法規則の追加になる傾向がある。

それは、文法規則と言えるレベルではなく俗に言うパッチ（対症療法）的な規則である。

モデル化、ルール化、テスト／評価のサイクルを繰り返すことによって、この問題はさらに大きくなる。パッチにパッチを重ねることにより、新たな副作用を呼び、翻訳精度の向上が不可能となってしまう。

一方、文法開発プロセスはサイクリックであるため、頻発する文法規則の追加／削除及び変更の作業は効率良く行なえることが必要となる。

文法規則が膨大である場合、既存の文法規則の改良は予想以上に難しくなる。特にそのダイナミックな動き（その文法規則がいつ発火されるか）を簡単に理解することは難しい。

このような理由により、文法規則には秩序立てられた体系が必要になる。

我々は、文法規則の体系を秩序立てるため、

- ・ 階層化
- ・ モジュール化

という設計方法を用いて文法規則を構築した。

4. 実現方法

4.1 基本設計

我々が行なった文法規則の「階層化」と「モジュール化」について、以下に説明する。

4.1.1 階層化

文法規則の設計で重要なことは、特殊な言語現象（非常に稀なケース）に対する処理が、全体の翻訳の処理の流れを崩さないようにすることである。特殊なものを扱え、かつ拡張性を保つ文法規則を実現するためには、まず全体の翻

訳の流れを崩さない核となる枠組と特殊なものを扱える枠組とを用意する必要がある。

我々は、文法規則を、翻訳処理の核となる（いわゆる骨格となる）規則と、実際に翻訳処理を実現する（いわゆる肉付けとなる）規則、さらに経験的に得られた規則、の3つに階層化することにより体系立てた。

- ・骨格となる文法規則のことを
 - (1)基本規則
- ・肉付けとなる文法規則のことを
 - (2)一般規則
- ・経験的に得られた規則のことを
 - (3)経験的規則

と名付けた。

- ・そしてさらに特定の言語現象に対処するため、一般規則の枠内に
 - (4)例外規則

を設けた。

その関係を図2に示す。

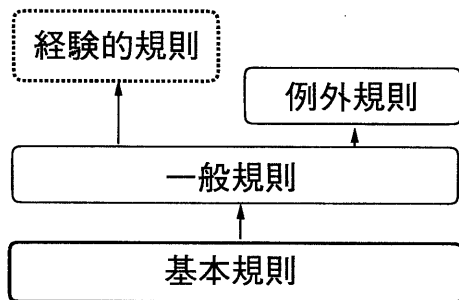


図2 「階層化」 各規則間の関係図

それぞれの規則について以下に述べる。

(1) 基本規則

基本規則は、翻訳処理の流れを制御している文法規則の集合である。

文法規則は、品詞の推定に始まり、句の処理、節の処理と処理が流れていく。各処理の先頭でまずこの基本規則が実行され、必要に応じて、

後述する一般規則等が呼び出され、翻訳処理を実現している。

例えば、品詞推定処理、名詞句解析処理、動詞句解析処理などの翻訳処理を規定する大枠の文法規則が本規則に相当する。

そのため、基本規則は、追加／削除及び変更が行なわれない文法規則の集合であり、文法規則の中で、最も中心的な役割を果たす。

(2) 一般規則

一般規則は、前述の基本規則から呼び出される文法規則である。基本規則が翻訳の流れを制御するのに対して、一般規則は翻訳処理を実現する文法規則の集合である。

例えば、「名詞＋格助詞→名詞句」や、「名詞＋副助詞→名詞句」という文法規則が一般規則であり、これらは、基本規則の名詞句解析処理から呼び出される。

「xxがooを…する。」のように構成が単純な文では、この一般規則と基本規則との二つの文法規則で翻訳処理が実行される。

一般規則には、さらに複雑な文に対応するための処理として、後述する経験的規則や例外規則を呼び出すための処理がある。これは、可能性の最も高い文法規則、またはある特定の処理に対応できなかった場合の文法規則を本規則がもった上で、経験的規則や例外規則を呼び出し処理を行なう、という形式をとる。

(3) 経験的規則

機械翻訳システムの入力文は、特別な制限を加えない限り、曖昧性を持ち、解釈が複数になることが多い。また、言語現象には、言語学で扱う文法理論の枠内に適合せず、大量の文を調査し分析した結果、得られるものが多く存在する。

経験的規則には、複数の解釈に対して適切な解析を行なうため、また言語学で扱う文法理論の枠外の言語現象を処理するための規則が用意されている。

例えば、名詞句解析処理における日本語並列句の係り先認識処理などがある。

(4) 例外規則

基本規則と一般規則との二つの文法規則で、翻訳処理の主要部分を実現されている。しかし基本規則と一般規則で実現される処理の流れに沿わない複雑な処理を必要とする言語現象もある。

例外規則は、この流れに沿わない言語現象を処理するために用意された文法規則である。

処理の流れに沿わない、特定の言語現象を例外規則内に押え込むことによって、基本規則や一般規則の変更を避けることが出来る。これは基本規則や一般規則の複雑化を避けることに貢献している。

例えば、呼応の解析、命令文の解析、助数詞の解析などがある。

4.1.2 モジュール化

モデルのルール化過程においては、

- ・モデルの働きに対応した扱いやすい単位で文法規則をまとめる。
- ・モデル間の相互作用（インターフェイス）を明確にする。

ことが重要である。

そのため、モデルの機能を文法規則の振舞いに対応させて特定の処理単位で表現する必要がある。この処理のまとまりをモジュールとする。これは、文法規則の構成に秩序を与え、保守性の向上を実現する。

モジュールは図3のような形式で表される。

ここでは、モジュールAはモジュール内の文法規則の実行の優先度の制御や規則の実行回数の制御を行なう。また、必要に応じて他のモジュールの呼び出しも行なう。

また、「階層化」の各規則と対応付けると、モジュールA、Bは基本規則に、モジュールa1、a2、a3は一般規則に相当する。モジュールa1'、a1''は一般規則または経験的規則、例外規則に相当する。

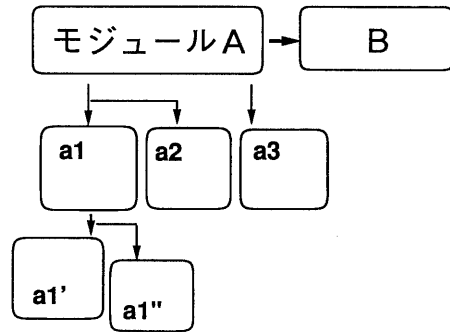


図3 モジュールの形式例

4.2 実例

「階層化」及び「モジュール化」について、格構造解析処理を例にして説明する。

我々が現在開発している日英機械翻訳システム[6]における格構造解析処理は、日本語表層格と句本体の意味情報によって文の格構造を決定するという格文法理論に基づいた手法をとっている。

4.2.1 階層化の実例

本機械翻訳システムで実現されている格解析処理における文法規則の階層化の例を示す。また各規則の日本語における例を表1に示す。

(1) 基本規則

- (a) 格と用言のかかり受け関係は交差しない。
- (b) 格は後出の用言にかかる。
- (c) 一つの用言がとる格要素は1種類につき1つである。(一文一格の法則[7])

(2) 一般規則

- (a) 必須格はその格をとる用言に優先的にかかる。
- (b) 自由格はその格から最も近い用言に優先的にかかる。

(3) 例外規則

- (a) 受身文の表層格の変換処理

(4) 経験的規則

- (a) “、”をもつ格はその格から最も遠い用言に優先的にかかる。

(1) 基本規則

a. 格の非交差の法則

A社は B社に 鉄鋼部門から 撤退すると 断言した。

b. 格は後出の用言にかかる

A社は 計画して B社を 買収した。

c. 一文一格の法則

A社は B社が 重要だと 判断した。

(2) 一般規則

a. 必須格はその格をとる用言に優先的にかかる

A社は B社に 鉄鋼部門から 撤退すると 断言した。

b. 自由格はその格から最も近い用言にかかる

A社は B社とともに 鉄鋼部門から 撤退すると 断言した。

(3) 経験的規則

a. ”、”をもつ格はその格から最も遠い用言に優先的にかかる

A社は、B社とともに 鉄鋼部門から 撤退すると 断言した。

(4) 例外規則

a. 受身文の表層格変換処理

A社が B社により 買収された

<本例の説明>

正しい かり先
 可能性のある かり先
 かかってはいけない かり先
 既に決定済み

表1 階層化の実例

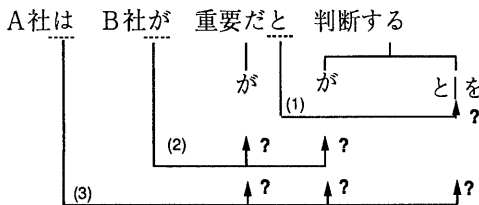
4.2.2 格解析処理におけるモジュール化の実例

本機械翻訳システムで実現されている格解析処理における文法規則のモジュール化の例を表2に示す。

この例では、助詞の特徴をひとまとまりの文法規則として構築できるように、助詞の表層ごとに文法規則をモジュール化している。

助詞における表層格の推定処理

(最右部の助詞から左の方向へ
向かって格解析処理は進む)



- (1) 表層『と』の処理
- (2) 表層『が』の処理
- (3) 表層『は』の処理

一文一格の法則
格の非交差の法則

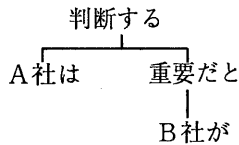


表2 モジュール化の実例

翻訳精度向上が促進された。

今後、この拡張性を利用して、微細なまた特殊な言語現象に対する処理の追加を行ない、本機械翻訳システムの翻訳精度向上に努めたい。

しかし、その一方で、機械翻訳システムが一般に普及し、大量かつ短いライフサイクルで出回るようになると、現在のようなエキスパートによる開発だけでは到底対応できなくなることが予想される。これは、

- ・ 文法開発時における分業化の促進
- ・ 翻訳結果の評価方法の確立
- ・ 翻訳対象文の仕様の明確化
- ・ 開発環境やツールの整備

のような高度な対応を行なうことで解決していかねばならない実用化機械翻訳システムの新たな課題である。

参考文献

- [1]宮原他: "実用性の向上を目指した機械翻訳システムの改良(2)", 情処学会第42回全国大会2C-8, 1991
- [2]"機械翻訳システムの実用化に関する調査研究", 日本電子工業振興協会, 1991
- [3]伊藤他: "類似文検索機能を備えた翻訳支援システム", 情処学会第39回全国大会3G-8, 1989
- [4]"機械翻訳システムの開発と実用に関する実態調査", 日本電子工業振興協会, 1989
- [5]成田: "機械翻訳における構造処理能力の評価", 情処学会NL研究会69-1, 1988
- [6]坂本他: "日英/英日機械翻訳システム PENSEE", bit別冊 機械翻訳, 共立出版, 1988
- [7]J.M.Fillmore, 田中春美訳: "格文法の原理", 三省堂, 1975

5. おわりに

本稿では、実用化機械翻訳システムの文法開発プロセスにおける翻訳精度向上のための各工程の問題点とその解決策を提示した。そして、その問題点を解決するための一手法として「階層化」「モジュール化」をとり入れた機械翻訳システムの実現法について述べた。

この手法を採用することにより、文法規則間の副作用を必要最低限にする文法規則の追加/削除及び変更が可能となり、文法規則の拡張性が高まった。その結果、改良作業が容易になり