TR-IT-0318

多言語話し言葉翻訳システム (TDMT) における構成素境界解析

Constituent Boundary Parsing for Multi-lingual Spoken-language Translation System (TDMT)

山田 節夫　　山本 和英　　古瀬 蔵 *

Setsuo Yamada　Kazuhide Yamamoto　Osamu Furuse

1999 年 12 月

概要

多言語話し言葉翻訳処理では、多様な表現を扱える頑健性、円滑なコミュニケーションのための実時間性、いろいろな翻訳ペアに適用できる汎用性、が要求される。本レポートでは、表層パターンのみの照合を行なう構成素境界解析を提案し、この解析と用例利用型処理を組み合わせた新しい変換主導翻訳 (TDMT) について述べる。また、TDMT が、頑健で、実時間で、汎用性が高いことを「旅行会話」を翻訳対象とした実験により示す。

ATR 音声翻訳通信研究所

ATR Interpreting Telecommunications Research Laboratories

* NTT サイバーソリューション研究所

NTT Cyber Solutions Laboratories

目次

第 1 章

はじめに

多言語話し言葉翻訳システムの処理には，文法から逸脱した表現などを含めた多様な表現を扱える頑健性，円滑なコミュニケーションのための実時間性，原言語と目的言語の様々なペアに適用できる汎用性，が必要である．多様な話し言葉表現をカバーするために詳細な構文意味規則を大量に記述する規則利用型 (rule-based) 処理は，多言語翻訳にとっては経済的な手法でない．一方，用例利用型 (example-based) 処理は，翻訳例の追加により翻訳性能を向上させていく汎用性の高い手法である．ただし，生データに近い状態の翻訳例をそのまま使うと，入力文に類似する翻訳例が存在しない場合が多くなる，翻訳例を組み合わせて翻訳結果を作り上げるには高度な処理が必要になる，などの問題が起こり，多様な表現に対して高精度の翻訳を実現することが困難になる．そこで，単純な構文構造や意味構造へ加工した用例を組み合わせて利用すれば，単純な解析を使うことによって頑健性も汎用性も高い翻訳処理が実現できる．

筆者らは，パタン照合 (pattern matching) による構文解析と用例利用型処理を用いた変換主導型機械翻訳 (Transfer-Driven Machine Translation, 以下， TDMT と呼ぶ) を話し言葉の翻訳手法として提案し，「国際会議に関する問い合わせ会話」を対象とする日英翻訳に TDMT を適用した (古瀬, 隅田, 飯田 1994)．しかし，この時点の TDMT は，頑健性，実時間性，汎用性においてまだ問題があった．

文献 (古瀬他 1994) では，多様な表現をカバーするために，表層パタンと品詞列パタンの使い分け，パタンを適用するための入力文の修正，などを行なっていた．例えば，名詞列について，ある場合は複合名詞を表すのに品詞列パタンを照合させ，別の場合は助詞を補完して表層パタンを照合させていた．しかし，どのようにパタンを記述すべきか，どのような場合にどのように入力文を修正すべきか，などの基準が不明瞭であった．そのため，誤った助詞を補完したり，補完の必要性を正確に判別できなかったりする場合があり，多言語翻訳へ展開するための汎用性に問題を残していた．また，限られた長さの複合名詞を品詞列パタンにより記述していたため，任意の長さの複合名詞を扱うことができないなど，頑健性にも問題があった．さらに，解析途中で構文構造候補を絞り込むことができない構文解析アルゴリズムを採用していたため，構文的な曖昧性の多い複文などに対して処理時間が増大するという実時間性の問題もあった．

本レポートでは，これらの問題を解決するために，表層パタンのみを用いた統一的な枠組で，パタンの記述や照合，入力文の修正を行なう構成素境界解析 (constituent boundary

parsing) を提案し，構成素境界解析を導入した新しい TDMT が多言語話し言葉翻訳 (Furuse, Kawai, Iida, Akamine, and Kim 1995; 山本, 古瀬, 飯田 1996; Sumita, Yamada, Yamamoto, Paul, Kashioka, Ishikawa and Shirai 1999) に対して有効な手法であることを評価実験結果により示す．また，構成素境界解析では，チャート法に基づくアルゴリズムで逐次的 (left-to-right) に入力文の語を読み込んで，解析途中で候補を絞り込みながらボトムアップに構文構造を作り上げることにより，効率的な構文解析が行なえることも示す．現在は，「国際会議に関する問い合わせ会話」よりも場面状況が多様である「旅行会話」を翻訳対象とし，日英双方向，日韓双方向などの多言語話し言葉翻訳システムを構築している．システムは，構成素境界解析と用例利用型処理を組み合わせた新しい TDMT の枠組により，多様な表現の旅行会話文を話し手の意図が理解可能な結果へ実時間で翻訳することができる．

パタンや用例を利用する頑健な翻訳手法として，原言語と目的言語の CFG 規則を対応させたパタンを入力文に照合させる手法 (渡辺・武田 1998)，詳細な構文意味規則を利用する翻訳を併用する手法なども提案されている (Brown 1996; 加藤 1995; 白井, 松島, 井上, 松尾, 矢部, 内野 1997)．前者は，表層語句だけでなく細かい属性を使ってパタンを記述することがあり，パタンの記述は必ずしも容易でない．また，解析中で競合する CFG 規則が多くなり処理時間が増大しやすい．後者は，入力文がパタンや用例にヒットすれば高品質の翻訳結果を得られるが，多様な入力文に対して高いヒット率を実現するのは容易ではない．また，多言語翻訳へ展開する際に，様々な言語ペアの翻訳に対して詳細な構文意味規則をそれぞれ用意するのも容易でない．これらの手法に比べて，TDMT は，表層パタンのみの照合を行なうので，実時間性の点で有利である．パタンの記述も容易であり，パタンを組み合わせることにより，他の翻訳手法を併用しなくても多様な入力文に対応でき，頑健性においても，多言語翻訳を実現する汎用性においても有利である．

以下，2 節で構成素境界解析と用例利用型処理を組み合わせた TDMT の枠組，3 節でパタンによる構文構造の記述，4 節で構成素境界解析による構文構造の導出，5 節で用例利用型処理による最尤の原言語構文構造の決定法と目的言語への変換，6 節で解析途中での構文構造候補の絞り込み，について説明し，7 節で日英双方向と日韓双方向の話し言葉翻訳の評価実験結果により，本レポートで提案する TDMT の有効性を示す．

なお，本レポートは文献 (古瀬, 山本, 山田 1999) に基づいて作成したものである．

第 2 章

TDMT の枠組

TDMT は，単純な表層パタンと用例で記述した変換知識の情報を用いて構成素境界解析と用例利用型処理を行なう．構成素境界解析と用例利用型処理は構文解析や変換などを行なう TDMT の中心的処理である．本節では，変換知識について説明したあと，TDMT の翻訳処理の概要について述べる．

2.1 変換知識

変換知識は，3節で説明する構成素境界パタンにより表した原言語表現が，用例を訳し分け条件としてどのような目的言語表現に対応するかを記述する．変換知識の作成は，原言語パタンごとに，システムが翻訳できるようなデータ形式に翻訳例を加工して行なう (この作業を以下，翻訳訓練と呼ぶ)．例えば，「京都に来てください」→“*Please come to Kyoto*” という翻訳例の原言語部分から，「X てください」と「X に Y」という原言語パタンを抽出し，それぞれの原言語パタンについて変換知識を作成する．「X に Y」では，X と Y の具体的な語の組 (京都, 来る) に対して目的言語パタンは “Y′ *to* X′” になるという以下のような日英の変換知識を作る [1]．X′ は X の対訳を示す．

$$
\begin{aligned}
\text{X に Y} \quad \Rightarrow \quad & \text{Y}' \textit{ to } \text{X}' \text{ ((京都, 来る),(空港, 行く)...)}, \\
& \text{Y}' \textit{ at } \text{X}' \text{ ((三時, 来る), ...)}, \\
& \vdots
\end{aligned}
$$

この変換知識は，「空港に行く」→“*go to the airport*” や「三時に来る」→“*come at three o'clock*” のような翻訳例の翻訳訓練結果も含んでいる．

TDMT では，変換知識の原言語パタンを用いて，入力文に適合する原言語構文構造の候補を作る．また，変換知識の用例と目的言語パタンを用いて，最尤原言語構文構造とその変換結果である最尤目的言語構文構造を決定する．なお，原言語パタンには，意味距離計算の対象となる入力文中の語を決めるために主部 (head) となる部分がどこであるかという

[1] 本レポートでは，X や Y のようなパタンの変項を具体化する語の組で，変換知識の中で訳し分け条件として記述されているものを用例と呼ぶ．2.1節の「X に Y」に関する変換知識の例では (京都, 来る) や (空港, 行く) が用例である．

情報を与える（5.1節参照）．目的言語パタンには，生成処理を助けるための情報を与える（5.2節参照）．

2.2 翻訳処理の概要

本レポートで提案する TDMT の構成を図 2.1に示す．

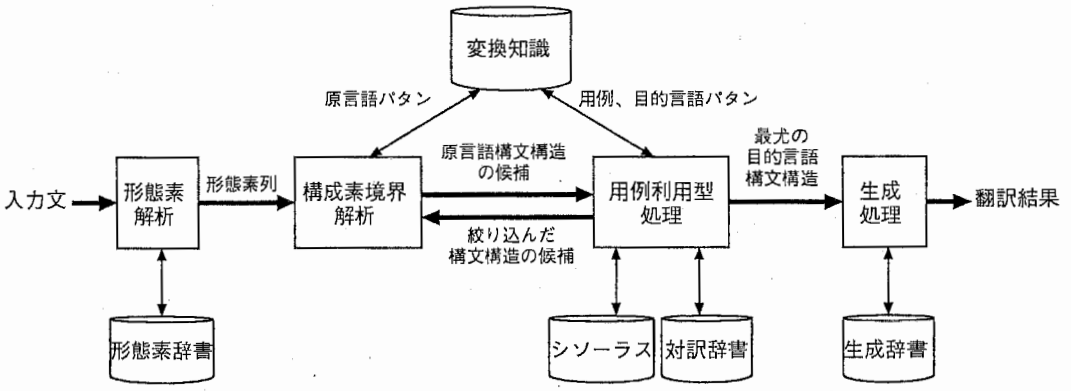

図 2.1: TDMT の構成

入力文を形態素解析した後，構成素境界解析では，変換知識の原言語パタンを組み合わせて入力文に適合する原言語構文構造の候補を作る．

用例利用型処理では，構成素境界解析から送られた原言語構文構造候補の各パタンごとに，意味距離計算の対象となる入力文中の語の組に対して，変換知識の各用例との意味距離をシソーラスを参照して計算する．最小の意味距離を与える用例を類似用例と定義する．この類似用例の意味距離を元に構文構造のスコアを求め，最尤の原言語構文構造を決定する．

構成素境界解析の途中で入力文の部分に対する構文構造候補ができた場合，用例利用型処理で構文構造のスコアを計算して候補を絞り込みながら，構成素境界解析を進めていく．構成素境界解析は入力文全体の原言語構文構造の候補を最終的に出力し，この候補の中から用例利用型処理で最尤のものを決定する．

用例利用型処理では，さらに，入力文全体の最尤の原言語構文構造について，構造を構成する各パタンからは変換知識の中で類似用例が与える目的言語パタンへ変換し，構造の終端の語句からは対訳辞書の中で対応する目的言語の語句へ変換して，最尤の目的言語構文構造を作る [2]．最後に，生成処理で，生成辞書を参照するなどして，最尤の目的言語構文構造から翻訳結果を出力する．

TDMT は，表層パタンの照合による構成素境界解析と用例利用型処理を組み合わせたことにより多言語話し言葉翻訳システムを構築する上で，以下のような利点を持つ．

[2]原言語構文構造を作りながら目的言語構文構造へ変換することも可能であるが，現在は，翻訳処理の省力化のため，枝刈りされる原言語構文構造についての変換・生成は行なっていない．翻訳入力の途中で部分的な翻訳結果を出力する同時翻訳機構では，構文解析と変換・生成を同時に行なう必要がある．

- 多様な話し言葉表現の構文構造を単純なパタンの組み合わせにより記述できる．（頑健性）
- 構文構造が単純であり，解析途中で候補を絞り込みながら構文構造を作り上げることにより，効率的な構文解析ができる．（実時間性）
- 変換知識の記述が容易であり，構成素境界解析と用例利用型処理は単純で言語に依存しない手法なので，様々な言語ペアの翻訳に対応できる．（汎用性）

第 3 章

構成素境界パタン

構成素境界パタンは，変項と構成素境界により成り，文や名詞句など意味的にまとまった語句の構文構造を表す (Furuse and Iida 1994)．変項は，X や Y などの記号により表し，構成素に対応する．構成素として変項を具体化するのは，内容語と，構成素境界パタンに照合する語句である．構成素境界は，機能語または品詞バイグラムマーカにより表し，構成素を関係づけたり修飾したりする．構成素境界解析では，構成素境界をキーにして構文構造を作っていくため（4.2節参照），構成素境界のない「X Y」のような二つの変項が隣接するパタンは認めず[1]，構成素の間には必ず構成素境界を置く．

以下，本節では，構成素境界パタンを用いた構文構造の記述方法について説明する．

3.1 構成素境界としての機能語

構成素境界パタンの中で構成素境界を表す表層語句は原則として機能語であり，内容語は構成素となるので構成素境界には使用しない．この制限により，パタンの種類が膨大になるのを防ぐことができる．

英語の前置詞，日本語や韓国語の助詞は頻出の機能語であり構成素境界となる．例えば，英語語句 "*go to Kyoto*" において，前置詞 "*to*" が構成素境界として二つの構成素 "*go*" と "*Kyoto*" の間にあると考え，構成素境界パタン "X *to* Y" を用いて図 3.1の (a) のように構文構造を記述する．日本語語句「こちらは観光局」においても，機能語「は」が構成素境界として二つの構成素「こちら」と「観光局」の間にあり，構成素境界パタン「X は Y」を用いて図 3.1の (b) のように構文構造を記述する．

3.2 構成素境界としての品詞バイグラムマーカ

3.2.1 品詞バイグラムマーカの挿入

[1] 目的言語表現は，構成素境界パタンでない「X′ Y′」のようなパタンでも表すことができる．構成素境界は構文解析のために使い，生成では必ずしも必要としない．

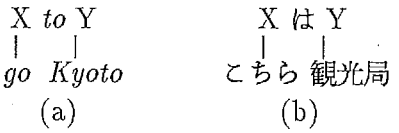

図 3.1: 構文構造（機能語が構成素境界）

英語語句“*I go*”は，二つの構成素“*I*”と“*go*”より成る．しかし，この二つの構成素の間には表層語句は存在しない．このような場合，形態素解析で品詞が確定した後に，品詞バイグラムマーカを二つの構成素の間に挿入する．前方の構成素の最後の語の品詞を A，後続する構成素の先頭の語の品詞を B とすると，<A-B> を品詞バイグラムマーカと定義する．本レポートでは，A と B を品詞の英語名で表すことにする．

連接する品詞 A と品詞 B の間に品詞バイグラムマーカ <A-B> を構成素境界として挿入する条件を以下に示す．

[1]A も B も，前後の構成素を関係づける格助詞や前置詞のような品詞でない．

[2]A が，後続の構成素を修飾する連体詞や冠詞のような品詞でない．

[3]B が，前にある構成素を修飾する日本語や韓国語の助動詞や接尾語のような品詞でない．

例えば，「こちら <pronoun-particle> は <particle-noun> 観光局」や“*go* <verb-preposition> *to* <preposition-propernoun> *Kyoto*”のような品詞バイグラムマーカの挿入は [1] の条件に抵触するので認めない．「その」と「ホテル」の間や“*the*”と“*bus*”の間は [2] に，「行き」と「ます」の間や「鈴木」と「さん」の間は [3] に，それぞれ抵触するので品詞バイグラムマーカは挿入しない．品詞バイグラムマーカは，本節の条件により機械的に挿入することができ，単語名でなく品詞名を使うので種類を限定することができる [2]．

3.2.2 品詞バイグラムマーカを用いた構文構造記述

英語語句“*I go*”の“*I*”と“*go*”はそれぞれ代名詞と一般動詞であり，<pronoun-verb> を構成素境界としてそれらの間に挿入する．この結果，“*I go*”は“*I* <pronoun-verb> *go*”に修正され，パタン“X <pronoun-verb> Y”に照合可能になる．従って，“*I go*”の構造は図 3.2の (a) のように記述できる．

[2]用例利用型処理（5節参照）により高精度の構文解析や変換を実現するためには変換知識の各パタンにできるだけ多くの用例を付与することが望ましい．そこで，品詞バイグラムマーカを「X<*-*>Y」のように一本化して用例を集約することも考えられる．しかし，英語のパタン“X <pronoun-verb> Y”が照合するのは“*I* <pronoun-verb> *go*”のような単文レベルの表現にほぼ限定されるというように，品詞バイグラムマーカの挿入位置が構成素境界パタンの構造レベル（3.4節参照）に関係する場合があるので，現在はマーカを挿入位置の前後の品詞で区別している．

また，日本語の話し言葉では，「こちら観光局」のように助詞がしばしば省略される．この語句は「こちら」と「観光局」という二つの構成素より成る．「こちら」は代名詞，「観光局」は普通名詞なので，<pronoun-noun> を構成素境界として「こちら」と「観光局」の間に挿入する．修正された「こちら <pronoun-noun> 観光局」は「X<pronoun-noun>Y」に照合可能になる．品詞バイグラムマーカ <pronoun-noun> が「は」と同様の働きをすることにより，助詞が脱落していない「こちらは観光局」と同様の構造を助詞脱落表現についても図 3.2の (b) のように記述することができる．

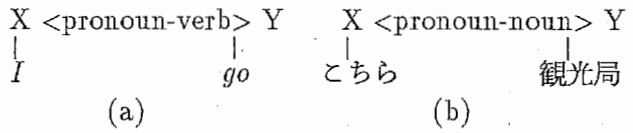

図 3.2: 構文構造（品詞バイグラムマーカが構成素境界）

文献 (古瀬他 1994) の TDMT は，「空港バス」や「会場入り口」など複数の名詞が連続した複合名詞を “$NOUN_1$ $NOUN_2$” という品詞列パタンにより表し，「X の Y」のような表層パタンの場合とは異なる照合のメカニズムを使っていた．品詞バイグラムマーカの導入により，複数の名詞が連続した複合名詞も，「X<noun-noun>Y」のように構成素境界パタンで表現でき，表層パタンの照合のみで構文解析を行なうことが可能になった．すなわち，「空港バス」のような複合名詞も「こちら観光局」のような助詞脱落表現も，品詞バイグラムマーカにより構文構造を記述できる．本レポートで提案する品詞バイグラムマーカの挿入により，助詞脱落表現に具体的な助詞を補完する手法 (古瀬他 1994) で生じた，補完する助詞を誤る，補完すべきでない時に助詞を補完する，などの問題を解決することができる．

3.3 多様な構成素境界パタン

変項の間に必ず構成素境界を置けば，「その X」，「X に Y」，「X から Y まで Z」など，変項の数に制限なく構成素境界パタンを作ることができる．

また，「明日までに行く」という語句では機能語である助詞「まで」と「に」が連続している．「まで」と「に」の間は 3.2.1節の条件 [1] に抵触するので品詞バイグラムマーカを挿入する必要はなく，機能語を連続させて構成素境界とした「X までに Y」のような構成素境界パタンを作ることができる．

3.4 構成素境界パタンの組み合わせ

構成素境界パタンに照合する語句は，構成素として別の構成素境界パタンの変項を具体化することができる．すなわち，構成素境界パタンを組み合わせることにより構文構造を作ることができる．任意の長さの複合名詞も，「X<noun-noun>Y」のような構成素境界パタンの組み合わせにより構文構造を記述できる．

しかし，パタンの組み合わせ方によってはありえない構文構造ができるので，構文解析の品質や効率を上げるためにこのような構造を排除する必要がある．このため，本レポートでは，パタンを構造レベルによって分類し，各構造レベルのパタンの変項を具体化できる語句について，そのサブ構造レベルと品詞を表 3.1のようにあらかじめ指定する [3]．これにより，パタンの組み合わせ方を制限し，ありえない構文構造を排除することができる．

表 3.1: 構造レベルの関係

構造レベル	変項を具体化できるサブ構造レベルと品詞
複文，重文	複文, 重文, 単文, 動詞句, ...
単文	動詞句, 名詞句, 複合名詞, ...
動詞句	動詞句, 名詞句, 複合名詞, 一般動詞, ...
名詞句	名詞句, 複合名詞, 普通名詞, 固有名詞, ...
複合名詞	複合名詞, 普通名詞, ...

例えば，“*I go to Kyoto*”という文の構文構造は，“X <pronoun-verb> Y”と“X *to* Y”というパタンの組み合わせになる．“*I go to Kyoto*”の正しい構造は図 3.3の (a) であり，(b) の構造は排除しなくてはならない．“X <pronoun-verb> Y”を単文レベル，“X *to* Y”を動詞句レベルのパタンに指定し，表 3.1のように動詞句の下部構造を制限すれば，“X <pronoun-verb> Y”は“X *to* Y”の下部構造とはなりえないので，(b) の構造は排除される．

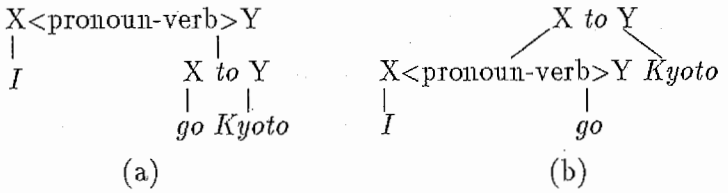

図 3.3: *I go to Kyoto* の構文構造

[3]これは各構造レベルで，変項すべてについて適用される緩やかな制限である．英語のパタン“X <pronoun-verb> Y”の X は名詞性のパタンや語でしか具体化できないなど，より厳しい制限を特定の変項についてローカルに与えることもできる．

第 4 章

構成素境界解析

構成素境界解析は，相互情報量を用いて再帰的に構成素境界を検知して構文構造を求める頑健な構文解析手法としても提案されているが (Margerman and Marcus 1990)，統計処理への依存が強く，文法情報をほとんど利用しないため解析精度に問題があった．本レポートで提案する構成素境界解析は，意味的にまとまった語句について構成素境界パタンを作り，構造レベルでパタンを分類するなど，単純で緩やかな文法制約を与えることにより高精度の構文解析を可能にする．

本節では，チャート法に基づくアルゴリズムで，逐次的に入力文の語を読み込んでボトムアップに構文構造を作り上げる構成素境界解析について説明する．

4.1 活性弧と不活性弧

チャート法は活性弧と不活性弧を組み合わせることにより入力文の構文構造を作る．図 4.1の (a) のような内容語による構造，構成素境界パタンのすべての変項が具体化された (b) と (c) のような構造は不活性弧に対応する．⇑ は，構文解析で読み込み中の語を指す走査カーソルである．入力文の構文構造は，入力文全体をカバーする不活性弧に対応する．構成素境界パタン中に具体化されていない変項がある図 4.2の (d) と (e) のような構造は活性弧に対応する．

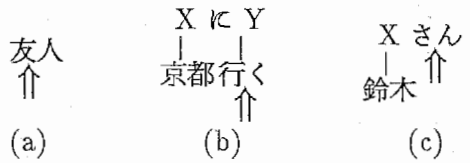

図 4.1: 不活性弧に対応する構造

チャート法は，部分的な構文解析結果を弧で表すことにより同じ解析を繰り返すのを回避し，効率的な構文解析を行なう．さらに，構成素境界パタンを使ったチャート法の構文解析では，表層をキーとして弧を張っていくので，競合する構成素境界パタンが少ない．従って，張られる弧の数も少ないため，処理時間をより一層抑えることができる．

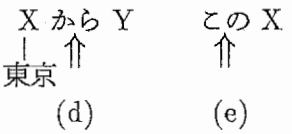

図 4.2: 活性弧に対応する不完全な構造

4.2 構文解析

「友人とハワイに来週行きます」という日英翻訳の入力文を例にとって，TDMT の構成素境界解析を説明する．

まず，形態素解析により入力文の各語の品詞を次のように決定する．

友人	と	ハワイ	に	来週	行き	ます
普通名詞	助詞	固有名詞	助詞	普通名詞	動詞	助動詞

3.2.1節の条件を満たす品詞バイグラムマーカは，普通名詞と動詞の間の <noun-verb> のみであり，入力文は「友人とハワイに来週 <noun-verb> 行きます」に修正される．修正された入力文に対し，以下のアルゴリズムに従って，逐次的にボトムアップの構成素境界解析を行なう．以下では，語と語の間に節点を置き，左から k 番目の語の左隣には節点 $k-1$ が，右隣には節点 k があるものとする．弧は節点から節点に張るものとする．

(0) 先頭の語に走査カーソルを設定し，$k := 1$ として，(i) へ．

(i) 走査カーソルの指す語が名詞や動詞などの内容語であれば，節点 $k-1$ から節点 k に不活性弧を張り，(iii) へ．そうでなければ，(ii) へ．

(ii) 走査カーソルの指す語が構成素境界 α_1 であれば，構成素境界から構成素境界パタンへの対応表を参照することにより，構成素境界パタンを検索する．検索されたすべてのパタンについて，その形式に応じて (ii.a)~(ii.e) のいずれかの処理を行なったうえで，(iii) へ．パタンが検索できなければ，(iv) へ．

(ii.a) 「$X\alpha_1Y$」，「$X\alpha_1Y\alpha_2Z$」，「$X\alpha_1\alpha_2Y$」のように，α_1 の左が変項一つのみであり，α_1 が右端でないパタンが検索された場合，そのパタンの α_1 の左隣の変項を，節点 j から節点 $k-1$ （ただし，$j < k-1$）に張られた不活性弧で具体化できれば，検索されたパタンに関する活性弧を節点 j から節点 k に張る．

(ii.b) 「$X\alpha_1$」のように，α_1 の左が変項一つのみであり，α_1 が右端であるパタンが検索された場合，そのパタンの α_1 の左の変項を，節点 j から節点 $k-1$ に張られた不活性弧で具体化できれば，検索されたパタンに関する不活性弧を節点 j から節点 k に張る．

(ii.c) 「α_1X」，「α_1Xα_2」のように，α_1 が左端であるパタンが検索された場合，検索されたパタンに関する活性弧を節点 $k-1$ から節点 k に張る．

(ii.d) 「Xα_0Yα_1Z」，「X$\alpha_0\alpha_1$Y」のように，α_1 の左に別の構成素境界があり，α_1 が右端でないパタンが検索された場合，検索されたパタンに関する活性弧が，α_1 より左のみ具体化されて節点 j から節点 $k-1$ に張られていれば，検索されたパタンに関する活性弧を節点 j から節点 k に張る．

(ii.e) 「α_0Xα_1」のように，α_1 の左に別の構成素境界があり α_1 が右端であるパタンが検索された場合，検索されたパタンに関する活性弧が，α_1 より左が具体化されて節点 j から節点 $k-1$ に張られていれば，検索されたパタンに関する不活性弧を節点 j から節点 k に張る．

(iii) 節点 i から節点 k（ただし，$i < k$）に新しく張られた不活性弧が，節点 h から節点 i（ただし，$h < i$）に張られた活性弧を構成するパタンの中のまだ具体化されていない最左の変項を具体化できれば，さらに節点 h から節点 k に新しい不活性弧または活性弧を張る．新しい弧が張れなくなるまでこの操作を繰り返し，(iv) へ．

(iv) 走査カーソルの指す語が入力文の最後の語であれば，解析終了．そうでなければ，走査カーソルを右へ一語移動させ，$k := k+1$ として，(i) へ．

(ii) で参照する対応表は，システムが持つ変換知識の原言語パタンからあらかじめ機械的に作成しておく．例文の構成素境界解析において検索される構成素境界パタンを表 4.1に示す．

表 4.1: 構成素境界パタンの検索

構成素境界	構成素境界パタン	(パタンの構造レベル)
と	X と Y	(名詞句，動詞句)
に	X に Y	(動詞句)
<noun-verb>	X<noun-verb>Y	(動詞句)
ます	X ます	(単文)

図 4.3は入力文に対して弧が張られていく過程を示すチャートである．実線は不活性弧を，点線は活性弧を示し，弧のできる順序を示す番号により弧を識別する．

先頭の語「友人」は内容語であり，不活性弧 (1) を張る．次の語「と」により「X と Y」の X を (1) で具体化させた活性弧 (2) と (3) を張る．「X と Y」は (2) では動詞句のパタン，(3) では名詞句のパタンである．「ハワイ」により不活性弧 (4) を張る．(3) の「X と Y」の Y を (4) で具体化し，不活性弧 (5) を張る．次の語「に」から検索された動詞句パタン「X に Y」の X を (4) と (5) でそれぞれ具体化し，活性弧 (6) と (7) を張る．「来週」

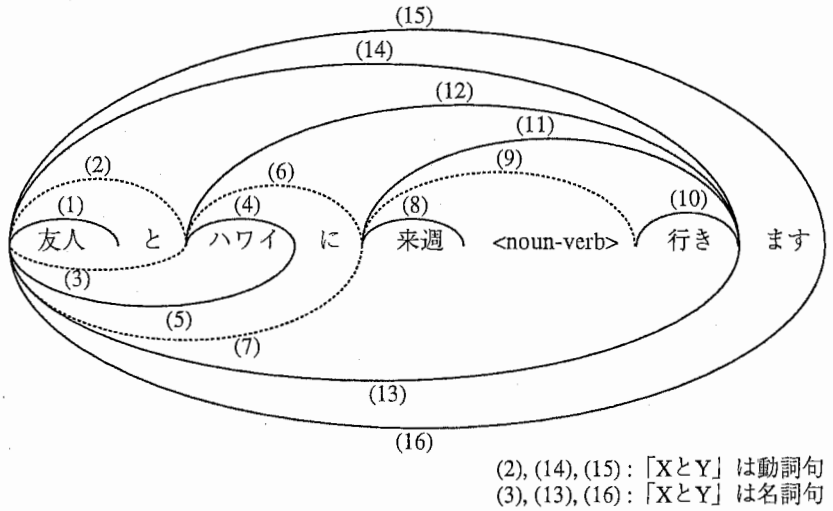

図 4.3: 構文解析の過程を示すチャート

により不活性弧 (8) を張る．<noun-verb> から検索された「X<noun-verb>Y」の X を (8) で具体化し，活性弧 (9) を張る．「行き」により不活性弧 (10) を張り，(9) の「X<noun-verb>Y」の Y を (10) で具体化し，不活性弧 (11) を張る．(6) と (7) の「X に Y」の Y を (11) で具体化し，それぞれ不活性弧 (12) と (13) を張る．さらに，(2) の「X と Y」の Y を (12) で具体化し，不活性弧 (14) を張る．

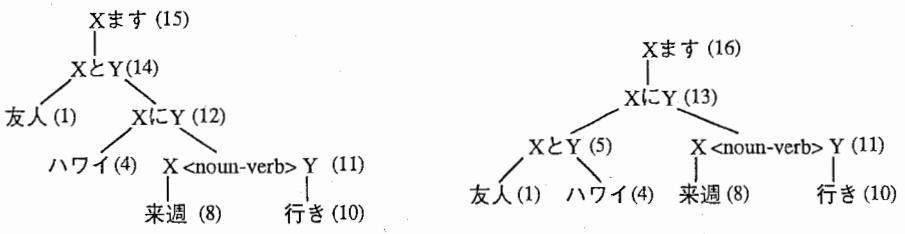

図 4.4: 入力文の構文構造の候補

入力文の最後の語「ます」から検索された「X ます」の X を (14) と (13) で具体化し，それぞれ不活性弧 (15) と (16) を張る．すべての語を読み込み終えて，これ以上新たな弧が張れない状態になり，解析は終了する．入力文全体をカバーする不活性弧 (15) と (16) が入力文の構文構造の候補に対応する．図 4.4に入力文の構文構造の候補を示す．パタンに付随する番号は，そのパタンを最上部とする構文構造が対応する不活性弧を示す．

第 5 章

用例利用型処理

本節では，構成素境界解析で得られた入力文の構文構造の候補から，意味距離計算によって最尤の目的言語構文構造を決定する用例利用型処理について，4.2節の例文を使って説明する．

5.1 意味距離計算

現在，TDMT では，シソーラス上での意味属性の位置関係により単語間に 0~1 の意味距離を与え (Sumita and Iida 1992)，構成素境界パタンに関する意味距離を，各変項についての単語間の意味距離の合計値としている．不活性弧 (11) を構成する「X<noun-verb>Y」では，X と Y を具体化する語の組 (来週, 行く) を意味距離計算の対象として [1]，「X<noun-verb>Y」に関する変換知識の用例との意味距離を計算する．例えば，(来週, 行く) と用例 (明日，来る) の間の意味距離は，「来週」と「明日」の間の意味距離と，「行く」と「来る」の間の意味距離の合計値である．

構成素境界パタンに照合する語句が上部の構成素境界パタンの変項を具体化している場合，主部の語の組を対象として用例との意味距離を計算する．構成素境界パタンで主部となる部分の情報は変換知識に記述しておき，主部は下部の構造から上部の構造へ伝搬するという性質を利用して，主部の語を機械的に求めることができる．不活性弧 (12) では「X に Y」の X と Y を，「ハワイ」と「来週 <noun-verb> 行き」でそれぞれ具体化する．「X<noun-verb>Y」では Y を主部に定めているとすると，「来週 <noun-verb> 行き」の主部は「行き」であり，不活性弧 (12) の「X に Y」に関する意味距離計算の対象は (ハワイ, 行く) となる．

5.2 最尤原言語構文構造の決定と目的言語への変換

[1] 意味距離計算は表記形「行き」でなく標準形「行く」に対して行なう．

用例利用型処理では，構文構造を構成する各構成素境界パタンについて類似用例を変換知識の中から求める．類似用例の与える情報により，最尤の原言語構文構造を決定し，その構造を目的言語に変換して，最尤の目的言語構文構造を得る．不活性弧 (15) と (16) に対応する構文構造を構成する構成素境界パタンについて，意味距離計算の結果を表 5.1のように仮定する．

表 5.1: 意味距離計算の結果

構文構造の最上部のパタン (太字は主部)	対応する不活性弧	意味距離計算の対象	意味距離計算の結果		
			類似用例	目的言語パタン	意味距離
X と **Y** (動詞句)	(14)	(友人, 行く)	(社長, 行く)	Y′ *with* X′	0.34
X と **Y** (名詞句)	(5)	(友人, ハワイ)	(京都, 奈良)	X′ *and* Y′	1.01
X に **Y**	(12),(13)	(ハワイ, 行く)	(京都, 行く)	Y′ *to* X′	0.18
X<noun-verb>**Y**	(11)	(来週, 行く)	(明日, 来る)	Y′ X′	0.12
X ます	(15),(16)	(行く)	(行く)	*I will* X′	0.00

類似用例が与える意味距離を，構文構造を構成する構成素境界パタンについてすべて合計した値を，構文構造のスコアと定義し，このスコアが最小のものを最尤の構文構造とする (古瀬他 1994)．不活性弧 (15) に対応する構文構造では，「X と Y」(動詞句)，「X に Y」，「X<noun-verb>Y」，「X ます」で類似用例が与える意味距離，0.34，0.18，0.12，0.00 を合計した 0.64 がスコアとなる．不活性弧 (16) に対応する構文構造では，「X と Y」(名詞句)，「X に Y」，「X<noun-verb>Y」，「X ます」で類似用例が与える意味距離 1.01，0.18，0.12，0.00 を合計した 1.31 がスコアとなる．従って，不活性弧 (15) に対応する構文構造が最小のスコアを持ち，入力文全体についての最尤の原言語構文構造となる．

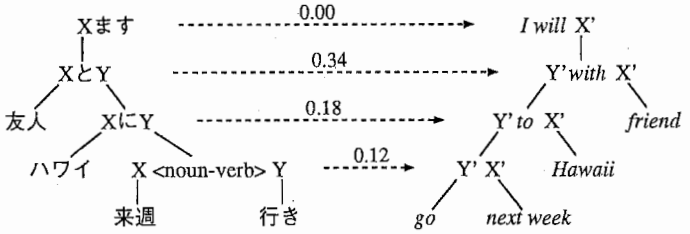

図 5.1: 最尤原言語構文構造の変換

最尤の原言語構文構造の各構成素境界パタンを，変換知識の中で類似用例が訳し分け条件となって与える目的言語パタンへと変換することにより，最尤の目的言語構文構造を作る．不活性弧 (15) に対応する構文構造では，各構成素境界パタンは表 5.1の 5 列目に示す目的言語パタンに変換される．内容語の「友人」，「ハワイ」，「来週」，「行き」は，対訳

辞書を参照して“*friend*”, “*Hawaii*”, “*next week*”, “*go*” にそれぞれ変換され[2]，図5.1に示す目的言語構文構造ができる．矢印の上の数字は，各構成素境界パタンのスコアである．

用例利用型処理で得られた最尤の目的言語構文構造は，原言語構文構造の性質を受け継いでいるため，そのまま線条化すると，“*I will go next week to Hawaii with the friend*” となってしまう．そこで，“Y′ X′” の “X′” は時間格，“Y′ *to* X′” の “*to* X′” は場所格，というような情報をあらかじめ変換知識の目的言語パタンに与えておいたうえで，語順や活用などの調整を生成処理で行ない，以下のような英語文を出力する．

“*I will go to Hawaii with the friend next week*”

[2]TDMT システムは，内容語に対して，対訳辞書に記述されたデフォルトの対訳語句を与えているが，意味距離計算の結果の類似用例によってはデフォルト以外の対訳語句を与えている (古瀬他 1994; 山田, 山本, 飯田 1998).

第 6 章

解析途中での構文構造候補の絞り込み

意味距離計算により構文構造のスコアを得るためには，下部の構造での主部の語を確定させ意味距離計算の対象を決定する必要がある．不活性弧は，構文構造を構成する構成素境界パタンのすべての変項が具体化されている構造であり，構成素境界パタンの主部の語をすべて求めることができるので意味距離計算の対象が決定し，構文構造のスコアが得られる．TDMT では，処理時間を短縮するために，入力文の同じ部分に対して作られる不活性弧をスコアにより順位づけし，上位 n 個 (n-best) の不活性弧のみを保持して構文解析を進めていく．すなわち，解析途中で構文構造候補の絞り込みを行なう．保持した不活性弧にはスコアと主部の情報を与え，上部の構造で意味距離計算による構文構造候補の絞り込みが容易にできるようにする．意味距離計算により解析途中で構文構造候補を絞り込むには，4.2節のアルゴリズムのようなボトムアップの解析が必要である．

TDMT システムは現在，1-best をデフォルトとして解析途中での構文構造候補の絞り込みを行なっているが，n の値は容易に変更可能である．例えば，4.2節の入力文の解析の途中で，「友人とハワイに来週 <noun-verb> 行き」に対して，二つの不活性弧 (13) と (14) ができる．保持する不活性弧を 1-best にして構文解析を行なうと，スコアの良い (14) のみが「友人とハワイに来週 <noun-verb> 行き」について保持され，「X ます」の X を (14) で具体化した (15) のみが入力文の構文構造に対応する．(13) は途中で枝刈りされるので (16) に対応する構文構造は作られない．

第 7 章

多言語話し言葉翻訳の評価実験

本節では，構成素境界解析と用例利用型処理を組み合わせたTDMTシステムに対する，日英双方向と日韓双方向の話し言葉翻訳の評価実験結果について述べる．

7.1 言語データベースからのシステムデータ構築

TDMTシステムの翻訳対象は，話し言葉翻訳を使用する場面を想定した「旅行会話」とし，TDMTシステムの翻訳訓練文と評価文を選定するために，ホテルの予約，ホテルの紹介，ホテルでのサービス，乗物の切符購入，道案内，交通手段の問い合わせ，観光ツアーの案内など旅行会話全般のトピックに渡る言語データベースを構築した (Furuse, Sobashima, Takezawa, and Uratani 1994)．この言語データベースは，通訳を介したバイリンガル模擬会話，基本表現を網羅するために机上で作成した対訳表現集より成る．この言語データベースに形態素のタグづけを行なうことによりTDMTシステムの形態素辞書を構築している．表 7.1は，TDMTシステムの主要データである形態素辞書と変換知識について，評価実験時の規模を翻訳訓練文の概要とともに示す．

表 7.1: TDMTシステムの規模

	日英	日韓	英日	韓日
形態素辞書の語彙数 (概算)	13000		8000	4000
翻訳訓練文数 (異なり)	2932	1543	2865	613
翻訳訓練文の平均語数 (異なり)	10.0	9.5	8.5	8.0
変換知識のパタンの種類	776	591	1177	330

7.2 評価実験の内容

旅行会話での多言語話し言葉翻訳のシステム性能を把握するために，言語データベースの中でTDMTシステムが翻訳訓練していないバイリンガル模擬会話から，評価文を無作為

表 7.2: 評価文（ブラインドテスト）

	日英，日韓	英日	韓日
のべ文数	1225 (9.7 語 / 文)	1341 (7.1)	1174 (8.1)
異なり文数	1001(11.4 語 / 文)	1019 (8.8)	1004 (9.1)

抽出して，評価実験を行なった．評価文は，表 7.2に示すように，各言語ペアの翻訳で異なり 1000 文以上である．日本語を入力とする日英と日韓の翻訳については，比較検討のため，同じ文を使って評価実験を行なった．

7.2.1 評価項目

評価項目は，翻訳品質，構文解析，処理時間である．

翻訳品質に関しては，複数の尺度で採点する方法 (長尾・辻井 1985) や，様々な言語現象を含む評価文の翻訳結果が評価項目をクリアしているかどうかを調べ，システム改良の参考データを求める方法 (池原, 白井, 小倉 1994) などが提案されている．ただし，これらはほとんど日英間の書き言葉翻訳を対象としており，多言語話し言葉翻訳についての評価方法は提案されていない．筆者らは，話し言葉翻訳という性格上，どのような言語的な誤りがあったかよりも，話し手の言いたいことが聞き手にどの程度伝わったかという観点が重要であると考え，システム開発者よりもシステム使用者の視点に立って翻訳品質を評価した．以下の翻訳成功率を設定し，各言語ペアの翻訳について，原言語に堪能な目的言語のネイティブ話者３名が採点した結果の平均値を求めた．

翻訳成功率Ａ:
話し手の言いたいことのすべてが問題なく聞き手に伝わっている
「**問題なし**」と判定された文の割合

翻訳成功率Ｂ:
話し手の言いたいことの最低限必要な内容が聞き手に伝わっている
「**理解可能**」と判定された文の割合

構成素境界解析による構文解析結果の評価では，入力文全体の構造を正しく解析できていれば成功，一部でも誤った構造になっていれば失敗と判定し，構文解析成功率を求めた．

処理時間は，Common Lisp で記述したプログラムをコンパイルした TDMT システムについて，SPARCstation10 上で計測した．

7.2.2 評価実験の前提

評価実験は以下の前提で行なった．

- システムへの入力は，文字列でなく正解形態素列とし，形態素解析の性能（山本，河井，隅田，古瀬 1997）と TDMT の性能を独立に評価することにした．処理時間も形態素解析の時間を除いて計測した．また，話し言葉翻訳という前提を考慮して，音声として現れない句読点，コンマ，ピリオドなどは入力に含めなかった．

- 入力文で同じ部分に対して保持する不活性弧を 1-best にして構文構造候補を絞り込みながら構成素境界解析を行なった．これは，構文構造を多く保持しても，表 7.3に示すように翻訳結果が変わるのは少数であり，翻訳結果が変わって品質が向上したのはごく少数だったという予備実験の結果による．

表 7.3: 構文構造候補の絞り込みの影響

		1-best	5-best	10-best
1-best の時と比較した	日英	0 %	5.9 %	6.1 %
翻訳結果の差分割合（のべ）	英日	0 %	5.0 %	5.1 %
全評価文の平均処理時間	日英	0.52 秒	0.70 秒	0.81 秒
（形態素解析の時間を除く）	英日	0.30 秒	0.48 秒	0.66 秒

7.3 評価実験結果

表 7.4に，全評価文に対する翻訳成功率と構文解析成功率を示す．

表 7.4: 翻訳成功率と構文解析成功率（全評価文）

		日英	日韓	英日	韓日
翻訳訓練文数	異なり	2932	1543	2865	613
翻訳成功率 A	のべ	45.3 %	60.4 %	43.4 %	47.4 %
（問題なし）	異なり	34.2 %	51.7 %	35.0 %	39.9 %
翻訳成功率 B	のべ	78.5 %	93.0 %	83.8 %	92.2 %
（理解可能）	異なり	73.9 %	91.5 %	81.2 %	91.1 %
構文解析成功率	のべ	77.8 %	70.5 %	74.6 %	60.0 %
	異なり	72.8 %	63.9 %	66.6 %	53.4 %

どの言語ペアの翻訳についても翻訳成功率 B は高く，TDMT システムが，話し手の意図が理解可能なレベルの多言語翻訳を多くの旅行会話文に対して実現していることが示された．TDMT システムの日英，日韓，英日，韓日の翻訳について，評価文に対する翻訳実行例を付録に示す．

日韓と韓日については，翻訳訓練文が少ないにもかかわらず特に高い翻訳成功率を達成している．構文解析成功率は，翻訳訓練文数が多い日英と英日が高く，訓練文数が最小だった韓日が最も低い．

図 7.1は，翻訳に要したCPU timeを各形態素数ごとに平均した値により処理時間を示す．形態素解析の時間は含めていない．翻訳訓練文数が多い日英と英日は他の翻訳に比べて処理時間が少し長いが，いずれの言語ペアの翻訳でも実時間の処理を実現している．

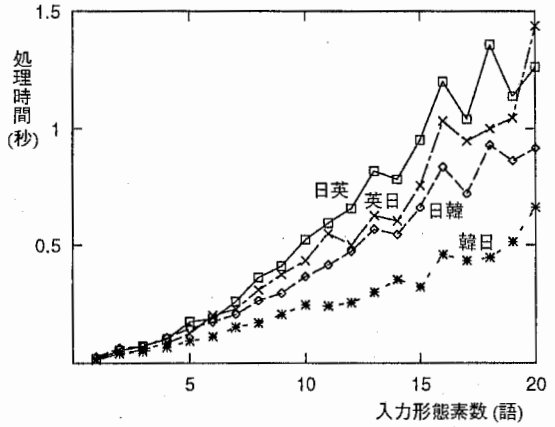

図 7.1: 入力形態素数と処理時間

他の話し言葉翻訳システムの多くは，音声で入出力を行なう音声翻訳システムの翻訳コンポーネントとして構築され，限定されたそれぞれの翻訳対象で音声認識結果を翻訳の入力としているので，TDMTシステムの翻訳性能との優劣を単純に決めることはできない．しかし，本レポートの提案手法を採用したTDMTシステムは，対象とする語彙数の多さ，トピックの広さ，扱う表現の多様さなど適用範囲の点で優位と言える．

例えば，音声翻訳システムASURA(Advanced Speech Understanding and Rendering system at ATR)の翻訳コンポーネントは，素性構造のトランスファ方式を採用し「国際会議に関する問い合わせ会話」を対象として日本語から英語とドイツ語への翻訳を行なってる．目的指向型電話会話の日本語基本表現の約90%をカバーしているが (浦谷、菊井、田代、田窪、定延、成田 1993)，語彙数は約1500語と少なく，音声認識の結果を翻訳入力とした場合の評価結果のみ報告されている (森元，田代，竹澤，永田，谷戸，浦谷，鈴木，菊井 1996).

また，中間言語方式の翻訳コンポーネントを持つ音声翻訳システムJANUSは，「会議の日程調整」を対象として英語，ドイツ語，スペイン語の間の翻訳を行なう[1]．語彙数は約3000～4000語であり，テキスト入力の発話に対するブラインドテストでは，翻訳結果の約80%が理解可能と判定されている (Lavie, Levin, Zhan, Taboada, Gates, Lapata, Clark, Broadhead, and Waibel 1997)．しかし，これは，日英間の翻訳に比べて類似した言語ペアの翻訳についての評価結果であり，翻訳対象の表現は，会話の話題と進行を強く制限することにより意味的曖昧性が抑えられている．

[1]「旅行会話」を翻訳対象とするシステムの研究も始まっている．また，日本語や韓国語を目的言語とする翻訳についても検討されている．

7.4 構成素境界解析の効果の評価

本節では，翻訳処理における構成素境界解析の効果を分析するために行なった評価実験の結果について述べる．

7.4.1 翻訳成功率と構文解析成功率の関係

表 7.5は，構文解析の成功あるいは失敗で評価文を分けてそれぞれの翻訳成功率を調べた結果である．図 7.2，7.3に，入力形態素数ごとの翻訳成功率と構文解析成功率を，日英と日韓の翻訳についてそれぞれ示す．

表 7.5: 構文解析結果ごとの翻訳成功率

	構文解析	日英	日韓	英日	韓日
翻訳成功率A（問題なし）	成功	57.1 %	76.0 %	54.8 %	67.2 %
	失敗	4.0 %	22.9 %	9.7 %	17.8 %
翻訳成功率B（理解可能）	成功	88.6 %	96.5 %	90.8 %	96.8 %
	失敗	43.0 %	84.6 %	63.4 %	85.2 %

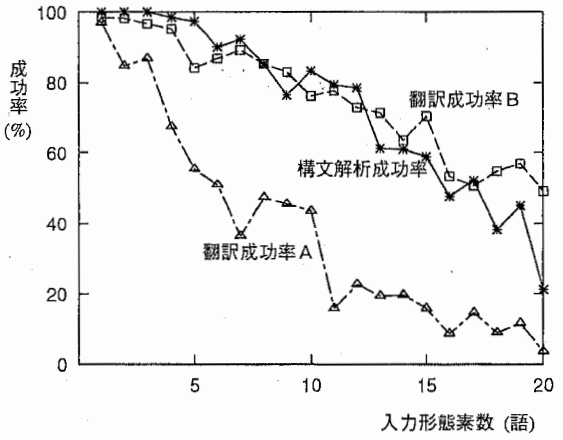

図 7.2: 入力形態素数ごとの翻訳成功率と構文解析成功率（日英）

いずれの言語ペアの翻訳においても，高精度の構文解析が高品質の翻訳結果につながっており，翻訳訓練文の追加などにより構成素境界解析の精度をさらに高めていく必要があることが示された．

日英と英日については，翻訳成功率A（問題なし），翻訳成功率B（理解可能）ともに，構文解析に失敗の影響を受けやすい傾向があった．一方，語順，構文構造，省略表現などで類似する言語ペアの翻訳である日韓と韓日は，構文解析成功率が低下しても，高い翻訳

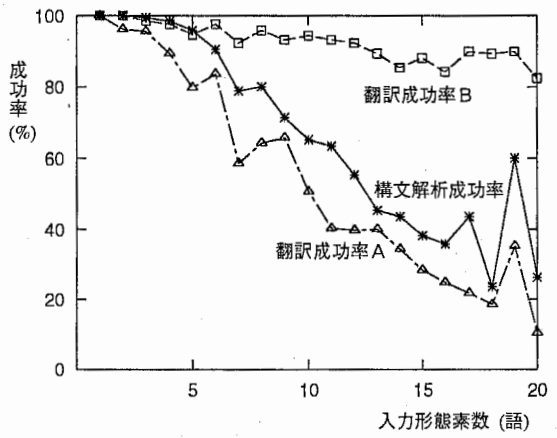

図 7.3: 入力形態素数ごとの翻訳成功率と構文解析成功率（日韓）

成功率B（理解可能）を維持していることが示された．しかし，翻訳成功率A（問題なし）については，構文解析成功率の低下の影響を受けることが図 7.3により示された．すなわち，日韓や韓日においても，より高品質の翻訳結果を得るためには，正しい依存関係を構文解析により求めて正しい訳し分けを行なうなどの必要があり，言語的類似性に頼りすぎるべきではない (金・崔 1998).

7.4.2 構成素境界パタンの組み合わせ方の制限の効果

構成素境界パタンの組み合わせ方の制限によってありえない構文構造を排除することの効果を調べるために，7.2節の表 7.2の評価文の翻訳結果が組み合わせ方の制限の有無でどれだけ違うかという実験を行なった．翻訳結果に違いがあった割合はのべ計算で日英 42.2%，日韓 22.8%，英日 43.3%，韓日 15.5% であり，特に，日英と英日で制限の影響が大きいことが示された．すべての言語ペアの翻訳においてパタンの組み合わせ方を制限したほうが翻訳品質が良い場合が多く，制限の効果が示された．日韓と韓日において翻訳結果が違った割合が小さいのは，日本語と韓国語の語順が類似しており，TDMT システムが持つパタンの種類が少なかったためである．表 7.6に，日英と英日についてパタンの組み合わせ方の制限の有無で翻訳結果が違った例を示す.

表 7.6: パタンの組み合わせ方の制限の有無で翻訳結果が違った例

	日英	英日
入力	はい大阪水上バス交通でございます	does it stop at the Kyoto Kanko Hotel
翻訳結果（制限あり）	Yes this is Osaka Aqua-bus	京都観光ホテルで止まりますか
翻訳結果（制限なし）	This is yes Osaka Aqua-bus	京都観光ホテルでの止まりますか

7.4.3 品詞バイグラムマーカの効果

7.2節の表7.2の評価文のうち，品詞バイグラムマーカを含む構成素境界パタンを使って翻訳を行なった文の割合を調べたところ，のべ計算で日英57.4%，英日77.3%と，品詞バイグラムマーカを用いた構文構造の記述がどちらの翻訳でも頻繁に行なわれていたことが示された．日英に比べて英日での使用割合が大きかったのは，日本語では助詞を介して格関係を構成することが多いのに対して，英語では主格や目的格は述部との間に前置詞を介さないため品詞バイグラムマーカを使用したことが原因である．

さらに，日英で，品詞バイグラムマーカを含む構成素境界パタンを使って構文解析が成功した文の数を調べたところ，のべ463文（全評価文1225文の37.8%）であった．すなわち，品詞バイグラムマーカの導入により構文解析成功率を40.0%から77.8%に向上させたことになる．

また，日英の評価文の中で，助詞脱落表現を含む文は28文（全評価文の2.3%）であった[2]．日英で助詞脱落表現部分について正しい構造が得られたのは，「様子分かる」，「熱出る」，「番組ここで調べる」などの表現を含む23文であり，これは助詞脱落表現を含む文の82.1%に相当した．

これらの結果は，品詞バイグラムマーカが，多様な表現の構文構造の記述，構文解析の精度向上に大きく貢献していることを示す．

[2]「明日行く」のような副詞的名詞句による述部修飾，「三日かかる」のような数量名詞句による述部修飾などは助詞脱落表現にカウントしなかった．

第 8 章

おわりに

表層パタンのみの照合による構成素境界解析を提案し，構成素境界解析と用例利用型処理を組み合わせた変換主導型機械翻訳 (TDMT) の新しい実現手法について述べた．英双方向と日韓双方向の話し言葉の評価実験の結果により，TDMT システムが多様な表現の旅行会話文を話し手の意図が理解可能な結果へ実時間で翻訳でき，本レポートで提案した TDMT が多言語話し言葉翻訳に有効であることを示した．高精度の構文解析が高品質の翻訳結果の重要な要素であることも評価実験結果により示した．今回の評価実験において，複文構造での格関係，複合名詞，等位接続詞表現などで，構成素境界解析の失敗が目立った．効率的な翻訳訓練，変換知識の記述の改良などにより，これらの表現を中心としてさらに構文解析成功率を向上させることが今後の課題である．

日英双方向と日韓双方向の翻訳に加えて，現在，日本語からドイツ語や中国語への翻訳 (Paul, Furuse, and Iida 1996; Yamamoto 1999) についても TDMT の適用を試みており，これらの翻訳についても評価実験を行ないながら，多言語話し言葉翻訳に対する TDMT の汎用性を確認していく予定である．また，話し言葉翻訳を音声認識と統合した音声翻訳では，入力文を分割して理解可能な部分を最大限出力する機構 (Furuse, Yamada, and Yamamoto 1998) や同時翻訳機構など，入力文を時系列に読み込んで処理することが想定される．本レポートで提案した構成素境界解析はこの要求に添うことができる．今後，構成素境界解析を基盤にして，音声処理との統合，長文入力を漸進的に翻訳する同時翻訳機構など高度な話し言葉翻訳技術にも取り組んでいく予定である．

付録 A

評価文に対する TDMT システムの翻訳実行例

3 名の評価者の判定が一致した翻訳結果で分類して，TDMT システムの翻訳実行例を示す.

A.1 日英

● 「問題なし」として翻訳成功

入力： それではジョンフィリップス様あしたの朝八時にお待ちしております

出力： well, mr. john phillips, we are looking forward to seeing you at eight o'clock tomorrow morning

---

入力： じゃそのパッケージのお値段はいくらぐらいになるんですか

出力： well, how much will the price of the package be ?

入力： かしこまりましたではチェックアウトのために十分前にフロントにお越しください

出力： all right, well, please come to the front desk ten minutes before for check-out

---

● 「理解可能」として翻訳成功，「問題なし」としては翻訳失敗

入力： 仁王像は東大寺の入口を守っていますからいつでもご覧になれますよ

出力： the statue of nio has kept the entrance todaiji. you can any time see

---

入力： プライベートビーチのあるホテルリゾートホテルでしたらどこでもよろしいんですよ

出力： anywhere is good for the hotel resort hotel with a private beach

---

入力： お友達のおみやげにね安い化粧品とか雑貨がほしいんですけどどっかお勧めの店ってありません

出力： i'd like to have a cheap cosmetic and the sundries as a friend's souvenirs. there isn't any recommendation store

---

A.2　日韓

●「問題なし」として翻訳成功

入力：　はい八月十三日の土曜日から八月二十日の土曜日まで一週間お願いします

出力：　"예 팔 월 십삼 일 토요일부터 팔 월 이십 일 토요일까지 일 주일 부탁합니다.

入力：　十月二十八日の金曜日ですけれども何時ごろにご到着のご予定でしょうか

出力：　시 월 이십팔 일 금요일입니다만 몇 시쯤에 도착할 예정이십니까?

入力：　ではパックの予約と同時にビデオカメラとシーディープレーヤーの貸し出しの予約もお願いいたします

出力：　그럼 팩 예약과 동시에 비디오 카메라와 시디 플레이어의 대출 예약도 부탁합니다.

●「理解可能」として翻訳成功，「問題なし」としては翻訳失敗

入力：　今ホテルのすぐ近くまで来てるんですけどここからどうやってそちらに行ったらいいのか教えてほしいんですけど

出力：　지금 호텔 바로 근처까지 와 있는데요 여기서 어떻게 그쪽으로 가면 되는지 가르치셨으면 하는데요.

入力：　ですが会席料理という京都の季節の味をうまくコーディネートした京料理をお勧めしたいのですが

出力：　하지만 가이세키 요리라는 교토의 계절의 맛을 맛있히 코디네이트하는 교 요리를 권해 드리고 싶은데요.

入力：　お一人様八千円からございまして私どもはこちらの方をお勧めしております

出力：　한 분 팔천 엔에서 있어서 저희들은 이쪽을 권해 드리고 있습니다.

A.3 英日

●「問題なし」として翻訳成功

入力： the suite is seven hundred dollars per night and the twin room is three hundred dollars per night

出力： スイートは一泊七百ドルです，ツインルームは一泊三百ドルです.

入力： yes my name's john phillips and i rented a nissan march from you yesterday

出力： はい，私の名前はジョン・フィリップスです，昨日あなたから日産マーチを借りました.

入力： yes can you tell me how to get to the stadium from osaka jr station please

出力： はい，どうやってジェイアールの大阪駅からスタジアムに行けば良いか教えて頂けますか.

●「理解可能」として翻訳成功，「問題なし」としては翻訳失敗

入力： so i turn right at the kyoto style restaurant and i will see your inn on the left hand side

出力： それでは，右京都スタイルのレストランに曲がります，左手側に旅館を見ます.

入力： it's only about two hundred meters away and there's an information board outside the station to direct you

出力： 約二百メーターだけ遠くです，教えるために駅付近の案内板があります.

入力： it begins at eight p.m. and you will be sure to pick up your tickets at the ticket window before seven thirty p.m.

出力： 午後八時に始めます，午後七時三十分の前に必ず切符の窓でチケットを受け取ります.

A.4　韓日

●「問題なし」として翻訳成功

入力：　그 다음에 출발하는 새마을호는 열여덟 시에 출발하는 기차가 동대구역에 이십일 시 이 분에 도착합니다.

出力：　その次に出発するセマウル号は十八時に出発する列車がトンテグ駅に二十一時二分に着きます.

入力：　같은 날 열 시 십 분에 출발하는 비행기에는 자리가 있습니다.

出力：　同じ日の十時十分に出発する飛行機には席があります.

入力：　전화 번호는 공칠오의 육육하나의 칠하나하나하나입니다.

出力：　電話番号は零七五の六六一の七一一一です.

●「理解可能」として翻訳成功，「問題なし」としては翻訳失敗

入力：　단지 요금은 일단 손님이 부담을 하시고 나중에 일본에 돌아가셔서 보험 회사에 청구를 하십시오.

出力：　料金はいったんお客様が負担をされますしあとに日本に帰られて保険会社請求をしてください.

入力：　시내에 있는 롯데 호텔하고 잠실에 있는 롯데 월드 호텔이 있는데 어느 쪽이신지 알고 계십니까?

出力：　市内にあるロッテホテルとチャムシルにあるロッテワールドホテルがあるのですがどちらかご存知ているのですか.

入力：　이 서류를 한국에 빨리 보내야 하는데 제일 빨리 보낼 수 있는 방법은 뭐죠?

出力：　この書類を韓国に早く送らなければならないのですが一番早く送れる方法は何ですね.

参考文献

Brown, R. (1996). "Example-Based Machine Translation in the Pangloss System." In *Proceedings of COLING-96*, pp. 169–174.

古瀬蔵, 隅田英一郎, 飯田仁 (1994). "経験的知識を活用する変換主導型機械翻訳." 情報処理学会論文誌, **35** (3), pp. 414–425.

Furuse, O. and Iida, H. (1994). "Constituent Boundary Parsing for Example-Based Machine Translation." In *Proceedings of COLING-94*, pp. 105–111.

Furuse, O., Kawai, J., Iida, H., Akamine, S., and Kim, D.B. (1995). "Multi-lingual Spoken-Language Translation Utilizing Translation Examples." In *Proceedings of NLPRS-95*, pp. 544–549.

Furuse, O., Sobashima, Y., Takezawa, T., and Uratani, N. (1994). "Bilingual Corpus for Speech Translation." In *Proceedings of AAAI'94 Workshop 'Integration of Natural Language and Speech Processing*, pp. 84–91.

Furuse, O., Yamada, S., and Yamamoto, K. (1998). "Splitting Long or Ill-formed Input for Robust Spoken-language Translation." In *Proceedings of COLING-ACL'98*, pp. 421–427.

古瀬蔵, 山本和英, 山田節夫 (1999). "構成素境界解析を用いた多言語話し言葉翻訳." 自然言語処理, **6** (5), pp. 63–92.

池原悟, 白井諭, 小倉健太郎 (1994). "言語表現体系の違いに着目した日英機械翻訳機能試験項目の構成." 人工知能学会誌, **9** (4), pp. 569–579.

加藤直人 (1995). "定型パタンを含む文の機械翻訳手法." 情報処理学会論文誌, **36** (9), pp. 2081–2090.

金泰完, 崔杞鮮 (1998). "日韓機械翻訳システムの現状分析および開発への提言." 自然言語処理, **5** (4), pp. 127–149.

Lavie, A., Levin, L., Zhan, P., Taboada, M., Gates, D., Lapata, M., Clark, C., Broadhead, M., and Waibel, A. (1997). "Expanding the Domain of a Multi-lingual Speech-to-Speech Translation System." In *Proceedings of ACL-EACL'97 Workshop 'Spoken Language Translation*, pp. 67–72.

Margerman, D. and Marcus, M. (1990). "Parsing a Natural Language Using Mutual Information Statistics." In *Proceedings of AAAI-90*, Vol. 1, pp. 984–989.

森元逞, 田代敏久, 竹澤寿幸, 永田昌明, 谷戸文廣, 浦谷則好, 鈴木雅実, 菊井玄一郎 (1996). "音声翻訳実験システム (ASURA) のシステム構成と性能評価." 情報処理学会論文誌, **37** (9), pp. 1726–1735.

長尾眞, 辻井潤一 (1985). “Mu プロジェクトにおける日英翻訳結果の評価.” 自然言語処理研究会 47–11, 情報処理学会.

Paul, M., Furuse, O., and Iida, H. (1996). “Japanese-to-German Spoken-language Translation Utilizing Empirical Linguistic Knowledge.” 情報処理学会第 53 回全国大会講演論文集, 4L-13.

白井諭, 松島英之, 井上浩子, 松尾三津恵, 矢部孝幸, 内野一 (1997). “市況速報記事を対象とした日英翻訳システムの構成.” 情報処理学第 55 回全国大会講演論文集, 5J-5.

Sumita, E. and Iida, H. (1992). “Example-based Transfer of Japanese Adnominal Particles into English.” *IEICE Transactions on Information and Systems*, **E75-D** (4) pp. 585–594.

Sumita, E., Yamada, S., Yamamoto, K., Paul, M., Kashioka, H., Ishikawa, K. and Shirai, S. (1999). “Solutions to Problems Inherent in Spoken-Language Translation: The ATR-MATRIX Approach.” In *Proceedings of Machine Translation Summit VII '99 (MT SUMMIT VII)*, pp. 229–235.

浦谷則好、菊井玄一郎、田代敏久、田窪行則、定延利之、成田一 (1993). “話し言葉の日英翻訳システムの評価法.” 情報処理学会第 46 回全国大会講演論文集, 6B-4.

渡辺日出雄, 武田浩一 (1998). “用例ベース処理を用いたパタンベース翻訳システム.” 言語処理学会第 4 回年次大会発表論文集, pp. 488–491.

山田節夫, 山本和英, 飯田仁 (1998). “「協調融合機械翻訳」における訳語選択.” 言語処理学会第 4 回年次大会発表論文集, pp. 508–511.

山本和英, 古瀬蔵, 飯田仁 (1996). “用例に基づく日韓の対話翻訳処理機構.” 情報処理学会第 53 回全国大会講演論文集, 4L-10.

山本和英, 河井淳, 隅田英一郎, 古瀬蔵 (1997). “単語と品詞の混合 n-gram を用いた形態素解析.” 情報処理学会第 54 回全国大会講演論文集, 1C-2.

Yamamoto, K. (1999). “Proofreading Generated Outputs: Automated Rule Acquisition and Application to Japanese-Chinese Machine Translation.” In *Proceedings of International Conference on Computer Processing of Oriental Languages (ICCPOL'99)*, pp. 87–92.