

非線形言語モデルに基づく文型パターン翻訳に 関する研究

2018年1月

坂田 純

概要

本研究では、非線形言語モデルに基づく日英文型パターン機械翻訳方式を提案し実装する。

非線形言語モデルでは、話者の視点等の主観的な表現は、様相表現等の要素的な表現のみならず、語順や語形そして品詞の種類等による全体としての関係、つまり文構造によっても担われているとしている。またこのような文構造の持つ抽象的な意味は、単純な部分構造の組み合わせでは再現できないとしている。そこで非線形言語モデルでは、部分構造を組み合わせで文構造を組み立てるのではなく、文構造の持つ抽象的な意味を保存するように作成された文型パターンを用いて文構造を得る。文型パターンは対訳文対ごとに原言語パターンと目的言語パターンの対として作成される。そのため原言語の文型パターンの適合により、入力文全体に対する構文解析のような複雑な処理なしに文構造が抽出でき、対となる目的言語の文型パターンの使用により、文構造の抽出と目的言語への構造変換が同時に達成される。非線形言語モデルでは、文構造の持つ意味を文型パターンとして抽出するために、他の要素に置き換えても文構造の持つ意味が変化しない要素を「線形要素」、置き換えにより変化してしまう要素を「非線形要素」と定義している。この定義に従い、すでに重文複文文型パターン辞書が作成されている。文型パターン辞書では、日英の言語分析や機械翻訳等への多岐の利用を視野に入れ、文型パターンの作成と統語的意味的分類の付与が行われている。また文型パターンの線形要素は変数として記述されており、変数化の範囲に応じて単語、句、節の3レベルの文型パターンが作成されている。

本研究では、このうち単語レベルと句レベルの文型パターンを用いる日英機械翻訳方式を提案し実装した。非線形言語モデルに基づく文型パターン翻訳方式は機械翻訳において初の試みであり、まず文型パターンの入力文全体に対する適合率や、翻訳精度を調査した。翻訳実験より、単語レベル文型パターン翻訳方式は文型パターンの適合率が約11%と低いが、文型パターンが適合した場合は翻訳精度が高いことが明らかになった。これに対し句レベル文型パターン翻訳方式ではパターン適合率は約28%に改善したが、翻訳精度は単語レベルに比べて大幅に低下していた。句レベル文型パターン翻訳方式の翻訳精度の改善と節レベル文型パターン翻訳方式の実装を行う必要がある。

次に統語的意味的分類の記述の妥当性と、単語レベル文型パターン翻訳方式と類似の翻訳システムである tree-to-string の SMT との間で、作成される文構造の比較を行った。調査の結果、統語的意味的分類の不適切な付与が多数存在することが明らかになった。また統語的意味的分類の言語分析への利用では重文複文全体の表現に対する新たな記述項目の追加が必要と示唆された。さらに単語レベル

句型パターン翻訳方式は tree-to-string の SMT よりも高い精度で日英間の文構造変換が行われていることが明らかになった。

目次

第1章	序論	1
1.1	研究背景	1
1.2	本研究の目的	3
第2章	関連研究	4
2.1	構文解析	4
2.1.1	句構造解析	4
2.1.2	ATN	6
2.1.3	係り受け解析	7
2.2	機械翻訳分野での関連研究	7
2.2.1	翻訳システムの分類	7
2.2.2	多段翻訳方式	9
2.2.3	tree-to-string の SMT	12
2.2.4	翻訳メモリを用いるシステム	12
第3章	非線形言語モデル	14
3.1	言語の規範としての意味	14
3.2	言語の非線形性	15
3.3	線形要素と非線形要素の定義	15
3.4	文構造の持つ意味の具体例	16
3.5	重文複文への非線形言語モデルの適用	17
3.6	線形要素の抽象度と文構造の持つ意味の抽象度	18
3.6.1	変数の意味属性による細分化と文構造の持つ意味の抽象度	18
3.6.2	線形要素の範囲と文構造の持つ意味の抽象度	18
3.7	線形要素と非線形要素の入れ子構造	18
第4章	重文複文文型パターン辞書	20
4.1	文型パターン作成の基準	20
4.2	文型パターンの具体例	22
4.3	日英の構造変換のための変数化	24

4.4	文型パターン辞書の統計量	25
4.5	統語的意味的分類と英語構文分類	25
4.6	意味的等価変換方式	25
4.7	翻訳に利用するための英語文型パターンの記述の変換	29
第5章	非線形言語モデルに基づく日英文型パターン翻訳方式	31
5.1	文型パターン翻訳の概要	31
5.1.1	文型パターンの適合率	32
5.1.2	3レベルの文型パターンの組み合わせ方	33
5.1.3	ルールベースかつコーパスベースの翻訳システムとしての指針	33
5.2	形態素解析	34
5.3	文型パターンの照合	34
5.3.1	SPMによる文型パターンの照合方法	36
5.3.2	文型パターンの照合と構文解析の違い	37
5.3.3	文型パターンの照合例	38
5.4	文型パターンの絞込み	39
5.5	英語訳出文の生成	40
5.5.1	変数に対応する要素の局所的翻訳	40
5.5.2	関数に従った変換処理	41
5.5.3	複数訳語候補からの訳語選択	41
5.6	複数文型パターンが適合した場合の処理	42
5.7	文型パターン翻訳方式のこれまでの取り組み	42
第6章	単語レベル文型パターン翻訳方式	44
6.1	単語レベル文型パターン翻訳方式の概要	44
6.2	日本語文型パターンの照合	44
6.3	日本語文型パターンの絞込み	45
6.4	単語レベルの英文生成	46
6.5	単語辞書	48
6.6	単語レベル文型パターンによる英文生成の例	48
6.7	翻訳実験	49
6.7.1	実験条件	49
6.7.2	比較システム	50
6.7.3	作成された単語辞書	50
6.7.4	パターン適合率	50
6.7.5	翻訳精度の評価	50
6.7.6	翻訳結果の具体例	53

6.7.7	文長等の統計量	54
6.7.8	比較システムと併用した場合の翻訳精度	55
6.8	考察	56
6.8.1	翻訳精度が低い原因	56
6.8.2	意味属性コードを用いたパターン絞込みの効果	58
6.8.3	パターン適合率の改善	59
6.8.4	言語モデルの学習文増加の効果	60
6.8.5	単語翻訳確率を使用しなかった場合の翻訳精度	60
6.9	今後の調査課題	61
6.10	単語レベル文型パターン翻訳方式のまとめ	62
第7章	句レベル文型パターン翻訳方式	63
7.1	句レベル文型パターン翻訳方式概要	63
7.2	句パターン辞書	64
7.2.1	句パターンの抽出	65
7.2.2	句パターンの種類と変換処理	65
7.2.3	句パターンの例	66
7.3	形態素解析	67
7.4	句レベル日本語文型パターンの照合	67
7.4.1	句レベル文型パターン照合例	69
7.5	日本語文型パターンの絞込み	70
7.6	句レベルの英文生成	71
7.6.1	単語変数に対応する要素の処理	72
7.6.2	句変数に対応する要素の処理	72
7.7	英文生成の例	73
7.8	翻訳候補文からの出力文の選択	74
7.9	翻訳実験	74
7.9.1	実験条件	74
7.9.2	比較システム	74
7.9.3	作成された句パターン	74
7.9.4	パターン適合率	75
7.9.5	評価結果	76
7.10	考察	77
7.10.1	使用した文型パターンの適切さ	77
7.10.2	不適切な文型パターンを使用した要因	78
7.10.3	句変数に対する意味属性による絞込みの問題	80
7.10.4	句変数に対応する要素の翻訳	81

7.10.5	句パターンが未適合の場合の文型パターンの適切さ	81
7.10.6	単語レベル文型パターン翻訳方式との翻訳精度の比較	82
7.11	句レベル文型パターン翻訳方式のまとめ	84
第 8 章	節レベル文型パターン翻訳方式	85
8.1	節レベル文型パターン翻訳を実装するために必要な事項	85
8.2	節レベル文型パターン翻訳方式の手順	86
8.3	形態素解析	86
8.4	文型パターン照合	86
8.5	文型パターンの絞込み	86
8.6	節レベルの英文生成	87
8.7	節レベル文型パターン翻訳における調査課題	87
第 9 章	統語的意味的分類と文構造の持つ意味に関する考察	88
9.1	統語的意味的分類の記述形式の妥当性	88
9.1.1	英語文型パターンの適切さと翻訳精度による分類	89
9.1.2	節の意味属性の一致率	89
9.1.3	変数に対応する要素の割合	90
9.1.4	具体例を用いた分析	90
9.1.5	統語的意味的分類に対する調査のまとめ	92
9.2	文型パターン翻訳と tree-to-string の SMT における文構造の違い	93
9.2.1	tree-to-string の SMT において作成される文構造	94
9.2.2	比較システムにおける文構造の階層数	95
9.2.3	使用された翻訳ルールの適切さ	95
9.2.4	不適切な翻訳ルール使用の要因	96
9.2.5	翻訳ルールの選択	97
9.2.6	両システムにおける文構造による翻訳精度への効果	98
9.3	文構造の持つ意味に関する調査課題	99
第 10 章	文型パターン翻訳方式の現状と今後の研究方針	100
第 11 章	まとめ	101
付録 A	本論文で使用した文型パターンの要素の説明	109
付録 B	文型パターン翻訳方式の実験環境	112

目次

2.1	句構造解析の例	5
2.2	遷移ネットワークの例	6
2.3	句構造表現と係り受け構造表現	8
3.1	重文と複文の表現構造	19
4.1	意味的等価変換方式	28
5.1	文型パターン翻訳の概要	32
5.2	形態素解析結果	35
5.3	文型パターンの遷移ネットワークと文法規則の遷移ネットワークの例	37
6.1	単語レベル文型パターン翻訳方式	45
6.2	形態素解析結果	46
6.3	意味属性による名詞単語変数の適合範囲の例	47
6.4	単語レベル文型パターンによる英文生成の例	49
7.1	句レベル文型パターン翻訳方式	64
7.2	句パターンの抽出	65
7.3	形態素解析結果	70
7.4	句レベルの英文生成の例	73
9.1	比較システムで作成される日英の文構造	94

表 目 次

2.1	CKY法の規則の一部	5
2.2	係り受け文法の規則の一部	7
2.3	一般的な翻訳方式	9
2.4	結合価パターンの記述例	11
4.1	本論文で使用した文型パターンの要素の説明	22
4.2	句と節の変数化の範囲	22
4.3	文型パターンの記述例	24
4.4	日英単独と日英セットでの変数化の違い	25
4.5	パターン化の対象となる日本語文の分類例	26
4.6	文種ごとの文数, 日本語文の平均形態素, 英語文の平均単語数	26
4.7	統語的意味的分類の説明	27
4.8	統語的意味的分類と英語構文分類の具体例	27
4.9	真理項	28
4.10	節間意味分類の分類項目	29
4.11	節の意味属性の分類項目	30
4.12	翻訳のための記述の変換	30
5.1	パターン照合結果	39
5.2	3レベルの文型パターンの適合の有無	39
5.3	変数に対応する要素の局所的翻訳に使用するデータベース	41
6.1	パターン照合結果	45
6.2	単語辞書の例	48
6.3	単語辞書の日本語見出し語数	51
6.4	自動評価結果	51
6.5	人手評価の基準とその例	52
6.6	人手評価結果	53
6.7	カッパ係数の値	53
6.8	翻訳結果の例	54

6.9	翻訳ルールの例	55
6.10	日本語文の文長(文字数)	55
6.11	自動評価結果(ハイブリッド)	56
6.12	複合的な表現の問題により翻訳精度の低い例	57
6.13	不適切な訳語選択により翻訳精度の低い例	58
6.14	パターン絞込みを行わなかった場合の自動評価結果	59
6.15	パターン絞込みにより適切な文型パターンが取り除かれた割合	59
6.16	パターン絞込みにより不適切な日英文型パターンが取り除かれた例	60
6.17	パターン絞込みにより適切な日英文型パターンまで取り除かれた例	61
6.18	自動評価結果(言語モデル学習文追加)	61
6.19	自動評価結果(単語翻訳確率なし)	61
7.1	句パターン作成のための記述の変換	66
7.2	VP-VPパターンの例	67
7.3	日本語 VP パターンの例	67
7.4	日本語 NP パターンの例	68
7.5	日本語 AJP パターンの例	68
7.6	日本語 AJVP パターンの例	69
7.7	日本語 ADVP パターンの例	69
7.8	複数の適合の仕方が可能な例	70
7.9	パターン照合により得られた日英文型パターンと変数情報	71
7.10	句パターン数	75
7.11	自動評価結果	76
7.12	人手評価結果	76
7.13	本翻訳方式において人手評価値の高い例1	77
7.14	本翻訳方式において人手評価値の高い例2	78
7.15	英語文型パターンの適切さと人手評価値の関係	79
7.16	不適切な文型パターンを使用した原因とその内訳	79
7.17	(4)と(5)の両方の原因を持つ例	80
7.18	句変数に対応する要素が慣用的表現である例	82
7.19	句パターンが不適合の例	83
7.20	対比較評価結果	84
9.1	節の意味属性の一致率	90
9.2	入力文における変数に対応する要素の割合とその内訳	90
9.3	分析を行った具体例	92
9.4	文構造の持つ意味の抽出例	93

9.5	文構造における階層数とその内訳	95
9.6	使用された不適切な翻訳ルールの種類による分類	96
9.7	日英で対応の取れていない誤った翻訳ルール	97
A.1	本論文でを使用した文型パターンの変数の説明	109
A.2	本論文でを使用した文型パターンの関数の説明	110
A.3	本論文でを使用した文型パターンの字面の説明	110
A.4	本論文でを使用した文型パターンの記号の説明	111
B.1	文型パターン翻訳方式の実験環境	112
B.2	単語レベル文型パターン翻訳方式の実行時間	112
B.3	句レベル文型パターン翻訳方式の実行時間	113

第1章 序論

1.1 研究背景

機械翻訳のタスクは、すでに完成された原言語文を目的言語文に変換することである。文は語が組み合わさったものであり、人間による文の理解または他言語への翻訳は、文を文頭から分解しながら、そして同時に要素を組み合わせながら行われている。人間が文の意味を文の分解と合成により理解できるのは、要素の足し合わせごとに、元の要素に還元されない新たな意味を生成（創発）しているからである。一般に、分割するごとに新たな性質の要素となり、これらの性質を足し合わせても元の性質とはならないこと、逆に要素が新たに加わるごとに全体として新たな性質としてあらわれることを非線形性と呼ぶ[1]。言語表現の非線形性、つまりある表現に他の語が加わるたびに、各語の意味の合算に還元できない表現全体の意味が新たに生成すること、このことは一般に認められている。

機械翻訳においても、文の分解がまず行われ、その後の個々の処理とその結果の結合を経て目的言語文が生成される。しかしながら計算機に意味そのものの生成や理解をさせることは困難であり、機械翻訳では、できるだけ意味が消失しない形で文の分解、再構築を行う方策が採られている。一般には、文構造とその構造の中の要素に分割後、各処理を施してから、目的言語文の生成が行われている。

自然言語の文は何らかの構造を持つが、発話または表記される時点で、1次元の並びに変換されている[2]。自然言語処理では一般に、構文解析によりそのような文の構造を復元する。構文解析では、語と語の関係を規定しているルール（文法規則）を記述し、それに基づいて文構造の解析を行う。文の意味はそれぞれの人間に生成されているとしても、文の意味が複数の人間に共有され得るのであるから、共通の意味の生成を可能にする規範が存在するはずである。文法規則や文構造がこのような規範を担っているとみなすことができる。文構造と文の意味の関係は明確にはわかっていないが、文構造は各単語の関係を規定することで、単語だけでは表現することのできない事態や話者の視点を表現する機能を持つと考えられる。

文構造を抽出、変換し、訳出文を生成する翻訳方式を、構文トランスファー方式と呼ぶ。近年において主流の機械翻訳方式を務めたのが構文トランスファー方

式であり、この方式は言語的知識を元にした各種変換ルールの適用を精密に制御することで成り立っている。それぞれの言語において文法規則は異なることから、機械翻訳を行うには言語間での文法規則の対応とそれに基づく文構造の変換が必要である。一般に構文トランスファー方式は、句構造解析または係り受け解析技術を利用して文構造を抽出し変換を行っている。

構文トランスファー方式は比較的高い翻訳精度を持つが、日英のような性質の大きく異なる言語間では、十分な翻訳精度はまだ得られていない。構文トランスファー方式は構文解析結果を利用し文構造の変換を行っていることから、上記のことは次の二点の問題を示唆している。

1. 構文解析結果が文構造の持つ機能を適切に保存していない。
2. 解析と変換の分離により、目的言語への構造変換の過程で文構造の持つ機能が消失する。

この二点の問題の解決を目的として、非線形言語モデルに基づく文型パターン翻訳方式が考案された。

非線形言語モデルでは、文構造そのものが、単語だけでは表現することのできない事態や話者の認識の仕方を表現する機能を、「意味」として持つと考えている。さらに単純な部分構造の組み合わせでは、一文全体の文構造の持つ意味が再現できないと仮定している [3]。また他の要素に置き換えても文構造の持つ意味が変化しない要素を「線形要素」、置き換えにより変化してしまう要素を「非線形要素」と定義している。そしてこの定義に従って文型パターンを作成し翻訳に利用することで、文構造の持つ意味を保存したまま精度の高い翻訳文が得られると想定している。文型パターンは対訳文対ごとに原言語パターンと目的言語パターンの対として作成する。そのため原言語側の文型パターンを入力となる原言語文に当てはめることで、対となる文型パターンにより目的言語側の文構造が得られることになる。

すでに非線形言語モデルに基づく「重文複文文型パターン辞書」（通称「鳥バンク」）[4] が作成されている。この文型パターン辞書では、日英の言語分析や機械翻訳等への多岐の利用を視野に入れ、日英文型パターンの作成および統語的意味的分類 [5] の付与が行われている。翻訳の際に、文型パターンの適合により一般の構文解析とは異なる形で文構造を抽出することになるが、それだけでなく文構造に対応する形で意味を付与できれば、明示的に文構造の持つ意味を抽出することができるはずである。日本語の重文と複文の文構造の持つ意味の明示的な抽出のため、統語的意味的分類の記述形式が定められ、対訳文ごとに付与されている。そして統語的意味的分類を媒介項として、原言語と目的言語間で類似の文構造による多対多の構造抽出を行う意味的等価変換方式が提案されている。

1.2 本研究の目的

本研究では非線形言語モデルに基づく日英文型パターン翻訳方式を提案し実装する。非線形言語モデルに基づく文型パターンはすでに、重文複文文型パターン辞書において作成されている。非線形言語モデルに基づく文型パターン翻訳方式は機械翻訳初めての試みであるため、まずパターン適合率や翻訳精度等の翻訳性能を調査する。そして翻訳結果を用いて統語的意味的分類の記述の妥当性を検討し、意味的等価変換方式への利用が可能であるか検討する。また類似する翻訳方式である tree-to-string の SMT との間で、作成された文構造を比較し、日英間の文構造変換の精度とその翻訳への影響を調査する。

文型パターン辞書では変数化の範囲に応じて単語、句、節の3レベルの文型パターンが作成されている。本研究ではこのうち単語レベル文型パターン翻訳方式と、句レベル文型パターン翻訳方式を実装する。単語レベル文型パターン翻訳方式の実装において新たに単語辞書の作成を行い、訳語選択に利用する確率値に単語翻訳確率を加えた。句レベル文型パターン翻訳方式は単語レベル文型パターン翻訳方式を元に実装した。句レベル文型パターンは句変数を持つ。句変数に対応する要素に対し、単語レベル文型パターンから抽出した句パターンと、単語レベル文型パターン翻訳方式を再帰的に利用することで翻訳を行った。節レベル文型パターン翻訳は未実装であるが、句変数に対応する要素の翻訳と同様の方法を、節変数に対応する要素の翻訳に適用することで、実装が可能である。

文型パターン辞書では、重文と複文の文構造の持つ意味を、統語的意味的分類として付与している。統語的意味的分類の言語処理や言語分析への利用のためには、統語的意味的分類の記述形式の妥当性と付与精度を確かめる必要がある。単語レベル文型パターン翻訳の結果を用いて、翻訳精度と統語的意味的分類の間の関連性を調査した。そして翻訳結果を用いた分析より、記述形式の妥当性と統語的意味的分類の付与精度を調査した。また文型パターン対による文構造の抽出と変換の効果を調べるため、tree-to-string の SMT で作成される文構造と文型パターン対の違いを検討した。具体的には両翻訳システム間における、文構造の階層性と文構造変換の精度の違いを調査した。

以下、第2章で関連研究についてまとめる。第3章で非線形言語モデル、第4章で重文複文文型パターン辞書を説明する。第5章で3レベルの文型パターンに共通する、日英文型パターン翻訳方式の手順を記述し、第6章で単語レベル文型パターン翻訳方式の実装と実験結果を、第7章で句レベル文型パターン翻訳方式の実装と実験結果を示す。第8章で節レベル文型パターン翻訳方式の実装について説明する。第9章で統語的意味的分類と文構造の持つ意味の関係を調査する。第10章で文型パターン翻訳方式の現在の状況と今後の展開をまとめる。最後第11章で本研究のまとめとする。

第2章 関連研究

2.1 構文解析

非線形言語モデルに基づく文型パターン翻訳方式は、構文トランスファー方式の問題点の改善を目的に提案された。構文トランスファー方式では文構造の抽出に、構文解析結果を利用している。文型パターンによる文構造の抽出と変換の効果を調査するためには、構文解析により得られる文構造と文型パターンとの違いを考察する必要がある。本節では構文解析技術のうち、構文トランスファー方式で一般に用いられる句構造解析と係り受け解析について説明する。

2.1.1 句構造解析

自然言語の文は何かしらの構造をもったものであるが、表記された文は1次元の単語の並びに変換されている。構文解析はそのような文の構造を復元する処理である。構文解析では、まず文法規則を記述し、それに基づいて入力文を解析し、その構造を明らかにする [2]。

構文解析のための文法規則は一般に文脈自由文法が用いられる。文法規則は (2.1) 式のように定義される。

$$G = \langle V_X, V_C, V_T, P_D, P_P, \sigma \rangle \quad (2.1)$$

$$V_X : \text{非終端記号}, V_C : \text{前終端記号}, V_T : \text{終端記号} \quad (2.2)$$

$$P_D : \text{辞書規則}, A \rightarrow B \quad A \in V_C, B \in V_T \quad (2.3)$$

$$P_P : \text{句構造規則}, A \rightarrow \alpha \quad A \in V_X, \alpha \in (V_C \cup V_X)^+ \quad (2.4)$$

$$\sigma (\in V_X) : \text{初期記号} \quad (2.5)$$

自然言語に適用する場合は、 V_T が単語の集合、 V_C が品詞の集合、 V_X が句記号 (NP, VP 等) となる。

具体例として、構文解析アルゴリズムの中でも、もっとも基本的なボトムアップアルゴリズムである CKY 法の文法規則の一部を表 2.1 に示す。CKY 法の文法

は、辞書規則と (2.6) 式に示す句構造規則のみで与えられる。

$$A \rightarrow BC \quad A \in V_X, B, C \in (V_C \cup V_X) \quad (2.6)$$

ボトムアップアルゴリズムでは文法規則を右辺から左辺への書き換え規則とみて、入力文 (ボトム) からトップノード (S) を作り出す。「I saw a girl with a telescope」の句構造解析結果を図 2.1 に示す。このような句構造解析で得られた文構造と文の意味の間に、何かしらの対応関係が存在すると考えられている。

実際には自然言語の文法を、純粋に文脈自由文法のみで記述することは困難であり、自然言語へ適用するために、何かしらの文脈自由文法の拡張が行われている。例えば英語の数・人称の一致を扱う場合、一致の制約を文法規則に与える一つの方法として、各非/前終端記号に数、人称などを表す素性を付与し、適用条件としてそれらの素性を調べるといった方法がとられている。

表 2.1: CKY 法の規則の一部

句構造規則	辞書規則
$S \rightarrow NP VP$	$N \rightarrow I girl telescope$
$NP \rightarrow DET N$	$NP \rightarrow I girl telescope$
$NP \rightarrow NP PP$	$V \rightarrow saw$
$VP \rightarrow VP NP$	$VP \rightarrow saw$
$VP \rightarrow VP PP$	$DET \rightarrow a$
$PP \rightarrow PREP NP$	$PREP \rightarrow with$

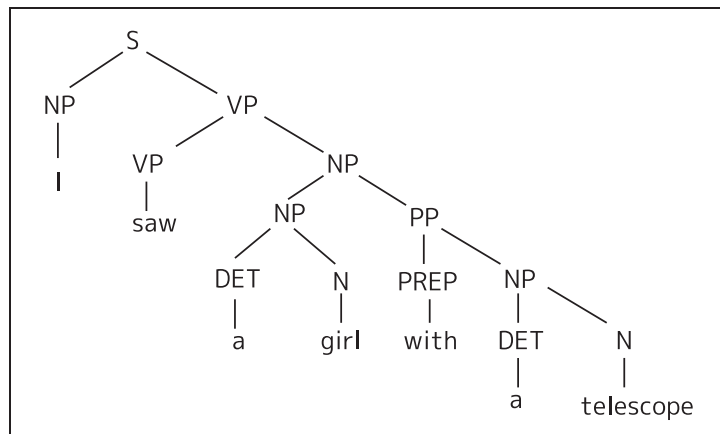


図 2.1: 句構造解析の例

2.1.2 ATN

状態遷移のネットワークとして、句構造の文法記述を可能にした枠組みを拡張遷移ネットワーク (Augmented Transition Network, ATN) と呼ぶ [2, 6]. 遷移ネットワークは状態を示すノードと遷移の経路を示すアークからなり、アークには前終端記号がラベルとして与えられる. 図 2.2 に遷移ネットワークの例を示す. 図 2.2 の「jump」は、品詞がなくてもノードを移動できることを示す. 図 2.2 上が文法規則「NP → (DET) ADJ* N」の遷移ネットワークであり、下が「S → NP V NP*」のネットワークである. 「(DET)」は「DET」が省略可能であること, 「ADJ*」は「ADJ」の 0 回以上の繰り返しを示す. 例えば, 「the/DET girl/N」や「pretty/ADJ little/ADJ children/N」等が図 2.2 上の遷移ネットワークによって NP として受理される. 遷移ネットワーク上では、一番左の初期状態から出発し、入力の単語の品詞に一致するラベルを持つアークがあれば、そのアークを通して次の状態に遷移する. これを繰り返して最終状態 (pop) に到達すれば、それまでの入力単語列が正しく受理されたということになる.

本論文で使用するパターン検索システムでは、この ATN 法を採用してパターン照合を行っている. パターン照合での ATN 法の利用方法は 5.3.1 節で説明する.

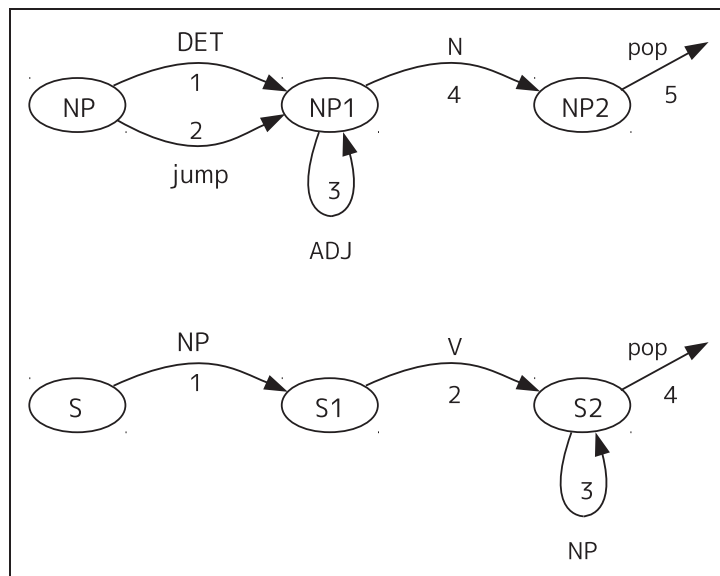


図 2.2: 遷移ネットワークの例

2.1.3 係り受け解析

句構造解析は主にヨーロッパ言語を元に考え出された解析方法である。日本語は省略が多く語順が比較的自由であり、句構造規則のみで日本語文の基本的構造を規定するには、大量の規則を用意する必要がある。日本語においては、古くから係り受け文法という考え方がある [2]。係り受け文法では、ある文節が他の文節に係るという形式で文構造を表現する。係り受け可能な二つの文節を規定したものが文法規則となる。係り受け解析では、係り受け文法を用いて、文の可能な係り受け構造を抽出する。係り受け文法の規則の一部を表 2.2 に示す。

図 2.3 に、「私は机の上にペンを置いた」の句構造表現と係り受け構造表現を示す。図 2.3 の右下の (c) に示したように、係り受け解析結果も木構造として表現できる。トップノードにあたるのは係られる先の文末表現“置いた”であり、日本語では文末表現に向かって文頭から収束させることに対応している。

表 2.2: 係り受け文法の規則の一部

係る文節のタイプ	受ける文節のタイプ
名詞+格助詞「が」	動詞 (ガ格の格要素を支配する動詞)
名詞+格助詞「を」	動詞 (ヲ格の格要素を支配する動詞)
動詞 連用形	動詞, 形容詞, 形容動詞
動詞 連体形	名詞
動詞+述語接続助詞	動詞, 形容詞, 形容動詞
副詞	動詞, 形容詞, 形容動詞
連体詞	名詞

2.2 機械翻訳分野での関連研究

2.2.1 翻訳システムの分類

一般に用いられる機械翻訳方式の区分の一つに、ルールベースとコーパスベースへの二分がある。ルールベースの“ルール”は人手による言語分析を通して得られたヒューリスティックなルールを指し、このようなルールをやはりヒューリスティックな適用規則に従って組み合わせる方式をルールベース翻訳と呼ぶ。構文トランスファー方式、ピボット方式、知識型翻訳等がルールベース翻訳に含まれる [2]。現在のルールベース翻訳は構文トランスファー方式がほとんどであり、構文トランスファー方式を指してルールベース翻訳と呼ぶことが多い。次に対訳文

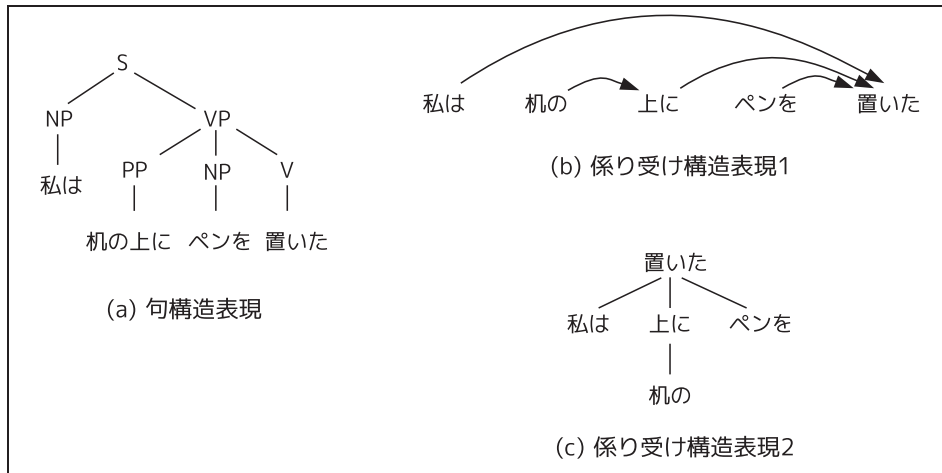


図 2.3: 句構造表現と係り受け構造表現

対等のコーパスを収集し、そのコーパスから主に自動で翻訳のためのルールを抽出し利用する方式がコーパスベース（データベース）の方式である [7]。フレーズベース SMT (statistical machine translation) やニューラルネットワーク翻訳がこれにあたる。

一方、どちらかに明確に区分できない翻訳方式が存在し、用例型翻訳の一部のシステムや文型パターン翻訳が中間的な翻訳方式である。またこれらの各方式もそれぞれの翻訳システムごとに、ルールとコーパス使用の度合いによりさらに細分することができる。文型パターン翻訳は、収集した対訳文対ごとに文型パターンを作成することからコーパスベースとみなすことができる。しかし変数や関数の処理はヒューリスティックな処理を多数組み込むことから、ルールベース翻訳の一種とみなすことも可能である。

ルールベースとコーパスベースの区分とは別に、構文情報を利用するシステムとそうでないシステムで区分することが可能である。ルールベースの構文トランスファー方式のみならず、一部の用例型翻訳や統語情報に基づく SMT、そして文型パターン翻訳が構文情報を利用するシステムに含まれる。一般的な構文情報を利用する翻訳システムは、句構造解析や係り受け解析技術を用いて文構造を抽出する。これらの解析技術は、言語的知識をもとにする汎化された部分ルールの組み合わせで、入力文の文構造を抽出している。これに対し文型パターン翻訳は、対訳文対ごとに作成しておいた文型パターンを当てはめることで、入力文の文構造の抽出を達成している。

表 2.3 に主要な翻訳方式の分類を示す。表 2.3 はルールベースかコーパスベースか、構文情報を使用するかないかで分類している。システムの呼称の仕方は一般

には未統一であり，翻訳システムの分類の仕方も統一的な方法があるわけではない¹．表 2.3 中の用例型翻訳は構文解析を用いて文構造変換を行うシステムのことを示す．このような用例型翻訳として KyotoEBMT システム [8] 等が挙げられる．用例型翻訳のうち構文解析を利用しない方式は，類似する翻訳メモリを用いる方式 [9, 10] に含めている．

次節から文型パターン翻訳方式と類似する方式について個別に説明する．構文トランスファー方式のうち，単文のパターンを使用する多段翻訳方式を 2.2.2 節で説明する．次に統語情報に基づく SMT のうち，原言語に対して句構造解析を行う tree-to-string の SMT について 2.2.3 節で説明する．最後に 2.2.4 節で翻訳メモリを利用する翻訳方式を説明する．

表 2.3: 一般的な翻訳方式

翻訳方式	ベース	構文情報
構文トランスファー方式	ルール	使用
文型パターン翻訳	中間	使用
用例型翻訳	中間	使用
統語情報に基づく SMT	コーパス	使用
フレーズベース SMT	コーパス	なし
翻訳メモリ	コーパス	なし
ニューラル MT	コーパス	なし

2.2.2 多段翻訳方式

文型パターンを用いる機械翻訳方式として，“結合価パターン”を用いた多段翻訳方式 (ALT-J/E) が考案され実装されている [11]．ただし，この翻訳方式は，単文対象の結合価パターンを主変換ルールとして使用する，構文トランスファー方式の一つとみなすべきであり，以下の点において文型パターン翻訳方式と異なる．

1. 重文や複文を対象とする場合は構文解析を用いて文構造を取り出し，結合価パターンの適用場所を選定する．
2. 結合価パターンでは動詞を字面として固定し，単文構造の持つ意味の明示的抽出は行わない．

¹できるだけ他の論文との整合性を図ってはいるが，表 2.3 の分類はあくまで著者の判断による分類である．

3. 様相など話者の主観を表現すると考えられる主体的表現は結合価パターンの中に含めない。

多段翻訳方式では、言語表現は主体的表現と客体的表現の2種類の表現から構成されるとしている。主体的表現は話者の主観的感情や意思を直接表現するもので、日本語では助詞、助動詞等が用いられる。客体的表現は話者により対象化されて捉えられた表現であり、名詞や形容詞の他、事態を表す文の場合もある。多段翻訳方式では主体的表現と客体的表現を分離し、客体的表現の目的言語への変換の後に、英文生成段階で主体的表現を変換された客体的表現に組み込む。主体的表現と客体的表現の分離は、話者の認識の仕方が反映される度合いの違いを元に分離されており、客体的表現にも話者の見方がその表現の構造に反映している。客体的表現に対し、構造への話者の見方の反映の度合いに対し、3レベルの構造変換を適用している。適用される変換方式は、慣用的表現変換方式、意味的結合価パターン変換方式、汎用パターン変換方式である。このうち意味的結合価パターン変換方式で使用されるのが結合価パターンであり、このパターンは単文対象の日英の文パターン対である（表2.4）。

日本語主体的表現は英語との対応関係から、法、時制、相、態などに分類されている。日本語主体的表現の解析により主体的表現の属性を決定し、日英の対応表から英語表現の変換の仕方を決定する。これに対し、文型パターン翻訳では主体的表現は文型パターンの中に含み、英語表現の変換処理は文型パターンの要素である関数として指定する（4.1節と4.2節参照）。

多段翻訳方式の翻訳手順を下に示す。

- 1 形態素解析：分ち書き、単語文法カテゴリの決定
- 2 係り受け解析：文要素間の関係の決定
- 3 J/J変換：日本語内での表現変換
- 4 単文抽出：係り受け解析結果から各用言の支配範囲を決定
- 5 単文解析
 - 5.1 述部解析：様相等を抽出し、平文化する
 - 5.2 名詞句解析：名詞句、複合語の意味構造の決定
- 6 埋め込み文解析：連体修飾等の意味構造を決める
- 7 平文英語変換：パターン対辞書で客体的表現を変換

- 8 接続解析：用言間の関係を決める
- 9 最適結果選択：文全体の多義を絞り込む
- 10 単位文生成
 - 10.1 基本構造生成：英文全体の構造を決定する
 - 10.2 副詞句生成：様相，時制，動詞等から副詞句訳決定
 - 10.3 名詞句生成：句構造と複合語構造の変換，埋め込み文組み込み
- 11 接続構造生成：接続属性，主語の有無等に応じて単位を接続
- 12 様相・時制構造生成：助動詞，不定詞の挿入，語形/構文変形など
- 13 英文調整：縮約，外置変形，冠詞付与，その他

構文トランスファー方式の翻訳手順は主に，解析，変換，生成の3過程に区分できる．上記手順では主に1から6までの手順が解析手順にあたり，7から9までが変換，10から13までが生成過程にあたる．

表 2.4 に結合価パターンの例を示す [12]．表 2.4 中の $N1$ ， $N2$ が名詞変数であり，日本語名詞変数に続く括弧内の“人”，“休暇”，“生活”等が意味属性である．結合価パターンは単文対象のパターンであるため，節間の接続表現はパターンの中に含まない．また主観的表現はパターンの外部で処理するため，様相等の表現もパターンの中に含まれていない．これらに対し文型パターン辞書の文型パターンは，節間の接続表現と様相表現をパターン中に含むことで，重文と複文に対し文型パターンのみで適合が可能になっている．

表 2.4: 結合価パターンの記述例

見出し語	日本語パターン	英語パターン
送る	$N1$ (人)が $N2$ (休暇)を送る	$N1$ spend $N2$
送る	$N1$ (人)が $N2$ (生活)を送る	$N1$ live $N2$
起こす	$N1$ (人)が $N2$ (人)を起こす	$N1$ wake $N2$
起こす	$N1$ (人)が $N2$ (訴訟)を起こす	$N1$ file $N2$

2.2.3 tree-to-string の SMT

構文解析結果を利用する翻訳方式に、統語情報に基づく統計翻訳や用例型翻訳が存在する。これらのうち、文型パターン翻訳と類似点の大きい翻訳方式に、統語情報に基づく SMT の一つである、tree-to-string の SMT (TTS-SMT) がある [13, 14, 15]。TTS-SMT では、原言語側であらかじめ句構造解析を行い、その解析結果と目的言語表現との間で、対となる翻訳ルールを作成する。翻訳ルールは対訳学習文から自動で抽出し、原則として比較的短い翻訳ルールの組み合わせにより、原言語文の句構造を組み立てると共に目的言語文を生成する。生成された目的言語文は木構造になっており、多数の木構造の中から、確率値により最尤の候補が翻訳文として選択される。

対訳文から直接翻訳ルールを抽出することと、原言語側の文構造の抽出と目的言語側への構造変換が同時に行われる点で、文型パターン翻訳と類似する。ただし SMT システムとして翻訳ルールの網羅性を重視し、フレキシブルに翻訳ルールの作成と適用を行う点で、正確性を重視し文型パターン対の厳格な適用を行う文型パターン翻訳とは異なっている。

一般に TTS-SMT の翻訳ルールは非終端記号と終端記号の両方を含んでよいいため、非終端記号化可能な要素を限定することで、文型パターンに類似する翻訳ルールが作成可能である。その場合は翻訳ルール適用の網羅性が減少し、網羅性を重視する TTS-SMT の基本設計と反するため、一般に非終端記号化の品詞による制限等は行われていない。

本論文では文型パターン翻訳方式との翻訳精度の比較に、TTS-SMT のシステムの一つである Travatar を利用する。また 9.2 節で、travatar で作成される文構造と文型パターンの違いを検討する。

2.2.4 翻訳メモリを用いるシステム

対訳文対などの形式で集積された多言語資源を機械翻訳に利用する場合に、そのような言語資源を一般に翻訳メモリ（または翻訳用例）と呼ぶ。もしほとんどの入力文に直接適合するほど大量に翻訳メモリを収集できれば、翻訳メモリの適合だけで精度の高い翻訳文が得られるであろう。しかし語の組み合わせ数の爆発により、実際の翻訳メモリの適合率はごくわずかである。そのため翻訳メモリを機械翻訳に利用する場合に、入力文と完全一致する対訳データが存在する場合は、その対となる目的言語文を出力文とし、部分的に一致しない箇所がある場合は、一致しない箇所だけ局所的に翻訳し、目的言語文の対応箇所に挿入する方法等がとられている。この場合は文型パターン翻訳と類似するが、目的言語への変換処理

等の規定を持たないため、語形変換等の処理が必要な場合に、適切な訳出文を得ることが難しい。

第3章 非線形言語モデル

本研究で使用する重文複文文型パターン辞書は、非線形言語モデルに基づいて作成されている。言語の非線形性は一般に認められているが（3.2節）、機械翻訳においては文の分解と合成（要素合成原理と呼ぶ）が必要となるため、言語の線形性を前提にした処理が機械翻訳の中心になっている。そして各翻訳方式がそれぞれの方法を用いて、要素合成原理の問題を緩和するための処理を組み込んでいる。その中で非線形言語モデルは、文構造の非線形性（3.3節）に着目して考案された機械翻訳（言語処理）方式である¹。

本章では非線形言語モデルについて説明する。まず言語の非線形性を述べた後、重文と複文への非線形言語モデルの適用について説明する。

3.1 言語の規範としての意味

言葉や文の「意味」を定義することは難しく、いまだ「意味」の統一された定義は存在しない。本論文では「文の意味」を、「その文を読み書きしたときに、話者または読者において主観的に生成または理解された事態や認識」とおく。文は個人によってその都度理解されるので、全ての人間において完全一致する「文の意味」は存在しないであろう。その一方で、文が意思の伝達のために使用されていることは事実であり、話者の認識の伝達を可能にする社会的に共有される規範としての側面を、言語が有していると考えられる。池原は[16]の論文において、言語表現によって生起される話者の認識を言語表現の「本来の意味」、主体間での共通認識を可能にする社会的に共有された言語規範を「規範としての意味」として区別している。文の「本来の意味」を計算機で扱うことは、現在の技術においては極めて困難であるが、文の「規範としての意味」を処理することは、現在でも十分に可能である。文の「本来の意味」そのものは扱わなくても、原言語の表

¹一般に機械翻訳方式は要素合成処理が必要となるため、線形言語モデルとも呼べるがこのような呼び方は通常行われぬ。非線形言語モデルは文構造の非線形性に注目することから“非線形”と名付けられている。非線形言語モデルは、要素合成原理の問題を構文トランスファー方式よりも積極的に緩和する方式であり、要素合成処理そのものはやはり必要である。

現を持つ「規範としての意味」を目的言語での表現に変換できれば、人間による「本来の意味」の生成が可能と考えられる。

各言語体系における文法は、社会的に共有される言語規範を体系化して明示したものであり、「規範としての意味」を計算機で扱うには、何かしらの方法で、文法を計算機で処理することが必要と考えられる。構文解析も自然言語での文法を計算機で扱うための手法の一つとみなすことができる。文型パターンは、構文解析とは異なる方法で、文法規則に基づく文構造を抽出する。

3.2 言語の非線形性

一般に、分割するごとに新たな性質の要素となり、これらの性質を足し合わせても元の性質とはならないこと、逆に要素が新たに加わるごとに全体として新たな性質としてあらわれることを非線形性と呼ぶ[1]。一方、線形性は、各要素の性質の足し合わせが全体の性質となることである。言語表現の非線形性は、ある表現に他の語が加わるたびに、各語の意味の合算に還元できない表現全体の意味が新たに生成することと定義できる。同時に、言語表現の線形性は、各語の意味の合算が表現全体の意味となることと定義できる。

新たな意味の生成の際、語と語を結合するための規則と共に、結合後に新たに生成する意味にも規則性がある。非線形言語モデルでは、語の結合ごとに生成する意味のうち、話者の対象認識のあり方が、最終的に文構造の形式で表現されているとみなしている[16]。確かに文の意味そのものは、一語でも消去や置き換えを行うことで異なる意味に変容する。しかし文構造の保存する、対象認識のあり方のような抽象的意味は、文型パターンを用いることで変容することなく抽出できると考えている。そして他の語への置き換え（または抽象化）により、文構造の持つ意味が変容する要素を非線形要素、変容しない要素を線形要素と定義している(3.3節)。文型パターンは線形要素と非線形要素の区別を元に作成される。

3.3 線形要素と非線形要素の定義

文型パターンは、大別して、抽象化される要素と抽象化されない要素の二種類の要素からなり、文型パターンを作成するには要素の抽象化の基準を定める必要がある。非線形言語モデルでは、文構造においても非線形性が存在すると仮定し、文構造の持つ意味の保存の視点から、要素の抽象化の基準を定めている。具体的には、他の要素への置き換えにより文構造の持つ意味が変化してしまう要素は抽象化せず、置き換えによっても文構造の持つ意味は変化しない要素を抽象化する(具体例による説明は3.4節)。

言語表現の非線形性により，文は要素に分解するごとに，それぞれの要素が新たな意味を生成しており，文の意味は分解により変容してしまう．一方，文の意味は各要素への分解だけでなく，要素の品詞への抽象化によっても変容することから，文の分解と要素の抽象化を類似の操作とみなすことが可能である．そこで非線形言語モデルでは，抽象化により文構造の持つ意味が変容する要素を非線形要素，抽象化により変容しない要素を線形要素と定義している．本来は文の要素に線形性と非線形性の区別は付けられないが，文の意味ではなく文構造の持つ意味を対象にすることと，文構造の持つ意味を文型パターンで抽出することから，非線形要素と線形要素の区別が可能となっている．

3.4 文構造の持つ意味の具体例

非線形言語モデルでは，単語だけでは表現することのできない事態や話者の認識の仕方を表現する機能を，文構造が「意味」として持つと考えている．文構造の持つ意味を「富士山は大山より高い。」と「Mt. Fuji is higher than Mt. Daisen.」の例で説明する．

この対訳文対は，一見して二つの等質のもの（ここでは「山」）をその性質で比較していることがわかる．そこでこの対訳文対の文構造の持つ意味として，例えば「二者比較」が考えられる [17]．この二つの文が「二者比較」を行っている文とわかるのは，単語だけではなく，文全体での語順や語形による文構造を見ているためである．文型パターンは主に品詞，語順，語形に着目して一文の文構造を抽出する．線形要素を変数（品詞を持つ）に置き換えることで，この対訳文対から，「<名詞 1> は <名詞 2> より <形容詞 3>。」と「<名詞 1> is <形容詞 3 比較級> than <名詞 2>。」の日英文型パターンが得られる．日英文型パターン共に，文の一部の要素を抽象化したため，元の文の意味そのものは保存されていない．しかしながら「二者比較」のような抽象的な意味は保存されている．日本語文型パターンでは，名詞と形容詞が特定の助詞“は”と“より”によって，この語順で連接されることにより，「二者比較」の意味が保存されている²．英語文型パターンにおいても，名詞と形容詞がbe動詞“is”と接続詞“than”によってこの語順で連接されることにより，「二者比較」の意味が保存されている．この例の日本語文では，抽象化された要素<名詞 1> や<形容詞 3> 等に置き換えられた“富士山”や“高い”等が線形要素にあたり，置き換えられなかった要素“は”や“より”等が非線形要素にあたる．英語文でも同様に置き換えられた要素が線形要素であり，置き換

²“は”を“が”に置き換えても「二者比較」は保存される．ただしこの置き換えにより，「主語の特定」という別の意味が生じる．このように類似の用法を持つ助詞であっても，置き換えによって何らかの文構造の持つ意味の変化を生じさせることが多いため，助詞の抽象化は行っていない．

えられなかった要素が非線形要素である。実際の日英機械翻訳への利用のためには、線形要素の変数化だけでなく、活用形を持つ単語の汎化、品詞活用形の指定、冠詞の削除等の処理が必要となる。

文型パターンを用いて翻訳を行えば、原言語の文型パターンを各入力文に適合させ、対となる目的言語の文型パターンを用いることで、文構造の持つ意味を保存したまま構造変換が達成されることになる³。そして英語文構造の持つ意味と局所的に翻訳された各単語の意味により、原言語文と同様の意味が訳出英語文において生成される⁴。

3.5 重文複文への非線形言語モデルの適用

重文と複文では、節と節の関係に話者の対象認識のあり方が最もよく現れると考えられる。文型パターン辞書では、日本語重文複文における文構造の持つ意味を、節間関係を規定する表現構造⁵に注目して、統語的意味的分類として付与している[5]。重文と複文は複数の節を持つため、一文の文構造の持つ意味を、例えば「二者比較」のような単純な表現で表すことが難しい。そのため統語的意味的分類は統語的分類（重文と複文の区別）、節の意味属性、節間意味分類、節間キーワードの4項目に分けて記述している（4.5節）。節の意味属性は、節の表現構造の持つ意味ではなく、述部の用言の意味属性が付与されている。

一方、文型パターンを翻訳に直接利用する場合、変数に適合した要素の局所的な翻訳と挿入のため、変数化可能な要素は原言語-目的言語間で直接対応する要素に限られる。文型パターンは翻訳への直接への利用も目的としたため、対訳文によっては、日英単独では線形要素として記述可能な要素が、非線形要素として記述されている場合が存在する（4.3節）。

³原言語の文型パターンの適合により、文に対する構文解析のような複雑な処理なしに文構造が抽出でき、対となる目的言語の文型パターンの使用により、文構造の抽出と目的言語への構造変換が同時に達成される。

⁴機械翻訳で生成するのは英語文で、その意味の生成は人間による。

⁵表現構造とは非線形性を持つ言語表現一般であり、文構造、節構造、句構造等を含む。

3.6 線形要素の抽象度と文構造の持つ意味の抽象度

3.6.1 変数の意味属性による細分化と文構造の持つ意味の抽象度

変数は品詞情報を持つが、品詞情報だけでは、文構造の持つ意味を文型パターンが適切に保存できない可能性がある⁶。そのため変数には、より細分されたカテゴリー分類である意味属性を付与し、翻訳の際に、文型パターン適合のための制約に利用する（4.2節，5.4節）。文型パターンが保存する文構造の持つ意味の抽象度は、変数の抽象度に相関して変化する。したがって階層性を持つ意味属性の使用により、文型パターンが保存する文構造の持つ意味の抽象度を、段階的に変化させることが可能である。

3.6.2 線形要素の範囲と文構造の持つ意味の抽象度

重文複文を対象にする場合、単語や句のみでなく節も線形要素としてみなすことが可能である。同じ対訳文から作成した文型パターンでも、変数化範囲により文構造の持つ意味の抽象度が変化する。全体としてみると変数化の範囲は、節変数、句変数、単語変数の順に大きく、文型パターンの抽象度もこの順に大きい。変数化範囲の増大による文型パターンの抽象度の増大は、個々の文型パターンの適合可能性を増大させるが、同時に、文構造の持つ意味を文型パターンが適切に保存できない可能性も増大させると考えられる。変数化範囲の大きさと、文型パターンが保存する文構造の持つ意味の抽象度の関係は、まだ明らかになっていない。

3.7 線形要素と非線形要素の入れ子構造

重文と複文を対象とする文型パターン辞書では、重文と複文においては、節と節の関係に話者の対象認識のあり方が最もよく現れると考え、節間の関係を規定する表現に着目して文型パターン化が施されている。その結果、線形要素には節と句も含まれている。ここで、表現構造を分割するごとに各要素の意味も変容して行くことを鑑みると、節や句も文から分離されたことで節や句単独の意味を持つことになる。そしてそれら節や句の表現構造の持つ意味も新たに立ち現れると考えられる。このことより、文では線形要素とみなされた要素も、その内部にさ

⁶変数は他の要素に置き換え可能な要素であるが、どのような要素でも置き換え可能なわけではない。例えば「富士山は大山より高い」で“富士山”を“立山”に置き換えても文構造の持つ意味は変わらないが、“私”に置き換えるとそもそも文として意味がよく理解できなくなる。

らに非線形要素と線形要素の入れ子構造が現れることになる[18]。したがって重文と複文の表現構造を，図 3.1 のように図示することができる。

句レベルと節レベルの文型パターンを翻訳に使用する場合は，句変数と節変数に対応する要素を局所的に翻訳する必要がある。上記の考えに基づき句レベルと節レベルの文型パターン翻訳方式では，句と節の表現構造の持つ意味を保存して局所的な翻訳ができるように，文型パターンの記述から句パターンと節パターンを抽出して翻訳に使用する。句パターンの抽出と句レベル文型パターン翻訳については7章で，節パターンの抽出と節レベル文型パターン翻訳については8章で説明する。

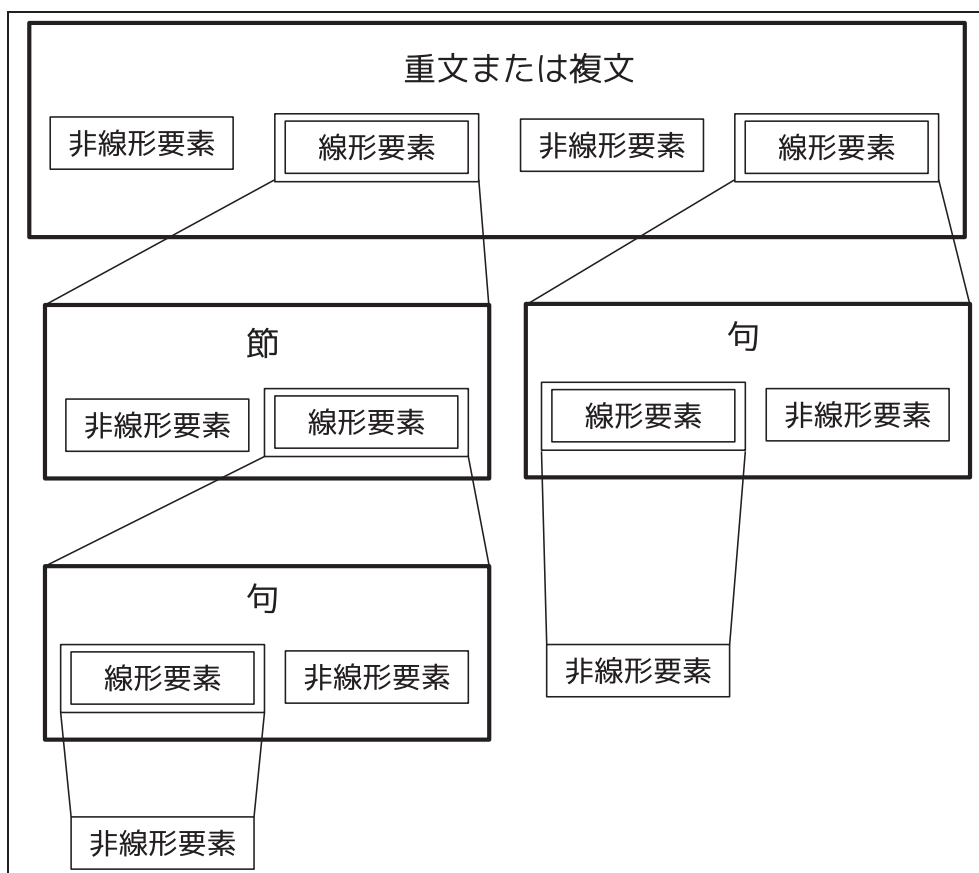


図 3.1: 重文と複文の表現構造

第4章 重文複文文型パターン辞書

非線形言語モデルに基づき、重文複文文型パターン辞書が作成されている。その目的を二つに大別すると、一つは複雑な文構造を持つ重文と複文に対し、文構造の持つ意味を明示的に抽出し、言語分析や機械翻訳に利用可能とすることである。文構造の持つ意味は統語的意味的分類として記述されている。もう一つは、文構造の持つ意味を保存する文型パターンを作成し、この文型パターンを用いて機械翻訳を行うことである。

重文複文文型パターン辞書では、述部を2つか3つ持つ日本語文とその英訳文約12万文対を対象に、文型パターン化と統語的意味的分類の付与が施されている。対訳文の収集はさまざまな辞書の例文から行っている¹。文型パターンでは線形要素は変数に変換されており、変数化の範囲に合わせ、単語レベル、句レベル、節レベルの文型パターンが作成されている。文型パターン数は、単語レベルで約12万、句レベルは約8万、節レベルでは約2万5千である。文型パターン辞書は主に次の3つの構成部位からなる。

1. 日英文型パターン
2. 統語的意味的分類
3. 英語構文分類

本章ではこれら3つの構成部について説明する。また統語的意味的分類を利用して日英間の文構造の対応付けを行う、意味的等価変換方式を説明する。最後に、翻訳に利用するために行った、英語文型パターンの記述の変換について説明する。

4.1 文型パターン作成の基準

重文複文文型パターン辞書は、言語分析への利用と機械翻訳への利用の二つを目的として作成されている。統語的意味的分類は節間の関係に着目して付与項目

¹収集に利用した辞書の詳細は、論文 [19] の表 12 を参照。約 15 万文対から重複する対訳文対や不適切と判断された対訳文対を削除している。

を設定しており，節を変数化した節レベル文型パターンが統語的意味的分類の形式と最も整合性が高い．しかし変数化要素を日英で直接対応する要素に限定したため，日英で節と節で直接対応する要素の選定が必要となり，この選定は容易ではない．そのため文型パターン作成は，日英間での対応箇所を決定しやすい単語変数のみを持つ単語レベル文型パターンから始め，順次，句レベル，節レベルの順に作成された．

単語レベル文型パターンの作成において，日本語線形要素は接続詞を除く自立語とし，それ以外の要素を日本語非線形要素としている [3]．英語線形要素は日本語線形要素と直接対応する要素としており，具体的には名詞，代名詞，定冠詞，動詞，形容詞，副詞がこれにあたる．英語非線形要素は前置詞，疑問詞，接続詞となる．線形要素ごとに変数化され，各変数の日英の対応は同一の変数番号で取られる．翻訳への利用のため，日英で直接対応する線形要素のみを変数化している (4.3 節)．

日本語非線形要素のうち，様相を表現する助詞等は英語用言の語形に対応することが多い．この場合，日本語文型パターンではこの日本語表現を様相関数に変換し，英語文型パターンでは変数に語形関数を付与している．また，主に日本語文型パターンにおける適合範囲拡大のため，日英文型パターンに各種記号を付与している．以上より文型パターンは変数，記号，関数，字面で記述されており，このうち変数が線形要素，関数と字面が非線形要素にあたる．表 4.1 にこれら要素の特徴をまとめる．

句レベル文型パターンは，あらかじめ定められた句変数化の範囲の定義に従い，単語レベル文型パターンを元に作成する．同様に節レベル文型パターンは，節変数化の範囲の定義に従い，句レベル文型パターンを元に作成する．句変数化と節変数化の範囲を表 4.2 に示す．句変数化の基準は非常に複雑なため，一部の基準と例のみを示す．句変数化は日英間で句と句が一对一で対応する場合のみ，節変数化も同様に節と節が一对一で対応するときのみ可能である．この基準により対象とする範囲が大きいほど変数化の基準が厳しくなり，そのため単語変数，句変数，節変数の順に変数化された箇所が減少する．結果として句変数を持つ句レベル文型パターンの数は，単語レベル文型パターンよりも減少し，さらに節変数を持つ節レベル文型パターンの数は句レベル文型パターンよりも減少している．

変数化された範囲が大きいほど，文構造の抽象度が高いことになり，全体としてみると単語レベル，句レベル，節レベルの順に抽象度が高い．文構造の抽象度が高いほど適用範囲が増大するが，その分，文構造の持つ意味を保持できない可能性が高まる．文構造の抽象度が高く，それでお文構造の持つ意味が保存されていることが理想である．潜在的には節レベルが最も理想に近いが，文構造の持つ意味を保持できる度合いは明らかになっていない．

表 4.1: 本論文で使用した文型パターンの要素の説明

要素	線形	非線形	説明
変数	○	×	入力文の任意の要素と適合可能.
関数	×	○	表記の揺れの回収や語形の指定等の機能を持つ.
字面	×	○	文構造の決定に重要な要素.
記号	なし	なし	適合範囲の拡張の機能を持つ.

表 4.2: 句と節の変数化の範囲

対象	例	範囲の定義
名詞句	試験の結果 大きな山	格助詞と名詞の連鎖した表現を対象. 用言に修飾された名詞も名詞句に含む.
動詞句	公園を散歩する	主格を除く1つ以上の格要素を伴った用言について、動詞とその配下にある格要素すべての範囲の表現を対象.
形容詞句	非常に美しい	動詞句の場合に準じる.
形容動詞句	とても複雑な	動詞句の場合に準じる.
副詞句	徒歩で	名詞句の場合に準じる.
節	彼は毎日公園を散歩する	用言とそれに付随する格要素、副詞要素の範囲とし、様相等の関数類は対象外.

4.2 文型パターンの具体例

文型パターンの具体例を表 4.3 に示す. 各文型パターン中の変数に続く「(符号, …)」等の記述は, 意味属性であり, パターン照合や英文生成では除去してから使用する. これらの記述はパターンの絞込みに利用する². 日本語文型パターン中の“/”は挿入可能な要素を持つ文節記号であり, 本来はこの直後に離散記号が続く. 離散記号の使用により大幅に適合する文型パターンの数が増大するが, 英語文型パターンに対応する要素をもたないために, 翻訳での使用が困難である. 本論文では使用しないため記述していない. またプログラム上の処理のために, 元の文型パターン辞書から記述を変更した場合があります (4.7 節), 本論文では変更後の表記で文型パターンを記載する.

表 4.3 中の日本語文型パターン原文と英語文型パターン原文を元に 3 レベルの文

²文型パターン辞書ではコード化されて記述されている. 本論文ではコード化する前の記述で示す.

型パターンが作成されている。単語レベル日本語文型パターンを例にとると、N1 や V4 等が変数であり、“が”、“なつてから”等が字面である。変数は入力文の任意の要素に適合可能であるが、品詞の制約があるため例えば名詞変数 N1 には名詞単語のみ適合する。“#1{”から“,”を挟んで“}”までの記述が順序任意記号である。語順は文構造の持つ意味を決定する重要な構造であるため、文型パターンは記述要素の順序が厳密に決定されている。ただし日本語では、規則に従う範囲で順序の入れ替えが可能である場合が多く、この事象に対応するため順序任意記号が使用されている³。この例では“,”までの“N1 が”と“,”の後ろの“N2 に”の記述が入れ替え可能なことを示しており、例えば“信号が青に”と“青に信号が”のどちらでもこの要素に適合可能である。動詞変数 V4 に続く“^meirei”と“.meireigo”が関数である。“V4^meirei”が動詞 V4 の命令形が適合可能なことを示しており、“V4.meirei”が V4 の後に命令表現が続くことを示している。“(要素1|要素2)”はどちらの要素でも適合可能な要素選択記号であり、この例では上記関数と合わせ、例えば“渡れ”と“渡りなさい”のどちらでもこの要素に適合する。

次に単語レベル英語文型パターンを例にとると、V4 や N3 が変数であり、“after”と“turn”が字面、“^base”が関数である。日本語文型パターンが入力文に適合させるためのパターンであるのに対し、英語文型パターンは日本語文型パターンとの対応を示すためのパターンである。文型パターンを翻訳に使用する場合は、英語文型パターンでの変数の品詞や関数はその品詞や語形への変換を示しており、例えば“V4^base”は日本語変数 V4 に適合した日本語要素を、英語動詞の原形に変換することを示している。文型パターンで使用される記述要素は多数あり、全ての要素を説明することは本論文の範囲を越えるため、本論文に記述した要素のみ、説明を付録の表にまとめる⁴。

表 4.3 の例において、日本語原文は第一節と第二節が助詞“てから”で接続された時間の継起を表す文であること、そして文末表現“なさい”による命令文であることがわかる。それに対応して英語原文は接続詞“after”で時間の継起が表現され、動詞原形“Cross”で文が始まることで命令文であることが表現されている。3 レベルの文型パターン共に、日本語文型パターンにおいては字面“てから”と関数“^meirei”および“.meireigo”により表現されている。英語文型パターンでも 3 レベル共に、字面“after”で時間の継起が、文頭の“変数^base”で命令形が表現されている。単語レベルから節レベルにかけて、変数化された要素はより抽象化されているが、時間の継起の表現と命令文であることは 3 レベルの文型パターン共に保持されている。

³同様の目的の記号に移動可能記号がある。

⁴その他の記述については [20] の「意味類型パターン記述言語仕様書」を参照。

表 4.3: 文型パターンの記述例

文 ID	AC000213-00
日本語文型パターン原文	信号が青になってから道路を渡りなさい。
英語文型パターン原文	Cross the street after the traffic light turns green.
単語レベル	/#1{/N1(符号, 標識, その他多数) が,/N2(色彩) に }
日本語文型パターン	/なつてから/N3(道路) を/cf(V4(状態, 属性, その他多数) ^meirei V4(状態, 属性, その他多数).meireigo)。
単語レベル	V4^base N3 after N1 turn N2 .
英語文型パターン	
句レベル	/N1(符号, 標識, その他多数) が VP2(属性, 因果関係, その他多数)(てから でから)/(VP3(状態, 属性, その他多数)^meirei VP3(状態, 属性, その他多数) .meireigo)。
日本語文型パターン	
句レベル	VP3^base after N1 VP2 .
英語文型パターン	
節レベル	/CL1(属性, 因果関係, その他多数)(てから でから)
日本語文型パターン	/(CL2 状態, 属性, その他多数)^meirei CL2(状態, 属性, その他多数).meireigo)。
節レベル英語文型パターン	CL2^base after CL1.

4.3 日英の構造変換のための変数化

文型パターンは、日英間での文構造変換が直接可能なように作成されているため、日英で直接対応する線形要素のみを変数化している。結果として、日英単独での文型パターン化では変数化可能な要素が、実際の文型パターンでは一部字面として記述されている。表 4.4 に、日英単独で文型パターンを作成した場合と実際の文型パターンの変数化（日英セットで変数化）の例を示す。表 4.4 では簡略化のため、変数化のみを行った結果を示している。なお表 4.4 中の日英単独の場合の変数番号は、日英間の対応は持たない。

日本語原文の“にっこり 笑つ”と対応するのは、英語原文の“a big smile”であり、日英の句として対応をとることができる。しかし単語としての対応を見ると、“にっこり”と“笑つ(笑う)”それぞれに対応する英語単語を確定することが困難である。以上により、表 4.4 の単語レベル文型パターンでは、“にっこり 笑つ”と“a big smile”が変数化されずに字面として記述されている。それに対し日英単独では、日本語文型パターンにおいて“にっこり”が *ADV*、“笑つ”が *V* に、英語文型パターンにおいて“big”が *AJ* に、“smile”が *N* に変数化可能である。一方、日英セットでの文型パターンにおいても、句レベル文型パターンでは“にっこり 笑つ”と“a big smile”の対応を取ることができるので、“にっこり 笑つ”が *VP* に、“a big smile”が *NP* に変数化されている（表 4.4 の句レベル文型パターン）。

表 4.4: 日英単独と日英セットでの変数化の違い

日本語原文	彼女はにっこり笑ってぼくにあいさつした。
英語原文	She greeted me with a big smile.
日英単独	
単語レベル日本語文型パターン	N1 は ADV2 V3 て N4 に V5 た。
単語レベル英語文型パターン	N1 V2 N3 with a AJ4 N5.
日英セット	
単語レベル日本語文型パターン	N1 は にっこり 笑って N2 に V3 た。
単語レベル英語文型パターン	N1 V3 N2 with a big smile .
句レベル日本語文型パターン	N1 は VP2 て VP3 た。
句レベル英語文型パターン	N1 VP3 with NP2.

4.4 文型パターン辞書の統計量

重文複文文型パターン辞書は、日本語重文複文とその英語対訳文の約 12 万文対を対象にして作成されている。この日英対訳文対は、述部を 2 つもしくは 3 つ持つ日本語文とその対訳文を対象に、様々なコーパスから集められている。パターン化の対象となる日本語文は、重文または複文の区別と述部数により、5 種類の文種に分類されている。なお、重文または複文の区別と述部数が、日本語文とその英語対訳文の間で一致しないことも、多々見られる。表 4.5 に日本語文の分類例を示す。

文型パターンは変数化の範囲に応じて 3 レベルで作成されている。文型パターン数は、単語レベルで 121,913、句レベルは 79,477、節レベルでは 25,482 である。

表 4.6 に、重文複文文型パターン辞書における、文種ごとの、文数、日本語文の平均形態素数（句読点を含む）、英語文の平均単語数（カンマとピリオドを含む）を示す。

4.5 統語的意味的分類と英語構文分類

重文と複文の文構造の持つ意味は、統語的意味的分類として記述されている。表 4.7 に統語的意味的分類の項目とその説明、また英語構文分類の説明を示す。表 4.3 の日英原文に対する統語的意味的分類を表 4.8 に示す。

4.6 意味的等価変換方式

4.5 節で示した統語的意味的分類を用いて、多対多の日英表現間の対応付けを行う、意味的等価変換方式が提案されている [5, 21]。統語的意味的分類は、日本語

表 4.5: パターン化の対象となる日本語文の分類例

文種	述部数	説明および具体例	
1	2	説明	文接続を 1ヶ所持つ文 (重文)
		日本語文例	彼は愛車を駆って横浜へ行った。
		英語文例	He drove his car to Yokohama.
2	3	説明	文接続を 2ヶ所持つ文 (重文)
		日本語文例	わたしは給料を日本円でもらっているの で円相場が上がって助かりました。
		英語文例	I am paid in yen, so the rise in the yen exchange rate has helped me.
3	2	説明	埋め込み文を 1つ持つ文 (複文)
		日本語文例	私の父は新しく買ったパソコンを愛用している。
		英語文例	My father loves using his new personal computer.
4	3	説明	埋め込み文を 2つ持つ文 (複文)
		日本語文例	これは人々に愛唱されている古い民謡の一つです。
		英語文例	This is one of the old folk songs people love to sing.
5	3	説明	文接続と埋め込み文を各 1つ持つ文 (重文かつ複文)
		日本語文例	最初の小説の成功に味をしめて 彼は 2 作目を書く決心をした。
		英語文例	The taste of success from his first novel encouraged him to begin a second.

表 4.6: 文種ごとの文数, 日本語文の平均形態素, 英語文の平均単語数

文種	1	2	3	4	5	全体
文数	55509 (45.5%)	5664 (4.6%)	42492 (34.9%)	5638 (4.6%)	12610 (10.3%)	121913
日本語文の平均形態素数	12.1	14.8	12.3	16.6	15.3	12.8
英語文の平均単語数	11.1	14.0	10.3	13.3	13.2	11.3

重文複文における文構造の持つ意味として, 対訳文対ごとに付与されている. 文型パターン翻訳では, 入力文に適合する文型パターンのみによる表現構造の抽出が可能であるが, 真理項 (統語的意味的分類) を媒介とすることで, 意味の同じ文構造をまとめて提示することができると考えられている. 意味的等価変換方式の概念図を図 4.1 に示す. 意味的等価変換方式は機械翻訳への利用と共に, 言語分析に用いるための言語知識の表現方法として考案されている.

日英の文型パターンの間を媒介する真理項は, 意味的等価変換方式の考案当初では, 表 4.9 のような階層性を持つ単一項として考えられていた. しかし日本語重文複文への適用のため, 重文複文文型パターン辞書の作成時に, 表 4.9 のような真理項の代わりに, 統語的意味的分類を付与している. 統語的意味的分類のうち, 節間意味分類と節の意味属性の分類項目を, それぞれ表 4.10 と表 4.11 に示す. 節間意味分類と節の意味属性共に 4 層で構成されているが, 表 4.10 は 2 層まで, 表

表 4.7: 統語的意味的分類の説明

項目	説明
統語的分類	節の数および重文と複文の区別を5種類で分類.
統語構造分類	節間の係り受け構造による分類
第一節の意味属性	節の用言による意味分類. 4階層で構成. “/”で階層を区切り, 抽象度は左が高く右が低い.
第二節の意味属性	同上
第三節の意味属性	同上
節間意味分類	節間の意味的關係. 4層で構成. 一層目が統語的分類, 2, 3層目が機能的分類, 4層目が用いられた表現による分類.
節間キーワード	節間意味分類を決定する日本語表現.
英語構文分類	日英の構文構造の対応関係を記述.

表 4.8: 統語的意味的分類と英語構文分類の具体例

統語的分類	述部二つを持つ重文
統語構造分類	副詞型
第一節の意味属性	現象事象の表現/発生消滅/発生/なし, 変化の現象事象の表現/なし/なし/なし
第二節の意味属性	社会的活動の行為/労働/土木建築業/なし, 移動行為の表現/行き来の移動/なし/なし, その他多数
節間意味分類	副詞節 (副詞的連用節)/時/事態の時/後
節間キーワード	てから
英語構文分類	日本語第一節が英語構文では“after”を用いた従属接続詞相当

4.11 は1層のみを示している.

真理項(統語的意味的分類)を検索項目として, 入力した真理項と一致する文型パターンを出力する, 意味検索プログラムがすでに作成されている⁵. 意味的等価変換方式を言語分析に利用する場合は, 入力文の真理項を人間が決定して検索することが可能であるが, 機械翻訳に使用する場合は入力文の真理項を自動で決定する必要がある. さらに意味検索で得られた文型パターンを用いて翻訳を行うには, 本研究で行う文型パターン翻訳方式とは異なる翻訳方法が必要である. そのため意味的等価変換方式を用いた機械翻訳は, まだ実装されていない. 現時点では, 入力文の真理項の決定と, 意味検索で得られた文型パターンを用いた翻訳を人手で行う, という条件で翻訳可能性が試されたのみである [22].

[22]の論文では, 統語的意味的分類の各項のうち, 最も適切に同じ意味の文構造を抽出できる項の選定が行われており, 従属節(第一節)の意味分類の使用が最も効果的であることが示されている. しかし統語的意味的分類の付与精度や, 項目の記述形式の妥当性等はまだ調査されていない. 統語的意味的分類を, 機械翻訳

⁵意味検索プログラムの“意味”は“文構造の持つ意味”のことである.

または言語分析へ利用するためには、まずこれらの調査が必要である (9.1 節)。

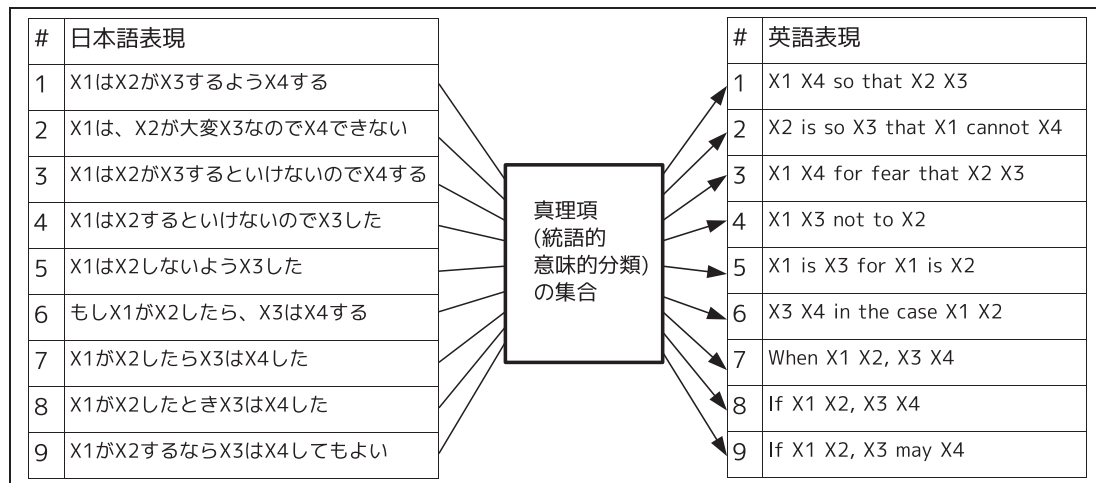


図 4.1: 意味的等価変換方式

表 4.9: 真理項

レベル 1	レベル 2	レベル 3
比較	同一, 比喩, 同関係, 同級, 添加, 同様, 以上, 以下, 換言, 比較級, 対比, 倍数, 相違, 割合, 択一, 最上級複数	漠然, 一般, くどい, 重々しい, 反事実, 限度, 所属, 原因, 譲歩, 性質, 継続, 推量, 感慨, 関連, 嗜好, 状況, 状態, 引用 (説明), 対句的, 代替, 定義づけ, 適正 (忠告, 禁止, 勧誘, 命令), 伝聞, 転換, 結果, 決定, 限定, 使役, 事実, 時点, 自動的, 主題, 十分, 同時, 判明, 反復 (習慣), 比例, その他多数

表 4.10: 節間意味分類の分類項目

#	第一層	第二層
1	補足節, 補語相当節	名詞節 引用節
2	名詞節, 連体修飾節	補足語修飾節 内容節 縮約的修飾節 機能的表現 の型接続 その他
3	副詞節, 副詞的連用節	時 因果関係 条件譲歩 付帯状況・様態 逆接 目的 程度 前提 手段 二者関係 相関 主観感情 場面 その他
4	並列節, 並列型連用節	順接的並列 逆接的並列
計	4 種	24 種

4.7 翻訳に利用するための英語文型パターンの記述の変換

文型パターン辞書の記述そのままでは、翻訳において不適切である事例が見つかっている。表 4.12 に翻訳のための英語文型パターンにおける記述の変換処理を示す。各要素の説明は付録に示す。なお節レベル文型パターンでは多くの記述および処理形式が未定のままとなっており、これについては 8 章で説明する。

表 4.11: 節の意味属性の分類項目

#	第一層	述部の品詞
1	知覚と情緒の表現	動詞
2	知的な行為の表現	
3	日常生活の行為	
4	地域社会生活行為	
5	社会的活動の行為	
6	現象事象の表現	
7	変化の事象の表現	
8	移動行為の表現	
9	対物行為の表現	
A	性状規定の表現	形容詞
B	だ文	形容動詞
計	11 種	3 種

表 4.12: 翻訳のための記述の変換

事項	変換処理例	説明
英語原形字面 品詞変換関数	'eat' → @eat V(N1) → V1	プログラム処理用に変更。 この例の左辺は日本語名詞変数 N1 が英語動詞変数 V1 と対応することを示す。プログラムの処理に合わせ右辺に変更。
英語様相関数	Do V1.not → Do not V1^base	プログラム処理に合わせ、パターンの記述をあらかじめ変更。
不定冠詞	(a an) A/I N2 → A/I N2	形容詞変数の前の不定冠詞を削除。
be 動詞の字面処理	M1 be → M1 @be	出現形字面 be を原形字面@be に変換。
have の字面処理	M1 have → M1 @have	出現形字面 have を原形字面@have に変換。
関数付与 (英)	V1 → V1^present	英語動詞を変数化する場合に現在形を指定する関数を付与。

第5章 非線形言語モデルに基づく日 英文型パターン翻訳方式

5.1 文型パターン翻訳の概要

本論文では、非線形言語モデルに基づく文型パターン翻訳方式（本翻訳方式）を実装する。文型パターンを日英機械翻訳に利用する場合、まず入力文に適合する日本語文型パターンを検索し、文型パターンにおいて変数に対応する部分を局所的に翻訳し、対となる英語文型パターンに挿入することで出力文を得る。文型パターンは変数化範囲の違いに伴い3レベルで記述されているが、この手順は3レベルの翻訳方式いずれにも共通する。3レベルに共通するシステムの全体図を図5.1に示す。また手順とその概要を以下に示す。

1. 入力文の形態素解析を行い、品詞コードと意味属性コードを付与する。
2. 日本語パターン検索システムを用いてパターン照合を行い、入力文に適合する日英文型パターンとその変数情報を得る。
3. 意味属性を用いて、日本語文型パターンの絞込みを行う。
4. 英文生成システムにより、得られた日英文型パターンを用いて変数対応部分を局所的に翻訳し、最尤の訳語の組み合わせを選択することで英語文を生成する。

単語レベル文型パターンは単語変数のみを持ち、単語変数に対応する要素の局所的翻訳には単語辞書を使用する。また句レベル文型パターンは単語変数と句変数を、節レベル文型パターンは単語変数、句変数、節変数を持つ。そのため句レベル文型パターン翻訳では、句変数に対応する部分に対し句パターンを用いて局所的に翻訳し、節レベル文型パターン翻訳では、節変数に対応する部分に対し節パターンを用いて翻訳する。パターン照合では3レベルの文型パターン同時に検索することが可能であるが、翻訳時の処理の違いから、それぞれのレベルごとに一つの文型パターン翻訳方式として実装する。単語レベル文型パターン翻訳は6章、句レベル文型パターン翻訳は7章、節レベル文型パターン翻訳は8章で説明する。

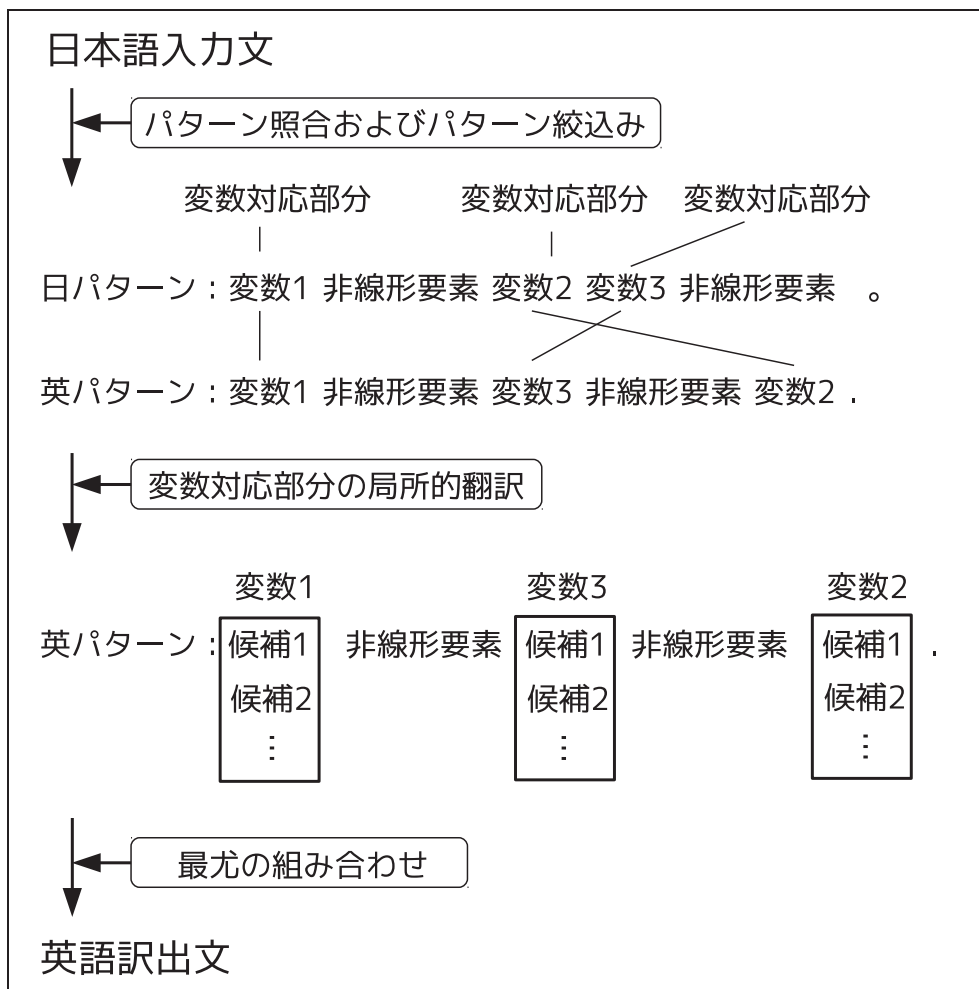


図 5.1: 文型パターン翻訳の概要

5.1.1 文型パターンの適合率

構文トランスファー方式では，できるだけ一般性を持つルールを作成して使用することが指針の一つである．それに対し文型パターン翻訳では，過度の汎化をすることなしに翻訳を行うことを目的とし，各対訳文対ごとに文型パターンの作成を行っている．一つ一つの文型パターンの適合率よりも，大量の文型パターン全体として入力文に適合するいくつかの文型パターンが得られればよいという立場に立っている．そのため本論文ではパターン適合率を，文型パターン集合全体として文型パターンが適合する文が，入力文のうちどれくらいの割合となるかとする．

実際に翻訳に使用する文型パターンは，パターン照合とパターン絞込みを行っ

た後に残っている文型パターンである。したがって、本論文では以後パターン適合率を、入力文に対する、パターン照合とパターン絞込みの後、少なくとも1つの文型パターンが適合している文の割合と定める。

5.1.2 3レベルの文型パターンの組み合わせ方

文型パターン翻訳方式は文型パターンが適合したときのみ翻訳が可能となるため、翻訳性能として、入力文のうち一つでも文型パターンが適合する文の割合が重要となる。文型パターン数が多いほど文型パターンの適合率は高くなるため、文型パターン数のみで判断すると、パターン適合率は単語レベル、句レベル、節レベル文型パターンの順に小さくなる。変数に適合可能な範囲は節、句、単語の順に小さくなるため、一文型パターンあたりの適合可能性は節レベル、句レベル、単語レベルの順に小さくなる。以上のことから実際に調べるまで3レベルの文型パターン間の適合率の大小はわからない。文型パターンの適合率は単語レベルが一番低く、句レベルと節レベルが同程度に高いことが明らかになっている [23]¹。ただし [23] の論文の時点では節レベル文型パターンの数が現在の約半数しか作成されていなかったため、現在は句レベルより節レベルの文型パターンの方が適合率が高いと考えられる。単純に翻訳性能を問題にする場合は、適合率の高い句レベルと節レベルの文型パターン翻訳方式が翻訳精度も高いのであれば、それ単独で用いればよい。逆に適合率の低い単語レベル文型パターン翻訳が翻訳精度が高いのであれば、まず単語レベル文型パターン翻訳方式を適用し、単語レベル文型パターンが適合しなかった文に対し句レベルと節レベルの文型パターン翻訳方式を適用することになる。また異なる翻訳システムよりも翻訳精度が高いのであれば、文型パターンが適合するときは文型パターン翻訳を行い、適合しないときは他の翻訳システムを使用するハイブリッド方式が可能である。

5.1.3 ルールベースかつコーパスベースの翻訳システムとしての指針

2.2.1 節で述べたように、文型パターン翻訳システムは、ルールベースかつコーパスベースの翻訳システムである。文型パターンの作成では、線形要素と非線形要素の区別や、変数化や関数化の基準を人手による言語分析を通して詳細に決定しており、またそれに伴い変数や関数の処理を細かく規定したと合わせ、ルールベースのシステムとみなせる。一方で収集した対訳文対から直接文型パターン

¹離散記号の使用等により今回とは条件がいくつか異なる。

を作成したことと、コーパス内での頻度情報などのデータを用いて訳語選択等を行うことから、コーパスベースのシステムともみなせる。

訳語選択や文型パターン選択にコーパスから得た統計値を利用しているが、適切な選択を行うためには、その目的に適した統計値の抽出と、あらかじめ選択候補から不適切な候補をある程度は除いておくことが必要となる。文型パターン翻訳では、この不適切な候補の削減を手で細かく設定したルールを用いて行っている。ただし手で作成したルールのみで最終候補を一つに決定することは困難であり、最終的には一つの候補が統計値により選択される。そのため統計値による選択の余地ができるように複数の候補を残すことが好ましい。ルールベース的手法とコーパスベース的手法の組み合わせは一般に試みられているが、両手法の結合の効果を得るためには、両手法の比重をヒューリスティックに決定しなければならない。このことは文型パターン翻訳においても同様である。

5.2 形態素解析

現存する機械翻訳方式はほぼ全て、まず入力文に対し形態素解析を行う。日本語の翻訳の場合は、入力文を各形態素の分かち書きの形式に変換する。文型パターン翻訳でもパターン照合の前に形態素解析を行うが、それに加えて形態素の品詞と文節境界の決定が必要となる。またパターン絞込みのために、名詞および用言の形態素への、意味属性の付与が必要となる。

図5.2に、入力文「信号が青になってから横断歩道を進みなさい。」に対する、形態素解析と品詞情報等の付与を行った結果の例を示す。図5.2の1行目において、“信号”は形態素であり、“用言性名詞”が品詞、“{符号, 標識, 電力機器}”は日本語彙大系[24]の意味属性である。各形態素の直前に付してある“/”は、この形態素の直前が文節境界であることを示しており、“+”は文節境界ではないことを示している。文節境界は記号の適合において必要となる。

5.3 文型パターンの照合

形態素解析の後、入力文に対し文型パターンの照合を行い、翻訳に使用する文型パターンを選択する。パターン照合において、日本語文型パターンの全ての必須な要素（字面、変数、関数）が、入力文の形態素と対応する場合に、その日本語文型パターンが入力文に適合する。文型パターンの変数は任意の要素に適合可能な部位であるが、品詞と意味属性を持つため、変数への適合はこれらの制限の元に行う必要がある。本論文では文型パターンの照合において、変数の品詞によ

-
1. /信号 (用言性名詞, { 符号, 標識, 電力機器 })
 2. +が (格助詞)
 3. /青 (一般名詞, { 色彩 })
 4. +に (格助詞)
 5. /なっ (自動詞, 成る, 成っ, { 生成, 属性, その他多数 })
 6. +てから (接続助詞)
 7. /横断歩道 (一般名詞, { 道路 })
 8. +を (格助詞)
 9. /進み (自動詞, 進む, 進み, { 属性変化, 物理的移動, 身体動作 })
 10. +なさい (補助動詞)
 11. +。 (句点)
 12. /nil
-

図 5.2: 形態素解析結果

る制約のみを使用し、意味属性による制約はパターン照合の後に文型パターンの絞込みにより代用する。

文型パターンの照合には、入力文に対し、文型パターンの字面、関数、変数、記号の各要素が適合可能な部位を検出することが必要になる。文型パターン辞書の記述形式に対応可能なパターン照合システムとして、すでに構造型パターン検索システム SPM[25] が作成されており、本研究ではパターン照合に SPM を使用する。SPM は ATN 法 (2.1.2 節) による構文解析機能を有しており、入力文に対し部分的な構文解析を行いながら、適合可能な文型パターンを全て検出する。文型パターンの出力は非線形要素の割合が多い順に出力する。そしてパターン照合により、入力文に適合した日本語文型パターンおよび対応する英語文型パターンと、変数情報を得る。

記号、関数、変数の適合範囲を、構文解析により決定するため、文法規則の記述の仕方により文型パターンの適合結果が変化する。特に句変数と節変数の適合範囲の決定の仕方は、句レベルと節レベルの文型パターン翻訳において翻訳精度と直結している。SPM では適合可能な文型パターンを取り残しなく検索するため、適合のための規則は、文型パターン作成時の変数化の範囲よりも広く取られている。また句変数と節変数に適合した部分の局所的な翻訳は句パターンと節パターンを用いるため、句パターン (7.2 節) と節パターンの照合にも SPM を使用する。

5.3.1 SPMによる文型パターンの照合方法

SPMはATN法(2.1.2節)をベースにパターン照合を行う。パターン照合に使用する文型パターンは全て、あらかじめ遷移ネットワーク化しておく。入力文に対しpopアークまでたどり着いた文型パターン(文型パターンの遷移ネットワーク)をその入力文に適合したとみなし、popアークまでたどりついた文型パターン全てをパターン照合結果として出力する。

図5.3に、句レベル文型パターン「/N1がVP2(てから|でから)/(VP3^meirei|VP3.meireigo)」の遷移ネットワーク(P1)と、文法規則のネットワークの一部を示す。文法規則ネットワーク中の“NN”は名詞，“J”は格助詞を示し，“V^”は動詞でかつ語形の制約があることを示す。文型パターンのネットワークにおけるアークは、変数、字面、関数、記号を持つ。記号により複数の適合が可能な場合は、ネットワーク中では分岐として展開する。例えばこの文型パターンの“(てから|でから)”は、VP2の後に、“てから”と“でから”のどちらかが続くことを示しており、遷移ネットワークP1ではノード5から6の二つのアークで展開されている。

パターン照合では、文型パターンの遷移ネットワークにおいて、前方から後方のpopに向かってネットワークを遷移していく。アーク上に付した変数や字面等がアークを移動するための条件となっており、入力文の要素がその条件を満たすときアークを移動する。アークの条件を満たすかどうかの判断は形態素解析結果の文節境界“/”，形態素，品詞コードを用いて行われる。変数と関数は文法規則としてネットワークの形で記述されており，文型パターンの遷移ネットワークでのアークが変数と関数の場合は，各文法規則ネットワークへ移動する。その文法規則ネットワーク内を遷移しpopまでたどり着けた場合に，変数や関数として正しく受理されたとみなし，文型パターンの遷移ネットワークのアークに戻る。この動作を繰り返し，文型パターンの遷移ネットワークにおいて最終的にpopにたどり着いたとき，その文型パターンが入力文に適合したと判断する。文型パターンのネットワークの途中で遷移に失敗した場合は，その文型パターンが入力文に適合しなかったと判断する。

入力文「信号が青になってから横断歩道を進みなさい。」を例に，この文型パターン(遷移ネットワークP1)の適合の仕方を示す。まずP1の先頭のノード1に入り，文節境界記号(図5.2の1行目の“/”)の一致によりノード2に遷移する。P1のノード2からノード3へのアークN1において，文法規則ネットワークNに移動する。入力文の“信号”の品詞は“用言性名詞”であり(図5.2)，ネットワークNにおけるNNのアークの条件を満たすため，ノード11からノード12に遷移しpopする。この動作により，“信号”をP1のN1として受理し，P1のノード3に遷移する。入力文の次の形態素は“が”であり，P1のノード3からノード4に遷移する。P1のVP2において，ネットワークVPに移動する。同様の手順で，入力文

の“青になっ(成る)”によりVPのノード13からノード16に移動し、popする。名詞“青”がアークのNNの条件を満たすため、ノード13からノード14に移動する。次に格助詞“に”がアークのJの条件を満たすため、ノード14からノード15に移動する。最後に動詞“なっ(成る)”がアークの“V^”を満たすため、ノード15からノード16に移動し、popしてP1のアークVP2に戻る。これらの動作により、“青になっ(成る)”がVP2として受理されて、P1のノード4からノード5に移動する。以下、同様の動作を繰り返すことで、最終的にP1のノード10に移動してpopし、この文型パターンP1が入力文に適合したと判断される。

SPMでは変数や関数として受理された入力文の要素を保存し、変数情報として文型パターンと共に出力する。この例ではN1に“信号”，VP2に“青になっ(成る)”，VP3に“道を渡り(渡る)”が対応するとして出力される。

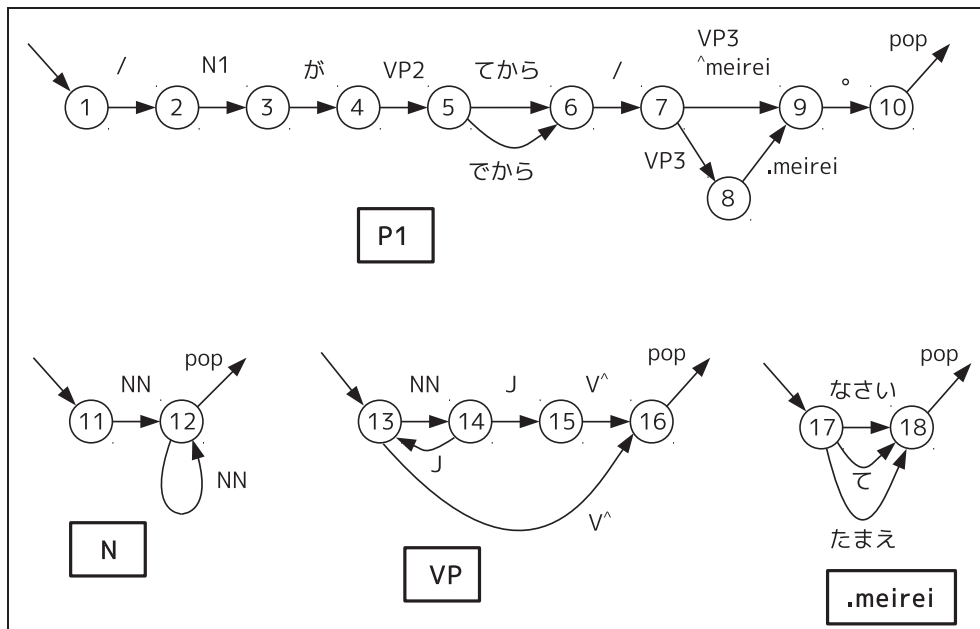


図 5.3: 文型パターンの遷移ネットワークと文法規則の遷移ネットワークの例

5.3.2 文型パターンの照合と構文解析の違い

文型パターンの照合では、文型パターンを遷移ネットワークとして表現するため、形式的には構文解析を行って入力文の文構造を抽出することに類似する。しかしながら、構文解析と比べ、文型パターンの照合には次の違いがある。

1. 重文と複文を受理するための、非常に長い遷移ネットワーク（各文型パターンの遷移ネットワーク）が、多数存在する²。
2. 遷移ネットワークのアークに、字面が多数出現する。

構文解析では、文を受理する遷移ネットワーク（図 2.2 下の例など）はごく少数である。それに対し文型パターンの照合では、文を受理する遷移ネットワークの数は、文型パターン数と同数である。構文解析では比較的小さな構造を組み合わせて文構造を抽出することを目的としている。一方、文型パターンでは、一つの構造で文構造を抽出することを目的としたため、一文全体を受理可能な非常に長い遷移ネットワークとなっている。

また構文解析では、ノード間を遷移するアークには品詞が付されており、その品詞を満たす形態素であればアークを遷移可能である。それに対し文型パターンではアークに字面が付されている場合が多数存在し、その字面の場合のみアークの遷移が可能である。したがって、pop まで遷移して、入力文を受理できる文型パターンの遷移ネットワーク数はわずかである。このことは、個々の文型パターンの適合能力が低いことを示している。

以上より、文型パターンは、一文全体に一つの構造を対応させて文構造を抽出可能となっているが、個々の文型パターンの適合可能性が低く、文型パターン翻訳には大量の文型パターンが必要であることを示している。

5.3.3 文型パターンの照合例

表 5.1 に、図 5.2 の形態素結果に対する、3 レベルの文型パターンの照合結果を示す。表 5.1 では 4.2 節の表 4.3 に示した文型パターンの照合結果のみを示す。表 5.1 において、句レベル文型パターンの変数情報における、“VP2=青になっ(成る){成る}” を例にとると、形態素列“青になっ(成る)” が動詞変数 VP2 に対応していることを示している。また“{成る}” は、この形態素列内部での係り先を示している。

文型パターンにおいて、変数は任意の要素に適合可能な要素であるが、品詞の区別や単語、句、節の区別を持つ。変数に適合可能な範囲は単語、句、節の順に大きくなるため、単語レベル、句レベル、節レベルの順に文型パターンの適合可能性が大きくなる。表 5.2 に入力文と 3 レベルの文型パターンの適合の有無を示す。

²単語レベル、句レベル、節レベル文型パターンの順に遷移ネットワークが長い。節レベル文型パターンでは、構文解析での文を受理するネットワークと長さの変わらない場合も存在する。

表 5.1: パターン照合結果

単語レベル文型パターン	
日本語文型パターン	/#1{/N1(符号, 標識, その他多数) が,/N2(色彩) に }/なってから /N3(道路) を/(V4(状態, 属性, その他多数)^meirei V4(状態, 属性, その他多数).meireigo)。
英語文型パターン	V4^base N3 after N1 turn N2.
変数情報	N1=信号, N2=青, N3=横断歩道, V4=進み (進む)
句レベル文型パターン	
日本語文型パターン	/N1(符号, 標識, その他多数) がVP2(属性, 因果関係, その他多数)(て から でから)/(VP3(状態, 属性, その他多数)^meirei VP3(状態, 属 性, その他多数).meireigo)。
英語文型パターン	VP3^base after N1 VP2.
変数情報	N1=信号, VP2=青になつ(成る){成る}, VP3=横断歩道を進み (進 む){進む}
節レベル文型パターン	
日本語文型パターン	/CL1(属性, 因果関係, その他多数)(てから でから)/(CL2(状態, 属 性, その他多数)^meirei CL2(状態, 属性, その他多数).meireigo)。
英語文型パターン	CL2^base after CL1.
変数情報	CL1=信号が青になつ(成る){成る}, CL2=横断歩道を進み (進む){ 進む}

5.4 文型パターンの絞込み

文型パターンの変数の持つ意味属性による制約は、その制約なしでパターン照合を行い、得られた文型パターンに対し意味属性による絞込みを行うことで代替する。意味属性は日本語語彙大系 [24] の意味属性を用いる。日本語語彙大系の意味属性は、対象の見方や捉え方が、一般名詞意味属性、固有名詞意味属性、用言意味属性として約 3,000 属性に分類、体系化されている。このうち一般名詞意味属性および用言意味属性を、日本語文型パターンの絞込みに用いる。一般名詞意味属性は 12 階層、2,715 分類で、用言意味属性は 4 階層、36 分類で記述されている。

表 5.2: 3 レベルの文型パターンの適合の有無

入力文	単語レベル	句レベル	節レベル
信号が青になってから道を渡りなさい。	○	○	○
信号が青になってから落ち着いてアクセルを踏め。	×	○	○
きょうは午後家に帰ってからあなたの弟が東京へ行きなさい。	×	×	○
明日晴れたら釣りにいこう。	×	×	×

以下の手順で日本語文型パターンの絞込みを行う。

1. 入力文および適合した日本語文型パターンから，変数部分の意味属性を抽出。
2. 入力文の形態素の意味属性を用いて，日本語文型パターンの変数に対する適合範囲を決定。
3. 日本語文型パターンの全ての変数において，各変数の意味属性が各適合範囲に含まれる場合，この文型パターンを翻訳に使用。

日本語語彙体系の意味属性では，階層により抽象度が異なっている³。本論文では，文構造の持つ意味の決定に，変数の意味属性の抽象度が大きく作用していると仮定した。そのため意味属性による適合範囲の決定には，ルートノードの方向にたどる階層を決定し，その階層の子孫を適合範囲とした。変数の種類により，親ノードをたどる階層を変えている。単語レベル文型パターンに対するパターン絞込みは6.3節で，句レベル文型パターンに対するパターン絞込みは7.5節で説明する。

5.5 英語訳出文の生成

変数に対応する要素の局所的翻訳と訳語選択を行い，英語訳出文を得る。本研究では文型パターン辞書の記述に対応した英文生成システムを作成した。英文生成システムは主に次の三つの機能を実装しており，この順に従って英文を生成する。

1. 変数に対応する要素の局所的翻訳。
2. 関数に従った変換処理。
3. 複数翻訳候補からの訳語選択。

5.5.1 変数に対応する要素の局所的翻訳

3レベルの文型パターン翻訳において，変数に対応する要素の局所的翻訳に使用する辞書等のデータベースを，表5.3にまとめる。句レベル文型パターン翻訳では，単語レベル文型パターン翻訳で使用するデータベースに加え，句変数に対応する日本語要素の翻訳に句パターンを使用する。また節レベル文型パターン翻訳ではさらに，節変数に対応する日本語要素の翻訳に節パターンを使用する。

³厳密に抽象度で階層分けを行ったわけではないので，同じ階層でも場所によりある程度の抽象度の違いがある。

句変数と節変数に対応する日本語要素の翻訳には、単語レベル句型パターン翻訳方式を再帰的に利用する。単語辞書の抽出と使用方法は6.4節と6.5節で、句パターンの抽出とその利用は7.2節と7.6節で説明する。節パターンについては8.1節で概要を示す。

表 5.3: 変数に対応する要素の局所的翻訳に使用するデータベース

レベル	データベース
単語レベル	単語辞書, 限定詞辞書, 活用形辞書, 英語品詞タガー (品詞タグ付け器)
句レベル	句パターン辞書, その他単語レベルと同様
節レベル	節パターン辞書, その他句レベルと同様

5.5.2 関数に従った変換処理

英語変数に関数が付与されている場合、その関数に指示される語形に局所翻訳結果を変換する。プログラム上では、英語活用形辞書により複数の語形と英語品詞情報を得て、その品詞情報をもとに関数に指定された語形以外の訳語候補を削除する。

関数の処理は単語変数、句変数、節変数の違いによって異なる。単語変数に対する関数処理は6.4節で、句変数に対する関数処理は7.6.2節で説明する。節変数に対する関数処理は現在検討中であり省略する。

5.5.3 複数訳語候補からの訳語選択

句型パターンを用いた日英翻訳では、パターン照合における変数に対応する要素を局所的に翻訳し、英語句型パターンの変数部分にそれらの結果を挿入することで翻訳文を得る。変数に対応する要素の翻訳で得られた複数の候補に対し、単語翻訳確率および単語連鎖確率 (tri-gram) による訳語選択を行い、英語文を生成する。

一文全体に対し、(5.1)式に従って確率値を計算し、最尤の単語組み合わせを一つ選択する。tri-gram値が存在しない3単語連鎖には(5.5)式のペナルティー値を与える。単語翻訳確率は、変数化された日英要素の対全体から、各単語の出現頻度を用いて計算する(5.4)式のスコアで代用する。単語辞書に複数単語で登録されている場合は、その複数単語を一つの単語とみなしてスコアを計算する。式中の e

は英語単語, j は日本語単語を示す. n はその英語文の単語数であり, 英語複合表現や冠詞の有無等により可変である. $C(e, j)$ は日英単語対の共起出現回数, $C(j)$ は日本語単語 j の出現回数, $C(e)$ は英語単語 e の出現回数である. t は総 3 単語連鎖数を示す.

$$\operatorname{argmax}_e P(e)P(j|e) \quad (5.1)$$

$$P(e) = \prod_{i=1}^n P(e_{i+2}|e_i e_{i+1}) \quad (5.2)$$

$$P(j|e) \simeq \prod_{i=1}^n \operatorname{Score}(j_i, e_i) \quad (5.3)$$

$$\operatorname{Score}(j_i, e_i) = \frac{C(e_i, j_i)}{C(j_i)} \times \frac{C(e_i, j_i)}{C(e_i)} \quad (5.4)$$

$$\operatorname{Penalty} = \frac{1}{\exp(3)} \times \frac{1}{t} \quad (5.5)$$

句変数と節変数に対応する要素の翻訳には, 単語レベル文型パターン翻訳を再帰的に利用するため, 3レベルの翻訳方式共に上式で計算する. 英語文型パターンは, 変数の語形を指定する関数を持ち, また不定冠詞の削除等が行われている. そのため英語単語における語形や冠詞の有無を選択するための処理を行った後に訳語選択を行う.

5.6 複数文型パターンが適合した場合の処理

通常の機械翻訳タスクは, 1つの入力文につき1つの出力文を得ることである. また入力文の前後の情報等は与えずに翻訳することが一般的である. 本論文でも通例に従い, 一入力文につき一出力文を決定する. 一入力文に複数の文型パターンが適合した場合は, 複数の文型パターンを用いて複数の翻訳候補文を得る. その後に (5.1) 式に示した確率値が最大の翻訳候補文を一つ出力文として選択する.

5.7 文型パターン翻訳方式のこれまでの取り組み

文型パターン翻訳方式はまず名詞句と動詞句を対象に, 文型パターンの記述形式に対応する翻訳機能が実装された [26, 27]. そしてそれらの結果を元に, [28] の論文により文型パターン翻訳方式の試作が行われた. [28] の論文ではクローズドテストが行われ, 評価は英語訳出文の SVOC の関係の構築のみを行っている. また英語パターン表記の不適切さや訳語選択の問題が指摘されている.

使用する文型パターンの絞込みとして, 用言の意味属性の使用による方法が試みられており, 絞込みの効果が確認されている [29]. 句レベル文型パターンを対象

に，多変量解析を用いた最適な文型パターンの選択が試みられており，その効果が確認されている [30].

句レベル文型パターン翻訳と節レベル文型パターンへの利用を視野に入れて，単語レベル文型パターンから句パターン辞書の抽出が行われている [31, 32]. そして句レベル文型パターン翻訳の試作が行われたが，文型パターンと句パターンの適合率の低さと，翻訳精度の低さが報告されている [33].

その他，文型パターン翻訳方式における意識効果の評価 [34] や，離散記号の使用を想定した格助詞相当句の抽出と翻訳 [35] も行われている. [34] の論文では文型パターン翻訳による意識の出力可能性が確認された. [35] の論文により，副詞相当表現への離散記号の適用と翻訳への利用の可能性が示唆された⁴.

⁴英語文型パターンへの挿入位置の決定が未解決のため本研究では利用していない.

第6章 単語レベル文型パターン翻訳方式

6.1 単語レベル文型パターン翻訳方式の概要

本章では単語レベル文型パターン翻訳方式について説明する。システム全体のモデル図を図6.1に、手順とその概要を以下に示す。

1. 入力文の形態素解析を行い、各形態素への品詞と意味属性を付与する。
2. 日本語パターン検索システムを用いてパターン照合を行い、入力文に適合する単語レベル日英文型パターンとその変数情報を得る。
3. 意味属性を用いて、日本語文型パターンの絞込みを行う。
4. 単語レベル英文生成システムにより、単語レベル日英文型パターン、単語辞書を用いて英語文を生成する¹。
5. 単語翻訳確率と単語連鎖確率を用いて、最尤の文を出力文として選択する²。

6.2 日本語文型パターンの照合

パターン照合に構造型日本語パターン検索システム SPM[25] を使用する。入力文「彼の姉がああ美しいとは予想しなかった。」の形態素解析結果を図6.2に示す。この入力文に対するパターン照合により得られた、日英文型パターンと変数情報を表6.1に示す。

¹[28] の論文では、電子辞書から抽出した情報を元に単語辞書を作成し、翻訳に使用している。本研究では、文型パターンの作成段階で抽出された、変数化された要素の対を元に単語辞書を作成し、翻訳に使用している。

²[28] の論文では、単語連鎖確率のみを用いている。

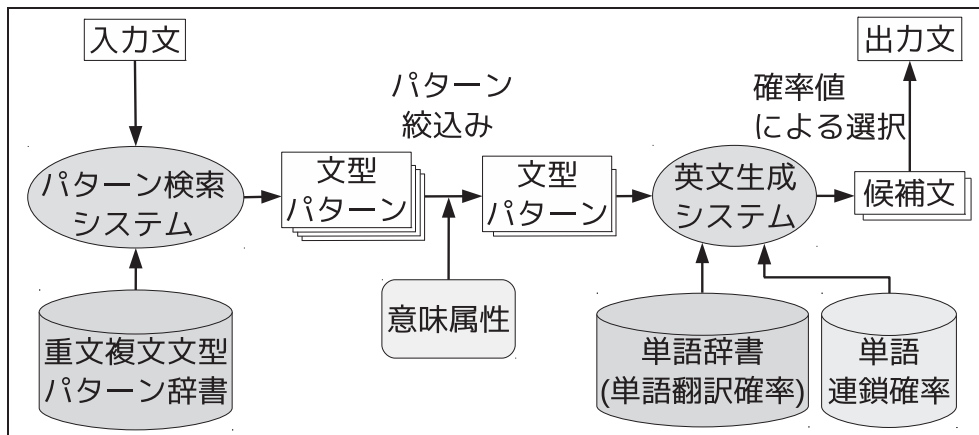


図 6.1: 単語レベル文型パターン翻訳方式

表 6.1: パターン照合結果

日本語文型パターン	</N1(人間)は>/N2(他称単数男, 男)の/N3(母, 女)が/ああ /AJ4(属性)とは/V5(思考動作, 感情動作).hitei.kako.
英語文型パターン	<I N1> never V5 ^{past} N2 ^{poss} N3 to be so AJ4 .
変数情報	N2=彼 { 他称単数男, 男 }, N3=姉 { 姉, 女 }, AJ4=美しい { 属性 }, V5=予想し (予想する) { 思考動作 }

6.3 日本語文型パターンの絞込み

日本語文型パターンの変数を持つ、意味属性による適合に対する制約は、文型パターン適合後のパターン絞込みにより行う。以下の手順で日本語文型パターンの絞込みを行う。

1. 入力文および適合した日本語文型パターンから、変数部分の意味属性を抽出。
2. 入力文の形態素の意味属性を用いて、日本語文型パターンの変数に対する適合範囲を決定³: 形態素が名詞の場合は、その意味属性が5層以下のとき、4層目の親属性とその子孫属性全てを適合範囲とする。4層以上のときはその属性と子孫属性全てを適合範囲とする。用言の場合は、その属性と子孫属性

³ごく少数ではあるが、複合名詞などにおいて、複数の形態素が一つの単語変数に適合する場合が存在する。その場合は変数に対応する形態素列中の係り先の形態素の意味属性を使用する（詳細は7.5節参照）。

-
1. /彼 (人称代名詞, { 他称単数男, 男 })
 2. +の (格助詞)
 3. /姉 (一般名詞, { 姉, 女 })
 4. +が (格助詞)
 5. /ああ (一般副詞型名詞)
 6. /美しい (形容詞, { 属性 })
 7. +と (格助詞)
 8. +は (副助詞)
 9. /予想し (他動詞サ変活用, 予想する, { 思考動作 })
 10. +なかつ (助動詞打消, ない)
 11. +た (助動詞過去)
 12. +。 (句点)
 13. /nil
-

図 6.2: 形態素解析結果

全てを適合範囲とする。形態素が複数の意味属性を持つときは、それぞれの意味属性で適合範囲を求め、それら全ての和を適合範囲とする。

3. 日本語文型パターン of 全ての変数において、各変数の意味属性が各適合範囲に含まれる場合、この文型パターンを翻訳に使用。

本論文では、文構造の持つ意味の決定に、変数の意味属性の抽象度が大きく作用していると仮定している。さらに一般名詞意味属性において、文構造の持つ意味の決定に最も関与すると思われる抽象度を、4階層目とした。

上記手順2の例を、表6.1の変数情報“N3=姉”の意味属性“{ 姉 }”を対象に、図6.3に示す。名詞形態素“姉”の意味属性“{ 姉 }”は第9層にあり、4層の“{ 人 }”までたどり、その子孫の意味属性全てが適合範囲となる。表6.1の日本語文型パターンの変数“N3”は、意味属性“{ 母 }”を持つことから、この適合範囲に含まれる。

6.4 単語レベルの英文生成

文型パターンを用いた日英翻訳では、パターン照合における変数に対応する要素を局所的に翻訳し、英語文型パターンの変数部分にそれらの結果を挿入することで翻訳文を得る。単語レベルの英文生成システムでは、単語辞書を用いて要素を翻訳し、単語翻訳確率および単語連鎖確率 (tri-gram) による訳語選択を行い、

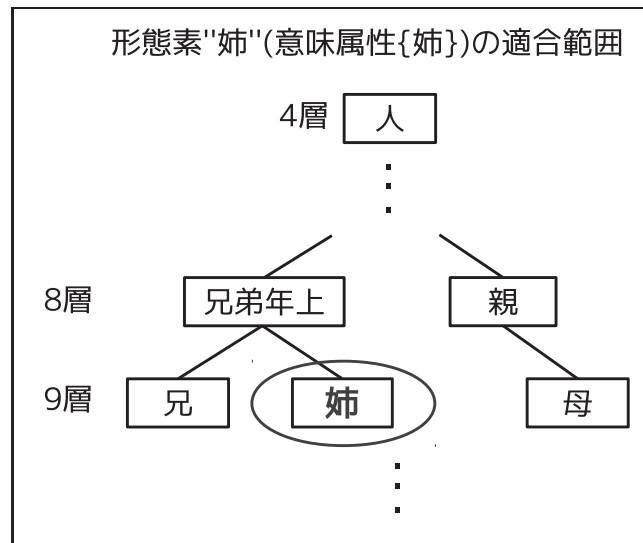


図 6.3: 意味属性による名詞単語変数の適合範囲の例

英語文を生成する。単語辞書は、文型パターンの作成段階で抽出される、日英の変数化された要素の対から作成する。日英の単語共に複合的表現（複合名詞、群動詞等）を含む。単語辞書は6.5節で説明する。

一文全体に対し、式(5.1) (5.5.3節)に従って確率値を計算し、最尤の単語組み合わせを一つ選択する。

英語単語における語形や冠詞の有無を選択するために、以下の処理を行った後に訳語選択を行う。

1. 変数に対応する形態素の翻訳（日本語形態素の検索と対応する英語単語の抽出）の際、英語文型パターンの変数と同じ品詞を持つ英語単語を訳語候補とする。
2. 活用形を持つ日本語形態素の検索は原形で行い、得られた英語単語が活用形を持つ場合は、活用形辞書を用いて全活用形を訳語候補とする。ただし英語変数に関数が付与されている場合は、関数に指示される活用形のみ訳語候補とする。
3. 英語文型パターンにおける変数が名詞または形容詞の場合は、日本語形態素の検索で得られた英語単語に冠詞を付与したのも訳語候補とする。名詞変数の場合はさらに複数形も訳語候補とする。

6.5 単語辞書

翻訳に使用する単語辞書は、文型パターンの作成段階で抽出される、日英の変数化された要素の対から作成する。文型パターン作成において、日英の変数化する要素の間で品詞の区別の制限は設けていないため、日英の変数間の品詞が異なる場合が多い（単語レベル文型パターン中の変数で約15%）。一般の電子辞書のデータから作成した単語辞書では、日英で品詞の異なる場合の単語は少ないため、結果として未知語が増加する。また記述形式の統一がとれていない等の問題も存在するため、本論文では単語辞書も文型パターンから作成した。日本語単語変数の品詞は10種類であり、英語名詞変数の種類は4種類である。単語辞書の作成は日本語変数の品詞の種類に合わせて10種類である。作成した単語辞書の例を表6.2に示す。

表 6.2: 単語辞書の例

日本語単語 (品詞)	英語単語 (品詞)	単語翻訳確率
蹴る (動詞)	hit(動詞)	0.01
	kick(動詞)	0.29
	stamp out(動詞)	0.05
	turn down(動詞)	0.06
翻訳 (名詞)	translation(名詞)	0.53
	translate(動詞)	0.20
入国する (動詞)	enter(動詞)	0.01
	enter the country(動詞)	0.33
	get through customs(動詞)	0.17
	entrance(名詞)	0.03

6.6 単語レベル文型パターンによる英文生成の例

入力文「彼の姉がああ美しいとは予想しなかった。」の翻訳例を図6.4に示す。翻訳には表6.1の文型パターンとその変数情報を用いる。

変数に対応する形態素に対し、単語辞書を用いた翻訳を行い、複数の訳語候補を得る。例えば、 $V5^{\wedge}past$ に対応する形態素は「予想する」であり、単語辞書から得られた訳語候補に関数を適用した結果、「predicted, expected, …」が得られる。同様に $N2^{\wedge}poss$ に対し「his, he's」が得られる。 $N3$ に対する翻訳結果は「sister, elder

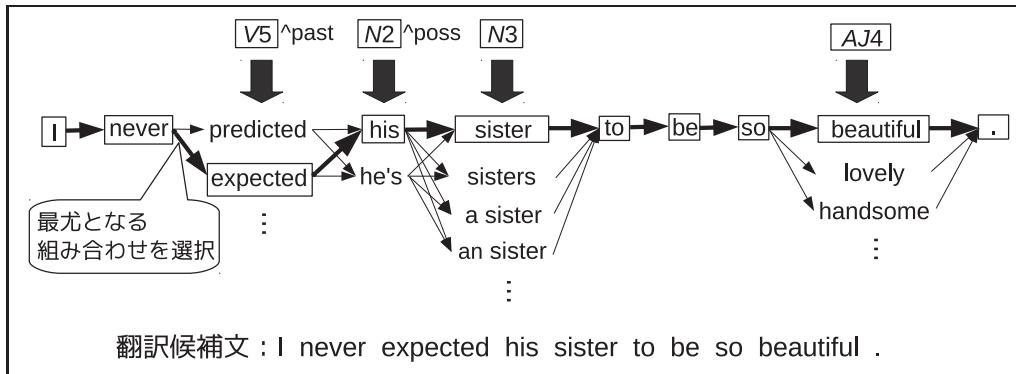


図 6.4: 単語レベル文型パターンによる英文生成の例

sister, ...」であり、訳語候補は冠詞を付与した候補を加え、「sister, a sister, an sister, the sister, the sisters, elder sister, ...」となる。AJ4 に対しては「beautiful, lovely, handsome, a beautiful, ...」が得られる。そして確率計算により最尤となる単語組み合わせを選択し、翻訳候補文「I never expected his sister to be so beautiful.」を得る。

パターン絞込み後に複数文型パターンが得られている場合は、全ての文型パターンを用いて翻訳を行い、確率値が最大の翻訳候補文を一つ、出力文として選択する。

6.7 翻訳実験

本翻訳方式は、非線形言語モデルに基づいて作成された文型パターンを用いることにおいて、機械翻訳における初の試みである。そのためパターン適合率 (5.1.1 節) や、文型パターンが適合した場合にどれだけの翻訳精度が得られるかは未知数であり、まずこれらの調査を行う。

6.7.1 実験条件

単語レベル文型パターン翻訳方式の有効性を検証するため、入力文に重文複文文型パターン辞書の日本語文型パターン原文を用いて翻訳実験を行う。文型パターン辞書の日英文型パターン約 12 万から抽出した、10 万の日英文型パターンを翻訳実験に使用する。この 10 万文型パターン対から抽出した、変数化された要素の対を用いて、単語辞書の作成と単語翻訳確率の学習を行う。

単語連鎖確率（単語 tri-gram 値）の学習は，この英語文型パターン原文 10 万文を用いて行う．学習に使用しない残りの日本語文型パターン原文約 2 万文から，5,000 文を抽出しテスト文に用いる．

6.7.2 比較システム

翻訳精度の比較には，構文解析により文構造を抽出する翻訳システムのうち，tree-to-string の SMT（TTS-SMT）を用いる．TTS-SMT のシステムには Travatar[13] を使用する．市販の構文トランスファー方式によるシステムは，構文解析結果や翻訳過程での処理方法が明確ではない．そのため構文解析結果と最終的に使用された英語文構造がわかり，かつ処理方法が明確に示されている Travatar を比較に用いる．

形態素解析と日本語句構造解析には，Travatar で推奨されている KyTea[36] と Ckylark[37] を用いる．翻訳モデルの学習は，文型パターン辞書の日英原文 10 万文対（単語レベル文型パターン翻訳方式の実験で使用する文型パターンの原文対）を用いる．言語モデルの学習は，この英語文型パターン原文 10 万文を用いて行う．パラメータはデフォルト値を使用する．パラメータチューニングには，文型パターン辞書の対訳文 1,000 文対を用いる．

6.7.3 作成された単語辞書

作成された単語辞書の日本語見出し語数を表 6.3 に示す．名詞が約 61% と最も多く，次いで動詞が約 13%，形容詞が約 12% であった．この 3 種類の単語辞書で約 86% を占めていることがわかる．

6.7.4 パターン適合率

入力 5,000 文のうち，パターン照合と意味属性を用いたパターン絞込みに成功した文は，572 文であった．よって本実験では，パターン適合率は約 11% (572/5,000) となる．この値は，10 万の単語レベル文型パターンでは，パターン適合率がまだ低いことを示している．

6.7.5 翻訳精度の評価

文型パターン翻訳では，一文全体に適合する文型パターンが得られたときのみ翻訳を行う．そのため本実験では，文型パターン翻訳方式と比較システム共に，文

表 6.3: 単語辞書の日本語見出し語数

品詞	見出し語数	割合 (%)
<i>N</i> (名詞)	33,601	61
<i>ND</i> (用言性名詞)	682	1
<i>NUM</i> (数詞)	1,801	3
<i>TIME</i> (時詞)	1,108	2
<i>REN</i> (連体詞)	534	1
<i>GEN</i> (限定詞)	7	0
<i>V</i> (動詞)	7,114	13
<i>AJ</i> (形容詞)	6547	12
<i>AJV</i> (形容動詞)	1,541	3
<i>ADV</i> (副詞)	1,816	3
総計	54,751	100

型パターンが適合した文のみを対象として評価を行う。

重文複文文型パターン辞書において、日本語原文が同じでも対訳英文が異なる場合は、異なる対訳文対とみなしている。したがって学習文が、入力文と同じ日本語原文を含んでいる場合がある。文型パターンが適合した572文中、学習文に同じ日本語文を持つ文は193文であり、持たない文は379文であった。上記379文の翻訳結果に対し、BLEU[38], TER[39], METEOR[40], RIBES[41]の4つの評価法を用いて自動評価を行った⁴。自動評価の結果を表6.4に示す。表6.4より、全ての評価法において単語レベル文型パターン翻訳方式が比較システムを上回っていることがわかる。

表 6.4: 自動評価結果

	BLEU	TER	METEOR	RIBES
単語レベル	0.290	0.574	0.568	0.771
比較システム	0.217	0.641	0.489	0.756

次に上記379文からランダムに100文抽出し人手評価を行った。自動評価では、その値から翻訳精度の程度を読み取るのは困難である。単語レベル文型パターン翻訳方式の翻訳精度を明示するため、adequacyによる5段階の絶対評価を行う。また人手評価値は単語レベル文型パターン翻訳方式の問題点の分析に利用する。評

⁴TERは誤り率(Error rate)を用いた評価法であり、値が低いほど翻訳精度が高いことを示す。

評価者は日本語ネイティブの3名であり、評価者Aが大学院生準拠レベル、評価者BがTOEIC高得点者（800点以上取得）、評価者Cが翻訳家である。評価者には評価例、入力文、出力文、参照文を提示する。出力文はどちらの翻訳システムの結果かわからないように提示する。評価の際は、辞書、インターネット等を自由に参考にしてよい。

人手評価の基準と評価例を表6.5に示す。また人手評価の結果を表6.6に示す。表6.6において、各評価値における数値は、その評価値の文数を示している。また人手評価値の2次重み付きカッパ係数を表6.7に示す。

表6.7より、どの評価者の間でもカッパ係数は $\kappa = 0.6$ 前後であり、人手評価値の一致度は比較的高い。表6.6の評価者平均の結果から、単語レベル文型パターン翻訳方式では翻訳精度の高い文（評価4または5）が約半数（51.0文）を占める。また単語レベル文型パターン翻訳方式の平均値は3.44であり、比較システムの2.54を大きく上回っている。

表 6.5: 人手評価の基準とその例

評価5	入力文の意味を正しく理解できる。
入力文	彼の言うことはいくらか本当だ。
翻訳結果	what he says contains some truth .
参照文	What he says, is true to a certain extent.
評価4	一部不適切な部分があるが、概ね理解できる。
入力文	資本主義と社会主義は相入れない概念だ。
翻訳結果	capitalism and socialism is different idea .
参照文	Capitalism is incompatible with socialism.
評価3	入力文の意味が何となく読み取れる。
入力文	バレーボールをやって突き指をした。
翻訳結果	i did some volleyball and sprained finger.
参照文	While playing volleyball, I sprained one of my fingers.
評価2	部分的な理解にとどまる、または意味が入力文と大きく異なる。
入力文	先生のお陰で大学に入る事が出来ました。
翻訳結果	the teacher 's to enter the university .
参照文	thanks to my teacher, I was able to get into the university.
入力文	その病院は駅からあまり遠くないところにある。
翻訳結果	the hospital is close to the station.
参照文	The hospital is not far from the station.
評価1	入力文の意味をほとんど理解できない。
入力文	なんでもいたしますが、それだけはいたしかねます。
翻訳結果	everything i am , but it is made .
参照文	I would do anything but that.

表 6.6: 人手評価結果

	評価 5	評価 4	評価 3	評価 2	評価 1	平均
評価者 A						
単語レベル	44	6	17	23	10	3.51
比較システム	18	3	14	36	29	2.45
評価者 B						
単語レベル	19	23	20	23	15	3.08
比較システム	13	4	16	23	44	2.19
評価者 C						
単語レベル	31	30	23	13	3	3.73
比較システム	12	21	29	30	8	2.99
評価者平均						
単語レベル	31.3	19.7	20.0	19.7	9.3	3.44
比較システム	14.3	9.3	19.7	29.7	27.0	2.54

表 6.7: カッパ係数の値

	評価者 A-B	評価者 A-C	評価者 B-C
単語レベル	0.597	0.599	0.568
比較システム	0.751	0.675	0.599

6.7.6 翻訳結果の具体例

表 6.8 に単語レベル文型パターン翻訳方式の人手評価の値が高い例を示す。人手評価の低い例は、原因分析の後、表 6.12 と表 6.13 に示す。表中の「参照文」は入力文の対訳英語文である。「日本語文型パターン」と「英語文型パターン」は、出力文の生成に使用した日本語文型パターンと英語文型パターンであり、「日本語文型パターン原文」と「英語文型パターン原文」は、この文型パターンの作成元となった日本語原文と英語原文である。比較のため表 6.8 の下 2 行に、同入力文における比較システムの出力文とその人手評価値を示している。

表 6.8 の例では、入力文と日本語文型パターン原文共に、「行為 A をして行為 B した」の日本語文構造であり、かつ行為 B が行為 A の目的となっている文である。そして行為 B を不定詞句で表現した英語文型パターンを用いることにより、人手評価 5 の結果が得られている。

これに対し比較システムの出力文は、文法的な間違いを含むため評価値は低い。

表 6.9 に比較システムで使用された翻訳ルールの一部を示す。1 番目のルールの“x0:動詞 p”が2番目のルールで展開され、2番目のルールの“x0:動詞 p”が3番目のルールで展開されている。どの翻訳ルールも正しいルールであるが、この入力文においては3番目のルールの使用が不適切である。文接続表現の“て”が、2番目のルールにおいて“x0:動詞 p”の中に含まれていることが原因の一つと考えられる。このことは日本語構文構造の問題を示唆している。

表 6.8: 翻訳結果の例

入力文	店員は娘を言いくるめてその指輪を買わせた。
参照文	The salesman inveigled the girl into buying the ring.
出力文	the clerk coaxed her to buy the ring .
人手評価値	評価者 A:5, 評価者 B:5, 評価者 C:5
日本語文型パターン	/\$1^{}/M1(自称単数女, 他多数) は }/N2(他称単数男, 男) を /V3(精神的移動)(て で)\$1/GEN4/N5(企業, 他多数) を / (V6(属性, 他多数).sase.kako V6(属性, 他多数)^sase.kako)。
英語文型パターン	M1 V3^past N2^obj to V6^base DEF4 N5.
変数情報	N1=店員, N2=娘, V3=言いくるめ(言いくるめる), GEN4=その, N5=指輪, V6=買わ(買う)
日本語文型パターン原文	私は彼を説得してその家を買わせた。
英語文型パターン原文	I convinced him to buy the house.
比較システム の出力文	the clerk and tried her daughter to buy the ring .
評価値 (比較システム)	評価者 A:2, 評価者 B:3, 評価者 C:2

6.7.7 文長等の統計量

文型パターンが適合した場合における、入力文の文長（文字数）を表 6.10 に示す。表 6.10 より、文型パターンが適合した入力文の文長は、文型パターン辞書の平均の文長より小さいことがわかる。文型パターンの適合条件は入力文が長くなるほど厳しくなるため、文型パターンの適合した文では、短い入力文の割合が多く

表 6.9: 翻訳ルールの例

翻訳ルール (日本語ルール → 英語ルール)	
1	root(動詞 p(x0:動詞 p 補助記号 (“。 ”))) → x0 “.”
2	動詞 p(動詞 p(助詞は p(名詞 (“店員”) 助詞は (“は”)) x0:動詞 p) x1:動詞 p) → “the” “clerk” x0 x1
3	動詞 p(x0:助詞を p 動詞 p(x1:動詞 助詞で (“て”))) → “and” x1 “her” x0

なつたと考えられる。また人手評価を行った各入力文ごとに平均の人手評価値を求め、文長との相関係数を求めた。単語レベル文型パターン翻訳方式では $r = -0.04$ 、比較システムでは $r = -0.16$ となり、両システム共に人手評価値と文長との相関はみられなかった。以上より、短い入力文の方が文型パターンが適合しやすいが、文型パターンが適合した場合は、文長は翻訳精度にほとんど影響を与えていないことがわかる。

次に使用された文型パターンの種類数を調べたところ、評価を行った 100 文中、98 種類の文型パターンが使用され、そのうち 2 つの文型パターンが 2 回ずつ使用されていた。このことより、多数の入力文に使用される特定の文型パターンの存在は確認されなかった。

表 6.10: 日本語文の文長 (文字数)

	平均文長 (文字数)
パターン適合 (379 文)	15.64
評価文 (100 文)	15.66
文型パターン辞書 (121,913 文)	22.59

6.7.8 比較システムと併用した場合の翻訳精度

単語レベル文型パターン翻訳方式の翻訳精度は、文型パターンが適合した場合は高い。したがって他のシステムと併用すれば、そのシステム単体よりも翻訳精度が向上する。文型パターンが適合したときは単語レベル文型パターン翻訳方式を用い、適合しなかったときは比較システムを用いた場合の自動評価結果 (5,000 文) を表 6.11 に示す。表 6.11 の “ハイブリッド” が単語レベル文型パターン翻訳方式と比較システムを併用した場合の結果である。全ての自動評価法において、若干ながら評価結果が向上していることがわかる。

表 6.11: 自動評価結果 (ハイブリッド)

	BLEU	TER	METEOR	RIBES
ハイブリッド	0.139	0.751	0.688	0.435
比較システム	0.133	0.757	0.686	0.427

6.8 考察

翻訳実験の結果より，単語レベル文型パターン翻訳方式はパターン適合率は低いが，文型パターンが適合した場合は翻訳精度が高いことが明らかになった．しかしパターン適合率の低さにより，他のシステムと併用したときの効果はわずかである．単語レベル文型パターン翻訳方式の効果を上げるには，文型パターンが適合したときの翻訳精度をさらに向上させることと，パターン適合率を増加させることが必要である．

本節ではまず翻訳精度が低い原因を調査し，翻訳精度とパターン適合率の改善方法を示す．次に追加実験を行い，言語モデルの学習文を増加した場合と，単語翻訳確率を使用しなかった場合の翻訳精度を調査する．

6.8.1 翻訳精度が低い原因

文型パターンが適合したときでも，評価者平均において約半数の 49.0 文の翻訳精度が低い（評価 1 から 3）．単語レベル文型パターン翻訳方式において翻訳精度が低い理由は，文型パターンに関連する問題と，変数に対応する形態素の翻訳に関連する問題の二つに大別できる．

人手評価を行った 100 文のうち，3 人の評価者において評価が高い（評価 4 または 5）か低い（評価 1 から 3）かが一致した文は，61 文であった．61 文の内訳は，評価の高い文が 31 文，低い文が 30 文であった．この 61 文に対し，使用した英語文型パターンが適切であったか調査した．適切な英語文が得られる可能性があるときを“適切”，そうでない場合を“不適切”と判断した．61 文中 44 文が適切であり 17 文が不適切であった．評価の高い 31 文は全て 44 文に含まれており，不適切な文型パターンを用いて翻訳精度の高い文は存在しなかった．

以上より翻訳精度の低い 30 文のうち，文型パターンに関連する問題が 17 文，変数に対応する部分の翻訳に関連する問題が 13 文（44 - 31 文）である．前者の問題による 17 文のうち，文型パターン記述ミスや形態素解析失敗等の理由によるものを除くと，不適切な文型パターンの使用は 7 例であった．表 6.12 に，この 7 例のうち，複合的な表現が分割されて変数に対応したことによる例を示す．

入力文の“情の細やかな(細やかだ)”は複合的な表現であり、単一の表現で言い換えるなら、例えば“繊細な”である。この文型パターンでは、“情”と“細やかな(細やかだ)”が別々の単語変数に対応しているため、不適切な訳語選択により不適切な出力文が得られている。これら7文は、パターン絞込みにおける意味属性の使用方法の変更により、翻訳精度が改善される可能性がある。また句レベル文型パターンの使用により、適切な出力文が得られる可能性がある。

後者の問題による13文は、主に未知語と不適切な訳語の選択による。単語辞書の見出し語の増加や、訳語選択におけるスムージング手法の最適化等により、ある程度の改善が可能である。表6.13に不適切な訳語選択により翻訳精度の低い例を示す。表6.13の例では、入力文中の“出す”の訳語が“opening”となっており、この入力文においては不適切である。この例では例えば“giving”が適切であり、訳語候補にこの語が存在するが、確率値による選択により“opening”が選択されている。また出力文中でbe動詞の複数形“are”が選択されており、この例では“is”が適切である。単語連鎖確率は3単語の連鎖確率のみを使用しているため、be動詞直前の“orders”との関係により複数形の“are”が選択されたと考えられる。この例でもスムージング手法の最適化等により適切な訳出文が得られる可能性は残るが、本論文で使用した訳語選択方法では、適切な訳出文が得られる可能性は非常に低いと思われる。

表 6.12: 複合的な表現の問題により翻訳精度の低い例

入力文	彼女は情の細やかな人だ。
参照文	She is a warm and considerate woman.
出力文	She has tinier sensibilities .
人手評価値	評価者 A:2, 評価者 B:3, 評価者 C:3
日本語文型パターン	/N1(他称単数男, 男) は/N2(関心, 他多数) の/AJV3(属性)^rentai!(人 ひと ヒト)だ。
英語文型パターン	N1 @have^present AJ3 N2 .
変数情報	N1=彼女, N2=情, AJV3=細やかな(細やかだ)
日本語文型パターン原文	彼は趣味の多方面な人だ。
英語文型パターン原文	He has many-sided interests.

表 6.13: 不適切な訳語選択により翻訳精度の低い例

入力文	命令を出すことは私の権限ではない。
参照文	It is outside my province to give orders.
出力文	the opening of my orders are not my province .
人手評価値	評価者 A:1, 評価者 B:2, 評価者 C:3
日本語文型パターン	/N1(欲) を/V2(属性, 属性変化, その他多数)^rentai!ことは #1[/N4(自称単数女, 対称単数, その他多数)の]/N5(目的, 方面).#da は (ない 無い)。
英語文型パターン	N2 of N4^poss N1 @be not #1[N4^poss] N5 .
変数情報	N1=命令, V2=出す, N4=私, N5=権限
日本語文型パターン原文	食欲を満たすことは私の目的ではない。
英語文型パターン原文	The gratification of my appetite is not my aim.

6.8.2 意味属性コードを用いたパターン絞込みの効果

意味属性コードを用いたパターン絞込みを行わなかった場合、パターン照合で少なくとも一つ文型パターンが適合した文は、1,030 文であった。このうち学習文と同じ日本語文を持たない文は 837 文であり、この 837 文の自動評価結果を表 6.14 に示す。表 6.4 の単語レベル文型パターン翻訳の結果と比べ、表 6.14 では全ての評価法で翻訳精度が大きく下回っている。この結果は、意味属性コードによるパターン絞込みにより、適切な文パターンを選択していることを示している。ただし不適切な文型パターンを取り除いたことにより、翻訳可能な入力文が大幅に減少する副作用が生じている（パターン絞込みなしの 1,030 文からパターン絞込みありの 572 文に減少）。

表 6.15 に、パターン絞込みにより、適切な文型パターンまで取り除かれ翻訳されなかった文の割合を示す。表 6.15 では、パターン絞込みにより文型パターンが一つも適合しなかった 458 文（1,030 文から 572 文を除いた文）を対象に、パターン絞込みを行わなかった場合の出力文に使用された文型パターンが、各入力文の翻訳に適切であるかを調査した。調査文は 458 文から 100 文をランダムに抽出した。表 6.15 の結果から、パターン絞込みにより、半数近く（47/100 文）の場合で、適切な文型パターンを用いた出力文まで取り除いていることがわかる。したがってパターン絞込みは適切な文型パターンを選択する効果を持つと共に、過剰な選

択によりパターン適合率を不必要に減少させていることがわかる。そのため単語レベル文型パターン翻訳方式での意味属性コードによる適合範囲の決定の仕方が、不適切である可能性がある。意味属性コードを用いてできるだけ不適切な文型パターンだけを取り除くには、意味属性コードによる適合範囲をより精密に決定する必要があると考えられる。

表 6.16 に、パターン絞込みにより不適切な文型パターンが取り除かれた例を示す。表 6.16 の出力文は、パターン絞込みを行わなかった場合の出力文である。入力文は「動作の結果」を表現する文だが、「同時進行的な行為」を表現する英語文型パターンが使用されたため、翻訳精度が低くなっている。

表 6.17 に、パターン絞込みにより適切な文型パターンまで取り除かれた例を示す。取り除かれた出力文の人手評価値（評価者 A）は 5 である。この出力文が取り除かれた理由は、入力文の形態素“人”に付与された意味属性が“単位”または“個数回数等”であるためである。この例では、この形態素に例えば“人”や“人間”等の意味属性が付与されれば適切である。この例は入力文の形態素への意味属性の付与失敗の例とみなすことができる。入力文の形態素への意味属性の付与精度を向上させるには、意味解析技術の導入が必要になると考えられる。

表 6.14: パターン絞込みを行わなかった場合の自動評価結果

	BLEU	TER	METEOR	RIBES
単語レベル翻訳	0.183	0.694	0.469	0.713

表 6.15: パターン絞込みにより適切な文型パターンが取り除かれた割合

	適切	不適切
文数	47	53

6.8.3 パターン適合率の改善

6.7.4 節の結果より、パターン適合率は約 11% と低い。単語レベル文型パターン翻訳では文型パターン数増加の副作用は少ないため、パターン適合率を改善する確実な方法は、新たな文型パターンの作成である。またパターン絞込みによる、適切な文型パターンまで取り除いてパターン適合率を低下させる副作用が明らかになっており、パターン絞込み方法の修正によっても、ある程度の改善は可能である。

表 6.16: パターン絞込みにより不適切な日英文型パターンが取り除かれた例	
入力文	階段で足を踏み外して落ちた。
出力文	I fell missing legs in the stairs.
参照文	I missed a step and fell down the stairs.
人手評価値	評価者 A:2
日本語文型パターン	/</N1(人間)は> #1{N2(公共機関, 博物館)で, /N3(創作物形状, 像・書画)を }/V4(知覚動作, 思考動作)(て で)#5[/TIME6(常時)]/V7(状態, 属性変化, 利用).kako.
英語文型パターン	<I N1> V7 ^{past} #5[N6] V4 ^{ing} N3 in N2.
変数情報	N2=階段, N3=足, V4=踏み外し(踏み外す), V7=落ち(落ちる)
日本語文型パターン原文	博物館で美術品を眺めて終日過ごした。
英語文型パターン原文	I spent the entire day looking at the works of art in the museum.

6.8.4 言語モデルの学習文増加の効果

言語モデルの学習文は単言語のみでよく、収集が容易である。そこで言語モデルの学習文を追加した場合の翻訳精度を調査した。表 6.18 に言語モデルの学習文 18 万文を追加した場合の自動評価結果 (379 文) を示す。追加の学習文には、辞書の例文から抽出した日本語単文の対訳英文を用いた。表 6.4 と表 6.18 の比較より、単語レベル文型パターン翻訳方式と比較システム共に言語モデルの学習文追加の効果はみられなかった。単語翻訳確率の信頼性が十分ではない可能性があり、その場合は単語翻訳確率のためのデータ数を増加させなければ、言語モデルの学習文追加の効果が現れないと考えられる。

6.8.5 単語翻訳確率を使用しなかった場合の翻訳精度

訳語選択に単語翻訳確率を使用しなかった場合の、379 文の自動評価結果を表 6.19 に示す。表 6.4 の結果に比べ、全ての評価法で結果が低下していることから、若干ながら単語翻訳確率使用の効果があると考えられる。

表 6.17: パターン絞込みにより適切な日英文型パターンまで取り除かれた例	
入力文	彼はイギリス人ではなくてアメリカ人です。
取り除かれた出力文	He is not english , but the american .
参照文	He is not an Englishman but an American.
人手評価値	評価者 A:5
日本語文型パターン	/\$1^{\wedge}\${/N1(他称単数女, 女) は }/N2(師匠, 教師, その他多数) ではなくて \$1/N3(看護婦, 女).#da。
英語文型パターン	N1 @be not N2 , but N3 .
変数情報	N1=彼, N2=イギリス人{人}, N3=アメリカ人{人}
日本語文型パターン原文	彼女は先生ではなくて看護婦です。
英語文型パターン原文	She is not a teacher, but a nurse.

表 6.18: 自動評価結果 (言語モデル学習文追加)

	BLEU	TER	METEOR	RIBES
単語レベル	0.290	0.575	0.564	0.769
比較システム	0.218	0.648	0.487	0.757

6.9 今後の調査課題

実験結果より単語レベル文型パターン翻訳方式の翻訳精度の高さが確認された。次に、文型パターンによる文構造の持つ意味の保存と抽出が、どのように翻訳精度に影響するか検証を行う必要がある。この検証によってさらに、日本語話者の対象認識のあり方がどのような英語表現に対応するか分析することが可能である。

その他、パターン絞込みにおける意味属性の使用方法の検証が必要である。意

表 6.19: 自動評価結果 (単語翻訳確率なし)

	BLEU	TER	METEOR	RIBES
単語レベル	0.277	0.588	0.552	0.759

意味属性による制約の強さと翻訳精度の関係を検証することにより，変数化の抽象度と文構造の持つ意味の関係が検証できる。

6.10 単語レベル文型パターン翻訳方式のまとめ

単語レベル文型パターンを用いた日英機械翻訳方式の実装を行った。翻訳実験を行ったところ，文型パターンの適合率は約 11% と低いことが明らかになった。しかしながら，評価を行った 100 文において評価者平均 51.0 文で翻訳精度が高く，文型パターンが適合した場合は，単語レベル文型パターン翻訳方式が tree-to-string の SMT の翻訳精度を上回っていることが示された。

文型パターンの適合率の大幅な改善は困難であるが，意味属性コードによる適合範囲をより精密に決定することで，ある程度の適合率の改善が可能と考えられる。

問題分析を行った 61 文において，翻訳精度の高い文は 31 文であり，精度の低い文は 30 文であった。翻訳精度の低い 30 文のうち，文型パターンに関連する問題が 17 文，変数に対応する部分の翻訳に関連する問題が 13 文であった。前者の問題の 17 文から，形態素解析ミス等による文を除くと 7 文であった。これら 7 文は，複合的な表現が分割されて変数に対応したこと等が原因であり，パターン絞込みにおける意味属性の使用方法の変更により，翻訳精度が改善される可能性がある。後者の問題による 13 文は，主に未知語と不適切な訳語の選択による。単語辞書の見出し語の増加や，訳語選択におけるスムージング手法の最適化等により，ある程度の改善が可能である。

第7章 句レベル文型パターン翻訳方式

7.1 句レベル文型パターン翻訳方式概要

句レベル文型パターンは句変数を持つため，句レベル文型パターン翻訳方式の実装には，句変数に対応する要素の局所的翻訳が必要となる．非線形言語モデルでは線形要素の翻訳方法は任意である．句変数に対応する要素は，単語の場合に比べて大きく，単語変数の場合のように句辞書を作成し使用することは現実的ではない．

3.7節で示したように，線形要素も入れ子状に非線形要素と線形要素の内部構造を持ち得る．そこで本論文では，文型パターンを元に，文型パターンと同様の記述形式の句パターンを作成し，句変数に対応する要素の局所的翻訳に利用する．プログラム上では，単語レベル文型パターン翻訳方式を呼び出して句局所翻訳を行う．

システム全体のモデル図を図7.1に，手順とその概要を以下に示す．

- 1 入力文の形態素解析を行い，各形態素への品詞と意味属性を付与する．
- 2 日本語パターン検索システムにより，入力文に対しパターン照合を行い，入力文に適合する句レベル日英文型パターンとその変数情報を得る．
- 3 意味属性を用いて，日本語文型パターンの絞込みを行う．
- 4 英文生成システムにより，日英文型パターン，単語辞書，句パターンを用いて英語文を生成する．
 - 4.1 単語変数に対応する形態素は，単語辞書を用いて翻訳する．
 - 4.2 句変数に対応する形態素列は，句パターンを用いて翻訳する．その際，単語レベル文型パターン翻訳方式を利用する．
 - 4.3 単語翻訳確率と単語連鎖確率を用いて，訳語候補から最尤の組み合わせを選択する．

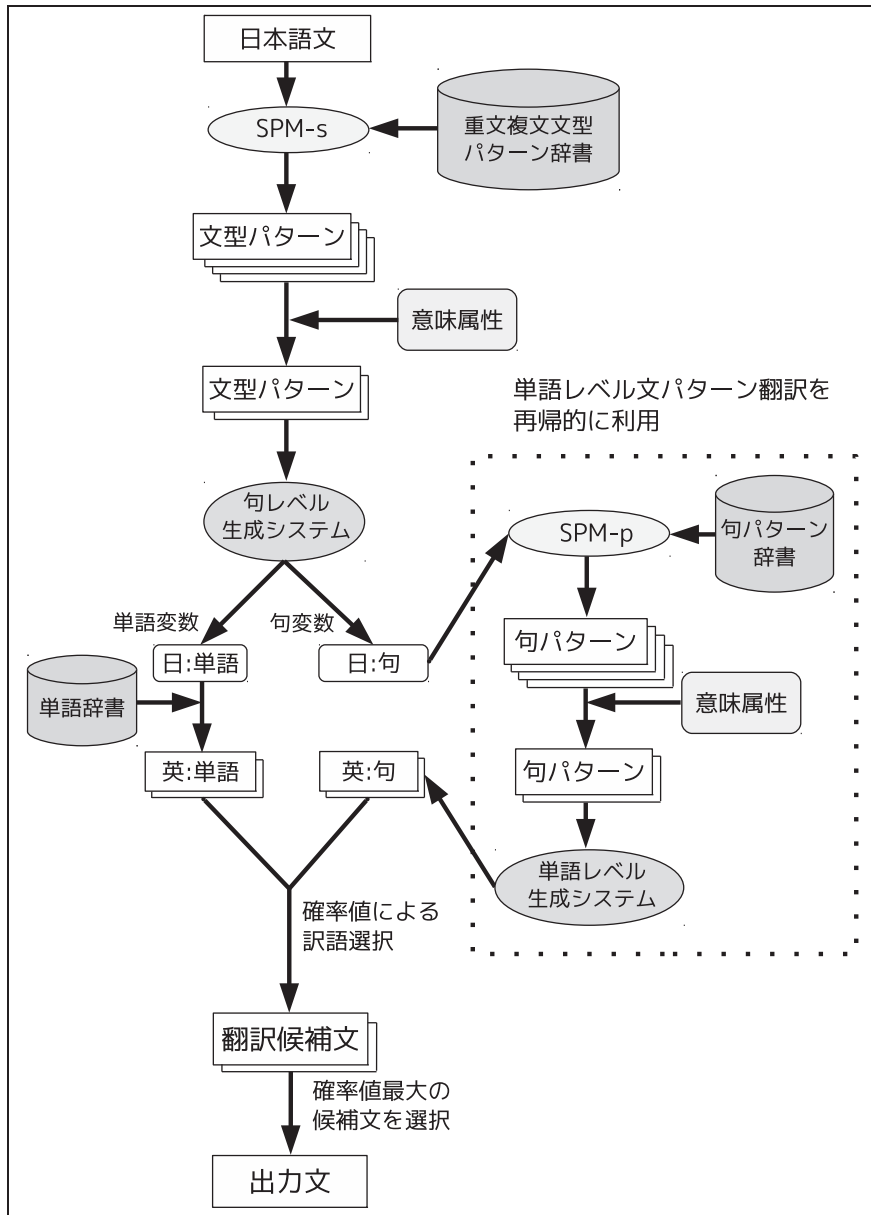


図 7.1: 句レベル文型パターン翻訳方式

7.2 句パターン辞書

句レベル文型パターン翻訳方式では、句変数に対応する要素を、句パターンを用いて翻訳する。句パターンは、句レベル文型パターンの句変数化された要素と、単語レベル文型パターンにおいてその句変数と対応する部位の記述を用いて自動

抽出する。抽出の際， unnecessary 記号や関数は削除する。以上より句パターンは一部の記号や関数を除き，単語レベル文型パターンと同一の記述形式を持つ。

7.2.1 句パターンの抽出

句レベル文型パターンの作成は，単語レベルを元にして行っており，句変数化された要素と，その元となった単語レベル文型パターンの記述との対応をとることができる。図 7.2 に，表 4.3 の文型パターンにおける対応と，作成される句パターンを示す。図 7.2 において，まず単語レベルの日本語文型パターンと句レベルの日本語文型パターンとの間の対応から，句変数化された日本語要素に対応する記述 “N2 になっ” と “N3 を (V4)” が抽出できる。同様に英語文型パターンから “turn N2” と “V4^base N3” が抽出できる。これらの記述から unnecessary 部分を削除し記述を修正することで，この例では図 7.2 下部に示した 2 つの句パターンが作成できる。

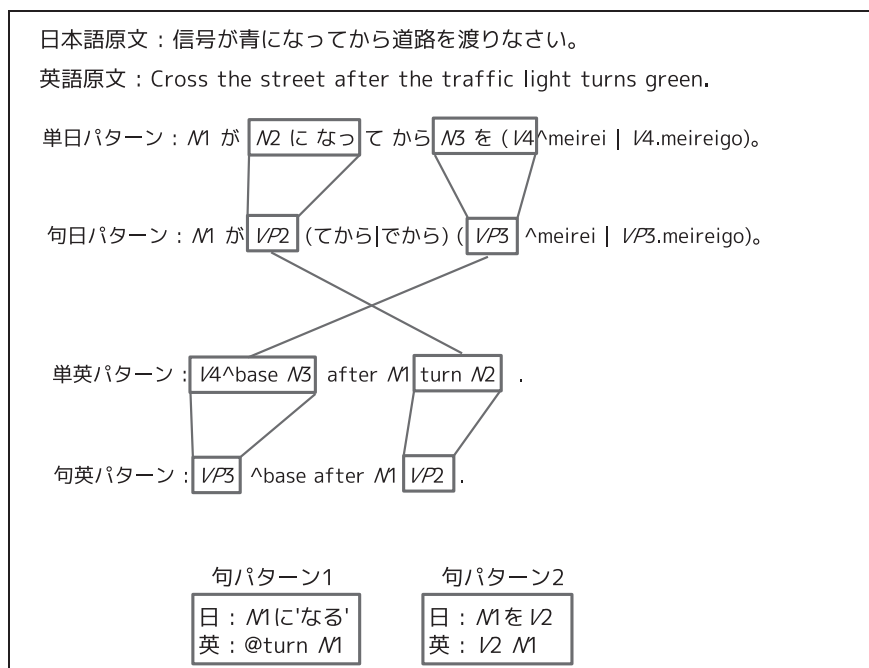


図 7.2: 句パターンの抽出

7.2.2 句パターンの種類と変換処理

句パターンは日本語句変数 5 種類と英語句変数 4 種類の組み合わせにより，20 種類に分類する。日本語句変数は名詞句 (NP)，動詞句 (VP)，形容詞句 (AJP)，

形容動詞句 (*AJVP*)，副詞句 (*ADVP*) の 5 種類，英語句変数はこれらから形容動詞句を除いた 4 種類である．また *VP(日)–VP(英)* パターンでは動詞を変数化する場合と原形字面化する場合の二種類のパターンを作成する．そのため合計 21 種類の句パターン辞書が作成される (具体例は 7.2.3 節)．*VP(日)–VP(英)* パターンにおいて，動詞を原形字面化することでその動詞のみが適合可能となり，対応する英語動詞が一つに決定されることで不適切な訳語選択が減少すると考えられる．それとは逆に動詞句への適合可能性が減少するため，動詞を変数化した句パターンも作成し使用する．

句パターンの作成の際，文型パターンの記述そのままでは句パターンとして不適切な場合が存在する．句パターン作成時の元の文型パターンからの記述変換処理を表 7.1 にまとめる．

表 7.1: 句パターン作成のための記述の変換

対象	変換の例	説明
番号の付け替え	#2[N3] の N4 → #1[N1] の N2	変数または記号の番号を 1 から付け替え．
末部の動詞字面	にっこり笑つ → にっこり'笑う'	日本語動詞句パターン (<i>VP-NP</i> 等) で動詞が字面の場合は原形字面に変換．
整合の取れていない要素	M1 を (V2 → M1 を V2	V2 の直前の “(” は V2 の後ろの処理とセットだが，その部位は抽出していないため記号の整合性がとれなくなっており，“(” は削除．
句パターンに必要な関数	AJ1 N2^pron → AJ1 N2	N2 に付与されている “^pron” は直前の変数を代名詞化する関数であり，句パターンでは不要であるため削除．

7.2.3 句パターンの例

VP(日)–VP(英) パターンの例を表 7.2 に示す．なお句パターンの変数も意味属性を持つが，記述の簡単化のために除いて記述する．日本語 *VP* の残りのパターンの例を表 7.3 に示す．日本語 *NP* の句パターンの例を表 7.4 に，*AJP* の例を表 7.5 に，*AJVP* の例を表 7.6 に，*ADVP* の例を表 7.7 に示す．“日原表記” と “英原表記” はそれぞれ日本語，英語原文での句変数化された要素を示している．

表 7.2: VP-VP パターンの例

日原表記	パソコンを使い
英原表記	using a personal computer
< VP - VP(動詞原形字面化)>	
日パターン	N1 を'使う'
英パターン	@use N1
< VP - VP(動詞変数化)>	
日パターン	N1 を V2
英パターン	V2 N1

表 7.3: 日本語 VP パターンの例
< VP - NP>

日原表記	一般に知ら
英原表記	common knowledge
日パターン	ADV1 V2
英パターン	AJ1 N2
< VP - AJP>	
日原表記	学校に遅れる
英原表記	late for school
日パターン	N1 に V2
英パターン	AJ2 for N1
< VP - ADVP>	
日原表記	語調を強め
英原表記	emphatically
日パターン	語調を V1
英パターン	ADV1

7.3 形態素解析

句レベル文型パターン翻訳で使用する形態素解析の記述形式は、単語レベル文型パターン翻訳と同様である。図 7.3 に入力文の形態素解析結果を示す。

7.4 句レベル日本語文型パターンの照合

句レベル文型パターンの照合にも SPM を使用する。句レベル文型パターンの照合では、入力文の要素が句変数に対応可能かどうか決定する必要がある。SPM では ATN 法による構文解析機能によりこのタスクを達成している (5.3.1 節)。複数

表 7.4: 日本語 NP パターンの例

< NP-NP >		< NP-VP >	
日原表記	あの建物	日原表記	ホテルの経営
英原表記	that buiding	英原表記	running a hotel
日パターン	GEN1 N2	日パターン	N1 の N2
英パターン	AJ1 N2	英パターン	V2 N1
< NP-AJP >		< NP-ADVP >	
日原表記	ざん新なデザイン	日原表記	長い間
英原表記	novel in design	英原表記	for a long time
日パターン	AJV1 ^{rentai} N2	日パターン	AJ1 ^{rentai} TIME2
英パターン	ADV1 in AJ2	英パターン	for AJ1 N2

表 7.5: 日本語 AJP パターンの例

< AJP-NP >		< AJP-VP >	
日原表記	水より比重の大きい	日原表記	協調性がない
英原表記	a specific gravity greater than water	英原表記	lack the will to cooperate
日パターン	N1 より N2 の AJ3 ^{rentai}	日パターン	N1 が AJ2
英パターン	N2 AJ3 ^{er} than N1	英パターン	V2 the will to V1 ^{base}
< AJP-AJP >		< AJP-ADVP >	
日原表記	いちばん良い	日原表記	時には楽しい
英原表記	the most suitable	英原表記	fun sometimes
日パターン	#[N1]AJ2 ^{rentai}	日パターン	#[TIME1 には] AJ2 ^{rentai}
英パターン	#[AJ1 st] AJ2	英パターン	N2 #1[ADV1]

の解釈が可能な場合は、全ての解釈を採用し、可能な限り多くの文型パターンが適合するように実装されている。

表 7.8 に複数の適合の仕方が可能な例を示す。句変数には単語（形態素一つ）も適合し得るので、表 7.8 の例では“瓦”や“ライオン”が名詞句変数 NP に適合している。また動詞句変数に適合可能な要素を認識するための規則には、格要素と格助詞も含まれているため、動詞句変数には主部を含む節が適合し得る。そのため表 7.8 の例では、「が格」を含む“腹が減っ”や「は格」を含む“ライオンは水に映っ”の節表現も動詞句変数 VP に適合する。節が VP に適合可能なことの翻訳への影響は明らかになっていない。これについては 7.10.5 節で触れる。

表 7.6: 日本語 AJVP パターンの例

< AJVP-NP >		< AJVP-VP >	
日原表記	検索が容易だ	日原表記	外見は豪華だ
英原表記	easy access	英原表記	looks fancy
日パターン	N1 が AJV2	日パターン	N1 は AJV2
英パターン	AJ2 N1	英パターン	V1 AJ2
< AJVP-AJP >		< AJVP-ADVP >	
日原表記	とても静かな	日原表記	なんともユーモラスだ
英原表記	very quiet	英原表記	pretty humorously
日パターン	#1[ADV1]AJV2^rentai	日パターン	#[ADV1]AJV2
英パターン	ADV1 AJ2	英パターン	#ADV1 ADV2

表 7.7: 日本語 ADVP パターンの例

< ADVP-NP >		< ADVP-VP >	
日原表記	今年の夏	日原表記	大変ありがたく
英原表記	last summer	英原表記	most appreciated
日パターン	#1[TIME1 の] TIME2	日パターン	#1[ADV1]ADV2
英パターン	#1[V1] N2	英パターン	#1[ADV1] V2
< ADVP-AJP >		< ADVP-ADVP >	
日原表記	いつでもすぐ	日原表記	一行ずつ空けて
英原表記	immediate	英原表記	on every other line
日パターン	いつでも ADV1	日パターン	1 N1 ずつ (空け あけ) て
英パターン	AJ1	英パターン	on every other N1

7.4.1 句レベル文型パターン照合例

図 7.3 の形態素解析結果に対する，句レベル文型パターンの照合結果を表 7.9 に示す．表 7.9 の例では，動詞句変数 VP2 に“計算機を利用し (利用する)”が，VP3 に“難解な数式を解い (解く)”が対応しており，これらの要素を句パターンを用いて局所的に翻訳する．また表 7.9 の変数情報中の“[利用し]”は，VP2 に対応する要素“計算機を利用し (利用する)”の内部での係り先が“利用し (利用する)”であることを示している．同様に VP3 内部の係り先は“解い (解く)”であり，文型パターンの絞込みではこれら係り先の形態素の意味属性を使用する (7.5 節)．

-
1. /彼 (人称代名詞, { 他称単数男, 男 })
 2. +は (副助詞)
 3. /計算機 (一般名詞, { コンピュータ })
 4. +を (格助詞)
 5. /利用し (他動詞, 利用する, 利用し, { 利用 })
 6. /難解な (形容動詞, 難解だ, 難解な, { 属性 })
 7. /数式 (一般名詞, { 式 })
 8. +を (格助詞)
 9. /解い (他動詞, 解く, 解い, { 思考動作, 身体動作, 属性変化 })
 10. +た (助動詞)
 11. +。 (句点)
 12. /nil
-

図 7.3: 形態素解析結果

7.5 日本語文型パターンの絞込み

句レベル文型パターン翻訳においても、意味属性による文型パターンの絞込みを行う。句レベル文型パターン翻訳では、単語変数と句変数に対して、パターン絞込みのための適合の制約をかける。意味属性は階層化されており、単語変数と句変数で適合範囲を決めるために最適な階層数が異なり得る。現時点で句変数に対する適合範囲をどのようにとるべきか明らかではないため、本研究では単語変数に対する適合範囲を句変数にも用いる。文型パターン絞込みの手順は以下の通りとする。

表 7.8: 複数の適合の仕方が可能な例

日本語表現	適合可能なパターン記述	変数情報
家の上の瓦	$NP1$	$NP1 = \text{家の上の瓦}$
	家の上の $NP1$	$NP1 = \text{瓦}$
腹が減っ	$N1$ が $V2$	$N1 = \text{腹}, V2 = \text{減っ}$
	$VP1$	$VP1 = \text{腹が減っ}$
ライオンは水に映っ	$NP1$ は $VP2$	$NP1 = \text{ライオン}, VP2 = \text{水に映っ}$
	$VP1$	$VP1 = \text{ライオンは水に映っ}$

表 7.9: パターン照合により得られた日英文型パターンと変数情報

日本語文型パターン	/N1は/VP2(所有的移動, 利用, 精神的移動)^renyou/VP3(属性変化, 身体動作, 思考動作).kako。
英語文型パターン	N1 VP3^past by VP2^ing.
変数情報	N1=彼 { 他称単数男, 男 }, VP2=計算機を利用し {[利用し], 利用 }, VP3=難解な数式を解い {[解い], 思考動作, 身体動作, 属性変化 }

1. 入力文および適合した日本語文型パターンから、変数部分の意味属性を抽出。
2. 入力文の形態素の意味属性を用いて、日本語文型パターンの変数に対する適合範囲を決定：適合範囲は単語レベル文型パターンと同様（名詞変数では4層以下の子孫，用言の変数の場合はその属性と子孫）。複数形態素が一つの変数に対応している場合は，その形態素列の中の係り先の形態素の意味属性を用いる。
3. 日本語文型パターンの全ての変数において，各変数の意味属性が各適合範囲に含まれる場合，この文型パターンを翻訳に使用。

句変数に対応する要素は複数の形態素であることがほとんどである。複数の形態素が一つの変数に対応する場合は，SPMの出力する変数内部の係り先の形態素を用いて上記2の変数に対する適合範囲を決定する。表7.9の変数VP2に対応する要素“計算機を利用し”を例に説明する。“計算機を利用し”の内部での係り先は“利用し(利用する)”であり，句レベル文型パターンの変数VP2に対する適合範囲は，“利用し(利用する)”の用言意味属性“利用”を用いて決定する。用言意味属性“利用”は第4段の属性であるため，適合範囲はこの意味属性“利用”のみとなる。表7.9の日本語文型パターンの変数“VP2”は，意味属性“利用”を持つことから，この適合範囲に含まれる。

7.6 句レベルの英文生成

句レベル文型パターン翻訳でも単語レベルと同様に，変数に対応する要素を局所的に翻訳し，英語文型パターンの変数部分にそれらの結果を挿入することで翻訳文を得る。そして確率値を用いた訳語選択により英語文を生成する。句レベル文型パターン翻訳では，単語変数に加え，句変数に対応する要素の局所的な翻訳

が必要になる。非線形言語モデルでは線形要素の翻訳方法は任意であり、本論文では句パターンを用いて翻訳を行う。句パターンは単語レベル文型パターンの記述を元に作成されているため、単語レベル文型パターン翻訳方式を利用して翻訳が可能である。また句変数に対応する要素の翻訳における訳語選択は、局所的翻訳時には行わず、句レベルの英語文型パターンに非決定的に訳語を代入しておき、最後に単語変数に対応する要素の翻訳結果と同時に行う（7.7節の具体例参照）。

7.6.1 単語変数に対応する要素の処理

単語変数に対応する要素の翻訳は単語レベル文型パターン翻訳と同様にして行う。

7.6.2 句変数に対応する要素の処理

句変数に対応する形態素列は、句パターンを用いて翻訳する。その際、再帰的に単語レベル文型パターン翻訳方式を利用する。ただし句変数に適合した要素が1形態素の場合は単語変数として処理する。句変数に対応する要素の翻訳手順を次に示す。

1. 日本語パターン検索システム（SPM）により、各句変数に対応する要素に対しパターン照合を行い、適合する句パターンとその変数情報を得る。
2. 意味属性コードを用いて、句パターンの絞込みを行う。絞込み後に100以上の句パターンが残る場合は、非線形要素の多い順に上位100句パターンのみを翻訳に使用する。
3. 単語レベル英文生成システムにより、句パターンを用いて英語句を生成する。句パターン内部の変数に関数が付与されている場合は関数処理を行う。
4. 句パターンを用いて得られた英語句を句レベル英文生成システムに戻す。

適合した句レベル文型パターンにおいて、句変数に関数が付与されている場合は、句の翻訳結果に対し語形の変換を行う。単語変数と句変数に付与される関数は同じであるが、単語変数と句変数では対応部分の局所的翻訳結果の構造が大きく異なるため、同じ関数が付与されていたとしても、単語変数と句変数で異なる語形変換処理が必要となる場合が存在する。VPに関数が付与されている場合は、句変数に対応する要素の翻訳結果のうち、最初に現れる動詞のみ語形を変換する。その他の句変数に関数が付与されている場合は、単語変数と同様の処理を行う。

訳語選択は句レベル英語文型パターンに非決定的に代入した後に、単語変数に対応する要素の翻訳結果と共に一括して行う。

7.7 英文生成の例

入力文「彼は計算機を利用し難解な数式を解いた。」の翻訳例を図 7.4 に示す。翻訳には、表 7.9 の句レベル文型パターンとその変数情報を用いる。

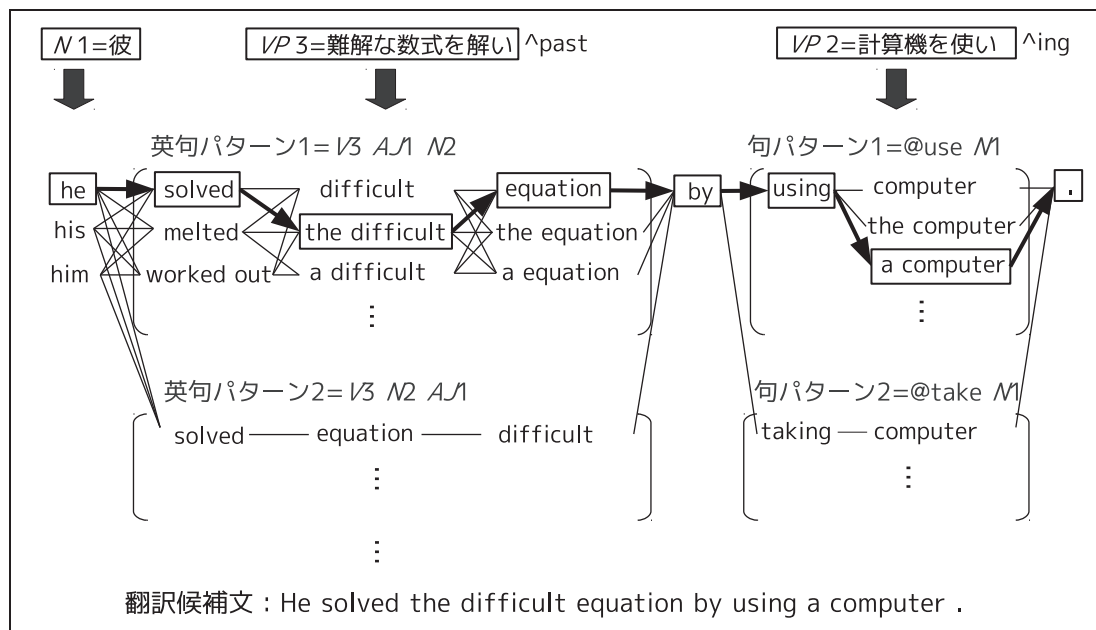


図 7.4: 句レベルの英文生成の例

変数に対応する形態素列に対し、単語辞書または句パターンを用いて翻訳を行い、複数の訳語候補を得る。例えば、 $VP2 \wedge ing$ に対応する要素は「計算機を使い」であり、一つ目の句パターン " $@use N1$ " を用いた翻訳結果は $[[\text{"using"}], [\text{"computer"}, \text{"the computer"}, \text{"a computer"} \dots]]$ であり、二つ目の句パターン " $@take N1$ " を用いた翻訳結果は $[[\text{"taking"}], [\text{"computer"} \dots]]$ である。原形字面 " $@use$ " は翻訳の際、語形が展開されて、 $[\text{"use"}, \text{"uses"}, \text{"using"}, \text{"used"} \dots]$ が訳語候補となる。この文型パターンでは句変数 $VP2$ に動詞の ing 形を指定する $\wedge ing$ が付与されているので、 $VP2$ の翻訳候補に関数の処理を適用し、 ing 形の "using" が訳語候補として残っている。そして確率計算（一文全体の単語翻訳確率と単語連鎖確率の乗算値）により最尤となる単語組み合わせを選択し、翻訳候補文「He solved the difficult equation by using a computer.」を得る。

7.8 翻訳候補文からの出力文の選択

パターン照合とパターン絞込みにより複数の日本語文型パターンが得られている場合は、SPMにおける上位3つの日英文型パターンに対し、7.6節の翻訳を行う。一つの英語文型パターンから一つの翻訳候補文が得られるため、最大3つの翻訳候補文が得られる。これらの翻訳候補文から、確率計算の値（6.4節の値と同じ）が最大のものを一つ、出力文として選択する。

7.9 翻訳実験

7.9.1 実験条件

句レベル文型パターン翻訳方式の有効性を検証するため、入力文に重文複文文型パターン辞書の日本語文型パターン原文を用いて、翻訳実験を行う。文型パターン辞書の日英文型パターン原文10万文対（単語レベル文型パターンの実験で使用した原文対）から作成された、65,171の句レベル文型パターンを翻訳実験に使用する。単語辞書の作成と単語翻訳確率の学習には、この10万日英原文対をもとにした単語レベル文型パターンから抽出した、変数化された要素の対を用いる。句パターンの抽出は、この10万日英原文対から作成された単語レベルおよび句レベル文型パターンから行う。単語連鎖確率（単語 tri-gram 値）の学習は、この10万文型パターンの英語原文10万文を用いて行う。学習に使用しない残りの日本語文型パターン原文約2万文から、1,000文を抽出しテスト文に用いる。

7.9.2 比較システム

比較システムは単語レベル文型パターン翻訳の場合と同様に、Travatarを使用する。その他学習文等も単語レベル文型パターンの場合と同様である。

7.9.3 作成された句パターン

表7.10に作成された句パターンの数を示す。VP(日)-VP(英)が最も多く、V原型字面化した場合とV変数化した場合共に約4万句パターン対であった。次に日本語NP句パターンが最も多く約2万5千句パターン対であった。V原型字面化したVP(日)-VP(英)句パターンを除いて考えると、句パターン全体に占める割合は、VP(日)-VP(英)句パターンが約57%を占め、日本語名詞句パターンは32.5%を占める。約90%の句パターンがVPとNPで占められていることになる。句パター

ンの数と句レベル文型パターン中の句変数の数はほぼ比例関係にあるため、句パターンの適合率や句の局所的翻訳の精度は、この二つの句パターンの影響が圧倒的に大きいことを示している。

表 7.10: 句パターン数

句パターンの種類	句パターン数
<i>VP</i> (日)- <i>VP</i> (英): <i>V</i> 原型字面化	41,505
<i>VP</i> (日)- <i>VP</i> (英): <i>V</i> 変数化	43,359
<i>VP</i> (日)- <i>NP</i> (英)	3,153
<i>VP</i> (日)- <i>AJP</i> (英)	1,750
<i>VP</i> (日)- <i>ADV</i> <i>P</i> (英)	107
<i>VP</i> (日)-その他 3 種類 (英) 合計	4,990
<i>NP</i> (日)- <i>NP</i> (英)	24,160
<i>NP</i> (日)- <i>VP</i> (英)	458
<i>NP</i> (日)- <i>AJP</i> (英)	99
<i>NP</i> (日)- <i>NP</i> (英)	50
<i>NP</i> (日) 合計	24,767
<i>AJP</i> (日)- <i>NP</i> (英)	188
<i>AJP</i> (日)- <i>VP</i> (英)	379
<i>AJP</i> (日)- <i>AJP</i> (英)	1359
<i>AJP</i> (日)- <i>ADV</i> <i>P</i> (英)	48
<i>AJP</i> (日) 合計	1974
<i>AJVP</i> (日)- <i>NP</i> (英)	131
<i>AJVP</i> (日)- <i>VP</i> (英)	256
<i>AJVP</i> (日)- <i>AJP</i> (英)	523
<i>AJVP</i> (日)- <i>ADV</i> <i>P</i> (英)	15
<i>AJVP</i> (日) 合計	925
<i>ADV</i> <i>P</i> (日)- <i>NP</i> (英)	15
<i>ADV</i> <i>P</i> (日)- <i>VP</i> (英)	10
<i>ADV</i> <i>P</i> (日)- <i>AJP</i> (英)	26
<i>ADV</i> <i>P</i> (日)- <i>ADV</i> <i>P</i> (英)	171
<i>ADV</i> <i>P</i> (日) 合計	222
総計	117,742

7.9.4 パターン適合率

入力 1,000 文のうち、文型パターンとの照合と意味属性コードを用いた文型パターンの絞込みに成功した文は、521 文であった。この 521 文のうち、全ての句変数適合部分に対し句パターンが適合した文は 277 文であった。本論文では以後パターン適合率を、入力文に対する、パターン照合とパターン絞込みの後、少なくとも 1 つの文型パターンが適合し、かつ全ての句変数適合部分に句パターンが適合している文の割合と定める。よって本実験では、パターン適合率は約 28% (277/1,000)

となる．単語レベル文型パターンを用いた場合（約 11%）に比べ，大幅にパターン適合率が改善していることがわかる．しかしそれでなお，パターン適合率は十分な割合に達していない．

7.9.5 評価結果

本実験でも文型パターン翻訳方式と比較システム共に，文型パターンが適合した文のみを対象として評価を行う．

文型パターンが適合した 277 文中，学習文に同じ日本語文を持つのは 25 文であり，持たない文は 252 文であった．上記 252 文の翻訳結果に対し，単語レベル文型パターン翻訳の時と同じ 4 つの評価法（6.7.5 節）を用いて自動評価を行った．自動評価の結果を表 7.11 に示す．表 7.11 より，全ての評価法において句レベル文型パターン翻訳方式は比較システムより評価値が若干低い．

表 7.11: 自動評価結果

	BLEU	TER	METEOR	RIBES
句レベル翻訳	0.131	0.737	0.419	0.680
比較システム	0.138	0.700	0.437	0.707

次に上記 252 文からランダムに 100 文抽出し人手評価を行った．人手評価は adequacy による 5 段階評価を行う．人手評価の基準は表 6.5 に示した基準を用いる．評価者は 6.7.5 節の評価者 A の 1 名で行う．人手評価の結果を表 7.12 に示す．表 7.12 において，各評価値における数値は，その評価値の文数を示している．本翻訳方式では評価 4 と 5 を合わせても 27 文であり，翻訳精度の高い文はあまり多くない．ただし平均値では比較システムを上回っている．

以上より自動評価と人手評価の間で評価が逆転している．自動評価法は，方式の大きく異なるシステム間では値が大きく異なり得ることにより，システム間の直接の比較が難しいことが指摘されている．そのため人手評価の方が翻訳精度を反映している度合いが大きいと考えられる．

表 7.12: 人手評価結果

	評価 5	評価 4	評価 3	評価 2	評価 1	平均値
句レベル翻訳	18	9	21	36	16	2.77
比較システム	10	3	15	47	25	2.26

表 7.13 と表 7.14 に本翻訳方式の人手評価の値が高い例を示す。比較のため表 7.13 と表 7.14 の下 2 行に、同入力文における比較システムの出力文とその人手評価値を示している。評価の低い例は、7.10 節において、表 7.17 と表 7.18 に示す。

表 7.13: 本翻訳方式において人手評価値の高い例 1

入力文	先生は学生にもっと勉強するようにと言いました。
参照文	The teacher told the students to study more.
出力文	The teacher told the students to study harder .
人手評価値	5
日本語文型 パターン	/\$1^{}/N1(他称単数男, 男) は }/N2(自称単数女, 他称単数, その他多数) に \$1!VP3(属性, 自然現象, その他多数).suitei と \$1!VP4(因果関係, 身体動作, その他多数).kako。
英語文型パ ターン	$N1 VP4^{past} N2^{obj} to VP3^{base}$.
変数情報	$N1=$ 先生, $N2=$ 学生, $VP3=$ もっと勉強する, $VP4=$ 言い(言う)
日本語文型パ ターン原文	彼は私に中に入るようにと手で招いた。
英語文型パ ターン原文	He beckoned me to come in.
比較システム の出力文	i told the teacher student to study harder .
評価値 (比較 システム)	2

7.10 考察

本節では、句レベル文型パターンを使用することの効果と問題点を明確にするための調査を行う。句レベル文型パターンに由来する問題と、本研究で用いた具体的方法の問題の区別が必要となる。

7.10.1 使用した文型パターンの適切さ

まず、出力文に使用された文型パターンがその入力文の翻訳に対し適切であったか調査した。英語文型パターンを用いて、適切な翻訳文を出力可能だと判断で

表 7.14: 本翻訳方式において人手評価値の高い例 2

入力文	仕事を済ませてからシャワーを浴びた。
参照文	After I finished my work, I took a shower.
出力文	On finishing the job , he took the shower .
人手評価値	5
日本語文型パターン	/VP1(属性, 因果関係, その他)(てから でから)</N2(人間)は >!VP3(相対関係, 行動, その他).kako。
英語文型パターン	On VP1 ^{ing} , <he N2> VP3 ^{past} .
変数情報	VP1=仕事を済ませ(済ませる), VP3=シャワーを浴び(浴びる)
日本語文型パターン原文	学校を出てから実業について。
英語文型パターン原文	On leaving school, he went into business.
比較システム の出力文	i was in the shower after we finish our work .
評価値 (比較システム)	3

きる場合を“適切”，そうでないときは“不適切”とした。人手評価を行った 100 文における，英語文型パターンの適切さと人手評価値の関係を表 7.15 に示す。

表 7.15 より，半数近い 44 文で不適切な文型パターンを使用しており，この 44 文全てで翻訳精度が低い（評価 1 から 3）。また適切な文型パターンを使用した場合でも，約半数の文（56 文中 29 文）で翻訳精度が低く，これらは変数に適合した要素の翻訳または訳語選択の問題による。以上より，テスト文全体の翻訳精度が高くない原因は，不適切な文型パターンを使用する割合が高いことと，変数に適合した要素の翻訳または訳語選択で不適切なことの二点である。重文複文文型パターン辞書の思想的背景を考慮すると，前者の問題がより重大である。

7.10.2 不適切な文型パターンを使用した要因

不適切な文型パターンを使用した原因には，文型パターン作成に使用した日英原文対が日英で対応していない等の，句レベル文型パターン翻訳と直接関係しない原因も含まれている。このような原因を持つ文は 19 文あり，本節では省略する。

表 7.15: 英語句型パターンの適切さと人手評価値の関係

	評価5	評価4	評価3	評価2	評価1	合計
適切	18	9	12	14	3	56
不適切	0	0	9	22	13	44

分析の対象は不適切な句型パターンを使用した44文から上記19文を除いた25文とする。この25文においては、入力文と使用した日本語句型パターン原文の表現する事象が大きく異なっていた。このことは意味属性を用いたパターン絞込みでは、入力文に類似する文から作成された句型パターンを選択できていないことを示している。パターン絞込みが適切な句型パターンの選択方法として機能しないと、多様な原因により各入力文の翻訳において各句型パターンが不適切となる。表7.16に句型パターンが不適切な理由とその内訳を示す。

表 7.16: 不適切な句型パターンを使用した原因とその内訳

原因(大分類)	原因	数
句変数化範囲	(1) 句型パターン作成時の句変数化の範囲の問題	2
	(2) パターン照合時の句変数化認識部分の範囲	11
英語句型パターンの構造	(3) 英語句型パターンで指定される語形が入力文の翻訳では不適切	7
	(4) 英語句型パターンで字面で記述されている助詞や接続詞がその入力文の翻訳には不必要	4
	(5) その他英語句型パターンの構造が不適切	5
その他	(6) 複文の入力文に重文から作られた日本語句型パターンが適合	2
	(7) 省略要素と補完記号に関連する問題	7

(4)と(5)の両方の原因を持つ具体例を表7.17に示す。入力文と出力文を見比べると、出力文からは入力文の意味がほとんど読み取れない。最大の原因は英語句型パターンの“@be^past VP2(群衆に向け)^ed”の部分で、受動態への変換を指定する記述になっているため、この部分の翻訳結果が“were given to the crowd”となっている。また“into N3(放水する)”の部分も、この入力文では不適切である。日本語パターン原文の“兵糧攻めにあつ(会う)”という表現は受身の意味も含んでおり、英語パターン原文ではそのニュアンスが含まれるように“were starved into”の受身表現で訳されている。したがって日英句型パターンでこの表現に対応する部分

は、日本語文型パターンで“VP2”，英語文型パターンで“@be^past VP2^ed”と記述されている。入力文で VP2 に適合した“群衆に向け(向ける)”の表現は受身の意味を含まないため、上記の英語文型パターン中の該当部分の記述により受身形に変換されると、入力文とは異なる意味の表現となる。また“starve [目的語] into [名詞]”で“[目的語]に飢えさせて [名詞]させる”の意味を持つ熟語であり、“into [名詞]”は“starve”と結びついている表現形式であるため、“starve”またはこの表現形式と同じ形式を持つ英語動詞が訳語候補として存在しなければ適切な翻訳は困難である。このことからこの表現は句変数化すべきでないが、日英原文で直接対応の取れる要素を変数化するという基準によっては、句変数化可能な要素と可能でない要素の区別をつけることが非常に難しい。

表 7.17: (4) と (5) の両方の原因を持つ例

入力文	機動隊は群衆に向けて放水した。
出力文	The riot police were given to the crowd into drainage .
参照文	The riot police turned the water cannon on the crowd.
日本語文型パターン	/\$1^{M1(軍)は}\$!VP2(存在,属性,その他多数)(て で)\$1/(V3.kako ND3をした)。
英語文型パターン	M1 @be^past VP2^ed into N3 .
変数情報	M1=機動隊, VP2=群衆に向け(向ける), V3=放水し(放水する)
日本語文型パターン原文	反乱軍は兵糧攻めにあって降服した。
英語文型パターン原文	The rebels were starved into submission.
人手評価値	1

7.10.3 句変数に対する意味属性による絞込みの問題

文型パターンの句変数に付与された意味属性は、句の中での係り先の形態素の意味属性である。またパターン絞込みにおいて、句変数に対応する要素から抽出して句変数への制約に用いるのも、係り先の形態素の意味属性である。これらのことは、句の意味属性を句の中の一つの形態素の意味属性で決定していることを意味している。しかし非線形言語モデルの考え方に従うと、文を句へと分割した場合は分割された一つの句全体で一つの意味を持つのであり、その中の一要素だ

けで句の意味は決定することができない。以上のことより句変数への意味属性の付与の仕方と入力文からの句に対する制約の抽出の仕方が、言語の非線形性を考慮に入れていないことになる。非線形言語モデルに基づく翻訳方式であるかぎり、言語の非線形性に基いて句変数の意味属性を決定する必要がある。句変数に付与すべき意味属性の体系がどのようなものであるか、今の段階でははっきりわからないことと、単純に量による人的なコストの問題から人手で句変数に意味属性を付与し直すのは困難である。word-embedding 等によって、ある程度の精度で付与できる可能性があるため、今後は何かしらの機械学習による付与の方法を調査する必要があると考えられる。

7.10.4 句変数に対応する要素の翻訳

文型パターンが適切な文は 56 文あり、そのうち約半数の 29 文で翻訳精度が低い。これは変数に対応する要素の翻訳結果の問題である。さらに句パターンの問題と訳語選択の問題があるが、この二つの問題は密接に関係しており、明確に区分することは難しい。また句変数に対応する要素が慣用表現的である場合も多々存在し、その場合は句パターンの前に慣用句辞書を適用しなければ適切な翻訳は難しい。まず句変数に対応する要素の翻訳を適切に行うために何が必要であるかを、明らかにする必要がある。

表 7.18 に、句変数に対応する要素が慣用的表現であるために、適切な翻訳が難しい例を示す。この例では“目のまわる”が日本語動詞句変数 VP1 に対応しており、その翻訳結果が“turn the eye”となっている。この“目のまわる”は本当に目が回るのではなく、忙しい程度が甚だしいことを表現するための慣用的な表現であり、直訳では適切な英語表現を得ることが難しい。

7.10.5 句パターンが未適合の場合の文型パターンの適切さ

非線形言語モデルでは線形要素に対応する要素の翻訳方式は任意のため、適合する句パターンが得られなかった場合に、異なる方式で翻訳することも可能である。句パターンが適合しなかった 20 文に対し、その文型パターンで精度が高い翻訳文がえられるか調査したところ、13 文で文型パターンが不適切なため句局所翻訳の結果によらず精度の高い翻訳文を得る可能性が存在しなかった。そのため句パターンが適合しなかった場合に異なる翻訳方式を用いることは現時点では有用ではない。

上記 13 文のうちの 1 例を表 7.19 に示す。これは表 7.8 の一例と同じ、句変数に節が対応していることが原因である。句レベル文型パターンの動詞句変数は、主

表 7.18: 句変数に対応する要素が慣用的表現である例

入力文	目のまわるほど忙しい。
出力文	It is busy enough to turn the eye .
参照文	I am in a whirl of business.
日本語文型 パターン	/VP1(身体動作) ^{rentai} (ほど 程)/AJ2(属性, 知覚状態) #1(.genzai .kako)。
英語文型パ ターン	It @be#1(^{present} ^{past})AJ2 enough to VP1 ^{base} .
変数情報	VP1=目のまわる,AJ2=忙しい
使用句パター ン(VP1)	日: N1 の V2, 英: V2 N1
句パターンの 変数情報	N=目,V2=まわる
日本語文型パ ターン原文	耳がちぎれるほど寒い。
英語文型パ ターン原文	It is cold enough to freeze your ears off.
人手評価値	1

格表現を含まないように厳格に変数化されており、文型パターン作成時の仕様と大きく異なる適合の仕方をしては、文構造の持つ抽象的意味を保存する構造変換は達成できないことを示している。

ただし不適切な句パターンが適合する場合は、上記の例のように構造に一定の傾向があるために、その場合を機械的に判定しやすい。句変数への要素の対応の仕方に問題が存在しないと判断できる場合は、異なる局所的翻訳方法を採用することで、文型パターンの適合率をある程度改善することが可能である。

7.10.6 単語レベル文型パターン翻訳方式との翻訳精度の比較

単語レベル文型パターン翻訳方式と句レベル文型パターン翻訳方式の翻訳精度の違いを比較する。句レベル文型パターン翻訳方式の翻訳実験に使用した1,000文中で、単語レベル文型パターン翻訳方式で出力文が得られたのは99文(約10%)であった。このうち、入力文が学習文に同一の日本語文を持たない文は、66文で

表 7.19: 句パターンが不適合の例

入力文	レストランで男が食い逃げするのを見た。
出力文	I saw to レストランで男が食い逃げする .
参照文	At the restaurant I saw a man leave without paying his bill.
日本語文型パターン	/</N1(人間)は>!VP2(身体動作, 思考動作)^rentai! のを / V3(結果, 知覚動作).kako.
英語文型パターン	<I N1> V3^past to VP2^base .
変数情報	VP2=レストランで男が食い逃げする, V8=見 (見る)
日本語文型パターン原文	葉書に宛名を書くのを忘れた。
英語文型パターン原文	I forgot to address the postcard.

あった¹。そして 66 文中で句レベル文型パターンが適合している文は 56 文であった。この 56 文に対し、単語レベル文型パターン翻訳方式と句レベル文型パターン翻訳方式の間で対比較評価を行った。対比較評価では、評価基準に adequacy を用い、二つの翻訳方式のうちどちらがより適切な翻訳文を得ているか、入力文ごとに評価した。評価は評価者 A のみが行っている。評価結果を表 7.20 に示す。表 7.20 において、“句レベル○” が、句レベル文型パターン翻訳の方が適切な場合であり、“単語レベル○” が、単語レベル文型パターン翻訳の方が適切な場合である。“差なし” は両翻訳方式間で翻訳精度に差がみられない場合であり、“同一” は出力結果が両翻訳方式で同一であった場合である。

表 7.20 では、句レベル文型パターン翻訳がより適切な文は 5 文で、単語レベル文型パターン翻訳がより適切な文は、句レベルよりは多い 13 文であった。この結果から、単語レベル文型パターン翻訳方式の方が句レベル文型パターン翻訳方式よりも翻訳精度が高いと考えられる。

¹実験に使用した 1,000 文のうち、学習文に同一の日本語文を持つ入力文は、34 文であった。単語レベル文型パターン翻訳では、1 文 (34 - 33 文) において自己日本語文型パターンが適合していないことになる。これは日本語文型パターンの記述ミスが原因であった。また句レベルでは 9 文 (34 - 25 文) で自己日本語文型パターンが適合していないことになる。この原因は日本語文型パターンの記述ミスと日本語句パターンの記述ミスによる。

表 7.20: 対比較評価結果

	句レベル○	単語レベル○	差なし	同一
文数	5	13	32	6

7.11 句レベル文型パターン翻訳方式のまとめ

本章では句レベル文型パターン翻訳方式を実装し、翻訳実験を行った。句レベル文型パターンは句変数を持つため、句レベル文型パターン翻訳方式では、句パターンと単語レベル文型パターン翻訳方式を利用することで、句変数に対応する要素の翻訳を行った。翻訳実験の結果適合率は約28%であり、単語レベル文型パターン翻訳の11%に比べ大幅に向上した。しかし翻訳精度は比較システムよりは若干高いが、単語レベル文型パターンに比べ低い。翻訳精度の低さは不適切な文型パターンの適合と句変数に対応する要素の翻訳精度の低さが主な原因である。不適切な文型パターンの適合は、句変数に対する意味属性によるパターン絞込み方法が適切ではないことが原因である。不適切な文型パターンを使用した場合、精度の高い翻訳文を得る可能性は非常に低いため、まず適切な文型パターンの絞込み方法を見つける必要がある。次に句変数に対応する要素の局所的翻訳の問題は、句パターン抽出の精度向上や関数の処理の追加等が必要である。

第8章 節レベル文型パターン翻訳方式

現在、節レベル文型パターン翻訳方式は実装中であるため、本章では実装のための仕様を説明する。節レベル文型パターンは単語、句、節の3つの変数を持つ。そのため節レベル文型パターンを用いて翻訳を行うためには、これらの変数に対応する要素の翻訳が必要となる。節変数に対応する要素の翻訳には、句変数の場合と同様に節パターンを抽出し翻訳に利用する。翻訳手順は句レベル文型パターン翻訳方式と同様にして翻訳できるが、句レベル文型パターン翻訳において明らかになった問題が、そのまま節レベル文型パターン翻訳の問題点として当てはまる可能性が高い。ただし句に比べ節の境界の決定は曖昧さが少ないため、同様の手順を用いるだけで句レベル文型パターン翻訳よりも高い翻訳精度が得られる可能性がある。また節変数に対応する要素の範囲は句よりも大きく、同時に節パターンの対応可能な要素の範囲も句パターンよりも広いため、文型パターンの適合率の改善が見込める。

統語的意味的分類は節間表現や節の意味属性の形式で付与されているため、文構造の持つ抽象的意味の観点からは、節レベル文型パターンが最も統語的意味的分類との関連性が高い。節レベル文型パターンの翻訳性能の調査と問題分析から、統語的意味的分類と文構造の持つ意味の関係についても新たな知見を得られる可能性がある。

8.1 節レベル文型パターン翻訳を実装するために必要な事項

句レベル文型パターンの場合と同様に節レベル文型パターン翻訳では節パターン辞書を抽出する必要がある。節パターン辞書は、節レベル文型パターンと句レベル文型パターンの対応をとることで句レベルの節パターンが、単語レベル文型パターンとの対応をとることで単語レベルの節パターンが抽出できる。句変数対応部位の局所的翻訳の精度が十分ではないため、まずは単語レベルの節パターン

の抽出が妥当である。

節変数に付与された関数の処理は、単語変数と句変数の場合に比べ処理が複雑になる。また新たな関数や変数の追加が必要となる可能性がある。例えば表 4.3 の節レベル文型パターンでは、英語原文文頭の命令表現 “Cross the street” が “CL2^base” の形式に変換されているが、“^base” は本来動詞原形を指定するだけの関数であり、この場合は動詞原形で始まる命令表現に変換しなければならない。したがって語形の変換だけでなく語順の変更も必要となる可能性がある。

8.2 節レベル文型パターン翻訳方式の手順

単語レベルの節パターンを使用する場合は、句レベル文型パターン翻訳方式と同様の手順となる。句レベルの節パターンを使用する場合は、節変数に対応する要素の局所的な翻訳に句レベル文型パターン翻訳方式を利用する。この場合は節パターンを用いた翻訳の内部でさらに句パターンを用いた翻訳を行う。単語レベルの節パターンと句レベルの節パターンは、同時に両方の節パターンを使用する場合と、単語レベル、句レベルの順に節パターン照合を行う場合の、二つの組み合わせ方が可能である。

また節レベル文型パターンも単語変数と句変数を持ち得るので、単語変数の場合は単語レベル文型パターン翻訳と同様に単語辞書を用いて、句変数の場合は句レベル文型パターン翻訳と同様に句パターンを用いて局所的に翻訳する。

8.3 形態素解析

単語レベル、句レベル文型パターン翻訳方式と同様である。

8.4 文型パターン照合

節レベル文型パターンを用いるパターン照合を行うことを除いて、単語レベル、句レベル文型パターン翻訳方式と同様である。

8.5 文型パターンの絞込み

単語レベル、句レベル文型パターン翻訳方式と同様にして可能であるが、句レベル文型パターンの場合以上に不適切な文型パターンが絞込み後に残る可能性が

ある。入力文の文構造の持つ意味を文型パターンを用いて抽出するという、文型パターン翻訳本来の視点からは、何かしらの方法でより適切な絞込みの方法が必要となる。

8.6 節レベルの英文生成

句レベル文型パターン翻訳方式に加え、節変数に対応する要素を節パターンを使用して翻訳する。句レベルの節パターン（句変数を持つ）を使用する場合は、句レベル文型パターン翻訳方式を利用し、単語レベルの節パターン（単語変数のみを持つ）を使用する場合は単語レベル文型パターン翻訳方式を利用する。

節変数に付与された関数の処理は現時点では決定できない。節レベル文型パターン翻訳方式を実装した後、実験を通して各関数の処理方法を決定する必要がある。

最後に、複数候補からの訳語選択は句レベル文型パターン翻訳と同様の方法で行う。

8.7 節レベル文型パターン翻訳における調査課題

翻訳性能の調査だけでなく、文構造の持つ意味に関連した調査が必要となる。そして調査結果を元にして、各種の処理方法や追加の処理を決定する必要がある。調査課題を下にまとめる。

- 節レベル文型パターンおよび節パターンの適合率
- 入力文に対する翻訳精度および節に対する翻訳精度
- 適切な関数の処理方法
- 節レベル文型パターンの絞込み方法
- 単語変数と句変数を持たず、節変数のみを持つ文型パターンによる翻訳
- 文構造の持つ意味と節レベル文型パターンの持つ意味の関係

第9章 統語的意味的分類と文構造の持つ意味に関する考察

非線形言語モデルでは、話者における対象への主観的な視点は、様相表現等の要素だけでなく、語順や語形、品詞の種類等、それらの一文全体での関係、つまり文構造として表現されると仮定している。そして非線形言語モデルに基づく文型パターンによって、文構造の持つ意味を保存したまま文構造変換が達成されると想定している。さらに統語的意味的分類によって、文構造の持つ意味を明示的に抽出できると期待している。

現時点ではこれらはまだ仮定の域を出ない。これら仮定の証明は非常に難しいため、この仮定の妥当性の検証はほとんど行われていない。本章ではまず、統語的意味的分類の記述形式の妥当性等、調査可能な項目から調査を行い、文構造の持つ意味と文型パターンの関係を明らかにするための方策を模索する。

9.1 統語的意味的分類の記述形式の妥当性

非線形言語モデルに基づき、重文複文文型パターン辞書が作成されている。その目的の一つは複雑な文構造を持つ重文と複文に対し、文構造の持つ意味を明示的に抽出し、言語分析や機械翻訳に利用可能とすることである。文構造の持つ意味は、統語的意味的分類として付与されている。統語的意味的分類の付与は、現在、統語的分類や節間意味分類のように、複数の項目に分けて記述されている。この分類情報を用いた、文型パターンの選択可能性に関する調査は行われているが[22]、現在の記述形式の妥当性やその有効性の調査は、その困難さによりまだ行われていない。ところで単語レベル文型パターンを用いた翻訳において、文型パターンが適合した場合は翻訳精度が高いことが明らかになっている。そこで本節では、統語的意味的分類の記述形式の妥当性を検証するため、文型パターン翻訳の翻訳結果を利用して調査を行う。調査項目を以下にまとめる。

1. 入力文と文型パターン間の、節の意味属性の一致率
2. 入力文における変数に対応する要素の割合

3. 具体例を用いた分析から記述形式の問題点を抽出

調査には6章の単語レベル文型パターン翻訳の結果を利用する。また3人の評価者のうち評価者Aの評価結果のみを使用する。

9.1.1 英語文型パターンの適切さと翻訳精度による分類

人手評価を行った100文中、適切な英語文型パターンを使用した文は75文、不適切な文型パターンを使用した文は25文であった。この25文中、形態素解析エラーや文型パターン記述ミス等による文は18文であった。これら18文を除き、残り82文を調査の対象にする。英語文型パターンが適切で翻訳精度の低い文は、主に未知語と不適切な訳語選択による。しかし未知語と訳語選択の問題を、文型パターンの問題と明確に区別することは難しい。次節からは、以下の3通りに分類して調査を行う。人手評価値4または5を翻訳精度が高い、評価値1から3を翻訳精度が低いとみなす。

- (a) 文型パターンが適切かつ翻訳精度が高い:50文
- (b) 文型パターンは適切だが翻訳精度が低い:25文
- (c) 文型パターンが不適切なため翻訳精度が低い:7文

9.1.2 節の意味属性の一致率

6章の翻訳実験は、学習に用いなかった日本語原文をテスト文に使用したため、各入力文が文型パターンと統語的意味的分類を持つ。統語的意味的分類に対し、入力文の文型パターンと翻訳に使用された文型パターンの間で一致する割合を調べた。このうち統語的分類、節間意味分類、節間キーワードの3つは一致率が90%を越えていた。表9.1に、節の意味属性の調査結果を示す。表9.1では、全一致は全ての節の意味分類が一致した文の数、部分一致は少なくとも一つの節が一致した文の数、全不一致は一致した節がない文の数、一致率は全一致節数/全節数を示している。なお統語的分類に記述ミスがあった文が4文あり、これらは以後分析から除く。

表9.1において、(c)ではほとんどが全不一致であり、逆に(a)では全不一致は少数しかみられない。したがって節の意味属性が入力文と文型パターン間で一致するほど翻訳精度が高くなることを示している。しかし(a)においても一致率は約60%であり、高い翻訳精度を得るのに、節の意味属性の一致が必須ではないことを示している。

表 9.1: 節の意味属性の一致率

	全一致	部分一致	全不一致	一致率
(a)	19	21	8	0.62(63/102)
(b)	10	8	6	0.58(28/48)
(c)	0	1	5	0.08(1/12)

9.1.3 変数に対応する要素の割合

文型パターンの長所は、抽象化された文構造でありながら元の文構造の持つ意味を保存できることである。しかし日英で直接対応の取れる要素を変数化したため、日英で個別に変数化した場合に比べ、変数化要素の割合が低下している。変数化要素の割合が低下するほど、文型パターンの保存する文構造の持つ意味は抽象度を失っていき、文構造を抽出することの効果が減退していく。したがって文型パターン翻訳方式は、文型パターンの変数の割合が低下するごとに、翻訳メモリを用いた翻訳方法に近似していくことになる。ただし日英対訳文対間での要素の対応を、自動で適切にとることは難しく、人手で対応を調べて変数化を行った文型パターンの方が翻訳精度は高くなると考えられる。

表 9.2 に、入力文における変数に適合した形態素の割合を示す。表 9.2 中の数は数値 % 以上 ~ 数値 % 未満の文数を示している。表 9.2 において、(a) と (b) の間で大きな差は見られない。また変数化要素の極端に少ない文型パターン (~20%) を使用したのはわずか 5 文であり、文構造の抽象度をある程度高く保った状態で、翻訳への効果をもたらしているとみなせる。

表 9.2: 入力文における変数に対応する要素の割合とその内訳

%	~20	20~30	30~40	40~50	50~
(a)	4	4	14	24	2
(b)	1	2	8	9	4
(c)	0	1	2	1	2

9.1.4 具体例を用いた分析

本節では具体例を用いて特徴や問題点を抽出する。表 9.3 に、(a) における、入力文と使用文型パターン間で節が全不一致かつ変数に対応する要素の割合が 40%

以上の例を示す¹。

まず“入力文の各種情報”からわかる事項を述べる。第二節の意味属性のうち、“地域社会生活の行為”と“移動行為の表現”は日本語表現“ため息を漏らす”の意味属性として不適切である。むしろ“知覚と情緒の表現”と“日常生活の行為”を付与すべきである。節の意味属性は用言の意味属性を元にして付与している。節の意味属性の過剰な付与または不足は、調査を行った入力文と日パターン原文の約半数（85/156）においてみられたため、節の意味属性を付与し直す必要がある。また日パターンの動詞変数 V3 の意味属性のうち、“属性”、“結果”、“生成”、“感情動作”、“思考動作”は、この日本語文では不適切である。変数の意味属性は自動で付与しており、用言の変数の意味属性は多義性を考慮した付与方法を用いるべきと考えられる。

次に“使用文型パターンの各種情報”を見ると、節間意味分類の“副詞的引用節/因果関係/原因/時間的原因”が不適切である。節間意味分類は節間キーワードを元にして付与されており、節間意味分類の過剰な付与または不足は、約20%（33/156）においてみられた。節の意味属性に比べて少ないが、こちらも付与し直す必要がある。以上より現在の付与精度では、統語的意味的分類の翻訳への利用は困難と考えられる。また複数項目への分割や4階層の記述により、統語的意味的分類を用いて言語分析を行うには熟練を要する可能性が高い。言語分析への利用には、複数の項目をまとめる新たな記述項目の追加が必要と思われる。

最後に、入力文と使用文型パターンの各種情報を比較し、考察を行う。入力文と日パターン原文では、言明されている事象が明らかに異なっている。そこで表9.3の節の意味属性と統語的意味的分類を用いて、人手で文全体の文構造の持つ意味を抽出すると、一例として表9.4の記述が得られる。表9.4における節の意味属性と節間意味分類は、表9.3から適切と判断できる記述を選択して用いている。表9.4より、入力文と日パターン原文共に時間的経過を表現する文構造を持ち（“一節の行為の後に二節の行為”）、そのためこの文型パターンにより高い翻訳精度の文が得られている。以上より、適切な節の意味属性と節間意味分類を付与できれば、現在の統語的意味的分類の記述形式のまま、文型パターンの絞込み等に利用可能と考えられる。また表9.4の例は、統語的意味的分類の新たな記述項目の表現形式を示唆するが、詳細な検討が必要のため、稿を改めて記述する。

¹節の意味属性は4階層のうち抽象度の最も高い1階層目のみ記述している。日パターンは日本語文型パターンで英パターンは英語文型パターンを示す。入力文の各種情報における日パターン原文と英パターン原文は、入力文と参照文と同じため省略。

表 9.3: 分析を行った具体例

入力文	母は家計簿を見てため息を漏らした。
参照文	As my mother looked over the family finances, she sighed.
出力文	Mother looked household account book and breathed a sigh .
人手評価	○
入力文の各種情報	
日パターン	/\$1^{}/N1(女, 母) は }/N2(文書類, 帳) を/V3(属性, 結果, 身体動作, 生成, 知覚動作, 感情動作, 思考動作)(て—で)\$1/N4(感嘆, 呼吸) を/漏らした。
英パターン	As N1 V3 ^{past} over N2, N1 ^{pron} V4 ^{past} .
統語的分類	第一節が従属節の述部二つの重文
一節意味属性	知覚と情緒の表現, 知的な行為の表現, 日常生活の行為
二節意味属性	知的な行為の表現, 地域社会生活の行為, 現象事象の表現, 移動行為の表現
節間意味分類	副詞的引用節/因果関係/原因/時間的原因, 並列節/順接的並列/総記の並列/なし
節間キーワード	て
使用文型パターンの各種情報	
日パターン	/</N1(人間) は >/N2(作物, 野菜) を/V3(属性変化, 身体動作)(て で)</N4 を >V5(状態, 身体動作)#6(.genzai .kako)。
英パターン	<You N1> V3 N2 and V5#6(^present ^past) <them N4^obj>.
変数情報	N1=母, N2=家計簿, V3=見(見る), N4=ため息, V5=漏らし(漏らす)
日パターン原文	野菜を煮て食べる。
英パターン原文	You boil the vegetables and eat them.
統語的分類	第一節が従属節の述部二つの重文
一節意味属性	日常生活の行為
二節意味属性	日常生活の行為
節間意味分類	入力文と同じ
節間キーワード	て

9.1.5 統語的意味的分類に対する調査のまとめ

重文複文文型パターン辞書の統語的意味的分類に対し、文型パターン翻訳の翻訳結果を用いて記述形式の検討を行った。分析結果から、統語的意味的分類の不適切な付与が多数存在することが明らかになった。考察から適切な統語的意味的分類を付与すれば、現在の記述形式で文型パターン翻訳に利用可能であることが示唆された。しかし言語分析への利用では、新たな記述項目の追加が必要と示唆された。

表 9.4: 文構造の持つ意味の抽出例

入力文の統語的意味的分類	
一節意味属性	日常生活の行為
二節意味属性	日常生活の行為, 知覚と情緒の表現
節間意味分類 1	因果関係
文構造の意味 1	一節の行為 (原因) により二節の行為および表現 (結果)
節間意味分類 2	順接的並列
文構造の意味 2	一節の行為の後に二節の行為
日パターン原文の統語的意味的分類	
一節意味属性	日常生活の行為
二節意味属性	日常生活の行為
節間意味分類	順接的並列
文構造の意味	一節の行為の後に二節の行為

9.2 文型パターン翻訳と tree-to-string の SMT における文構造の違い

構文解析結果を利用する翻訳方式は比較的高い翻訳精度を持つことが一般に知られているが、日英のような性質の大きく異なる言語間では、十分な翻訳精度はまだ得られていない。

文型パターン翻訳方式と tree-to-string の SMT 共に入力文の文構造を抽出し利用する点で共通する。文型パターン翻訳方式では、入力文に適合する日本語文型パターンが入力文から抽出された文構造にあたり、その対となる英語文型パターンが変換された英語の文構造にあたる。tree-to-string の SMT では、入力文の句構造解析結果を文構造として、適合する翻訳ルールを組み合わせることで、出力英語文においても階層的な文構造が得られている。本章では、文型パターン翻訳方式における、文型パターン対による文構造の抽出と変換の効果調べるため、tree-to-string の SMT で作成される文構造と文型パターン対の違いを検討する。

調査には6章の単語レベル文型パターン翻訳と比較システム (Travatar) の結果を利用する。また3人の評価者のうち評価者 A の評価結果のみを使用する (6.7.5 節)。

9.2.1 tree-to-string の SMT において作成される文構造

tree-to-string の SMT である Travatar (比較システム) においても, 入力文に適合する句構造と, それに対応する英語文構造が作成される. また日英文型パターンにおける変数は翻訳ルールにおける非終端記号に相当し, 各翻訳ルールは大別して次の3つに分類できる.

ルール1 文全体の翻訳ルール

ルール2 非終端記号を含む句翻訳ルール

ルール3 非終端記号を含まない句翻訳ルール

図9.1に比較システムにおいて作成される日英の文構造の例を示す. 日本語の文構造は品詞情報を持つ句構造であるが, それらは記述の簡略化のため省略した. 比較のため同入力文で使用された文型パターンを図9.1の下部に示している.

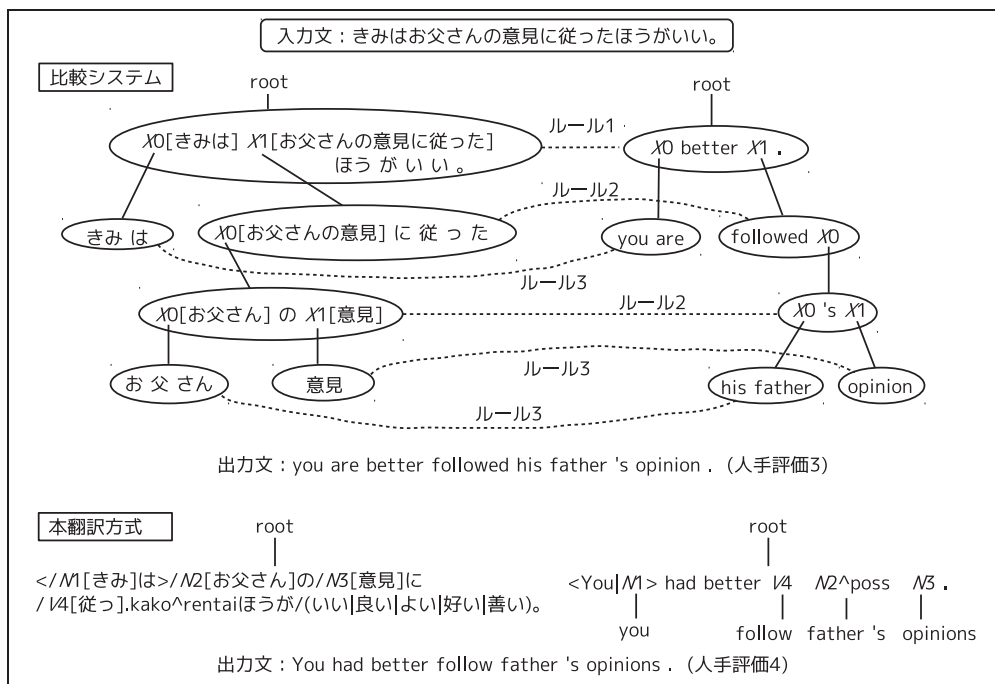


図 9.1: 比較システムで作成される日英の文構造

この例は root と終端記号を含め, 5階層の木構造となっている. ルール1とルール2による階層構造が文型パターンと対比可能であり, 文型パターンはルール2を持たない3層限定 (root と終端記号を含む) の木構造とみなすことができる. ルー

ル3は、本翻訳方式における形態素の翻訳およびその訳語選択と、ほぼ同等とみなせる。実際には非終端記号は品詞情報を持つため、翻訳ルールの適合には品詞による制約がある。

9.2.2 比較システムにおける文構造の階層数

人手評価を行った100文のうち、未知語を含まない文は68文であった。この68文中、句構造解析に失敗していた文は12文であり、この12文を除いた56文に対し文構造の階層数を調査した。

56文の階層数による分類を表9.5に示す²。最小で3階層、最大で7階層であり、平均4.6階層であった。また3階層で構成された文は5つだけであったため、ほとんどの文構造が、ルール1の適用後にルール2を適用して多階層で構成されていることを示している。階層数と人手評価値の相関係数は $r = -0.14$ （有意水準5%の棄却限界値は0.26）であり、これらの間の相関はほとんど見られなかった。

表 9.5: 文構造における階層数とその内訳

評価値	階層数					合計
	3	4	5	6	7	
5	2	6	7	1	0	16
4	0	2	0	0	0	2
3	1	3	6	0	0	10
2	1	5	8	2	1	17
1	1	4	6	0	0	11
合計	5	20	27	3	1	56

9.2.3 使用された翻訳ルールの適切さ

次に使用された翻訳ルールが適切であったか調査した。上記56文に対する、使用された不適切な翻訳ルールの種類（ルール1から3）と人手評価値との関係を表9.6に示す。各値は文数を示しており、“なし”は不適切なルールの使用がない文数を示している。翻訳ルールの探索はrootから下層に向かって行われるため[14, 15]、上層で不適切な翻訳ルールが適合した場合は、その下層においてどの翻訳ルール

²root直下の「日:X0。→英:X0。」の翻訳ルールは階層数のカウントから除いた。また56文の評価値は、表6.6の各文の評価値を再度用いている

表 9.6: 使用された不適切な翻訳ルールの種類による分類

評価値	ルール 1	ルール 2	ルール 3	なし	合計
5	1	1	0	14	16
4	0	2	0	0	2
3	3	6	1	0	10
2	3	11	3	0	17
1	7	4	0	0	11
合計	14	24	4	14	56

を用いても適切に翻訳できない可能性がある。そのため複数の不適切なルールの使用がある場合は、上層から下層のルールの順に、つまりルール 1 から 3 の順に振り分ける。不適切な翻訳ルールには、「日英で対応の取れていない誤ったルール³」と、「ルールそのものとしては良いがその入力文における使用では不適切なルール」の二通りがある。

表 9.6 において、不適切なルールの使用により不適切な文構造が作成されたのは 38 文（ルール 1 の 14 文とルール 2 の 24 文）であり、その大半の 34 文（ルール 1 とルール 2 の評価 1 から 3 の合計）で翻訳精度が低い。そして残り 18 文（ルール 3 の 4 文となしの 14 文）のうち 14 文（なしの評価 5 の 14 文）で高い翻訳精度が得られていることから、tree-to-string の SMT でも文全体の構造を利用することにより高い翻訳精度を得る効果があるとみなせる。ただしあまりにも適切な文構造を得る割合が低いため（56 文中 18 文）、現時点では文構造の利用が翻訳精度の向上にさほど寄与していない。

SMT では、不適切なルールが 2 つ以上組み合わせあって、偶然に適切な表現となる事例がしばしば見られる。表 9.6 では、不適切な文構造が作成された上記 38 文中、不適切なルールの組み合わせにより偶然翻訳精度が 5 となったのは、わずか 2 文である。この結果も、tree-to-string の SMT において高い翻訳精度を得るには、適切な文構造の作成が重要であることを示唆している。

9.2.4 不適切な翻訳ルール使用の要因

表 9.6 で不適切な翻訳ルールが使用された 42 文（ルール 1 から 3 の合計 42 文）において、合計 48 個の不適切な翻訳ルールの使用が生じていた。この 48 個の内

³日英翻訳では、日本語で省略された要素を、何かしらの翻訳結果として英語文に挿入する必要が多々見られる。そのため厳密には日英で対応がとれていない翻訳ルールでも、例えば「日:X0 → 英:he X0」のような一部のルールは、対応の取れた正しいルールとみなした。

訳は、対応のとれていない誤ったルールの使用が27個で、残り21個がその文における使用では不適切な場合であった。表9.7に、日英で対応の取れていない誤った翻訳ルールの例を示す。表中の右矢印の左側が日本語側で、右側が英語側である。なお翻訳ルールにおける句構造と品詞情報は省略した。

句構造解析は68文中56文で成功しているため(9.2.2節)、句構造解析失敗が誤った翻訳ルール作成の主要因とは考え難い。日英対訳文対間のアライメント情報をもとに翻訳ルールを作成しているため、対訳文対間におけるアライメントの精度の低さが、不適切な翻訳ルールおよび不適切な文構造の作成される最大の原因と考えられる。また、日英間でアライメント付けされなかった形態素列をどのように翻訳ルールに含めるかが、翻訳ルールの正確性に影響するため、翻訳ルールを抽出するアルゴリズムの問題ともみなせる。以上よりアライメントの精度の向上と、翻訳ルールに含める形態素列の境界の最適化により、翻訳精度が向上する可能性がある。

表 9.7: 日英で対応の取れていない誤った翻訳ルール

誤った翻訳ルールの例	ルールの種類
X0 X1 者だ → X0 X1	ルール 1
X0 になった → X0	ルール 2
わかる → do not know	ルール 3

ただし、対応のとれた翻訳ルールが多数作成されたとしても、その入力文においては不適切な翻訳ルールが選択される可能性が残る。図9.1における、ルール2の「X0に従った → followed X0」のルールがこれにあたる。入力文は提案の文であり、動詞“follow”が現在分詞か、“had better”に続くときは原型であるべきだが、翻訳確率により過去形の“followed”が選択されている可能性が高い。これは階層を下に降るときに、文法的な制約が伝達されずに消失してしまうことによると考えられる。この例は、多階層の文構造を用いて翻訳を行う場合に、階層間において文法的制約を伝達する必要性があることを示唆している。

9.2.5 翻訳ルールの選択

比較システムにおいても、まず一文を包含する翻訳ルール(ルール1)を適合させるが、文型パターン翻訳に比べ、ルールを作成および適合するための制約が少ない。これによりルール1とルール2共に、適合における柔軟さがもたらされるが、逆に適切なルールを選択する困難さが増大する。表9.6のルール1とルール2

で不適切な翻訳ルールが使用された 38 文のうち、「ルールそのものとしては良いがその入力文における使用では不適切なルール」が使用されたのは 20 文であった。調査を行った 56 文から、ルール 1 とルール 2 において「日英で対応の取れていない誤ったルール」を使用した 18 文を除くと、約 53% (20/38) の文で、その入力文においては不適切なルールの選択により、不適切な文構造が作成されている。比較システムにおいても、確率値と異なる翻訳ルールの選択基準が必要となる可能性がある。

9.2.6 両システムにおける文構造による翻訳精度への効果

評価に使用された 100 文中で、適切な英語文型パターンを使用したのは 75 文であった。75 文中 50 文で翻訳精度が高いことから、本翻訳方式では適切な英語文構造を用いて高い翻訳精度を得ていることがわかる。このことより文型パターンにより日英間の文構造変換が適切に行われたため、高い翻訳精度を得たと考えられる。また表 9.6 において、56 文中 18 文で適切な文構造が得られ、18 文中 14 文で翻訳精度が高いことから、比較システムでも文構造を抽出することによる翻訳への効果が存在するとみなせる。しかし比較システムでは不適切な英語文構造の作成される割合が高く、そのため翻訳精度が低くなっている。これらの結果は、両システム間におけるアライメントの精度と変換（翻訳）ルールの抽出方法の違いが原因と考えられる。

比較システムにおいて、多階層的に木構造を作成することによる、翻訳精度への具体的な効果は、明らかにはならなかった。分析結果より、多階層的な木構造を利用する場合は、階層間における文法的制約の伝達の必要性が示唆されている(9.2.4 節)。一方、単語レベル文型パターンは少数階層 (3 階層) 限定の木構造とみなせる。表 6.6 の結果は、3 階層限定の文パターン対でも、高い翻訳精度を得ることが十分可能であることを示している。一文全体の文構造の変換を、一つの変換ルール (単語レベル文型パターン) で行うため、階層間の文法的制約の伝達等を考慮しなくてよいことが、翻訳精度の高い要因の一つと考えられる。

句レベル文型パターン翻訳では、句変数に対応する要素に、単語変数を持つ句パターンを適合させることから、句レベル文型パターン翻訳で得られる文構造は 4 階層の木構造とみなせる。また単語レベルの節パターンを使用する節レベル文型パターンは 4 階層、句レベルの節パターンを使用する節レベル文型パターンは 5 階層の木構造とみなせる。以上のことより句レベルと節レベルの文型パターンにおいても、階層を上層から下層に降るときに文法的制約の伝達が必要であると考えられる。

9.3 文構造の持つ意味に関する調査課題

非線形言語モデルにおける「文型パターンによる文構造の持つ意味の保存による精度の高い翻訳の実現」という仮定を検証するためには、「文構造の持つ意味」の具体的な抽出と、日英での文構造の対応を明らかにすることが必要となる。下に調査すべき項目をまとめる。

1. 節構造の持つ意味と節の意味分類の関係
2. 日本語文構造と英語構文構造の関係
3. 重文複文全体の文構造の持つ意味を表現可能な、統語的意味的分類の新たな記述項目の策定

非線形言語モデルは当初、単文を対象にその構造の持つ抽象的な意味に着目して考案されており、むしろ分類項の付与は単文に対する方が行いやすい。重文と複文では節が単文にあたると考えて良く、節構造の持つ抽象的な意味を抽出できるはずである。統語的意味的分類では節の用言の意味属性が用いられているが、重文と複文全体の文構造の持つ意味は、節構造の持つ意味と節間表現の意味によって生成されとも考えられる。したがって1の調査が必要となる。

9.1節での分析と考察は、日本語の統語的意味的分類のみを対象としており、翻訳への効果を検証するために、2の調査が必要である。そして文構造の持つ意味の保存による翻訳への効果を示すためには、3の調査が必要になると考えられる。

第10章 文型パターン翻訳方式の現状 と今後の研究方針

第6章の実験結果から、単語レベル文型パターン翻訳方式では、文型パターン適合率は約11%と低いが、文型パターンが適合した場合は高い翻訳精度（人手評価平均3.44）を持つことが明らかになった。また第7章の実験結果から、句レベル文型パターン翻訳方式では、文型パターン適合率は約28%と単語レベルに比べ改善されているが、翻訳精度は単語レベルよりも低かった（人手評価平均2.77）。句レベル文型パターン翻訳の翻訳精度は比較システムよりは高いため、実際の翻訳への利用では、単語レベル文型パターン翻訳、句レベル文型パターン翻訳、比較システムの順に適用するハイブリッド方式が実現可能である。句レベル文型パターン翻訳方式の翻訳精度向上と共に、節レベル文型パターン翻訳方式の実装によりパターン適合率の向上を目指す予定である。その後は述部が4つ以上の長文に使用できるように、文の分割と文型パターンの適合の組み合わせ方等を調査する必要がある。

統語的意味的分類の記述の妥当性の調査から、統語的意味的分類の不適切な付与が多数存在することと、言語分析への利用では新たな記述項目の追加が必要であることが示唆された。また文型パターン翻訳方式と tree-to-string の SMT における文構造の違いの調査から、文型パターンにより日英間の文構造変換が適切に行われたため、高い翻訳精度を得ていることが示唆された。文型パターンにおける日英の表現構造の関係、そして文型パターンから具体的に文構造の意味を取り出せるか、さらなる調査を行うことで言語分析や意味的等価変換方式への利用が期待できる。

第11章 まとめ

本研究では、非線形言語モデルに基づく日英文型パターン翻訳方式を提案し実装した。まず単語レベル文型パターン翻訳を実装し、それを元に句レベル文型パターン翻訳を実装した。実験結果から、単語レベル文型パターン翻訳は、パターン適合率が低い（パターン適合率約 11%）が翻訳精度が高い（人手評価平均 3.44）こと、それに対し句レベル文型パターン翻訳はパターン適合率は単語レベルより大幅に改善している（パターン適合率約 28%）が、翻訳精度は大きく低下している（人手評価平均 2.77）ことが明らかになった。句レベル文型パターン翻訳の翻訳精度の低い最大の原因は、不適切な文型パターンを使用している割合が高いことであった。翻訳精度改善にはまず適切な文型パターンの絞込み方法を探す必要がある。またパターン適合率をさらに改善する必要があると考えられる。また重文複文の文構造の持つ意味として付与されている、統語的意味的分類に対し記述の妥当性を調査した。調査結果より統語的意味的分類の不適切な付与が多数存在することが明らかになった。さらに文型パターン翻訳方式と、構文解析結果を利用する tree-to-string の SMT における文構造の違いを調査した。調査結果から、文型パターンによる文構造変換の精度が高く、そのため高い翻訳精度を得ていることが示唆された。今後は非線形言語モデルの仮説である、「文型パターンによる文構造の持つ意味の保存と変換」について、検証する必要がある。

非線形言語モデルは、文構造の持つ抽象的な意味を、文型パターンを用いて取り出し利用することを目的として考案されている。まず本研究の基盤となる、非線形言語モデルについて第 3 章で、重文複文文型パターン辞書について第 4 章で説明した。本研究で使用した重文複文文型パターン辞書は変数化の範囲に応じて、単語レベル、句レベル、節レベルの 3 レベルの文型パターンを持つ。これら 3 レベルの文型パターンは変数化の範囲が異なるため、変数に対応する要素の局所的な翻訳方法が大きく異なる。第 5 章においてまず 3 レベルの文型パターンに共通する翻訳手順をまとめた。以下、翻訳実験と調査の結果を章ごとにまとめる。

第 6 章で単語レベル文型パターン翻訳方式の実装を説明し、翻訳実験を行った。翻訳実験より文型パターン適合率は約 11% と低いが、文型パターンが適合した場合は、単語レベル文型パターン翻訳が高い翻訳精度を持つことが明らかになった。

adequacy による 5 段階の人手評価では、単語レベル文型パターン翻訳の平均値は 3.44 であり比較システムの 2.54 を大きく上回った。また 61 文中で翻訳精度の高い文は 31 文で精度の低い文は 30 文であり、約半数の文で精度の高い翻訳文が得られた。この 61 文中で適切な英語文型パターンが使用されたのは 44 文であり、不適切な文型パターンが使用されたのは 17 文であった。そして評価の高い 31 文は適切な文型パターンが使用された 44 文の中に全て含まれていた。このことは単語レベル文型パターンにより、文構造の持つ意味が消失しないように適切な文構造変換が行われていることを示唆している。

第 7 章では句レベル文型パターン翻訳方式を実装し、翻訳実験を行った。句レベル文型パターン翻訳では、句変数に対応する要素に対し、句パターンを用い単語レベル文型パターン翻訳方式を再利用することで翻訳を行った。句レベル文型パターンの適合率は約 52% であったが、句パターンの適合まで含めると、パターン適合率は約 28% であった。単語レベル文型パターン翻訳に比べ、パターン適合率は大幅に増加したが、それでお十分ではないことを示している。また翻訳精度も人手評価平均で 2.77 であり、比較システムの 2.26 よりは高いが、単語レベル文型パターン翻訳に比べると大きく精度が下がっていることがわかる。評価を行った 100 文中 44 文で不適切な句レベル文型パターンを使用しており、これが翻訳精度が低い最大の要因であった。翻訳精度改善にはまず適切な文型パターンの絞込み方法を探る必要があると考えられる。また適切な文型パターンを使用した場合でも、56 文中 29 文で翻訳精度が低い。句パターンの作成精度とその改善方法の調査が必要である。

第 8 章では節レベル文型パターン翻訳方式の実装方法を示した。節レベル文型パターン翻訳では節変数に対応する要素の翻訳を節パターンを使用して行う。そのため句レベル文型パターン翻訳の問題点が節レベル文型パターン翻訳にも当てはまる。その問題とは別に、節変数や関数の処理に節レベル文型パターン固有の処理が必要と考えられる。句レベル文型パターン翻訳で明らかになった問題点の改善とともに、節レベル文型パターン翻訳の実装により問題点を明らかにし改善する必要がある。

第 9 章において統語的意味的分類の記述の妥当性と、文型パターン翻訳方式と tree-to-string の SMT における文構造の違いを調査した。単語レベル文型パターン翻訳の翻訳結果を用いた分析から、統語的意味的分類の不適切な付与が多数存在することが明らかになった。また統語的意味的分類の言語分析への利用では重文複文全体の表現に対する新たな記述項目の追加が必要と示唆された。さらに文構造の違いの調査から、文型パターンによる文構造変換の精度が高く、そのため高い翻訳精度を得ていることも示唆された。今後は文型パターンによる文構造の持つ意味の保存の翻訳への効果を明らかにするため、文型パターンにおける日英の

表現構造の関係，そして句型パターンから具体的に文構造の意味を取り出せるか，さらなる調査が必要である．

謝辞

本研究を行うにあたって多数の方のご協力を頂きました。重文複文文型パターン辞書の作成は科学技術振興機構（JST）の戦略的基礎研究事業（CREST）で行っており、文型パターン辞書作成に携わった皆様に感謝いたします。また非線形言語モデルの考案に携わった皆様に感謝いたします。特にその指導的立場であられた故池原悟氏には、研究のご指導もいただきました。深く感謝いたします。非線形言語モデルに基づく言語分析の研究と、文型パターン翻訳に関する研究には、鳥取大学自然言語処理研究室（旧計算機工学講座 C 研究室）に所属する多くの方々のご助力を頂きました。研究室でご助力いただいた皆様に感謝いたします。特に英文生成システムの実装に多大なご助力を頂きました鳥取大学徳久雅人講師，石上真理子氏，水田理夫氏に感謝いたします。翻訳結果の人手評価を行って頂いた尾崎かおる氏，今仁優希氏に感謝いたします。

本研究を進めるにあたりご指導頂きました鳥取大学徳久雅人講師，鳥取大学村上仁一准教授，鳥取大学村田真樹教授に心よりお礼申し上げます。また本論文をまとめるにあたりご指導くださいました鳥取大学木村周平教授に深くお礼申し上げます。

参考文献

- [1] 蔵本由紀. 非線形科学. 集英社新書, 2007.
- [2] 長尾真. 岩波講座ソフトウェア科学 15 自然言語処理. 岩波書店, 1996.
- [3] 池原悟, 阿部さつき, 徳久雅人, 村上仁一. 非線形な表現構造に着目した重文と復文の日英文型パターン化. 自然言語処理, Vol. 11, No. 3, pp. 69–95, 2004.
- [4] 鳥バンク (Tori-Bank). <http://unicorn.eecs.tottori-u.ac.jp/toribank>.
- [5] 池原悟, 阿部さつき, 竹内奈央, 徳久雅人, 村上仁一. 意味的等価変換方式のための重文複文の統語的意味的分類体系について. 情報処理学会研究報告自然言語処理, Vol. 2006, No. 124, pp. 1–8, 2006.
- [6] S. C. Shapiro. Generalized Augmented Transition Network Grammars For Generation From Semantic Networks. *The American Journal of Computational Linguistics*, Vol. 8, No. 1, pp. 12–25, 1982.
- [7] 奥村学監修, 渡辺太郎, 今村賢治, 賀沢秀人, Neubig Graham, 中澤敏明. 自然言語処理シリーズ 4 機械翻訳. コロナ社, 2014.
- [8] J. Richardson, R. Dabre, C. Chu, F. Cromières, T. Nakazawa, and S. Kurohashi. KyotoEBMT System Description for the 2nd Workshop on Asian Translation. In *Proc. of the 2nd Workshop on Asian Translation (WAT2015)*, pp. 54–60, 2015.
- [9] P. Koehn and J. Senellart. Convergence of Translation Memory and Statistical Machine Translation. In *proc. of the Second Joint EM+/CNGL workshop*, pp. 21–31, 2010.
- [10] 高橋博之, 川崎立八, 牧田光晴, 樋口重人, 藤井敦, 石川徹也. 日英特許公報を用いた対訳辞書および翻訳メモリの構築. 情報処理学会研究報告自然言語処理 (NL) , Vol. 2003, No. 57, pp. 39–46, 2003.

- [11] 池原悟, 宮崎正弘, 白井諭, 林良彦. 言語における話者の認識と多段翻訳方式. 情報処理学会, Vol. 28, No. 12, pp. 1269–1279, 1987.
- [12] 金出地真人, 徳久雅人, 村上仁一, 池原悟. 結合価文法による動詞と名詞の訳語選択能力の評価. 自然言語処理, Vol. 11, No. 3, pp. 149–164, 2004.
- [13] G. Neubig. Travatar: A Forest-to-String Machine Translation Engine based on Tree Transducers. In *Proc. of the ACL Demonstration Track*, pp. 91–96, 2013.
- [14] Y. Liu, Q. Liu, and S. Lin. Tree-to-String Alignment Template for Statistical Machine Translation. In *Proc. of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the ACL*, pp. 609–616, 2006.
- [15] H. Mi, L. Huang, and Q. Liu. Forest-Based Translation. In *Proc. of the ACL-08:HLT*, pp. 192–199, 2008.
- [16] 池原悟. 非線形な言語表現と文型パターンによる意味の記述. 情報処理学会研究報告自然言語処理, Vol. 2004, No. 1, pp. 139–146, 2004.
- [17] 池原悟. 言語で表現される概念と翻訳の原理. 電子情報通信学会技術研究報告. TL, 思考と言語, Vol. 103, No. 487, pp. 7–12, 2003.
- [18] 池原悟, 徳久雅人, 村上仁一, 佐良木昌, 池田尚志, 宮崎正弘. 非線形な重文複文の表現に対する文型パターン辞書の開発. 情報処理学会研究報告自然言語処理, Vol. 2005, No. 117, pp. 157–164, 2005.
- [19] 村上仁一, 藤波進. 日本語と英語の対訳文対の収集と著作権の考察. 第一回コーパス日本語学ワークショップ, pp. 119–130, 2012.
- [20] 意味類型パターン記述言語仕様書. <http://unicorn.eecs.tottori-u.ac.jp/toribank/download/shiyou.pdf>.
- [21] 池原悟, 佐良木昌, 宮崎正弘, 池田尚志, 新田義彦, 白井諭, 柴田勝征. 等価的類推思考の原理による機械翻訳方式. 電子情報通信学会技術研究報告. TL, 思考と言語, Vol. 102, No. 491, pp. 7–12, 2002.
- [22] 中村聡, 村上仁一, 徳久雅人, 池原悟. 「意味」による文型パターン検索方式の最適化. 言語処理学会第12回年次大会発表論文集, pp. 632–635, 2006.

- [23] 池原悟, 徳久雅人, 竹内(村本) 奈央, 村上仁一. 日本語重文・複文を対象とした文法レベル文型パターンの被覆率特性. 自然言語処理, Vol. 11, No. 4, pp. 148–178, 2004.
- [24] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系. 岩波書店, 1997.
- [25] 徳久雅人, 村上仁一, 池原悟. 重文・複文文型パターン辞書からの構造照合型パターン検索. 情報処理学会研究報告自然言語処理(NL), Vol. 2006, No. 124, pp. 9–16, 2006.
- [26] 吉岡篤史, 徳久雅人, 村上仁一, 池原悟. 日英対訳パターンを用いた名詞句翻訳. 言語処理学会第12回年次大会発表論文集, pp. 580–583, 2006.
- [27] 石上真理子, 徳久雅人, 村上仁一, 池原悟. 結合価パターンを用いた動詞句の翻訳可能性の調査. 言語処理学会第11回年次大会発表論文集, pp. 364–367, 2005.
- [28] 石上真理子, 水田理夫, 徳久雅人, 村上仁一, 池原悟. 関数・記号付き文型パターンを用いた機械翻訳の試作と評価. 言語処理学会第13回年次大会発表論文集, pp. 67–70, 2007.
- [29] 前田春奈, 村上仁一, 徳久雅人, 池原悟. 用言意味属性を用いた適合文型パターンの絞り込み. 言語処理学会第12回年次大会発表論文集, pp. 576–579, 2006.
- [30] 原真一朗, 村上仁一, 徳久雅人, 池原悟. 日英機械翻訳における多変量解析を用いた最適パターンの選択. 言語処理学会第12回年次大会発表論文集, pp. 268–271, 2006.
- [31] 原田真. 日英重文複文パターン辞書からの節と句のパターン抽出. 2007年度鳥取大学卒業論文, 2008.
- [32] 坂田純. 非線形言語モデルに基づくパターン翻訳システムにおける句パターン辞書の改良. 2008年度鳥取大学卒業論文, 2009.
- [33] J. Sakata, M. Tokuhisa, and J. Murakami. Phrase-Level Pattern-Based Machine Translation Based on Analogical Mapping Method. In *Proc. of the 8th International Conference on NLP(japTAL2012)*, pp. 28–39, 2012.
- [34] 水田理夫, 徳久雅人, 村上仁一, 池原悟. 重文・複文文型パターン辞書による意識の可能性. 電子情報通信学会ソサイエティ大会講演論文集, pp. S55–S56, 2007.

- [35] 吉田大蔵, 徳久雅人, 村上仁一, 池原悟. 格助詞およびその相当表現のパターン翻訳の試み. 電子情報通信学会技術研究報告. TL, 思考と言語, Vol. 109, No. 413, pp. 13–18, 2010.
- [36] Graham Neubig and Shinsuke Mori. Word-based Partial Annotation for Efficient Corpus Construction. In *Proc. of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, pp. 2723–2727, 2010.
- [37] Y. Oda, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. Ckylark: A More Robust PCFG-LA Parser. In *Proc. of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT) (Software Demonstrations)*, pp. 41–45, 2015.
- [38] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. of the 40th Annual meeting on the Association for Computational Linguistics*, pp. 311–318, 2002.
- [39] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proc. of the 7th Conference of the Association for Machine Translation in the Americas*, pp. 223–231, 2006.
- [40] S. Banerjee and A. Lavie. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *Proc. of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pp. 65–72, 2005.
- [41] 平尾努, 磯崎秀樹, 須藤克仁, Duh Kevin, 塚田元, 永田昌明. 語順の相関に基づく機械翻訳の自動評価法. 自然言語処理, Vol. 21, No. 3, pp. 421–444, 2014.

付録A 本論文で使用した文型パターンの要素の説明

本論文に使用した文型パターンの変数の説明を表 A.1 に、関数の説明を表 A.2 に、字面の説明を表 A.3 に、記号の説明を表 A.4 にまとめる。

表 A.1: 本論文で使用した文型パターンの変数の説明

要素	説明
変数	線形要素。変数名，番号の順に記述し，同じ番号を持つ変数が日英で対応する。品詞と意味属性を持つ。日英で品詞の異なる要素でも変数化可能。
単語変数	品詞の同じ任意の単語に適合。複合名詞や複合動詞等にも適合可能。3レベルの文型パターンで使用。
<i>N</i>	名詞変数。日英で使用。
<i>AJ</i>	形容詞変数。日英で使用。
<i>V</i>	動詞変数。日英で使用。
<i>GEN</i>	限定詞変数。日でのみ使用。
<i>DEF</i>	定冠詞変数。英でのみ使用。
<i>AJV</i>	形容動詞変数。日でのみ使用。
<i>ND</i>	用言性名詞変数。日でのみ使用。
<i>NUM</i>	数詞変数。日でのみ使用。
<i>TIME</i>	時詞変数。日でのみ使用。
<i>REN</i>	連体詞変数。日でのみ使用。
<i>GEN</i>	限定詞変数。日でのみ使用。
<i>ADV</i>	副詞変数。日英で使用。
句変数	品詞の同じ任意の句または単語に適合。句レベルおよび節レベル文型パターンで使用。
<i>VP</i>	動詞句変数。日英で使用。
<i>NP</i>	名詞句変数。日英で使用。
<i>AJP</i>	形容詞句変数。日英で使用。
<i>AJVP</i>	形容動詞句変数。日でのみ使用。
<i>ADV</i>	副詞句変数。日英で使用。
<i>CL</i>	節変数。日英で使用。節レベル文型パターンのみで使用。
意味属性	変数のより細分化された分類。変数番号の直後の“()”内に記述。名詞意味属性と用言意味属性の2種類がある。

表 A.2: 本論文で使用した文型パターンの関数の説明

要素	説明
関数	非線形要素. 表記のゆれを吸収するための様相関数 (“ <code>.</code> 関数”) と, 語形を指定するための語形関数 (“ <code>^</code> 関数”) がある.
<code>.meireigo</code>	命令表現が前の要素に続くことを示す.
<code>.hitei</code>	否定表現が前の要素に続くことを示す.
<code>.kako</code>	過去表現が前の要素に続くことを示す.
<code>.sase</code>	使役表現が前の要素に続くことを示す.
<code>.suitei</code>	推定表現が前の要素に続くことを示す.
<code>^meirei</code>	直前の変数が命令形であることを示す.
<code>^rentai</code>	直前の変数が連体形であることを示す.
<code>^base</code>	直前の変数が原形であることを示す.
<code>^er</code>	直前の変数が比較級であることを示す.
<code>^past</code>	直前の変数が過去形であることを示す.
<code>^ed</code>	直前の変数が過去分詞であることを示す.
<code>^sase</code>	直前の変数が “せる” または “させる” を含むことを示す.
<code>^present</code>	直前の変数が現在形であることを示す.
<code>^pron</code>	直前の変数が代名詞であることを示す.
<code>^poss</code>	直前の変数が所有格であることを示す.
<code>^obj</code>	直前の変数が目的格であることを示す.

表 A.3: 本論文で使用した文型パターンの字面の説明

要素	説明
字面	非線形要素. パターン原文の文字で記述.
“が”, “なってから” 他	日本語出現形字面.
“after”, “turn” 他	英語出現形字面.
“ ’笑う ’”, “ ’使う ’” 他	日本語原形字面. その動詞の全ての活用形に対応.
“@be”, “@eat” 他	英語原形字面. その動詞の全ての語形が候補になり訳語選択において一つ選択される.

表 A.4: 本論文で使用した文型パターンの記号の説明

要素	説明
記号	適合範囲の拡大の役割を持つ。
/	文節境界を示し、本来は“/”の次に離散記号が続くが、本論文では使用しないため離散記号は省略する。
#1{ 要素 1, 要素 2}	順序任意記号。要素 1 と要素 2 が入れ替わっても適合可能なことを示す。
!	分節境界記号。挿入可能な要素のない文節を表す。
(要素 1 要素 2)	要素選択記号。日英で使用。日本語の場合は、要素 1 と要素 2 のどちらでも適合可能なことを示す。英語では要素 1 と要素 2 が訳語の候補として扱われる。
<M1 は > および <I M1>	補完記号。日英でセットで使用される。M1 に対応する語が入力文にあるときは英語文型パターンにおいて M1 が選択され、ないときは、“I” が選択される。
# 番号 [要素]	任意要素記号。日英でセットで使用。パターン照合の際、適合する要素があってもなくてもよい要素。
\$1^{ { 要素 } }	移動可能記号。日でのみ使用。{ } 内の要素が \$1 の位置にある場合も適合可能であることを示す。
\$1 # 番号 (関数 1 関数 2)	日英セットで使用。日本語で関数 1 と 2 のどちらかが適合したとき、英語側で対応する関数が選択される。

付録B 文型パターン翻訳方式の実験環境

単語レベル，句レベル文型パターン翻訳方式の実験環境を表 B.1 に示す。

表 B.1: 文型パターン翻訳方式の実験環境

仮想マシン	VirtualBox ver. 3.0.4_OSE r50677
ゲスト OS	VineLinux 4.2 32bit 版
ゲストメモリ	4GB
ゲストプロセッサ数	2
プログラム記述言語	Ruby ver. 1.8.5

表 B.2 に，単語レベル文型パターン翻訳方式における入力 5,000 文の翻訳時間を示す。文型パターンが適合して，実際に翻訳を行ったのは 572 文である。表 B.2 と表 B.3 では “time” コマンドの “real” の結果のみを示す。なお，入力 1 文ごとの最大メモリ使用量の平均（10 文の平均）は，約 36Mbyte であった。

表 B.2: 単語レベル文型パターン翻訳方式の実行時間

形態素解析，品詞コードと意味属性コードの付与	0m49.777s
パターン照合，パターン絞込み	15m53.673s
単語レベル文型パターン翻訳 (572 文)	29m15.782s
翻訳候補文からの出力文の選択	0m1.613s
合計	47m0.845s

表 B.3 に，句レベル文型パターン翻訳における入力 1,000 文の翻訳時間を示す。文型パターンが適合して，実際に翻訳を行ったのは 277 文である。入力 1 文ごとの最大メモリ使用量の平均（10 文の平均）は，約 68Mbyte であった。

表 B.3: 句レベル文型パターン翻訳方式の実行時間

形態素解析, 品詞コードと意味属性コードの付与	0m3.377s
パターン照合, パターン絞込み	7m57.070s
句レベル文型パターン翻訳 (277 文)	138m53.415s
翻訳候補文からの出力文の選択	0m3.447s
合計	146m57.309s

研究業績一覧

主論文

1. 坂田純, 村上仁一. 非線形言語モデルに基づく文型パターンを用いる日英機械翻訳方式. 電子情報通信学会, Vol. J101-D, No.1, January, 2018.
2. J. Sakata, M. Tokuhisa, and J. Murakami. Phrase-Level Pattern-Based Machine Translation Based on Analogical Mapping Method. In *8th International Conference on NLP(japTAL2012), Lecture Notes in Artificial Intelligence*, Subseries of *Lecture Notes in Computer Science*, LNAI, Vol. 7614, pp. 28-39, 2012.
3. J. Sakata, J. Murakami, M. Tokuhisa, M. Murata. Machine Translation Method Based on Non-Compositional Semantics (Word-Level Pattern-Based MT). In *The 2015 Conference of the Pacific Association for Computational Linguistics (PACLING2015)*, Part of the *Communications in Computer and Information Science book series (CCIS)*, vol. 593, pp. 225-237, 2016.

参考論文

1. 坂田純, 徳久雅人, 村上仁一. 意味的等価変換方式による句レベルパターン翻訳方式の調査. 言語処理学会第 18 回年次大会発表論文集, pp. 271-274, 2012.
2. 坂田純, 徳久雅人, 村上仁一, 村田真樹. 意味的等価変換方式に基づく単語レベルパターン翻訳方式の評価. 言語処理学会第 20 回年次大会発表論文集, pp. 298-301, 2014.
3. 坂田純, 村上仁一, 徳久雅人, 村田真樹. 文構造の利用による翻訳への効果の調査. 言語処理学会第 22 回年次大会発表論文集, pp. 1021-1024, 2016.
4. 坂田純, 村上仁一. 重文複文文型パターン辞書の統語的意味的分類に関連する調査. 言語処理学会第 23 回年次大会発表論文集, pp. 489-492, 2017.