

要件効果構造に基づく制限言語モデルを用いた法律文解析  
— 並列構造の推定方法について —

高尾宜之 平松寛司 永井秀利 中村貞吾 野村浩郷

九州工業大学 情報工学部

E-mail: {takao,hiramatu,nagai,teigo,nomura}@dumbo.ai.kyutech.ac.jp

本稿では、既に報告した要件効果構造モデルに基づく解析方法の枠組に取り入れる並列構造の推定法について述べる。提案する推定法は、統語解析をボトムアップに進めることを前提にして段階的に評価していくものである。まず、並列構造の解析結果に対し、経験則に基づく制約を用いて、間違っただけの構文構造を除去する。次に、並列要素の長さ、表層的・深層的類似性などに基づく評価を行い、並列構造の範囲を推定する。そして文全体の解析が終了した時点でその評価値がもっとも高いものを範囲推定の最適解として選ぶ。本研究では、この推定手法に対する評価実験を行ない、本手法における有効性について検証した。

Syntactic Analysis for Legal Sentences  
Based on Legal Condition-Effect Structure  
— A Method for Analyzing Coordinate Structure —

Yoshiyuki TAKAO Kanji HIRAMATSU Hidetoshi NAGAI  
Teigo NAKAMURA Hirosato NOMURA

Department of Artificial Intelligence

Kyushu Institute of Technology

Iizuka, 820, Japan

E-mail: {takao,hiramatu,nagai,teigo,nomura}@dumbo.ai.kyutech.ac.jp

Coordinate structures appear frequently in law sentences and it is difficult to determine the extent of each constituents of these structures in whole sentences correctly. In this paper, we extend a parsing method for controlled linguistic models based on condition-effect structure of legal sentences to be able to estimate the extent of coordinate structures. It uses many heuristics such as length of each constituent, similarity of expression, and so on and other particular restrictions of coordinate structure of legal sentences to assign the evaluation value to each constituent of syntactic structure, and accept the parse with highest evaluation value as the most suitable result.

## 1 はじめに

法律文は、現代社会のあらゆる状況に対応するため、大量に制定され続けており、これを実際に全て把握しようとすれば多大な労力が必要となる。このような背景の下、法律文エキスパートシステムや法律 CAI システムなどの法的支援システムの必要性が高まっている [1]。

この法的支援システムで扱う基礎データを整備するためには、多くの条件節や並列構造を持つ複雑な法律文の文構造を的確に捉え、その中から情報を抽出し、計算機向きのデータに変換する必要がある。我々は、これを実現するために従来から法律文に含まれる様々な言語的特徴に着目し、法律文制限言語モデルとして構築し、それに基づいて解析する方法について提案してきた [5][6]。

さて、その解析手法の枠組の中での並列構造の扱いについては、構成要素中の文構造解析において、制約を用いて解析結果の曖昧性を削減する方針により対処してきた。しかし、これらの制約の適用は構成要素内の体言や用言などの局所範囲に制限されており、構成要素全体、さらには法律文全体の構造の中で並列構造の範囲を正しく推定することに關しては不十分であった。

本研究では、従来提案してきた要件効果構造モデルに基づく解析手法の枠組に法律文の並列構造の推定法を新しく取り入れ、法律文解析の拡充を図る。また、2つの法律文に対する解析実験を行なうことで、本手法の有効性を検証する。

なお、分析の対象として「国際動産売買契約に関する国連条約（ウィーン統一売買法、全101条、318文）」[2]と「条約法に関するウィーン条約（全85条、322文）」[3]の2つの法律文を用いる。本稿では、特に断りのない限り、法律文とはこれらの条文を指すものとする。

## 2 従来における並列構造の解析

我々は、従来の研究において、法律文の骨格構造として存在する要件効果構造に着目し、それに基づく制限言語モデルと法律文解析について述べてきた [5][6]。

並列構造の解析については、このモデルを用いた解析の枠組をベースにし、構成要素内の構造解析を行う際に法律文の言語的制約と類似性に基づく制約をかけることにより、解析結果の曖昧性を極力排除する立場をとってきた。以下にその制約の例を挙げる。

### 1. 接続詞の結合強度に基づく並列構造の埋込関係

同タイプの接続詞において、「又は」<「若しくは」や「並びに」<「及び」のような用法に基づく結合強度の関係に着目し、結合強度の強い並列構造は結合強度の弱い並列構造を埋め込むことができないという制約をかける。

### 2. 読点制約

体言並列内に係り受け制限として使われている読点を含めてはならない。

### 3. 概念類似度に基づく評価

シソーラスを用いて前部と後部の末尾にある二単語間の概念類似度を評価式により導き出し、その類似度が閾値を越えていれば接続可能、越えていなければ接続不可能とする

以上の制約により、通常の構文解析で生じる曖昧性はかなり削減されるとされてきた。しかし、その一方でこの制約に基づく並列構造の解析には以下に挙げる問題点が含まれていた。

- 2つの並列構造の並列キーに含まれる接続詞の間に明確な結合強度の関係がない場合、1の制約はかけることができない(例:「又は」と「及び」)。
- 用言並列や部分並列においてはスコープ内に係り受け制限の機能を持つ読点を含む場合もあるので、2の制約はそれらに適用できない。
- 概念的な類似度が低くても並列されている場合(例:「時期、場所及び方式」のような列挙並列)があり、閾値により間違ったスコープ推定とされる場合がある。しかし、低く設定すればあらゆるパターンを接続可能としてしまい制約としての効果が出ない。

このようにこれらの制約を用いた推定方法だけでは的確に並列構造のスコープを捉えることができない。これらの問題に共通するのは、全て局所的な部分でしか制約がかけられていないことに起因するものである。従って、この問題を解決するためには、大域的な部分から制約をかけ、さらにスコープに対する推定を行わなければならない。

この大域的な部分から推定するものとして、黒橋らが提案する構造的類似性に着目した推定手法がよく知られている [7][8]。これは、係り受け解析する前に、三角行列上で文節間の類似度を求め、ダイナミックプログラミングの手法を用いて並列要素のバランスが整ったものを優先して範囲を推定するものである。黒橋らによると、82%以上の精度で推定可能と報告しているが、この手法を法

律文に適応した場合を考えると、あまり良い成果は望めない。その理由として、法律文の並列構造は並列要素のバランスが悪く、かつ、多重の埋め込み関係を持つものが多く存在しているため、あらかじめ表層的な部分からそれらを推定するのは困難であると考えられるからである。

以上より、法律文の並列構造の推定処理においては、局所・大域部分の両方を考慮した推定手法が必要であると言える。

### 3 法律文における並列構造の推定法

本節では、従来の手法の問題点を解決するために新しく考案した並列構造の推定法について述べる。前節で考察したように、法律文の並列構造を正しく推定するためには、局所的な観点から行う推定と大域的な観点から行う推定の両方が必要となる。

本手法は、ボトムアップで行う統語解析上に組み込むことを前提にして、要件部や効果部などの構成要素を大域レベル、そしてその中に含まれる文構造の構成素となる体言や用言などを局所レベルとして分け、その段階別に制約と評価方法を用いることで並列構造の推定を行う。

まず、局所レベルでは、間違っただけの推定結果に対しては、経験則に基づく制約で排除し、それ以外のものについては評価計算を行う。そして、大域レベルでは、呼応表現のように解析結果を全体的に見渡さなければ判定が難しいものを調べ判定する。

ここで評価する基準は並列構造内の要素間の結び付きの強さを表す度合(結合強度)と現時点での推定範囲の正しさを表す度合(範囲評価)の2つである。この2つの評価値は、並列構造を持つノードに割り当てるようにし、解析木が出来上がる度に、親ノードへ次々伝搬させていくようにする。

最後に解析が終了した時点でトップノードが持つ範囲評価がもっとも高い解析結果を最適解として選ぶ。

以下、この推定手法の基本手順である制約、範囲推定に対する評価計算、最終的な解の選出法について詳しく説明する。

#### 3.1 経験則に基づく制約

本節では、接続タイプ別における制約について局所レベル、大域レベルの順に示す。なお、部分並列に対しては、制約となる特徴が特に見られなかったので設けていない。

##### ○ 局所レベルの制約

- 要件表現の呼応に関する制約(体言並列のみ)  
要件表現で使われる名詞は主に条件的な事柄を述べるために専門的に使われる単語(法令用語)である。よって、そのような法令用語とその他の名詞とが並列されることはない。したがって、並列キーの自立語が要件表現の名詞「場合」「とき」の場合、後部の末尾の単語も必ず同じ「場合」「とき」でなければ後部の推定が間違っているものとする。

##### ● 格フレームに関する制約(用言並列のみ)

例えば「AがBし、Cする」のような用言句があるとする。もし、ここで「Aが」が「Cする」にとって必須格でなく、かつ、「Aが」が「Cする」に係るような依存構造として捉えている場合、後部にあたる「Cする」の格フレームを用いて「Aが」の格標識を基にそれが表す格関係とのマッチングを行なう。もしその格フレーム上の意味素性と「A」の意味素性がマッチしない場合は、間違っているもので、その構文構造を除去する。

##### ● 同一単語の並列に対する制約

法律文においては、「売主又は売主」のように同じ単語が並列されることはないので、単語同士の間並列のときに前部と後部の単語の基本形を調べ、もし同じであるならば、その構文構造を除去する。

##### ● 並列構造の埋め込み関係に対する制約

埋め込み関係を持つ2つの並列構造の間には、後で述べる結合強度に基づき、「結合強度の強い並列構造が結合強度の弱い並列構造を含むことができない」という関係を保たなければならない。これは、法律文の作成において慎重に接続詞の使い分けがなされていることに基づいている。結合強度の求め方については後述する。

##### ○ 大域レベルの制約

##### ● 呼応表現に関する制約

もし呼応表現の並列キーとなるものが現時点で捉えているスコープ内に含まれているならばそれと呼応する末尾の語が構成要素内に現れていないか確認する。もし、あれば現時点での推定範囲は間違っているもので、その結果を除去する。

##### ● 指示代名詞に基づく範囲制約

一般的に後部に「その」「それらの」といった指示代名詞が現れた場合、前部は構成要素の範

囲内で最長一致をとることが多い。そこで、指示代名詞が後部にある場合、前部が構成要素の範囲内で最長かどうか調べる。

### 3.2 評価計算

本節では、前節で述べた制約にかからなかったものに対して行う評価方法について述べる。スコop推定の評価には、条文構造に含まれる種々の表現や構文構造に対する特徴が複雑に関わっていることを考慮しなければならないので、ある種の条件を満たしていれば現時点での推定結果が正しい可能性が高いといった経験則を用いることにする。

この評価で求める評価値は、並列構造間の関係を決定するための基準となる結合強度と、推定範囲の正確さを表す範囲評価の2つとする。

以下に結合強度と範囲評価の求め方について説明する<sup>1</sup>。

#### 3.2.1 結合強度

人間が並列構造を捉える際に、「売主又は買主」のように前部と後部のバランスが取れており、かつ、その間に見られる類似性が高ければ高い程、それ自体をひとまとまりの構造として捉えることができる。一方、句や節同士の並列の中に見られるような前部と後部のバランスが悪かったり、類似性が低い場合は、その並列構造自体をまとまっているとして認識することはあまりない。

この並列構造に対する「まとまり」の認識度を並列要素の結び付きの強さに置き換えて考えてみると、前述した例では、前者は結び付きが強く、後者は結び付きが弱いと言える。ここで、この結び付きの度合を「結合強度」と呼ぶことにする。

結合強度を基準にして並列構造の埋め込み関係を見てみると、結合強度が強い(ひとまとまりとして捉えやすい)並列構造は結合強度の弱い(まとまっているとは捉えにくい)並列構造の中に埋め込まれていることが分かる。

本手法では、並列構造間の関係を決定するための基準として、この結合強度を導入する。この基準を用いて、同一文中に複数現れる並列構造間の関係を決定していくことにする。

この結合強度を判定する上で、以下の3点を主な判定基準とした。

1. 前部と後部に対する文節数の違い
2. 前部と後部の末尾の文節の類似度

<sup>1</sup>なお、これらの評価ポイントは分析対象から約30文選び、その中で正しく推定できるよう調整し設定した。

### 3. 並列キーの読点の有無

1は構造的なバランスに対する評価であり、2はその並列されている要素の表層的・深層的な面での類似性に対する評価である。また、3は読点により結合強度が弱まることを考慮した評価である。

以下に結合強度の求め方について述べる。

#### 1. 前部と後部に対するバランスの評価

本手法では、前部と後部のバランス性を評価するため、前部と後部にそれぞれ含まれる自立語の数の差を基に評価することにする。

具体的には、まず前部と後部の自立語の数をそれぞれ求める。そして、 $10 - (\text{前部と後部の自立語の差}) \times 2$ によりバランス性の評価を求める。もし、この結果がマイナスになる場合は0とする。この計算式は、対象文の並列構造を分析した結果から、人間が構造的に似ていると判断できる前部と後部の自立語の単語の差はおおよそ4~5個であるという仮定を考慮した上で立てている。

なお、並列構造が前部か後部に存在している場合、主にそれ自体がまとまった一つの単位として捉えられている場合が多いので、その場合は構造の長さを1に近似できるとの仮定の基に数える。しかし、並列構造のバランスが悪い場合、それを含む並列要素の構造全体の長さに影響を与えらることも考えられるので、ここではその前部と後部の差に1加えることで埋め込まれている並列構造のの長さを求めることにする。

#### 2. 前部と後部の末尾の文節の間の類似性評価

ここでは、単語の類似性の評価を黒橋らの計算方法を基に以下の手順で評価を行う。

- (1) 品詞が一致している場合、名詞なら2ポイント、動詞なら5ポイントを与える
- (2) 自立語が同じ単語である場合、15ポイントを与える。なお、この場合は、以下の文字の並びと概念の類似度の評価は行わない。
- (3) 自立語が名詞である場合に限り、自立語の文字列が部分的に一致する場合、一致した文字数×2(最大10ポイントで打ち切る)を加える。
- (4) 自立語が名詞である場合に限り、2つの意味関係を調べるためにシソーラス上での概念類似度を以下に述べる評価式を基に求める。文字列

の一致度と合計し、10ポイントで打ち切る<sup>2</sup>。

- (5) 並列要素の末尾に現れる助詞・助動詞の判定を部分並列と用言並列についてのみ行なう。もし一致すれば3ポイント加算する。

(1)でポイント設定を名詞と動詞で分けたのは、動詞に対して、(3)と(4)の評価ができないので、できる限り体言並列と用言並列との間に整合性を保たせることが望ましいという考えによるものである。

(3)の手順で概念の類似度を求めるときの評価式は、従来の研究で使っていた式と同じ

$$Sim(w_i, w_j) = \frac{d_c \times 2}{d_i + d_j} \times k$$

を使って、計算することにする。この計算式において、 $d_i, d_j$ がシソーラス上における二つの語の深さを、そして $d_c$ がそれらの共通の上位語(ノード)の深さを表している。 $k$ は、並列要素のバランス評価点との間につりあいを保たせるための係数であり、その値は10としている。

類似度計算をする上で基本となるシソーラスについては、法律固有(特に契約に関する条文全般)の概念を体系化していると思われる法的オントロジー(法的概念階層木)[1]を利用することにした。

(5)は、呼応表現のような表現に類似性を認めるために行なうポイント計算である。

### 3. 並列キーに含まれる読点の有無

読点は、意味的な区切りを表すために打たれる。よって、その読点が並列キーに含まれるということは、それだけ前部と後部の結合強度は読点がない場合よりも弱まっていると考えられる。そこで、もし並列キーに読点が含まれているならば、今まで計算してきた結合強度の値から、4ポイント減点する。

#### 3.2.2 範囲評価

現時点でのスコープ推定がどれくらい正しいのかの基準を解析結果に持たせるため、範囲評価を導入する。この評価基準は、結合強度で評価

<sup>2</sup>前部と後部のどちらかの要素全体が並列構造である場合、この類似度評価は適応できないため、(4)のみ評価するようにする。

する並列要素のバランスと類似性と経験則の3つである。並列要素における構造的バランスと類似性を評価基準として設けているのは、あまりにバランスが悪かったり、類似度が低い場合の範囲推定は誤りである可能性が高いと考えることができる。さらに経験則に従う評価を行うことにより、より詳細に範囲推定の評価が可能となる。

以下に範囲評価を求める際に行われる経験則の評価を示す。

#### (a) 構造バランスと類似度に基づく再評価

バランスの悪い並列構造を正しく推定させるために、以下の評価を付け加える。もし、バランスに対する評価値が6以下の場合でも、前部と後部の末尾の単語の類似度が8ポイント以上であるならば、それは並列されている可能性が高いのでボーナスとして(前部と後部の差)×2ポイント加算する。ただし、このボーナスポイントは上限10ポイントとする。

#### (b) 読点制約に基づく評価

体言並列のスコープ中には、係り受け制限としての読点を含むことはまずあり得ない。そこで、係り受け制限としての読点を末尾に持つ格要素が、スコープ内に含まれている場合、3ポイント減点する<sup>3</sup>。

#### (c) 接続詞に基づく並列構造間の関係評価

##### (i) 2つの並列構造の接続詞が異なる場合

構成要素内中に「又は」と「若しくは」をそれぞれ並列キーとする並列構造の間に埋め込み関係がない場合、スコープ推定が間違っていると考えられるので6ポイントペナルティを与える。この経験則は「並びに」と「及び」についても当てはまる。

##### (ii) 2つの並列構造の接続詞が同じ場合

同じ接続詞を並列キーとして持つ2つの並列構造が埋め込み関係にあるとき、結合強度に基づくものに付け加えて、それぞれの並列タイプが用言か体言か考慮することにより、より正確に推定することができる。並列タイプが同じ場合には、そのまま結合強度により判定する。並列タイプが異なる場合は、用言並列のほうが体言並列よりも結合強度が弱いという経験則を用いて判定する。もしその経験則に当てはまらない場合は、3ポイント減点する。しかし、体言並列のほうの並列キーに読点が含まれる場合、こ

<sup>3</sup>ただし、後部に指示代名詞がある場合は、前部を最長一致で捉えるため、この評価は適用しない。

の経験則は適用しない。

#### (d) 依存構造に関する評価

体言並列中において、後部に用言が修飾句となつて入り込んでいるとき、その体言並列にその用言句と同じようなものが係る構造はほとんど見られなかった。そこで、もし推定結果にこのような構造が見られた場合、6ポイントペナルティを与える。

### 3.3 最適解の選出

結合強度と範囲評価は、局所レベルの段階から順を追って評価されていく。構成要素中に並列構造が1つしかない場合は、局所レベルで計算した評価値を親ノードへ単に伝搬していくことで、その並列構造の評価値を構成要素内の推定評価として、さらには文全体における推定評価として表すことができる。

一方、複数の並列構造が構成要素内中に存在する場合は、親ノードに対する評価値が更新されることになる。

複数の並列構造が存在する場合には、2つのパターンが考えられる。1つは、子ノードの中だけに並列構造が存在する場合(図1(A))である。この場合、親ノードに対する結合強度はその子ノードの結合強度をそれぞれ合わせもつようにし、範囲評価はその子ノードの範囲評価を加算することで求める。

もう1つは、親ノードと子ノードの両方に並列構造が存在する場合(図1(B))である。この場合、以下の手順で親ノードの評価値を求める。まず、親ノードに対する結合強度と範囲評価を求める。そして子ノードに対する結合強度との比較を行なう。この時、違反した場合は、推定誤りとしてその構造自体を除去する。次に、親ノード自体に対する結合強度と子ノードの結合強度を両方合わせ持つようにし、それを親ノードのもつ結合強度とする。範囲評価については、現時点での親ノードの範囲評価と子ノードの範囲評価とを合計したものが親ノードに新しい範囲評価となる。

このように、解析木における親ノードと子ノードの間に対して評価値の伝搬を行い、最終的に法律文全体の解析が終了した時点の解析木のトップノードに割り当てられた範囲評価が最も高いものを並列構造の範囲推定が正しくできた最適解として選ぶ。

以上までの推定法を解析木に適用した例を図2に示す。図2の例では、「[売主及び買主]」の部分に修飾構造が入る場合と入らない場合とで範囲評

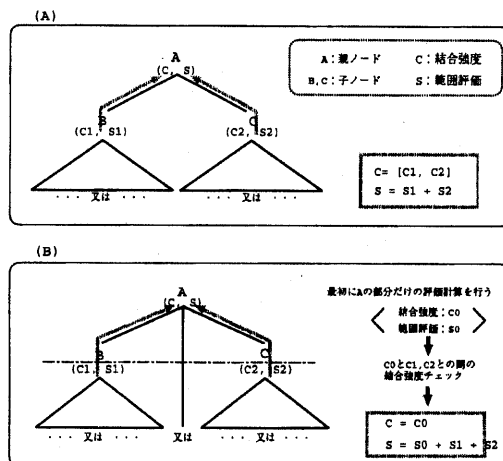


図1: 親ノードと子ノード間における評価値の伝搬

価が変わることを表している。この場合、修飾構造が入らない構造が正しいので、範囲評価により正しく並列構造が推定できていることが分かる。

## 4 実験と考察

本章では、本稿で提案した解析手法を従来のシステム[6]にインプリメントし、その実験とその結果、及びその結果に対する考察について述べる。

### 4.1 実験方法

本稿で提案した並列構造の推定処理に対する評価実験について以下に述べる。

実験に用いる対象文は、ウィーン統一売買法とウィーン条約から並列構造を含む文を無作為に抽出した184文とした。また、本稿で用いる辞書・文法の登録数は、辞書が1801個、文法が64個、格フレームが262個、シソーラス上の概念定義が293個である。

### 4.2 実験結果と考察

実験対象文184文の解析結果について、人手で評価を行った。ここでは、最適解に対する正解率を主に評価している。最適解として選ばれなくても全解析結果中に正しく範囲を推定している場合もあるが、ここでは、本手法が有効に機能していないとして推定間違いとした。その結果は、正しく範囲を推定した割合が全体の81%であり、そのうち、最適解が一意に出力された割合が38%、複数出力された場合が43%であった。このように複数

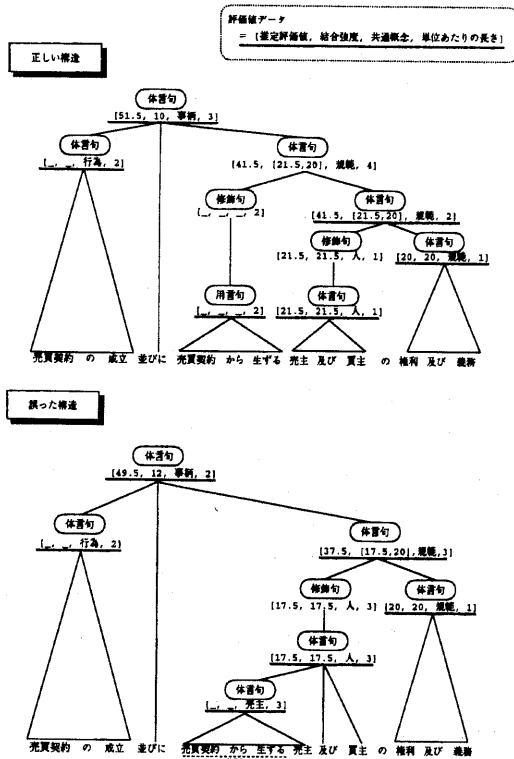


図 2: 正しい構造木と誤った構造木に対する評価値

出力されたもののほうがかなり多くなった原因としては、推定範囲に曖昧性が生じるといった直接的原因よりも、係り受けに関する曖昧性から影響をうける間接的原因によるものが多かったことが挙げられる。このことを考慮すると、本論文で述べた推定法は十分有効であると言える。

次に並列構造が埋め込み関係を持つ場合と持たない場合とに分類し、対象文毎にまとめた結果をそれぞれ表 1 と表 2 に示す。埋め込み関係を持たない場合の一意決定の 47% と比べ、埋め込み関係を持つ場合は 20% とかなり差が現れた。これは、埋め込み関係を持つ場合の方が、前に述べた通り、用言並列中に多くの修飾構造を含んでおり、その部分の依存構造に曖昧性が生じたためと考えられる。また、埋め込み関係を持つ方が推定誤りを多く起こしているのは、列挙型の並列が原因となっていることがほとんどであった。このような文は、人間でも専門知識がないとすぐには判断できないような難解な構造を持っていた。

表 1: 解析結果の評価 (埋込関係を持たない文)

単語数	ウィーン統一充買法			ウィーン集約			文数 (%)
	9~20	20~40	40~64	13~20	20~40	40~59	
一意に推定・正解	10	17	3	9	19	0	58(46)
複数存在(正解含む)	5	16	8	3	15	2	49(40)
推定誤り	1	4	6	0	4	2	17(14)
	16	37	17	12	38	4	124

表 2: 解析結果の評価 (埋込関係を持つ文)

単語数	ウィーン統一充買法			ウィーン集約			文数 (%)
	5~20	20~40	40~69	17~20	20~40	40~57	
一意に推定・正解	2	4	1	3	1	1	12(20)
複数存在(正解含む)	2	6	5	0	10	6	29(48)
推定誤り	0	1	4	0	6	8	19(32)
	4	11	10	3	17	15	60

以下に実験で推定誤りを起こしたのものの中で特に目立った問題について述べる。

1. 列挙的に並列されている要素の長さが極端に異なる文の誤推定

列挙的に並列されている個々の要素がほぼ同じ長さで並列されている場合は的確に推定することができる。しかし、長さが全く異なる場合、用言並列や体言並列でも表層的・深層的類似性が極端に低いため、最適解を求めることは極めて困難である。以下の例文では、「署名」から「加入」まで読点で区切られている要素全てが同じ並列構造の範囲内で列挙されている。しかし、本システムでは列挙要素をそれぞれ「代金」、「支払」、「物品の品質」、「数量、引渡の場所及び時期」として推定してしまった。この原因は、概念関係に関する類似性が大きく影響を与えていると思われる。すなわち、概念の類似度が高くなるように推定してしまったため、本来構造的に認識する列挙型の並列構造を誤って推定してしまったと考えられる。人間がこのような並列構造を認識する際には、読点を基に範囲を推定している。よって、このような列挙型の並列に対応できるように、今以上に読点に関する評価を行なっていくことが必要と考えられる。

【例】

・・・、特に代金、支払、物品の品質及び数量、引渡の場所及び時期、一方当事者の相手方に対する責任の限度、又は紛争の解決方法に関するものは、・・・

2. 用言並列と構成要素の範囲の判定間違い

以下に示す例文における並列構造には、用言

並列が存在しているが、このスコープ推定を行う場合、図3に示すように、「それが十分明確であり、」と対応する後部がどれか決める際に曖昧性が生じる。つまり、「承諾があった」、「拘束される」、「～申込者の意思が示されている」の3つの候補があるため、どれと対応しているか一意に決定するのは難しい。

このうち、もし後部が「承諾があった」として推定された場合、局所要件部内に読点を含む並列構造として捉えられるので、「～示されているときは」を要件部としてまとめる段階で並列構造の範囲の判定を行うことにより、読点の有無に関するペナルティから推定評価値を下げることができ、その結果推定誤りを起こさずに済む。

しかし、後部が「拘束されるとの」として捉えた場合、この読点の有無に関する制約は用いることができない。よって、「拘束される」と「～示されている」の2つの候補の中から選ぶ場合、他の手がかりが必要である。なお、本システムでは、「拘束される」までを後部と誤って推定してしまった。

この例文に限定すれば、動詞の動作性と状態性を辞書データに登録し、それを基に判定することで改善される。例えば、前部の「～である」は状態性を表す動詞であるから、それに対応する状態性の動詞は、「～ている」が末尾につく「～示されている」であるから、それを後部として捉えるように評価することにより、推定誤りは避けることができると考えられる。

次に、本手法全体に対して考察する。本手法は、並列要素の長さのバランス性が低い場合でも適切なポイントを与えることによって推定範囲の評価を高めるようにしている。この評価を高めるときに重要となるのが、シソーラスである。もしこのシソーラスの精度が悪いと、並列構造の範囲推定に対する評価をうまく修正することはできない。よって、本手法の推定処理の精度はシソーラスの善し悪しによって左右されてしまうと考えられる。

また、実験対象文には見られなかったが前部が長く後部が短いような体言並列はうまく推定することができない。これは、本手法が前部が短く後部が長い場合には、効果を発揮することに起因する。つまり、シソーラスデータを用いることで後部の終点は、並列キーに含まれる自立語を基に推定することはできるが、前部の始点は構造的なバランスが悪い場合簡単には推定できない。

評価ポイントの設定については、今回の実験により、ほぼ適当に表現されていると思われる。しかし、これらの設定はあくまで人間の主観により設定されたものであるため、今後、学習データから自動的に評価ポイントを獲得していくことが望ましい。

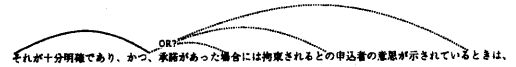


図3: 範囲推定における曖昧性

## 5 むすび

本稿では、従来の要件効果構造モデルに基づく解析手法の枠組に組み込む並列構造の推定手法を提案した。本手法を実際にシステム上に実装し、解析実験を行った結果、本手法が法律文の並列構造を推定する上で有効であることが分かった。

今後の課題としては、本手法で推定誤りを頻繁に起こしていた列挙型の並列や用言並列に対する推定処理の強化が挙げられる。

## 参考文献

- [1] 平成7年度科研究費重点領域研究「法律エキスパート」研究成果報告書、1996
- [2] 国際的動産売買契約に関する国連条約(ウィーン統一売買法)
- [3] 条約法に関するウィーン条約(ウィーン条約)
- [4] 田中、川添、成田、法律条文の標準構造、情報処理学会自然言語処理研究会報告、NL97-12、1993
- [5] 平松、永井、中村、野村、要件効果構造に基づく法律文制限言語モデルと法律文解析、情報処理学会自然言語処理研究会報告、NL115-4、1996
- [6] 平松、永井、中村、野村、要件効果構造に基づく要件効果構造に基づく法律文統語構造解析、情報処理学会自然言語処理研究会報告、NL118-7、1997
- [7] 黒橋、長尾、長い日本語文における並列構造推定、情報処理学会論文誌、Vol.33、No.8、pp1022-1031、1992
- [8] 黒橋、長尾、並列構造の検出に基づく長い日本語文の構文解析、情報処理学会自然言語処理研究会報告、NL88-1、1992