

## 統語森係り受け解析による最尤構文木の推定

釜谷 聡史 知野 哲朗 木村 和広  
(株)東芝 研究開発センター

〒212-8582 神奈川県川崎市幸区小向東芝町 1

{satoshi.kamatani,tetsuro.chino,kazu.kimura}@toshiba.co.jp

ロバストな文脈自由文法に基づく構文解析では、文法規則の大規模化に伴い、時に膨大な構文木が生成される。一方、構文木ベースの意味解析は、意味的な曖昧性を解消する有効な手段の一つである。しかし、構文木各々に対して意味解析を施し優先性を比較することは、時間コストが極めて高いという問題があった。そこで本稿では、複数の構文木を圧縮共有したデータ構造である圧縮共有統語森に対し、一括して構文木ベースの係り受け解析を行う手法を提案する。さらに、統語森上での係り受け解析結果に基づき、各構文木間に優先度を与え、効率的に最尤構造を選出する手法について述べる。本稿では、係り受けの尤度を評価する指標として共起確率を用いた。また、口語を対象とした実験で構文木数の 66.9%削減を実現し、本手法の有効性を確認した。

## Dependency Analysis on an Syntactic Forest and Estimation of the Optimum Parse Tree

KAMATANI Satoshi, CHINO Tetsuro and KIMURA Kazuhiro

TOSHIBA Corporation, Corporate Research and Development Center  
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki, 212-8582, Japan,  
{satoshi.kamatani,tetsuro.chino,kazu.kimura}@toshiba.co.jp

The syntax analysis, based on robust context-free grammar, often generates huge number of syntax trees. On the other hand, it is well-known that the semantic analysis on syntactic structures is effective for semantic disambiguation. But, to analyze whole syntactic trees, we had to pay so expensive costs. In this paper, we propose the efficient dependency analysis method on the syntactic forest, which packs and shares all parse trees for one sentence. With this method and co-occurrence probability, we can assign priorities between each tree in the forest. After proposition, we will also report about experimental result that achieves 66.9% reductions of parse trees.

# 1. はじめに

音声対話システムなど、話し言葉を扱うシステムでは、種々の発話を受理する頑健な解析器が必要となる。自然言語の解析において用いられる文脈自由文法(Context Free Grammar)は、曖昧性に対して頑健に解析するために、その文法規模を拡大すると、時に、極めて膨大な数の構文解析結果を生成することが知られている。

一方、構文解析の結果である構文木に従って意味解析を行い、意味的に尤もらしい構造を選択する手法は、曖昧性を解決する有効な手段の1つである。しかし、構文木に基づく、文法規則主導の意味解析(Syntax Driven Semantic Analysis)は、基本的に一つの正しい構文解析結果に対して適用することを想定している[1]。よって、複数の構文木に対する意味解析をする場合、個々に処理を施す必要があるため非常に効率が悪く、対象の数に応じて時間コストが爆発的に増大する。このため、可能な解釈である全ての構文木の中から、最良の解釈を効率的に選び出すメカニズムが必要になる。

そこで本稿では、一つの文に対する全ての構文木を圧縮共有して保持するデータ構造である圧縮共有統語森(Packed-Shared Syntax Forest. 以下、統語森と呼ぶ)[2]に対して、係り受け解析を一括して施す手法を提案する。これにより、内包する全ての構文木に対して係り受け解析を施すことができる。

更に本稿では、提案手法と共起確率を用いた、

最尤構文木推定実験の結果を報告する。

# 2. 係り受け解析のための統語森構造

## 2.1. 従来の統語森

統語森は、複数の構文木が共通して持つ部分木を共有(部分木共有)し、以下の条件を満たす部分木の根を併合(局所曖昧性圧縮)することにより、効率的に構文木の集合を保持できるデータ構造である。

- a) 句の非終端記号が同じ
- b) 支配する入力文の範囲が同じ

図1に示した4つの構文木を、圧縮共有した統語森を図2に示す。図2において、中間節点のラベルは、その句の非終端記号を示す。図2では、局所曖昧性圧縮の様子を分かり易くするため、節点Xに、x1、x2と補助的な名前を付記して示しているが、実際にはどちらも同じ非終端記号Xの節点である。また、節点Lでは、節点y1を含む構造と節点s2を含む構造との間で、節点の共有がなされている。以下説明のために、部分木共有をした節点を共有節点、局所曖昧性圧縮をした節点を圧縮節点、節点と節点を結ぶ弧をアークと呼ぶことにする。

図1、図2において、太線で示したアークは、その根側の節点の支配する入力範囲における主辞を表している。各部分木における主辞の伝播は、文法規則に記述した主辞の継承関係を基に規定する。このように、構文木に主辞の概念を導入することで、構文木上の係り受け関係は、各節点間を結ぶ主辞ではないアーク上の関係と

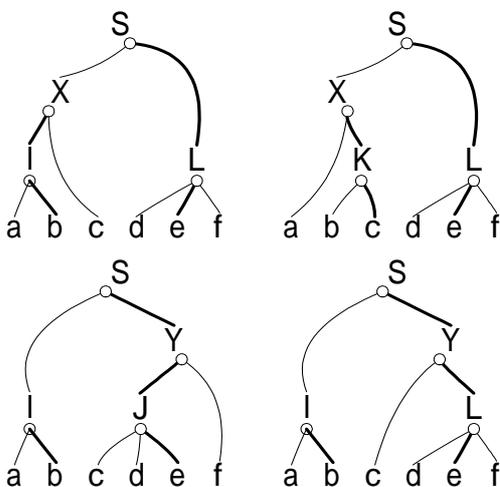


図1 可能な構文木の集合

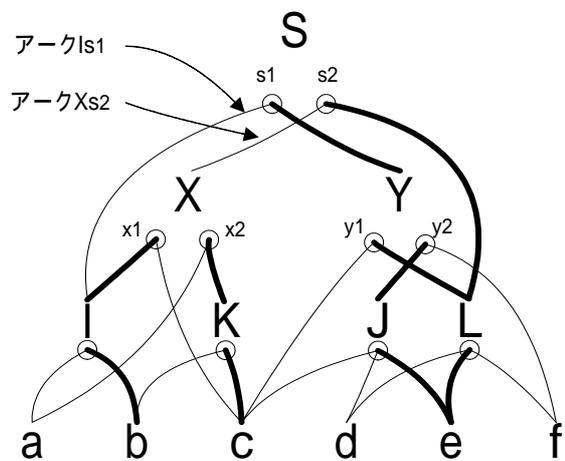


図2 図1の構文木に対する統語森

して捉えられる。例えば、図2のアーキ  $1s_1$  は、入力単語  $b$  を受け側主辞、入力単語  $e$  に係り側主辞とする係り受け関係と対応付けられる。

ところが、図2のアーキ  $Xs_2$  では、入力単語  $b$ 、入力単語  $c$  の2つが、係り側主辞となりえる。つまり、1つのアーキを複数の係り受け可能性が共有していることになる。結果、このアーキを選んだときの係り受け関係を、一意に決めることができない。このことが、統語森における係り受け解析を困難にする要因であった。

## 2.2. 主辞構造付き統語森

前節で述べたような、係り受け関係が一意に決まらないアーキは、圧縮節点と他の節点間に作られる。従来の局所曖昧性圧縮の条件では、その注目節点が支配する入力範囲における主辞が異なっても、節点を持つ非終端記号が等しければ、文法的に振る舞いが同じであると判断し、併合していたことに起因する問題である。そこで、局所曖昧性圧縮の条件に次項を加える。

c) 句の主辞となる主構成素が同じ

この追加条件により、図1の構文木は、図3のように圧縮共有される。図3においても、便宜上、節点  $X_1$ 、 $X_2$  と名前を変えて示しているが、実際には、支配する入力範囲が同じであり、かつ、非終端記号が同じ  $X$  である節点を示している。ここで、図2と異なり、節点  $X_1$ 、 $X_2$  が圧縮されていないことに注意されたい。第3の条件を追加することにより、従来の統語森構造に比べ、圧縮率はやや悪化するが、係り受け関係が一意に決まる構造となる。また、条件追加後の統語森が包含する構文木は等価である。

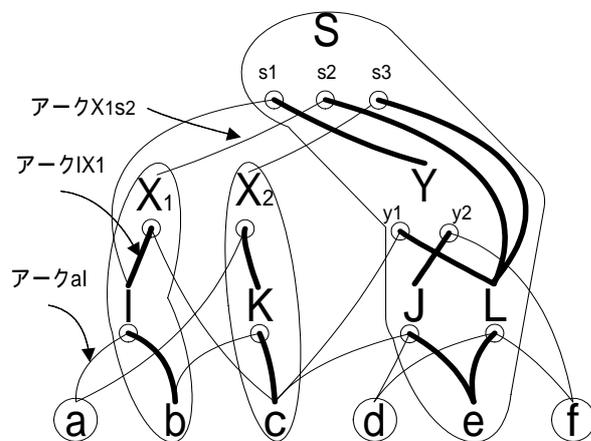


図3 主辞構造付き統語森

このような構造を作ることによって、図3に示すように、各入力単語が主辞として支配する統語森の領域を、互いに重なることなく切り分けることが可能となる。以下、この各入力単語の支配領域を、主辞の“島”と呼ぶことにする。この島付きの構造を捉えることで、統語森上の係り受け関係を、主辞が異なる2つの島間の関係として扱うことができる。言い換えれば、島と島とを結ぶアーキに対応付けることができる。例えば、アーキ  $al$  は、入力単語  $a$  を係り側主辞、入力単語  $b$  を受け側主辞とした係り受け関係に対応付けられる。

## 3. 統語森係り受け解析

### 3.1. 直接係り受けと間接係り受け

前節で述べたような、2つの島が直接結ばれ、一方の島の主辞が他方の島の主辞単語に係る関係を、直接係り受けと呼ぶことにする。例えば、図3におけるアーキ  $X1s2$  に対応する係り受けは、子節点側の主辞  $b$  と親節点側の主辞  $e$  との間の直接係り受けであるという。また同時に、アーキ  $X1s2$  は、アーキ  $IX1, al$  を経由して、主辞  $a$  が主辞  $e$  に係っている関係をも、併せて表現していると見ることができる。これを、間接係り受けと呼ぶことにする。本研究における係り受け解析では、一般的な直接係り受けのみではなく、この間接係り受けも考慮に入れて解析を進め、最尤構造を推定する指標とする。

間接係り受けの考え方を導入することで、ある部分構造が、他の部分構造、あるいは単語に係るような構造を、双方の間に関係の存在しないような構造よりも優先させることができる。例えば図4で、節点  $S_2$  を含む構造を見ると、「私は」が「行く」に直接係っている。一方、「私が学校に行く途中に」という節全体として、「会った」に係っていると見ることもできる。すなわ

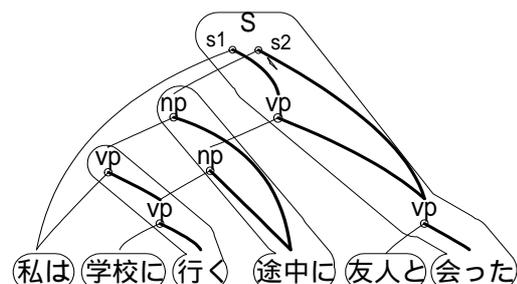


図4 間接係り受け関係

ち、「私が」が間接的に「会った」に係っていると見ることが可能である。一方、節点 S1 を取ると、「私は」は、「会う」は直接係るが、「行く」に係らないと見ることができる。

上記のような、間接係り受け関係の導入により、表層的に遠い係り受け関係よりも、近い係り受け関係が優先され易くなる。統計的に、単語に複数の係りが想定される場合、その 98.6% が、1 番目、2 番目、及び、最末尾の受け側候補に係るという統計が報告されており、特に 1 番目の受け側候補に限っても 7 割前後で係り受けが起こることが報告されている[3]。これに対して、間接係り受けに基づくスコアの差が、表層的に近い係り受け関係より表層的に遠い係り受けを優先させる際の尤度の差として有効に働くと考えられる。例えば、図 4 の例では、「私が「学校」へ「行」き、「私が」が「友人」に「会った」、節点 S2 を含む構造が優先される。

### 3.2. 係り受けスコア

統語森において、ある節点を選ぶと、これを親節点として接続するアークが決まり、更に子節点との間に発生する係り受け関係を決めることができる。よって、係り受けが発生するアークの親節点側に、その節点の係り受け尤度を表すスコアを付与する。

本研究で用いる係り受けスコアは、共起確率を基に計算する。共起関係は、前節で述べた直接係り受けの関係に対応するため、間接係り受けのスコアを、直接計算することはできない。そこで、共起確率から計算される直接係り受けのスコアを  $S_d$  として、次式で近似する。

$$S = \frac{S_d}{(\alpha + \beta + \gamma)^2} \quad (1)$$

$S_d$  : 直接係り受けに対するスコア

: 係り受け間の島の数

: 係り受け間の節数

: 間接係り受け補正項

$S_d$  の詳細な計算手法は、4 節において述べ、ここでの詳細な説明は省略する。  $\alpha$  は、ある係り側単語が、受け側単語に係るまでに経由した島の数である。これにより、間接係り受けであっても、構造的な遠さによりその値に差をつけることができる。  $\beta$  は、同じ節(cclause)内での

係り受けに対して、節を越えた係り受けにペナルティを与えるためのパラメータである。  $\gamma$  は、間接係り受けのスコアを、直接係り受けのスコアよりも、必ず低くなるように補正する項である。本稿では、解析対象の統語森が持つ島の数である、  $\beta$  の最大値とした。

係り受けスコアは、各入力単語に注目した場合に、直接/間接係り受けが発生する節点、全てに対して評価し、足し込んでいく。ただし、ある節点において、全く同一の受け側と係り側単語の組み合わせで係り受けが発生した場合、(1)式で評価した最大値を足し込む。

### 3.3. スコアの集計と最尤構造推定

今、前節で与えた式(1)により、図 3 の統語森に、図 5 に示すように係り受けスコアが付与されたとする。これは、各節点、特に曖昧節点において、局所的にどの節点が尤もらしいかを示した指標となる。例えば、図 3 では、圧縮節点 Y で圧縮された節点 y1 に 0 点、y2 に 2 点が付与されていることから、y2 の方がより尤もらしいことを表している。

ついで、葉節点から根節点方向へ向かって、この係り受けスコア得点を合算していけば、各節点を選んだとき、その節点以下が支配する部分木の尤もらしさを得ることができる[4]。この時、各節点のスコアは、圧縮された子節点に付与された得点と自身に付与されたスコアとの和で更新する。ただし、圧縮節点では、圧縮されている全ての子節点を持つスコアの最大値と自身のスコアとの和で更新する。例えば、図 5 の

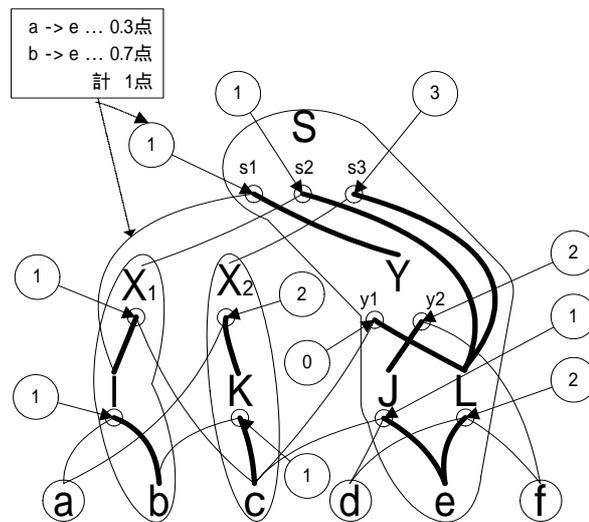


図 5 係り受けスコアの付与

節点  $X_i$  のスコアは、自身のスコア 1 点と、子節点  $l$  のスコア 1 点とを足し合わせた 2 点となる。また、図 5 の節点  $S1$  のスコアは、自身のスコア 1 点と、節点  $l$  のスコア 1 点、節点  $y_1, y_2$  の内、最大のスコア 3 点とを足し合わせた 5 点となる。図 5 の統語森に対して、スコアを集計した結果を図 6 に示す。

上記処理により、各節点には、その節点を選択した場合に得られる、最大の係り受けスコアが付与される。よって、根節点から葉節点方向へ、最大スコアを持つ節点を選び出していくことで、最尤構造を選び出すことが可能である。以降では、以上の処理を意味選好処理と呼ぶ。

意味選好処理後の構造は、統語森内でスコアが同点のものが一つもなければ単一の構文木となり、同点のものがあれば、入力された統語森から、解決できた曖昧性を除去した部分統語森となる。後者の場合、係り受けで除去できない曖昧性が構文木に含まれていた、あるいは、共起情報が不足していたなどの問題で、曖昧性が完全には除去できなかったことを示している。

## 4. 共起に基づく係り受けスコア

### 4.1. 共起確率

本研究では、鳥澤が提案した手法[5]に基づいて学習したモデルを用いて、共起確率の推定をする。鳥澤らの手法では、まず、動詞  $v$  と名詞  $n$  の格関係  $rel$  における共起確率  $p(v, n, rel)$  を、その生起が隠れた意味カテゴリのクラス  $c \in C$  ( $C$  は想定される全クラスの集合) によってコントロールされていると仮定する。次いで、ク

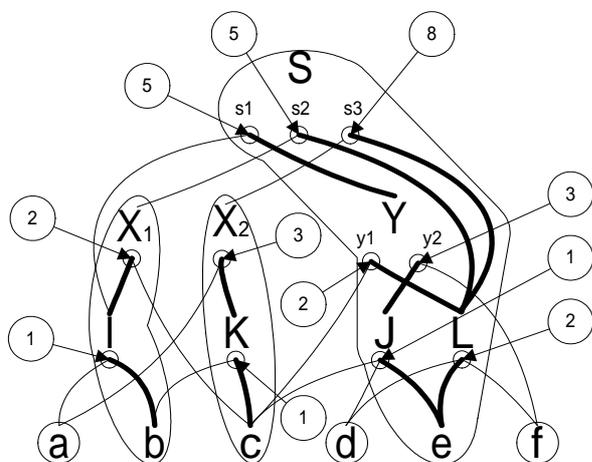


図 6 係り受けスコアの集計結果

ラス数を指定し、EM アルゴリズムを用いて、単語の共起、および、語のクラス分けを学習する。

本稿では、学習パターンに形容詞・形容動詞と名詞(連体修飾時には  $rel$  を「が」格として学習)及び複合名詞( $rel$  を“empty”として学習)を加え、これらの間の係り受け尤度も評価できるように拡張した。このため、本稿では、 $m$  を係り側単語、 $h$  を受け側単語として、その共起確率を次式でモデル化する。

$$p(<h, m, rel >) = \sum_{c \in C} p(<h, rel > | c) p(m | c) p(c) \quad (2)$$

学習には、毎日新聞 7 年分の記事、約 700 万文から延べ約 2,800 万共起を抽出し、内、共起頻度が一定の閾値を越えた約 1,900 万共起を学習データとして利用した。このとき、受け側単語は 110,973 種類、係り側単語は 132,316 種類、共起の組み合わせは 3,544,989 種類であった。また、学習するクラスの個数は 100 とした。

学習によって得られたモデルにおける、確率値  $p(<h, m, rel >)$  の最大値、最小値、平均値、分散を表 1 に示す。

### 4.2. 係り受けスコアの計算手法

ATR-バイリンガル旅行会話コーパス(SLDB)[6]から、抜き出した発話の内、次節の実験で用いるテスト用例文 100 文を除いた 21,642 文が含む共起に関する、前節で述べたモデルにおける共起確率の分布を調査した。抜き出した例文 21,642 文が含む共起の異なり数は 8,546 であった。この内、6,547 共起について、共起確率を計算することができた。この時得られた確率値の分布を図 7 に示す。

図 7 によれば、このモデルによって与えられる共起確率は、直接係り受けの尤度や優先性を評価するには各値のばらつきが大きい。そこで、本稿の実験では、確率を、その指数部に基づいて、0~1.0 点にマップして利用する。まず、次式により確率の指数部  $E_p$  を得る。ここで  $f(x)$

表 1 共起確率値の分布

最大値	$4.05 \times 10^{-4}$
最小値	0.0
平均値	$5.31 \times 10^{-12}$
標準偏差	$4.71 \times 10^{-9}$

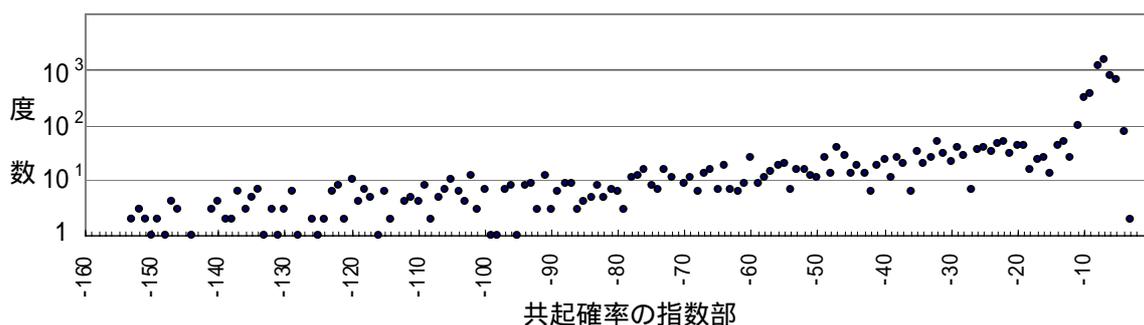


図7 共起確率の指数部ごとの度数

は  $x$  の整数部を取り出す関数とする。

$$E_p(<h, m, rel >) = f(-\log_{10}(p(<h, m, rel >))) \quad (3)$$

次いで、直接係り受けスコア  $S_d$  を次式により計算する。ここで、 $m$  は図7に示された確率の指数部の平均値とし、 $m=18.1$  である。また、 $k$  はスコアの減衰率を規定するパラメータである。

$$S_d(<h, m, rel >) = \frac{1}{1 + \exp(k \times (E_p(<h, m, rel >) - m))} \quad (4)$$

ただし、(4)で計算されるスコア  $S_d$  の値が閾値 EPS を下回るときは、EPS をスコアとする。また、前節で述べたように、共起モデルは直接係り受けに関するモデルであるため、間接係り受けに対するスコアを得ることはできない。そこで、格関係  $rel$  を無視して検索した、 $E_p(<h, m, rel >)$  の最大値をもって、スコア  $S_d$  を近似する。

$$S_d'(<h, m >) = \max_{rel} (E_p(<h, m, rel >)) \quad (5)$$

また、(4)式で  $S_d$  が計算できなかった場合も、格関係  $rel$  を無視して、間接係り受けと同列に扱って計算することとし、(4)式で計算したスコアを、(1)式で補正したものを与える。

## 5. 実験

### 5.1. 実験方法

ATR-バイリンガル旅行会話コーパス (SLDB) [6] から、100 文を抜き出して評価用例文とした。抜き出した例文は、平均文長 24.0 文字、平均単語数 14.9 単語であった。

構文解析には、話し言葉向けに独自に作成し

た頑健な文脈自由文法に用いた。また、解析器には、本文法による解析を可能とするように拡張した、一般化 LR 解析法 (Generalized LR Parsing) [2] ベースの解析手法を用いた。また、本文法は、構文的に適格性の低い部分木を構築する文法規則を、他の一般規則と区別することができ、これに基づいて統語的に優先される構文木を選択することが可能である。以下、この処理を統語選好処理と呼び、意味選好処理に先んじて施す。ついで、統語選好後の統語森に対して、本稿提案の意味選好処理を施す。

係り受けスコアの計算には、4 節で述べた共起モデルを用いた。(4)式で用いる各パラメータは、 $k=0.2$ 、 $EPS=10^{-20}$  とした。ただし、名詞句「AのB」や連語構造などに見られる、名詞-名詞間の共起は、その組み合わせが膨大に考えられ、また構文的にも種々の係り受け組み合わせが可能である。このため、名詞-動詞間の係り受けに比べてスコアが低く計算されやすい傾向がある。そこで、本実験においては、名詞-名詞間の係り受け曖昧性解決せず、統語森からの選好処理では、曖昧な構造をそのままの形で含む出力をするようにした。

意味的に尤もらしい構造が選好されていることを評価する基準として、人手で係り受けの正解を定めた。これに基づき、意味選好処理後の統語森が内包する各構文木が、正解係り受けをいくつを内包していたかを評価し、これを内包率と呼ぶことにする。内包率は、統語選好後の統語森  $F$  が含む構文木を  $t$  として、次式で与える。

$$\text{内包率} = \frac{\sum_{t \in F} \text{正解係り受け数}}{\text{統語森が内包する構文木数}} \quad (5)$$

## 5.2. 実験結果

### 5.2.1. 評価対象の選別

実験に用いたテスト用例文 100 文は、形態素解析の結果、および、構文解析の結果である統語森が内包する係り受け曖昧性にに基づき、表 2 のように分類できる。

本実験に用いた解析システムや文法は、テスト用例文向けに調整されたものではないために、形態素 / 構文解析で誤ったものが 4 文あった。加えて、受理時および統語選好時において、正解である係り受け関係を一つも含まないものが 3 文あった。これら 7 文は、提案手法の枠組み外での誤りであるため、以降の評価では対象としないこととした。

提案手法は、構文木が表す係り受けの尤度を比較することで、最尤構造を推定するものであるため、係り受け関係に曖昧性のない構文木間に優先順位を付けることはできない。感動詞のみの文などの係り受け関係が存在しないものや、統語選好処理までに構文木が一つに絞り込まれたものがこれにあたり、93 文中 20 文あった。本実験の目的は、曖昧性をもつ統語森に対して、意味的な優先順を付けることであるから、これらの文も以降の評価では対象としないこととし、残りの 73 文について論じる。

### 5.2.2. 最尤構造推定の評価

まず、提案手法の計算効率について評価する。意味選好処理時になされた単語間係り受け関係の評価回数は、一文あたり約 8.3 回であった。これに対して、統語森が内包する全ての構文木、個々に対して評価すると、一文あたり約 244.5 回の計算が必要である。構文木全体を評価しながら、大きく計算量を減らし、効率的に推定処理がなされていることが確認できた。

表 2 解析結果の分類

形態素解析失敗	2
構文解析失敗	2
受理時正解係り受け含まず	2
統語選好処理後正解係り受けなし	1
係り受けなし	8
統語選好処理後構文木数 = 1	12
統語選好処理後構文木数 > 1	73
計	100

次に、提案手法による意味的曖昧性解消の有効性について評価する。受理時、統語選好処理後、意味選好処理後の各段階における、統語森が内包する構文木数の変化を表 3 に示す。提案手法により、統語選好処理後に比べて、構文木数を平均で約 23.9% に削減している。

最尤構造として選出した構造が、確かに係り受け的に尤もらしい構造であることを確認する。評価に用いた 73 文中、意味選好処理により構文木数を削減できた文は 53 文である。これらの内、内包率が変化しなかった、あるいは、向上した文は 50 文であった。

内包率が悪化した 3 文は、名詞句の曖昧性解決をしなかったため、正解係り受けを含む構造が絞り込まれても誤った構造が多く残り、結果として内包率が悪化する傾向にあった。これは、係り受けスコアを計算する共起確率のモデルを改善し、名詞間の係り受け曖昧性を解消することで解決できる可能性がある。

統語選好処理後、意味選好処理後の正解係り受け内包率の変化を表 4 に、その分布の様子を図 8 に示す。図 8 では、矩形中央の横線が内包率の平均値を、縦線が内包率の分布範囲を、矩形自身が平均値から 1 の範囲を示している。まず、表 4 より、意味選好処理の結果、内包率が上昇していることを確認できる。

以上より、意味的な質を落とすことなく構文木数を削減できたことが確認できる。

## 6. 考察と今後の課題

本稿では、共起確率を基とした、係り受けスコアを用いて、係り受けに関する最尤構造を推

表 3 各選好段階における構文木数

	平均構文木数
受理時	10349.96
統語選好処理後	51.74
意味選好処理後	10.74

表 4 各選好段階における内包率

	内包率
受理時	0.411
統語選好処理後	0.574
意味選好処理後	0.673

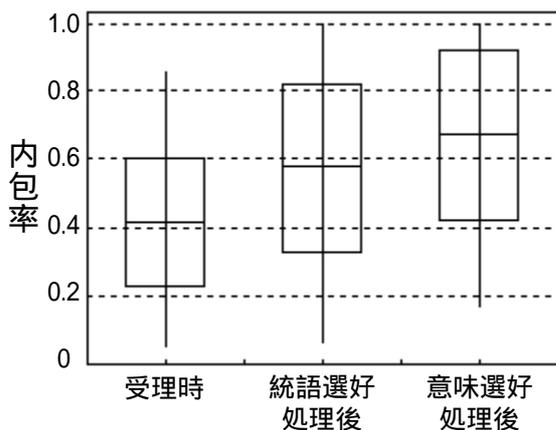


図8 各選好フェイズにおける内包率の分布

定した。共起情報は、係り側と受け側に着目したときの共起の強度であり、局所的な尤もらしさを提供する。これに対して、文全体として、尤もらしい係り受けの組み合わせを選ぶ問題では、複数の係り側・受け側の組み合わせの中から最尤の組み合わせを選び出す、大域的な尤もらしさを比較するべきである。そのためには、複数の係り受け候補間で、直接その尤度を比較可能なモデルを検討することが必要である。

5節で述べた実験では、名詞 - 名詞間の係り受け曖昧性に関して、解決をしなかったために、意味選好処理後の構文木数が十分削減しきれていない。また、受身などによる格の交代、一文一格などの経験的に得られている言語知識が、係り受けの評価に反映されていない。これらの知識を柔軟に組み入れると共に、構文木の部分構造を捉えて、評価方法を変えるような枠組みも重要になると考えている。

本手法は、統語森の形式で入出力が可能であり、かつ、その他の構造をひつよとしないため、構文的・意味的尤度を評価し最尤構造を推定する種々の手法と組み合わせで適用できる利点がある。例えば、確率 GLR 解析[7]を併用し、構文的にも意味的にも尤もらしい構造を推定するなどの方法も考えられる。

## 7. おわりに

本稿では、入力に対する可能な全ての解釈(構文木)の集合を圧縮共有した構造である統語森を対象とし、一括して係り受け解析を施す手法を提案した。

また、口語文を対象とした実験では、提案手法に基づく意味選好処理によって、統語選好処理後に比べ、構文木数を約 66.1%削減し、かつ、係り受け解析処理を構文木個々に対して施す場合に比べて、計算量を約 96.3%削減できるという結果を得た。これにより、提案手法が、全構文解析結果を対象にしながら、係り受け的に尤もらしい構造の効率的な選好処理を可能としたことを確認した。

## 参考文献

- [1] 八木豊, 野呂智哉, 橋本泰一, 徳永健伸, 田中穂積, “単語の共起情報を利用した文法主導の係り受け解析”, 情報処理学会研究報告 SIG-72, pp.1-8, (2003).
- [2] M.Tomomita, “Generalized LR Parsing”, Kluwer Academic Publishers, Norwell, Massachusetts, (1991).
- [3] KANAYAMA Hiroshi, TORISAWA Kentaro, MITSUISHI Yutaka, and TSUJII Jun ichi, “Statistical Dependency Analysis with an HPSG-based Japanese Grammar”, *Proc. of NLPRS*, pp. 138-143, (1999).
- [4] 田村直之, M.J.W.Bos, 吉見毅彦, 西田収, J.Jelinek, 村上温夫, “チャート構造上での意味処理メカニズム”, 情報処理学会研究報告, NL-81, pp.25-32, (1991).
- [5] Torisawa, K., “An supervised Method for Canonicalization of Japanese Postpositions”, *Proc. of the 6th Natural Language Processing Pacific Rim Symposium*, pp.211-218, (2001).
- [6] 竹沢寿幸, 白井諭, 大山芳史, “バイリンガル旅行会話コーパスに見られる話し言葉の特徴分析”, 情報処理学会研究報告, NL-141, pp.137-144, (2001).
- [7] K.Inui, V.Sornlertlamvanich, H.Tanaka, and T.Tokunaga, “Probabilistic GLR parsing”, *Proc. of the 5th International Workshop on Parsing Technologies*, pp.123-134, (1997).
- [8] 平川秀樹, “最適解探索に基づく日本語意味係り受け解析”, 情報処理学会論文誌, Vol.43, No.03 - 002, (2002).