

## 単語の共起情報を利用した文法主導の係り受け解析

八木豊 野呂智哉 橋本泰一 徳永健伸 田中穂積  
東京工業大学大学院 情報理工学研究科 計算工学専攻  
{yutaka, noro, taiichi, take, tanaka}@cl.cs.titech.ac.jp

与えられた構文木に沿って解析を進める Syntax-Directed または Syntax-Driven な文法主導の解析手法は、広範囲の文を対象としたときに実用上の問題があるとこれまで言われてきた。そのため、近年では統計的な自然言語処理の手法が多く利用されている。しかし、純粋に統計的な自然言語処理を用いた手法には飽和現象が見え始めたというのも事実である。そこで、もう一度、Syntax-Directed な手法の良さを見直し、それを自然言語処理に応用すべき時期にきていると我々は考えている。本論文では、我々が現在開発中である CFG ベースの大規模な日本語文法をもとに Syntax-Directed な手法で日本語の解析を行い、本手法の有効性を確認するための前実験の結果について報告する。

### Syntax-Directed Dependency Analysis using very Large Japanese Grammar and Lexical Co-Occurrence Relation

Yutaka Yagi, Tomoya Noro, Taiichi Hashimoto, Takenobu Tokunaga, Hozumi Tanaka  
Department of Computer Science, Graduate School of Information Science and Engineering,  
Tokyo Institute of Technology  
{yutaka, noro, taiichi, take, tanaka}@cl.cs.titech.ac.jp

Syntax-Directed analysis or Syntax-Driven analysis, method of analysis following a given parse tree, had problems in practice when parsing general sentences. Therefore, recently, many methods based on statistical information were proposed in natural language processing to improve the analysis. However it is also a fact that the improvement by these methods has been saturated. Nonetheless, we believe that the merit of Syntax-Directed analysis can be increased, and as such it should be applied to natural language processing This paper proposes Syntax-Directed analysis using large-scale Japanese context-free grammar and reports preliminary results.

## 1 はじめに

人工言語のコンパイラの世界では、早くから Syntax-Directed または Syntax-Driven Semantic Analysis[7] の方法が実用化され、コンパイラコンパイラの一つである Yacc などではオブジェクトコード生成に使われていた [1]。この方法によれば、与えられた文の構文木に沿って、その先の意味解析などを進めることが可能である。構文木は、部分から全体に向けて（あるいはその逆に）木が再帰的に構成されているため、Principle of Compositionality（構成原理）あるいは Fregean Principle（フレーゲの原理）に基づく意味解析を行なうのに都合の良い構造をしている。部分から全体の意味を計算するために、下位の木の節点に計算された意味は上位の木の節点に送られる。その計算が再帰的に（上位に向かって）繰り返されることで、文全体の意味が根の節点に計算される。Knuth の提案した属性文法（attribute grammar）は、こうした計算方法をさらに一般化したものである。木の各節点の計算結果は、原理的にはそれより上位（全体）の節点にも下位（部分）の節点にも計算結果を送ることができる。人工言語の世界では、LR 文法など文法に制限を付与し、構文解析結果が一意に決まるように、文法に制限をつけている。したがって、構文解析結果に曖昧性が生じることはないことに注意したい。

自然言語の世界でも、Syntax-Driven Semantic Analysis（SDSA）の方法が幾つか採用されている。Syntax-Driven とは文法主導で規則主体（rule base）の手法であることに注意したい。例えば典型的な Montague 文法では、文の構造（分析木）に与える範疇文法のカテゴリーに沿って  $\lambda$  変換を繰り返すことで、部分から全体の意味をシステムティックに計算する意味解釈の方法が導入されている [4]。最近では、言語理論にもこの手法が採り入れられている。GPSG[5]、LFG[3]、HPSG [11, 13] などのユニフィケーションベースの言語理論では、句構造にしたがって、部分の素性構造を次々にユニファイして文全体の素性構造を計算する。計算を意識した言語理論の構築を目指しているのである。Alshawi らの Core Language Engine [2] でベースとした DCG は [10]、論理プログラムの枠組にこの考え方を組み込んだものである。いずれの言語理論にしても、実用に耐える大規模な文法の開発それ自体は今

後の課題であり、それらの自然言語処理への本格的な応用は、未だ研究段階にある。

上記した SDSA では、正しい構文解析結果に、この解析手法を適用することのみが想定されている。言語理論の骨子であり、高速なアルゴリズムが開発されている CFG を用いた構文解析（syntactic parsing）では、CFG の規模が大規模になるにつれて、一つの文から、極めて多数の、時として天文学的な数の構文解析結果が得られることが稀ではない。天文学的な数の構文解析結果のおおのほに SDSA を施すことには、時間の問題がある。たとえそれが可能であったとしても、多数の SDSA の結果の中の、どれが正しい結果であるかを判定しなければならないという別の問題もある。これらもまた SDSA の適用を妨げる大きな要因の一つであった。

一方、骨子となる CFG 規則の数を極めて少数に抑え、構文解析結果の数を制限しようとする文法理論が HPSG である。その代償として LFG 同様、語彙項目記述を豊富にする必要があり、項目数の大規模化とともに、その記述内容についても研究段階にあるなど、幾つかの問題を抱えることになった。さらに、多種多様な文の解析が可能で（文法でカバー可能な文の範囲が広い）、SDSA を前提にした大規模 CFG の開発が遅れていることも問題である。

筆者の知る限り、SDSA の計算機構を備え、得られた構文解析結果に何らかの方法でスコア付けすることが可能なパーザとして、Pratt が開発した LINGOL システムがある [12, 19]。しかしスコア付けの方法はパーザの使用者に任せられおり、明確な基準は示されていない。

以上に述べた様々な理由により、SDSA の方法による規則ベースの自然言語処理の手法は、広範囲の文を対象とした時、実用上問題があるとして、多くの研究者から見放されてきたように思える。1990 年代の自然言語処理の研究者が、統計的な自然言語処理に新たな活路を見い出そうとした理由の一つが、ここにあると我々は考えている。実際、これまで統計的な手法は、旧来の規則ベースの手法と比較して、華々しい成果をあげてきた。しかし、純粹に統計的な手法による自然言語処理に飽和現象が見え始めたというのも事実である。本論文の目的は、困難ではあっても、もう一度、CFG ベースの SDSA の手法の良さを見直し、それを自然言語処理に心

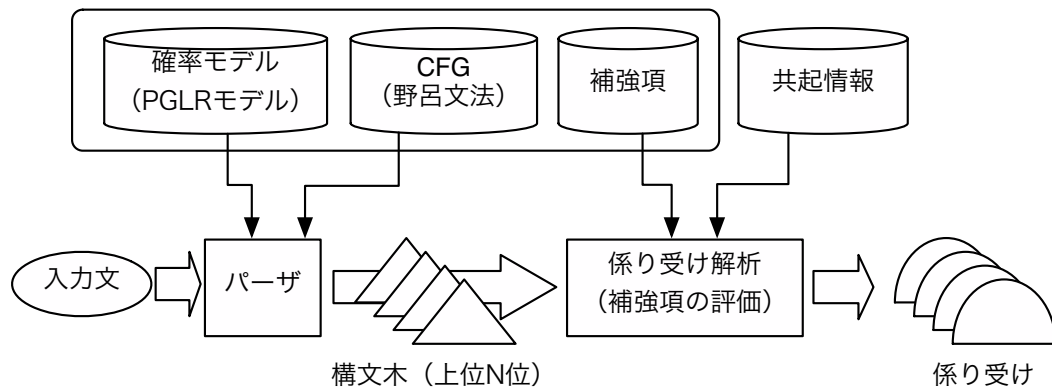


図 1: 補強項を利用した係り受け解析の概略

用すべき時期にきていることを示すことにある。SDSA の手法は、多種多様で広範な文の解析が可能な CFG 規則がありさえすれば、さらに解析精度を向上させたいとき、何をどのようにすれば良いかが（暗箱化した統計的手法に比べて）比較的透明であるという利点を持つからである。

本論文では、SDSA をベースとした係り受け解析を行うことにより、自然言語処理における SDSA の有効性や可能性について検討した結果を報告する。

第 2 節では、SDSA ベースの係り受け解析の概略について述べる。第 3 節では、本研究で使用する野呂コーパスと野呂文法についての概要を述べる。第 4 節では、係り受け関係に対するスコアの計算方法について述べる。第 5 節では、野呂文法を使った係り受け解析の実験方法とその結果について述べる。第 6 節では、本研究についての考察と今後の課題についてまとめる。

## 2 SDSA のフィージビリティ研究

SDSA には、幾つかの優れた特徴があると我々は考えている。文法の規模が大きくなればなるほど、天文学的な数の構文解析結果が出力されることが、SDSA の適用を妨げる大きな要因であることを前章で述べた。これが第 1 の問題である。ところが（多種多様な文の構文解析を行なうことが可能な）大規模文法の作成が実際には困難であることも既に指摘した。これが SDSA を巡る第 2 の問題である。大規模な文法の存在を前提にした SDSA による文法主導の意味解析が実質不可能であったというのが、現状である。幸いにして我々は、野呂の開発した大規模日本

語文法（野呂文法 [8, 9]）を持っており、前述した第 2 の問題をクリアすることができる。

野呂文法は、朝日新聞の記事に若干の変更を加えた EDR コーパスに現れる文を構文解析する目的で作成したものである。野呂文法の被覆率はおよそ 97% であり、日本語の新聞記事に関する広範囲の文の解析が可能である。野呂文法を使った各構文解析結果に対して、PGLR 確率モデルに基づくスコア付けを行ない [6]、上位 10 位以内の構文解析結果の中に、正しい解析結果が含まれる割合を調べたところ、およそ 92% であった。上位 100 位では 98% を越える。したがって、PGLR モデルによる構文解析結果に確率的なスコア付けをすることで、上位 10 位以内あるいは 100 位以内の構文解析結果に対してのみ SDSA を施すことで、ほぼ十分であると考えられるので、スコア付けの基準はさておき、第 1 の問題がクリアできる。

SDSA による解析の結果を他の統計的な手法を用いた結果と比較するために、本研究では、SDSA による文の係り受け解析を行ない、その解析精度を比較の尺度とする。ただし本研究は、SDSA のアプローチの有効性を確認するための前実験として、今回はおよそ次の手順で比較的粗い実験を行う (図 1)。

1. 与えられた CFG の各規則の補強項に係り受け解析用の手続きを記述する。
2. CFG と PGLR モデルを利用して、構文解析を行う。
3. 構文解析結果の上位 10 位の構文木に沿って、ボトムアップに補強項を評価し、係り受け関係のスコアを計算する。

| 野呂コーパス      |         |
|-------------|---------|
| 文数          | 8,912   |
| 形態素数 (異なり)  | 178,340 |
| 形態素数 (のべ)   | 22,347  |
| 一文当りの平均形態素数 | 20.01   |

| 野呂文法           |         |
|----------------|---------|
| 被覆率            | 97.24 % |
| 再現率            | 95.77 % |
| 上位 1 位の文の正解率   | 60.25 % |
| 上位 10 位の文の正解率  | 92.25 % |
| 上位 100 位の文の正解率 | 98.67 % |

$$\begin{aligned} \text{被覆率} &= \frac{\text{解析木が少なくとも一つは出力される文の数}}{\text{解析した文の総数}} \\ \text{再現率} &= \frac{\text{出力された解析木の中に正しい木が存在する文の数}}{\text{解析した文の総数}} \\ \text{上位 } N \text{ 位の文の正解率} &= \frac{\text{解析結果上位 } N \text{ 位以内に正しい木が存在する文の数}}{\text{解析した文の総数}} \end{aligned}$$

表 1: 野呂コーパスから抽出した文法の評価実験

4. 係り受け解析結果の上位の係り受けを出力する。

まず、PGLR による確率的なスコア付けをした構文解析結果上位 10 個を選び、それらに SDSA を施す。SDSA による意味解析といっても（名詞と助詞のペアからなる）助詞句と動詞との間の共起関係に関する統計データを主として用いて、係り受けの強さを確率的なスコアでランク付けする。係り受けの強さとして、係り受けに関する距離情報を用いたいところであるが、その詳細な情報は利用しない。受身や使役による格の交替も考慮しない。10 個の構文解析結果に SDSA を施してえた 10 個の結果の、最上位に位置する結果に対して係り受けの精度を測定する。したがって、本研究では、単語の共起データの一部を利用した係り受け解析を行なっているのではなく、本格的な意味解析を行なっている訳ではないことに注意したい。しかし、SDSA の有効性を確認する前実験としてはこれで十分であると考えている。この問題については、第 6 章で再び考察する。

### 3 野呂コーパスと野呂文法

本研究で使用した文法は、野呂が EDR コーパスをベースに開発しているコーパス（野呂コーパス）[8, 9] から規則を自動抽出することで作成した（野呂文法）。野呂コーパスは、文に対して句構造を付与しているが、構文解析結果の増加を

抑制するために複合名詞句、連体修飾句、並列修飾句の構造を制限している。

我々は、野呂コーパスから抽出された文法の性能を測るため予備実験を行なった。野呂コーパス (8,912 文) を 10 分割し、1 つをテスト用、残り 9 つを文法抽出用とし、10 ホールドクロスバリデーションにより評価を行った。評価尺度として、文法の被覆率、再現率、PGLR モデル [6, 17] による上位 1, 10, 100 位の文の正解率を利用した。PGLR モデルは、文法抽出に利用した文により学習した。各評価尺度の定義と実験結果を表 1 に示す。

表 1 より、野呂文法の被覆率はおおよそ 97 % であり、広範囲の文の解析が可能であることがわかる。野呂文法を使った構文解析結果に対して、PGLR モデルに基づくスコアづけを行ない [6, 17]、上位 10 位以内の構文解析結果の中に、正しい解析結果が含まれる割合は、約 92 % であり、この文法の有効性を確認した。

## 4 文節の係り受け関係のスコア

### 4.1 野呂文法と補強項

野呂コーパスから自動的に抽出した文法には、人手により補強項を付与した。補強項は、大きく 2 種類に分類される。一つは、係り受け関係のスコアを計算するための手続きである。もう一つは、文節区切りを決定するための手続きである。野呂コーパスに付与された構文構造は句

構造であるため，そのままでは文節の係り受け関係を決定することができない．そのため，単語列とヒューリスティックから文節区切りを決定する手続きを補強項として付与した．

#### 4.2 文全体の係り受け関係のスコア

野呂文法には，文節の係り受け関係を表す規則は，3種類に分類できる．一つは連体修飾句や並列句が名詞句に係る規則，もう一つは副詞などの連用修飾句が動詞句に係る規則，最後に名詞句と助詞で構成される助詞句が動詞句に係る規則である．そして，それぞれ分類に当てはまる規則に対して，人手により手続きを付与した．

構文木から導出される文全体の係り受け関係のスコアは，3種類の係り受け関係のスコアの和によって求める．

$$\begin{aligned} Sco(\text{全体}) = & Sco(\text{連体修飾句}) \\ & + Sco(\text{連用修飾句}) \\ & + Sco(\text{助詞句}) \end{aligned}$$

#### 4.3 連体修飾句の係り受け関係のスコア

野呂文法では，解析木の増加を防ぐため，連続する連体修飾句はすべて右下りの木に制限されており，連体修飾句の正しい係り受け関係が構文木の曖昧性として表現されない．そのため，連体修飾句の係り受け解析を行う場合には，連体修飾関係の曖昧性を解消しなければならない．

しかし，本論文では，すべての連体修飾句は最も近い名詞句に係るとした(図2)．連体修飾句が最も近い名詞句に係るように決定したのは，コーパスに出現する3つ以上の連体修飾句からなる名詞句における連体修飾句の係り先を分析したところ，最も近い名詞句に係る事例が最も多かったためである．

このように連体修飾句の係り受け関係を短絡的に決定したため，連体修飾句の係り受け関係のスコアは0とし，全体の係り受け関係のスコアには反映させないようにした．

$$Sco(\text{連体修飾句}) = 0$$

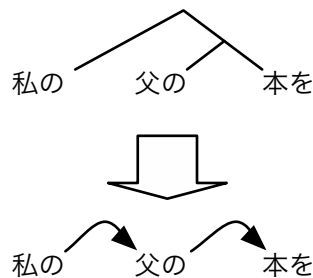


図 2: 連体修飾句の係り受け関係

#### 4.3.1 連用修飾句の係り受け関係のスコア

動詞の連用形，副詞などの連用修飾句と動詞句の共起事例をもとに連用修飾句のスコアを決定する．しかし，動詞と動詞，副詞と動詞などの共起事例は，名詞と動詞，名詞と名詞などの共起事例に比べ，獲得が非常に困難である上，事例数が少ない．本研究では，連用修飾句と動詞の十分が共起事例が獲得できなかったため連用修飾句の係り受け関係のスコアは，0とした．

$$Sco(\text{連体修飾句}) = 0$$

#### 4.3.2 助詞句の係り受け関係のスコア

構文木中に含まれる助詞句の係り受け関係のスコアは，各助詞句の係り受け関係のスコアの平均値で求める．

$$Sco(\text{助詞句}_1, \dots, \text{助詞句}_n) = \frac{\sum_{i=1}^n Sco(\text{助詞句}_i)}{n}$$

各助詞句の係り受け関係のスコアは，係り元である助詞句の中の格要素となる名詞句の主辞  $n$ ，助詞  $p$ ，係り先の動詞  $v$  の共起頻度をもとにスコアを決定する．

$$\begin{aligned} Sco(\text{助詞句}_i) & = Sco(n, p, v) \\ & = \alpha \times \frac{Fre(n, p, v)}{\sum_{v'} Fre(n, p, v')} \end{aligned}$$

$Fre(n, p, v)$  は  $n, p, v$  の共起頻度を， $\alpha$  は重みを表す．共起事例  $(n, p, v)$  は，白井 [16] が単語の共起を利用した語彙モデルの実験で用いたデータを利用した．その共起事例は，RWC コーパスと EDR 日本語共起辞書から抽出した延べ 7,964,320 組，異なりで 2,218,589 組である．RWC コーパスには構造は付いていないので，いくつかのヒュー

リストックスを使って、係り受け関係にあると思われる個所から事例を取り出している。

また、助詞句と述語の間に埋め込み文がある場合には、重み  $\alpha$  の値を小さくすることで助詞句がより近くの述語に係る様にした。ただし、主辞が「は」の場合、助詞句が読点を含む場合はこの限りではない。将来的には、この部分にも距離を考慮した確率値を与える必要がある。

## 5 評価実験とその考察

### 5.1 評価実験

野呂コーパスのすべての文 (8,912 文) から文法を抽出し、野呂コーパスからテスト用にランダムに 100 文を選択した。そして、テスト用の 100 文に対して、PGLR モデルを用いて構文解析を行った。PGLR モデルの学習には、野呂コーパスからテスト用 100 文を除いた 8,812 文を使用した。PGLR モデルによる生成確率上位 10 位までの構文木に対して、係り受け解析を行い、スコアが 1 位の係り受け解析結果を評価した。ただし、複数の係り受け解析結果のスコアが同じ場合には、PGLR モデルによる生成確率が高い構文木の出力結果を優先した。テスト文の正しい係り受け関係は、人手により付与した。その実験結果を表 2 に示す。

表 2 中の“係り受け A”とは、文末の文節を除く文節の係り受け正解率であり、“係り受け B”とは、文末から 2 文節を除く文節の係り受け正解率である。“文正解率”とは、文全体の文節の係り受け関係の正解率である。“PGLR”とは、係り受け関係のスコアを使用せず、PGLR モデルによる生成確率が 1 位の構文木から得られる係り受け解析の結果である。これが、ベースラインとなる。

### 5.2 評価実験結果の考察

前節では、SDSA のアプローチによる係り受け解析の評価実験方法とその結果について述べた。その実験結果 (表 2) からわかるように、PGLR のみによる係り受け正解率が約 90% 前後と非常に高い。これから、野呂文法が構文解析や係り受け解析に適した文法であることがわかる。さらに、非常に単純な計算方法により係り受けの

スコア付けしたにも関わらず、すべての項目に対して精度の向上が確認できた。特に、文正解率に関しては、約 7% も精度が向上している。我々は、この実験からは、SDSA の有効性を証明するまでには至っていないが、その可能性を見出すことができると考えている。

しかし、本研究は、まだまだ多くの課題を残している。一つ目は、係り受け関係のスコアの計算方法の改良である。本実験では、連体修飾句、連用修飾句の係り受け関係のスコアは 0、助詞句の係り受け関係のスコアは共起確率の平均値という単純な手法を使用した。白井 [16]、工藤 [18]、内元 [15, 14] をはじめとする先行研究で提案されているように統計的モデルを構築し、係り受け関係のスコアを計算する必要がある。

一つ目の課題と関連するが、二つ目は、係り受け解析に用いる統計的情報の充実である。名詞、助詞、動詞の共起事例 (異なりで約 200 万) を用いたが、テストで共起事例がなくスコアが計算されないケースが多くみられた。今後、さらなる共起事例の獲得が必要であると考えている。また、名詞、助詞、動詞の共起事例のみでなく、副詞と動詞、名詞と名詞、動詞と動詞などの共起事例も抽出し、利用できるようにしたい。その他にも、受け身や使役などによる格の交替、係り受けの距離、シソーラスを利用した語彙の意味的情報などを SDSA の手続きとして取り入れる必要もある。

三つ目は、実験規模の問題である。本実験では、テスト文を 100 文として評価を行った。この実験規模は、非常に小規模であるため SDSA の有効性を示すには至っていない。しかし、我々は、実験結果から SDSA によるアプローチが有効である可能性があると考えている。今後、野呂コーパス、野呂文法を含め実験規模の拡大と先行研究との正確な比較により評価を行う必要がある。

四つ目は、本格的な意味解析への移行である。本来、SDSA は、意味解析の枠組みであり、我々の目標も意味解析にある。本論文では、SDSA の枠組みを利用して係り受け解析を行うことで、意味解析の枠組みとして SDSA の可能性を調べた。その結果、十分に SDSA は意味解析の枠組みとして有効であるという考えている。今後は、SDSA ベースの意味解析手法の開発へも研究を進めていきたい。

|      | 係り受け A  | 係り受け B  | 文正解率    | 文節区切り不一致文数 |
|------|---------|---------|---------|------------|
| PGLR | 91.32 % | 89.61 % | 61.54 % | 9          |
| 共起   | 92.98 % | 91.61 % | 68.82 % | 7          |

$$\begin{aligned} \text{係り受け A} &= \frac{\text{正しい係り受け関係を決定できた文節数 (ただし, 文末の文節を除く)}}{\text{総文節数} - (\text{テスト文の総数} \times 1)} \\ \text{係り受け B} &= \frac{\text{正しい係り受け関係を決定できた文節数 (ただし, 文末の 2 文節を除く)}}{\text{総文節数} - (\text{テスト文の総数} \times 2)} \\ \text{文正解率} &= \frac{\text{正しい係り受け関係をすべて決定できた文の数}}{\text{テスト文の総数}} \end{aligned}$$

表 2: 評価実験

## 6 考察とまとめ

本研究では, SDSA の有効性を確認するための前実験という位置付けで実験を行なった. 文法主導の SDSA に対して PGLR による言語モデル, 単語の共起関係についての統計情報を用いているので, 文法主導 (ルールベース; 規則主導) を解析の骨組みとしつつ, 解析結果のランク付けには統計情報を用いている. したがって, 本論文で提案した SDSA の方式は, 文法主導に統計的手法を加味したハイブリッド方式であるといえる. 実験結果によれば, 本稿で提案した SDSA の方式は将来, 飽和状態に到達しつつあると考えられる統計的な方式に, 大きなブレイクスルーをもたらす可能性を秘めていると考えている.

SDSA の有効性を確認するための前実験から, 以下のことが今後の課題である.

1. SDSA に基づく本格的な意味解析の実験に進むことは十分意味がある.
2. SDSA による前実験として行なった係り受け解析について, さらに解析精度の向上を見込むことができる. その主な要因を以下にあげる.
  - (a) 集めた単語の共起データが非常に不足している. 大量の品詞付きコーパスから共起データを集める必要がある. 現在, 名詞と助詞の対 (助詞句) と動詞の共起データを主として用いた係り受け解析を行なっているが, この対の他に, 用言 (動詞; 形容詞; 副詞) と用言, 助詞「の」を介した名詞

と名詞の単語対などを考慮する必要がある. それにより, 係り受け精度の向上が見込まれる. これらの共起データは, 必ずしも隣接している必要がないので, データの収集が音声認識用の単語バイグラム, トライグラムのように簡単には収集できないかも知れない.

- (b) 不足している共起データを補うための, いわゆるスムージング技術を考えなければならない.
- (c) 単語と単語の共起データの利用に際して, 受身形や使役形に対する「格の交替現象」が考慮されていない. これは, 我々の SDSA の枠組では容易に取り込むことができる.
- (d) 深層格レベルでは「1 文 1 格の原則」が多くの場合成立しており, それが意味解析で有用な役割を果たすことが良く知られている. 本実験では, 単語の共起関係のみでこの原則を考慮していない. これは, 繰り返し述べたように, これまでの SDSA の前実験は, 本格的な意味解析を行なう実験ではないからである. 辞書項目の記述を充実させることにより, このような原則は容易に SDSA の枠組に組み込むことができる. 今後, 動詞記述に結合値情報や深層格情報を組み込み, 辞書項目を充実させた本格的な (SDSA による) 意味解析に取り組む意義は十分ある.

- (e) 係り受け解析で重要な「係り受け非交差の原則」は、CFGとして自然に実現されている。
  - (f) 係り受けの距離の問題についても今後詳しく検討する必要がある。
  - (g) 野呂文法は、SDSAによる解析に不都合な規則が若干含まれている。構文解析に後続するSDSAの立場からの文法体系の修正を行なう必要がある。これについては、すでに検討が始まっている。
3. SDSAを施した係り受け解析に付与する確率的スコアの計算モデルの検討を行なう必要がある。現在、SDSAによる解析結果のスコアに、PGLRモデルから得た確率値は補助的にしか反映されていない<sup>1</sup>。音声認識では音響モデルのスコアと単語認識モデルのスコアを適当に配分して全体のスコアを計算する。我々のモデルでも、PGLRのスコアとSDSAのスコアを適当に按分したスコアを用いることが考えられる。これは今後の研究課題である。最終的には、白井 [16] が提案した両者を統合化した確率モデルを考えることも必要かも知れない。

## 参考文献

- [1] A.V. Aho, R. Sethi, and J. Ullman. *Compilers, Principle, Techniques, and Tools*. Addison Wesley, 1986.
- [2] H. Alshawi, editor. *Core Language Engine*. The MIT Press, 1992.
- [3] J.W. Bresnan and R.M. Kaplan. Lexical-functional grammar; a formal system for grammatical representation. In Joan W. Bresnan, editor, *The Mental Representation of Grammatical Relation*, pp. 173–281. The MIT Press, 1978.
- [4] David R. Dowty, Robert E. Wall, and Stanley Peters. *Introduction to Montague Semantics*. D.Reidel Publishing Company, 1981.
- [5] Gazdar G., E. Klein, G. Pullum, and I. Sag. *Generalized Phrase Structure Grammar*. Havard University Press, 1985.
- [6] K. Inui, V. Sornlertlamvanich, H. Tanaka, and T. Tokunaga. Probabilistic GLR parsing. In *Proc. of the fifth International Workshop on Parsing Technologies*, pp. 123 – 134, 1997. Also in Harry Bunt et al. (eds.). *Advances in Probabilistic and Other Parsing Technologies*, Kluwer Academic Publishers, pp.85-104(2000).
- [7] Daniel Jurafsky and James H. Martin. *SPEECH and LANGUAGE PROCESSING*. Prentice-Hall, Inc., 2000.
- [8] 野呂智哉, 岡崎篤志, 徳永健伸, 田中穂積. 大規模日本語文法構築に関する一考察. 言語処理学会 第 8 回年次大会, pp. 387–390, 2002.
- [9] 野呂智哉, 八木豊, 橋本泰一, 徳永健伸, 田中穂積. 大規模日本語文法の開発に関する諸問題. 言語処理学会 第 9 回年次大会, pp. 121–124, 2003.
- [10] F.C.N Pereira and D.H.D. Warren. Definite clause grammar for language analysis—a survey of the formalism and comparison with augmented transition networks. *Artificial Intelligence*, Vol. 13, No. 3, pp. 231–278, 1980.
- [11] Carl Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, 1994.
- [12] V.R. Pratt. Lingol—a progress report. In *Proc. of the 4th International Conference on Artificial Intelligence*, pp. 372–381, 1975.
- [13] Ivan A. Sag and Thomas Wasow. *Syntactic Theory: A Formal Introduction*. CSLI Publications, 1999.
- [14] 内元清貴, 村田真樹, 関根聡, 井佐原均. 後方文脈を考慮した係り受けモデル. 自然言語処理, Vol. 7, No. 5, pp. 3–17, 2000.
- [15] 内元清貴, 関根聡, 井佐原均. 最大エントロピー法に基づくモデルを用いた日本語係り受け解析. 情報処理学会論文誌, Vol. 40, No. 9, pp. 3397–3407, 1999.
- [16] 白井清昭. 統計情報を利用した統合的自然言語解析. PhD thesis, 東京工業大学, 1998.
- [17] 白井清昭, 植木正裕, 橋本泰一, 徳永健伸, 田中穂積. 自然言語解析のためのMSLRパーザ・ツールキット. 自然言語処理, Vol. 7, No. 5, pp. 93–112, 2000.
- [18] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842, 2002.
- [19] 田中穂積, 佐藤泰介, 元吉文男. 自然言語処理のためのプログラミングシステム—拡張lingolについて. 電子通信学会誌, Vol. J60-D, No. 12, pp. 1061–1068, 1977.

<sup>1</sup>SDSAの解析結果の最上位のスコアが同一値のとき、PGLRモデルのスコアを優先させるという反映のさせ方をいう。