統計的日本語文解析における種々の統計量の扱いについて

白井清昭　乾健太郎　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
{kshirai,inui,take,tanaka}@cs.titech.ac.jp

Abstract

本稿では，形態素解析・構文解析・多義性解消からなる複合的問題に統計的手法を適用するために，個別の問題に対する既存の解決法をどのように拡張し，組み合わせればよいかについて論じ，構文モデル・語彙モデル・語義モデルからなる統合的言語モデルを提案する．提案するモデルは，確率文法，係り受け関係の距離の分布，隣接する品詞の従属関係，語彙的従属関係を反映する．モデルの主な特徴は，語彙的従属関係を評価するための従属係数という統計量を導入することにより既存手法のいくつかに確率的解釈を与え，それをモデルに組み込んだ点，格フレーム構造を導入することにより構文的依存関係と意味的依存関係が同形でない場合に対処できるようにした点である．

1 はじめに

統計的手法は形態素解析 [14]，構文解析 [15]，多義性解消 [3]，照応解消 [6] などさまざまな問題に適用され，比較的望ましい成果が得られている．しかしながら，これらの手法の多くは個別の問題に特化した統計モデルを用いているので，形態素解析から文脈処理までを含む複合的な問題にどのように適用すればよいかは自明ではない．たとえば，PP-attachment 問題については，「動詞: v，名詞: n_1，前置詞: p，名詞: n_2」という単語列における前置詞句の係り先を決定するための統計的手法がいくつか提案されているが [7, 15]，これらの手法では特定の文脈に特化した確率モデルを用いているので，自然言語解析におけるその他の問題を含む複合的な問題への適用が困難である．形態素解析が実際には構文や意味の情報を用いなければ完全には解くことができないように，自然言語解析に内在する種々の問題はいずれも独立に解くことはできない．個別の問題に対する解決法を洗練する努力を続けるとともに，その成果を統合して複合的問題に適用する方法を議論し，統合化の容易性という観点から個別の解決法を評価する必要がある．統合化を容易にするには，個別の解決法で用いられている統計量や解のスコアを組み合わせて利用するための共通の理論的基礎が必要であろう．本稿では，その第一歩として，日本語文解析における複合的問題に対する解に確率論的意味論を持つスコアを与える方法について論じる．

本稿で扱う問題は，形態素解析，構文的曖昧性解消，語義曖昧性解消の組み合わせである．もちろん，これらの問題を解くにはさらに文脈や語用論的情報が必要な場合が少なくない．また，これらの問題を解くことが自然言語解析の最終的な目的なのでもない．文脈や語用論的情報の扱いについては今後の課題である．ここでは，形態素解析，構文的曖昧性解消，語義曖昧性解消からなる複合的問題に対し，文内の情報を用いて解の候補を順序づけるという問題を考える．

入力文字列を A，A を生成する単語列の一つを W，W を生成する品詞列の一つを L，さらに L を生成する構文構造（句構造と係り受け構造のいずれを仮定してもよい）の一つを R とする．R は，品詞列 L を生成する構文規則のインスタンス r_i の集合で表すことができる．また，i 番目の語 w_i の語義の一つを s_i とする．

$$A = \{a_1, \ldots, a_l\}$$
$$W = \{w_1, \ldots, w_m\}$$
$$L = \{l_1, \ldots, l_m\}$$
$$R = \{r_1, \ldots, r_n\}$$
$$S = \{s_1, \ldots, s_m\}$$

$P(A|W) = 1$，$P(L|R) = 1$ が成り立つので，形態素・構文解析は

$$P(R, L, W|A) = \alpha_A \cdot P(R, L, W) = \alpha_A \cdot P(R, W) \quad (1)$$

によって解 R の順位づけを行う問題と見なすことができる（α_A は A に依存する定数）．また，これに多義性解消も加えると，

$$P(S, R, L, W|A) = \alpha_A \cdot P(S, R, W) \quad (2)$$

によって解 R，S の順位づけを行う問題になる．

(1)，(2) の同時分布は直接学習することができないので，訓練事例から学習可能な周辺分布を用いて近似することが必要になる．ここでは，多くの先行研究と同様，種々の条件付き独立を仮定し，同時分布を周辺分布の積で近似する方法を考える．近似の精度を上げるには，条件つき独立の仮定を制限し，周辺分布の母数を大きくすればよいが，その場合学習に必要な訓練事例数が増える．

したがって，我々の問題は，現在または将来入手できる妥当な量の訓練事例に対し，母数を学習可能な大きさに抑えながら，近似の精度をできるだけ高くするような条件つき独立の仮定を求めることであるといえる．

2 既存手法に対する考察

まず，先行研究で提案された統計モデルをいくつかとりあげ，その利点と問題点を検討する．

2.1 確率文法

$r_i \in R$ を品詞を終端記号とする文脈自由文法の規則とし，さらに品詞 l から単語 w を生成する文脈自由規則 $l \rightarrow w$ の集合を考えると，(1) の同時分布は (4) のように近似できる（確率文脈自由文法；PCFG）．

$$P(R,W) = P(R) \cdot P(W|L,R) \tag{3}$$

$$= \prod_{i=1}^{n} P(r_i|r_1^{i-1}) \cdot \prod_{i=1}^{m} P(w_i|w_1^{i-1},R)$$

$$\approx \prod_{i=1}^{n} P(r_i) \cdot \prod_{i=1}^{m} P(w_i|l_i) \tag{4}$$

ただし，$P(r_i)$ は規則 r_i の適用確率である．適用確率は共通の左辺を持つ規則の集合で正規化されている．w_i^j は単語列 $\{w_i,\ldots,w_j\}$ を表す．

PCFG は文法規則の適用確率に文脈自由性を仮定しているため，PP-attachment 問題など多くの構文的曖昧性を解消することができない．これに対し，これまでに少なくとも以下のような 4 通りの方策が提案されている．

第 1 は，規則適用確率が構文木導出の履歴に従属する確率モデルを用いるアプローチである．たとえば，Black は，導出履歴 r_1^{i-1} を同値類に分類する関数 $E(r_1^{i-1})$ を与え，$P(r_i|r_1^{i-1}) \approx P(r_i|E(r_1^{i-1}))$ と近似するモデルを提案している [1]．また， Kita らは規則適用の隣接 bigram を用いて $P(r_i|r_1^{i-1}) \approx P(r_i|r_{i-1})$ と近似するモデルを提案している [9]．ただし，これらはいずれも $P(R)$ を与えるモデルであり，後述する語彙的従属性は扱っていない．

第 2 は，CFG 規則の作り方を工夫するアプローチである．たとえば，Sekine らは，非終端記号を文 S と名詞句 NP に限定し，他の非終端記号をすべて終端記号に展開した CFG 規則を Penn Treebank から自動獲得し，それを用いて構文解析するアプローチを提案している [18]．これは，非終端記号による品詞列情報の抽象化を制限することにより，PCFG の文脈自由性の問題の解消をねらったもので，よい結果も報告されている．ただし，この方式では，規則数が爆発的に大きくなるので，文法の保守・拡張およびパラメタ推定の容易性に問題がある．また，文法獲得のための訓練事例が十分にないと，文法のカバレジも十分に上がらない．

第 3 は，単語共起の統計量をモデルに導入するアプローチである．一般的な PCFG では，語を品詞に抽象化する際に語の情報が捨象されるため単語の共起関係（以下，語彙的従属関係または語彙的従属性）は無視される．しかしながら，構文構造の優先度が語彙的従属性に強く依存することは多くの先行研究から明らかである．語彙的従属性を統計的に定量化し，構文解析に利用する手法についてはすでにさまざまな試みが報告されている．これについては，2.2 節で述べる．

第 4 は，構文構造の優先度に距離の尺度を導入するアプローチである．文節間の係り受け関係では，位置的に近い文節間の係り受け関係の方が遠いものより高い頻度であらわれることが Maruyama らによって確かめられている [13]．また，小林によると，「（名詞）の（名詞）の（名詞）」という形の名詞句の約 8 割が

[[[（名詞）の]（名詞）の]（名詞）]

というような係り受け構造をもつ [10]．したがって，係り受け関係にある単語間の距離の分布を確率モデルに採り入れることは有効であると予想される．これについては，2.3 節で述べる．

2.2 語彙的従属関係

2.2.1 語の共起情報を利用するモデル

Hindle らは，「動詞: v，名詞: n_1，前置詞: p，名詞: n_2」という単語列における前置詞句の係り先を決定する統計的手法を提案している [7]．Hindle らのモデルでは，v，n_1 が与えられたとき，v に係る前置詞 p が存在する確率 $P(p|v) \cdot P(\mathrm{null}|n_1)$ と n_1 に係る前置詞 p が存在する確率 $P(p|n_1)$ を比較することによって係り先を推定する．また，Ratnaparkhi らはこれと同じ問題に対し，$P(d|v,n_1,p,n_2)$（d は前置詞句の係り先が v か n_1 かを表す二値の確率変数）というモデルを用い，最大エントロピー法によってモデルのパラメタを推定する手法を提案している [15]．これらの手法は，語彙的従属性を確率モデルに採り入れた点で評価できるが，限定された問題に特化したモデルを用いているため，一般の文解析への拡張性に乏しい．

語彙的従属性の定量化に相互情報量を用いるアプローチも広く見られる．たとえば，小林は名詞 n_1，n_2 間の相互情報量 $\log \frac{C(n_1,n_2)}{C(n_1)\cdot C(n_2)}$ を用いて複合名詞句内の名詞間の係り受け解析を行う手法を提案している [10]．この手法では，係り受け構造の各候補のスコアを，係り受け関係にある名詞の組の相互情報量の幾何平均で与えており，その有効性も実験により示されている．相互情報量を用いる手法は一般の係り受け解析にも応用できるようにみえるが，スコアの意味論が明確でないため，他の統計量（たとえば，構文的優先度や単語の出現頻度など）と組み合わせるのが困難であると予想される．

一般の文解析を対象とする統計的手法に語彙的従属性を採り入れる研究もいくつか報告されている．たとえば，

Li らは，語彙的従属性に基づく解の優先度と構文的選好に基づく解の優先度を独立に求める方式を提案している．タグつきの入力文 L，W に対する構文構造のスコアは次のように計算される．

$$S(R|L,W) = S_t(T) \cdot S_r(R) \tag{5}$$

$$S_t(T) = \left(\prod_{\langle w_h, c, w_c \rangle \in T} P(w_c|w_h, c) \right)^{1/m} \tag{6}$$

(5) の第 1 項は語彙的尤度 (lexical likelihood) と呼ばれる尺度で，語彙的従属性に基づくスコアに当たる．第 2 項は長さ確率 (length probability) と呼ばれる構文的選好に基づくスコアである．語彙的尤度は，主辞 w_h の格 c に格要素 w_c が出現する確率 $P(w_c|w_h, c)$（以下，格要素の導出確率）の幾何平均である（式 (6)）．幾何平均を求めるのは，

A number of companies sell and buy by computer.

のような文を解析する際に，構文構造の候補によって格要素の導出確率を掛ける回数が異なるのを補正するためである．このように，Li らのモデルは，語彙的従属性と構文的優先度の組み合わせ方の一例を示しており，その有効性も実験的に確認されているが，小林の手法と同様，スコアの意味論が明確でないという問題が残る．

2.2.2 PCFG ベースのモデル

これに対し，PCFG をベースに語彙的従属性を評価する確率モデルが Hogenhout らや田辺らによって提案されている [8, 20]．たとえば，Hogenhout らの確率モデルでは，規則 $X \rightarrow Y_1 \ldots Y_m$ について，各記号 Y_i が支配する構成素の意味主辞 h_i に依存した適用確率 $P(Y_1{:}h_1, \ldots, Y_m{:}h_m|X)$ を与える．これによって，

動詞句 → 後置詞句:を 動詞句:食べる
後置詞句 → 名詞:ケーキ 後置詞:を

のような規則を用いて「を」と「食べる」,「ケーキ」と「を」の共起の強さを確率モデルに反映させることができる．ただし，厳密には，規則 $X \rightarrow Y_1{:}h_1 \ldots Y_m{:}h_m$ で X を展開した場合，次に Y_i を展開するときは Y_i の意味主辞 h_i が固定されているので，Y_i を展開する規則の適用確率は $P(Z_1{:}h_{11}, \ldots, Z_{m'}{:}h_{1m'}|Y_1{:}h_1)$ で与えられなければならない．このモデルは，PCFG の自然な拡張によって語彙的従属関係を表現することができるという良い性質を持っているが，モデルのパラメタの種類が構文規則の種類と共起する意味主辞の種類の組み合わせになるため，パラメタ空間が爆発的に大きくなる恐れがある（これについては Hogenhout らが様々なパラメタスムージングの方法を試みている）．たとえば，2.1 節で触れたように導出履歴を規則の適用確率の前件に追加しようとする場合，パラメタ数の増加はもとの PCFG の場合に比べ深酷である．

語彙項目と部分構文木の組を構文規則とする LTAG に規則の適用確率を付与した確率モデル SLTAG [16, 17] も語彙的従属性を評価するモデルになっているが，上で述べた PCFG ベースのモデルと同様，パラメタの種類が構文規則の種類と共起する意味主辞の種類の組み合わせになる．

2.2.3 Collins のモデル

Collins は，品詞づけされた単語列を日本語の文節に当たるような構成素にグルーピングし，構成素間の係り受け関係を決定する確率モデルを提案している [5]．この手法では，グルーピングで得られる構成素列を $B = \{b_1, \ldots, b_{m'}\}$，構文構造 R において b_i が関係 r_i で $d(b_i)$ に係るとし，$P(R|L,W)$ を次のように推定する．

$$\begin{aligned} & P(R|L,W) \\ &= P(B|L,W) \cdot P(d(b_1), r_1, \ldots, d(b_m), r_m|B, L, W) \\ &\approx P(B|L,W) \cdot \prod_{i=1}^{m} P(d(b_i), r_i|B) \end{aligned} \tag{7}$$

$$P(d(b_i){=}b_j, r_i|B) \approx \alpha_i \cdot F(r_i|b_i, b_j) \tag{8}$$

ただし，$F(r_i|b_i, b_j)$ は b_i，b_j が同じ文に存在するとき b_i が b_j に関係 r_i で係る確率，α_i は正規化のための係数である．Collins の確率モデルは，語彙的従属関係を反映するだけでなく，2.3 節 で述べるように「近い構成素に係りやすい」という距離の尺度を自然に取り込むことができるというよい性質を持っており，英語文の解析では良い実験結果も報告されている．しかしながら，次のような問題もある．

次の例を考えよう．

(a) [[[[瞳] の] 大きい] 少女] の] 写真
(b) [[[瞳] の] 大きい][[少女] の] 写真

この例では，構文構造の候補として少なくとも上の (a)「大きい」が「少女」に係る構文と，(b)「大きい」が「写真」に係る構文があり得る．Collins の確率モデルによると，この 2 つの候補の確率の比は $P(\text{d(大きい)} = \text{少女}|B, L, W)$ と $P(d(\text{大きい}) = \text{写真}|B, L, W)$ の比として計算される．ところが，(a) の構文では,「大きい」のは「瞳」であって「少女」でない．この場合,「大きい」と「少女」の依存関係よりも「瞳」と「少女」の「所有関係」に着目して (a) の構造を選択するべきだろう．このように，Collins の手法は，構文レベルの係り受け構造と意味レベルの依存構造が同形でない場合に，うまく確率を比較できないという問題がある．

2.3 係り受けの距離

2.1 節の最後で述べたように，係り受け関係にある単語間の距離の分布を確率モデルに採り入れることは有効であると予想される．距離の分布を考慮した確率モデルの例には Hogenhout らのモデル [8]，Li らのモデル [11]，Collins のモデル [5] があげられる．

Hogenhout ら，Li らのモデルは PCFG をベースにしている．Hogenhout らモデルでは，規則 $X \rightarrow Y_1 \ldots Y_m$ について，各記号 Y_i が支配する単語数（または非終端記号の数）n_i に依存した適用確率 $P(Y_1{:}n_1, \ldots, Y_m{:}n_m|X)$ を与える．これによって，たとえば

動詞句 → 後置詞句:n_1 動詞句:n_2

のような規則の場合，n_2が小さいほど後置詞句が近くの動詞に係ることを表せるので，その分布を学習すれば係り受けの距離を考慮した確率モデルが得られると期待できる．ただし，2.2 節で述べたのと同様の理由で，規則の適用確率は $P(Y_1{:}n_1, \ldots, Y_m{:}n_m|X{:}n)$（ただし，$n = \sum_i n_i$）で与えられることに注意する必要がある．このモデルは PCFG の自然な拡張になっているが，モデルのパラメタ数が CFG 規則と長さの分布との組み合せになり，推定パラメタの数が増大するといった問題がある．また，この他にも次のような問題が存在する．

例として，

PP → PP PP | N の
N → A | B | C | D

という文法と入力「A の B の C の D の」を考える．図 1 のような係り受け構造の候補 (a) と (b) を比較するとき，両者は「B の」の係り先が異なるだけなので，語彙的従属関係を考慮しない確率モデルでは係り受け関係の距離が短い (a) の方が優先されるはずである．しかしながら，Hogenhout らのモデルでは，(a) と (b) の確率の比は

$P(\mathrm{PP{:}6, PP{:}2|PP{:}8}) \cdot P(\mathrm{PP{:}4, PP{:}2|PP{:}6})$

と

$P(\mathrm{PP{:}4, PP{:}4|PP{:}8})$

の比で与えられることになるので（$P(\mathrm{PP{:}2, PP{:}2|PP{:}4) = 1}$ であることに注意），(a) は必ずしも優先されない．この例から推測されるように，係り受け構造を二分木で表そうとすると，各節点の左右の子の長さが均等に近い木の方が，可能な部分木の候補の数が少なくなり，結果として高い確率が与えられやすい．このことは，Hogenhout らのモデルが「近い単語に係りやすい」という現象を必ずしもうまく反映しないことを示唆する．これについては，同様に PCFG をベースにした Li らのモデルにも当てはまると予想される．

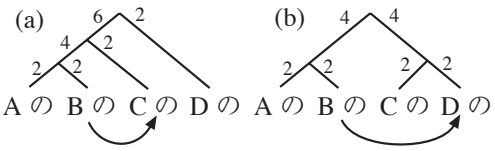

図 1: 係り受けの曖昧性と構成素の長さ

これに対し，Collins のモデルは，要素 b_i が他の要素 b_j に係る確率を直接学習し，係り受け構造の確率を計算するので，(8) を (9) のように拡張することにより，b_i と b_j の距離（δ_{ij}）の情報を自然な形で確率モデルに取り込むことができる．

$$P(d(b_i){=}b_j, r_i|b_i, b_j, \delta_{ij}) \approx \alpha_i \cdot F(r_i|b_i, b_j, \delta_{ij}) \quad (9)$$

一方，係り受け関係の距離の分布は，GLR 構文解析法に確率を導入した GPLR [2] の確率モデルを用いても反映させることができると考えられる．GPLR では，LR 表の各状態 s_i において，先読み語が l_j であり，そのときのアクションが a_k である確率 $P(l_j, a_k|s_i)$ を学習し，入力文の解析で起こる状態遷移列の確率をこれらの確率の積で近似する．例として，図 2 の l_j の手前まで解析が終了した状態を考える．次のアクションは，先読み語 l_j をシフトするか，X，Y を右辺に持つ規則によってこれらをレデュースするかのいずれかである．レデュースの場合は X の主辞の係り先が（Y の主辞に）決まるが，シフトの場合は X の主辞が Y をとびこえてその先の語に係ることになる．したがって,「近くの語に係りやすい」という現象は，図 2 のような状態における shift/reduce コンフリクトの確率分布に自然に反映されると期待できる．

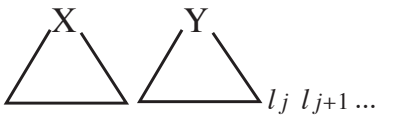

図 2: GLR 法で l_j の手前まで解析が終了した状態

2.4 形態素レベルの曖昧性

統計的形態素解析では，HMM に基づく確率モデル

$$P(L,W) \approx \prod_{i=1}^{m} P(l_i|l_{i-1}, l_{i-2}) \cdot P(w_i|l_i)$$

を用いる手法が代表的である [14]．しかしながら，これは L の生成確率を与えるモデルなので，同じく L の導出確率 $P(R,L)$ を与える PCFG ベースの確率モデルと組み合わせるためには，何らかの工夫が必要である．

一方，GPLR[2] は LR 表の各状態における先読み語を予測するモデルになっており，これは品詞間の bigram の前件事象をさらに LR 表の状態で分割したものに対応する．したがって，GPLR を使えば，構文構造の確率モデルと品詞の隣接関係のモデルを自然な形で統合できると考えられる．

2.5 まとめ

以上の議論を総合すると，次のような結論が得られる．

- PCFG ベースのモデルでは，規則の各記号をそれが支配する主辞の種類で区別することによって，語彙的従属関係を反映した確率モデルを得ることができる．ただし，その場合，パラメタの種類が CFG 規則の種類と単語の種類の組み合わせになり，パラメタ空間が組み合せ的に大きくなるという問題が深

酷である．また，係り受け関係の距離の分布をうまく反映する方法が今のところ発見されていない．さらに，PCFG ベースのモデルは品詞列 L を生成するモデルであるので，同じく品詞列 L を生成する HMM ベースのモデルと組み合わせることが困難であると考えられる．

- これに対し，GPLR は構文構造の確率分布と品詞の隣接関係の確率分布をうまく統合した確率モデルを与えることが期待できる．さらに，係り受け関係の距離の分布を確率モデルに取り込むことも期待できる．しかしながら，語彙的従属性を考慮するためには何らかの工夫が必要である．
- 一方，単語列を前件とするボトムアップな確率モデルである Collins のモデルは，単語間の共起の分布や係り受け関係の距離の分布を自然に反映させることができるが，構文的依存構造と意味的依存構造が同形でない場合に係り受け構造の候補間の比較がうまくできないことがある．

3 統合的言語モデル

前節の議論に基づき，以下のような統合的確率モデルを提案する．

まず，任意の R に対して一意に決まる格フレーム構造 T を考えよう ($P(T|R,W)=1$)．格フレーム構造は 3 つ組 $\tau=\langle t_i,t_j,t_k\rangle$ または$\tau=\langle t_i,\rho_j,t_k\rangle$ の集合で表現される．t_i は番号 i を持つ単語 w_i を指すポインタである．ただし，w_i の番号 i は，次の条件 (10) を満たすように付与されるものとする [1]．

$$\forall\langle t_i,t_j,t_k\rangle\in T \quad i<j<k \tag{10}$$

前者のタイプの 3 つ組 $\langle t_i,t_j,t_k\rangle$ は，主辞 w_i が格 w_j を持ち，その格要素が w_k であることを表す．たとえば，$W=\{w_5=$デパート$,w_4=$で$,w_3=$本$,w_2=$を$,w_1=$買う$\}$ のとき，$T=\{\tau_1=\langle t_1,t_4,t_5\rangle,\tau_2=\langle t_1,t_2,t_3\rangle\}$ である．このタイプの 3 つ組は，格関係などの統語的関係が表層の語によって明示的に示される場合に用いる．一方，後者のタイプの 3 つ組 $\langle t_i,\rho_j,t_k\rangle$ は，w_i と w_k が統語的関係 ρ_j にあることを表す．ρ_j は常に R によって一意に与えられるものとする．たとえば，$W=\{w_3=$昨日$,w_2=$買っ$,w_1=$た$\}$ のとき $T=\{\langle t_1,\rho_1=$助動詞修飾$,t_2\rangle,\langle t_1,\rho_1=$助動詞修飾$,t_3\rangle\}$ のような 3 つ組の集合が考えられる．ρ_j は，英語における主語や目的語のように，統語的関係が構文規則によって与えられる場合に用いる．格フレーム構造は各構文規則に付加された補強項によって決定的に生成されるものとする．

[1] この条件を満たす番号のつけ方が必ず存在するかどうかは今後の調査により明らかにする必要がある．

いま，$P(T|R)=1$ と仮定したので，1 節で設定した問題 (2) は次式のように展開できる．

$$\begin{aligned}P(S,R,W)&=P(S,T,R,W)\\&=P(R)\cdot P(W|T,R)\cdot P(S|T,R,W)\end{aligned}\tag{11}$$

$P(R)$，$P(W|T,R)$，$P(S|T,R,W)$ をそれぞれ構文モデル，語彙モデル，語義モデルとよぶ．以下，各モデルについて述べる．

3.1 構文モデル

$P(R)$ は単語間の共起関係を無視した，構文の構造的優先度を決めるモデルである．これについては，品詞の bigram，係り受け関係の距離の分布を自然に統合することができる統計的 GLR 法の確率モデル [2] を用いる．他の選択肢としては，PCFG の規則の適用確率を導出履歴に依存して変化させるというアプローチが考えられる [1, 9]．たとえば，最左導出における規則の適用確率が，その規則に支配される構成素の左隣りの品詞に従属するという確率モデルを考えれば，品詞の隣接関係を考慮したモデルになる．

3.2 語彙モデル

3.2.1 語彙モデル

語彙モデル $P(W|T,R)$ は以下のように展開できる．

$$P(W|T,R)=\prod_{i=1}^{m}P(w_i|w_1^{i-1},T,R)\tag{12}$$

まず，格フレーム構造の 3 つ組の第 1 項のみに現れるポインタに対応する語（典型的な語は動詞，形容詞，名詞）の導出を考える．これら主辞 w_i の導出は品詞 l_i のみに依存するとする．

$$P(w_i|T,R)\approx P(w_i|l_i)\tag{13}$$

この他，格フレーム構造に現れない語の導出も品詞のみに依存するとする．

次に，3 つ組の第 2 項に現れるポインタに対応する語（典型的な語は後置詞，後置詞相当語句，接続助詞）の導出を考える．任意の主辞 w_i について，w_i の格の集合を $V=\{w_j|\langle t_i,t_j,t_k\rangle\in T,R\}$ とし，これをあらためて $V=\{v_1,\ldots,v_n\}$ と書き，対応する品詞ラベルを $L'=\{l'_1,\ldots,l'_n\}$ と書くことにすると，V の導出確率は以下のように見積ることができる．

$$P(V|w_i,T,R)\approx\prod_{j=1}^{n}P(v_j|l'_i[w_i\{v_1^{j-1}\}])\tag{14}$$

$P(v_i|l'_i[w_i\{v_1^{i-1}\}])$ は，同一の主辞 w_i と共起する格の間の従属関係を考慮した格の導出確率である．また,「早く」と「起きる」の間の統語的関係「動詞修飾」のよう

に，3 つ組の第 2 項が表層の語に現れない統語的関係ρ_iである場合には，それを構文規則によって導出されたものとみなし，語彙モデルでは確率 1 で導出されると考える．

最後に，3 つ組の第 3 項に現れるポインタに対応する語 w_k（典型的な語は名詞，副詞，連体詞）の導出を考える．w_k について，w_k を格要素にとる主辞と格の組の集合を $U_k = \{w_i{:}w_j | \langle t_i, t_j, t_k\rangle \in T, R\}$ とすると，w_k の導出確率は式 (15) で与えられる．

$$P(w_k|w_1^{k-1}, T, R) \approx P(w_k|l_k[U_k]) \tag{15}$$

ただし，$P(w_k|l_k[U_k]) = P(w_k|l_k[w_{11}:w_{12}, \ldots, w_{n1}:w_{n2}])$ は，Li らの格要素の導出確率に対応するもので，品詞ラベル l_k を持つ語が存在し，それが n 個の主辞 w_{i1} とそれぞれ格関係 w_{i2} にあるとき，l_k から w_k が導出される確率を表す．とくに，$P(w_k|l_k[w_i{:}w_j])$ のとき，Li らのモデルと同じ格要素の導出確率になる．

3.2.2 従属係数

複数の主辞の格要素になっている語 w_i の導出確率は以下のように計算できる．まず，w_i が同時に 2 つの格 $w_{j1}{:}w_{j2}$，$w_{k1}{:}w_{k2}$ の格要素になる場合を考える．

$$\begin{aligned}
&P(w_i|l_i[w_{j1}{:}w_{j2}, w_{k1}{:}w_{k2}]) \\
&= \frac{P(l_i[w_{j1}{:}w_{j2}, w_{k1}{:}w_{k2}]|w_i) \cdot P(w_i)}{P(l_i[w_{j1}{:}w_{j2}, w_{k1}{:}w_{k2}])} \\
&\approx \frac{P(l_i[w_{j1}{:}w_{j2}]|w_i) \cdot P(l_i[w_{k1}{:}w_{k2}]|l_i, w_i) \cdot P(w_i)}{P(l_i[w_{j1}{:}w_{j2}]) \cdot P(l_i[w_{k1}{:}w_{k2}]|l_i])} \\
&= P(w_i|l_i) \cdot \frac{P(w_i|l_i[w_{j1}{:}w_{j2}])}{P(w_i|l_i)} \cdot \frac{P(w_i|l_i[w_{k1}{:}w_{k2}])}{P(w_i|l_i)} \\
&= P(w_i|l_i) \cdot L(w_i|l_i[w_{j1}{:}w_{j2}]) \cdot L(w_i|l_i[w_{k1}{:}w_{k2}])
\end{aligned} \tag{16}$$

ただし，ここでは，同一の語を格要素とする 2 つの主辞の間の条件つき独立性を仮定している．$L(w_i|l_i[w_{j1}{:}w_{j2}])$ は主辞 w_{j1} と格関係 w_{j2} に対する格要素 w_i の共起の強さを表す尺度で，これを $w_{j1}{:}w_{j2}$ に対する w_i の従属係数とよぶ．従属係数は次式で与えられる．

$$L(w_i|l_i[w_{j1}{:}w_{j2}]) = \frac{P(w_i|l_i[w_{j1}{:}w_{j2}])}{P(w_i|l_i)} \tag{17}$$

(16) より，従属係数については次式が成り立つ．

$$\begin{aligned}
&L(w_i|l_i[w_{j1}{:}w_{j2}, w_{k1}{:}w_{k2}]) \\
&\approx L(w_i|l_i[w_{j1}{:}w_{j2}]) \cdot L(w_i|l_i[w_{k1}{:}w_{k2}])
\end{aligned} \tag{18}$$

これを一般化すると，制約 $c_1, \ldots, c_n$ のもとでの単語 w_i の導出確率は以下のように近似できる．

$$P(w_i|l_i[c_1, \ldots, c_n]) \approx P(w_i|l_i) \cdot \prod_{j=1}^{n} L(w_i|l_i[c_j]) \tag{19}$$

ただし，c_j は，$w_j\{w_{p_1}, \ldots, w_{p_l}\}$ あるいは $w_{j1}{:}w_{j2}$ のいずれかで，前者は w_i が w_j に係る後置詞，後置詞相当語句，接続助詞などである場合，後者は w_i が w_{j1} の格 w_{j2} の格要素である場合に対応する．また，(19) が成り立つのは，制約 c_j が l_i，w_i のもとで条件つき独立と見なせる場合に限る．

(12) – (19) より，T，R から W を導出する確率は，単語間の従属関係を無視した導出確率 $P(w_i|l_i)$ と従属係数の積によって見積られることがわかる．w_i の導出確率に影響を与える独立な制約の集合を C_{w_i} とすると，$P(W|T, R)$ は以下のように計算できる．

$$P(W|T, R) \approx \prod_{i=1}^{m} P(w_i|l_i) \cdot \prod_{i=1}^{m} \prod_{c \in C_{w_i}} L(w_i|l_i[c]) \tag{20}$$

従属係数 $L(w_i|l_i[c])$ は w_i と c の独立性が高いとき 1 に近づく．両者に正の相関があれば値が大きくなり，負の相関があれば 0 に近づく．したがって，すべての従属係数を 1 とすると，(20) は語彙的従属関係を無視したモデルになる．

ここで注意したいのは，従属係数の対数が相互情報量に相当する点である．$P(w_i, w_{j1}{:}w_{j2})$ を $w_{j1}{:}w_{j2}$ の格要素が w_i である確率とすると，(17) より次式が成り立つ．

$$\begin{aligned}
\log L(w_i|l_i[w_{j1}{:}w_{j2}]) &= \log \frac{P(w_i|l_i[w_{j1}{:}w_{j2}])}{P(w_i|l_i)} \\
&= \log \frac{P(w_i, w_{j1}{:}w_{j2})}{P(w_i) \cdot P(w_{j1}{:}w_{j2})}
\end{aligned}$$

したがって，係り受け関係にある単語間の相互情報量の積によって係り受け候補のスコアを与えるモデル（たとえば，小林のモデル [10]）は，語彙モデル (20) のうち従属係数の積の部分だけを評価しているとみなすことができる．

3.2.3 例

例として,「妹と旅行に行く」という入力文に対する構文構造の候補の一つ

(a) [妹と] [旅行に] 行く

を考える．構文構造 (a) の導出に用いられた構文規則のインスタンスの集合を R，入力単語列を $W = \{w_5{=}$妹$, w_3{=}$と$, w_4{=}$旅行$, w_2{=}$に$, w_1{=}$行く$\}$，格フレーム構造を $T = \{\tau_1{=}\langle t_1, t_2, t_4\rangle, \tau_2{=}\langle t_1, t_3, t_5\rangle\}$ とすると，語彙モデルは次のように展開できる．

$$P(W|T, R) = \prod_{i=1}^{5} P(t_i{=}w_i, |t_1^{i-1}{=}w_1^{i-1}, T, R) \tag{21}$$

文の主辞「行く」の導出確率 $P(t_1{=}$行く $|T, R)$ はその品詞のみに依存すると仮定するので，(13) より

$$P(t_1{=}\text{行く}\ |T, R) \approx P(\text{行く}\ |\text{V}). \tag{22}$$

次に,「に」および「と」の導出について考える．これら後置詞句の導出は主辞にのみ従属すると仮定し，次のように見積る．

$$\begin{aligned}
&P(t_2=\text{に}, t_3=\text{と} \mid t_1=\text{行く}, T, R)\\
&\approx P(t_2=\text{に} \mid t_1=\text{行く}, \tau_1)\\
&\quad \cdot P(t_3=\text{と} \mid t_1=\text{行く}, t_2=\text{に}, \tau_2)\\
&= P(\text{に} \mid \text{P[行く}\{\}]) \cdot P(\text{と} \mid \text{P[行く}\{\text{に}\}]) \qquad (23)
\end{aligned}$$

最後に，格要素の導出を考える.「旅行」については，その導出が係り先の「行く」および格関係を表す「に」に従属し，その従属関係が無視できないと見なし，

$$\begin{aligned}
&P(t_4=\text{旅行} \mid t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, T, R)\\
&\approx P(t_4=\text{旅行} \mid t_1=\text{行く}, t_2=\text{に}, \tau_1)\\
&= P(\text{旅行} \mid \text{N[行く:に]}). \qquad (24)
\end{aligned}$$

「妹」の導出も同様である．

$$\begin{aligned}
&P(t_5=\text{妹} \mid t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, t_4=\text{旅行}, T, R)\\
&\approx P(\text{妹} \mid \text{N[行く:と]}) \qquad (25)
\end{aligned}$$

したがって，Li らの語彙的尤度（式 (6)）は，語彙モデルに含まれる項のうち，(24)，(25) のような格要素の導出確率だけを掛け合わせたものと解釈することができる．

次に，もう一つの構文構造 (b) の場合を考えよう．

(b) [[妹 と 旅行] に] 行く

(b) には $T' = \{\tau_1' = \langle \text{t}_1,\text{t}_2,\text{t}_4\rangle, \tau_2' = \langle \text{t}_1,\text{t}_2,\text{t}_5\rangle, \tau_3' = \langle \text{t}_4,\text{t}_3,\text{t}_5\rangle\}$ のような格フレームが対応すると考えられる．(a) との違いは，同一の名詞（w_5＝妹）が複数の 3 つ組の格要素に現れる点である.「と」,「に」についての格の導出確率，「旅行」についての格要素の導出確率はそれぞれ (23)，(24) と同様である.「妹」は τ_2'，τ_3' の両方に現れるので，

$$\begin{aligned}
&P(t_5=\text{妹} \mid t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, t_4=\text{旅行}, T', R')\\
&\approx P(t_5=\text{妹} \mid t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, t_4=\text{旅行}, \tau_2', \tau_3')\\
&= P(\text{妹} \mid \text{N[行く:に, 旅行:と]})\\
&\approx P(\text{妹} \mid \text{N}) \cdot L(\text{妹} \mid \text{N[行く:に]}) \cdot L(\text{妹} \mid \text{N[旅行:と]}) \qquad (26)
\end{aligned}$$

のようになる．これに対し Li らのモデルでは，(6) で各格関係ごとに格要素の導出確率を掛けるので，

$$P(\text{妹} \mid \text{N[行く:に]}) \cdot P(\text{妹} \mid \text{N[旅行:と]})$$

という計算をすることになる．これに確率論的意味論を与えるのは困難である．

3.2.4 構文的依存構造と意味的依存構造が乖離する場合

2.2.3 節で述べた，構文的依存構造と意味的依存構造が乖離する場合の問題に対しても，格フレーム構造を利用することによって対処できる可能性がある．たとえば，2.2.3 節の例では，単語列 $W = \{w_6 = \text{瞳}, w_5 = \text{の}, w_4 = \text{大きい}, w_3 = \text{少女}, w_2 = \text{の}, w_1 = \text{写真}\}$ について，(a),(b) それぞれの解釈に応じて次のような格フレーム構造を作ることができる．

(a) [[[[瞳] の] 大きい] 少女] の] 写真
$T = \{\langle t_6, \text{所有関係}, t_3\rangle, \langle t_4, t_5, t_6\rangle, \langle t_1, t_2, t_3\rangle\}$

(b) [[[瞳] の] 大きい][[少女] の] 写真
$T = \{\langle t_6, \text{所有関係}, t_1\rangle, \langle t_4, t_5, t_6\rangle, \langle t_1, t_2, t_3\rangle\}$

ただし,「所有関係」は R によって一意に特定される統語的関係である．語彙モデル $P(W|T,R)$ における (a) と (b) の確率の比は L(少女 |N[瞳 : 所有関係]) と L(写真 |N[瞳 : 所有関係]) の比になり，直観に合った比較ができる．

3.3 語義モデル

語彙モデル $P(S|T,R,W)$ は，式 (11) の第 2 項の単語導出の場合と同様，単語間の従属関係を無視した導出確率 $P(s_i|w_i)$ と従属係数の積によって見積もることができる．

$$P(S|T,R,W) \approx \prod_{i=1}^{m} P(s_i|w_i) \cdot \prod_{i=1}^{m} \prod_{c \in C_{w_i}} L(s_i|w_i[c]) \qquad (27)$$

3.4 パラメタの推定

本節で述べた確率モデルでは，構文モデルと語彙モデルが分かれているので，構文的優先度を与えるパラメタ（たとえば，GPLR において LR 表のアクションに割り当てられた確率）と語彙的従属関係（語彙レベルの従属係数）を与えるパラメタを個別に学習することができる．このことは，パラメタ数を抑えるだけでなく，パラメタ推定に必要な訓練事例の獲得のしやすさにも影響する．構文モデルと語彙モデルを分けないモデルでは，理想的には完全な係り受け解析を施したコーパスを訓練事例とする必要がある．一方，従属係数の学習では，コーパスの部分解析結果から信頼性の高い係り受け事例を抽出し，それを訓練事例とすることもできるので，必ずしも完全に構文解析されたコーパスを必要としない．構文モデルの学習には完全に構文解析されたコーパスが必要だが，語彙的従属関係と組み合わせる場合に比べパラメタ数が極端に少ないので，それほど大きなコーパスは必要でない．

従属係数のパラメタ数は，格要素の導出確率の場合，高々 (語彙の大きさ)^2 × (構文的関係の数) である．これの学習については，最大エントロピーを使う方法など，他稿 [19] で論じている．また，格の導出確率の場合は，(語彙の大きさ) × (格の組み合せ総数) である．これについては，格間の従属関係を考慮した導出確率を見積る方法がすでにいくつか提案されている [12, 21]．また，最大エントロピー法による推定も有効だと考えられる．

4 予備実験

本稿で提案した統合的言語モデルのふるまいを調べる手始めとして簡単な予備実験を行った．実験では，語彙モデル $P(W|T,R)$ の評価に焦点を当て，式 (28) に示すように構文モデルと語彙モデルのみを用いて確率を計算した．

$$P(T,R,W) = P(R) \cdot P(W|T,R) \tag{28}$$

また，$P(R)$ については，PCFG によって計算される解析木の生成確率をそのまま利用した．

語彙モデルの計算に用いる格フレーム構造の集合 T としては，動詞とその格要素の間の関係 $\langle v_i, p_j, n_k \rangle$ (v_i は動詞，p_j は格助詞または格助詞相当語句，n_k は名詞を表す) のみを考慮した．したがって，式 (20) における語彙モデルで考慮する従属係数は次の 2 種類である．

- 動詞に対する格の従属係数

$$L(p_j \mid \mathrm{P}[v_i]) = \frac{P(p_j \mid \mathrm{P}[v_i])}{P(p_j \mid \mathrm{P})} \tag{29}$$

- 動詞と格に対する格要素の従属係数

$$L(n_k \mid \mathrm{N}[v_i : p_j]) = \frac{P(n_k \mid \mathrm{N}[v_i : p_j])}{P(n_k \mid \mathrm{N})} \tag{30}$$

ただし，式 (29) の従属係数は，動詞が複数の格をとる場合にはそれらの従属関係は無視できるものとして計算している．これらの従属係数を語彙モデルに加えた場合と加えない場合の解析精度を比較することにより，本稿で提案した語彙モデルのふるまいを調査した．

4.1 実験手順

まず，式 (29),(30) で示した従属係数の計算方法について説明する．従属係数は (n_k,p_j,v_i) といった 3 つ組単語共起データから最尤推定する．推定に用いる共起データは EDR 共起辞書と RWC コーパスから収集した．EDR 共起辞書からは 403,329 組の共起データを抽出した．RWC コーパスはタグ付きコーパスであるので，名詞，助詞，動詞の共起関係が明示されているわけではないが，次に示す 2 つのヒューリスティクスを用いて 1,402,269 組の共起データを収集した．

- 「名詞 格助詞 動詞」という品詞列が現れた場合には，これらは共起しているとみなす．
- 文末に「名詞 助詞 動詞」という品詞列が現れた場合には，これらは共起しているとみなす．

収集した共起データの名詞の異り数は 89,578，助詞の異り数は 1,187，動詞の異り数は 14,560 であった．しかしながら，単語の共起の強さを表わす従属係数はパラメータ数が膨大で，これらを推定するのに十分な共起データが得られたわけではない．そこで，次のような手法を用いてデータスパースネスに対処した．

- 名詞，助詞，動詞のうち，特にデータがスパースなのは名詞 n_k である．そこで，n_k を意味クラス C_k を用いて抽象化した．意味クラスの導入に伴い，式 (19) に示した名詞の導出確率の計算を以下のように変更する．

$$\begin{aligned} &P(n_k \mid \mathrm{N}[v_i : p_j]) \\ &= \sum_{C_k} P(n_k \mid C_k) \cdot P(C_k \mid \mathrm{N}[v_i : p_j]) \quad (31) \\ &= \sum_{C_k} P(n_k \mid C_k) \cdot P(C_k \mid N) \cdot \\ &\quad L(C_k \mid \mathrm{N}[v_i : p_j]) \quad (32) \end{aligned}$$

  実験では，意味クラスとして分類語彙表の分類コードの上位 5 桁を使用した．
- 式 (30) において，前件事象 $v_i : p_j$ の出現頻度が小さければ小さいほど，推定される従属係数の信頼性も低いと考えられる．そこで，前件事象の出現頻度が十分大きい場合には従属係数を計算し，それ以外の場合の従属係数は 1 として語彙的従属性を無視した．実験においては，前件事象 $[v_i : p_j]$ の出現頻度が 500 以上の場合においてのみ従属係数を語彙モデルに反映した．

次に，EDR コーパスの中から，付加された構文木が二分木である例文 (22,780 文) を実験対象とし，この中からランダムに選択した 1,000 文をテストデータとした．テスト文の平均文長は 18.04 単語であった．次に，テストデータ以外の文を訓練データとし，これから式 (28) における $P(R)$ を計算する PCFG を学習した．さらに，学習した PCFG を用いてテスト例文の統語解析を行った．1 文当たり 458 個の解析木が生成された．実験結果を表 1 に示す．

表 1: 実験結果

	従属係数なし	従属係数あり
正解 (1 位)	43.90%	44.60%
正解 (5 位)	74.20%	74.90%
平均相対順位	0.50	0.46

表 1 において,「正解 (1 位)」の欄は確率が 1 位の解析木が正解である文の割合,「正解 (5 位)」の欄は確率の上位 5 位の解析木の中に正解が含まれる文の割合を示している．また,「平均相対順位」とは，正解となる解析木の確率による相対順位 (式 (33)) の平均を示している．

$$\frac{\text{正解の解析木の確率による順位}}{\text{生成した解析木の総数}} \tag{33}$$

4.2 解析結果の考察

前節の実験結果をみると，PCFG のみを用いて確率を計算した場合の解析精度は決して良いものとは言えな

い．また，従属係数を語彙モデルに追加しても解析精度の向上はほとんど見られなかった．そこで，前節の実験結果を詳しく分析し，解析に失敗する原因について考察した．以下，その主なものについて概説する．

1. 格フレーム構造の生成に不備があるもの
   - 使役，受け身による格の交替
     動詞が使役形や受け身形の場合は，それに係る格は表層とは異なる．例えば，実験においては，

     ··· わが国のモラトリアムを停止させた ···

     という例文に対して ⟨ 停止する, を, モラトリアム ⟩ という格フレーム構造を生成し従属係数を計算していた．しかしながら，この例文における「停止する」は受け身形になっており，これに係るヲ格は実際にはガ格に相当するので，⟨ 停止する, が, モラトリアム ⟩ という格フレーム構造に対して従属係数を計算するべきである．この問題は，動詞が受け身形または使役形の場合には，それに係る格を適切な格に変換して格フレーム構造を生成することで対処できると考えられる．
   - 述部が並列になっている場合
     文の述部が並列になっている場合には，1 つの後置詞句，特に「は」を伴って主題を表わす後置詞句が複数の用言に係る場合も考えられる．例えば，

     ··· 氏は重傷を負い後方に退く。

     という例文の場合は,「氏は」という主題を表わす後置詞句は「負い」と「退く」の両方に係るとみなすことができる．このような場合には,「氏」の係り先の解釈の違いに応じて次のような格フレーム構造を生成し，それらに対して従属係数を計算しなければならない．

     (a) 「氏」が「負う」のみに係る解釈
     ⟨ 負う, は, 氏 ⟩
     (b) 「氏」が「退く」のみに係る解釈
     ⟨ 退く, は, 氏 ⟩
     (c) 「氏」が両方に係る解釈
     ⟨ 負う, は, 氏 ⟩, ⟨ 退く, は, 氏 ⟩
   - 複合名詞，複合動詞
     実験では，複合名詞の主辞は一番最後の名詞，複合動詞の主辞は一番最初の動詞として格フレーム構造を生成していた．しかしながら，例えば「有毒/植物/群」という複合名詞に対しては，主辞は「植物」であると考えられるので，$\langle v_i, p_j,$ 植物 $\rangle$ といったような格フレーム構造を生成するべきである．この問題は，複合名詞や複合動詞が入力文中に存在する場合，それらの主辞を正しく同定してから格フレーム構造を生成することにより対処できる．
2. 格要素間の従属関係
   次の例文においては,「外国人」が「勉強する」に係るという解釈と「外国人」が「なる」に係るという解釈が存在する．

   外国人が日本語を勉強するための最適の
   日本語辞書となろう。

   実験では，それぞれの解釈の格フレーム構造 ⟨ 勉強する, が, 外国人 ⟩, ⟨ なる, が, 外国人 ⟩ に対する従属係数を比較した結果,「外国人」が「なる」という解釈に対して高い確率を与えてしまい，解析に失敗していた．この例文を正しく解析するには，⟨ なる, が, 外国人 ⟩ と ⟨ なる, と, 辞書 ⟩ という 2 つの格フレーム構造は共起しにくいというように，格要素間の依存関係を考慮しなければならない．この問題は，式 (15) で示した格要素の導出確率を，主辞がとる他の格要素にも依存するとして計算することにより解決できるが，このとき推定するパラメタ数が増大するといった問題も生じる．
3. 未知語が存在する場合
   未知語については，それが属する意味クラスを特定できないために従属係数を計算することができない．特に，固有名詞が含まれていることが原因で解析に失敗した例が多く見られた．未知語に対する従属係数を解析木全体の確率に反映させるには，未知語が現われる文脈 (未知語名詞の場合だったら係り先の動詞や表層格などの情報) からその意味クラスを推定する機構が必要となる．

5 おわりに

本稿では，形態素解析・構文解析・多義性解消からなる複合的問題に統計的手法を適用するために，個別の問題に対する既存の解決法をどのように拡張し，組み合わせればよいかについて論じた．とくに，確率文法，係り受け関係の距離，隣接する品詞の従属関係，語彙的従属関係に着目し，構文モデル・語彙モデル・語義モデルからなる統合的言語モデルを提案した．提案したモデルは，語彙的従属関係を評価するための従属係数という統計量を導入することにより既存手法のいくつかに確率的解釈を与え，それをモデルに組み込んだ点，格フレーム構造を導入することにより構文的依存関係と意味的依存関係が同形でない場合に対処できるようにした点が特徴的である．

語彙モデルは主辞・格・格要素それぞれの導出確率を個別に与えるモデルになっている．同一の主辞に対する複数の格の間の従属関係は格の導出確率に対応する従属係数で与えられる．もちろん，複数の格・格要素の組み

合わせに対し導出確率を与えるモデルを考えることもできる．宇津呂らの提案するモデルはその一例である [21]．

4 節で述べたように，本稿で提案した確率モデルの有効性を実証するためには，解決すべき問題が数多く残されている．今後は，これらの問題の解決策を検討していくと同時に，確率モデルのふるまいをさらに調査する種々の実験を進めていく予定である．

謝辞

有益なコメントをいただきました査読者の方，ならびに奈良先端大の松本裕治教授，宇津呂武仁氏，北陸先端大の Thanaruk Theeramunkong 氏に感謝いたします．

参考文献

[1] E. Black, F. Jelinek, J. Lafferty, D. M. Magerman, R. Mercer, and S. Roukos. Towards history-based grammars: Using richer models for probabilistic parsing. In *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics*, pp. 31–37. ACL '93, 6 1993.

[2] T. Briscoe and J. Carroll. Generalized probabilistic LR parsing of natural language (corpora) with unification-based grammars. *Computational Linguistics*, Vol. 19, No. 1, pp. 25–59, 3 1993.

[3] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer. Word sense disambiguation using statistical methods. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp. 264–270. ACL '91, 6 1991.

[4] COLING '92. *Proceedings of the 15th International Conference on Computational Linguistics*, 1994.

[5] M. J. Collins. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*. ACL '96, 1996.

[6] 江原暉将, 金淵培. 確率モデルによるゼロ主語の補完. 自然言語処理, Vol. 3, No. 4, pp. 67–86, 10 1996.

[7] D. Hindle and M. Rooth. Structural ambiguity and lexical relations. *Computational Linguistics*, Vol. 19, No. 1, pp. 103–120, 3 1993.

[8] W. R. Hogenhout and Y. Matsumoto. Experiments with using semantical categories in parsing systems. 言語処理学会第 2 回年次大会, 1996.

[9] K. Kita, T. Morimoto, K. Ohkura, S. Sagayama, and Y. Yano. Spoken sentence recognition based on HMM-LR with hybrid language modeling. *IEICE Trans. Inf. & Syst.*, Vol. E77–D, No. 2, 1994.

[10] 小林義行. コーパスを用いた日本語複合名詞の解析に関する研究. 博士論文, 東京工業大学, 1995. TR96-0002.

[11] H. Li. A probabilistic disambiguation method based on psycholinguistic. In *Workshop on Very Large Corpora*, pp. 141–154, 1996.

[12] H. Li and N. Abe. Learning dependencies between case frame slots. 人工知能学会言語・音声理解と対話処理研究会, No. SIG-NL-116-14, 1996.

[13] H. Maruyama and S. Ogino. A statistical property of Japanese phrase-to-phrase modification. *Mathematical Linguistics*, pp. 348–352, 1992.

[14] M. Nagata. A stochastic japanese morphological analyzer using a forward-dp backward-a n-best searchalgorithm. In *Proceedings of the 16th International Conference on Computational Linguistics*, Vol. 1, pp. 201–207. COLING '94, 1994.

[15] A. Ratnaparkhi, J. Reynar, and S. Roukos. A maximum entropy model for prepositional phrase attachment. In *Proceedings of the Workshop: Human Language Technology*, pp. 250–255. HLT '93, 1993.

[16] Philip Resnik. Probabilistic tree-adjoining grammar as a framework for statistical natural language processing. In *Proceedings of the 15th International Conference on Computational Linguistics* [4], pp. 418–424.

[17] Y. Schabes. Stochastic lexicalized tree-adjoining grammars. In *Proceedings of the 15th International Conference on Computational Linguistics* [4], pp. 425–432.

[18] S. Sekine and R. Grishman. A corpus-based probabilistic grammar with only two non-terminals. In *Proceedings of the International Workshop on Parsing Technologies '95*, 1995.

[19] 白井清昭, 乾健太郎, 徳永健伸, 田中穂積. 最大エントロピー法を用いた単語 bigram の推定. 情報処理学会自然言語処理研究会, No. NL–116, 1996.

[20] 田辺利文, 冨浦洋一, 日高達. 語の共起関係の文脈自由文法への取り込み法. EDR 電子化辞書利用シンポジウム論文集, pp. 25–31, 1995.

[21] 宇津呂武仁, 松本裕治. コーパスからの下位範疇化優先度の学習: 隠れ変数を用いた格の依存関係・格要素の汎化レベルの曖昧性の取り扱い. 信学技報, No. NLC–96, 1996.