自然言語処理 116－6
（1996. 11. 18）

種々の制約を統合した統計的日本語文解析

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

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．

The Integration of Statistics-based Techniques in the Analysis of Japanese Sentences

INUI Kentaro, SHIRAI Kiyoaki, TOKUNAGA Takenobu and TANAKA Hozumi
Department of Computer Science, Tokyo Institute of Technology
(2-12-1 Ôokayama Meguro Tokyo 152 Japan)

This paper presents a probabilistic language model for statistics-based morphological, syntactic and semantic analysis of Japanese sentences, which integrates some of the existing statistics-based techniques. The model couples structural preferences, morphological preference and lexical sensitivities on the theoretical basis.

1 はじめに

統計的手法は形態素解析 [13]，構文解析 [14]，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, 14]，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はない．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．

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．

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

一般に R が決まれば入力文の構文構造は一意に決まるから，形態素・構文解析は

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

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 は定数）．

ここで，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 $\tau = \langle t_i, \rho_j, t_k \rangle$ の集合で表現するものとする．t_i は番号 i を持つ単語 w_i を指すポインタである．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，$W = \{w_1 = \text{デパート}, w_2 = \text{で}, w_3 = \text{本}, w_4 = \text{を}, w_5 = \text{買う}\}$ のとき，$T = \{\tau_1 = \langle t_5, t_2, t_1 \rangle, \tau_2 = \langle t_5, t_4, t_3 \rangle\}$ 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 はつねに R によって一意に与えられるものとする．たとえば，$W = \{w_1 = \text{昨日}, w_2 = \text{買う}\}$ のとき $T = \{\langle t_3, \rho_1 = \text{副詞的修飾}, t_1 \rangle\}$ である．ρ_i 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る場合に用いる．いま，$P(T|R) = 1$ と仮定したので，(1) は次式のように書ける．

$$P(R, L, W|A) = \alpha \cdot P(T, R, W) \quad (2)$$

さらに，i 番目の語 w_i の語義の一つを s_i，その集合を

$$S = \{s_1, \ldots, s_m\}$$

とするとき，構文的曖昧性と語義曖昧性の解消問題は

$$P(S, R, L, W|A) = \alpha \cdot P(S, T, R, W) \quad (3)$$

によって解 S，R の順位づけを行う問題と見なせる．

ここで，(1)，(2)，(3) の結合確率はいずれも直接学習することができないので，コーパス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，結合確率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周辺分布のパラメタ数を大きくすればよいが，その場合学習に必要なデータ量が増える．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データに対し，パラメタ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付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確率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 $l \to w$ の集合を考えると，(1) の結合確率は (6) のように近似できる（統計的文脈自由文法；PCFG）．

$$P(R, L, W) = P(L, R) \cdot P(W|L, R) \quad (4)$$
$$= \prod_{i=1}^{n} P(r_i|r_1^{i-1}) \cdot \prod_{i=1}^{m} P(w_i|w_1^{i-1}, R) \quad (5)$$
$$\approx \prod_{i=1}^{n} P(r_i) \cdot \prod_{i=1}^{m} P(w_i|l_i) \quad (6)$$

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の問題を解消することができない．これに対し，構文木の導出の履歴に依存した規則適用確率を用いるモデルがいくつか提案されている [1, 2, 9]．ただし，これらはいずれも W に依存しない構文的な優先度 $P(L, R)$ を与える確率モデルである．

これに対し，単語と構文構造の組を構文規則とする LTAG に適用確率を与えるモデル [15] が提案されているが，次に述べるような語彙的従属関係を扱うものではない．

2.2 語彙的従属関係

構文的曖昧性の優先度は単語間の共起関係（以下，語彙的従属関係）に強く依存することが多くの研究者によって指摘されている [5, 7, 8, 11, 14]．

Hogenhout らは PCFG をベースとする確率モデルを提案している．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，たとえば

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，つぎに 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っているため，パラメタ空間が爆発的に大きくなる恐れがある．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（これについては Hogenhout らがさまざまなパラメタスムージングの方法を試みている）．同様な確率モデルは田辺ら [17] によっても提案されている．

これに対し，Li らは，語彙的従属関係と構文的優先度を独立に求める手法を提案している．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．

$$S(R|L, W) = S_t(T) \cdot S_r(R) \tag{7}$$

$$S_t(T) = \left(\prod_{\langle w_h, c, w_c \rangle \in T} P(w_c|w_h, c) \right)^{1/m} \tag{8}$$

(7) の第 1 項は語彙的尤度 (lexical likelihood) と呼ばれる尺度で，依存関係にある主辞の単語と補語の単語の間の共起の強さを計量する．第 2 項は長さ確率 (length probability) と呼ばれる尺度で，2.5 節で述べる Hogenhout らのモデルと類似するモデルである．語彙尤度は，主辞 w_h の格 c に補語 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 (8)）．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，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，語彙的従属関係と構文的優先度が独立に扱えるという利点があり，またその有効性も実験により示されているが，確率論的意味論を与えるのが難しいという問題もある．

2.3 格フレーム構造に基づく確率モデル

ここで，Li らの語彙的尤度を，格フレーム構造を利用した確率モデル上で解釈しなおしてみよう．「妹と旅行に行く」という入力文に対し，つぎのような構文構造が候補となる場合を考える．

(1) [妹と] [旅行に] 行く
(2) [[妹と] [旅行に]] 行く

この構文構造は品詞を終端記号とする文法規則と品詞から語を生成する語彙規則から生成できる．構文構造 (1) 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 $W = \{w_1 = \text{行く}, w_2 = \text{に}, w_3 = \text{と}, w_4 = \text{旅行}, w_5 = \text{妹}\}$ とする．ただし，単語 w_i の番号 i は，3.2 節 の条件 (21) を満たすように付与されている．構文 (1) の導出確率 $P(R, W)$ は式 (4) と同様に求められる．

$$P(R, W) = P(R) \cdot P(W|R) \tag{9}$$

以下，(9) の第 1 項 $P(R)$ を構文モデル，第 2 項 $P(W|R)$ を語彙モデルと呼ぶ．構文モデルは PCFG に代表される何らかの確率文法によって与えられるとする．語彙モデルは，$P(T|R) = 1$ なる格フレーム構造 $T = \{\tau_1 = \langle t_1, t_2, t_4 \rangle, \tau_2 = \langle t_1, t_3, t_5 \rangle\}$ を用いて次のように展開できる．

$$\begin{aligned} P(W|R) &= P(W|T, R) \\ &= \prod_{i=1}^{5} P(t_i = w_i, |t_1^{i-1} = w_1^{i-1}, T, R) \end{aligned} \tag{10}$$

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．

$$P(t_1=\text{行く}\,|T,R) \approx P(\text{行く}\,|\text{V}) \tag{11}$$

つぎに，「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すると，次のように計算できる．

$$\begin{aligned}
&P(t_2=\text{に}, t_3=\text{と}\,|t_1=\text{行く}, T, R)\\
&\approx P(t_2=\text{に}\,|t_1=\text{行く}, \tau_1)\\
&\quad \cdot P(t_3=\text{と}\,|t_1=\text{行く}, t_2=\text{に}, \tau_2)\\
&= P(\text{に}\,|\text{P}[\text{行く}\{\}]) \cdot P(\text{と}\,|\text{P}[\text{行く}\{\text{に}\}])
\end{aligned} \tag{12}$$

ただし，$P(w_j|\text{P}[w_i\{w_{p_1}^{p_k}\}])$ は，格 $\{w_{p_1},\ldots,w_{p_k}\}$ の係り先である主辞 w_i にさらに品詞ラベル P（格助詞）の別の語が係るとき，その P から w_j が導出される確率を表す．これを格要素の導出確率と区別して，格の導出確率と呼ぶ．

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すると，

$$\begin{aligned}
&P(t_4=\text{旅行}\,|t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, T, R)\\
&\approx P(t_4=\text{旅行}\,|t_1=\text{行く}, t_2=\text{に}, \tau_1)\\
&= P(\text{旅行}\,|\text{N}[\text{行く}:\text{に}]).
\end{aligned} \tag{13}$$

ただし，$P(w_k|l_k[w_{11}:w_{12},\ldots,w_{n1}:w_{n2}])$ 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．「妹」の導出も同様である．

$$\begin{aligned}
&P(t_5=\text{妹}\,|t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, t_4=\text{旅行}, T, R)\\
&\approx P(\text{妹}\,|\text{N}[\text{行く}:\text{と}])
\end{aligned} \tag{14}$$

したがって，Li らの語彙尤度は，語彙モデルに含まれる項のうち，(13)，(14) 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，構文構造 (2) の場合を考えよう．(2) 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．(1) 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 (12)，(13) と同様である．「妹」は τ_2'，τ_3' の両方に現れるので，

$$\begin{aligned}
&P(t_5=\text{妹}\,|t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, t_4=\text{旅行}, T', R')\\
&\approx P(t_5=\text{妹}\,|t_1=\text{行く}, t_2=\text{に}, t_3=\text{と}, t_4=\text{旅行}, \tau_2', \tau_3')\\
&= P(\text{妹}\,|\text{N}[\text{行く}:\text{に}, \text{旅行}:\text{と}])
\end{aligned} \tag{15}$$

のようになる．Li らのモデルでは，(8) で各格関係ごとに格要素の導出確率を掛けるので，

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

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．

2.4 意味レベルの依存構造

式 (1) の $P(R,L,W)$ はつぎのようにも展開できる．

$$P(R,L,W) = P(L,W) \cdot P(R|L,W) \tag{16}$$

たとえば，Chiang らは $P(R|L,W)$ を $P(R|L)$ で近似し，品詞列からボトムアップに構文構造を生成する確率モデルを提案している [4]．

また，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．

$$\begin{aligned}
&P(R|L,W)\\
&= P(B|L,W) \cdot P(d(b_1), r_1, \ldots, d(b_m), r_m | B, L, W)\\
&\approx P(B|L,W) \cdot \prod_{i=1}^{m} P(d(b_i), r_i | B)
\end{aligned} \tag{17}$$

$$P(d(b_i)=b_j, r_i|B) \approx \alpha_i \cdot F(r_i|b_i, b_j) \tag{18}$$

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，2.5 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利点も備えており，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，以下のような問題もある．

次の例を考えよう．

(1) [[瞳の] 大きい] 少女の] 写真

(2) [[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 (1)「大きい」が「少女」に係る構文と，(2)「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 $P(\text{d}(\text{大きい})=$少女$|B,L,W)$ と $P(d(\text{大きい})=$写真$|B,L,W)$ の比として計算される．ところが，(1) 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 (1) 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．

この問題は，2.3 節で述べた格フレーム構造を利用するモデルを使えば解決することができる．たとえば，上の例では単語列 $W=\{w_1=$写真$,w_2=$の$,w_3=$少女$,w_4=$大きい$,w_5=$の$,w_6=$瞳$\}$ に対し，(1) に対応する格フレーム

$$T=\{\langle t_6,\text{所有関係},t_3\rangle,\langle t_4,t_5,t_6\rangle,\langle t_1,t_2,t_3\rangle\}$$

と (2) に対応する格フレーム

$$T=\{\langle t_6,\text{所有関係},t_1\rangle,\langle t_4,t_5,t_6\rangle,\langle t_1,t_2,t_3\rangle\}$$

を作ることができる．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．2.3 節と同様に $P(W|T,R)$ を展開すると，(1) と (2) の構文確率 $P(R)$ が同じ場合，両者の比は P(少女 |N[瞳 : 所有関係]) と P(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．一方，Collins の手法は，各構成素ごとに係り先の確率分布を計算するので，上で述べたような問題を避けられない．

2.5 係り受けの距離

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 [12]．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．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るという問題の他につぎのような問題がある．

例として，

PP → PP PP，PP → N の

という文法と入力「A の B の C の D の」を考える (ただし，A，B，C，D は名詞)．図 1 のような係り受け構造の候補を比較するとき，(a) と (b) 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構文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

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

と

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

の比で与えられることになるので ($P(PP{:}2,PP{:}2|PP{:}4)=1$ 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．

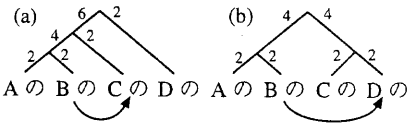

図 1: 係り受けの曖昧性と構成素の長さ

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，(18) を (19) 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．

$$P(d(b_i){=}b_j,r_i|b_i,b_j,\delta_{ij})\approx\alpha_i\cdot F(r_i|b_i,b_j,\delta_{ij}) \quad (19)$$

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 PGLR [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．PGLR 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．つぎのアクションは，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．

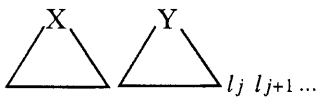

図 2: GLR 法で l_j の手前まで解析が終了した状態

2.6 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 [13]．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，PGLR[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，PGLR 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．

2.7 まとめ

以上の議論を総合すると，つぎ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いまのところ発見されていない．
- 一方，PGLR の場合，係り受け関係の距離の分布を確率モデルにとりこめることが期待できる．
- 単語列（文字列）の生成確率を与えるこれらのトップダウンな確率モデルでは，品詞から単語を導出する確率モデルに単語間の従属関係を反映させることにより，語彙的従属関係と構文的優先度を独立に学習するモデルを作ることができる．
- また，トップダウンな確率モデルでは，意味的依存構造を格フレーム構造で表現することにより，構文的依存構造と意味的依存構造が同形でない場合も扱うことができる．
- これに対し，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

前節の議論にもとづき，以下のような統合的確率モデルを提案する．まず，1 節で設定した問題 (3) を考え，これを以下のように展開する．

$$P(S,T,R,W) = P(R) \cdot P(W|T,R) \cdot P(S|T,R,W) \tag{20}$$

$P(R)$，$P(W|T,R)$，$P(S|T,R,W)$ をそれぞれ構文モデル，語彙モデル，語義モデルとよぶ．格フレーム構造は各構文規則 $r \in R$ 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．

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 bi-gram，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

$P(W|T,R)$ では品詞列から単語列を導出する．まず，条件 (21) を満たすように W の要素に番号を付ける [1]．

$$\forall \langle t_i,t_j,t_k \rangle \in T \quad i<j<k \tag{21}$$

このとき，

$$P(W|T,R) = \prod_{i=1}^{m} P(w_i|w_1^{i-1},T,R) \tag{22}$$

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．

$$P(w_i|T,R) \approx P(w_i|l_i) \tag{23}$$

つぎに，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計算される．

$$P(V|w_i,T,R) \approx \prod_{i=1}^{n} P(v_i|l'_i[w_i\{v_1^{i-1}\}]) \tag{24}$$

[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．

$P(v_i|l_i'[w_i\{v_1^{i-1}\}])$ 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．「早く」と「起きる」の間の統語的関係「副詞的修飾」のように，3つ組の第2項の統語的関係が表層の語に現れないものについては，それを構文規則によって導出されたもの見なし，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 の導出確率は次で与えられる．

$$P(w_k|w_1^{k-1}, T, R) \approx P(w_k|l_k[U_k]) \quad (25)$$

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．

$$\begin{aligned}
&P(w_i|l_i[w_{j1}:w_{j2}, w_{k1}:w_{k2}]) \\
&= \frac{P(l_i[w_{j1}:w_{j2}, w_{k1}:w_{k2}]|w_i) \cdot P(w_i)}{P(l_i[w_{j1}:w_{j2}, w_{k1}:w_{k2}])} \\
&\approx \frac{P(l_i[w_{j1}:w_{j2}]|w_i) \cdot P(l_i[w_{k1}:w_{k2}]|l_i, w_i) \cdot P(w_i)}{P(l_i[w_{j1}:w_{j2}]) \cdot P(l_i[w_{k1}:w_{k2}]|l_i])} \\
&= P(w_i|l_i) \cdot \frac{P(w_i|l_i[w_{j1}:w_{j2}])}{P(w_i|l_i)} \cdot \frac{P(w_i|l_i[w_{k1}:w_{k2}])}{P(w_i|l_i)} \\
&= P(w_i|l_i) \cdot L(w_i|l_i[w_{j1}:w_{j2}]) \cdot L(w_i|l_i[w_{k1}:w_{k2}]) \quad (26)
\end{aligned}$$

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．

$$L(w_i|l_i[w_{j1}:w_{j2}]) = \frac{P(w_i|l_i[w_{j1}:w_{j2}])}{P(w_i|l_i)} \quad (27)$$

(26) より，従属係数については次式が成り立つ．

$$\begin{aligned}
&L(w_i|l_i[w_{j1}:w_{j2}, w_{k1}:w_{k2}]) \\
&\approx L(w_i|l_i[w_{j1}:w_{j2}]) \cdot L(w_i|l_i[w_{k1}:w_{k2}]) \quad (28)
\end{aligned}$$

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．

$$P(w_i|l_i[c_1, \ldots, c_n]) \approx P(w_i|l_i) \cdot \prod_{j=1}^{n} L(w_i|l_i[c_j]) \quad (29)$$

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，分配則が成り立つのは，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かぎる．

(22) – (29) 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．

$$P(W|T,R) \approx \prod_{i=1}^{m} P(w_i|l_i) \cdot \prod_{i=1}^{m} \prod_{c \in C_{w_i}} L(w_i|l_i[c]) \quad (30)$$

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，(30) は語彙的従属関係を無視したモデルになる．また，2.2節の最後で述べた Li らのモデルの問題も，(14) を (29) により以下のように展開することによって解消できることに注意されたい．

$$\begin{aligned}
&P(\text{妹}|\text{N}[\text{行く}:\text{に}, \text{旅行}:\text{と}]) \\
&\approx P(\text{妹}|\text{N}) \cdot L(\text{妹}|\text{N}[\text{行く}:\text{に}]) \cdot L(\text{妹}|\text{N}[\text{旅行}:\text{と}])
\end{aligned}$$

また，従属係数の対数が相互情報量に当たることにも注意したい．$P(w_i, w_{j1}:w_{j2})$ を $w_{j1}:w_{j2}$ の格要素が w_i である確率とすると，(27) 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方法（たとえば，小林が提案している複合名詞句の係り受け解析法 [10]）は，語彙モデル (30) のうち従属係数の積の部分だけを評価していることになる．

3.3 語義モデル

(20) の第3項 $P(S|T,R,W)$ は，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．

$$P(S|T,R,W) \approx \prod_{i=1}^{m} P(s_i|w_i) \cdot \prod_{i=1}^{m} \prod_{c \in C_{w_i}} L(s_i|w_i[c]) \quad (31)$$

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，PGLR 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た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．これについては，最大エントロピーを使う方法など，他稿 [16] 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．これについても，最大エントロピー法による推定が有効だと考えられる．

4 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 [18]．

現在，提案したモデルのふるまいを調査する種々の実験を進めている．これについては別の機会に報告したい．

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] T. Chiang and K.Y. Su. Statistical models for deep-structure disambiguation. In *Workshop on Very Large Corpora*, pp. 113–124, 1996.

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

[13] 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.

[14] 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.

[15] Y. Schabes. Stochastic lexicalized tree-adjoining grammars. In *Proceedings of the 15th International Conference on Computational Linguistics*, Vol. 2, pp. 425–432. COLING '92, 8 1994.

[16] 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.

[17] 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.

[18] 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.