音声言語情報処理 32－3
（2000．7．14）

構文構造を反映した確率的言語モデル

森 信介 西村 雅史 伊東 伸泰
日本アイ・ビー・エム東京基礎研究所
〒242-8502 神奈川県大和市下鶴間 1623-14
{mori,nisimura,iton}@trl.ibm.co.jp

あらまし
本論文では、形態素を単位とする係り受けに基づく言語モデルを提案し、音声認識と構文解析の実験結果を報告する。我々が提案する言語モデルは、文を係り受け関係にある形態素の列とみなし、各形態素を文頭から順に予測する。ある時点での履歴は、部分的な構文解析の結果であり、形態素をノードとする木の列で表現される。このモデルでは、まず履歴である木の列のうち次の形態素に係る木の数を予測し、続いて、係る木の列から次の形態素を予測する。実験では、音声認識と構文解析を同時に行なう方法と音声認識結果の最尤の結果を構文解析する方法を比較した。その結果、音声認識結果と構文解析を同時に行なう方法の精度がより高かった。

キーワード 確率的手法 コーパス 構文解析 確率的言語モデル 音声認識

A Structural Stochastic Language Model

Shinsuke MORI, Masafumi NISHIMURA, Nobuyasu ITOH
IBM Research, Tokyo Research Laboratory, IBM Japan, Ltd.
1623-14 Shimotsuruma Yamatoshi Kanagawaken 242-8502 Japan
{mori,nisimura,iton}@trl.ibm.co.jp

Abstract
In this paper, we present a stochastic language model using dependency and report a result of speech recognition and systactic analysis. This model considers a sentence as a word sequence and predicts each word from left to right. The history at each step of prediction is a sequence of partial parse trees covering the preceding words. First our model predicts the partial parse trees which have a dependency relation with the next word and then predicts the next word from those trees. In our experiment, we compared two methods: a seamless combination of the speech recognition module and the parser, and a cascade combination. The result tells us that the accuracy by the seamless combination is more accurate than the cascade combination.

Key Words Stochastic Approach, Corpus, Parsing, Stochastic Language Model, Speech Recognition

1 はじめに

音声認識を端緒とする確率的手法は、自然言語処理の優れた方法論の一つである。実際、大語彙音声認識の際に使用される言語モデルの多くは n-gram モデルであり、英語などの形態素解析器の多くは品詞 n-gram モデルやその拡張に基づいている。[1, 2]。形態素解析は自然言語処理の最初の段階であり、確率的形態素解析器の精度は多くの応用に対して十分である。次の段階は文の構造を明らかにする構文解析である。最近、多くの確率的構文解析器が提案され、高い精度が報告されている。しかしながら、現状の精度は多くの応用には十分ではなく、さらなる精度向上が望まれる。

構文解析器の主な応用の一つとして、音声言語理解を目的とした認識結果の構文解析があろう。構文解析器と音声認識器を組み合わせることを考えた場合、構文解析器が生成的な確率的言語モデルを基礎とすることが望ましい。ここで、生成的とは、すべての可能な文字列に対する生成確率の和が 1 以下であることを意味する。言語モデルが生成的であれば、構文解析器と音声認識器を継目なく組み合わせることが可能になる。つまり、音声認識器の言語モデルを、構造を記述する言語モデルとし、従来の n-gram モデルよりも豊富な情報を用いて認識を行ない、同時に構文解析の結果を出力するのである。このような組合せが現実的なでない場合でも、音声認識器が N-best の認識結果をその確率とともに出力し、構文解析器はそれらを構文解析するとともに確率値を更新し、この結果得られる確率最大の文とその構造の組を出力することで、ほぼ同じ効果が期待される。

本論文では、形態素を単位とする係り受けに基づく言語モデルを提案し、音声認識と構文解析の実験結果を報告する。我々が提案する言語モデルは、文を係り受け関係にある形態素の列とみなし、各形態素を文頭から順に予測する。ある時点での履歴は、部分的な構文解析の結果である。これは、形態素をノードとする木の列であるが、我々が提案するモデルでは、まず履歴である木の列のうち次の形態素に係る木の数を予測し、続いて、係る木の列から次の形態素を予測する。実験では、音声認識と構文解析を同時に行なう方法と音声認識結果の最尤の結果を構文解析する方法を比較した。評価基準は、係り側と受け側の表記と品詞を含めた正しい係り受け関係の割合である。実験の結果、音声認識結果と構文解析を同時に行なう方法の再現率が 64.1% で適合率が 64.6% であり、音声認識結果の最尤の結果のみを構文解析する方法の再現率 (61.7%) と適合率 (62.4%) よりも高かった。

2 係り受けに基づく確率的言語モデル

この節では、我々が提案する係り受けに基づく確率的言語モデルについて述べる。係り受けを記述する多くの確率的言語モデルと異なり、我々のモデルは隠れマルコフモデルの一つである。我々のモデルでは、文を構成するそれぞれの形態素は文頭から順に予測される。予測の各段階での履歴は、基本的には、係り先が未定の形態素の列である。構文に対する心理言語学的研究 [3] によれば、文の各位置において、係り先が未定の形態素の数には上限がある。この上限は短期記憶のためのスロットの数によって規定され、7±2 程度であるとされる [4]。この制限を用いることで、有限状態機械に基づく確率的言語モデルを作成することが可能となる。

2.1 文のモデル

我々が提案するモデルの基本的なアイデアは、それぞれの形態素の予測においてより重要な情報は、直前の形態素列 (形態素 n-gram モデルなど) ではなく、予測される形態素と係り受け関係にある形態素であるとの直観である。例として、図 1に示される文構造と図 2に示される 6 番目の形態素「りんご」が予測される過程の文構造の仮説について考察する。図 2の上の部分解析木は、1 つのノードだけの木 (m_3 からなる t_b) と、2 つのノードからなる木 (m_1 と m_2 からなる t_a、m_4 と m_5 からなる t_c) がある。仮に最後の 2 つの木 (t_b と t_c) が次の形態素 (m_6) に係るとすると、この形態素はこれらの形態素から予測されるのがよいであろう。このような観点から、我々のモデルは、まず、次の形態素に係る部分解析木を予測し、次いで、これらの木から次の形態素を予測する。

ここで、本モデルを形式的に説明するために以下の定義を行なう。

- $\boldsymbol{m} = m_1 m_2 \cdots m_n$: 形態素列。形態素は文字列と品詞の対である。
- $\boldsymbol{t}_i = t_1 t_2 \cdots t_{k_i}$: 先行する i 個の形態素を覆う部分解析木の列。
- $\boldsymbol{t}_i^+$ 及び $\boldsymbol{t}_i^-$: 次の形態素に係る部分解析木の列、及び次の形態素に係らない部分解析木の列。係り受け関係は交差しないと仮定しているので $\boldsymbol{t}_i = \boldsymbol{t}_i^- \boldsymbol{t}_i^+$ である。
- $\langle \boldsymbol{t}\, m \rangle$: m を根とし $\boldsymbol{t}$ を根の部分木とする木。形態素 m_{i+1} がそれに係る木の列 ($\boldsymbol{t}_i^+$) から予測された後、係らない木の列 ($\boldsymbol{t}_i^-$) と新たに生成された木 ($\langle \boldsymbol{t}_i^+ m_{i+1} \rangle$) の連接が履歴となる。したがって $\boldsymbol{t}_{i+1} = \boldsymbol{t}_i^- \cdot \langle \boldsymbol{t}_i^+ m_{i+1} \rangle$ である。
- y_{max} : 係り先が未定の形態素の数の上限。

以上の定義のもと、本確率的言語モデルは以下のように

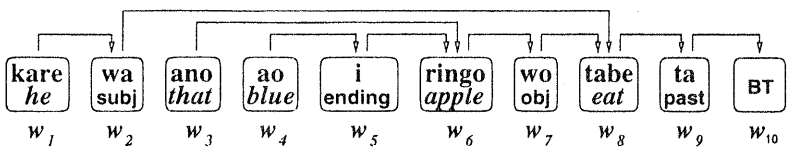

図 1: 文とその係り受け構造

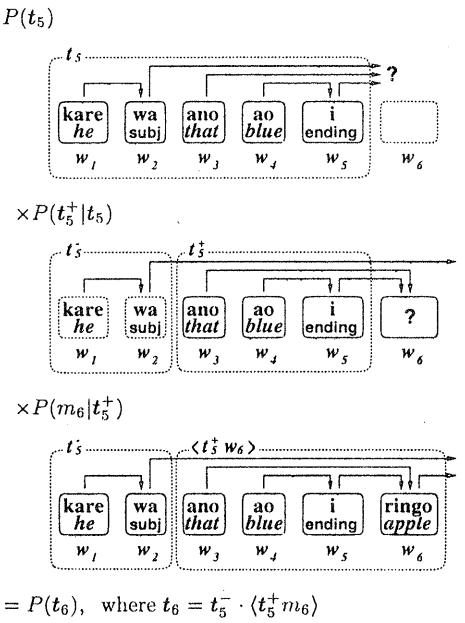

図 2: 部分解析からの形態素予測

定義される。

$$P(m) = \prod_{i=1}^{n} P(m_i|m_1 m_2 \cdots m_{i-1}) \approx \sum_{t_n \in T_n} \prod_{i=1}^{n} P(m_i|t_{i-1}^{+}) P(t_{i-1}^{+}|t_{i-1}) \quad (1)$$

ここで、T_n は n ノードの可能なすべての 2 分木の集合。この式の第 1 因子 ($P(m_i|t_{i-1}^{+})$) を形態素予測モデルと呼び、第 2 因子 ($P(t_{i-1}^{+}|t_{i-1})$) を状態予測モデルと呼ぶことにする。もう一度図 2について考察する。上段の図は 6 番目の形態素の予測の直後の状態である。状態予測モデルは、まず次の形態素に係る部分解析木を予測し、中段の図となる。次いで形態素予測モデルが、次の形態素をそれに係る部分解析木から予測し、下段の図となる。

すでに述べたように、形態素予測の各段階において、係り先が未定の形態素の数には上限があると考えられる。つまり、部分解析木列 (t_i) の要素数には上限がある。したがって、部分解析木の深さに上限があるとすると、可能なすべての状態の数は有限となる。この条件が満たされている限り、我々のモデルは隠れマルコフモデルである。

係り受け関係は交差しないことを仮定しているので、状態予測モデルは次の形態素に係る木の数を予測するだけで十分である。したがって、木の列 t_{i-1}^{+} の要素数を $y = |t_{i-1}^{+}|$ として、$P(t_{i-1}^{+}|t_{i-1}) = P(y|t_{i-1})$ となる。非交差の仮定から、最後の y 個の部分解析木が i 番目の形態素に係ることが分かる。形態素列に対する可能な解析木の数はその形態素数に対して指数関数的に増加するので、部分解析木の列の長さに上限がある場合でも、部分解析木の列がなす空間は広大である。このことは、データスパースネスの問題を引き起こす。この問題を避けるために、部分解析木の区別に利用するノードの深さを制限することとする。3節で述べる実験では、根とその子ノードのみを区別の対象とする。このように、最初のレベルの形態素と次のレベルの形態素を区別の対象とするモデルを P_{LL} と表す。したがって、実験に用いたモデルでは、次の形態素に係る木の数と形態素は、先行する形態素列を覆う部分解析木の深さ 2 以下の部分木を履歴と見なして予測される。もし、それぞれの形態素が次の形態素に係るという文構造を仮定すれば、本モデルは形態素 3-gram モデルと等価であることを付言しておく。

我々の提案するモデルを未知の入力に対して頑強にするために、n-gram モデルと同様の補間 [5] を行なう。木を区別する際の規則を緩和することで、より一般的なモデルが得られる。例えば、根とその子ノードの品詞のみを区別するとすれば品詞 3-gram モデルに似たモデル (以下 P_{PP} と表記) が得られる。根の形態素のみを区別し、その子ノードを無視すると形態素 2-gram モデルに似たモデル (以下 P_{NL} と表記) が得られる。平滑化の一手法として、形態素 3-gram に類似する P_{LL} を P_{PP} や P_{NL} などのより一般的なモデルと補間することが考えられる。実験では、以下の式が示すように一般化のレベルが異なる 7

つのモデルを補間した。

$$
\begin{aligned}
P(m_i|t_{i-1}^+) &= \lambda_6 P_{LL}(m_i|t_{i-1}^+) + \lambda_5 P_{PL}(m_i|t_{i-1}^+) \\
&\quad +\lambda_4 P_{PP}(m_i|t_{i-1}^+) + \lambda_3 P_{NL}(m_i|t_{i-1}^+) \\
&\quad +\lambda_2 P_{NP}(m_i|t_{i-1}^+) + \lambda_1 P_{NN}(m_i|t_{i-1}^+) \\
&\quad +\lambda_0 P_{m,0-\mathrm{gram}} \qquad (2)
\end{aligned}
$$

ここで、P_{YX} の X は深さ 1 のノードの区別のレベル (N: 区別しない、P: 品詞、L: 形態素) を示し、Y は深さ 2 のノードの区別のレベルを示す。また $P_{m,0-\mathrm{gram}}$ は、語彙 $\mathcal{M}$ に対する一様分布を表す ($P_{m,0-\mathrm{gram}} = 1/|\mathcal{M}|$)。

状態予測モデル ($P(y|t_{i-1})$) も全く同様に補間される。この場合、可能な事象は $y = 1, 2, \ldots, y_{max}$ であるので、$P_{y,0-\mathrm{gram}} = 1/y_{max}$ である。

本言語モデルが未知形態素を扱うことができるように、文字 2-gram モデルからなる未知語モデル [10] を付加した。語彙にない形態素を予測する場合、まずその品詞を予測し、次いで未知語モデルが品詞毎に異なる文字 2-gram モデルを用いて文字列の出現確率を計算する。

2.2 パラメータ推定

我々の提案するモデルは隠れマルコフモデルであるから、生コーパスから EM アルゴリズム [6] を用いてパラメータを推定することができる。このアルゴリズムを用いれば、生コーパスの出現確率が最大となるパラメータが推定される (F-B training)。この際、各文の構造は考慮されないので、結果として得られるモデルは、認識の目的には有効であるが、必ずしも構文解析に適切であるとは限らない。

構文解析にも適切なモデルを構築することを目的とした場合、構文構造が付与されたコーパスから、相対頻度による最尤推定を用いてパラメータを推定することが望ましい (Viterbi training)[7]。本報告では、認識とともに構文解析を行なうため、以下の式が示す Viterbi training を行なった。

$$
\begin{aligned}
P(m|t^+) &\overset{\mathrm{MLE}}{=} \frac{f(\langle t^+ m_i \rangle)}{\sum_m f(\langle t^+ m_i \rangle)} \\
P(y|t) &\overset{\mathrm{MLE}}{=} \frac{f(y,t)}{f(t)}, \quad \text{where } y = |t^+|
\end{aligned}
$$

ここで、$f(x)$ は事象 x の学習コーパスでの頻度を表す。

式 (2) の補間係数は、削除補間法 [5] によって推定される。

2.3 語彙化する形態素の選択

一般的に、形態素 n-gram モデルは品詞 n-gram モデルよりも予測力が高い。しかしながら、低頻度の形態素を語彙化することは、しばしばデータスパースネスの問題を引き起こし、モデルに悪影響を及ぼす恐れがある。例えば、

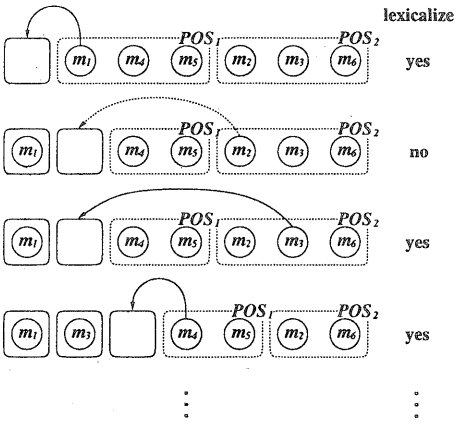

図 3: 語彙化のアルゴリズム

英語の形態素解析器 [8] では、頻度 100 以上の単語のみが語彙化されている。同様に、最も精度が高いとされる英語の構文解析器 [9] では、学習コーパスに 5 回以上出現する単語のみが語彙化される。このような理由から、本のモデルでは、語彙化する形態素を学習の時点で選択することとした。上述の形態素を区別の対象とするモデル (P_{LL} と P_{PL} と P_{NL}) では、選択された形態素のみを区別の対象とし、それ以外は品詞のみ区別する。選択の基準は、パラメータ推定とは別に用意された学習コーパスの一部であるヘルドアウトコーパスに対する構文解析の精度 (3節参照) とした。したがって、テストコーパスや未知の文の構文解析の精度を向上させると推定される形態素のみが語彙化される。この際のアルゴリズムは、以下の通りである (図 3参照)。

1. 初期状態では形態素はそれぞれの品詞に対応するクラスに属している。
2. すべての形態素は頻度の降順に整列され、この順に以下の処理が実行される。
   (a) 注目する形態素を仮に語彙化して、ヘルドアウトコーパスの解析精度を計算する。
   (b) もし、精度の向上が観測されればこの形態素を語彙化する。

このアルゴリズムの計算結果を部分解析木の区別に用いる。つまり、木の区別に際しては語彙化された形態素のみの文字列をチェックし、語彙化されなかった形態素は品詞のみ区別する。もし仮に、語彙化される形態素がなければ、$P_{NL} = P_{NP}$ であり、$P_{LL} = P_{PL} = P_{PP}$ となる。もし、形態素の併合も試みることとすれば、このアル

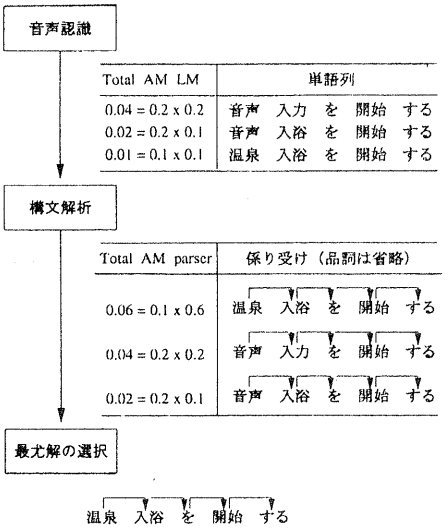

図 4: 同時処理の流れ図

表 1: コーパス

	文数	形態素数	文字数
学習	1,289	31,685	48,106
テスト	116	1,899	2,806

ゴリズムはトップダウンクラスタリングのアルゴリズムとなる。

3 評価

2節で述べた言語モデルに基づく構文解析器を構築し、音声認識と構文解析を同時に行なう方法と音声認識結果の最尤の結果を構文解析する方法を比較する実験を行なった。なお、解探索は、Viterbi アルゴリズムに基づいており、入力の文字数を n とすると、$O(n)$ の時間で最適の解を計算することができる。

3.1 実験の条件

実験に用いたのは、日本経済新聞の記事に含まれる文からなるコーパスである。各文は、形態素に分割され、構文構造が付与されている。学習コーパスの 1/10 を 2節で述べた語彙化アルゴリズムにおけるヘルドアウトコーパスとした。テストコーパスは、日本経済新聞の記事の 20 人の話者による読み上げ結果 116 文である (表 1参照)。予測の各時点で係り先が未定の形態素の数を学習コーパスについて調べた結果、この値を 10 とすることとした ($y_{max} = 10$)。

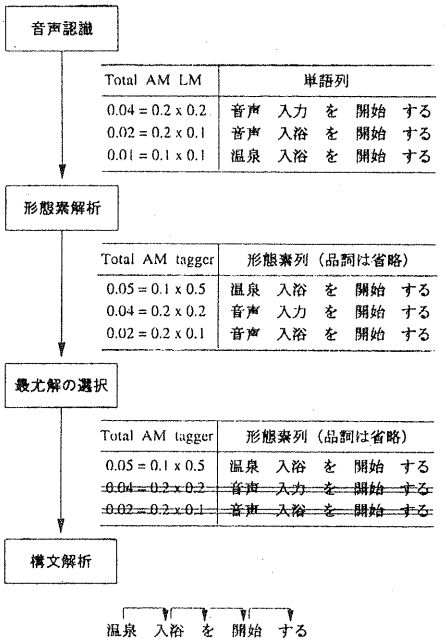

図 5: 逐次処理の流れ図

構文構造を反映した確率的言語モデルの効果を評価するために、大語彙音声認識装置 [11] によって得られた n-best 解に対して以下の 2 つの方法を比較した。各 n-best 解には、大語彙音声認識の大規模単語 3-gram モデルの尤度を除去することで得られる、音響モデルのみの尤度が付与されている。

同時処理 (図 4参照)

1. 音声認識の n-best 解のテキストを構文解析
2. 音響モデル部分の確率値と構文解析器の確率値との積が最大の文字列と構造の組を選択

逐次処理 (図 5参照)

1. 音声認識の n-best 解のテキストを形態素 2-gram モデル [10] により形態素解析
2. 音響モデル部分の確率値と形態素解析器の確率値の積が最大の形態素列を選択
3. 選択された形態素解析結果を、形態素区切りと品詞を固定して構文解析

逐次処理において、音声認識器の言語モデルの尤度を利用しない理由は、音声認識器が本モデルに比べて格段に大きい学習コーパスを用いて構築されていることである。

表 2: 認識と解析の精度

	構文解析		形態素解析
処理方法	再現率	適合率	%Accuracy
同時処理	64.1%	64.6%	74.6%
逐次処理	61.7%	62.4%	74.3%

構文解析結果の評価基準は、表記と品詞を含めた係り受け関係の精度である。つまり、推定された各係り受け関係について、係り側の表記と品詞と受け側の表記と品詞のすべてがコーパスに付与された結果と同じである場合のみ、正しい係り受けとみなす。この条件のもと、係り受け関係の精度は以下の値で評価する。

$$\text{再現率} = \frac{\text{正しい係り受け関係の数}}{\text{コーパスに付与された係り受け関係の数}}$$

$$\text{適合率} = \frac{\text{正しい係り受け関係の数}}{\text{推定した係り受け関係の数}}$$

形態素解析の精度は、解析結果を正解にするために必要な挿入と削除と置換の回数の最小値を正解の形態素数で割り、それを 1 から引いた値である。

3.2 評価

表 2は、同時処理と逐次処理のそれぞれの再現率と適合率である。この結果から、本言語モデルによる認識結果の選択とその構文解析では、これを同時に処理するほうが、認識結果の選択と構文解析を逐次処理するよりもよい結果となったことが分かる。したがって、音声認識結果を構文解析する場合には、音声認識器の言語モデルとして構文構造を反映した言語モデルを用いることが有効であると期待される。しかしながら、これには構文構造を付与した大規模なコーパスが必要であるという問題点がある。したがって、構文構造を効率良く付与する方法や、構文構造が部分的にのみ付与されているコーパスからモデルを構築する方法などが必要である。別の問題として、音響モデルと言語モデルの間での単語の単位の差異が挙げられる。本言語モデルの単位は形態素であり、音声認識器 [11] の認識単位よりも小さく、候補の絞り込みに本言語モデルを利用するのが困難である。この問題は、係り受け関係を保持した形態素列 (単語) を単位として本言語モデルを構築することで解決される。

4 結論

本論文では、係り受け構造を基礎とする確率的言語モデルについて述べた。このモデルでは、文は形態素の列とみなされ、それぞれの形態素は文頭から順に予測される。予測の各段階での履歴は、その時点までの形態素列を覆う部分解析木の列である。このモデルでは、まず、次の形態素に係る部分解析木を予測し、それから、次の形態素をそれに係る部分解析木から予測する。このモデルの評価するための実験として、音声認識と構文解析を同時に行なう方法と音声認識結果の最尤の結果を構文解析する方法を比較した。評価基準は、係り側と受け側の表記と品詞を含めた正しい係り受け関係の割合である。実験の結果、音声認識結果と構文解析を同時に行なう方法の再現率が 64.1% で適合率が 64.6% であり、音声認識結果の最尤の結果を構文解析する方法の再現率 (61.7%) と適合率 (62.4%) より高く、音声認識結果を構文解析する場合には、有望な方法であることが確認された。

参考文献

[1] Kenneth Ward Church. A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text. In *Proceedings of the Second Conference on Applied Natural Language Processing*, pp. 136–143, 1988.

[2] Evangelos Dermatas and George Kokkinakis. Automatic Stochastic Tagging of Natural Language Texts. *Computational Linguistics*, Vol. 21, No. 2, pp. 137–163, 1995.

[3] Victor H. Yngve. A Model and a Hypothesis for Language Structure. *The American Philosophical Society*, Vol. 104, No. 5, pp. 444–466, 1960.

[4] George A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, Vol. 63, pp. 81–97, 1956.

[5] Fredelick Jelinek, Robert L. Mercer, and Salim Roukos. Principles of Lexical Language Modeling for Speech Recognition. In *Advances in Speech Signal Processing*, chapter 21, pp. 651–699. Dekker, 1991.

[6] L. E. Baum. An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Process. *Inequalities*, Vol. 3, pp. 1–8, 1972.

[7] Bernard Merialdo. Tagging English Text with a Probabilistic Model. *Computational Linguistics*, Vol. 20, No. 2, pp. 155–171, 1994.

[8] Julian Kupiec. Augmenting a Hidden Markov Model for Phrase-Dependent Word Tagging. In *Proceedings of the DARPA Speech and Natural Language Workshop*, pp. 92–98, 1989.

[9] Michael Collins. Three Generative, Lexicalised Models for Statistical Parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pp. 16–23, 1997.

[10] 森信介, 山地治. 日本語の情報量の上限の推定. 情報処理学会論文誌, Vol. 38, No. 11, pp. 2191–2199, 1997.

[11] 西村雅史, 伊東伸泰, 山崎一孝. 単語を認識単位とした日本語の大語彙連続音声認識. 情報処理学会論文誌, Vol. 40, No. 4, pp. 1395–1403, 1999.