

# 音声入力による質問応答システムのための音声認識用言語モデルの検討

秋葉友良<sup>†</sup> 伊藤克亘<sup>\*†</sup> 藤井敦<sup>\*†</sup> 石川徹也<sup>‡</sup>

<sup>†</sup>産業技術総合研究所 <sup>‡</sup>図書館情報大学 <sup>\*</sup>科学技術振興事業団 CREST

e-mail: t-akiba@aist.go.jp

## 1 はじめに

質問応答 (QA) は、1999 年の TREC-8 にタスクとして採択されて以来、次世代の情報検索技術を目指した評価タスクとして注目されている。従来の情報検索タスクも音声入力に対応するように拡張されてきたが [4]、質問応答では入力が質問文というより話し言葉に近い表現が使用されることから、より音声入力に適したタスクであると考えられる。我々はこのような、音声入力を前提とした質問応答システムを開発中である。

質問応答システムへの入力となる検索者の発話、質問文という定型表現となる一方、QA の検索対象に関する多様な表現が使用される。そのため音声認識部では、これら性質の異なる 2 種の表現を同時に扱う言語モデルが必要となる。前者の定型表現には、記述文法を使い人手で特定タスクの言語モデルを作成することが考えられる [1]。データベースなどを対象としたタスクの限定された対話システムなどでは有効な方法であるが、QA タスクでは質問の対象が広範囲にわたるためすべてを人手で用意するのは現実的ではない。

一方、後者の検索対象に関する表現には、QA の知識源となる辞典や新聞記事などを利用して、そこから学習した N-gram 言語モデルを使うのが現実的である。しかし、これらを学習データとして使うだけでは、定型表現の質問表現を扱うには不十分である。質問文のための学習データを別途用意する必要があるが、収集のためのコストの面を考えると、これは容易ではない。この問題は、N-gram モデルのタスク適応の問題として捉えられてきており、比較的少数のタスク特有の学習データを使って、元のモデルをタスク用に再訓練する手法が提案されている ([6] など)。

これら 2 つのアプローチの利点を共に生かす方法として、N-gram と記述文法を何らかの方法で統合する方法が考えられる。鹿島ら [5] は、タスクの言語表現をモデル化した記述文法の間を、類似タスクから学習した bi-gram で補完する手法を提案している。

ところで、N-gram モデルは (確率付きの) 有限状態オートマトンと等価であり、正規言語の表現力を持つ。したがって、N-gram モデルの枠組の中で、正規文法<sup>1</sup> (ネットワーク文法) を表現することは可能なのである。そこで本稿では、汎用の N-gram モデルの中に、ネットワーク文法を持つ二値的制約を埋め込む手法を提案する。本手法により、N-gram による統計的知識と、記述文法による人手による知識を同時に一つのモデルに表すことが可能となり、広範囲の言語表現を N-gram でカバーしつつ、ある特定のタスクに用いる表現を記述文法で重点的に扱う言語モデルを獲得することが可能になる。また、提案法の言語モデルは従来の N-gram 表現形式で記

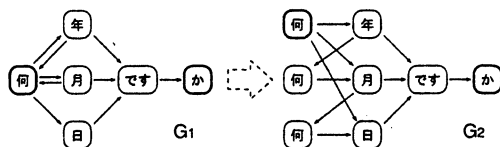


図 1: ネットワーク文法

述可能であり、言語モデルを置換えるだけで既存の音声認識デコーダでそのまま利用可能という、既存のシステムとの互換性の面でも優れた特徴を持つ。

## 2 ネットワーク文法の bi-gram モデル表現

単語 bi-gram は、単語を頂点とし、全ての単語間の有向弧に確率が付与された、重み付き (ループのある) 完全有向グラフと見ることができる。この時、有向弧の bi-gram 確率が 0 である場合、その単語連続があり得ないことを表すため、弧が存在しないことと等価である。したがって、単語を頂点として表現した任意のネットワーク文法から、有向弧が存在する場合は 0 でない確率値を、有向弧が存在しない場合は確率 0 を割り当て、単語 bi-gram で表現することが可能となる。もしネットワーク文法中で、ある単語から入出力する有向弧を、文脈に応じて変えたいのであれば、文脈の数だけ同じ単語を表す頂点を複製して表現すればよい。

このような、単語 bi-gram で表現されたネットワーク文法は、例文の集合から簡単に獲得可能である。例えば、年月日を尋ねる発話を表した以下の例文から文法を獲得することを考える。

何/年/です/か    何/年/何/月/です/か  
何/月/何/日/です/か

この 3 文から獲得できる接続可能な単語対は以下の通りである。

$A = \{ (何, 年), (何, 月), (何, 日), (年, 何), (月, 何), (年, です), (月, です), (日, です), (です, か) \}$

この単語対だけが接続可能であると考え、ネットワーク文法 ( $G_1$ ) は 4 つ組 ( $W_a, W_s, W_f, A$ ) で表現できる。ここで、 $W_a, W_s, W_f$  は、それぞれ、全単語集合、開始単語集合、終了単語集合であり、

$W_a = \{ 何 年 月 日 です か \},$   
 $W_s = \{ 何 \}, W_f = \{ か \}$

となる。 $G_1$  のグラフ表現を図 1 左に示す。

この時、ネットワーク文法は、以下の制約を満たす bi-gram として表現できる。

$$P(w_j|w_i) = \begin{cases} p(>0) & \text{if } (w_i, w_j) \in A \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where } \sum_{w_j \in W_a} P(w_j|w_i) = 1$$

文法  $G_1$  は「何年何年ですか」「何月何年ですか」「何年何日ですか」のような、意図されない言語表現までモデル化してしまう。そこで、文法作成者の持つ言語知識を利用して、好ましくない表現を排除し、図 1 右のようなネットワーク文法  $G_2$  に修正することを考える。新たに導入したノード (文脈) 毎

<sup>1</sup> 本手法で扱える記述文法は、(N-gram の表現力の制約の故に) 正規文法までである。自然言語の記述として広く用いられている文脈自由文法は、そのまま埋め込むことはできない。しかし、有限長の文法は必ず正規言語で表現できること、文脈自由文法を正規文法に近似するアルゴリズムが知られている [2] こと、などから実用上ほとんど問題はない。

に、新たな単語記号を導入して、次のような文法 ( $W_a, W_s, W_f, A'$ ) として表現する。

$$W'_a = W_a \cup \{ \text{何 1 何 2} \}$$

$A' = \{ (\text{何}, \text{年}), (\text{何}, \text{月}), (\text{何}, \text{日}), (\text{年}, \text{何 1}), (\text{何 1}, \text{月}), (\text{月}, \text{何 2}), (\text{何 2}, \text{日}), (\text{年}, \text{です}), (\text{月}, \text{です}), (\text{日}, \text{です}), (\text{です}, \text{か}) \}$

文法  $G_2$  は、「何年何月ですか」「何月何日ですか」のような、作成者の意図する表現だけを受理し、それ以外を排除する。このように、ネットワーク文法では、人の持つ言語知識を利用して、N-gram では獲得不可能な、単語間の長距離の依存関係も表現することが可能である。

### 3 ネットワーク文法と汎用 N-gram の統合

前節で述べた方法によって作成したネットワーク文法を、新聞記事などから学習した既存の N-gram モデルと統合する。N-gram がモデル化する単語列の中に、ネットワーク文法によって表される単語列が過不足無く現れる (ネットワーク文法開始単語から終了単語までのパスが存在するような単語列がある) 場合には、その単語列を優先するように、モデルを作成する。このような統合によって、N-gram により汎用の単語列が認識されるのと同時に、タスクに特有の表現を優先して認識することが可能になる。例として、QA タスクの入力には、答えを得るための次のような質問文が想定される。

1976 年に火星に軟着陸した探査機は何という名前でしたか

この入力中、文末の「何という名前でしたか」の部分は質問文に典型的に現れるパターンであり、ネットワーク文法でモデル化する。一方、「1976 年に火星に軟着陸した探査機は」の部分は汎用の N-gram モデルで扱う。

ここで問題となるのは、性質の異なる 2 種のモデルをどのように統合して 1 つのモデルで表現するかということである。ネットワーク文法は、単語接続が二値的で、陽に指定した単語接続以外は許さない。また、ネットワークの開始から終了までの単語列が過不足無く現れている必要がある。一方、N-gram モデルでは、全ての単語が互いに接続可能とし、接続の強弱は確率値で表現するように作成するのが普通である。

既存の N-gram とネットワーク文法を混合して、単純に N-gram モデルを作成すると、全ての単語が接続可能なモデルとなる。これは、N-gram モデルの学習には、スムージングが行われるためである。バックオフ・スムージングでは、高次の N-gram が存在しない場合、低次の N-gram で補間される。bi-gram で二値的制約を表現しても uni-gram で補間されるため、結局すべての単語とネットワーク内単語との間で接続可能となってしまう。一方、全くスムージングを行わないモデルを作成することも出来るが、その場合ネットワーク部の二値的制約は獲得されるが、N-gram 部にゼロ頻度問題が生じ、精度が落ちてしまう。

提案法の基本的アイデアは、統合モデルの部分に応じてスムージング方法を切り替えて学習を行うというものである。二値的な接続制約が必要な部分はバックオフスムージングを行わずに学習を行い、その他の部分はその部分の単語集合だけを元にバックオフスムージングを行う。以下では、既存の bi-gram とネットワーク文法から、統合 bi-gram を作成する

場合について説明する。tri-gram の場合も、同様の方法が適用可能である。

#### 3.1 ネットワーク文法の作成

まず、既存 N-gram の語彙を用いてネットワーク文法を作成する<sup>2</sup>。2 節で述べたように、ネットワーク文法は例文から獲得することもできるが、より良いモデルを作成するためには人手によるチューニングを行うのが好ましい。

開始単語と終了単語は、他のネットワーク内単語と区別して、それぞれ開始時、終了時のみ到達可能とするようにネットワークを構成する<sup>3</sup>。すなわちネットワーク内単語集合  $W_G$  は、互いに共通部分のない、開始単語集合  $W_{G_s}$ 、終了単語集合  $W_{G_f}$ 、中間に現れる単語集合  $W_{G_m}$  から構成されるとする。また、ネットワーク文法の単語は、元の N-gram の同じ単語と区別するために、別の単語記号を割り当てる。本稿では、N-gram 中の単語  $w$  に対して、ネットワーク文法の単語を、記号 '@' を付けて @w と表すことにする。

作成したネットワーク文法を元の N-gram とマージし、統合モデルとする。統合モデルの語彙  $W_A$  は、元の N-gram からの単語  $w \in W_U$  と、ネットワーク文法からの単語  $W_G (= W_{G_s} \cup W_{G_m} \cup W_{G_f})$  から構成される。 ( $W_U \cap W_G = \phi$ )

#### 3.2 単語列頻度の付与

統合モデルに、bi-gram 確率を付与するために、各単語対  $(w_i, w_j)$  の頻度  $C(w_i, w_j)$  と単語頻度  $C(w_j)$  を与える<sup>4</sup>。頻度  $C$  は、元の bi-gram モデルの単語列頻度  $C_0(w_i, w_j)$ ,  $C_0(w_j)$  を利用して獲得する (図 2)。

#### N-gram 内単語列の頻度

元の単語列の、bi-gram および uni-gram 頻度をそのまま使う。

$$C(w_i, w_j) = C_0(w_i, w_j) \quad w_i, w_j \in W_U$$

$$C(w_j) = C_0(w_j) \quad w_j \in W_U$$

#### ネットワーク文法内の単語列頻度

ネットワーク文法の有向弧の存在する単語対  $(@w_i, @w_j) \in A$  に対し、対応する N-gram モデル中の単語対  $(w_i, w_j)$  の頻度を与える。

$$C(@w_i, @w_j) = C_0(w_i, w_j) \quad (@w_i, @w_j) \in A$$

ただし、 $C_0(w_i, w_j) = 0$  となるような単語対  $(@w_i, @w_j)$  が存在する可能性があるため、ディスカウントを行う必要がある<sup>5</sup>。一方、開始単語  $@w_s$  を除いて単語頻度を 0 とする。

$$C(@w_i) = 0 \quad @w_i \in W_{G_f} \cup W_{G_m}$$

<sup>2</sup> 例えば、日本語ディクテーション基本ソフトウェア [8] の言語モデルは、2 万語と 6 万語のものがあり、特殊な固有名詞を使わない限り、ネットワーク文法を記述するには十分だと考えられる。もし N-gram 語彙に含まれない単語を使用する場合は未知語に対応づけることも可能である。また、N-gram の語彙を利用するのは、評価のために単語の単位を合わせるためでもある。

<sup>3</sup> 開始単語と終了単語のために、他の単語と区別する特別な単語記号を用意すればよい。

<sup>4</sup> 普通、N-gram 確率学習では、高次の単語列頻度と低次の単語列頻度が無矛盾なので、高次の単語列頻度だけを与える。一方、提案法での N-gram 確率学習では、各長さ毎に頻度を用意することに注意されたい。

<sup>5</sup> 利用するディスカウント方法は種々考えられるが、本稿では単純なラプラス法 (全ての頻度に 1 を加える) を用いた。

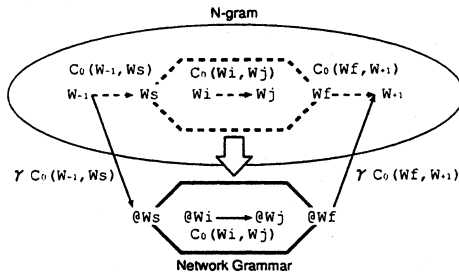


図 2: 統合モデルへの単語列頻度の付与

### N-gram とネットワーク文法を結び単語列頻度

ネットワーク文法の開始単語  $@w_s \in W_{G_s}$  に対し、対応する単語  $w_s$  の頻度を用いて、次のような頻度を与える。

$$C(w_i, @w_s) = \gamma C_0(w_i, w_s)$$

$$C(@w_s) = \gamma C_0(w_s)$$

$\gamma$  は、ネットワーク文法の表す単語列の、対応する N-gram モデルでの単語列に対する相対的な優先度を表す。1 以上の値を与え、大きな値を与えるほど、ネットワーク文法の表現が優先的に考慮される。後で述べるように、 $\gamma = 1$  としても、ネットワーク文法の表現は優先的に扱われるようになる。

ネットワーク文法の終了単語  $@w_f \in W_{G_f}$  から N-gram への単語列頻度も、同様にする。ただし、 $@w_f$  の単語列頻度は 0 とする。

$$C(@w_f w_j) = \gamma C_0(w_f w_j)$$

$$C(@w_f) = 0$$

### 3.3 モデルの学習

与えた頻度からモデルを学習する<sup>6</sup>。その際、予測する単語によってスムージング方法を切り替えて計算する(図 3)。一般に、バックオフスムージングが行われた bi-gram は、次の式で表される。

$$P(w_j|w_i) = \begin{cases} d_C(w_i, w_j) P_{ML}(w_j|w_i) & C(w_i, w_j) > 0 \\ \alpha(w_i) P(w_j) & C(w_i, w_j) = 0 \end{cases}$$

ここで、 $d_C(w_i, w_j)$ 、 $\alpha(w_i)$ 、 $P_{ML}(w_j|w_i)$  は、それぞれディスカウント係数、uni-gram への配分を正規化する関数、最尤推定による確率を表す。

まず、統合モデルの語彙  $W_A$  のうち、開始単語を除くネットワーク文法中の単語  $@w_j \in W_{G_m} \cup W_{G_f}$  を予測する確率  $P(@w_j|@w_i)$  を計算する。この時、コンテキストを構成する単語は  $@w_i \in W_{G_s} \cup W_{G_m}$  である。このモデルは、バックオフを行わないで求める。すなわち、bi-gram 頻度  $C(@w_i, @w_j)$  が存在するものに対してのみ bi-gram 確率を割り当て、 $\alpha(@w_i) = 0$  とする。 $W_{G_m} \cup W_{G_f}$  に再配分される uni-gram 確率は 0 となる。

次に、N-gram 内の単語とネットワーク開始単語  $w_j \in W_U \cup W_{G_s}$  を予測する確率  $P(w_j|w_i)$  を計算する。コンテキストを構成する単語は、 $w_i \in W_U \cup W_{G_f}$  である。このモデルは、普通にバックオフスムージング

<sup>6</sup> 提案法では、bi-gram 頻度と uni-gram 頻度を個別に与えており、その間に整合性がないので、確率計算において uni-gram 頻度  $C(w_j)$  と bi-gram のコンテキスト頻度  $C_c(w_i) = \sum_{w_j} C(w_i, w_j)$  を区別して用いる。例えば、 $P_{ML}$  は、bi-gram 頻度とそのコンテキスト頻度を用いて、 $P_{ML}(w_j|w_i) = C(w_i, w_j)/C_c(w_i)$  と計算する。

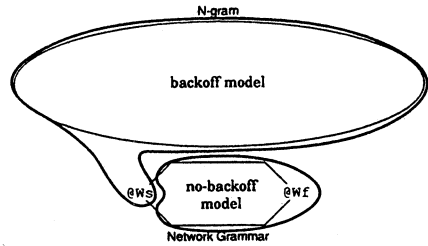


図 3: 統合モデルの学習

ングを行なって求める。その際、 $W_{G_m} \cup W_{G_f}$  にはすでに uni-gram 確率 0 が割り当てられていることに注意し、残りの単語集合  $W_U \cup W_{G_s}$  だけを使って再配分を行なう。

### 3.4 モデルの性質

以上の方法で作成した統合言語モデルは、例えば ARPA 形式などの、従来の N-gram モデルと同じフォーマットで表現できる。そのため、N-gram を言語モデルとして利用する既存の音声認識デコーダでそのまま利用できる。そして以下に示すように、N-gram とネットワーク文法の両方の性質を併せ持ったモデルとして利用できる。

- ・N-gram からネットワーク文法内部の単語を予測する確率は必ず 0 となる。開始単語を除く Network 内単語  $@w_j \in W_{G_m} \cup W_{G_f}$  の予測モデルは、bi-gram のみが用いられ、uni-gram は 0 となる。Network 内単語で、かつ bi-gram 確率の存在する(有向弧の存在する)単語からのみ予測可能となる。N-gram 内単語  $w_i \in W_U$  からの N-gram 確率は、

$$P(@w_j|w_i) = \alpha(w_i) P(@w_j) = \alpha(w_i) \cdot 0 = 0$$

となり、 $w_i$  から  $@w_j$  への遷移は生じない。

- ・ネットワーク文法内部から N-gram 単語を予測する確率は必ず 0 となる。終了単語を除くネットワーク内単語  $@w_i \in W_{G_s} \cup W_{G_m}$  に対し、 $\alpha(@w_i) = 0$  となる。ネットワーク内単語  $@w_i$  から N-gram 内単語  $w_j$  への頻度  $C(@w_i, w_j)$  は必ず 0 なので、

$$P(w_j|@w_i) = \alpha(@w_i) P(w_j) = 0 \cdot P(w_j) = 0$$

したがって、 $@w_i$  から  $w_j$  への遷移は生じない。

- ・ネットワーク文法の開始単語/終了単語では N-gram との接続が可能。N-gram 内単語からネットワーク開始単語、およびネットワーク終了単語から N-gram 内単語は、通常の N-gram モデルと同様のスムージングされた確率値が割り当てられる。したがって、全ての N-gram 内単語からネットワーク文法開始単語への遷移、文法終了単語から全ての N-gram 内単語への遷移が可能である。

- ・N-gram 内の単語列だけから成る文に割り当てられる確率値の順序関係は保存される。N-gram 内単語列に与える頻度は、元の N-gram モデル学習用の頻度と同じである。

- ・同じ単語列では、ネットワーク文法を通る単語列が優先される。ネットワーク文法内単語列を含む文には、それに対応する(同じ音素列を持つ) N-gram 内単語だけで構成された文が必ず存在する。既存の認識デコーダは、文の確率をパスの最大確率で近似する方法(ビタビ・アルゴリズム)が普通であるので、認識時には両者の確率値を比較し、高い方が採用されることになる。両者の bi-gram 確率は同じ頻度か

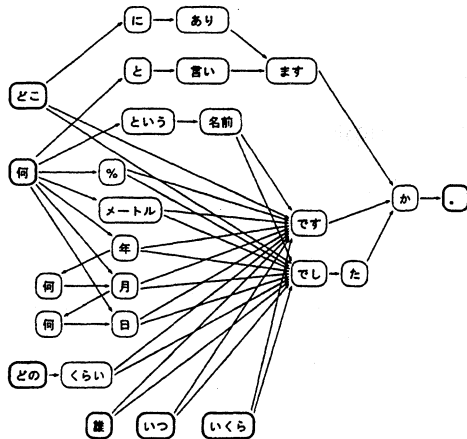


図 4: QA タスク定型表現の文法

ら学習されているが、ネットワーク文法内単語列は、接続しない単語 (特にネットワーク外の単語) を予測するための確率配分がないこと、バックオフを行わないで学習したモデルであること、から相対的に高い確率値が割り当てられることになる。また、ネットワーク文法開始単語への単語列頻度を  $\gamma$  で調節することで、N-gram に対する優先性をコントロールすることができる。

#### 4 実験

QA タスクの質問文を想定した定型表現を受理可能なネットワーク文法 (図 4) を作成し、新聞記事 111 か月分から学習した 2 万語 bi-gram および tri-gram と統合、ネットワーク文法統合モデル (net) を作成した。今回は、既存の N-gram モデル作成ツール [3] を用いて、近似的に作成した<sup>7</sup>。  $\gamma$  は 2 とした。また比較のため、新聞記事のみから学習した N-gram モデル (base) を作成した。スムージング手法は、共に Witten-Bell 法を用いた。

評価データには、新聞記事 100 文 (NP) と QA タスク用質問文 50 文 [7] (QA) を、男性 2 人女性 2 人によって読み上げた音声データを用いた。作成したネットワーク文法は、29 単語と比較的小規模のものであるが、質問文のうち 72% の 36 文 (QA') が、この文法のモデル化する表現を含んでいた。

デコーダには大語彙音声認識デコーダ julius [8] のバージョン 3.2 を使用し、音響モデルには 2000 状態 16 混合性別非依存 triphone を、言語モデル重みは新聞記事 N-gram での最適値を用いた。探索アルゴリズムの変更は行っていない<sup>8</sup>。

<sup>7</sup> 以下の手順で作成した。(1) palmkit [3] の idngram2lm を修正。頻度読み込み部を修正して、N-gram 頻度を長さ N 毎に別々に読み込めるように変更。同時に、各頻度のコンテキスト頻度を求め、確率計算に使用するように変更。(2) 単語集合を  $W_U \cup W_{G_s} \cup W_{G_f}$  と  $W_{G_s} \cup W_{G_m} \cup W_{G_f}$  に分割して、前者からスムージングモデルの ARPA 形式を、後者からスムージングしないモデルの ARPA 形式を作成。(3) 結合部の確率値に注意しながら両モデルをマージ。(4) 言語モデル評価ツール (evallm) を用いて、正しいモデルとなっていることを確認。

<sup>8</sup> ネットワーク文法内部の単語と N-gram モデル内の単語の間での N-gram 確率は必ず 0 となり、この間の単語間遷移は起こり得ない。この単語間遷移の探索を抑制するようにデコーダを修正することで、より効率的な探索を行うことができるようになると思われる。

表 1: 実験結果

n-gram	評価データ	言語モデル	COR	ACC
2	NP (100)	base	81.9	79.4
		net	81.9	79.4
	QA (50)	base	72.9	69.4
		net	74.2	70.4
	QA' (36)	base	71.3	67.2
		net	73.8	69.6
3	NP (100)	base	90.4	87.8
		net	90.0	87.8
	QA (50)	base	85.0	80.3
		net	85.1	80.5
	QA' (36)	base	84.4	79.9
		net	85.2	81.0

COR=単語正解率 (%), ACC=単語正解精度 (%)

実験結果を表 1 に示す。新聞記事の認識精度を下げることも無く、質問文の精度が向上することを確認した。bi-gram にくらべ、tri-gram の場合にはあまり効果が現れなかった。既存の実装方法との整合性に問題が生じている可能性も考えられ、今後調査が必要である。また、今回の提案法の tri-gram との統合モデルでは、ネットワーク文法部は bi-gram のままであったが、tri-gram を用いることも理論上可能であり、今後検討していきたい。

#### 5 まとめ

音声入力による質問応答システムのために、定型的な質問文表現と多様な検索対象に関する言語表現を、同時にモデル化した言語モデルを提案した。提案法は、N-gram 言語モデル表現を利用しつつ、その内部に記述文法で表現した二値的な制約を統合する手法である。N-gram では不可能であった単語間の長距離の依存関係も、記述文法を使って表現可能である。提案法で作成した言語モデルは、従来の N-gram モデルと互換性があり、既存の大語彙音声認識デコーダでそのまま利用できるという特徴を持つ。提案法により、新聞記事から作成した言語モデルと、質問文を表現したネットワーク文法を統合した言語モデルを作成し、既存の音声認識デコーダで認識実験を行った。元の汎用 N-gram 言語モデルが対象とする新聞記事の読み上げに対する認識精度を下げることもなく、記述文法で扱うタスク特有の質問文に対し認識精度が向上することを確認した。

#### 参考文献

- [1] T. Akiba and K. Itou. Semi-Automatic Language Model Acquisition without Large Corpora. In Proc. of ICSLP-2000, vol.4, pp.49-52, 2000.
- [2] F.C.N.Pereira and R.R.Wright. Finite-state approximation of phrase-structure grammars. In Proc. of ACL 1991, pp.246-255, 1991.
- [3] 伊藤彰則, 好田正紀. 単語およびクラス n-gram 作成のためのツールキット. 信学技報, SP2000-106, pp.67-72, 2000.
- [4] 伊藤克互, 秋葉友良, 藤井敦, 石川徹也. 音声入力型テキスト検索システムのための音声認識. 日本音響学会講演論文集, pp.193-194, Oct. 2001.
- [5] 鹿島博品, 河原達也. 複合的言語制約に基づくキーフレーズスポッティングによる対話音声理解. 信学技報, SP2000-106, pp.115-120, 2000.
- [6] 駒谷和範, 河原達也, 清田陽司, 黒橋禎夫, Pascale Fung. 柔軟な言語モデルとマッチングを用いた音声によるレストラ探索システム. 信学技報, SP2001-113, pp.67-72, 2001.
- [7] 佐々木 裕, 磯崎 秀樹, 平博順, 廣田 啓一, 賀沢秀人, 平尾努, 中島浩之, 加藤恒昭. 質問応答システムの比較と評価. 信学技報, NLC2000-24, pp.17-24, 2000.
- [8] 鹿野清宏, 伊藤克互, 河原達也, 武田一哉, 山本幹雄 (編). 音声認識システム. オーム社, 2001.