

# 述語項構造を用いた文変換とフィルタリングに基づく 音声対話用言語モデル

吉野 幸一郎      森 信介      河原 達也

京都大学 情報学研究科

## 概要

情報検索型音声対話システムのための音声認識用言語モデルを、係り受け構造と述語項構造を用いた文変換と、述語項構造を用いたフィルタリングによって構築する手法について述べる。提案手法では、まず検索対象となる書き言葉の文書に対して、係り受け構造・述語項構造と述語の単純な変換ルールを用いて、質問文への文変換を行う。さらに、述語項構造に着目した情報構造の抽出を行い、ドメインにおける重要な情報構造のパターン（述語項構造テンプレート）を自動で定義する。このテンプレートを用いて文選択を行うことで、タスク・ドメインに適合した言語モデルを構築する。

この手法により構築した言語モデルと、典型的な従来手法である新聞記事とQAサイトの質問文を混合したモデルと比較して音声認識実験を行った結果、最大16.9%のWER（単語誤り率）削減と、18.0%のパープレキシティ削減を実現した。文選択のみでも、7.8%のWER削減と5.2%のパープレキシティ削減を実現し、述語項構造テンプレートによるフィルタリングの効果を確認した。

## 1 はじめに

Web上に集積する情報は日々爆発的に増加しており、それらに対する情報アクセスを行うインタフェースの登場が望まれている。こうした情報へのアクセスはキーワード型検索が主であるが、キーワード型検索ではユーザの漠然とした情報要求、自然文での検索に対する要求などに応えることが難しい。そこで我々は、ユーザの意図・嗜好を対話的に顕在化しながら情報を提示するインタフェースとして、雑談形式による情報検索型音声対話システムの研究を行なっている[1, 2]。

音声対話システムのための音声認識には、タスクやドメイン、そして話し言葉のスタイルに適合した言語モデルが不可欠である。これまで研究されてきた音声対話システム用言語モデルの構築方法は、検索対象の文書からなる書き言葉コーパスに話し言葉調のコーパスを混合するもの[3]と、Webから話し言葉調のテキストやn-gramを収集し、これを追加するという枠組み[4, 5]が主流であった。これに対して想定される質

問文を、知識ベースや書き言葉形式のコーパスから生成し、言語モデルを構築しようというアプローチがある[6]。このアプローチは、適切に想定質問文を生成することができれば、高い音声認識精度の実現が期待できる。

また、情報検索型の対話システムにおいて重要な情報構造はドメインに依存しており[1]、限られた重要な情報構造を利用することが必要である。これまでの研究では、このような情報構造は人手で構築したものを利用することが一般的であったが[7, 8]、Webに存在する様々なドメインに対して自動で適用することが困難である。

そこで本研究では、これらの問題を解決する手法を提案する。第一に、文の係り受け構造と述語項構造を利用して、述語に対する単純な変換規則によって発話形式への文変換を行う。第二に、先行研究[1]で提案した、述語項構造に基づく情報構造の自動定義を用いてこれらの文のフィルタリングを行う。これによって当該ドメインでの対話において有用な情報構造を持つ文を抽出し、音声対話における音声認識精度の向上を目指す。

## 2 手法の概要

提案手法の概要を図1に示す。まず、検索対象の文書からなるコーパス（本研究では新聞記事）に対して、係り受け解析・述語項構造解析を行う。この解析には、JUMAN<sup>1</sup>/KNP<sup>2</sup>を用いる。係り受け構造とは文節間の依存関係を示すものである。述語項構造とは述語、格要素、その格関係から成り立つ意味表現の単位である。これらの解析結果と元の文、さらに単純な変換ルールを用いて、質問文生成を行う。次に述語項構造解析結果を利用して、述語項構造テンプレートの学習を行う。最後に、変換された質問文コーパスに対して述語項構造テンプレートを用いたフィルタリングを行う。この処理によって、情報検索型音声対話システムに必要な十分な語彙を含む言語モデル用学習コーパスを構築する。

<sup>1</sup><http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html>  
(2012年1月現在)

<sup>2</sup><http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html>  
(2012年1月現在)

表 1: 変換ルール

Predicate type	POS	Rules
用言	動詞&動詞性接尾辞 イ形容詞 (形容詞) & イ形容詞性接尾辞 ナ形容詞 (形容動詞) & ナ形容詞性接尾辞	連用形 + (/ます + か   まし + たか/) 基本形 + (/です + か   でし + たか/) 語幹 + (/です + か   でし + たか/)
事態性名詞	一般名詞&名詞性接尾辞 サ変名詞&サ変名詞性接尾辞	原形 + (/です + か   でし + たか/) 原形 + (/し + ます + か   し + まし + たか/)

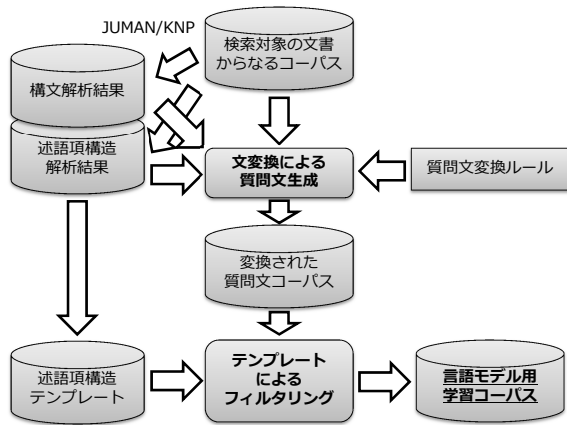


図 1: 提案手法の全体像

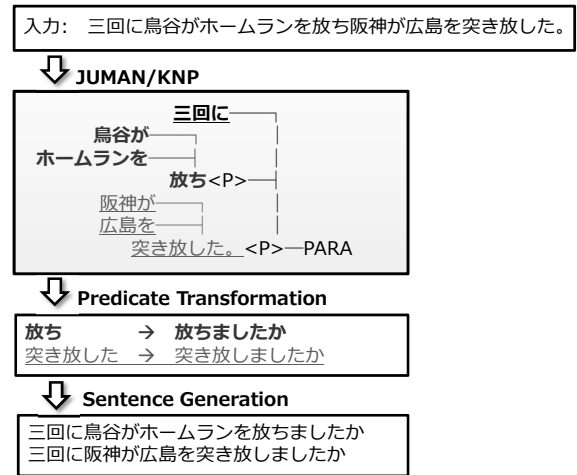


図 2: 文変換の手順

### 3 文構造を利用した文変換

まず、構文解析結果と述語項構造解析結果を用いた文変換について述べる。述べる。その概要を図 2 に示す。

#### 3.1 ルールによる述語の変換

本手法では、まず述語項構造における述語を特定する。述語は用言と事態性名詞に分類できる。文中の事態を指しうる表現として用言（動詞、形容詞）の他に名詞があることが知られており、これを事態性名詞と呼ぶ [9, 10]。それらに対し、表 1 に示すルールを適用し、質問文への変換を行う。この変換を現在形・過去形でランダムに行う。JUMAN/KNP の品詞大分類は益岡・田窪文法 [11] に準じ、用言には動詞、イ形容詞（形容詞）、ナ形容詞（形容動詞）が挙げられる。また、事態性名詞として取りうるものは、名詞細分類として一般名詞とサ変名詞がある。加えて述語になりうる品詞として接尾辞があるが、これらの扱いは接尾辞細分類で性質として与えられる品詞分類に準じる。

#### 3.2 係り受け木に基づく文の変換

次に、係り受け木を用いて述語に係る文節を選択する。係り受け木上で 3.1 節で特定された述語に対して係る文節を列挙し、その語順を保持したまま述語とあわせて出力する。図 2 の例では、「放ち」の枝となる文節は「鳥谷が」「ホームランを」「三回に」なのでこれらをあわせて「三回に鳥谷がホームランを放ちましたか」という文が生成される。この例では、「三回

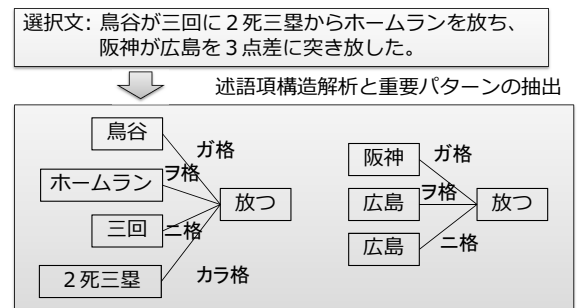


図 3: 述語項構造の例

に」は「放ち」「突き放した」の双方にかかるため、いずれの文変換にも利用される。

### 4 述語項構造テンプレートによるフィルタリング

音声対話システムの対象ドメインにおける情報構造を定義するため、述語項構造を利用する。述語項構造とは、図 3 に示されるような述語、要素、その格関係から成る意味表現の単位である。述語と係り受け関係にある要素に対して、ガ格、ヲ格、時間などの意味役割を付与している。

#### 4.1 ドメイン依存述語項構造テンプレートの抽出

パーザによって自動で抽出された全ての述語項構造が情報検索対話において重要なわけではなく、重要な

パターンはドメインに依存している。例えば、野球のドメインにおいては「A (ガ格), B (ニ格) - 勝つ」、  
「A (ガ格), B (ヲ格) - 打つ」などのパターンが、経済のドメインにおいては「A (ガ格), B (ヲ格) - 買う」「A (ガ格), B (ニ格) - 買収」などのパターンが重要である。

我々は、このようにドメインに依存した重要パターンを自動で抽出する手法を先行研究 [1] において提案しており、本研究でもその手法を用いる。先行研究においては Naive Bayes 法による述語項構造のパターン (述語項構造テンプレートと呼ぶ) 抽出手法が有効であることが示されており、これをフィルタリングに用いる。この手法では、単語  $w_i$  が出現した場合にそのドメインが  $t$  である確率を以下のように定義する。

$$P(t|w_i) = \frac{C(w_i, t) + D_t \gamma}{C(w_i) + \gamma}. \quad (1)$$

ここで  $\gamma$  はディリクレ過程 (Chinese Restaurant Process) を用いたサンプリング [12] によって推定されたスムージング係数である。 $D_t$  はドメイン  $t$  のコーパスサイズに対する正規化係数で、以下の式で与えられる。

$$D_t = \frac{\sum_j C(w_j, t)}{\sum_k C(w_k)}. \quad (2)$$

述語項構造テンプレートに対する評価スコアは、述語項構造中の述語 (p)、要素 (a)、意味役割 (s) について計算された  $P(t|w_i)$  の相乗平均により求める。その際、意味役割を述語と結合するか、要素と結合するかによって2つの評価尺度を提案する。すなわち、述語と意味役割を結合して1つの語彙として扱うものを ps とし、要素と意味役割を結合して1つの語彙として扱うものを sa とする。これらについて式 (1) の確率値を求めたものを利用し、評価スコア  $NB$  を以下のように定義する。

$$\begin{cases} NB(t|w_{ps-a,i}) = (P(t|w_{ps,i}) \times P(t|w_{a,i}))^{\frac{1}{2}} \\ NB(t|w_{p-sa,i}) = (P(t|w_{p,i}) \times P(t|w_{sa,i}))^{\frac{1}{2}} \end{cases} \quad (3)$$

統計的手法においては、特に人名などの固有表現でデータスパースネスの問題が生じ、学習データとテストデータにおけるミスマッチが固有表現に多く現れる。そこで、学習コーパス中に出現する固有表現に対してクラス化を行う。今回は人名、地名、組織名の3つの固有表現をクラス化する。このクラス化は以下の式で行われる。

$$P(t|N_j) = \sum_i P(t|w_i(w_i \in N_j))P(w_i|N_j). \quad (4)$$

ただし、この式にそのまま式 (1) の定義を当てはめると、スムージング項がそれぞれ  $\sum_i \gamma$  と  $\sum_i D_t \gamma$  になってしまう。そこで式 (1) からスムージング項を取り除き、

$$P'(t|w_i) = \frac{C(w_i, t)}{C(w_i)} \quad (5)$$

とした上でクラス化を行い、その後スムージングを行う。

$$P'(t|N_j) = \sum_i P'(t|w_i(w_i \in N_j))P(w_i|N_j) \quad (6)$$

$$= \frac{\sum_i C(w_i, t)}{\sum_j C(w_j)} \quad (7)$$

$$\approx \frac{\sum_i C(w_i, t) + D_{t,N_j} \cdot \gamma_{N_j}}{\sum_j C(w_j) + \gamma_{N_j}}. \quad (8)$$

このクラスにおける Naive Bayes の確率値を計算した後、式 (3) で定義される評価スコア  $NB$  を算出する。文全体の評価スコアは、以下の式で与える。

$$NB_s = \frac{\sum_{i=1}^n NB(i)}{n} \quad (9)$$

ここで  $i$  とは文  $s$  に含まれる述語項構造テンプレートであり、 $NB(i)$  はテンプレート  $i$  の評価スコアである。

## 4.2 テンプレートを用いたフィルタリング

4.1 節で定義された文の評価スコア ( $NB_s$ ) を使い、3章の手法で生成された質問文のフィルタリングを行う。 $NB_s$  を利用して質問文のソートを行い、言語モデル構築に利用する質問文をスコア上位から 1/1、1/2、1/4、1/8、1/16 と選択していくことによってフィルタリングの効果を確認する。

## 5 音声認識実験による評価

提案手法を音声認識実験によって評価した。評価セットとして [1] で作成された日本プロ野球ドメインの質問文 201 問を利用した。また、学習セットとして毎日新聞記事データベース (CD-毎日新聞データ集 2000 - 2009) のプロ野球関連記事 176,852 文を用いた。これらの文中から用言 342,322 個と事態性名詞 158,201 個が抽出され、それらを変換して生成した質問文 500,523 文を学習に利用した。

本研究では評価指標として単語誤り率 (WER) とパープレキシティを用いた。パープレキシティについては、質問文 500,523 文全てを学習に利用した場合 (カットオフ 5) の語彙 16,239 単語を認識対象語彙とした補正パープレキシティを用いた。

これらと比較する手法として、先述の新聞記事 176,852 文と、Yahoo! 知恵袋データ (第 2 版)<sup>3</sup> から抽出した野球カテゴリの質問文 481,243 文から構築した言語モデルを用意した。

### 5.1 実験結果

WER のグラフを図 4、パープレキシティのグラフを図 5 に示す。横軸は利用した学習テキストの量である。この結果、WER においては質問文のサイズを 1/16 した場合に、従来手法と比べて 16.9% の削減となった。また、述語項構造テンプレートを用いたフィ

<sup>3</sup>本研究の実施にあたり、ヤフー株式会社が国立情報学研究所に提供した「Yahoo! 知恵袋データ (第 2 版)」を利用した。

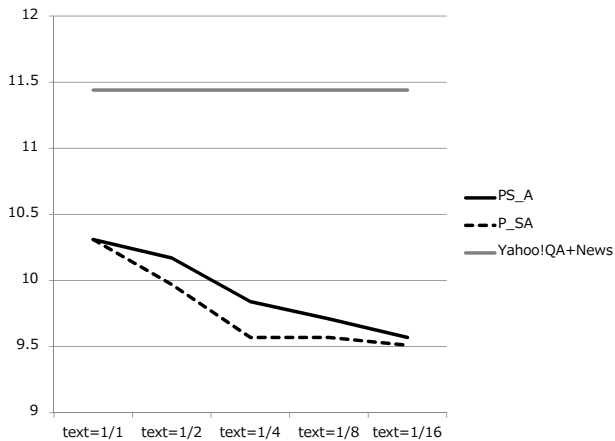


図 4: 単語誤り率 (WER)

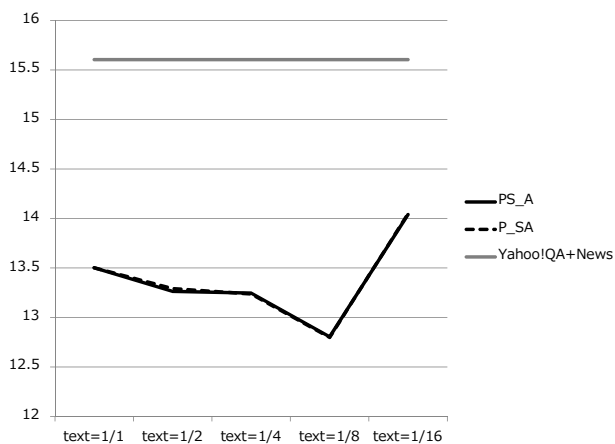


図 5: パープレキシティ

ルタリングを行わない場合 (text=1/1) と比較しても 7.8% の削減率が見られ、フィルタリングが有効であることが示された。パープレキシティにおいてもコーパスサイズを 1/8 とした場合に、従来手法と比較して 18.0%、フィルタリングを行わない場合と比較して 5.2% の削減が見られる。これらにより、本手法が音声対話用の言語モデルを構築する上で有効な手法であることが確認できた。

## 6 まとめ

文における係り受け構造と述語項構造に着目した文変換と、述語項構造テンプレートを利用したフィルタリングを用いた音声対話用言語モデルの構築手法を提案した。また、この手法が典型的な従来手法と比較して有効であることを示した。これにより、述語項構造を用いたドメイン依存の重要な情報構造の抽出が、音声対話において有効であることが示された。今後の課題としては、複数のドメインに対して大規模に本手法を適用すること、最適なフィルタリングのしきい値の設定法を検討することが挙げられる。

## 参考文献

- [1] 吉野幸一郎, 森信介, 河原達也. 述語項の類似度に基づく情報抽出・推薦を行う音声対話システム. 情報処理, Vol. 52, No. 12, pp. 3386-3397, 2011.
- [2] Tatsuya Kawahara. New perspectives on spoken language understanding: Does machine need to fully understand speech? In *Proc. IEEE-ASRU*, pp. 46-50, 2009.
- [3] 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ン検索システム. 情報処理学会研究報告, SLP-39-30, pp. 67-72, 2001.
- [4] 翠輝久, 河原達也. ドメインとスタイルを考慮した Web テキストの選択による音声対話システム用言語モデルの構築. 電子情報通信, Vol. J90-D, No. 11, pp. 3024-3032, 2007.
- [5] Ivan Bulyko, Mari Ostendorf, Manhung Siu, Tim Ng, Andreas Stolcke, and Özgür Çetin. Web resources for language modeling in conversational speech recognition. *ACM Trans. Speech Lang. Process.*, Vol. 5, No. 1, pp. 1:1-1:25, 2007.
- [6] Varga Istvan, 大竹清敬, 鳥澤健太郎, D.Saeger Stijn, 松田繁樹, 林輝昭. WWW テキストのみを用いたオープンドメイン質問応答用音声認識言語モデル. 言語処理学会第 17 回年次大会, NLP-D1-2, 2011.
- [7] Ralph Grishman. Discovery methods for information extraction. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 243-247, 2003.
- [8] L.Ramshaw and R.M.Weischedel. Information extraction. In *IEEE-ICASSP*, Vol. 5, pp. 969-972, 2005.
- [9] Jane Grimshaw. *Argument structure*. MIT Press, 1990.
- [10] 小町守, 飯田龍, 乾健太郎, 松本裕治. 名詞句の語彙統語パターンを用いた事態性名詞の項構造解析. 自然言語処理, Vol. 17, No. 1, pp. 141-159, 2010.
- [11] 益岡隆志, 田窪行則. 基礎日本語文法. くろしお出版, 1992.
- [12] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, Vol. 101, pp. 1566-1581, 2006.