2007－SLP－67 (1)
2007／7／20

フィラーの書き起こしのないコーパスからのフィラー付き言語モデルの構築

太田 健吾† 土屋 雅稔‡ 中川 聖一†
豊橋技術科学大学 情報工学系† / 情報メディア基盤センター‡
〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘1-1
kohta@slp.ics.tut.ac.jp, tsuchiya@imc.tut.ac.jp, nakagawa@slp.ics.tut.ac.jp

要旨

本稿では，フィラーを含まないコーパスから，フィラー予測モデルを利用してフィラー付き言語モデルを作成する方法を提案する．フィラー予測モデルは，周辺コンテキストを用いてフィラー挿入個所を推定するフィラー挿入モデルと，推定された箇所に挿入すべきフィラーを周辺のコンテキストに基づいて予測するフィラー選択モデルの2つのモデルからなる．日本語話し言葉コーパスと国会会議録に対する評価実験の結果，本提案手法は，フィラーを含む正確な話し言葉コーパスから作成した3-gramモデルにきわめて近い言語モデルを再現できることを示す．

キーワード 話し言葉，フィラー，言語モデル，講演音声，国会会議録

Construction of Language Model with Fillers from Corpus without Fillers

Kengo Ohta, Masatoshi Tsuchiya, Seiichi Nakagawa
Department of Information and Computer Sciences Information Media Center ,
Toyohashi University of Technology
1-1, Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, Japan
kohta@slp.ics.tut.ac.jp, tsuchiya@imc.tut.ac.jp, nakagawa@slp.ics.tut.ac.jp

Abstract

This paper proposes a novel method to construct a spoken language model with fillers using a filler prediction model from a corpus without fillers. The filler prediction model consists of two models: a filler insertion model which predicts places where fillers should be inserted, and a filler selection model which predicts appropriate fillers for given places. The experiments against the corpus of spontaneous Japanese and Japanese National Diet Record show that language models constructed by the proposed method achieve quite near performance of the traditional 3-gram language model constructed from the exact spontaneous speech corpus including fillers.

Keywords Spoken Language, Filler, Language Model, Lecture Speech, Japanese National Diet Record

1 はじめに

近年，講義音声の自動要約やニュース音声のインデキシングなどの技術に対する需要が高まってきている [7][10]．これらを実現するには，対象となる音声の発話スタイルとドメインが一致し，かつ，フィラーなどの話し言葉特有の現象にも対応した言語モデルを備えた大語彙音声認識器が必要である．そのような言語モデルを構築する最も単純な方法は，対象とする音声と同一ドメインの大規模な話し言葉コーパスから，言語モデルを構築するという方法である．しかし，そのようなコーパスを整備する作業は極めて高コストであり，あらゆるドメインに対して，条件を満たすコーパスを入手できると仮定することは非現実的である．

このような状況に対処するため，対象とする音声とは異なるドメインの話し言葉言語モデルと，対象とする音声と同一のドメインの書き言葉言語モデルを組み合わせる手法が提案されている [2]．また，秋田・河原ら [1] は，まったく同一の内容を対象とした書き言葉と話し言葉のパラレルコーパスから，書き言葉を話し言葉に変換する統計的なモデルを学習

し，書き言葉言語モデルを話し言葉言語モデルに変換する手法を提案している．

それに対して，書き言葉コーパスと話し言葉コーパスの中間的なコーパスとして，フィラーなどの話し言葉特有の現象が省略されている**不正確な話し言葉コーパス**に注目する．このようなコーパスは，議事録や速記録の形で広く作成されており，話し言葉特有の現象も正確に書き起されている話し言葉コーパスに比べて，比較的容易に入手可能である．例えば，国立国会図書館は，1947年以後の全ての国会の会議録を公開している[1]．このようなコーパスは，書き言葉コーパスよりも話し言葉に近いコーパスと考えられるので，話し言葉特有の現象に対応した言語モデルを作成するという目的には，書き言葉コーパスよりも適していると期待される．ただし，不正確な話し言葉コーパスは，フィラーや言い淀み，言い直しなどの話し言葉特有の現象のほとんどが省略されているため，言語モデルの学習を行う前にそれらを復元する必要がある．

本稿では，話し言葉特有の現象の中でも最も発生頻度の高い現象であるフィラーに注目し[6]，対象とする音声とは異なるドメインの正確な話し言葉コーパスからフィラー予測モデルを学習し，この予測モデルに基づいて，対象とする音声と同一のドメインの不正確な話し言葉コーパスに対してフィラーの復元を行い，フィラーが復元されたコーパスから言語モデルを学習するという手法を提案する．

2 フィラー予測モデル

2.1 フィラー予測モデルの定式化

フィラーを含まないコーパスから，フィラーに対応した言語モデルの作成方法を考える．例として，文(1)のようなフィラーを含まない文から，フィラーに対応した言語モデルを作成する方法を考える．

(1) この画面を見ると・・・

この場合，2つの方法が考えられる．第1の方法は，文(1)からフィラーを含まない言語モデルを学習しておき，その言語モデルをフィラーに対応した言語モデルに変換するという方法である．第2の方法は，文(1)中の適切な個所にフィラーを挿入して，文(2)のようなフィラーを含む文を作成し，その文からフィラーに対応した言語モデルを学習するという方法である．

(2) この画面をえーと見ると・・・

第1の方法では，対象とする言語モデルに対応した変換規則または変換モデルが必要となり，別種の言語モデルを利用するためには，変換規則または変換モデルを作成し直す必要がある．それに対して，第2の方法では，フィラーの挿入個所と種類を予測するモデルが必要になるが，そのようなモデルさえ得られれば，言語モデルの変更には容易に対応可能である．そこで本稿では，第2の方法によるフィラーを含む言語モデルの構築方法を提案する．

文(1)を文(2)のように書き換えるために，フィラーの挿入個所と種類を予測するモデルを**フィラー予測モデル**と呼ぶ．実際の正確な話し言葉コーパス[4]を対象とする分析から，フィラーには多様な派生形が存在することが分かっており，フィラーの挿入箇所と種類を同時にモデル化すると，データスパースネスが生じる恐れがある．そこで，本稿では，フィラーの挿入箇所と種類は独立に推定できるという仮定をおく．すなわち，フィラーを挿入する箇所を推定する**フィラー挿入モデル**と，推定された箇所に挿入するべき適当なフィラーを選択する**フィラー選択モデル**，という2つのモデルの組み合わせとしてフィラー予測モデルを定式化する．

2.2 フィラー挿入モデル

フィラー挿入モデルとは，ある形態素列が与えられた時に，その形態素列中においてフィラーを挿入すべき箇所を推定するモデルである．本稿では，このモデルを，形態素列を対象とし，個々の形態素に対して，その形態素の直後にフィラーを挿入するべきかどうかというラベルを付与するという，系列ラベリング問題として定式化する．例えば，文(1)を文(2)に変換する場合には，最初に文(1)を形態素列に分解し，図1のように個々の形態素に対して，直後にフィラーを挿入すべきである場合にはラベルFを付与し，フィラーを挿入すべきではない場合にはラベルOを付与する．

形態素列	この 画面 を 見る と ・・・
	(文頭) 連体詞 名詞 助詞 動詞 助詞
ラベル列	O O O F O O ・・・

図1: フィラー挿入モデル

本稿では，このような問題を解くフィラー挿入モデルを，Conditional Random Field(CRF)[3]を用いて作成する．CRFは，隠れマルコフモデルなどのモデルと比べて柔軟な素性設計が可能であり，また，比較的少量の学習データでも良い性能を示すことが知られている識別モデルである．

CRFでは，形態素列Xに対するラベル列Yの条件付き確率$P(Y|X)$を，次式のように表す．

$$P(Y|X) = \frac{1}{Z(X)} \exp\left(\sum_i^n \sum_a \lambda_a f_a(X_i, Y_i)\right) \quad (1)$$

ここで，f_aは素性関数，λ_aは素性関数に対する重み，$Z(X)$は正規化項である．

[1] http://kokkai.ndl.go.jp/

2.3 フィラー選択モデル

フィラー選択モデルは，適当な形態素列とフィラーの挿入箇所が指定された時に，挿入すべき適当なフィラーを選択するモデルである．本稿では，単純に，周囲の形態素やモーラなどの文脈 h に対してフィラー f が生起する条件付き確率 $P_s(f|h)$ を，フィラー選択モデルとして用いる．条件付き確率 $P_s(f|h)$ は，Witten-Bell スムージングを適用して [5]，次式のように推定する．

$$P_s(f|h) = \begin{cases} \frac{c(h,f)}{c(h)+r(h,f)} & \text{if } c(h,f) > 0 \\ \frac{r(h,f)}{c(h)+r(h,f)} \cdot P_s(f|h') & \text{otherwise} \end{cases}, \quad (2)$$

ただし，$c(h,f)$ はフィラーを含む正確な話し言葉コーパスにおいて文脈 h とフィラー f が同時に生起する頻度，$c(h)$ は文脈 h の生起する頻度，$r(h,f)$ は文脈 h の直後に現れるフィラーの種類の数である．文脈 h' は，文脈 h から条件を 1 つ取り除いた文脈である（バックオフ）．

3 フィラー予測モデルを用いたフィラーつき言語モデルの構築

本稿では，フィラーを含まない不正確な話し言葉コーパスから，フィラーに対応した話し言葉言語モデルを作成する手順として，以下のような手順を提案する．

1. フィラーを含む正確な話し言葉コーパス (以後，**学習コーパス**と呼ぶ) から，フィラー予測モデルを構築．この部分は，更に以下の 2 段階に分けられる．
   (a) フィラー挿入モデルの構築．
   (b) フィラー選択モデルの構築．
2. フィラーを含まない不正確な話し言葉コーパス (以後，**開発コーパス**と呼ぶ) に対してフィラー予測モデルを適用し，フィラーを付与したコーパスを作成．
3. フィラーを付与したコーパスから，言語モデル（トライグラム）を構築．

本節では，この処理の詳細について述べる．

最初に，学習コーパスからフィラー挿入モデルを構築する．学習コーパスに対して，個々の形態素の直後がフィラーであるか否かを表すラベルを付与した上で，フィラーを取り除く．例えば，文 (2) を学習コーパス中の文とすると，図 1 のような学習データが得られる．この学習データに基づいて，形態素列 X に対するラベル列 Y の条件付き確率 $P(Y|X)$ を CRF を用いて求める．CRF の学習用プログラムとしては CRF++[2] を用いた．素性としては，直前 2 形態素，直後 2 形態素，および現在の形態素，それぞれの品詞，直前の 2 モーラの情報などの組合わせを用いる ([9])．

次に，学習コーパスからフィラー選択モデルを構築する．本稿では，単純に，周囲の形態素やモーラなどの文脈 h を条件として，フィラー f が生起する条件付き確率 $P_s(f|h)$ を，フィラー選択モデルとして用いる．この条件付き確率は，学習コーパスから式 (2) に基づいて求められる．ただし，フィラーは，発音上の揺れによる派生形が生じやすい．例えば，日本語話し言葉コーパス (以下，**CSJ** と略記する)[4] には 151 種類のフィラーが出現しているが，これらの多くは，長音・促音の有無や語尾音節の繰り返しなどの発音上の揺れによる派生形である．出現頻度が非常に小さい派生形について信頼できる条件付き確率 $P(f|h)$ を推定することは困難であるため，発音が類似しているフィラーは同一のものと見なして条件付き確率を求める．

次に，ここまでの手順によって得られたフィラー予測モデルを用いて，開発コーパスにフィラーを挿入する．具体的には，開発コーパス中のそれぞれの形態素 $x_i (i = 1, 2, \ldots)$ に対して，以下の処理を行う．

1. 形態素列 X 中のそれぞれの形態素 x_i の直後にフィラーが挿入される確率 $P(y_i = \mathrm{F}|X)$ を次式により求める．
$$P(y_i = \mathrm{F}|X) = \sum_{\{Y|y_i=\mathrm{F}\}} P(Y|X). \quad (3)$$
一様でランダムな確率変数 Q_i(ただし，$0 \leq Q_i \leq 1$) が，$Q_i \leq P(y_i = \mathrm{F}|X)$ を満たすとき，形態素 x_i の直後にフィラーを挿入するため，次のステップに進む．そうでなければ，次の形態素に進む．
2. あるフィラー $f_j (j = 1, 2, \ldots, |F|)$ が次式を満たすとき，そのフィラー f_k を形態素 x_i の直後に挿入する．
$$\sum_{j=1}^{k-1} P_s(f_j|h_i) \leq Q'_i < \sum_{j=1}^{k} P_s(f_j|h_i) \quad (4)$$
ただし，Q'_i は一様でランダムな確率変数 ($0 \leq Q'_i \leq 1$)，h_i は形態素 x_i 周辺の文脈である．

ここで，一様でランダムな確率変数 Q_i, Q'_i は，フィラー生起の不規則性を模倣するために導入している．これにより，まったく同一のコーパスを用いた場合でも，上述の手順によって作成されたコーパス中のフィラーの位置や種類は一定とはならない．そのため，次節以降では，10 回の試行の結果を平均した結果を実験結果として示す．

このようにして得られたフィラーを付与したコーパスから，言語モデルとして形態素 3-gram モデル

[2] http://chasen.org/~taku/software/CRF++/

表 1: 学会講演と模擬講演の比較
(辞書は模擬講演から作成)

テストコーパス	未知語率
模擬講演	0.86 %
学会講演	2.51 %

表 2: 実験データ諸元

	学習コーパス	開発コーパス	テストコーパス
ドメイン	模擬講演	学会講演	学会講演
講演数	1715	937	50
収録時間 (hour)	329.9	258.4	16.0
総文数	498k	363k	22K
総単語数	3,606K	3,109K	170K
語彙数	41K	29K	8K
フィラー発生頻度	175K	174K	11K
フィラー発生率	4.8%	5.6%	6.7%

を構築することは，非常に容易である．なお，実際の実験においては，頻度順に上位 20,000 語の語彙のみを用い，残りの低頻度語は未知語と見なして処理した．

4 日本語話し言葉コーパスを対象とする実験

本節では，CSJ を学習コーパスおよび開発コーパスとして用いた実験結果について述べる．CSJ は，話し言葉特有の現象を含めて正確に書き起された話し言葉コーパスである．テストコーパスとして CSJ の学会講演を用いる場合には，CSJ からフィラーを取り除いたコーパスを開発コーパスとして用いると，会議録や議事録を開発コーパスとして用いる場合よりも理想的な結果が得られると考えられる．

4.1 実験条件

CSJ は，学会講演・模擬講演・対話・朗読という 4 種類の部分コーパスに分けることができる．この内，模擬講演の一部 (1665 講演) から作成した辞書を用いて，模擬講演と学会講演それぞれの 50 講演の未知語率を求めると，表 1 のように大きく異なる結果が得られる．よって，学会講演と模擬講演は，たがいにドメインの異なるコーパスと考えることができる．

そこで，本節の実験では，CSJ の模擬講演を学習コーパスに用い，学会講演を開発コーパスとテストコーパスの 2 つに分割して用いた．それぞれのコーパスの諸元を表 2 に示す．ただし，開発コーパスとして用いる学会講演については，実験前にフィラーを削除しておき，フィラーを含まない不正確な話し言葉コーパスを模擬した．

作成した言語モデルの評価には，テストコーパスに対するテストセットパープレキシティ PP と補正テストセットパープレキシティ PP^* を用いた．補正テストセットパープレキシティ PP^* は，テストコーパス中に出現した未知語率を考慮した尺度であり，テストコーパス中に出現した未知語の延べ頻度を o，異なり数を m，総単語数を n とすると，次式によって定義される [8]．

$$\log_2 PP^* = \log_2 PP + \frac{o}{n}\log_2 m \qquad (5)$$

また，テストセットパープレキシティ PP をフィラー部分のみについて計算した PP_F と，フィラー以外の部分について計算した PP_O も補助的な尺度として用いた．PP_F は，テストセット w_1^n 中でフィラーが n_F 回出現し，それらの集合を F とした場合，次式によって計算される．

$$H_F = -\frac{1}{n_F} log \prod_{w_i \in F} P(w_i|w_{i-2}w_{i-1}), PP_F = 2^{H_F} \qquad (6)$$

同様に，PP_O は，テストセット w_1^n 中でフィラー以外の単語が n_O 回出現し，それらの集合を O とした場合，次式によって計算される．

$$H_O = -\frac{1}{n_O} log \prod_{w_i \in O} P(w_i|w_{i-2}w_{i-1}), PP_O = 2^{H_O} \qquad (7)$$

なお，w_{i-2}，w_{i-1} を O の要素に限定するか，フィラーも許すかどうかによって PP_O が多少異なってくる（透過モデルの是非）．

4.2 フィラー挿入モデルの評価

最初に，フィラー挿入モデルのみの性能評価を行うため，フィラーの種類の違いを区別せず，全てのフィラーを同一視した実験を行った．結果を表 3 に示す．表 3 より，フィラー挿入モデルとして形態素トライグラムや，品詞トライグラム，単純なフィラーのユニグラム確率を用いた場合のいずれと比較しても，CRF を用いた場合の結果が良いことが分かる．特に，フィラー部分への対応の差が，PP_F の差として顕著に現れている．また，この結果は，開発コーパスからフィラーを取り除かずに作成した場合の結果 (目標値) とかなり近い．よって，フィラー挿入モデルとして CRF を用いた提案手法は，実際の話し言葉にかなり近い言語モデルを再現できると言える．また，ドメインに強く依存するような名詞や動詞・形容詞の表層形を素性として用いない場合でも，性能はほとんど低下していない．

さらに，直前 2 形態素および直後 2 形態素の基本形の文字列・品詞と挿入個所直前の 2 モーラを素性とする CRF をフィラー挿入モデルとして用いた場合について，学習コーパスの分量とテストセットパープレキシティの関係を図 2 に示す．図 2 より，200 講演程度の学習コーパスで十分な性能の言語モデルが作成できることが分かる．ただし，200 講演を下回ると性能は徐々に低下していき，特に学習コーパスが 10 講演以下まで少なくなると，十分な性能は得られなくなる．

表 3: フィラー挿入モデルの比較

フィラー挿入モデル	素性					フィラー頻度	PP	PP^*	PP_F	PP_O
	直前2形態素，直後2形態素および現在の形態素				挿入箇所直前の2モーラ					
	表層形の文字列			品詞						
	名詞	動詞/形容詞	その他							
CRF	○	○	○	○	○	152614	60.5	68.3	13.7	67.7
	×	○	○	○	○	151269	60.7	68.5	14.0	67.8
	×	×	○	○	○	153722	60.9	68.7	14.0	68.0
形態素トライグラム						134234	62.9	70.7	17.1	69.3
品詞トライグラム						155463	63.5	71.7	16.3	70.4
ユニグラム						148452	67.6	76.3	29.3	72.0
フィラーを除去していない正確な開発コーパスから作成した言語モデル						175253	59.5	67.1	10.9	67.6

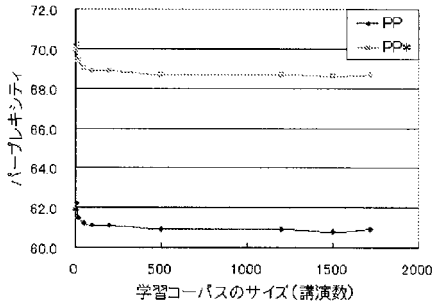

図 2: フィラー挿入モデルの性能の変化

4.3 フィラー選択モデルの評価

次に，フィラー挿入モデルとフィラー選択モデルを統合した提案手法全体の評価を行うため，フィラー挿入モデルとして，CRF，形態素トライグラムや品詞トライグラムおよびユニグラムを用いた場合，および，フィラー選択モデルとして，形態素トライグラムやモーラトライグラム，品詞トライグラムおよびユニグラムを用いた場合を組み合わせた実験を行った．結果を表 4 に示す．

表 4 より，フィラー挿入モデルとフィラー選択モデルの両方のモデル化において，厳密に周囲のコンテキストを考慮してモデル化を行っている手法が，周囲のコンテキストを考慮しないでモデル化を行っている手法に比べて，かなり良いことが分かる．また，素性としてドメインに依存しそうな名詞や動詞の形態素情報は用いずに品詞情報だけにしても性能差はない．CRF ではこのような有用な情報が自動学習されていると言える．その結果は，開発コーパスからフィラーを取り除かずに作成した場合（目標値）の結果とかなり近い．以上より，周囲のコンテキストを考慮したフィラー挿入モデルとフィラー選択モデルを組み合わせたフィラー予測モデルによって，実際の話し言葉にかなり近い言語モデルを再現できることが分かる．

表 4: フィラー選択モデルの比較

フィラー挿入モデル	フィラー選択モデル	フィラー頻度	PP	PP^*
CRF	形態素トライグラム	153722	70.6	79.6
	モーラトライグラム	153722	70.7	79.8
	品詞トライグラム	153722	70.5	79.6
	ユニグラム	153722	71.7	81.0
形態素トライグラム	形態素トライグラム	134234	72.7	81.8
	モーラトライグラム	134234	72.8	82.0
	品詞トライグラム	134234	72.6	81.7
	ユニグラム	134234	73.8	83.1
品詞トライグラム	品詞トライグラム	155463	73.2	82.7
ユニグラム	ユニグラム	148452	79.7	90.1
開発コーパスから言語モデル作成		175253	67.9	76.6

5 国会会議録を対象とした実験

本節では，実際に存在する不正確な話し言葉コーパスとして，国立国会図書館によって公開されている国会会議録を用いた実験結果について述べる．

5.1 実験条件

国立国会図書館によって公開されている国会会議録は，国会における各種会議の討議内容を対象として，フィラーや言い直し・言い淀みなどの話し言葉特有の現象は省略されている不正確な話し言葉コーパスである．

まず，テストコーパスとして，2007 年に衆議院にて行われた会議から，フィラーが CSJ と同程度に出現している会議を 4 件選び，それぞれの会議から 30 分ずつを抜き出して，合計 2 時間のコーパスを用意した．前述の通り，このコーパスには，フィラーや言い直しなどの話し言葉特有の現象は書き起こされていない．そのため，その会議の録画に含まれている音声情報を参照して，人手でフィラーの挿入を行った．なお，フィラーを挿入しようとする個所に同時に言い直しや言い淀みなどが生じていた場合には，その修正も同時に行った．これにより，少なくともフィラー出現個所の周辺部分に関しては，正確な話し言葉コーパスと見なせるテストコーパスが得られた．言語モデル作成用開発コーパスとテストコーパスの諸元を表 5 に示す．

開発コーパスとしては，1999 年から 2007 年にかけての衆議院における会議から 1083 件を抽出して

表 5: 国会会議録を対象とする比較

言語モデル作成用コーパス	総単語数	異なり語数	未知語率	フィラー頻度	PP	PP^*	PP_F	PP_O
模擬講演	3.6M	29k	13.22%	175.3k	251.4	630.0	96.4	270.0
国会会議録	36M	55k	7.93%	0	60.5	89.9	N/A	60.5
模擬講演＋国会会議録（従来法）	39.6M	69k	1.01%	175.3k	63.9	69.7	1567.4	50.5
フィラーを挿入した国会会議録	38M	55k	1.03%	1871.1k	53.7	57.4	102.8	51.2
国会会議録（テストコーパス）	22k	2k	0%	1.8k				

使用した．ただし，テストコーパスにおいて発言している話者は含まれないようにした．

5.2 フィラー予測モデルの評価

言語モデル作成用のコーパスを，(1)CSJ の模擬講演を用いた場合，(2) 国会会議録にフィラーを挿入せずに用いた場合，(3)CSJ の模擬講演と国会会議録を混合して用いた場合，(4) 提案手法により国会会議録にフィラーを挿入したコーパスを用いた場合，という 4 通りに変化させた場合の比較結果を，表 5 に示す．なお，言語モデルの語彙サイズは 20k である．

表 5 より，ドメインの異なる正確な話し言葉コーパス (1) や，ドメインが一致している不正確な話し言葉コーパス (2) をそのまま単独で用いても，不十分な性能しか得られないことが分かる．それに対して，提案手法によって作成されたコーパス (4) を用いると，補正テストセットパープレキシティ PP^* が 36%改善された言語モデルが得られる．また，提案手法による言語モデルの性能は，ドメインの異なる正確な話し言葉コーパスとドメインが一致している不正確な話し言葉コーパスを単純に混合するという既存手法 (3) を用いた場合の結果よりもパープレキシティは 17.6%改善されている[3]．PP_F を比較しても分かるように，フィラーへの対応において (3) と (4) の差は非常に大きい．

なお，PP_O で比較すると，(3) や (4) と比べて (2) の方が小さくなっていることがわかる．通常であれば，言語モデルにフィラーを含まない (2) の方が，フィラー以外の単語の生起確率が大きくなり，従って，PP_O は小さくなるはずである．今回そのようにならなかったのは，(2) で PP_O を計算する際，トライグラム確率 $P(w_i|w_{i-2}w_{i-1})$ の履歴 $w_{i-2}w_{i-1}$ にフィラーが含まれるときに必ず起こってしまうバックオフの影響によるものであると考えられる．

6 おわりに

本稿では，フィラーを含む正確な話し言葉コーパスが十分に得られない状況のもとで，フィラーを考慮した言語モデルを構築するための手法として，フィラー予測モデルを用いる方法を提案した．提案手法は 2 段階からなり，最初に，正確な話し言葉コーパスからフィラー予測モデルを作成し，次に，このモデルから与えられる確率に基づいてフィラーを挿入したコーパスから言語モデルを構築した．日本語話し言葉コーパスを対象とした実験により，提案手法は，実際の正確な話し言葉コーパスから作成された言語モデルにかなり近い言語モデルを作成できることを示した．また，国会会議録を対象とした実験により，提案手法が実際の会議録に対しても有効であることを示した．今後は，フィラーを 1 段階で予測する方法の検討および実際の音声認識における提案手法の有効性を検討していく予定である．

[3] コーパスを混合するときの重みとしては 1 : 1 ～ 10 : 1 の間で変化させた実験を行い，最も良い性能を示した 1 : 1 の場合の結果を表 5 に載せている．

参考文献

[1] Yuya Akita and Tatsuya Kawahara. Efficient estimation of language model statistics of spontaneous speech via statistical transformation model. In *Proc. of ICASSP*, Toulouse, France, May 2006.

[2] Thomas Hain, John Dines, Giulia Garau, Martin Karafiat, Darren Moore, Vincent Wan, Roeland Ordelman, and Steve Renals. Transcription of conference room meetings: An investigation. In *Proceedings of INTERSPEECH*, pp. 1661–1664, 2005.

[3] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proc. of ICML*, pp. 282–289, 2001.

[4] Kikuo Maekawa. Corpus of Spontaneous Japanese: Its design and evaluation. In *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, Tokyo, Japan, 2003.

[5] I. H. Witten and T. C. Bell. The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory*, Vol. 37, No. 4, pp. 1085–1094, Jul 1991.

[6] 太田健吾, 土屋雅稔, 中川聖一. 講義・講演音声におけるフィラー，言い淀み，倒置の発生頻度の分析. 日本音響学会 2006 年秋季研究発表会講演論文集, 2006.

[7] 岩野公司, 広畑誠, 新中庸介, 古井貞熙. 重要文抽出による音声自動要約手法とその客観評価法についての検討 (要約, 検索, 認識・理解・対話・一般). Vol. SP2005-20, pp. 1–6, 2005.

[8] 中川聖一, 赤松裕隆. 未知語を含む文集合のパープレキシティの算出法—新補正パープレキシティ—. 日本音響学会研究発表会講演論文集, Vol. 1998, No. 2, pp. 63–64, 19980901.

[9] 土屋雅稔, 太田健吾, 中川聖一. フィラー予測モデルに基づくフィラー付き言語モデルの構築. 第 1 回音声ドキュメントワークショップ論文集, pp. 81–88, 2007.

[10] 藤井敦, 伊藤克亘, 秋葉友良, 石川徹也. 音声言語データの構造化に基づく講演発表の自動要約. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 173–177, 2001.