

TR-IT-0235

連続音声認識用ベースライン言語モデル

Baseline Language Model for Continuous Speech Recognition

政瀧 浩和	高橋 一裕	塚田 元	谷垣 宏一
Masataki Hirokazu	Takahashi Kazuhiro	Tsukada Hajime	Tanigaki Kouichi
山本 博史	西野 敦士	中村 篤	
Yamamoto Hirofumi	Nishino Atsushi	Nakamura Atsushi	

1997.9.12

自然発話音声を対象とした音声翻訳システム構築のために、音声認識用言語モデルの検討を進めている。本報告は、我々、第1研究室の言語モデルグループが提供する、現時点でのベースライン言語モデル仕様について、その作成条件、及びベースライン言語モデルを用いた音声認識実験の結果を中心にまとめたものである。

©ATR 音声翻訳通信研究所

©ATR Interpreting Telecommunications Research Laboratories

もくじ

1	はじめに	2
2	データベース概要	2
3	A T Rの連続音声認識システム用言語モデル	3
3.1	可変長 N-gram (複合 N-gram) 概要	3
3.2	可変長 N-gram の生成・評価	4
4	サーチのパラメータ、及びモデルの選択	5
4.1	ビーム幅と 1st パス言語重みの決定	6
4.2	2nd パス言語重みの決定	6
4.3	言語モデルによる認識率の変化	6
4.4	結論	7
5	評価実験	7
5.1	全テストセットの認識率評価	7
5.1.1	認識率評価	7
5.1.2	不要語の削除	8
5.2	認識-翻訳接続テストセットの評価	9
6	結論・今後の予定	10

1 はじめに

自然発話音声を対象とした音声翻訳システム構築のために、音声認識用言語モデルの検討を進めている。本報告は、我々、第1研究室の言語モデルグループが提供する、現時点でのベースライン言語モデル仕様について、その作成条件、及びベースライン言語モデルを用いた音声認識実験の結果を中心にまとめたものである。言語モデル作成のための具体的なアルゴリズムの詳細については、本報告の主たる目的とするところではないため割愛した。これらアルゴリズムの詳細については、末尾にまとめた文献を参照されたい。言語モデルグループでは、これまで、1997年2月時点での中間的な研究成果として得た確率言語モデル、及び認識用単語発音辞書を、暫定版ベースライン言語モデルとして提供してきた。本報告により、正式な意味での第1版ベースライン言語モデルを提供することとなる。暫定版ベースライン言語モデルからの主な差分は以下の3点である。

- 学習データの増量による、確率言語モデルの高精度化
- 可変長 N-gram クラス数・ビーム幅・言語重みの最適パラメータの選択
- 冗長語（「間投詞」・「その他」に属する語）に対する処理の検討

また、単語の発音を正確にするため、音声データベースの音素書き起こしデータに基づいて、辞書の各単語について多種発音化を行った。

本報告では、第1版ベースライン言語モデルの性能を検証した結果を報告する。第2章では、評価、及び学習に用いたデータベースについて述べる。第3章では、言語モデルの概要、及びその作成条件、評価を示す。第4章では、サーチのパラメータおよび言語モデルの決定を行い、第5章では、音声認識実験における評価結果を示す。最後の6章では、本報告の総括を行うと共に、今後の認識率向上のための課題について述べる

2 データベース概要

現在、音声翻訳通信研究所では、音声認識および機械翻訳の研究用として、旅行会話ドメインにおける自然発話データベース（SLDB, SDB1~4）を収集している [5][6]。8月現在で、2,664対話（申込・受付は別対話としてカウント）のデータが存在し、597630語（のべ）、7,221語（異なり）である。ただし、文頭・文末の記号は単語とみなすが、句点・括弧等の記号は除いている。

これらのデータベースの内、認識の評価用として次のようなテストセットを設定している。

SDB - S1,S2,S4

旅行会話のテストセットで。特に、S1, S2は、ホテルの部屋の予約タスクからなる¹、S1,S2,S4を合わせると、SDBの旅行会話で使われているプロットの分布が、反映された会話セットとなっている。

SLDB - TDMTa,TDMTb,TDMTc

TDMTの訓練会話のうち、SLDBに含まれているもの。a, b, cはそれぞれ、発呼者、被呼者、英日通訳者を表す。

- TDMT2a, TDMT2b

TDMTの評価会話のうち、SLDBに含まれているもの。a, bはそれぞれ、発呼者、被呼者を表す。

¹S1を言語モデル open, S2を言語モデル closed として認識実験に用いられてきたという経緯がある。

- SLTA1

音声翻訳統合システムのオープンテスト用のホテル予約会話。SLDB 中のホテルの部屋を予約する会話 (TA*) から、次の基準で 24 会話を選んだ。

- * 最低限の音声の品質を確保
- * 男女の話者が混じっている (M=8/F=17)
- * 日英 / 英日がバランスしている
- * プロットが、SLDB の残りのホテル予約会話と大まかにバランスしている。
- * TDMT の訓練等に使用されていない。(今後も使われない)
- * 第四研究室で進められている音声認識用の文法の開発に使われていない。

認識評価に関しては言語モデル Open で評価を行うため、言語モデルテストセットの対話を除いた旅行会話の全データを学習セットとして使用した。ただし、テストセットの別側のパートの会話も含んでいる (例えば、TAC70021 の申込側のみがテストセットの場合は、TAC70021 の受付側の会話は学習セットに含まれる)。これらのセットのデータ量を表 1 に示す。

表 1: SLDB,SDB データベースのテストセットと学習セット

セット	対話数	発話数	単語数
SDB テストセット	42	551	7,377
SLDB テストセット	108	1,411	24,318
学習セット	2,514	30,112	565,935
合計	2,664	32,074	597,630

暫定版言語モデル構築の際は、学習単語がのべ 414,326 語であり、今回の第一版言語モデルでは、学習データ量が約 37% 増加した。

3 ATRの連続音声認識システム用言語モデル

3.1 可変長 N-gram (複合 N-gram) 概要

ATR 音声翻訳通信研究所第一研究室では、現在、「可変長 N-gram(複合 Bigram)」を用いた連続音声認識を行っている [1][2]。本モデルは、次の 3 種類のクラスを単位としたの Bigram 確率モデルである。

- A) 品詞クラス
- B) 独立した 1 単語のみで構成されるクラス
- C) 1 接続単語列のみで構成されるクラス

出現頻度の小さい単語に対しては、品詞クラスとして扱うことでパラメータ推定の信頼性を増し、出現頻度の大きい単語に対しては、可変長フレーズの N-gram として扱うことにより、モデルの精度を高めることができる。従って、従来の単語 N-gram と比較して、少量のテキストデータでの精度の良いモデルを得ることができる。

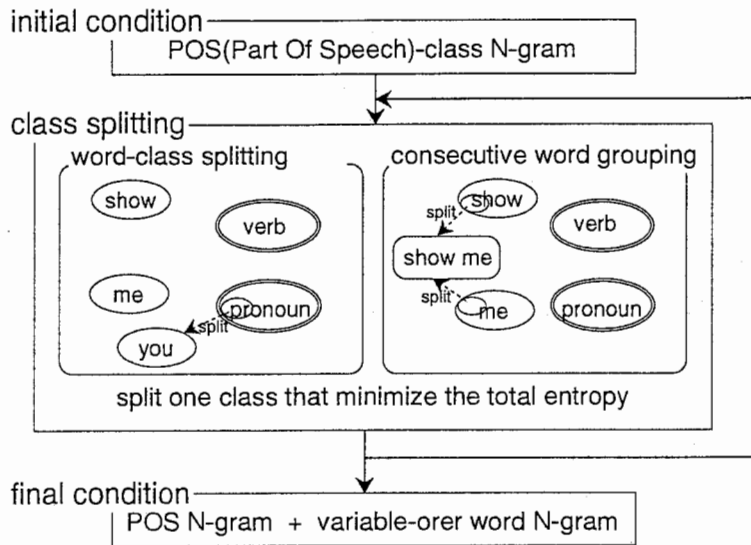


図 1: 可変長 N-gram の生成過程

可変長 N-gram による文の生成確率は、下式のクラス N-gram の形で与えることができる。

$$P(w_1^L) = \prod_{t=1}^K P(ws_t|c_t) \cdot P(c_t|c_{t-1}) \quad (1)$$

但し、 ws_t は文章を上記のクラス分類を用いた場合の、 t 番目の単語列（単独の単語も含める）を意味する。また、文章の K は単語列の個数を表し、 $K \leq L$ である。例として、次の文章（7単語）を考える。

「わたくし - 橋本 - と - 言 - い - ま - す」

「橋本」は出現頻度が低いいため、固有名詞クラスとして扱う方が適切であると考えられる。「わたくし」及び「と」は日本語の文章で頻繁に出現する単語であるため、品詞クラスより分離して単独で扱う。また、「言 - い - ま - す」は日本語で頻繁に用いられるフレーズであるため、結合させて一単位として扱う方が効果的であると考えられる。従って、この文章の生成確率は、次の式で与えられる。

$$\begin{aligned}
 P(w_1^L) &= P(\text{わたくし}|\{\text{わたくし}\}) \cdot P(\{\text{わたくし}\}) \\
 &\cdot P(\text{橋本}|\{\text{固有名詞}\}) \cdot P(\{\text{固有名詞}\}|\{\text{わたくし}\}) \\
 &\cdot P(\text{と}|\{\text{と}\}) \cdot P(\{\text{と}\}|\{\text{固有名詞}\}) \\
 &\cdot P(\text{言います}|\{\text{言います}\}) \cdot P(\{\text{言います}\}|\{\text{と}\})
 \end{aligned}$$

3.2 可変長 N-gram の生成・評価

可変長 N-gram は品詞 Bigram を初期状態とし、エントロピーの最小化基準により、単語の品詞クラスからの分離、の 2 種類のクラスを分離を逐次的に行うことにより得られる。詳しくは参考文献 [1][2] を参考のこと。

可変長 N-gram の生成は、初期状態として活用形・活用型を含めた 158 品詞クラスを用い、0,200,500,1000 のクラス分離のデータを採取した。なお、学習データに出現しない単語、及びクラス遷移に対する対策として、削除補完法 [3] を用い 0 でない確率を与えた。

パープレキシティの比較を表 2 に示す。

表 2: テストセットパープレキシティによる言語モデルの評価

モデル	Number of Parameters	SDB Testset Perplexity	SLDB Testset Perplexity
単語 Bigram	67,788	18.52	19.92
単語 Trigram	208,304	15.93	14.68
可変長 N-gram 0	9,523	31.53	40.27
200	17,493	17.42	23.17
500	34,129	15.52	19.45
1000	54,818	15.07	17.79

表 2より、可変長 N-gram は分離クラス数に比例してパープレキシティは減少し、1000 クラス分離した段階では、いずれのテストセットにおいても単語 Bigram よりも低い値を示す。単語 Trigram と比較すると、SDB テストセットに関しては、可変長 N-gram の方がパープレキシティが低いが、SLDB テストセットにおいては高くなる。しかし、パラメータ数は格段に少なく、また、インプリメントの容易さ、探索空間の狭さ等から判断すると、可変長 N-gram は非常に有利である。

4 サーチのパラメータ、及びモデルの選択

統計的手法による連続音声認識は、

$$\arg \max_W P(W|X) = \arg \max_W P(X|W)P(W) = \arg \max_W \frac{+ \log P(X|W) \log P(W)}{\quad} \quad (2)$$

(X : 観測データ W : 単語系列)

の式を満たす単語列の候補 W を認識結果とするが、通常、言語の Log スコアに係数 (言語重み, λ) を乗じて

$$\arg \max_W \log P(X|W) + \lambda \log P(W)$$

とするのが一般的である。正確な認識率を得るためには、この言語重みを適切に決定する必要がある。

また、現在第一研究室では、探索の手法として、ビームサーチの手法を用いているが、ビーム幅により、認識率・計算時間は大きく変化する。ビーム幅が狭い場合は、最適解が求まりにくいため認識率が低下するが、逆にビーム幅を広くした場合は、探索空間が爆発的に増大し、容量の不足・計算時間の増大を招く。従って、ビーム幅に対しても、適切な値を決定する必要がある。

本章では、認識評価実験に先立って、言語重み・ビーム幅の2種類のパラメータを決定するために行った実験結果を示す。実験は、次の順序で行った。

1. 1st Pass の言語重み、ビーム幅の決定 → 4.1 節
2. 2nd Pass の言語重みの決定 → 4.2 節
3. 言語モデルの決定 → 4.3 節

なお、本章全体を通して、次の条件で実験を行っている。

- 音響モデル: ベースライン音響モデル Ver 1. 性別非依存モデル ([8])
- 言語モデル: 可変長 N-gram 分離クラス数 1000 (4.3 節では 0-1000 に変化させる)
- サーチ: 単語グラフサーチ [7] (ATRLattice, r04r04)
- テストセット S1,S2 16 会話

4.1 ビーム幅と 1st パス言語重みの決定

最初に、認識率に最も大きな影響を与える、ビーム幅と 1st パスの言語重みを次のように変化させ、最適値を決定した。

- ビーム幅: 80, 90, 100
- 1stPass 言語重み: 6.0, 8.0, 10.0
- 2ndPass 言語重み: 1stPass と同一値

それぞれ条件における、S1,S2 テストセットの認識率（単語 Accuracy）を表 3 に示す。ただし、表中の単語 Accuracy の欄における - の記号は、探索空間の増大のためメモリ容量が不足し、認識できない対話があったことを意味する。

表 3: ビーム幅・1st パス言語重みと認識率との関係

言語重み	ビーム幅	Top Accuracy	Net Accuracy
6.0	70	67.7	78.3
6.0	80	-	-
6.0	90	-	-
6.0	100	-	-
8.0	70	62.7	67.1
8.0	80	70.5	77.1
8.0	90	74.3	83.3
8.0	100	-	-
10.0	70	67.7	78.3
10.0	80	70.4	81.9
10.0	90	70.1	75.4
10.0	100	74.0	81.6

表 3 より、次の 2 つの条件の時に、ほぼ最高の認識率が得られることが分かった。

- 言語重み 8.0、ビーム幅 90
- 言語重み 10.0、ビーム幅 100

4.2 2nd パス言語重みの決定

現在の探索手法では、1st パスでは弱い言語重みにして認識候補を多数求め、2nd パスでは言語重みを強くし正確な認識結果を得るといった戦略をとっている。

本節では、2nd パスでの言語重みを決定するための実験を行った。1st パスのビーム幅、言語重みは、前節で結果の良かった 2 種類の条件を用いた。なお、2nd パスはビーム幅によらず最適解を求めることができるため、ためビーム幅は 90 に固定している。

認識結果を表 4 に示す。

表 4 より、最も認識率が良かったのは、次の条件の場合である。

- 1st Pass ビーム幅: 90
- 1st Pass 言語重み: 8.0
- 2nd Pass 言語重み: 20.0

4.3 言語モデルによる認識率の変化

第 4.3 節では、パープレキシティにより言語モデルの評価を行ったが、パープレキシティと認識率とは相関があるとされているが、パープレキシティが低ければ、必ず認識率が向上するとは限らない。そこで、本節では、認識率により言語モデルを評価した。サーチのパラメータは前節で決定した値を用いた。

表 4: 2nd Pass 言語重みと認識率との関係

1stPass 言語重み	2ndPass 言語重み	ビーム幅	単語 Accuracy
8.0	8.0	90	74.3
8.0	10.0	90	75.8
8.0	12.0	90	76.1
8.0	14.0	90	75.3
8.0	16.0	90	76.3
8.0	20.0	90	77.3
8.0	24.0	90	77.0
8.0	32.0	90	76.4
10.0	10.0	100	74.0
10.0	12.0	100	74.9
10.0	14.0	100	75.6
10.0	16.0	100	76.1
10.0	20.0	100	76.3
10.0	24.0	100	76.5
10.0	32.0	100	76.0

表 5: 可変長 N-gram の分離クラス数と認識率との関係

分離クラス数	単語 Accuracy
0	58.0
200	71.8
500	73.7
1000	77.3

表 5より、パープレキシティでの評価と同様、分離クラス数が大きくなるのに比例して認識率が向上することが確認できた。可変長 N-gram の分離クラス 1000 の場合を用いた場合が最も認識率が高い。

4.4 結論

これまでの実験より、ベースライン言語モデルとして、次のモデル、及びパラメータを使用して認識評価を行う。

- 言語モデル：可変長 N-gram 分離クラス 1000
- 1st Pass ビーム幅: 80
- 1st Pass 言語重み: 8.0
- 2nd Pass 言語重み: 20.0

5 評価実験

5.1 全テストセットの認識率評価

5.1.1 認識率評価

前節で求めた条件により、全テストセットの認識率を評価した（実験に用いたスクリプトは付録 C）に掲載している）。なお、本表で Real Time Factor は、発話時間に対して何倍の時間がかかったかを表す値である。ただし、実験にはランニ

ングマシン DEC ALPHA STATION 500(500MHz)を使用した。また、TDMT2テストセットに関しては、波形ファイル・トランスクリプションファイルが整備されていない対話が多かったため、今回は実験を行わなかった。

各テストセット毎の認識率を表6に示す。

表6: 各テストセットにおける認識率

テストセット	Top 単語 Accuracy	Net 単語 Accuracy	Real Time Factor
S1	77.03	82.94	4.12
S2	77.00	81.90	4.11
S4	67.28	74.25	4.53
SDB テストセット平均	70.57	82.94	4.38
TDMTa	75.44	81.15	5.69
TDMTb	74.39	79.60	4.77
TDMTc	81.23	84.94	4.02
SLTA1	77.88	83.50	5.49
SLDB テストセット平均	76.97	82.26	5.17
全平均	74.66	80.37	4.85

暫定版の言語モデルは、SDBテストセットの単語 Accuracy が1位候補で 67.3% ネット内で 75.2%であったのに対し、本ベースライモデルは、1位候補で 70.6% ネット内で 82.9%と、それぞれ約 3.3%、及び 7.7% 向上している。なお、各会話毎の認識率、及び計算時間は付録A)の表9、9に掲載した。

5.1.2 不要語の削除

SLDB データベースは自然発話のデータを集めたものであるため、

- 言い淀み・言い間違い等の「その他」に属する語、
- 「あー」「えーと」等の「間投詞」

等の冗長語が含まれる。

「その他」の語は、データベース中で一度しか出現しないものがほとんどであり、現在の N-gram に基づいた言語モデルでは、有意義な統計量を得るのは困難である。従って、「その他」の語は認識が困難であり、逆に、湧き出し誤り等の悪影響を及ぼす可能性もある。従って、「その他」の語を辞書から削除することが望ましい。

また、「間投詞」は、「あ」と「あー」や、「えっとー」と「えーとー」のように、極めて類似した語があり、これらは音響的にも言語的にも識別することが困難である。また、短い「あ」「え」等の間投詞は、書き起こしの際に省略されている場合も多く、正しい認識率が求められない可能性がある。従って、間投詞も評価の対象外とするのが好ましい。ただし、「その他」の語と異なり、データ中には大量に出現するため、辞書からは削除せず、認識は行すが、認識率評価時に計算の対象外とした。

各テストセット毎の認識率を表7に示す。

表7より、前節の結果と比較して、認識率は約 1.5% 向上した。当然、これは、認識結果が向上した訳ではないが、認識システムに翻訳等の後処理システムを結合させることを考えた場合こちらが妥当な評価であると思われる。

また、計算時間は約 10% 向上している。これは「その他」の語彙を削除した結果であるが、デモシステム等で計算の高速性が要求される場合は、固有名詞等の、認識の対象としては不適切な語彙を削除することによりさらなる高速化が図れると考えられる。

表 7: 各テストセットにおける認識率 (不用語の削除)

テストセット	Top 単語 Accuracy	Net 単語 Accuracy	Real Time Factor
S1	78.28	83.71	3.80
S2	78.04	83.33	3.79
S4	69.39	76.92	4.08
SDB テストセット平均	72.37	79.17	3.98
TDMTa	77.68	82.80	5.03
TDMTb	75.56	80.87	4.30
TDMTc	82.96	86.28	3.66
SLTA1	78.57	83.83	4.97
SLDB テストセット平均	78.23	83.22	4.65
全平均	76.15	81.78	4.38

5.2 認識-翻訳接続テストセットの評価

音声翻訳通信研究所では、最終デモに向けて音声認識と機械翻訳のシステムとを統合したシステムの試作が始まっており、次の9会話（申し込み側、受付側、通訳側の全パートを含む）を用いて認識-翻訳接続システムの性能評価を行ってきた。

TAS12008 TAS12010 TAS13005 TAS13009 TAS22001

TAS23001 TAS32002 TAS33001 TAS33011

本会話は、いずれも SLDB データベース内の会話で、タスクはホテル予約 (TA) に限定している。

次の3種類言語モデル・語彙について比較を行った。

MODEL A モデルの学習に学習セット全体を用い、語彙もデータベースの全語彙としたもの

MODEL B モデルの学習に学習セット全体を用い、語彙は SLDB データベースのホテル予約に出現する単語に限定したもの

MODEL C モデルの学習には SLDB のホテル予約タスクのデータのみを用い、語彙も SLDB ホテル予約タスクに出現する単語に限定したもの

各モデル毎の認識率を表 8 に示す。

表 8: 翻訳接続評価用テストセットの認識結果

モデル	役割	Top Accuracy	Net Accuracy	Real Time Factor
MODEL A	A,B	83.81	86.83	3.91
	C	81.23	84.94	4.02
	Total	82.57	85.92	3.97
MODEL B	A,B	83.94	87.55	2.60
	C	81.65	86.41	2.67
	Total	82.84	87.00	2.63
MODEL C	A,B	84.08	88.01	2.94
	C	84.59	88.31	2.89
	Total	84.33	88.15	2.92

モデルA, B, Cの順に認識率が良くなっている。しかし、モデルBはモデルAより0.3%、モデルCは1.8%程度の向上にとどまり、語彙限定、言語モデル学習タスクの限定による効果は、認識率向上への寄与は小さかった。ただし、計算時

間に関しては、モデルBはモデルAよりも33%、モデルCは26%短縮され、デモ等で計算の高速性を重視する際には有効である。

6 結論・今後の予定

本報告では、第一研究室における連続音声認識のための標準言語モデルの評価を行った。認識率は全テストセットの平均で74.4%であった。接続テストセット9会話において、語彙、及び学習データをホテル予約タスク限定した結果、82.6%から84.3%に向上し、計算時間も26%短縮された。

今後、認識率のさらなる向上のために、言語モデルとして、

- 補完（平滑化）の手法の変更
- タスク適応
- FSA, TDMT の知識等による強い言語制約を用いた再探索
- TDMT 体系の形態素による認識

等について検討を行う予定である。

参考文献

- [1] 政瀧, 松永, 匂坂: “連続音声認識のための可変長連鎖統計言語モデル”, 信学技報, SP95-73, pp1-6, 平成7年
- [2] H. Masataki, Y. Sagisaka: “Variable-order N-gram generated by word-class splitting and consecutive word sequence,” ICASSP96, Vol1. pp188-191, 1996.
- [3] F. Jelinek and R. L. Mercer: “Interpolated estimation of Markov Source Parameters from Sparse Data,” Proc. Workshop Pattern Recognition in Practice, pp.381-37, 1980.
- [4] Singer. H & Ostendorf. M.: “Maximum Likelihood Successive State Splitting,” Proc. ICASSP-96, Vol.2, pp.601-604, May 1996.
- [5] T. Morimoto, N. Uratani, T. Takezawa, O. Furuse, Y. Sobashima, H. Iida, A. Nakamura, Y. Sagisaka, N. Higuchi and Y. Yamazaki: “A Speech and Language Databases for Speech Translation Research,” ICSLP, pp1791-1794, 1994.
- [6] A. Nakamura, S. Matsunaga, T. Shimizu, M. Tonomura and Y. Sagisaka: “Japanese speech database for robust speech recognition,” ICSLP, pp 2199-2192, 1996.
- [7] 清水, 山本, 松永, 匂坂: “単語グラフを用いた自由発話音声認識” 信学技報, SP95-88, pp49-54, 平成7年
- [8] H. Singer, M. Tonomura, Q. Huo, J. Ishii, T. Fukada, M. Schuster, H. Yamamoto: “Baseline Acoustic Models for the Spoken Language Database(SLDB),” ATR Technical Report, TR-IT-0206, 1997.

付録A) 認識率一覧

表 9: 認識率一覧 (1)

会話番号	役割	話者性別	Top Accuracy	Net Accuracy	Utterance Time	CPU Time	RTF
S1.ascii							
TAC70021	A	M	85.4	87.0	55.2	220.6	4.0
TAC70019	A	F	77.4	85.0	46.6	171.7	3.7
TAC70101	A	F	92.9	94.3	57.0	153.9	2.7
TAC70103	A	M	77.1	78.3	30.0	126.1	4.2
TAC70202	A	M	61.3	70.2	48.8	319.5	6.5
TAC70203	A	F	77.6	85.8	70.0	285.4	4.1
TAC70303	A	F	88.1	94.1	56.3	137.8	2.4
TAC70304	A	M	55.2	65.5	36.8	237.7	6.5
S2.ascii							
TAC70015	A	F	71.2	77.1	52.1	231.7	4.4
TAC70016	A	M	72.6	73.6	35.8	167.8	4.7
TAC70017	A	M	89.9	89.9	29.2	91.7	3.1
TAC70022	A	M	64.2	70.4	61.2	372.2	6.1
TAC70023	A	M	68.8	76.6	63.7	254.7	4.0
TAC70102	A	F	85.0	91.0	75.4	260.4	3.5
TAC70201	A	F	82.2	85.4	55.2	159.5	2.9
TAC70301	A	F	85.8	91.8	46.4	183.2	4.0
S4.ascii							
TCC70109	A	M	65.5	76.5	47.4	246.0	5.2
TCC70103	A	F	72.0	74.2	42.0	125.9	3.0
TCC71035	A	F	84.5	86.4	42.3	181.7	4.3
TCC71008	A	M	62.4	66.7	74.2	280.8	3.8
TCC71016	A	F	76.1	84.1	71.0	267.5	3.8
TCC71001	B	F	67.7	75.7	76.8	342.3	4.5
TCS70070	A	M	54.1	67.1	31.1	152.7	4.9
TCC70201	A	M	29.1	44.7	30.7	208.6	6.8
TCC70307	A	M	74.5	80.6	54.8	296.6	5.4
TCS70055	A	M	65.3	76.8	59.7	260.8	4.4
TCS70074	A	M	86.3	87.3	36.3	107.0	2.9
TCC70212	A	M	45.5	54.0	69.5	392.8	5.6
TCS70034	A	M	55.0	63.9	69.9	408.0	5.8
TCC71007	A	F	78.5	87.3	84.2	266.0	3.2
TCS70013	A	F	78.9	86.0	43.1	122.9	2.9
TSC71005	B	F	79.9	85.1	58.2	256.1	4.4
TCS70023	A	F	71.7	77.5	87.1	308.7	3.5
TCS70059	A	F	82.0	87.0	36.7	121.6	3.3
TCS70082	A	F	61.4	71.3	62.8	272.4	4.3
TCS70004	B	F	56.2	64.1	79.4	559.8	7.1
TCS70010	A	F	57.8	59.4	34.9	198.7	5.7
TCS70028	A	F	88.2	90.0	44.0	139.9	3.2
TCS70025	A	F	70.0	77.9	44.0	140.6	3.2
TSC71013	A	F	61.7	67.7	90.7	499.2	5.5
TCS70020	A	F	76.5	77.3	51.6	356.2	6.9
TCS70047	A	F	74.8	85.2	46.5	146.9	3.2

(次頁へつづく)

表 10: 認識率一覧(2)

会話番号	役割	話者性別	Top Accuracy	Net Accuracy	Utterance Time	CPU Time	RTF
TDMTa.ascii							
TAS22001	A	F	89.9	94.1	36.5	146.1	4.0
TBS22001	A	F	80.2	82.1	34.8	226.7	6.5
TCC22074	A	F	56.4	65.4	48.2	286.4	5.9
TGS22001	A	F	75.4	79.1	41.4	195.2	4.7
THS22003	A	F	86.5	92.5	41.4	211.6	5.1
TSS22001	A	F	80.8	84.9	46.7	246.4	5.3
TCC22011	A	F	60.7	69.7	54.0	301.7	5.6
TAS32002	A	F	85.6	88.0	37.6	133.0	3.5
TCC22031	A	F	76.6	81.4	63.9	376.0	5.9
TAS12008	A	F	92.1	93.1	32.1	139.1	4.3
TAS12010	A	F	88.8	92.1	23.0	119.7	5.2
TCS12020	A	F	71.1	78.1	39.0	167.6	4.3
TDS12004	A	F	74.5	86.3	30.8	142.0	4.6
TCS12025	A	F	80.4	84.7	46.2	290.9	6.3
TCC22034	A	M	69.1	77.6	39.7	373.3	9.4
TCS32017	A	M	73.9	79.3	26.4	118.8	4.5
THS12001	A	M	66.2	72.9	39.3	298.3	7.6
TOS12003	A	M	57.3	65.9	26.3	252.4	9.6
TDMTb.ascii							
TAS23001	B	F	80.6	82.7	72.0	294.9	4.1
TBS23001	B	F	73.7	76.7	45.4	212.2	4.7
TIS23001	B	F	77.4	81.6	67.8	401.8	5.9
TOS23006	B	F	68.6	74.5	58.8	472.1	8.0
TAS33001	B	F	90.2	92.7	79.6	239.2	3.0
THS33001	B	F	83.3	88.1	83.3	281.6	3.4
TKS33002	B	F	54.5	65.4	44.3	235.6	5.3
TRS33002	B	F	69.3	73.5	67.3	298.0	4.4
TAS13005	B	F	73.2	78.6	36.2	129.8	3.6
TAS13009	B	F	88.6	90.4	66.7	244.9	3.7
TCS13021	B	F	63.6	75.6	58.3	318.7	5.5
TCS13022	B	F	59.7	68.4	76.9	493.7	6.4
TAS33011	B	M	73.0	77.8	81.8	375.0	4.6
TDMTc.ascii							
TAS22001	C	F	77.2	81.9	68.0	204.1	3.0
TAS32002	C	F	74.2	78.5	62.7	284.5	4.5
TAS12008	C	F	80.9	87.2	42.7	312.6	7.3
TAS12010	C	F	92.3	93.4	59.5	185.2	3.1
TAS23001	C	F	80.3	81.8	43.7	147.5	3.4
TAS33001	C	F	89.9	94.2	41.8	137.2	3.3
TAS13005	C	F	82.3	83.9	17.1	72.3	4.2
TAS13009	C	F	69.0	76.1	34.7	152.6	4.4
TAS33011	C	F	86.5	88.5	58.0	226.6	3.9
SLTA1.ascii							
TAC22013	A	F	83.8	88.9	73.4	358.4	4.9
TAC23022	B	F	82.1	86.2	34.3	134.2	3.9
TAS12002	A	M	83.5	88.1	77.7	410.0	5.3
TAS12006	A	F	80.2	81.3	26.6	127.8	4.8
TAS12010	A	F	88.8	92.1	23.0	119.7	5.2
TAS12014	A	F	88.2	91.4	51.5	256.7	5.0
TAS12016	A	F	87.7	92.0	39.2	202.1	5.1
TAS12019	A	F	91.4	95.4	55.0	189.0	3.4
TAS12020	A	F	77.9	82.1	54.9	310.3	5.7
TAS13002	B	M	77.4	84.2	80.3	520.4	6.5
TAS13014	B	F	80.0	85.2	94.0	677.0	7.2
TAS13018	B	F	78.9	86.7	39.9	322.5	8.1
TAS13019	B	F	81.7	86.0	92.0	354.6	3.9
TAS13020	B	F	67.5	77.6	74.5	516.9	6.9
TAS13022	B	F	76.5	82.6	65.4	591.5	9.0
TAS13024	B	F	74.2	76.5	67.6	384.5	5.7
TAS32001	A	F	84.6	92.6	43.5	163.3	3.8
TAS32012	A	M	78.6	82.7	44.4	255.3	5.7
TAS32014	A	M	79.9	85.1	41.0	209.1	5.1
TAS32016	A	M	56.6	71.3	33.6	228.5	6.8
TAS33010	B	M	70.1	74.1	64.1	303.3	4.7
TAS33015	B	M	64.4	72.1	79.2	407.8	5.1
TAS33020	B	M	71.4	76.7	71.0	365.3	5.1
TAS33022	B	F	77.1	84.9	69.8	262.7	3.8
平均(合計)	-	-	74.7	80.4	5658.7	27449.7	4.85

付録B) 可変長 N-gram 学習時の初期品詞クラス

表 11: 品詞一覧

品詞	代表例
活用を伴わないもの (28種類)	
UTT START	(発話開始のシンボル)
UTT END	(発話終了のシンボル)
その他	はい <u>そ</u> そうです
サ変名詞	予約 <u>したい</u> んですが
引用助詞	到着は二時頃になる <u>と</u> 思うのですが
格助詞	部屋 <u>を</u> 予約したいのですが
感動詞	<u>もしもし</u> 鈴木と申しますが
間投詞	<u>えー</u> 十三日をお願いします
係助詞	お部屋 <u>は</u> ありますか
形容名詞	予約 <u>可能</u> ですか
固有名詞	<u>もしもし</u> <u>ニューヨークシティーホテル</u> です
終助詞	一泊でございます <u>か</u>
住所名	<u>東京</u> <u>都</u> <u>江戸川</u> 区です
準体助詞	お願いしたい <u>ん</u> です
人名	<u>鈴木</u> <u>直子</u> です
数詞	<u>二</u> 泊をお願いします
接続詞	それでは <u>人数</u> をお知らせ下さい
接続助詞	確認致します <u>ので</u> 少々お待ち下さい
接頭辞	何時頃 <u>ご</u> 到着になりますでしょうか
接尾辞	<u>二人</u> をお願いします
代名詞	<u>わたくし</u> 一人です
日時	<u>八月</u> の <u>十日</u> からをお願いします
普通名詞	<u>部屋</u> の予約をしたいんですが
副詞	確認致しますので <u>少々</u> お待ち下さい
副助詞	百ドル <u>ぐらい</u> の部屋をお願いします
並立助詞	住所 <u>と</u> 電話番号をお聞かせ下さい
連体詞	<u>どのような</u> お部屋がよろしいでしょうか
連体助詞	部屋 <u>の</u> 予約をしたいんですが
活用を伴うもの (5種類, 括弧内の数字は活用の種類数)	
形容詞 (3)	到着は <u>遅</u> くなりそうです
語尾 (67)	ジョンフィリップスで <u>す</u>
助動詞 (24)	ご用件を承り <u>ます</u>
補助動詞 (13)	シングルをお願い <u>いた</u> します
本動詞 (23)	何時ごろ <u>来</u> られますか