接続の方向性を考慮した 多重クラス複合Ｎ－ｇｒａｍ言語モデル

山本博史　匂坂芳典

(株) ATR 音声翻訳通信研究所
〒619-02 京都府相楽郡精華町光台 2-2
E-mail: yama@itl.atr.co.jp

あらまし　クラス２－ｇｒａｍにおける効率的なクラス分類を実際のコーパスから統計的に行うための手法を提案する。本手法では直前および、直後の単語への接続性を別の属性としてとらえ、各単語に対してその属性ごとに複数のクラスを割り当てる。これらのクラスは前後に接続している単語の分布に基づいて各々独立に作成されることによって、効率的でかつ信頼性の高いクラス分類となっている。さらにこの多重クラス２－ｇｒａｍを結合単語との多重複合Ｎ－ｇｒａｍに拡張することにより、千分の一以下の論理パラメーターサイズでパープレキシティ、単語認識率とも単語Ｎ－ｇｒａｍを上回る性能を示した。

キーワード　クラスＮ－ｇｒａｍ　可変長Ｎ－ｇｒａｍ　自動クラス分類　連鎖語

MULTI CLASS COMPOSITE N-GRAM LANGUAGE MODEL BASED ON CONNECTION DIRECTION

Hirofumi Yamamoto, Yoshinori Sagisaka

ATR Interpreting Telecommunications Res.Labs.
2-2 Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
yama@itl.atr.co.jp

Abstract　A new word-clustering technique is proposed to efficiently build statistically salient class 2-grams from language corpora. By splitting word neighboring characteristics into word-preceding and following directions, multiple (two-dimensional) word classes are assigned to each word. In each side, word classes are merged into larger clusters independently according to preceding or following word distributions. This word-clustering can provide more efficient and statistically reliable word clusters. Further, we extend it to Multi-Class Composite N-gram that unit is Multi-Class 2-gram and joined word. Multi-Class Composite N-gram showed better performance both in perplexity and recognition rates with one thousandth smaller logical parameter size than conventional word 2-grams.

key words　Class N-gram, Variable Order N-gram, Automatic Clustering, Joined Word

1. はじめに

連続音声認識のための統計的言語モデルとして単語Ｎ－ｇｒａｍモデルが広く用いられている。単語Ｎ－ｇｒａｍモデルはルールベースの文法的な制約に比べて効率的かつ柔軟な場合が多い。しかしながらコンパクトな認識システムを構築する場合、必要とするメモリの多さは重大な問題点である。一方統計的な観点からは観測データの少ない（あるいは存在しない）単語ペアに対しては統計的に正確な値を予測することが難しいという問題がある。これらの問題の解決法としてはクラスＮ－ｇｒａｍあるいは可変長Ｎ－ｇｒａｍによる手法が提案されている。[1]-[4]

これらの手法では単語に対するクラスの定義方法はモデルの精度、サイズに直結する重要な問題点となっている。実際総単語数４５９，３８３語、異なり単語数７，２２１の訓練セットにおけるテストセットパープレキシティが単語２－ｇｒａｍで１８．５１であったのに対し、品詞情報に基づく１５８種類のクラスを用いたクラス２－ｇｒａｍはサイズを単語２－ｇｒａｍの０．５％に縮めることができたものの、パープレキシティは３１．５３にとどまった。これらのことから精度の高いクラスＮ－ｇｒａｍの構築のためには、クラスの規定方法に工夫必要がある。

本稿では、単語に対するクラスの規定方法として単語の前方向と後方向の接続性を別の属性とみなし、各単語にその属性ごとに複数のクラスを割り当てる多重クラスを提案する。さらに、頻出する単語列を連鎖語として取り扱う多重クラス複合Ｎ－ｇｒａｍを提案する。実験によりこの多重クラス複合Ｎ－ｇｒａｍは、単語２－ｇｒａｍの千分の一以下の論理パラメータサイズでより高い認識性能を示すことが判明した。

2．多重クラス

2.1 クラス2－ｇｒａｍ

クラスＮ－ｇｒａｍにおいて直前のＮ－１個の単語列に続いて次単語が出現する確率は一般に次の式で与えられる。

$$P(W_N \mid C_N) \times P(C_N \mid C_{N-1}, C_{N-2}, \ldots, C_1) \quad (1)$$

（ここでC_Nは単語W_Nが属するクラスを表わす）

さらにＮ＝２としたクラス２－ｇｒａｍでは

$$P(W_2 \mid C_2) \times P(C_2 \mid C_1) \quad (2)$$

となり、これはクラス２－ｇｒａｍにおいて必要なパラメーター数がクラス数の２乗に比例することを表わしており、クラス数の増加によって急激にパラメーター数が増加することになる。

2.2 クラス２－ｇｒａｍにおけるクラスの問題点

ここでクラス２－ｇｒａｍにおけるクラスが単語間の接続性を表わすためにどのような働きをしているかをみてみることにする。一例としてある動詞を考える。動詞は活用形ごとに単語の接続性が異なるため活用形の６種類ごとに別々のクラスが割り当てられるのが普通である。しかしながら活用形が接続性に与える影響は後続の単語についてのみであり、先行の単語に対する接続性には影響を与えないと考えられる。このことは後続するクラスとして動詞が現われるかどうかを推定する場合には活用形を考慮したクラス分けが不要であることを示している。

一方動詞を自動詞と他動詞に分けた場合、これらの接続性の差は主として先行する格助詞の種類などに対して現われる。このため、先行するクラスとして動詞が現われる場合には不要なクラス分け（この場合は逆に活用形のみが重要であるため）と考えられる。

2.3 多重クラスＮ－ｇｒａｍ

上記の動詞の例が示すように、先行するクラスと後続するクラスでは接続性を表わすために有効なクラスが必ずしも一致しない。このため先行、後続で用いられる場合で用いるクラスを別にする方が効率的と考えられる。この考えを（２）式に対してあてはめると次のようになる。

$$P(W_2 \mid C_2^t) \times P(C_2^t \mid C_1^f) \quad (3)$$

ここでC^tは後続する場合のクラス（以下ｔｏ－クラスと呼ぶ）を表わし、C^fは先行する場合のクラス（以下ｆｒｏｍ－クラスと呼ぶ）を表わす。従来のクラスＮ－ｇｒａｍおいては自動詞、他動詞およびその活用形を表現するためには２（自動詞、他動詞の）×６（活用形の）＝１２通りのクラスが必要であり、パラメーター数はクラス数の自乗であるため１２×１２となる。これに対し本提案では２通りのｔｏ－クラスと６通りのｆｒｏｍ－クラスで表現可能で、かつパラメーター数もｔｏ－クラス数×ｆｒｏｍ－クラス数の１２と、きわめて効率的なクラス分類が可能となる。（３）式はＮ＞２の場合にも拡張可能で、クラスＮ－ｇｒａｍ同様直前のＮ－１個の単語列に続いて次単語が出現する確率は次の式で与えられる。

$$P(W_N \mid C_N^t) \times P(C_N^t \mid C_{N-1}^f, C_{N-2}^{f-1}, \ldots, C_1^{f-N+2}) \quad (4)$$

このように各単語に対して接続の方向性ごとに与えられる複数のクラスを多重クラスと呼び、この多重クラスを用いたクラスN－ｇｒａｍを多重クラスN－ｇｒａｍと呼ぶことにする。

3．クラスの自動分類

3.1 自動分類の目的

クラスN－ｇｒａｍにおけるクラス分類の指標としては、しばしば品詞情報が用いられる。品詞情報はコーパスに現われない単語に対しても割り当てることができるという利点がある。しかしながらN－ｇｒａｍにとって最も重要な単語間の統計的な接続特性を詳細に表しているとは必ずしも言い難い。また品詞情報自体は先見知識として与えられるものであるため、品詞情報自体の分類のしかたがクラスN－ｇｒａｍにおけるクラス分類の効率に直接影響を与えてしまうという問題点もある。従って統計的モデルの観点からは、実際のコーパスから単語間の接続性のみに着目してクラス分類を行うことが好ましい。

3.2 自動分類の方法

実際のコーパスからクラス分類を自動的に行う方法としては、[1]をはじめとするいくつかの提案がなされている。本稿においては以下に示すような手順で自動クラス分類を行った。分類の対象としたクラスは従来のクラスN－ｇｒａｍにおけるクラス、ｔｏ－クラス、ｆｒｏｍ－クラスの３つである。

1．一単語一クラスとする。

2．個々の単語またはクラス X に対して次のようなベクトルを考える。

$$V_t(X)=\{P_t(W_1|X),P_t(W_2|X),...,P_t(W_n|X)\} \quad (5)$$

$$V_f(X)=\{P_f(W_1|X),P_f(W_2|X),...,P_f(W_n|X)\} \quad (6)$$

ここで $P_t(W_i|X)$、$P_f(W_i|X)$ は単語またはクラス X から単語 W への後向きおよび前向きの２－ｇｒａｍの確率値を表す。従来のクラスN－ｇｒａｍにおけるクラスでは前後の接続を同時に考慮するため、接続属性を表すベクトルとしては次のものを用いる。

$$V(X)=\{V_t(X),V_f(X)\} \quad (7)$$

一方ｔｏ－クラス、ｆｒｏｍ－クラスにおいてはそれぞれ前向き、後向きの接続のみを考慮するため次のベクトルを用いることになる。

$$V(X)=V_t(X) \quad (8)$$

$$V(X)=V_f(X) \quad (9)$$

3．マージコスト $U_{new}-U_{old}$ が最小となるようなクラスのペアを選び、統合して一つのクラスとする。ここで

$$U_{new}=\sum_W P(W)\times D(V(C_{new}(W),V(W)) \quad (10)$$

$$U_{old}=\sum_W P(W)\times D(V(C_{old}(W),V(W)) \quad (11)$$

であり、C_{new} は統合後のクラス、C_{old} は統合前のクラスを表し、$D(V_C,V_W)$ はベクトル V_C と V_W のユークリッド距離の自乗を表すものとする。

4．２までの手順をあらかじめ定められたクラス数になるまで繰り返す。

4．多重クラス２－ｇｒａｍの性能評価

4.1 パープレキシティによる評価

前節で示した手順で求めたクラスを用いて多重クラス２－ｇｒａｍのパープレキシティによる評価を行った。比較対象としては、同じ10前節で求めたクラスを用いた通常のクラス２－ｇｒａｍ、および品詞情報に基づく１５８のクラスを用いたクラス２－ｇｒａｍを用いた。訓練セットは総単語数４５９，３８３語、異なり単語数７，２２１であり、評価は訓練セットに含まれない４１会話６，８２８語の評価セットで行った。なお、語彙に関しては評価セットに現われる総ての単語が含まれている。また多重クラス２－ｇｒａｍに関してはｔｏ－クラスとｆｒｏｍ－クラスの数は必ずしも一致している必要はないが、比較のために同一クラス数とした。図１．に示すように品詞情報に基づくクラス２－ｇｒａｍと自動クラス分類によるクラス２－ｇｒａｍでは同一クラス数でも

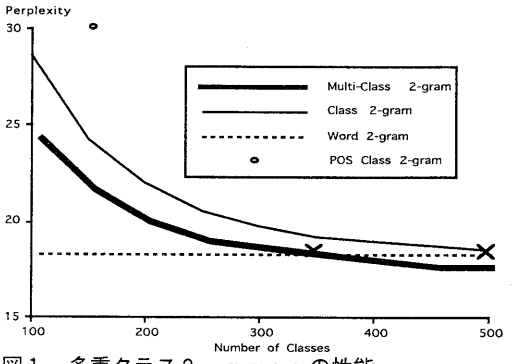

図１．多重クラス２－ｇｒａｍの性能

性能に大きな差がでている。このことからクラス分類の方法としては、品詞情報に基づくよりも実際のコーパス上での接続性に基づく分類の方が効率的であることがわかる。またクラス２－ｇｒａｍと多重クラス２－ｇｒａｍにおいては、単語２－ｇｒａｍと同等のパープレキシティを示す時のクラス数がクラス２－ｇｒａｍで約５００、多重クラス２－ｇｒａｍで約３５０（図中で×で示されている点）とクラス数で約３０％、論理パラメーター数で約半分に削減できていることがわかる。

4.2 連続単語認識による評価

パープレキシティは言語モデルの性能評価のための良い指標であるが、必ずしも単語認識率に直決しない。そこで、連続単語認識における性能評価もあわせておこなった。実験条件は次に示す通りである。

評価セット
　パープレキシティの評価と同じ４１会話
特徴量
　標本化周波数１６ｋ
　フレーム周期１０ｍｓｅｃ
　１２次のメルケプストラム、パワーおよびそれらの一次回帰係数の計２６次元
音響モデル
　ＭＬ－ＳＳＳによる音素環境依存８００状態５混合のＨＭｎｅｔ[5]
　男女依存モデルの自動選択
デコーディング[6]
　１パス時間同期ビタビサーチ
　２パス言語重みを変更した上でのフルサーチ

評価の対象としては単語２－ｇｒａｍ、クラス数５００のクラス２－ｇｒａｍ、そしてクラス数３５０の多重クラス２－ｇｒａｍに対しておこなった。なお、これらのモデルは前段で述べたように、いずれもほぼ同じパープレキシティを示している。評価の基準としては次に示す単語認識率で行った。

$$\frac{W-D-I-S}{W}$$

（Ｗ：正解単語数、Ｄ：脱落誤り数、
Ｉ：挿入誤り数、Ｓ：置換誤り数）

単語認識率は単語２－ｇｒａｍで６９．０５％、クラス２－ｇｒａｍで６９．７８％、多重クラス２－ｇｒａｍで７０．２９％であり、多重クラス２－ｇｒａｍが最も少ないパラメーター数であるにもかかわらず最も高い性能を示している。

5．多重クラス複合Ｎ－ｇｒａｍ

5.1 単語Ｎ－ｇｒａｍの導入

クラス２－ｇｒａｍは訓練データが少ない場合でも少ないパラメーターで頑健なモデルを構築することができるが、精度の面では単語Ｎ－ｇｒａｍに劣る。これに対し、単語Ｎ－ｇｒａｍでは信頼性のあるモデルの構築のためには多量の訓練データを必要とする。しかしながら少量の訓練データで単語Ｎ－ｇｒａｍを構築した場合でも出現回数の多いワードペアあるいはトリオ以上に関しては局所的には単語Ｎ－ｇｒａｍの値に対し信頼性が確保できる。従って出現回数の少ないワードペアに対してはクラス２－ｇｒａｍを用い、出現回数が多く、信頼性が確保できるワードペアあるいはトリオ以上に対しては単語Ｎ－ｇｒａｍを用いるという方法が考えられる。この方法では単語ペア A,B の出現回数が十分である場合、単語 X の直後に単語列 A,B,C が出現する確率は次の式で与えられる。

$$P(A\mid C(A))\times P(C(A)\mid C(X))\times P(B\mid A)\times P(C\mid C(C))\times P(C(C)\mid C(B)) \quad (12)$$

さらに多重クラス２－ｇｒａｍを用いた場合には（１２）式は次のようになる。

$$P(A\mid C_i(A))\times P(C_i(A)\mid C_f(X))\times P(B\mid A)\times P(C\mid C_i(C))\times P(C_i(C)\mid C_f(B)) \quad (13)$$

続いて単語ペア A,B を連鎖語として一つの単語 $A+B$ とするならば（１３）式は次のようになる。

$$P(A+B\mid C_i(A+B))\times P(C_i(A+B)\mid C_f(X))\times P(C\mid C_i(C))\times P(C_i(C)\mid C_f(A+B)) \quad (14)$$

ここで $C_i(A+B)$ は連鎖語 $A+B$ の前にどのような単語が来やすいかを表すクラスであるため、連鎖語の先頭の単語 A のそれに等しいと考えられ

$$C_i(A+B)=C_i(A) \quad (15)$$

とおくことができる。同様に連鎖語の後方向の接続性は最後尾の単語に等しいと考えられ

$$C_f(A+B)=C_f(B) \quad (16)$$

とおくことができる。（１５）、（１６）式を用いれば（１４）式は次のようになる。

$$P(A+B\mid C_i(A))\times P(C_i(A)\mid C_f(X))\times P(C\mid C_i(C))\times P(C_i(C)\mid C_f(B)) \quad (17)$$

（１７）式は単語２－ｇｒａｍが連鎖語の導入により多重クラス２－ｇｒａｍの形式を保ったまま、かつ新たなクラスを導入することなしに表現可能であることを示している。またこの際に増加したパラメーターは連鎖語$A+B$のユニグラムのみである。このことは単語Ｎ－ｇｒａｍの導入に関しても同様に言え、この場合はＮ連鎖語で単語Ｎ－ｇｒａｍが表現されることになる。増加するパラメーターに関しても２－ｇｒａｍの時と同様にＮ連鎖語のユニグラムのみである。Ｎ連鎖語により部分的に単語Ｎ－ｇｒａｍを導入した多重クラス２－ｇｒａｍを多重クラス複合Ｎ－ｇｒａｍと呼ぶことにする。

5.2 多重クラス複合Ｎ－ｇｒａｍの構築手順

多重クラス複合Ｎ－ｇｒａｍは次の手順で構築することができる。

1. 初期状態として多重クラス２－ｇｒａｍを与える。
2. 単語ペアのうち出現回数が一定値以上のものを連鎖語として辞書に加える。この連鎖語のｔｏークラスは先行単語のｔｏークラスと同じ、ｆｒｏｍークラスは後続単語のｆｒｏｍークラスと同じとする。
3. 新たに加わった連鎖語も含めて２．の手順を繰り返す。出現回数が一定値以上のものが存在しない場合は終了する。

5.3 品詞および可変長単語列の複合Ｎ－ｇｒａｍとの比較

クラス２－ｇｒａｍと単語Ｎ－ｇｒａｍの短所を補い合うモデルとしては品詞および可変長単語列の複合Ｎ－ｇｒａｍ[2]がある。品詞および可変長単語列の複合Ｎ－ｇｒａｍは品詞情報に基づくクラス２－ｇｒａｍをベースとし、クラスの中から単語を分離し、独立したクラスにするという操作と分離された単語同士から連鎖語を生成し新たなクラスとするという操作をエントロピーの減少を基準にして繰り返すものである。

品詞および可変長単語列の複合Ｎ－ｇｒａｍは良い性能を示すモデルであるが、次のような問題点があり、多重クラス複合Ｎ－ｇｒａｍではこの問題は解決されていると考えられる。

- クラスから分離された単語からでないと連鎖語を生成することができないため、すでに適切なクラス分類が行われている場合でもクラス分離を行う必要がある。
- クラスから分離された単語は通常十分大きなユニグラム出現数を持つが、バイグラムに対しては必ずしもそうとは言えないため新たにデータスパースの問題を引き起こすことがある。

表１．　品詞および可変長単語列の複合Ｎ－ｇｒａｍとの相違点

	品詞および可変長単語列の複合Ｎ－ｇｒａｍ	多重クラス複合Ｎ－ｇｒａｍ
分離の対象	クラスから単語	クラス２－ｇｒａｍから単語２－ｇｒａｍ
分離の基準	エントロピーの減少	単語ペアの出現回数
Ｎ－ｇｒａｍの表現単位	クラス２－ｇｒａｍ クラス－単語２－ｇｒａｍ 単語－クラス２－ｇｒａｍ 単語Ｎ－ｇｒａｍ	クラス２－ｇｒａｍ 単語Ｎ－ｇｒａｍ
パラメーター数の増分	分離単語数＋単語列数の自乗	連鎖語数

多重クラス複合Ｎ－ｇｒａｍと品詞および可変長単語列の複合Ｎ－ｇｒａｍとの違いをまとめると表－１のようになる。このうちＮ－ｇｒａｍの表現単位に関しては品詞および11変長単語列の複合Ｎ－ｇｒａｍの方が豊富であるが、そのうちどれが用いられるかは前後に来る単位がクラスか単語か単語列で決まってしまうため自由度という観点からは必ずしも高いとは言えない。また、分離の基準に関しては品詞および可変長単語列の複合Ｎ－ｇｒａｍで出現回数を基準とすることも、多重クラス複合Ｎ－ｇｒａｍでエントロピーを基準とすることも可能なため本質的な差ではない。

5.4 多重クラス複合Ｎ－ｇｒａｍの性能評価

まず多重クラス複合Ｎ－ｇｒａｍの性能評価としてパープレキシティによる評価を行った。多重クラス複合Ｎ－ｇｒａｍの初期状態としては４節で用いた多重クラス２－ｇｒａｍを用い、連鎖語導入のための出現回数は２０回と設定した。４節と同じ訓練セット、評価セットを用いた時の結果を図－２に示す。また比較対象として同一条件における単語２－ｇｒａｍ、単語３－ｇｒａｍ、そして初期品詞クラス１５８、分離クラス１０００の品詞および可変長単語列の複合Ｎ－ｇｒａｍの結果も同時に示す。

この結果からクラス数４００においてほぼ単語３－ｇｒａｍに近い性能を示していることがわかる。連鎖語導入のための出現回数が２０回の条件で導入された連鎖語の総数は２，

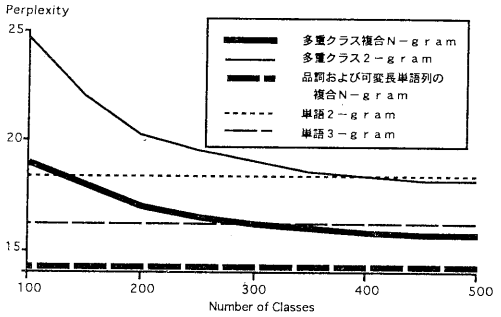

図２．　多重クラスＮ－ｇｒａｍの
パープレキシティ

表２．　多重クラス複合Ｎ－ｇｒａｍの性能

モデルの種類	論理パラメーター数	2－ｇｒａｍ以上のエントリ数	Perplexity	認識率
200 クラス	5.0×10^{4}	5.7×10^{3}	17.54	76.30
400 クラス	1.7×10^{5}	1.4×10^{4}	16.29	75.79
単語2－ｇｒａｍ	5.5×10^{7}	6.1×10^{4}	18.52	68.95
単語3－ｇｒａｍ	3.5×10^{11}	2.0×10^{5}	15.93	
品詞および可変長単語列の複合N－ｇｒａｍ	1.4×10^{6}	5.2×10^{4}	14.84	75.51

２１２単語であり、これらの連鎖語の訓練セットにおける出現回数、すなわち単語Ｎ－ｇｒａｍの使用回数は１１６，５２５で、これは全体の約２０％を占める。またこのうち３単語以上の連鎖語の出現回数、すなわち単語３－ｇｒａｍ以上の使用回数は６０，５２９回で、全体の約１０％である。続いて同様の条件のもとでのクラス数２００および４００の多重クラス複合Ｎ－ｇｒａｍの単語認識率による評価結果を表２に示す。多重クラス複合Ｎ－ｇｒａｍクラス数２００の場合においても単語２－ｇｒａｍに比べパープレキシティ、認識率ともに上回っており、その際に必要は論理パラメーター数はわずか千分の一以下となっている。またエントリサイズにおいても十分の一以下となっている。品詞および可変長単語列の複合Ｎ－ｇｒａｍとの比較においてもパープレキシティでは劣るものの４％以下の論理パラメーター数での同等以上の認識性能を示している。

６．まとめ

本稿ではクラスＮ－ｇｒａｍのための効率的クラス分類の方法として各単語に対して接続の方向性ごとに複数のクラスを割り当てる多重クラスを提案した。多重クラスを利用した多重クラス２－ｇｒａｍにおいては単語間の接続のみに着目したクラス分類を行っているため、パープレキシティ、単語認識率を劣化させることなく非常にコンパクトなサイズでモデルを表現することができる。

また多重クラス２－ｇｒａｍに対して部分的に単語Ｎ－ｇｒａｍを導入した多重クラス複合Ｎ－ｇｒａｍを提案した。多重クラス複合Ｎ－ｇｒａｍにおいては出現回数が多く、信頼性の高い単語Ｎ－ｇｒａｍが連鎖語の形で表現されており、クラス２－ｇｒａｍの頑健さと単語Ｎ－ｇｒａｍの精度を兼ね備えたモデルとなっている。また形式的にはあくまでも多重クラス２－ｇｒａｍの形をしているためデコーダにとっても非常に扱いやすいモデルとなっている。本モデルは一つの連鎖語（すなわち単語Ｎ－ｇｒａｍ）の導入にあたって増加するパラメーターはそのユニグラムのみであるため、多重クラス２－ｇｒａｍと同等のモデルサイズでより高性能がえられる。実験結果においては単語２－ｇｒａｍに対し論理パラメーター数で千分の一以下のサイズでより高いパープレキシティ、単語認識率が得られ、品詞および可変長単語列の複合Ｎ－ｇｒａｍとの比較においても４％程度の論理パラメーター数で同等以上の単語認識率を得ることができた。

参考文献

[1] Shuanghu Bai, Haizhou Li, Zhiwei Lin, Baosheng Yuan :"Building Class-based Language Models with Contextual Statistics, "Proc.ICASSP, vol. 1, pp. 173-176, 1998

[2] 政瀧浩和, 松永昭一, 匂坂芳典, "品詞および可変長単語列の複合Ｎ－ｇｒａｍの自動生成," 信学論, vol.J81-D-II, no.9, pp.1929-1936, Sep. 1998.

[3] P.F.Brown et al :"Class-Based n-gram Models of Natural Language, "Computational Linguistics,vol. 18, No. 4, pp. 467-479, 1992

[4] Sabine Deligne, Frederic Bimbot :"I anguage modeling by variable length sequences, "Proc ICASSP, vol. 1, pp. 169-172, 1995

[5] M.Ostendorf, H.Singer :"HMM topology design using maximum likelihood successive state splitting, " Computer Speech and Language, vol. 11, pp. 17-41, 1997

[6] Tohru Shimizu, Hirofumi Yamamoto, Hirokazu Masataki, Shoichi Matsunaga, Yoshinori Sagisaka : "Spontaneous Dialogue Speech Recognition Using Cross-Word Context Constrained Word Graphs," Proc.ICASSP, vol. 1, pp. 145-148, 1996