

Internal Use Only

TR-I-0148

HMM音韻認識における音節連鎖統計情報の利用

HMM Phoneme recognition Using Syllable Trigrams

花沢 利行, 川端 豪†, 伊藤 克亘††, 鹿野 清宏†††

Toshiyuki Hanazawa, Takeshi Kawabata,
Katsunobu Itoh and Kiyohiro Shikano

1990.2

内容梗概

音節の連鎖統計情報をHMM法による音韻認識と組み合わせることによって、語彙の仮定なしに、入力音声を音韻系列に変換する「音韻タイプライタ」の構成を試みた。音節の連鎖統計情報を用いることによって、文節音声中の音韻認識率が大きく改善されることを確認した。

ATR自動翻訳電話研究所

ATR Interpreting Telephony Reseach Laboratories

© (株) ATR自動翻訳電話研究所 1988

©1988 by ATR Interpreting Telephony Reseach Laboratories

†現在NTT基礎研究所, ††東京工業大学修士課程,

†††現在NTTヒューマンインターフェース研究所

目次

1. はじめに	1
2. システムの構成	2
2.1 HMM音韻モデル	2
2.2 音節連鎖の確率モデル	2
2.3 音韻タイプライタの構成	4
3. 言語モデルの評価尺度	6
4. 言語モデル作成用テキストデータベース	7
5. 評価用音声データベース	7
6. 言語モデルの評価実験	8
7. 音声認識実験	10
8. むすび	11
文献	12
付録	13
音韻タイプライタの出力結果 (全テキストデータベースを用いて言語モデルを作成した場合)	14
実験で用いた音節の種類	27
音節のパープレキシティ	28
音韻タイプライタ用の音節連鎖文法	29

1. はじめに

人間の発声した音声を言語的な情報なしに、機械によって正しく認識することは現在の音声認識技術レベルでは困難である。そこで通常の音声システムでは認識できる語彙を限定したりタスクに文法を仮定することで、認識候補を絞ることが行われている。この場合には辞書に登録されていない未知語や、文法上で許されない発声をした場合には認識が不可能となる問題がある。

一方、文節音声認識候補の絞り込みに、日本語音節連鎖の確率モデルが有効であることが、荒木ら¹⁾によって音節音声入力のシミュレーションによって示されている。音節連鎖確率モデルの利点は、音節連鎖の局所的な統計情報のみを用いるので、認識語彙を限定することなしに認識候補が絞れる点にある。荒木らの実験では、音節区間の境界は正しいものとして検討を行っている。実際の音声認識システムでは100%の境界検出を行うことは困難であり、この場合に対しては音節連鎖の確率モデルがどの程度有効であるかは明らかにされていない。

そこで本報告では、HMM (Hidden Markov Model)²⁾による音韻認識手法と音節連鎖統計情報による候補絞り込みに基づいて実際の音声認識システムを構成し、語彙の仮定なしに文節音声中の音韻を認識する「音韻タイプライタ」として動作させることを試みる。

音節連鎖の統計はATRで収集されたテキストデータベースを用いて行う。音節連鎖のモデルとしては0-gram から3-gram までを用いる。また削除補間法 (Deleted Interpolation)³⁾によるそれらの重みづけの検討も行う。

音韻認識部はHMM-LR連続音声認識システム⁴⁾⁵⁾を用い、語彙情報を含まず日本語の音韻構造を記述する簡単な文法を与えて動作させる。HMM音韻モデルは複数コードブックと継続時間制御を持つもの⁵⁾を用いる。

システムの評価は文節音声中の音韻認識実験によって行う。また参考のために語彙の仮定なしに得られた文節認識率についても述べる。

2. システムの構成

2.1 HMM音韻モデル

音韻認識に用いるHMMは、文献5)のものと同一である。

モデルの学習には、特定話者の発声した5240単語及び音韻バランス216単語中6)の音韻を用いる。音声試料は12kHzで標本化され、21.3msのHamming窓を用いて、フレーム(9ms)ごとに12次LPC分析される。続いて、複数のコードブックの各々を用いてVQコード列に変換する。VQコードブックは次の3種である。

WLR : スペクトル距離尺度7)

POW : パワー(スカラー)8)

DCEP : LPCケプストラムの時間差分(約50ms巾)9)

このVQコード列を学習データとして、各音韻モデルの学習を行う。音韻モデルとしては、子音に対しては3ループ、母音及び撥音に対しては1ループのHMMを用いる(図1)。また、同じ学習データを用いて、HMMの各状態の継続時間の分布を調べ、正規分布で近似しておく。認識時には、この分布を発声速度の変化に応じて補正し10)、HMMの状態の継続時間長制御を行う。

2.2 音節連鎖の確率モデル

音節連鎖の確率モデルとしてn-gramを用いる。音節系列 s_0, \dots, s_{N-1}, s_N に対するn-gramモデル L は(1)式のように定義される。

$$L = P(s_i | s_{i-(n-1)}, s_{i-(n-2)}, \dots, s_{i-1}) \quad (1)$$

もし、学習用テキストデータベースが無限にあれば、gram数が大きいほど利用できる情報量が増えるので強い絞り込み効果が期待できる。しかし、現実には有限のデータベースから確率値を求める場合には、gram数の増加に伴って確率値の推定精度が低下する。そこで現実的な方法として、0-gram(全ての音節が等確率で生成されるモデル)から3-gramまでをそれぞれの推定値に推定精度に応じた重みをつけ、それらを加算して用いることが考えられる。この重み係数を決定する方法として、削除補間法3)(Deleted Interpolation)を用いる。

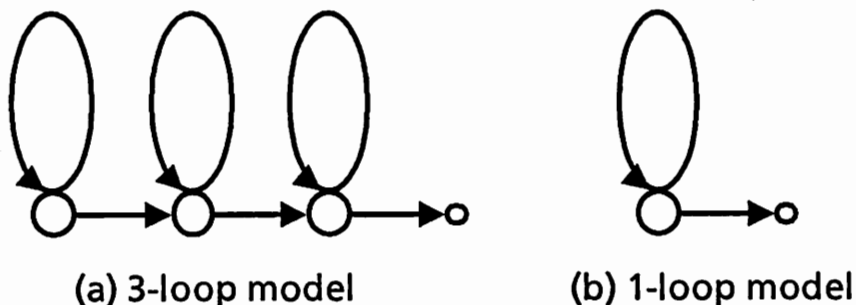


Fig.1 HMMs for phoneme recognition

[削除補間法]

言語モデル L を、音節連鎖の0-gram から 3-gramモデルまでの荷重平均として(2)式のように設定する。

$$L = \lambda^{(0)}P^{(0)} + \lambda^{(1)}P^{(1)} + \lambda^{(2)}P^{(2)} + \lambda^{(3)}P^{(3)} \quad (2)$$

$P^{(n)}$: n-gramモデル

$\lambda^{(n)}$: n-gramモデルに対する重み係数

解くべき問題は、 $\sum \lambda_k = 1$ の条件下で言語モデル L から、評価データが生成される確率、すなわち(3)式を最大化する λ_k を決めることである(最尤推定)。

$$P = \prod_{j=1}^N p(b_j) \quad (3)$$

$p(b_j)$: 言語モデル L からテキストデータ中の文節 b_j ($j=1 \sim N$)が生成される確率

テキストデータが文節 b_j ($j=1 \sim N$)の集合からなるとき、全文節を用いて音節連鎖のn-gramモデルを推定すると、式を最大化する λ_k は $\lambda_0 = \lambda_1 = \lambda_2 = 0, \lambda_3 = 1$ となる。これは学習と評価データが同じ場合には gram 数が大きいほどよいという事実と対応するが、この λ_k の値は学習外データに対しての最適値とはならない。

そこで、文節 b_j の生起確率を計算するときは、文節 b_j を除いたテキストデータから求めた連鎖確率値を用いることにする(削除補間法)。この条件下では、学習外データに対する推定を含むことになる。以下に、重み係数 $\lambda^{(k)}$ の推定手順を示す。

- ① 各文節 ($j=1 \dots N$) に対し②③を繰り返す。
- ② テキストデータから文節 b_j を削除し、0-gram~3-gram ($P^{(n)}, n = 0 \sim 3$) 確率を求める。
- ③ 文節 b_j の音節系列を $s_0, s_1, s_2, \dots, s_{K_j}$ とするとき、次式にしたがって n-gram の寄与率を計算する。

$$c_i^{(n)}(b_j) = \frac{\lambda^{(n)} P_i^{(n)}}{\lambda^{(0)} P_i^{(0)} + \lambda^{(1)} P_i^{(1)} + \lambda^{(2)} P_i^{(2)} + \lambda^{(3)} P_i^{(3)}}, \quad i = 0 \sim K_j \quad (4)$$

- ④ 寄与率 $c^{(n)}$ を i, j について平均し、 $\lambda^{(n)}$ を更新する。

$$new \lambda^{(n)} = \frac{\sum_{j=1}^N \sum_{i=1}^{K_j} c_i^{(n)}(b_j)}{\sum_{j=1}^N K_j} \quad (5)$$

これらの操作を、生成確率((3)式)が収束するまで繰り返す。

[直前の2つ組ごとに重み係数 λ を設定する削除補間法]

上述した削除補間法では、すべてのコンテキストで同じ係数 λ を用いたが、テキストデータ中での出現頻度の高い2つ組の次にくる音節に関する3-gramについては推定精度が高く、逆に出現頻度の低い2つ組の次にくる音節に関する3-gramについては推定精度が低いと考えられる。そこで式のように直前の2つ組ごとに重み係数 λ を設定する。

$$L = \lambda^{(0)}(i-1, i-2)P^{(0)} + \lambda^{(1)}(i-1, i-2)P^{(1)} + \lambda^{(2)}(i-1, i-2)P^{(2)} + \lambda^{(3)}(i-1, i-2)P^{(3)} \quad (6)$$

$\lambda^{(n)}(i-1, i-2)$: 直前の2つ組ごとのn-gramモデルに対する重み係数

簡単のため以後、このモデルをコンテキスト別削除補間法と呼ぶことにする。

[直前の2つ組の出現頻度ごとに重み係数 λ を削除補間法]

コンテキスト別削除補間法では2つ組の数だけ λ を推定する必要があり、テキストデータベースが十分でない場合には、 λ の推定精度が低くなる可能性がある。そこで2つ組の出現頻度がある範囲内で等しいものは式のように共通の重み係数 λ を設定する。

$$L = \lambda^{(0)}(F)P^{(0)} + \lambda^{(1)}(F)P^{(1)} + \lambda^{(2)}(F)P^{(2)} + \lambda^{(3)}(F)P^{(3)} \quad (7)$$

$\lambda^{(n)}(F)$: 直前の2つ組の出現頻度ごとのn-gramモデルに対する重み係数

出現頻度で分ける範囲として、以下の3つを設定した。

- $F = 0$: $0 \leq$ 直前の2つ組の出現頻度 < 10
- $F = 1$: $10 \leq$ 直前の2つ組の出現頻度 < 100
- $F = 2$: $100 \leq$ 直前の2つ組の出現頻度

簡単のため以後、このモデルを出現頻度別削除補間法と呼ぶことにする。

2.3 音韻タイプライタの構成

HMM-LR連続音声認識システム4)5)上に、音韻タイプライタを構築する。通常、HMM-LRシステムは語彙情報を含んだ文脈自由文法を与えて動作させるわけであるが、これを語彙情報を含まず日本語の音韻構造(例えば「子音が連続しない」「撥音は語頭にこない」等)を記述する簡単な文法と置き換えることによって、音韻タイプライタとしても動作させることができる。

認識を進めてゆく過程で、音韻の照合スコアはHMM音韻モデルからの尤度と、音節連鎖n-gramモデルからの尤度を荷重平均して求める。

HMM-LRシステムには、ビーム探索機構が組み込まれており、同時に保持される中間結果は、ある個数(ビーム幅 = 250)に制限されている。すなわち、認識の各段階で、各々の中間結果に対し次式の確率値を計算し、値の大きいものを候補として残していく。

[枝刈りのための中間(対数)確率値]

$$\ln P^{(total)} = 0.75 \times \ln P^{(hmm)} + 0.25 \times \ln P^{(gram)} \quad (8)$$

$P(hmm)$: HMMによって計算される確率

$P(gram)$: 音節連鎖情報から計算される確率

である。 $P(gram)$ は、その時点までの音節連鎖を s_0, \dots, s_{i-1}, s_i とするとき、次の手続きによって漸化的に計算される。

(処理する音韻が母音のとき)

$$\ln P_i^{(gram)} = \frac{(i-1)\ln P_{i-1}^{(gram)} + \ln P(s_i | \dots, s_{i-2}, s_{i-1})}{i} \quad (9)$$

(処理する音韻が子音のとき)

$$\ln P_i^{(gram)} = \frac{(i-1)\ln P_{i-1}^{(gram)} + \sum_s \ln P(s_i | \dots, s_{i-1})}{i} \quad (10)$$

但し、上式の Σ は、処理対象となっている子音を含むすべての音節に対する総和を表わすものとする。

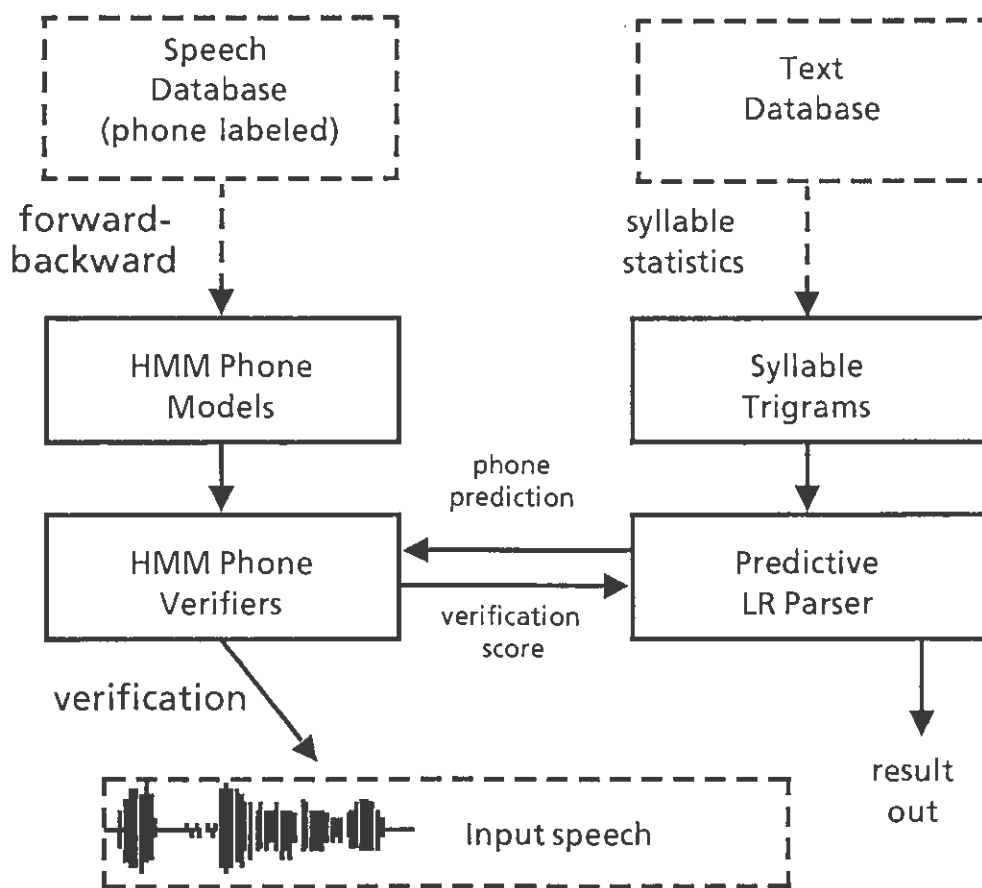


Fig.2 Schematic diagram of the phonetic typewriter

3. 言語モデルの評価尺度

本報告では、言語モデルの評価尺度として以下のものを用いる。

3.1 Coverage

評価データの音節連鎖を言語モデルがカバーする割合。

3.2 評価データ生成確率

評価データ生成確率 P_{test} は式のように計算される。

$$P_{test} = \frac{1}{N_s} \sum_{j=1}^{N_b} \log_2(p(b_j)) \quad (11)$$

N_s : 評価データ中の音韻数(各文節の最後の無音を1つとして数える。)

N_b : 評価データ中の文節数

$p(b_j)$: 評価データ中の i 番目の文節に対する言語モデルの確率値

評価データ生成確率 P_{test} は言語モデルから、評価データが生成される1音韻あたりの対数確率であり、この値が大きいほどよいモデルであるといえる。

3.3 評価データパープレキシティ(11)

$$F_{test} = 2^{-P_{test}} \quad (12)$$

評価データパープレキシティ F_{test} は式のように計算される。

P_{test} : 評価データ生成確率式

評価データパープレキシティは通常のパープレキシティとは定義が異なるが、評価データを解析する過程で言語モデルによって、予測される平均音韻数であると考えることができる。

4. 言語モデル作成用テキストデータベース

音節連鎖の統計は、ATRで収集されたテキストデータベースを用いて行う。このデータベースは書き言葉(雑誌, 新聞, 論説文)、話し言葉(キーボード会話, 電話会話)の両方を含んでおり、約32万音節からなっている。表1に出典と音節数を示す。但し本報告では短母音と長母音を別の音節と考えている。すなわち、/aa/, /ee/, /moo/等を1音節として扱っている。文節の区切り記号を1音節と考えると、音節の種類は259である。

5. 評価用音声データベース

言語モデルの評価と音声認識実験には、ATRで収集している音声データベース6)から成人男性1名(MAU)発声した25文章(279文節)を用いる。この文章は国際会議の参加登録に関する問い合わせを想定して収録されたもので、各文章とも文節ごとに区切って発声されている。文節発声の例を図2に示す。1文節あたりの平均音韻数は7.4である。

またこの評価データは、言語モデルを作成したテキストデータベース中でキーボード会話, 電話会話と近い発話内容である。

表1 言語モデル作成用テキストデータベース

出典	音節数
1. 雑誌記事	34,308
2. 朝日新聞「声の欄」	38,050
3. 論説文	42,449
4. キーボード会話	53,998
5. 電話会話	154,015
計	322,819

表2 評価用データの文節発声例

「第一回」「通訳」「電話」「国際」「会議に」
「参加の」「登録を」「ご希望される」「方は」
「所定の」「申込み」「用紙に」「住所」「氏名と」
「発表」「聴講の」「別を」「明記して」「国際」
「会議」「事務局まで」「お申し込み」「下さい」

[音響分析]

これらの音声は、2.1節で述べたHMMの学習データと同様に12kHzで標本化され、21.3msのHamming窓を用いて、フレーム(9ms)ごとに12次LPC分析される。各フレームの分析結果を、2.1節で述べた複数のコードブックを用いてファジィVQ(12)コードに変換する。ファジィVQを行う理由は、話者適応(13)(14)を行った場合でも認識ができるようにするためである。

6. 言語モデルの評価実験

4章で述べたテキストデータベースを用いて種々の言語モデルを作成し、5章で述べた文節データに対して評価を行った結果を表3に示す。表中でcoverageが100%でない言語モデルに対しては、音韻連鎖の生成確率が0になる場合があるので、評価データ生成確率は音節生成確率の最低値を 10^{-5} として計算した。

この表から3-gram等の言語モデルを用いることによって、評価データ生成確率が大きくなっている、すなわち評価データパープレキシティが小さくなっており、音韻候補の絞り込みが有効になされていることがわかる。

削除補間法を用いることによってさらに生成確率が大きくなっており、その効果が確認できる。2つ組の出現頻度に応じて重み係数を推定した削除補間法では、僅かではあるが生成確率が大きくなっている。このときの重み係数の値を表3に示す。直前の2つ組の出現頻度が小さい場合には、3-gramの重み係数は小さく、出現頻度が大きくなるにつれて3-gramの重み係数も大きくなっており、期待した結果が得られている。

一方2つ組のコンテキストごとに重み係数を推定した削除補間法では、すべてのコンテキストで同じ重み係数を用いた削除補間法よりも若干生成確率が小さくなっているが、これはコンテキストごとに重み係数を推定する場合、推定すべきパラメータの数が多くなり、推定精度が低下したためだと思われる。

学習データの違いに対する影響をみると、雑誌、新聞、論説文のみで言語モデルを作成した場合、評価データに対する3-gramモデルのcoverageは86%と低い値であり、評価データ生成確率も低くなっている。これは認識のタスクとかなり異なるテキストデータベースを用いて言語モデルを作成したためと思われる。この場合には削除補間法による効果が大きくなっている。また表3に示すように全データを用いて言語モデルを作成した場合に比べて、3-gramに対する重み係数が小さくなり、2-gramが補っていることがわかる。

表3. 種々の言語モデルの特性

学習データ	言語モデル	coverage(%)	評価データ 生成確率 (log ₂)	評価データ パープレキシティ
——	0-gram	100	-4.81	28.1
全データ	1-gram	100	-3.36	10.3
	2-gram	98.9	-2.55	5.87
	3-gram	96.4	-2.08	4.24
	0,1,2,3-削除補間	100	-1.98	3.94
	コンテキスト別 0,1,2,3-削除補間	99.8	-2.00	4.01
	出現頻度別 0,1,2,3-削除補間	100	-1.96	3.89
キーボード &電話	3-gram	95.1	-2.12	4.35
	0,1,2,3-削除補間	100	-1.98	3.96
雑誌&新聞 &論説	3-gram	85.5	-3.30	9.85
	0,1,2,3-削除補間	100	-2.68	6.41

注) coverageが100%でない言語モデルに対しては、音韻連鎖の生成確率が0になる場合があるので、評価データ生成確率は、音節生成確率の最低値を10⁻⁵として計算した。

表4. 種々の削除補間法における重み係数 λ の値

学習データ	言語モデル	直前の2つ組 の出現頻度	重み係数 λ			
			$\lambda^{(0)}$	$\lambda^{(1)}$	$\lambda^{(2)}$	$\lambda^{(3)}$
全データ	0,1,2,3-削除補間	共通の重み係数	0.001	0.022	0.062	0.915
	出現頻度別 0,1,2,3-削除補間	0 ≤ 頻度 < 10	0.002	0.021	0.458	0.518
		10 ≤ 頻度 < 100	0.001	0.053	0.071	0.875
		100 ≤ 頻度	0.001	0.013	0.014	0.971
雑誌&新聞 &論説	0,1,2,3-削除補間	共通の重み係数	0.002	0.043	0.145	0.810

7. 音声認識実験

前章で述べたいくつかの言語モデルを音声認識システムに組み込んで、6章で述べた文節音声の中の音韻認識実験を行った。

性能評価は音韻認識率によって行った。音韻認識率は、音韻タイプライタから出力された音韻系列と正しい音韻系列とのDPマッチングを行い、音韻の置換・挿入・脱落誤りの個数を求め、(13)式により計算した。

$$\text{音韻認識率} = \frac{\text{全音韻数} - \text{置換} - \text{挿入} - \text{脱落}}{\text{全音韻数}} \quad (13)$$

表5に認識実験の結果を示す。

全テキストデータを用いて言語モデルを作成した場合、音節 3-gram を用いることによって音韻認識率が73.2%から94.0%に向上することがわかる。また削除補間法を用いることによって94.9%、直前の2つ組の出現頻度に応じて重み係数を設定する削除補間法では95.1%の音韻認識率が得られた。直前の2つ組のコンテキ

表5. 音韻タイプライタの性能

学習データ	言語モデル	音韻認識率(%)	文節の累積認識率(%)	
			1	~5
———	0-gram	73.2	21.1	37.6
全データ	3-gram	94.0	76.7	88.5
	0,1,2,3-削除補間	94.9	78.5	92.1
	コンテキスト別 0,1,2,3-削除補間	94.8	78.9	91.8
	出現頻度別 0,1,2,3-削除補間	95.1	79.2	91.8
雑誌&新聞 &論説	3-gram	82.8	46.2	58.8
	0,1,2,3-D.I.	84.6	51.3	70.6

$$\ln P = 0.75 \times \ln P(\text{hmm}) + 0.25 \times \ln P(\text{ngram}) \quad \text{beam} = 250$$

ストごとに重み係数を設定する削除補間法では、音韻認識率は94.8%であり、コンテキスト共通の重み係数を用いる削除補間法からの改善はみられなかった。これらの結果は前節の言語モデルの評価結果とほぼ一致している。

一方、雑誌、新聞、論説文のみで言語モデルを作成した場合、テキストデータベース全体を用いた場合と比べて認識率が低下する。これは、①認識のタスクとかなり異なるテキストデータベースを用いて言語モデルを作成した、②言語モデル作成のためのテキストデータの不足等の理由が考えられる。実際、表3に示したように評価データに対する3-gramモデルのcoverageは85.5%と低い値であった。しかしこの場合にも確率モデルを用いないときよりも認識率が向上しており、削除補間法を用いることによってさらに認識率が向上することがわかる。

以上の検討から音節連鎖の確率モデルが、実際の音声認識システムに組込んだ場合でも有効であることが確認できた。但し今回の実験では言語モデルを作成するためのテキストデータベースが十分とはいえ、言語モデルの作成に必要なテキストデータ数の量、言語モデルのタスク依存性等について調べるにはより多くのテキストデータベースが必要と思われる。

8. むすび

音節の連鎖統計情報をHMM法による音韻認識と組み合わせることによって、語彙の仮定なしに、入力音声を音韻系列に変換する「音韻タイプライタ」の構成を試みた。音節の連鎖統計情報を用いることによって、文節音声の中の音韻認識率が大きく改善されることを確認した。

但し今回の実験では言語モデルを作成するためのテキストデータベースが十分とはいえ、言語モデルの作成に必要なテキストデータ数の量、言語モデルのタスク依存性等について調べるにはより多くのテキストデータベースが必要と思われる。

謝辞

研究の機会を与えて頂いたATR自動翻訳電話研究所 樽松明 社長に感謝いたします。認識実験を手伝って頂いた早稲田大学修士過程 北川英一郎氏に感謝します。また、日頃、熱心に討論頂くATR自動翻訳電話研究所の皆さまに感謝します。

文献

- 1) 荒木 哲郎, 村上 仁一, 池原 悟 : “2重音節マルコフモデルによる文節音節認識候補の曖昧さの解消効果”, 情報処理学会論文誌, Vol.30, No.4, PP467-477 (1989)
- 2) S.E. Levinson, L.R. Rabiner, M.M. Sondhi : “An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition”, BSTJ, Vol. 62, No. 4, pp.1035-1074 (1983)
- 3) F. Jelinek and R. Mercer : “Interpolated Estimation of Markov Source Parameters from Sparse Data”, in Pattern Recognition in Practice, E.S. Gelsema and L.N. Kanal, North Holland, pp381-397 (1980)
- 4) 北 研二, 川端 豪, 斎藤 博昭 : “HMM音韻認識と予測LRパーザを用いた文節認識”, 信学技報SP88-88, pp63-69 (1988)
- 5) 花沢 利行, 北 研二, 中村 哲, 川端 豪, 鹿野 清宏 : “HMM-LR音声認識システムの性能評価”, 信学技法, SP89-94, pp63-70 (1989)
- 6) (1985)武田 一也, 匂坂 芳典, 片桐 滋, 桑原 尚夫 : “研究用日本語音声データベースの構築”, 音響学会誌 Vol.44 No.10, pp747-754, (1988-10)
- 7) 杉山 雅英, 鹿野 清宏 : “ピークに重みを置いたLPCスペクトルマッチング尺度”, 信学論A Vol. J64-A No.5, pp409-416 (1981)
- 8) 相川 清明, 杉山 雅英, 鹿野 清宏 : “パワー情報とスペクトル情報を用いたベクトル量子化による単語音声認識”, 音声研究会資料S82-61, pp481-488 (1982-12)
- 9) K. Shikano, : “Evaluation of LPC Spectral Matching Measures for Phonetic Unit Recognition”, Carnegie Mellon University, (1986)
- 10) T. Kawabata, K Shikano : “Island-driven Continuous Speech Recognizer Using Phone-based HMM Word Spotting”, ICASSP89, Vol. S1, S9.7, pp.461-464 (1989)
- 11) F. Jelinek : “Self-organized Language Modeling for Speech recognition”, IBM T.J. Watson Research Center, Unpublished
- 12) H.P. Tseng, M.J. Sabin and E.A. Lee : “Fuzzy Vector Quantization Applied to Hidden Markov Modeling”, ICASSP87, pp 641-644, (1987)
- 13) 中村 哲, 鹿野 清宏, : “ファジィベクトル量子化を用いたスペクトログラムの正規化の検討”, 信学技報 SP87-123, pp7-12 (1988-02)
- 14) 中村 哲, 花沢 利行, 鹿野 清宏 : “ファジィベクトル量子化に基づく話者適応化のHMM音素認識による評価”, 音講論集 2-P-20, pp241-242 (1988-10)

付録

[音韻データベース出力結果]

----- No. 1 -----
input: d-a-i-i-Q-k-a-i
3-gram: d-a-i-i-Q-k-a-i
0,1,2,3-D.I.: d-a-i-i-Q-k-a-i

----- No. 2 -----
input: ts-u-u-y-a-k-u
3-gram: ts-u-u-y-a-k-u
0,1,2,3-D.I.: ts-u-u-y-a-k-u

----- No. 3 -----
input: d-e-N-w-a
3-gram: d-e-N-w-a
0,1,2,3-D.I.: d-e-N-w-a

----- No. 4 -----
input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 5 -----
input: k-a-i-g-i-n-i
3-gram: k-a-i-g-i-n-i
0,1,2,3-D.I.: k-a-i-g-i-n-i

----- No. 6 -----
input: s-a-N-k-a-n-o
3-gram: s-a-N-k-a-n-o
0,1,2,3-D.I.: s-a-N-k-a-n-o

----- No. 7 -----
input: t-o-u-r-o-k-u-o
3-gram: t-o-u-r-o-k-u-o
0,1,2,3-D.I.: t-o-u-r-o-k-u-o

----- No. 8 -----
input: g-o-k-i-b-o-u-s-a-r-e-r-u
3-gram: g-o-k-i-b-o-u-s-a-r-e-r-u
0,1,2,3-D.I.: g-o-k-i-b-o-u-s-a-r-e-r-u

----- No. 9 -----
input: k-a-t-a-w-a
3-gram: k-a-t-a-w-a
0,1,2,3-D.I.: k-a-t-a-w-a

----- No. 10 -----
input: sy-o-t-e-i-n-o
3-gram: sy-o-t-e-n-o
0,1,2,3-D.I.: sy-o-t-e-n-o

----- No. 11 -----
input: m-o-u-sh-i-k-o-m-i
3-gram: m-o-u-sh-i-k-o-m-i
0,1,2,3-D.I.: m-o-u-sh-i-k-o-m-i

----- No. 12 -----
input: y-o-u-sh-i-n-i
3-gram: y-o-u-sh-i-n-i
0,1,2,3-D.I.: y-o-u-sh-i-n-i

----- No. 13 -----
input: zy-u-u-sy-o
3-gram: zy-u-u-sy-o
0,1,2,3-D.I.: zy-u-u-sy-o

----- No. 14 -----
input: sh-i-m-e-i-t-o
3-gram: sh-i-m-e-i-t-o
0,1,2,3-D.I.: sh-i-m-e-i-t-o

----- No. 15 -----
input: h-a-Q-py-o-u
3-gram: h-a-Q-py-o-u
0,1,2,3-D.I.: h-a-Q-py-o-u

----- No. 16 -----
input: cy-o-u-k-o-u-n-o
3-gram: cy-o-u-k-o-u-n-o
0,1,2,3-D.I.: cy-o-u-k-o-u-n-o

----- No. 17 -----
input: b-e-t-s-u-o
3-gram: b-e-t-s-u-o
0,1,2,3-D.I.: b-e-t-s-u-o

----- No. 18 -----
input: m-e-i-k-i-sh-i-t-e
3-gram: m-e-i-k-i-sh-i-t-e
0,1,2,3-D.I.: m-e-i-k-i-sh-i-t-e

----- No. 19 -----
input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 20 -----
input: k-a-i-g-i
3-gram: k-a-i-g-i
0,1,2,3-D.I.: k-a-i-g-i

----- No. 21 -----
input: z-i-m-u-ky-o-k-u-m-a-d-e
3-gram: z-i-m-u-ky-o-k-u-m-a-d-e
0,1,2,3-D.I.: z-i-m-u-ky-o-k-u-m-a-d-e

----- No. 22 -----
input: o-m-o-u-sh-i-k-o-m-i
3-gram: o-m-o-u-sh-i-k-o-m-i
0,1,2,3-D.I.: o-m-o-u-sh-i-k-o-m-i

----- No. 23 -----

input: k-u-d-a-s-a-i
3-gram: k-u-d-a-s-a-i
0,1,2,3-D.I.: k-u-d-a-s-a-i

----- No. 24 -----

input: h-a-i
3-gram: h-a-i
0,1,2,3-D.I.: h-a-i

----- No. 25 -----

input: k-o-ch-i-r-a-w-a
3-gram: k-o-ch-i-r-a-w-a
0,1,2,3-D.I.: k-o-ch-i-r-a-w-a

----- No. 26 -----

input: d-a-i-i-Q-k-a-i
3-gram: d-a-i-i-Q-k-a-i
0,1,2,3-D.I.: d-a-i-i-Q-k-a-i

----- No. 27 -----

input: ts-u-u-y-a-k-u
3-gram: ts-u-u-y-a-k-u
0,1,2,3-D.I.: ts-u-u-y-a-k-u

----- No. 28 -----

input: d-e-N-w-a
3-gram: d-e-N-w-a
0,1,2,3-D.I.: d-e-N-w-a

----- No. 29 -----

input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 30 -----

input: k-a-i-g-i
3-gram: k-a-i-g-i
0,1,2,3-D.I.: k-a-i-g-i

----- No. 31 -----

input: z-i-m-u-ky-o-k-u-d-e-s-u
3-gram: z-i-m-u-ky-o-k-u-d-e-s-u
0,1,2,3-D.I.: z-i-m-u-ky-o-k-u-d-e-s-u

----- No. 32 -----

input: m-o-sh-i-m-o-sh-i
3-gram: m-o-sh-i-m-o-sh-i
0,1,2,3-D.I.: m-o-sh-i-m-o-sh-i

----- No. 33 -----

input: ts-u-u-y-a-k-u
3-gram: ts-u-u-y-a-k-u
0,1,2,3-D.I.: ts-u-u-y-a-k-u

----- No. 34 -----

input: d-e-N-w-a
3-gram: d-e-N-w-a
0,1,2,3-D.I.: d-e-N-w-a

----- No. 35 -----

input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 36 -----

input: k-a-i-g-i-e-n-o
3-gram: k-a-i-g-i-e-n-o
0,1,2,3-D.I.: k-a-i-g-i-e-n-o

----- No. 37 -----

input: s-a-N-k-a-o
3-gram: s-a-N-k-a-o
0,1,2,3-D.I.: s-a-N-k-a-o

----- No. 38 -----

input: m-o-u-sh-i-k-o-m-i-t-a-i-n-o-d-e-s-u
3-gram: m-o-u-sh-i-k-o-m-i-t-a-i-n-d-e-s-u
0,1,2,3-D.I.: m-o-u-sh-i-k-o-m-i-t-a-i-n-d-e-s-u

----- No. 39 -----

input: k-e-r-e-d-o-m-o
3-gram: k-e-r-e-d-o-m-o
0,1,2,3-D.I.: k-e-r-e-d-o-m-o

----- No. 40 -----

input: d-o-n-o
3-gram: d-o-n-o
0,1,2,3-D.I.: d-o-n-o

----- No. 41 -----

input: y-o-u-n-a
3-gram: i-u-y-o-u-n-a
0,1,2,3-D.I.: i-u-y-o-u-n-a

----- No. 42 -----

input: t-e-t-s-u-z-u-k-i-o
3-gram: t-e-t-s-u-z-u-k-i-o
0,1,2,3-D.I.: t-e-t-s-u-z-u-k-i-o

----- No. 43 -----

input: s-u-r-e-b-a
3-gram: s-u-r-e-b-a
0,1,2,3-D.I.: s-u-r-e-b-a

----- No. 44 -----

input: y-o-r-o-sh-i-i-n-o-d-e-sy-o-u-k-a
3-gram: y-o-r-o-sh-i-i-N-d-e-sy-o-u-k-a
0,1,2,3-D.I.: y-o-r-o-sh-i-i-N-d-e-sy-o-u-k-a

15

----- No. 45 -----
input: ts-u-u-y-a-k-u
3-gram: ts-u-u-y-a-k-u
0,1,2,3-D.I.: ts-u-u-y-a-k-u

----- No. 46 -----
input: d-e-N-w-a-n-o
3-gram: d-e-N-w-a-n-o
0,1,2,3-D.I.: d-e-N-w-a-n-o

----- No. 47 -----
input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 48 -----
input: k-a-i-g-i-n-i
3-gram: k-a-i-i-N
0,1,2,3-D.I.: k-a-i-i-N

----- No. 49 -----
input: s-a-N-k-a-s-u-r-u
3-gram: s-a-N-k-a-s-u-r-u
0,1,2,3-D.I.: s-a-N-k-a-s-u-r-u

----- No. 50 -----
input: t-a-m-e-n-i-w-a
3-gram: t-a-m-e-n-i-w-a
0,1,2,3-D.I.: t-a-m-e-n-i-w-a

----- No. 51 -----
input: sy-o-t-e-i-n-o
3-gram: sy-o-t-e-n-o
0,1,2,3-D.I.: sy-o-t-e-n-o

----- No. 52 -----
input: m-o-u-sh-i-k-o-m-i
3-gram: m-o-u-sh-i-k-o-m-i
0,1,2,3-D.I.: m-o-u-sh-i-k-o-m-i

----- No. 53 -----
input: y-o-u-sh-i-o
3-gram: y-o-u-sh-i-o
0,1,2,3-D.I.: y-o-u-sh-i-o

----- No. 54 -----
input: m-o-ch-i-i-t-e
3-gram: m-o-ch-i-i-t-e
0,1,2,3-D.I.: m-o-ch-i-i-t-e

----- No. 55 -----
input: s-a-N-k-a
3-gram: s-a-N-k-a
0,1,2,3-D.I.: s-a-N-k-a

----- No. 56 -----
input: t-o-u-r-o-k-u-s-u-r-u
3-gram: t-o-u-r-o-k-u-s-u-r-u
0,1,2,3-D.I.: t-o-u-r-o-k-u-s-u-r-u

----- No. 57 -----
input: k-o-t-o-g-a
3-gram: k-o-t-o-g-a
0,1,2,3-D.I.: k-o-t-o-g-a

----- No. 58 -----
input: h-i-t-s-u-y-o-u-d-e-s-u
3-gram: h-i-t-s-u-y-o-u-d-e-s-u
0,1,2,3-D.I.: h-i-t-s-u-y-o-u-d-e-s-u

----- No. 59 -----
input: k-a-i-g-i-n-i
3-gram: k-a-i-g-i-n-i
0,1,2,3-D.I.: k-a-i-g-i-n-i

----- No. 60 -----
input: h-a-Q-py-o-u-s-u-r-u-n-o-d-e-w-a
3-gram: h-a-Q-py-o-u-s-u-r-u-n-o-d-e-w-a
0,1,2,3-D.I.: h-a-Q-py-o-u-s-u-r-u-n-o-d-e-w-a

----- No. 61 -----
input: n-a-k-u-t-e
3-gram: n-a-k-u-t-e
0,1,2,3-D.I.: n-a-k-u-t-e

----- No. 62 -----
input: cy-o-u-k-o-u-s-u-r-u-d-a-k-e-d-a-t-o
3-gram: cy-o-u-k-o-u-s-u-r-u-d-a-k-e-d-a-t-o
0,1,2,3-D.I.: cy-o-u-k-o-u-s-u-r-u-d-a-k-e-d-a-t-o

----- No. 63 -----
input: h-i-y-o-u-w-a
3-gram: h-i-y-o-u-w-a
0,1,2,3-D.I.: h-i-y-o-u-w-a

----- No. 64 -----
input: i-k-u-r-a
3-gram: i-k-u-r-a
0,1,2,3-D.I.: i-k-u-r-a

----- No. 65 -----
input: k-a-k-a-r-i-m-a-s-u-k-a
3-gram: k-a-k-a-r-i-m-a-s-u-k-a
0,1,2,3-D.I.: k-a-k-a-r-i-m-a-s-u-k-a

----- No. 66 -----
input: g-o-h-a-Q-py-o-u-o
3-gram: g-o-h-a-Q-py-o-u
0,1,2,3-D.I.: g-o-h-a-Q-py-o-u

----- No. 67 -----

input: k-i-b-o-u-s-a-r-e-r-u
3-gram: k-i-b-o-u-s-a-r-e-r-u
0,1,2,3-D.I.: k-i-b-o-u-s-a-r-e-r-u

----- No. 68 -----

input: b-a-a-i-n-i-w-a
3-gram: b-a-a-i-n-i-w-a
0,1,2,3-D.I.: b-a-a-i-n-i-w-a

----- No. 69 -----

input: y-o-k-o-u-sy-u-u-d-a-i
3-gram: y-o-k-u-o-N-sy-a-g-u-r-a-i
0,1,2,3-D.I.: y-o-k-u-o-N-sy-u-d-a-i

----- No. 70 -----

input: t-o-u-r-o-k-u-ry-o-u-o
3-gram: t-o-u-r-o-k-u-i-ry-o-u
0,1,2,3-D.I.: t-o-u-r-o-k-u-i-ry-o-u

----- No. 71 -----

input: h-u-k-u-m-e-t-a
3-gram: h-u-k-u-m-a-r-e-t-a
0,1,2,3-D.I.: h-u-k-u-m-e-t-a

----- No. 72 -----

input: s-a-N-k-a
3-gram: s-a-N-k-a
0,1,2,3-D.I.: s-a-N-k-a

----- No. 73 -----

input: h-i-y-o-u-w-a
3-gram: h-i-y-o-u-w-a
0,1,2,3-D.I.: h-i-y-o-u-w-a

----- No. 74 -----

input: y-o-N-m-a-N-e-N-d-e-s-u
3-gram: y-o-N-m-a-N-e-N-d-e-s-u
0,1,2,3-D.I.: y-o-N-m-a-N-e-N-d-e-s-u

----- No. 75 -----

input: cy-o-u-k-o-u-n-o-m-i-n-o
3-gram: cy-u-u-o-u-k-o-u-m-i-N-n-o
0,1,2,3-D.I.: cy-o-u-k-o-u-n-o-g-i-n-o

----- No. 76 -----

input: b-a-a-i-w-a
3-gram: b-a-a-i-w-a
0,1,2,3-D.I.: b-a-a-i-w-a

----- No. 77 -----

input: t-o-u-z-i-ts-u-n-o
3-gram: t-o-u-z-i-ts-u-n-o
0,1,2,3-D.I.: t-o-u-z-i-ts-u-n-o

----- No. 78 -----

input: u-k-e-ts-u-k-e-m-o
3-gram: u-k-e-ts-u-k-e-n-o
0,1,2,3-D.I.: u-k-e-ts-u-k-e-n-o

----- No. 79 -----

input: k-a-n-o-u-d-e
3-gram: k-a-i-d-o-u-d-e
0,1,2,3-D.I.: k-a-n-o-u-d-e

----- No. 80 -----

input: y-o-k-o-u-sy-u-u-d-a-i-o
3-gram: ry-o-k-o-u-sh-i-n-a-i-y-o-u-o
0,1,2,3-D.I.: ry-o-k-o-u-sh-i-r-e-n-a-i-y-o-u-o

----- No. 81 -----

input: h-u-k-u-m-e-t-a
3-gram: h-u-k-u-m-a-r-e-t-a
0,1,2,3-D.I.: h-u-k-u-m-e-t-a

----- No. 82 -----

input: h-i-y-o-u-w-a
3-gram: h-i-y-o-u-w-a
0,1,2,3-D.I.: h-i-y-o-u-w-a

----- No. 83 -----

input: s-a-N-m-a-N
3-gram: s-a-N-w-a-N
0,1,2,3-D.I.: s-a-N-w-a-N

----- No. 84 -----

input: g-o-s-e-N-e-N
3-gram: g-o-s-e-N-e-N
0,1,2,3-D.I.: g-o-s-e-N-e-N

----- No. 85 -----

input: k-a-k-a-r-i-m-a-s-u
3-gram: k-a-k-a-r-i-m-a-s-u
0,1,2,3-D.I.: k-a-k-a-r-i-m-a-s-u

----- No. 86 -----

input: s-a-N-k-a
3-gram: s-a-N-k-a
0,1,2,3-D.I.: s-a-N-k-a

----- No. 87 -----

input: t-o-u-r-o-k-u-n-o
3-gram: t-o-u-r-o-k-u-n-o
0,1,2,3-D.I.: t-o-u-r-o-k-u-n-o

----- No. 88 -----

input: m-o-u-sh-i-k-o-m-i
3-gram: m-o-u-sh-i-k-o-m-i
0,1,2,3-D.I.: m-o-u-sh-i-k-o-m-i

61

----- No. 89 -----

input: y-o-u-sh-i-w-a
3-gram: y-o-u-sh-i-w-a
0,1,2,3-D.I.: y-o-u-sh-i-w-a

----- No. 90 -----

input: d-o-n-o
3-gram: d-o-n-o
0,1,2,3-D.I.: d-o-n-o

----- No. 91 -----

input: y-o-u-n-i
3-gram: i-u-y-o-u-n-i
0,1,2,3-D.I.: i-u-y-o-u-n-i

----- No. 92 -----

input: sh-i-t-e
3-gram: sh-i-t-e
0,1,2,3-D.I.: sh-i-t-e

----- No. 93 -----

input: t-e-n-i
3-gram: t-e-n-i
0,1,2,3-D.I.: t-e-n-i

----- No. 94 -----

input: i-r-e-r-e-b-a
3-gram: i-d-e-a-r-e-b-a
0,1,2,3-D.I.: i-e-r-e-b-a

----- No. 95 -----

input: y-o-r-o-sh-i-i-n-o-d-e-sy-o-u-k-a
3-gram: y-o-r-o-sh-i-i-N-d-e-sy-o-u-k-a
0,1,2,3-D.I.: y-o-r-o-sh-i-i-N-d-e-sy-o-u-k-a

----- No. 96 -----

input: o-n-a-m-a-e-t-o
3-gram: o-n-a-m-a-e-t-o
0,1,2,3-D.I.: o-n-a-m-a-e-t-o

----- No. 97 -----

input: g-o-zy-u-u-sy-o-o
3-gram: g-o-zy-u-u-sy-o-o
0,1,2,3-D.I.: g-o-zy-u-u-sy-o-o

----- No. 98 -----

input: o-sh-i-r-a-s-e
3-gram: o-sh-i-r-a-s-e
0,1,2,3-D.I.: o-sh-i-r-a-s-e

----- No. 99 -----

input: i-t-a-d-a-k-e-r-e-b-a
3-gram: i-t-a-d-a-k-e-r-e-b-a
0,1,2,3-D.I.: i-t-a-d-a-k-e-r-e-b-a

----- No. 100 -----

input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 101 -----

input: k-a-i-g-i-n-o
3-gram: k-a-i-g-i-n-o
0,1,2,3-D.I.: k-a-i-g-i-n-o

----- No. 102 -----

input: z-i-m-u-ky-o-k-u-k-a-r-a
3-gram: z-i-m-u-ky-o-k-u-k-a
0,1,2,3-D.I.: z-i-m-u-ky-o-k-u-k-a

----- No. 103 -----

input: sy-o-t-e-i-n-o
3-gram: sy-o-t-e-n-o
0,1,2,3-D.I.: sy-o-t-e-n-o

----- No. 104 -----

input: y-o-u-sh-i-o
3-gram: y-o-u-sh-i-o
0,1,2,3-D.I.: y-o-u-sh-i-o

----- No. 105 -----

input: o-k-u-r-a-s-e-t-e
3-gram: o-k-u-r-a-s-e-t-e
0,1,2,3-D.I.: o-k-u-r-a-s-e-t-e

----- No. 106 -----

input: i-t-a-d-a-k-i-m-a-s-u
3-gram: i-t-a-d-a-k-i-m-a-s-u
0,1,2,3-D.I.: i-t-a-d-a-k-i-m-a-s-u

----- No. 107 -----

input: zy-u-u-sy-o-w-a
3-gram: zy-u-u-sy-o-w-a
0,1,2,3-D.I.: zy-u-u-sy-o-w-a

----- No. 108 -----

input: o-o-s-a-k-a-sh-i
3-gram: o-o-s-a-k-a-sh-i
0,1,2,3-D.I.: o-o-s-a-k-a-sh-i

----- No. 109 -----

input: h-i-g-a-sh-i-k-u
3-gram: h-i-g-a-sh-i-k-u
0,1,2,3-D.I.: h-i-g-a-sh-i-k-u

----- No. 110 -----

input: sh-i-r-o-m-i
3-gram: sh-i-r-o-m-i
0,1,2,3-D.I.: sh-i-r-o-m-i

81

----- No. 111 -----

input: i-ch-i-n-o
3-gram: i-ch-i-n-o
0,1,2,3-D.I.: i-ch-i-n-o

----- No. 112 -----

input: n-i-n-o
3-gram: n-i-n-o
0,1,2,3-D.I.: n-i-n-o

----- No. 113 -----

input: r-o-k-u-zy-u-u-i-ch-i
3-gram: g-o-r-o-k-u-zy-u-u-i-ch-i
0,1,2,3-D.I.: g-o-r-o-k-u-zy-u-u-i-ch-i

----- No. 114 -----

input: z-i-d-o-u
3-gram: z-i-d-o-u
0,1,2,3-D.I.: z-i-d-o-u

----- No. 115 -----

input: h-o-N-y-a-k-u
3-gram: h-o-N-y-a-k-u
0,1,2,3-D.I.: h-o-N-y-a-k-u

----- No. 116 -----

input: d-e-N-w-a
3-gram: d-e-N-w-a
0,1,2,3-D.I.: d-e-N-w-a

----- No. 117 -----

input: k-e-N-ky-u-u-sy-o-d-e
3-gram: k-e-N-ky-u-u-sh-i-d-e
0,1,2,3-D.I.: k-e-N-ky-u-u-sh-i-d-e

----- No. 118 -----

input: n-a-m-a-e-w-a
3-gram: n-a-m-a-e-w-a
0,1,2,3-D.I.: n-a-m-a-e-w-a

----- No. 119 -----

input: t-a-k-a-r-a
3-gram: k-a-k-a-r-a
0,1,2,3-D.I.: k-a-k-a-r-a

----- No. 120 -----

input: i-ch-i-t-a-r-o-u-d-e-s-u
3-gram: i-sh-i-t-a-r-o-u-d-e-s-u
0,1,2,3-D.I.: i-sh-i-t-a-r-o-u-d-e-s-u

----- No. 121 -----

input: z-i-d-o-u
3-gram: z-i-d-o-u
0,1,2,3-D.I.: z-i-d-o-u

----- No. 122 -----

input: h-o-N-y-a-k-u
3-gram: h-o-N-y-a-k-u
0,1,2,3-D.I.: h-o-N-y-a-k-u

----- No. 123 -----

input: d-e-N-w-a
3-gram: d-e-N-w-a
0,1,2,3-D.I.: d-e-N-w-a

----- No. 124 -----

input: k-e-N-ky-u-u-sy-o
3-gram: k-e-N-ky-u-u-sy-o
0,1,2,3-D.I.: k-e-N-ky-u-u-sy-o

----- No. 125 -----

input: t-a-k-a-r-a-s-a-m-a-d-e-s-u-n-e
3-gram: t-a-k-a-r-a-s-a-m-a-d-e-s-u-n-e
0,1,2,3-D.I.: t-a-k-a-r-a-s-a-m-a-d-e-s-u-n-e

----- No. 126 -----

input: d-e-w-a
3-gram: d-e-w-a
0,1,2,3-D.I.: d-e-w-a

----- No. 127 -----

input: s-o-ch-i-r-a-n-i
3-gram: s-o-ch-i-r-a-n-i
0,1,2,3-D.I.: s-o-ch-i-r-a-n-i

----- No. 128 -----

input: m-o-u-sh-i-k-o-m-i
3-gram: m-o-u-sh-i-k-o-m-i
0,1,2,3-D.I.: m-o-u-sh-i-k-o-m-i

----- No. 129 -----

input: y-o-u-sh-i-o
3-gram: y-o-u-sh-i-o
0,1,2,3-D.I.: y-o-u-sh-i-o

----- No. 130 -----

input: o-k-u-r-a-s-e-t-e
3-gram: o-k-u-r-a-s-e-t-e
0,1,2,3-D.I.: o-k-u-r-a-s-e-t-e

----- No. 131 -----

input: i-t-a-d-a-k-i-m-a-s-u
3-gram: i-t-a-d-a-k-i-m-a-s-u
0,1,2,3-D.I.: i-t-a-d-a-k-i-m-a-s-u

----- No. 132 -----

input: k-i-z-i-t-s-u-t-o
3-gram: k-i-z-i-t-s-u-t-o
0,1,2,3-D.I.: k-i-z-i-t-s-u-t-o

19

----- No. 133 -----
input: k-a-i-zy-o-u-o
3-gram: k-a-i-zy-o-u
0,1,2,3-D.I.: k-a-i-zy-o-u

----- No. 134 -----
input: k-a-k-u-n-i-N-s-a-s-e-t-e
3-gram: k-a-k-u-n-i-N-s-a-s-e-t-e
0,1,2,3-D.I.: k-a-k-u-n-i-N-s-a-s-e-t-e

----- No. 135 -----
input: i-t-a-d-a-k-i-t-a-i-n-o-d-e-s-u-g-a
3-gram: i-t-a-d-a-k-i-t-a-i-N-d-e-s-u-g-a
0,1,2,3-D.I.: i-t-a-d-a-k-i-t-a-i-N-d-e-s-u-g-a

----- No. 136 -----
input: k-i-z-i-ts-u-w-a
3-gram: k-i-z-i-ts-u-w-a
0,1,2,3-D.I.: k-i-z-i-ts-u-w-a

----- No. 137 -----
input: s-e-N-ky-u-u-hy-a-k-u-ky-u-u-zy-u-u-n-a-n-a-n-e-N-n-o
3-gram: s-e-N-ky-u-u-hy-a-k-u-ky-u-u-zy-u-u-r-a-i-n-e-N-n-o
0,1,2,3-D.I.: s-e-N-ky-u-u-r-o-k-u-ky-u-u-zy-u-u-n-a-n-a-n-e-N-n-o

----- No. 138 -----
input: h-a-ch-i-g-a-ts-u
3-gram: h-a-ch-i-g-a-ts-u
0,1,2,3-D.I.: h-a-ch-i-g-a-ts-u

----- No. 139 -----
input: i-ts-u-k-a-k-a-r-a
3-gram: i-Q-k-a-k-a-r-a
0,1,2,3-D.I.: i-Q-k-a-k-a-r-a

----- No. 140 -----
input: y-o-u-k-a-d-e
3-gram: y-o-u-k-a-d-e
0,1,2,3-D.I.: y-o-u-k-a-d-e

----- No. 141 -----
input: k-a-i-zy-o-u-w-a
3-gram: k-a-i-zy-o-u-w-a
0,1,2,3-D.I.: k-a-i-zy-o-u-w-a

----- No. 142 -----
input: ky-o-u-t-o
3-gram: ky-o-u-t-o
0,1,2,3-D.I.: ky-o-u-t-o

----- No. 143 -----
input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 144 -----
input: k-a-i-g-i-zy-o-u-t-o
3-gram: k-a-i-g-i-zy-o-u-t-o
0,1,2,3-D.I.: k-a-i-g-i-zy-o-u-t-o

----- No. 145 -----
input: y-u-u
3-gram: y-u-r-u
0,1,2,3-D.I.: y-u-r-u

----- No. 146 -----
input: k-o-t-o-d-e
3-gram: k-o-t-o-d-e
0,1,2,3-D.I.: k-o-t-o-d-e

----- No. 147 -----
input: y-o-r-o-sh-i-i-d-e-sy-o-u-k-a
3-gram: y-o-r-o-sh-i-i-d-e-sy-o-u-k-a
0,1,2,3-D.I.: y-o-r-o-sh-i-i-d-e-sy-o-u-k-a

----- No. 148 -----
input: h-a-i
3-gram: h-a-i
0,1,2,3-D.I.: h-a-i

----- No. 149 -----
input: s-o-n-o
3-gram: s-o-n-o
0,1,2,3-D.I.: s-o-n-o

----- No. 150 -----
input: t-o-u-r-i-d-e-s-u
3-gram: t-o-o-r-i-d-e-s-u
0,1,2,3-D.I.: t-o-o-r-i-d-e-s-u

----- No. 151 -----
input: k-a-i-zy-o-u-n-o
3-gram: k-a-i-zy-o-u-n-o
0,1,2,3-D.I.: k-a-i-zy-o-u-n-o

----- No. 152 -----
input: ky-o-u-t-o
3-gram: ky-o-u-t-o
0,1,2,3-D.I.: ky-o-u-t-o

----- No. 153 -----
input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 154 -----
input: k-a-i-g-i-zy-o-u-e
3-gram: k-a-i-g-i-zy-o-u-e
0,1,2,3-D.I.: k-a-i-g-i-zy-o-u-e

20

----- No. 155 -----

input: i-k-u-n-i-w-a
3-gram: y-u-k-u-n-i-w-a
0,1,2,3-D.I.: y-u-k-u-n-i-w-a

----- No. 156 -----

input: ky-o-u-t-o-e-k-i-k-a-r-a
3-gram: ky-o-u-t-o-e-k-i-k-a-r-a
0,1,2,3-D.I.: ky-o-u-t-o-e-k-i-k-a-r-a

----- No. 157 -----

input: d-o-N-n-a
3-gram: d-o-N-n-a
0,1,2,3-D.I.: d-o-N-n-a

----- No. 158 -----

input: k-o-u-t-s-u-u
3-gram: k-o-u-t-s-u-u
0,1,2,3-D.I.: k-o-u-t-s-u-u

----- No. 159 -----

input: sy-u-d-a-N-g-a
3-gram: sy-u-d-a-N-d-a
0,1,2,3-D.I.: sy-u-d-a-N-d-a

----- No. 160 -----

input: a-r-i-m-a-s-u-k-a
3-gram: a-r-i-m-a-s-u-k-a
0,1,2,3-D.I.: a-r-i-m-a-s-u-k-a

----- No. 161 -----

input: ch-i-k-a-t-e-t-s-u-d-e
3-gram: ch-i-k-a-t-e-t-s-u-d-e
0,1,2,3-D.I.: ch-i-k-a-t-e-t-s-u-d-e

----- No. 162 -----

input: k-i-t-a-o-o-z-i-e-k-i-m-a-d-e
3-gram: k-i-t-a-o-o-z-i-e-k-i-m-a-d-e
0,1,2,3-D.I.: k-i-t-a-o-o-z-i-e-k-i-m-a-d-e

----- No. 163 -----

input: i-Q-t-e
3-gram: i-Q-t-e
0,1,2,3-D.I.: i-Q-t-e

----- No. 164 -----

input: s-o-k-o-k-a-r-a
3-gram: s-o-k-o-k-a-r-a
0,1,2,3-D.I.: s-o-k-o-k-a-r-a

----- No. 165 -----

input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 166 -----

input: k-a-i-g-i-zy-o-u-y-u-k-i-n-o
3-gram: k-a-i-g-i-zy-o-u-y-a-k-u-n-o
0,1,2,3-D.I.: k-a-i-g-i-zy-o-u-e-k-i-n-o

----- No. 167 -----

input: b-a-s-u-k-a
3-gram: b-a-s-u-k-a
0,1,2,3-D.I.: b-a-s-u-k-a

----- No. 168 -----

input: m-a-t-a-w-a
3-gram: m-a-t-a-w-a
0,1,2,3-D.I.: m-a-t-a-w-a

----- No. 169 -----

input: t-a-k-u-sh-i-i-o
3-gram: t-a-k-u-sh-i-i-o
0,1,2,3-D.I.: t-a-k-u-sh-i-i-o

----- No. 170 -----

input: r-i-y-o-u
3-gram: g-o-r-i-y-o-u
0,1,2,3-D.I.: g-o-r-i-y-o-u

----- No. 171 -----

input: d-e-k-i-m-a-s-u
3-gram: d-e-k-i-m-a-s-u
0,1,2,3-D.I.: d-e-k-i-m-a-s-u

----- No. 172 -----

input: ky-o-u-t-o-e-k-i-k-a-r-a
3-gram: ky-o-u-t-o-e-k-i-k-a-r-a
0,1,2,3-D.I.: ky-o-u-t-o-e-k-i-k-a-r-a

----- No. 173 -----

input: ky-o-u-t-o
3-gram: ky-o-u-t-o
0,1,2,3-D.I.: ky-o-u-t-o

----- No. 174 -----

input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 175 -----

input: k-a-i-g-i-zy-o-u-m-a-d-e
3-gram: k-a-i-g-i-zy-o-u-m-a-d-e
0,1,2,3-D.I.: k-a-i-g-i-zy-o-u-m-a-d-e

----- No. 176 -----

input: z-i-k-a-N-w-a
3-gram: z-i-k-a-N-w-a
0,1,2,3-D.I.: z-i-k-a-N-w-a

21

----- No. 177 -----

input: d-o-n-o
3-gram: d-o-n-o
0,1,2,3-D.I.: d-o-n-o

----- No. 178 -----

input: k-u-r-a-i
3-gram: k-u-r-a-i
0,1,2,3-D.I.: k-u-r-a-i

----- No. 179 -----

input: k-a-k-a-r-i-m-a-s-u-k-a
3-gram: k-a-k-a-r-i-m-a-s-u-k-a
0,1,2,3-D.I.: k-a-k-a-r-i-m-a-s-u-k-a

----- No. 180 -----

input: ch-i-k-a-t-e-t-s-u-t-o
3-gram: ch-i-k-a-t-e-t-s-u-t-o
0,1,2,3-D.I.: ch-i-k-a-t-e-t-s-u-t-o

----- No. 181 -----

input: b-a-s-u-o
3-gram: b-a-s-u-o
0,1,2,3-D.I.: b-a-s-u-o

----- No. 182 -----

input: n-o-r-i-t-s-u-i-d-a
3-gram: n-o-r-i-t-s-u-i-d-a
0,1,2,3-D.I.: n-o-r-i-t-s-u-i-d-a

----- No. 183 -----

input: b-a-a-i
3-gram: b-a-a-i
0,1,2,3-D.I.: b-a-a-i

----- No. 184 -----

input: m-a-ch-i-z-i-k-a-N-o
3-gram: n-a-i-ch-i-z-i-k-a-N-o
0,1,2,3-D.I.: n-a-i-ch-i-z-i-k-a-N-o

----- No. 185 -----

input: i-r-e
3-gram: i-u
0,1,2,3-D.I.: i-u

----- No. 186 -----

input: t-e
3-gram: e
0,1,2,3-D.I.: e

----- No. 187 -----

input: i-ch-i-z-i-k-a-N
3-gram: i-ch-i-z-i-k-a-N
0,1,2,3-D.I.: i-ch-i-z-i-k-a-N

----- No. 188 -----

input: t-e-e-d-o
3-gram: t-e-i-d-o
0,1,2,3-D.I.: t-e-i-d-o

----- No. 189 -----

input: k-a-k-a-r-i-m-a-s-u
3-gram: k-a-k-a-r-i-m-a-s-u
0,1,2,3-D.I.: k-a-k-a-r-i-m-a-s-u

----- No. 190 -----

input: k-a-i-zy-o-u
3-gram: k-a-i-zy-o-u
0,1,2,3-D.I.: k-a-i-zy-o-u

----- No. 191 -----

input: h-u-k-i-N-n-o
3-gram: h-i-k-u-i-N-n-o
0,1,2,3-D.I.: h-i-k-u-i-N-n-o

----- No. 192 -----

input: ch-i-z-u-y-a
3-gram: ch-i-z-u-y-a
0,1,2,3-D.I.: ch-i-z-u-y-a

----- No. 193 -----

input: ch-i-k-a-t-e-t-s-u-y-a
3-gram: ch-i-k-a-t-e-t-s-u-y-a
0,1,2,3-D.I.: ch-i-k-a-t-e-t-s-u-y-a

----- No. 194 -----

input: b-a-s-u-n-o
3-gram: b-a-s-u-m-o
0,1,2,3-D.I.: b-a-s-u-m-o

----- No. 195 -----

input: z-i-k-o-k-u-hy-o-u-w-a
3-gram: z-i-k-o-k-u-hy-o-u-w-a
0,1,2,3-D.I.: z-i-k-o-k-u-hy-o-u-w-a

----- No. 196 -----

input: z-i-m-u-ky-o-k-u-d-e-w-a
3-gram: z-i-m-u-ky-o-k-u-d-e-w-a
0,1,2,3-D.I.: z-i-m-u-ky-o-k-u-d-e-w-a

----- No. 197 -----

input: y-o-u-i-s-a-r-e-t-e
3-gram: y-o-u-i-s-a-r-e-t-e
0,1,2,3-D.I.: y-o-u-i-s-a-r-e-t-e

----- No. 198 -----

input: i-r-u-d-e-sy-o-u-k-a
3-gram: i-N-d-e-sy-o-u-k-a
0,1,2,3-D.I.: i-N-d-e-sy-o-u-k-a

22

----- No. 199 -----

input: s-a-N-k-a
3-gram: s-a-N-k-a
0,1,2,3-D.I.: s-a-N-k-a

----- No. 200 -----

input: t-o-u-r-o-k-u-n-o
3-gram: t-o-u-r-o-k-u-n-o
0,1,2,3-D.I.: t-o-u-r-o-k-u-n-o

----- No. 201 -----

input: zy-u-r-i
3-gram: zy-u-u-r-a-i
0,1,2,3-D.I.: zy-u-u-r-a-i

----- No. 202 -----

input: ts-u-u-ch-i-n-o
3-gram: ts-u-u-ch-i-n-a-o
0,1,2,3-D.I.: ts-u-i-ch-i-d-o

----- No. 203 -----

input: n-a-k-a-n-i
3-gram: n-a-i-k-a-N
0,1,2,3-D.I.: n-a-k-a-N

----- No. 204 -----

input: ky-o-u-t-o-e-k-i-k-a-r-a
3-gram: ky-o-u-t-o-e-k-i-k-a-r-a
0,1,2,3-D.I.: ky-o-u-t-o-e-k-i-k-a-r-a

----- No. 205 -----

input: ky-o-u-t-o
3-gram: ky-o-u-t-o
0,1,2,3-D.I.: ky-o-u-t-o

----- No. 206 -----

input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 207 -----

input: k-a-i-g-i-zy-o-u-m-a-d-e-n-o
3-gram: k-a-i-g-i-zy-o-u-m-a-d-e-n-o
0,1,2,3-D.I.: k-a-i-g-i-zy-o-u-m-a-d-e-n-o

----- No. 208 -----

input: k-o-u-ts-u-u-r-o-n-o
3-gram: k-o-u-ts-u-u-n-o
0,1,2,3-D.I.: k-o-u-ts-u-u-n-o

----- No. 209 -----

input: ry-a-k-u-z-u-y-a
3-gram: g-a-k-u-z-u-y-a
0,1,2,3-D.I.: g-a-k-u-z-u-y-a

----- No. 210 -----

input: z-i-k-o-k-u-hy-o-u-g-a
3-gram: z-i-k-o-k-u-hy-o-u-g-a
0,1,2,3-D.I.: z-i-k-o-k-u-hy-o-u-g-a

----- No. 211 -----

input: h-u-k-u-m-a-r-e-t-e
3-gram: h-u-k-u-m-a-r-e-t-e
0,1,2,3-D.I.: h-u-k-u-m-a-r-e-t-e

----- No. 212 -----

input: i-m-a-s-u
3-gram: i-m-a-s-u
0,1,2,3-D.I.: i-m-a-s-u

----- No. 213 -----

input: k-o-N-k-a-i-n-o
3-gram: k-o-N-k-a-i-n-o
0,1,2,3-D.I.: k-o-N-k-a-i-n-o

----- No. 214 -----

input: k-o-k-u-s-a-i
3-gram: k-o-k-u-s-a-i
0,1,2,3-D.I.: k-o-k-u-s-a-i

----- No. 215 -----

input: k-a-i-g-i-n-i
3-gram: k-a-i-g-i-n-i
0,1,2,3-D.I.: k-a-i-g-i-n-i

----- No. 216 -----

input: r-o-N-b-u-N-o
3-gram: r-o-N-b-u-N-o
0,1,2,3-D.I.: r-o-N-b-u-N-o

----- No. 217 -----

input: h-a-Q-py-o-u-sh-i-t-a-i-N-d-e-s-u-g-a
3-gram: h-a-Q-py-o-u-sh-i-t-a-i-N-d-e-s-u-g-a
0,1,2,3-D.I.: h-a-Q-py-o-u-sh-i-t-a-i-N-d-e-s-u-g-a

----- No. 218 -----

input: k-e-N-ky-u-u
3-gram: k-e-N-ky-u-u
0,1,2,3-D.I.: k-e-N-ky-u-u

----- No. 219 -----

input: b-u-N-y-a-w-a
3-gram: b-u-N-y-a-w-a
0,1,2,3-D.I.: b-u-N-y-a-w-a

----- No. 220 -----

input: ts-u-u-y-a-k-u
3-gram: ts-u-u-y-a-k-u
0,1,2,3-D.I.: ts-u-u-y-a-k-u

23

----- No. 221 -----

input: d-e-N-w-a-n-i
3-gram: d-e-N-w-a-n-i
0,1,2,3-D.I.: d-e-N-w-a-n-i

----- No. 222 -----

input: k-a-g-i-r-a-r-e-r-u-n-o-d-e-sy-o-u-k-a
3-gram: k-a-g-i-r-a-r-e-r-u-N-d-e-sy-o-u-k-a
0,1,2,3-D.I.: k-a-g-i-r-a-r-e-r-u-N-d-e-sy-o-u-k-a

----- No. 223 -----

input: i-i-e
3-gram: i-i-e
0,1,2,3-D.I.: i-i-e

----- No. 224 -----

input: t-s-u-u-y-a-k-u
3-gram: t-s-u-u-y-a-k-u
0,1,2,3-D.I.: t-s-u-u-y-a-k-u

----- No. 225 -----

input: d-e-N-w-a-d-a-k-e-d-e
3-gram: d-e-N-m-a-a-k-u-d-e
0,1,2,3-D.I.: d-e-N-w-a-d-a-k-e-d-e

----- No. 226 -----

input: n-a-k-u
3-gram: n-a-k-u
0,1,2,3-D.I.: n-a-k-u

----- No. 227 -----

input: o-N-s-e-i
3-gram: o-N-s-e-i
0,1,2,3-D.I.: o-N-s-e-i

----- No. 228 -----

input: n-i-N-sh-i-k-i
3-gram: n-i-N-sh-i-k-i
0,1,2,3-D.I.: n-i-N-sh-i-k-i

----- No. 229 -----

input: o-N-s-e-i
3-gram: o-N-s-e-i
0,1,2,3-D.I.: o-N-s-e-i

----- No. 230 -----

input: r-i-k-a-i
3-gram: g-i-k-a-i
0,1,2,3-D.I.: g-i-k-a-i

----- No. 231 -----

input: g-e-N-g-o
3-gram: g-e-N-o
0,1,2,3-D.I.: g-e-N-o

----- No. 232 -----

input: k-a-i-s-e-k-i
3-gram: k-a-i-s-e-k-i
0,1,2,3-D.I.: k-a-i-s-e-k-i

----- No. 233 -----

input: k-i-k-a-i
3-gram: sh-i-t-a-i
0,1,2,3-D.I.: sh-i-t-a-i

----- No. 234 -----

input: h-o-N-y-a-k-u-n-a-d-o-n-o
3-gram: h-o-N-y-a-k-u-n-a-d-o-n-o
0,1,2,3-D.I.: o-m-o-i-y-a-k-u-n-a-d-o-n-o

----- No. 235 -----

input: k-o-u-h-a-N-n-a
3-gram: k-o-u-h-a-N-n-a
0,1,2,3-D.I.: k-o-u-h-a-N-n-a

----- No. 236 -----

input: b-u-N-y-a-o
3-gram: b-u-N-y-a-k-u-o
0,1,2,3-D.I.: b-u-N-y-a-k-u-o

----- No. 237 -----

input: h-u-k-u-N-d-e
3-gram: h-u-k-u-N-d-e
0,1,2,3-D.I.: h-u-k-u-N-d-e

----- No. 238 -----

input: i-m-a-s-u
3-gram: i-m-a-s-u
0,1,2,3-D.I.: i-m-a-s-u

----- No. 239 -----

input: k-a-i-g-i-n-o
3-gram: k-a-i-g-i-n-o
0,1,2,3-D.I.: k-a-i-g-i-n-o

----- No. 240 -----

input: k-o-u-sh-i-k-i
3-gram: k-o-u-sh-i-k-i
0,1,2,3-D.I.: k-o-u-sh-i-k-i

----- No. 241 -----

input: g-e-N-g-o-w-a
3-gram: g-e-N-g-o-w-a
0,1,2,3-D.I.: g-e-N-g-o-w-a

----- No. 242 -----

input: n-i-Q-p-o-N-g-o-t-o
3-gram: n-i-Q-p-o-N-t-o
0,1,2,3-D.I.: n-i-Q-p-o-N-g-o-t-o

72

----- No. 243 -----

input: e-i-g-o-t-o
3-gram: e-i-g-o-t-o
0,1,2,3-D.I.: e-i-g-o-t-o

----- No. 244 -----

input: y-u-u
3-gram: y-u-r-u
0,1,2,3-D.I.: y-u-r-u

----- No. 245 -----

input: k-o-t-o-d-e-s-u-g-a
3-gram: k-o-t-o-d-e-s-u-g-a
0,1,2,3-D.I.: k-o-t-o-d-e-s-u-g-a

----- No. 246 -----

input: k-o-r-e-w-a
3-gram: k-o-r-e-w-a
0,1,2,3-D.I.: k-o-r-e-w-a

----- No. 247 -----

input: d-o-u
3-gram: d-o-u
0,1,2,3-D.I.: d-o-u

----- No. 248 -----

input: y-u-u
3-gram: y-u-r-u
0,1,2,3-D.I.: y-u-r-u

----- No. 249 -----

input: i-m-i-d-e-sy-o-u-k-a
3-gram: i-m-i-d-e-sy-o-u-k-a
0,1,2,3-D.I.: i-m-i-d-e-sy-o-u-k-a

----- No. 250 -----

input: e-i-g-o-n-o
3-gram: e-i-g-o-n-o
0,1,2,3-D.I.: e-i-g-o-n-o

----- No. 251 -----

input: h-a-Q-py-o-u-w-a
3-gram: h-a-Q-py-o-u-w-a
0,1,2,3-D.I.: h-a-Q-py-o-u-w-a

----- No. 252 -----

input: n-i-Q-p-o-N-g-o-n-i
3-gram: b-i-Q-k-u-o-N-n-a-N
0,1,2,3-D.I.: n-i-Q-p-o-N-m-o-N

----- No. 253 -----

input: z-i-d-o-u
3-gram: z-i-d-o-u
0,1,2,3-D.I.: z-i-d-o-u

----- No. 254 -----

input: h-o-N-y-a-k-u-s-a-r-e-r-u
3-gram: k-o-N-r-a-k-u-s-a-r-e-r-u
0,1,2,3-D.I.: o-N-r-a-k-u-s-a-r-e-r-u

----- No. 255 -----

input: y-o-u-n-i
3-gram: i-u-y-o-u-n-i
0,1,2,3-D.I.: i-u-y-o-u-n-i

----- No. 256 -----

input: n-a-Q-t-e
3-gram: n-a-Q-t-e
0,1,2,3-D.I.: n-a-Q-t-e

----- No. 257 -----

input: i-m-a-s-u
3-gram: i-m-a-s-u
0,1,2,3-D.I.: i-m-a-s-u

----- No. 258 -----

input: m-a-t-a
3-gram: m-a-t-a
0,1,2,3-D.I.: m-a-t-a

----- No. 259 -----

input: n-i-Q-p-o-N-g-o-n-o
3-gram: n-i-Q-p-o-N-n-o
0,1,2,3-D.I.: n-i-Q-p-o-N-g-o-n-o

----- No. 260 -----

input: h-a-Q-py-o-u-w-a
3-gram: h-a-Q-py-o-u-w-a
0,1,2,3-D.I.: h-a-Q-py-o-u-w-a

----- No. 261 -----

input: e-i-g-o-n-i
3-gram: e-i-g-o-n-i
0,1,2,3-D.I.: e-i-g-o-n-i

----- No. 262 -----

input: h-o-N-y-a-k-u-s-a-r-e-m-a-s-u
3-gram: k-o-N-r-a-k-u-s-a-i-m-a-s-u
0,1,2,3-D.I.: k-o-N-r-a-k-u-s-a-i-m-a-s-u

----- No. 263 -----

input: t-a-d-a-sh-i
3-gram: t-a-i-sh-i
0,1,2,3-D.I.: t-a-i-sh-i

----- No. 264 -----

input: y-o-k-o-u-sy-u-u-n-i
3-gram: ry-o-k-o-u-sh-i-N-n-i
0,1,2,3-D.I.: y-o-k-u-o-sh-i-e-r-u-n-i

25

----- No. 265 -----

input: ts-u-i-t-e-w-a
3-gram: ts-u-i-t-e-w-a
0,1,2,3-D.I.: ts-u-i-t-e-w-a

----- No. 266 -----

input: s-u-b-e-t-e
3-gram: s-u-b-e-t-e
0,1,2,3-D.I.: s-u-b-e-t-e

----- No. 267 -----

input: e-i-g-o-d-e
3-gram: e-i-g-o-d-e
0,1,2,3-D.I.: e-i-g-o-d-e

----- No. 268 -----

input: sh-i-Q-p-i-ts-u-sh-i-t-e
3-gram: sh-i-Q-p-a-ts-u-sh-i-t-e
0,1,2,3-D.I.: sh-i-Q-p-a-ts-u-sh-i-t-e

----- No. 269 -----

input: i-t-a-d-a-k-i-m-a-s-u
3-gram: i-t-a-d-a-k-i-m-a-s-u
0,1,2,3-D.I.: i-t-a-d-a-k-i-m-a-s-u

----- No. 270 -----

input: s-o-u-d-e-s-u-k-a
3-gram: s-o-u-d-e-s-u-k-a
0,1,2,3-D.I.: s-o-u-d-e-s-u-k-a

----- No. 271 -----

input: w-a-k-a-r-i-m-a-sh-i-t-a
3-gram: w-a-k-a-r-i-m-a-sh-i-t-a
0,1,2,3-D.I.: w-a-k-a-r-i-m-a-sh-i-t-a

----- No. 272 -----

input: d-e-w-a
3-gram: d-e-w-a
0,1,2,3-D.I.: d-e-w-a

----- No. 273 -----

input: m-o-u-sh-i-k-o-m-i
3-gram: m-o-u-sh-i-k-o-m-i
0,1,2,3-D.I.: m-o-u-sh-i-k-o-m-i

----- No. 274 -----

input: y-o-u-sh-i-n-o
3-gram: y-o-u-sh-i-n-o
0,1,2,3-D.I.: y-o-u-sh-i-n-o

----- No. 275 -----

input: k-e-N
3-gram: k-e-N
0,1,2,3-D.I.: k-e-N

----- No. 276 -----

input: y-o-r-o-sh-i-k-u
3-gram: y-o-r-o-sh-i-k-u
0,1,2,3-D.I.: y-o-r-o-sh-i-k-u

----- No. 277 -----

input: o-n-e-g-a-i
3-gram: o-n-e-g-a-i
0,1,2,3-D.I.: o-n-e-g-a-i

----- No. 278 -----

input: m-o-u-sh-i
3-gram: m-o-sh-i
0,1,2,3-D.I.: m-o-sh-i

----- No. 279 -----

input: a-g-e-m-a-s-u
3-gram: a-g-e-m-a-s-u
0,1,2,3-D.I.: a-g-e-m-a-s-u

26

[実験で用いた音節の種類頁]

Mar 3 18:12 1990 MOR A Page 1

a	de	uu	rou	myaa
i	do	ei	roo	myuu
u	ba	ee	waa	myou
e	bi	ou	gaa	myoo
o	bu	oo	gii	ryaa
ka	be	kaa	guu	ryuu
ki	bo	kii	gei	ryou
ku	pa	kuu	gee	ryoo
ke	pi	kei	gou	byaa
ko	pu	kee	goo	byuu
sa	pe	kou	zaa	byou
shi	po	koo	zii	byoo
su	kya	saa	zuu	gyaa
se	kyu	shii	zei	gyuu
so	kyo	suu	zee	gyou
ta	sha	sei	zou	gyoo
chi	shu	see	zoo	zyaa
tsu	sho	sou	daa	zyuu
te	cha	soo	dei	zyou
to	chu	taa	dee	zyoo
na	cho	chii	dou	pyaa
ni	nya	tsuu	doo	pyuu
nu	nyu	tei	baa	pyou
ne	nyo	tee	bii	pyoo
no	hya	tou	buu	faa
ha	hyu	too	bei	fii
hi	hyo	naa	bee	fei
hu	mya	nii	bou	fee
he	myu	nuu	boo	fou
ho	myo	nei	paa	foo
ma	rya	nee	pii	tii
mi	ryu	nou	puu	tuu
mu	ryo	noo	pei	dii
me	bya	haa	pee	duu
mo	byu	hii	pou	# /* end of word */
ya	byo	huu	poo	
yu	gya	hei	kyaa	
yo	gyu	hee	kyuu	
ra	gyo	hou	kyou	
ri	zya	hoo	kyoo	
ru	zyu	maa	shaa	
re	zyo	mii	shuu	
ro	pya	muu	shou	
wa	pyu	mei	shoo	
N	pyo	mee	chaa	
ga	Q	mou	chuu	
gi	fa	moo	chou	
gu	fi	yaa	choo	
ge	fe	yuu	nyaa	
go	fo	you	nyuu	
za	ti	yoo	nyou	
zi	tu	raa	nyoo	
zu	di	rii	hyaa	
ze	du	ruu	hyuu	
zo	aa	rei	hyou	
da	ii	ree	hyoo	

[音節パープレキシティ]

音節n-gramモデルの音節パープレキシティ F は次式で与えられる。音節パープレキシティが小さいほど絞り込みの効果が大きいとすることができる。

$$F = 2^H$$

$$H = - \sum_i \sum_j \sum_k p(i-1, \dots, i-(n-1)) \cdot p(i | i-1, \dots, i-(n-1)) \cdot \log p(i | i-1, \dots, i-(n-1))$$

$$= - \sum_i \sum_j \sum_k p(i, i-1, \dots, i-(n-1)) \cdot \log p(i | i-1, \dots, i-(n-1))$$

H : 1音節当たりのエントロピー

表A-1に今回の実験で、全テキストデータベースを用いて言語モデルを作成した場合の音節パープレキシティを示す。

表A-1 音節n-gramモデルのパープレキシティ

言語モデル	1音節当たりの エントロピー	音節 パープレキシティ
0-gram	8.02	259
1-gram	5.48	44.7
2-gram	4.13	17.5
3-gram	3.03	8.18

[音節連鎖の文法]

Oct 23 19:25 1989 /ifg1/G-TH/sp51.gra Page 1

```
;;  
;; Generate Japanese syllabic contexts  
;; Aug. 8.1989 by E.Kitagawa  
;;  
(<start> <--> (<_start>))  
(<_start> <--> (q1 <sequence> q2))  
(<sequence> <--> (<syllable_f>))  
(<sequence> <--> (<syllable_f> =))  
(<sequence> <--> (<sequence> <syllable>))  
(<syllable> <--> (<syllable_f>))  
(<syllable> <--> (<syllable_f> =))  
(<syllable> <--> (<syllable_n>))  
(<syllable> <--> (<syllable_n> =))  
(<syllable_f> <--> (w a))  
(<syllable_f> <--> (<vowel>))  
(<syllable_f> <--> (b <vowel>))  
(<syllable_f> <--> (g <vowel>))  
(<syllable_f> <--> (m <vowel>))  
(<syllable_f> <--> (n <vowel>))  
(<syllable_f> <--> (r <vowel>))  
(<syllable_f> <--> (z <vowel>))  
(<syllable_f> <--> (d <a_e_o>))  
(<syllable_f> <--> (t <a_e_o>))  
(<syllable_f> <--> (s <a_e_o>))  
(<syllable_f> <--> (h <a_e_o>))  
(<syllable_f> <--> (p <a_e_o>))  
(<syllable_f> <--> (k <a_e_o>))  
(<syllable_f> <--> (y <a_u_o>))  
(<syllable_f> <--> (by <a_u_o>))  
(<syllable_f> <--> (gy <a_u_o>))  
(<syllable_f> <--> (hy <a_u_o>))  
(<syllable_f> <--> (my <a_u_o>))  
(<syllable_f> <--> (ny <a_u_o>))  
(<syllable_f> <--> (ry <a_u_o>))  
(<syllable_f> <--> (zy <a_u_o>))  
(<syllable_f> <--> (sy <a_u_o>))  
(<syllable_f> <--> (py <a_u_o>))  
(<syllable_f> <--> (cy <a_u_o>))  
(<syllable_f> <--> (ky <a_u_o>))  
(<syllable_f> <--> (sh i2))  
(<syllable_f> <--> (h i2))  
(<syllable_f> <--> (ch i2))  
(<syllable_f> <--> (p i2))  
(<syllable_f> <--> (k i2))  
(<syllable_f> <--> (s u2))  
(<syllable_f> <--> (h u2))  
(<syllable_f> <--> (ts u2))  
(<syllable_f> <--> (p u2))  
(<syllable_f> <--> (k u2))  
(<syllable_f> <--> (sh ii))  
(<syllable_f> <--> (h ii))  
(<syllable_f> <--> (ch ii))  
(<syllable_f> <--> (p ii))  
(<syllable_f> <--> (k ii))  
(<syllable_f> <--> (s uu))  
(<syllable_f> <--> (h uu))  
(<syllable_f> <--> (ts uu))  
(<syllable_f> <--> (p uu))  
(<syllable_f> <--> (k uu))  
(<syllable_n> <--> (q t <a_e_o>))  
(<syllable_n> <--> (q s <a_e_o>))  
(<syllable_n> <--> (q p <a_e_o>))  
(<syllable_n> <--> (q k <a_e_o>))  
(<syllable_n> <--> (q sy <a_u_o>))  
(<syllable_n> <--> (q py <a_u_o>))
```

(<syllable_n> <--> (q cy <a_u_o>))
(<syllable_n> <--> (q ky <a_u_o>))
(<syllable_n> <--> (q sh i2))
(<syllable_n> <--> (q ch i2))
(<syllable_n> <--> (q p i2))
(<syllable_n> <--> (q k i2))
(<syllable_n> <--> (q s u2))
(<syllable_n> <--> (q ts u2))
(<syllable_n> <--> (q p u2))
(<syllable_n> <--> (q k u2))
(<syllable_n> <--> (q sh ii))
(<syllable_n> <--> (q ch ii))
(<syllable_n> <--> (q p ii))
(<syllable_n> <--> (q k ii))
(<syllable_n> <--> (q s uu))
(<syllable_n> <--> (q ts uu))
(<syllable_n> <--> (q p uu))
(<syllable_n> <--> (q k uu))
(<vowel> <--> (i))
(<vowel> <--> (ii))
(<vowel> <--> (e))
(<vowel> <--> (ee))
(<vowel> <--> (ei))
(<vowel> <--> (<a_u_o>))
(<a_u_o> <--> (a))
(<a_u_o> <--> (aa))
(<a_u_o> <--> (u))
(<a_u_o> <--> (uu))
(<a_u_o> <--> (o))
(<a_u_o> <--> (oo))
(<a_u_o> <--> (ou))
(<a_e_o> <--> (a))
(<a_e_o> <--> (aa))
(<a_e_o> <--> (e))
(<a_e_o> <--> (ee))
(<a_e_o> <--> (ei))
(<a_e_o> <--> (o))
(<a_e_o> <--> (oo))
(<a_e_o> <--> (ou))