

## 意味属性を利用したクラス N-gram 言語モデルの評価

池谷 晴生 福田 隆 山田 博文 桂田 浩一 新田 恒雄

豊橋技術科学大学 大学院工学研究科 〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1

E-mail: {ikeya, fukuda, yamada}@vox.tutkie.tut.ac.jp, {katsurada, nitta}@tutkie.tut.ac.jp

あらまし 一般に、クラス N-gram 言語モデルは (1) 小規模コーパスで学習ができる、(2) 未知語への対策が容易に行えるなどの利点を持つが、認識性能の観点からは改善が少ない。本報告では、日本語語彙大系中の意味属性をクラス情報として用いたクラス N-gram 言語モデルの評価結果を報告する。提案方式は (1) 学習コーパス中に現れる名詞を、日本語語彙大系を元に意味属性と対応付け、(2) 日本語語彙大系に出現しない単語を、分類対象外単語クラスとして一つのクラスにまとめることで、クラス N-gram を構築する。本方式は、体系化された意味属性を用いるため、パープレキシティなどに基づく自動クラスタリングと比較して、モデル構築時間が格段に低減する。評価では、まず、通常の単語 N-gram と提案のクラス N-gram モデルとの比較を行う。次に、クラスタリング手法に関する比較検討を行った後、意味属性を初期クラスとした自動クラスタリングの性能評価を行う。提案方式は、対話音声を対象とした音声認識実験で性能を大きく改善すると共に、クラスタリングにかかる膨大な計算を不要とする。

キーワード 音声対話認識, 言語モデル, クラス N-gram, 意味属性

### Evaluation of Class N-gram Language Model Based on Semantic Attributes

Haruki IKEYA, Takashi FUKUDA, Hirobumi YAMADA, Kouichi KATSURADA, and Tsuneo NITTA

Graduate School of Engineering, Toyohashi University of Technology 1-1 Hibirigaoka, Tempaku-cho, Toyohashi-shi, Aichi, 441-8580 Japan

E-mail: {ikeya, fukuda, yamada}@vox.tutkie.tut.ac.jp, {katsurada, nitta}@tutkie.tut.ac.jp

**Abstract** In general, a class N-gram model: (1) can be trained with small size of corpora, and (2) have a potential to deal with the problem of unknown words, however, it gives less improvement on performance. This paper describes an attempt to apply the class n-gram language model (LM), in which each lexicon corresponds to semantic attributes in a Japanese lexicon so-called "Goi-Taikei", to an ASR system. The proposed language model is designed by: (1) correlating nouns in training corpora with their semantic attributes, and (2) merging words that do not exist in the Goi-Taikei into a class. The proposed method with well-structured semantic attributes can be trained quickly as compared with auto-clustering methods based on perplexity. Experiments were conducted to evaluate the proposed language model by comparing the performance among LMs. Furthermore, an auto-clustering method with the semantic attributes in "Goi-Taikei" as initial clusters was investigated. We showed that the proposed method achieved significant improvement in the experiments of spoken dialog recognition, and could eliminate the computation needed at the clustering of a class N-gram.

**Keyword** Spoken Dialogue, Language Model, Class N-gram, Semantic Attributes

#### 1. はじめに

我々は、対話音声認識性能の向上を目的に、言語モデルに関する研究を行っている[1],[2],[3]。対話音声認識対象とする実験例としては、これまでチケット予約システム[4]、ホテル案内システム[5]、航空管制シ

テム[6]、および参議院会議の音声認識[7]などが報告されている。一般に、対話音声認識システムは、マルチタスク and/or マルチドメインで使用されることが多く、タスクと利用ドメインを限定できる用途以外で、コーパスを大量に用意することは難しい。そのため、

従来の単語 N-gram では認識性能に限界が生じる場合が多い。我々は先に、日本語の語彙体系を利用し、コーパス中の語彙と同じ意味属性を持つ語彙を追加することで、コーパスの強化を行い、単語 N-gram を補間する方法を試みたが、性能の改善は極わずかであった[2]。

そこで我々は、対話音声認識で一般に用いられているクラス N-gram モデルの利用を検討している。クラス N-gram モデルは、(1) 連鎖確率の対象が単語 (単語 N-gram モデル) に比べて、格段に絶対数が少ないクラスであるため、小規模コーパスでモデルを学習できる、(2) 対話コーパスに出現する未知語をクラスに割当て、N-gram 確率を付与することで、未知語の認識性能を向上できる、などの利点を持つ。一方、クラス N-gram モデルの性能は、よく利用されるクラスタリングに基づく設計を採用すると、設計方法に強く依存することから、これまで様々な手法が提案されている[8,9,10,11]。この中、森らはテストセットパープレキシティが向上するよう自動的に最適なクラスを探索するクラスタリング手法[10]を提案している (以後、自動クラスタリングと呼ぶ)。しかしながら、この方式では、単語間の共起関係に不整合が生じやすく、未知語のクラス割り当てについても一貫性が失われやすいという欠点を持つ。またクラスタリングに基づく設計は、コーパスが比較的小さい対話音声の場合でも、計算時間が膨大になることが問題である。

これに対し我々は、日本語語彙大系[12]中の意味属性を利用して、対話コーパス中の語彙を意味属性クラスに写像し、クラス N-gram 言語モデル (意味属性クラス言語モデル) を学習する方法を検討する。語彙の持つ意味属性を利用することにより、文脈中の意味を考慮したクラスタリングを行えるため、単語 N-gram に比べて共起関係の信頼度が向上すると期待される。また意味属性は、すでに語彙を体系的に分類したものを利用するため、クラスタリングに要する時間も極少ない。提案手法は、小規模の学習データで高速にモデルを構築し、認識性能を改善することを狙っている。

本報では、まず2節で、意味属性クラス・単語混在言語モデルの構築方法を示す。次に、3節で提案手法と従来手法との比較実験結果を示し、続いて4節では意味属性クラスのクラス数を、日本語語彙大系の木構造に沿って減少させた時の実験結果を示す。

## 2. 意味属性クラス言語モデル

### 2.1. クラス N-gram 言語モデル

クラス N-gram 言語モデルにおける単語  $w_i$  の出現確率は、クラス内の単語生起確率とクラス間の N-gram

確率との積として次式のように近似される。

$$P(w_i) = P(w_i | C_i)P(C_i | C_{i-N+1} \dots C_{i-1}) \dots (1)$$

ここで、 $C_i$  は単語  $w_i$  が属するクラスを表す。クラス N-gram は単語 N-gram と比べて、単語間の連鎖確率を直接扱わないため、未知語による認識性能の劣化が少ない。しかし一方で、語彙間の共起関係が弱まり易く、クラスタリング方法によっては認識性能が低下する。クラス N-gram に関する先行研究としては、品詞 N-gram と単語 N-gram を併用する方法[8]、前後の文脈を利用して自動クラスタリングする方法[9]、パープレキシティを利用する方法[10]などが提案されているが、これらの手法は単語間の共起関係に不整合を生じ易い。また、シソーラスを用いてクラス N-gram モデルを作成する手法[11]も提案されているが、単語 N-gram のスムージングの代替として利用するに留まっている。一方、単語の持つ意味に基づいてクラスタリングを行えば、共起関係に生じる不整合を緩和できるため、認識性能の向上が見込まれる。以下では、意味属性を用いた名詞のクラスタリング手法を提案し、この手法を適用して構築したクラス N-gram (以後、意味属性クラス N-gram と呼ぶ) の評価結果を述べる。

### 2.2. 自動クラスタリング手法に基づくクラスタリング

本報告では、クラスタリングに基づく言語モデル設計法から、パープレキシティを利用する自動クラスタリング手法[10]を取上げ、提案手法との比較を行う。自動クラスタリング手法の手順を以下に示す。

- (1) 一単語一クラスとする
- (2) 学習データを分割し、削除補間法により平均テストセットパープレキシティを算出する
- (3) 下記のルールに従ってクラスをマージする
  - ① 二つのクラスを実際にマージする
  - ② パープレキシティが平均テストセットパープレキシティより低くなる場合にのみ、マージを認める
  - ③ 全てのクラスが一度マージ対象となることで、クラスタリングを終了する

### 2.3. 意味属性に基づくクラスタリング

日本語語彙大系[12]では、普通名詞・固有名詞の意味属性 (約 3000) と用言の意味属性 (約 100) が体系的に分類されている (図 1)。今回は、日本語語彙大系を基に学習コーパス中の名詞 (普通名詞・固有名詞全て) のみを意味属性クラスに変換し、クラス N-gram 言語モデルを構築した。具体的には以下の手順で名詞

をクラスタリングする。

- (1) 学習コーパス中に現れる名詞を、日本語語彙大系中の意味属性に写像する
- (2) 日本語語彙大系に出現しない単語は、分類対象外クラスとして一つのクラスにまとめる

図1に例を示すように、「領土」に属する単語は「名詞-具体-場所-地域-領土」という階層化された意味体系に属しているが、ここでは意味属性「領土」にのみ帰属させる。今回は、名詞以外の語彙については、一単語一クラスとした。

## 2.4. 意味属性クラス N-gram の学習方法

日本語語彙大系に基づき語彙をクラスタリングする場合、一つの語彙が複数のクラスに属する場合がある。そのため、クラス間の連鎖確率推定を次式のように拡張する。

$$P(C_i | C_{i-2} C_{i-1}) = \max \{ P(C_{i,m} | C_{i-2,k} C_{i-1,l}) \}$$

$$= \max \left\{ \frac{N(C_{i,m} | C_{i-2,k} C_{i-1,l})}{\sum_m N(C_{i,m} | C_{i-2,k} C_{i-1,l})} \right\}$$

... (2)

ここで、 $k, l, m$  は当該単語が属するクラスのインデックスを表す。上式は単語  $w_{i-2}, w_{i-1}, w_i$  が複数のクラスに属する場合、全組み合わせ中の最大値を求めていることに相当する。単語のクラス内生起確率は、次式から推定した。

$$P(w_i | C_{i,m}) = \frac{N(w_{i,m} | C_{i,m})}{\sum N(w | C_{i,m})}$$

... (3)

これは、(2)式において最大尤度を与えるクラス  $C_{i,m}$  に属する単語  $w_i$  のクラス内生起確率を示す。

## 3. 評価実験

### 3.1. 実験概要

図2に実験システムの構成を示す。デコーダは、日本語ディクテーションシステム Julius 3.4.2 を使用した [13]。また音響モデルは、特徴パラメータに MFCC+ $\Delta$ t+ $\Delta$ P (25次元) を、HMM に 2000 状態の tri-phone モデル (対角化共分散, 性別非依存モデル, 混合数 16) を使用した。言語モデルの学習には、電総研(ETL)道案内対話音声コーパス [14] を利用し、学習コーパスを五つのサブセットに分割した後、意味属性クラス 3-gram モデルを構築する。実験では、① 1/5 の 1751 発話、② 2/5 の 3502 発話、③ 4/5 の 7003 発話の三つのコーパスについて言語モデルを作成する。評価には交差確認法を用いた。評価用音声試料には、道案内対話音声

コーパスの内、言語モデルの学習に使用しないデータから、話者 7 名、159 発話 (全発話時間 688[sec]) を利用している。

## 3.2. 実験結果と考察

### 3.2.1. 意味属性クラス言語モデルの評価

音声認識実験の結果を図3に示す。言語モデルとして、以下の三種類を比較した。

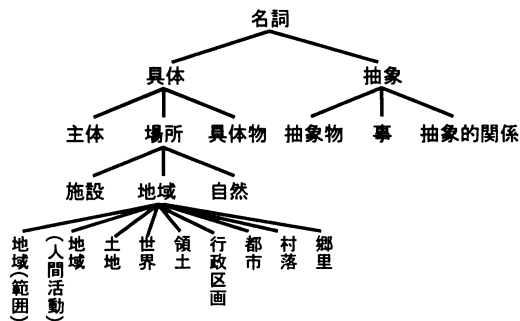


図1. 意味属性 (「地域」の例)

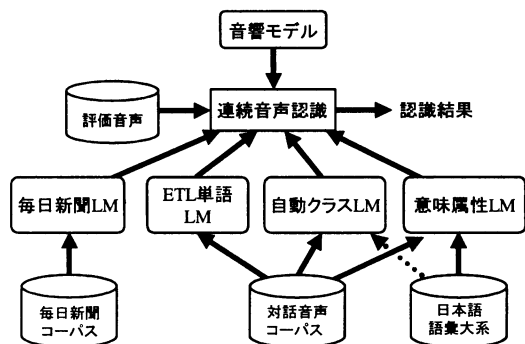


図2. 実験システムの構成

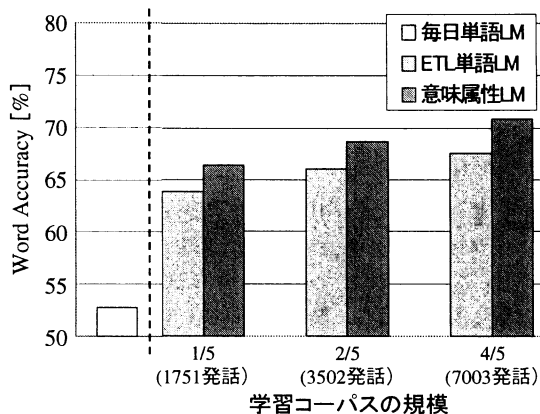


図3. コーパス規模による性能差

- a) 毎日新聞コーパス'91-97 (75ヶ月分) から作成した語彙数 20k の単語 3-gram (毎日単語 LM)
- b) ETL 道案内対話音声コーパスから作成した単語 3-gram (ETL 単語 LM)
- c) b) のコーパス中の名詞を、意味属性を基にクラスへ変換した意味属性クラス 3-gram (意味属性 LM)

まず、タスクの違いについて比較すると、道案内対話音声コーパスから作成した ETL 単語 LM と意味属性 LM は、評価データとは性質の異なる毎日新聞コーパスから作成した毎日単語 LM よりも高い性能を示した。次に、ETL 単語 LM と意味属性 LM を比較すると、学習コーパス規模にかかわらず、意味属性 LM が ETL 単語 LM よりも高い性能を示した。これは、意味属性を利用することで、N-gram の信頼度が向上したためと推測される。最後に、コーパスサイズについて比較すると、学習データが少なくなるにつれて、認識性能は全体的に減少傾向にあるが、いずれの場合も意味属性 LM は ETL 単語 LM より性能が上回った。

### 3.2.2. 意味属性クラス言語モデルの考察

表 1 に学習コーパスのサイズに対する意味属性 LM の (平均) クラス数を示す。クラス数の増加に伴い評価データ中の意味属性クラスのカバー率が向上し、認識性能の向上につながったと考えられる。表 2 は各言語モデルによるパープレキシティ (平均) である。意味属性 LM のパープレキシティは、ETL 単語 LM と同程度の値を示している。学習コーパスの増加に対して、性能が向上する一方、パープレキシティが減少していないが、これは道案内タスクで利用される発話が、比較的単純な文法構造を持つためと考えられる。

他方、表 1 より意味属性クラスに割り当てられる語彙数は、一クラスあたり平均二単語程度しかない。そのため、クラスへの単語割り当てのスパースさを解消することで、認識性能の向上が期待できる (意味属性クラス言語モデルのクラス数と認識性能の関係については 4 節で詳述する)。

### 3.2.3. 自動クラスタリング手法との比較

クラスタリング手法の違いによる比較結果を図 4 に示す。ここでは単語 3-gram を含む三つの言語モデルを比較した。ただし、各モデルは 3.1 で示した③4/5 サイズのコーパスを用いて学習している。

- a) 単語 3-gram (ETL 単語 LM)
- b) 自動クラスタリング手法によるクラス 3-gram (自動クラス LM)

- c) コーパス中の名詞を、意味属性に写像した意味属性クラス 3-gram (意味属性 LM)

ETL 単語 LM とクラス言語モデル (自動クラス LM, 意味属性 LM) を比較すると、クラス言語モデルは単語 N-gram よりも高い性能を示す。一方、クラス言語モデル内での比較では、意味属性 LM は自動クラス LM と同程度の認識性能を示した。この理由としては、先に述べたように道案内が単純なタスクで、自動クラスタリングに十分な量の学習データが、利用できているためと考えられるが、これについては、学習データを減少させた実験による確認が必要である。

### 3.2.4. 意味属性によるクラスの初期化

3.2.3 で用いた自動クラス LM のクラスタリング結果を見ると、店名や地名などの語彙が、複数のクラスに

表 1. 意味属性 LM 中のクラス数

言語モデル	意味属性 クラス数	語彙数 (名詞のみ)	語彙数 (全て)
意味属性 LM (1/5)	186	378	698
意味属性 LM (2/5)	239	526	945
意味属性 LM (4/5)	312	736	1298

表 2. コーパス規模によるパープレキシティ比較

言語モデル	発話数	パープレキシティ
毎日単語 LM		115.6
ETL 単語 LM (1/5)	1751	12.6
ETL 単語 LM (2/5)	3502	12.0
ETL 単語 LM (4/5)	7003	11.8
意味属性 LM (1/5)	1751	13.7
意味属性 LM (2/5)	3502	13.5
意味属性 LM (4/5)	7003	13.6

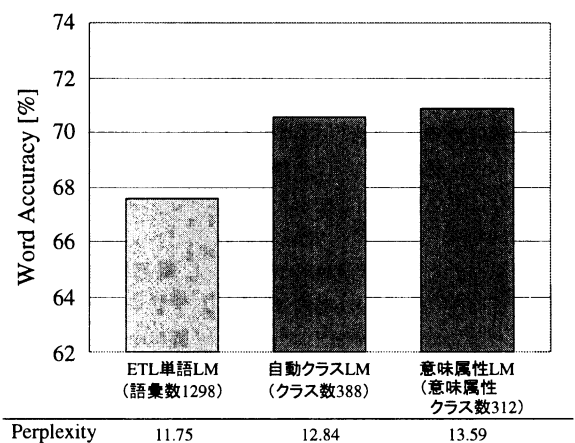


図 4. クラスタリング手法による性能比較

分かれ、それぞれで店名や地名クラスを構成していた。このことから、関連性の強い語彙セットを予め初期クラスターとして与え、関連性を持つ語彙同士が別クラスに分散しないようにすれば、文脈中の共起関係の信頼度が向上すると考えられる。そこで、3.3 に示した自動クラス LM の設計の際に、初期クラスターとして予め同じ意味属性を持つ語彙セットを与え、自動クラスターリングを行う方法（意味属性+自動クラス LM）について評価を行った。ただし、複数の意味属性を持つ語彙は、コーパス中の意味属性の頻度が最大となるものを、初期クラスとした。

意味属性+自動クラス LM の比較評価結果を図 5 に示す。同時に、クラスターリングに必要な計算量を比較するため、表 3 に実験環境を、また表 4 に各クラス言語モデルの構築に費やした時間をクラスターリング前後のクラス数と共に示した。まず認識性能をみると、図 5 から意味属性+自動クラス LM は、初期クラスを与えない自動クラス LM と同程度の単語精度を示した。次に計算量でみると、意味属性+自動クラス LM はクラスターリング時間を 30%以上縮めることができた。計算量の観点からみると、自動クラスターリング手法においても意味属性が有効であると言える。

## 4. 意味属性クラス数と性能の関係

### 4.1. 意味属性クラスの減少

3.2.2 で、意味属性クラス言語モデルは、語彙数に対してクラス数が約半分程度で構成されていることを示した（一クラス二単語程度）。本節では、意味属性クラスのクラス数を減少させたときの性能を比較する。

日本語語彙大系では、図 1 に示した木構造で意味属性を分類している。ここでは、上位層の意味属性が下位層の意味属性を包含することを踏まえ、クラス数を減少させる。以下にクラスを統合するルールを示す。

- (1) 末端の階層に属する意味属性クラスの内、頻度が最も少ないクラスを選択する、
- (2) 意味属性の木構造において、選択されたクラスを一つ上の階層のクラスにマージする、
- (3) 末端の階層に属する全ての意味属性が、設定した最小頻度以上になるまで(1) - (2) を繰り返す。

## 4.2. 実験概要

学習データは、① 1/5、② 2/5、③ 4/5 サイズのコーパスを利用し、さらに、末端のクラスの最小頻度を、2, 3, 4, 5, 10, 20, 30, 40, 50, 75, 100 に設定して、意味属性クラス言語モデルを作成した。

## 4.3. 実験結果と考察

末端の意味属性クラスの最小頻度毎に、性能（図 6）とパープレキシティ（図 7）を評価した。実験条件は 3 節と同じである。ここで、クラス数を減少させてい

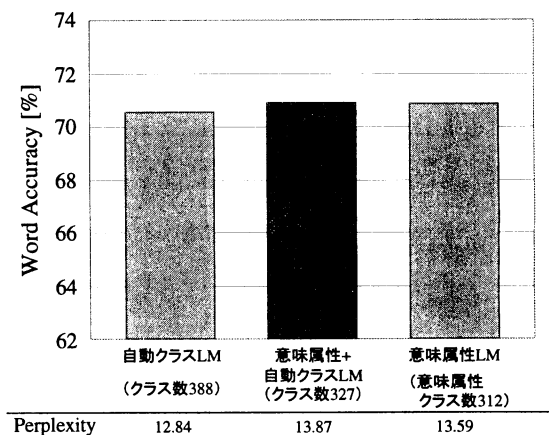


図 5. 意味属性を初期クラスに利用した場合の性能比較

表 3. 実験環境のマシンスペック

CPU	Intel Pentium4 (3.4GHz)
メモリ	2GB
OS	Linux kernel 2.4.18-14
HDD	300GB (10000 回転,RAID-0)

表 4. クラスターリングの所要時間とクラス数

	自動クラス LM	意味属性+自動クラス LM	意味属性 LM
クラスターリング前のクラス数	1298	984	736 (名詞の数)
クラスターリング後のクラス数	388	327	312
所要時間[分]	30768	21084	0.5

表 5. 末端の意味属性の頻度毎の意味属性クラス数

	Baseline	2	3	4	5	10	20	30	40	50	75	100
1/5 サイズ	186	178	170	163	157	134	109	93	85	78	63	55
2/5 サイズ	239	232	224	213	206	182	149	133	120	108	95	83
4/5 サイズ	312	303	292	280	275	246	212	186	172	157	136	117

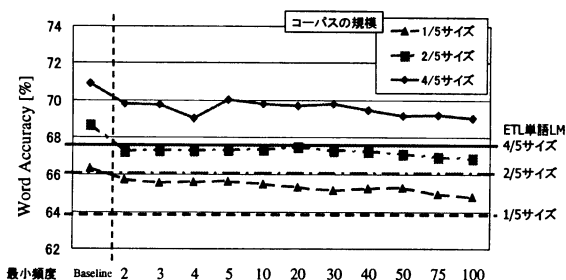


図 6. 末端の意味属性の頻度毎の性能比較

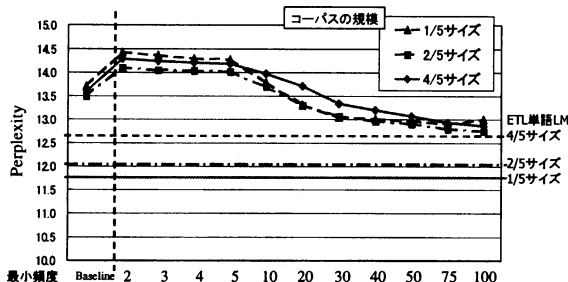


図 7. 末端の意味属性の頻度毎のパープレキシティ比較

ないモデルをベースラインとした。また、末端の意味属性クラスの最小頻度毎の意味属性クラス数を表 5 に示す。図 6 から、コーパス規模に依らず、ベースラインが最も高い性能を示すことが分かる。また、パープレキシティに関しては、ほとんど差がなかった(図 7)。意味属性クラス言語モデルが単語 3-gram よりも高い性能を示したのは、文脈中の意味を考慮したクラスが抽出された結果、共起関係の信頼度が向上したためと考えられる。

## 5. まとめ

意味属性を利用したクラス N-gram 言語モデルを提案し、評価結果を述べた。小規模コーパスを学習データとした音声認識実験の結果、提案手法は単語 N-gram を上回る性能を示した。一方、従来の自動クラスタリング手法と比較した結果は、性能差はわずかであった。また、意味属性を初期クラスとして与えた自動クラスタリング手法も性能上の改善はわずかであった。他方、提案手法は、クラスタリングに必要とされる膨大な計算時間が不要という大きな特長を持つ。今後は提案方式を利用して、対話音声の中のキーワードを抽出する実験[15]を行うと共に、講義音声などの別のドメインによる提案手法の利用・評価を検討したい。

## 文 献

- [1] 新田 恒雄, 浅見 弘道, 伊勢路 真吾, 福田 隆, 桂田 浩一, “汎用 LVCSR を用いた対話音声の認識” 情処研報 2002-SLP-41, pp.69-74, May 2002
- [2] 平川 竜士, 新田 恒雄, “意味属性を利用した小規模タスク用言語モデルの検討”, 電気関係学会東海支部連合大会, Nov. 2001
- [3] 池谷 晴生, 福田 隆, 山田 博文, 桂田 浩一, 新田 恒雄, “意味属性を利用したクラス N-gram 言語モデルの検討”, 音学講論 2-1-6, pp.47-48, Sep. 2004
- [4] 駒谷 和範, 上野 晋一, 河原 達也, 奥乃 博, “ユーザモデルを導入したバス運行情報案内システムの実験的評価”, 情処研究報告, SLP47-12, Num. 75 pp.59-64, Jul. 2003
- [5] 駒谷 和範, 鹿島 博晶, 田中 克明, 河原 達也, “複合的言語制約に基づくキーフレーズ検出を用いた汎用的なデータベース検索音声対話プラットフォーム”, 情処論文誌, Vol.44, No.5, pp.1333-1342, May 2003
- [6] 小川 厚徳, 今村 明弘, 外海 実, 中村 誠, 磯部 俊洋, 菅原 昌平, “日英シームレス音声認識技術による航空管制音声認識”, 情処研究報告, SLP-52, Num.74, pp.115-122, Jul. 2004
- [7] 加藤 正治, 小坂 哲夫, 好田 正紀, “参議院会議の音声認識”, 音学講論 2-1-2, pp.39-40, Sep. 2004
- [8] 山本 博史, 匂坂 芳典, “品詞 N-gram と単語 N-gram の融合言語モデル”, 信学技報 SP99-36, pp73-78, Jun. 1999
- [9] 山本 博史, 磯貝 俊太郎, 匂坂 芳典, “位置依存単語クラスと頻出単語列を用いた多重クラス複合 3-gram 言語モデル”, 信学技報 SP2001-33, pp13-18, Jun. 2001
- [10] 森 信介, 西村 雅史, 伊東 伸泰, “クラスに基づく言語モデルのための単語クラスタリング”, 情処論 Vol.38 No.11, pp2200-2208, Nov. 1997
- [11] 安藤 章男, 鈴木 順子, 白井 克彦, “シソーラスに基づく統計言語モデルの検討”, 信学技報 SP97-26, pp.59-66, Jun. 1997
- [12] NTTコミュニケーション科学基礎研究所, 日本語語彙大系, 岩波書店, 1997
- [13] <http://julius.sourceforge.jp/>
- [14] 伊藤 克亘, 秋葉 友良, 速水 悟, 田中 和世, “WOZ システムで収録した自由発話音声の対話データベース”, 音学講論 1-1-19, pp.37-38, Sep. 1998
- [15] 伊勢路 真吾, 福田 隆, 山田 博文, 桂田 浩一, 新田 恒雄, “日本語短・長音節単位の認識結果を用いた対話音声の中のキーワード検出”, 音学講論 3-Q-34, pp.211-212, Mar. 2004