

# 和音進行解析のための 語彙フリー無限グラムモデル

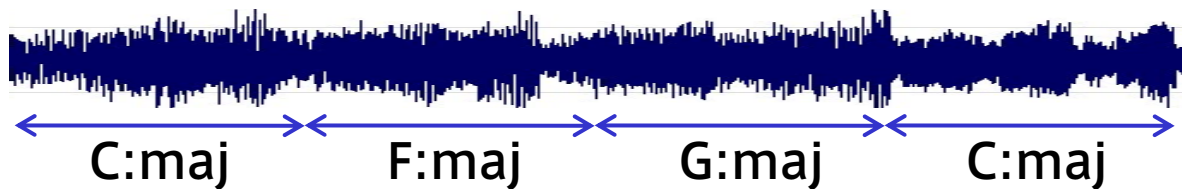
産業技術総合研究所

吉井 和佳 後藤 真孝

[k.yoshii@aist.go.jp](mailto:k.yoshii@aist.go.jp) [m.goto@aist.go.jp](mailto:m.goto@aist.go.jp)

# 研究の背景

- 基盤技術としての和音進行解析の重要性
  - 和音進行パターンは楽曲を特徴づける要素
    - 楽曲のジャンル [Perez-Sancho2009]
    - 作曲者のスタイル [Ogihara2008]
  - 音楽音響信号に対する和音系列認識への応用
    - 音響モデル：スペクトルの動的変化
    - 言語モデル：和音の連鎖に関する制約
  - 音楽音響信号に対する自動採譜への応用
    - 和音系列が分かれば音高候補の絞り込みが可能



CEG が演奏されている可能性が高い

GBD が演奏されている可能性が高い

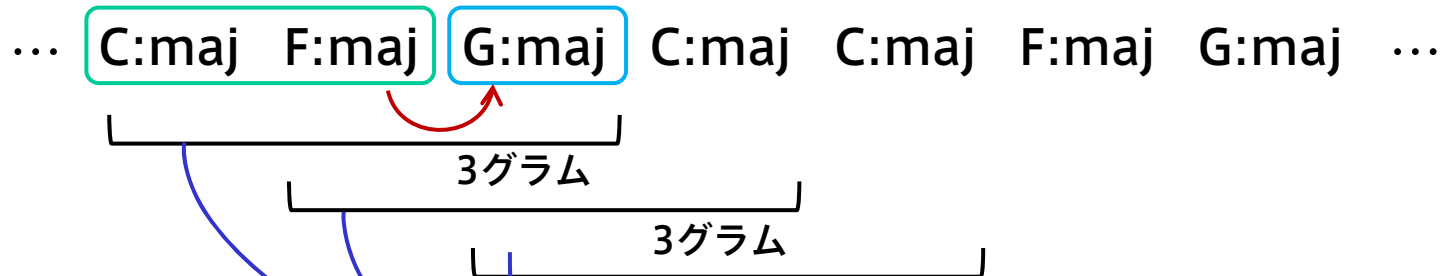
# 研究の目的

- 和音系列に対する高精度な言語モデルの学習

- $n$ グラムモデル**を利用

- $n$ グラム：ある系列中の長さ  $n$  のサブ系列
    - $n - 1$  次のマルコフ性を仮定

例：ある和音系列に対する 3 グラムモデル



P(和音系列)

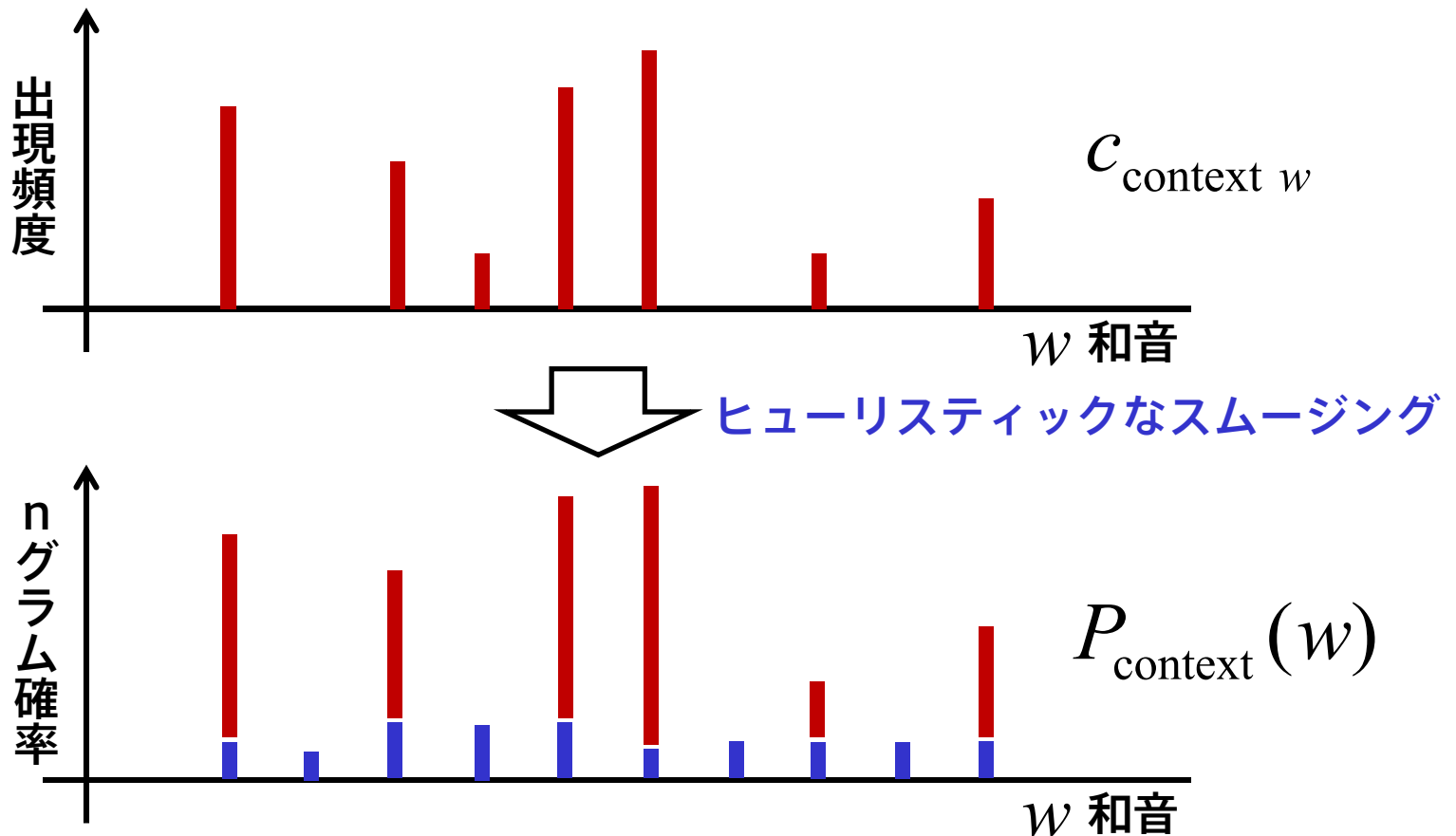
$$\begin{aligned} &= P(G:maj \mid C:maj F:maj) \\ &\times P(C:maj \mid F:maj G:maj) \\ &\times P(C:maj \mid G:maj C:maj) \\ &\times \dots \end{aligned}$$

学習データ (和音系列) から  
 $n$ グラム確率を推定したい

安直な方法：出現頻度を計算  
→ ゼロ頻度問題 (過学習)

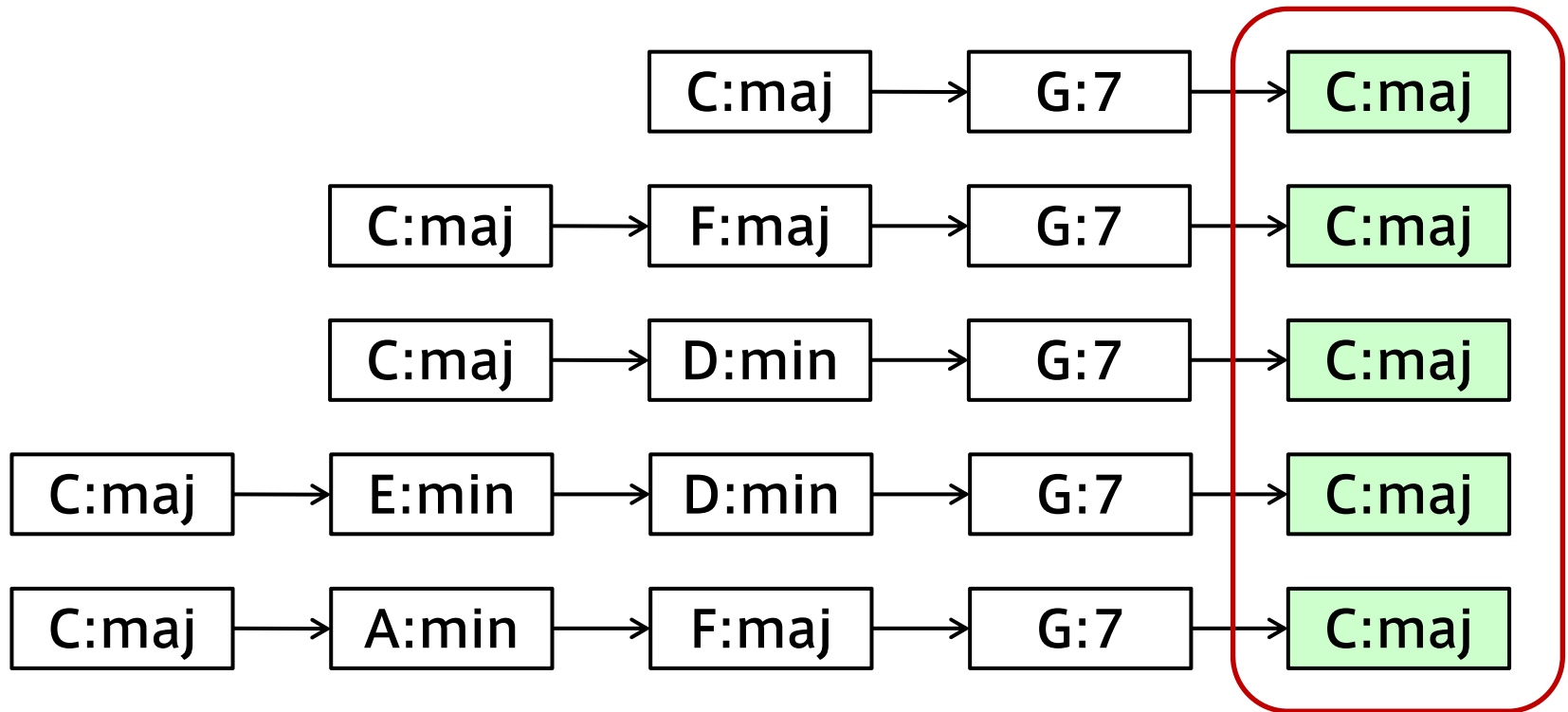
# 問題1: スムージング

- ゼロ頻度問題に対するヒューリスティックな対処
  - $n$ グラム確率分布を試行錯誤でスムージング
    - どのような基準のものに最適化すればいいか不明



# 問題2: $n$ の値

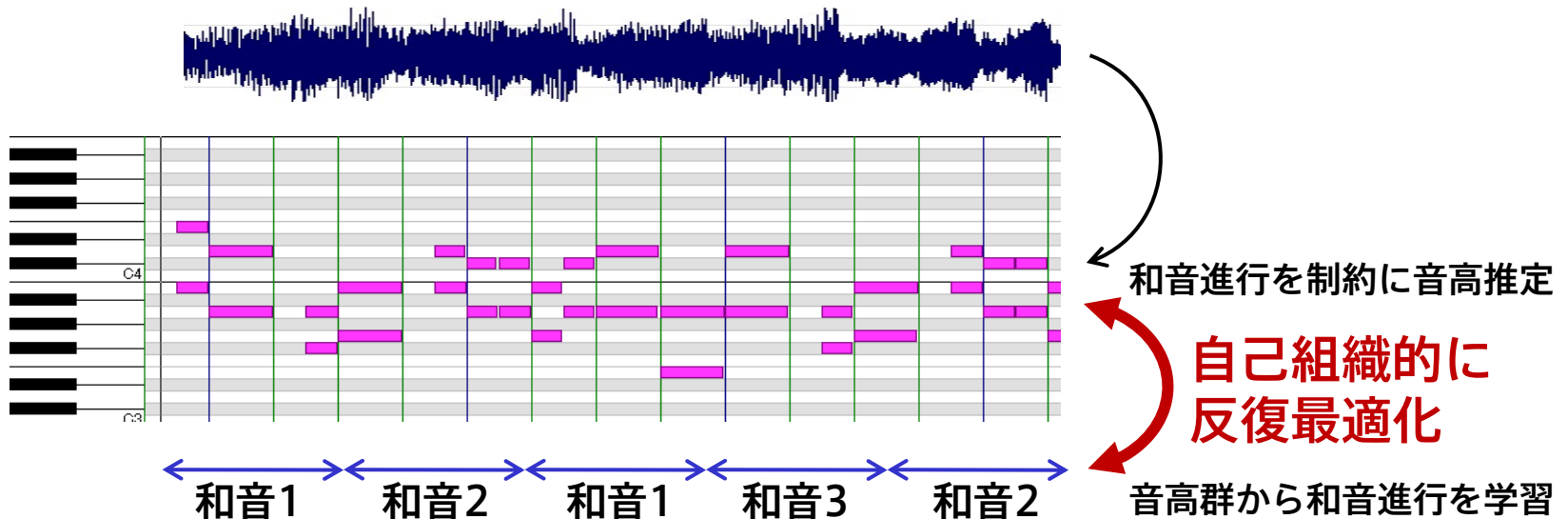
- 和音系列中の各和音のコンテキスト長は固定
  - 直前の何個の和音に影響を受けたかは常に一定
    - $n$ の値は固定・事前指定する必要あり



適切なコンテキスト長は異なるはず！

# 問題3: 和音語彙

- 音楽解析結果が和音の「ラベル」に依存
  - 和音系列認識タスク
    - 考慮すべき「コード」の恣意的な取捨選択が必要
      - maj, min, aug, dim, 7, 9, minmaj7, . . .
  - 自動採譜タスク
    - 音高の組み合わせを直接扱える方が望ましい場合あり
      - 教師なし音楽理解 (例: 「コード」の概念獲得)



# 課題とアプローチ

- 理論的なスムージングを行いたい
  - ノンパラメトリックベイズ化
    - $n$ グラム確率分布の階層的生成モデル [Teh2006]
- $n$ の値を固定する必要性をなくしたい
  - 無限グラム化
    - 各和音ごとにあらゆる $n$ の可能性を考慮 [持橋2007]
- 和音ラベルを定義する必要性をなくしたい
  - 語彙フリー化
    - 単語系列：文字のあらゆる並びを単語として許容 [持橋2009]
    - 和音系列：音高のあらゆる組合せを和音として許容 [本研究]

本研究の問題意識

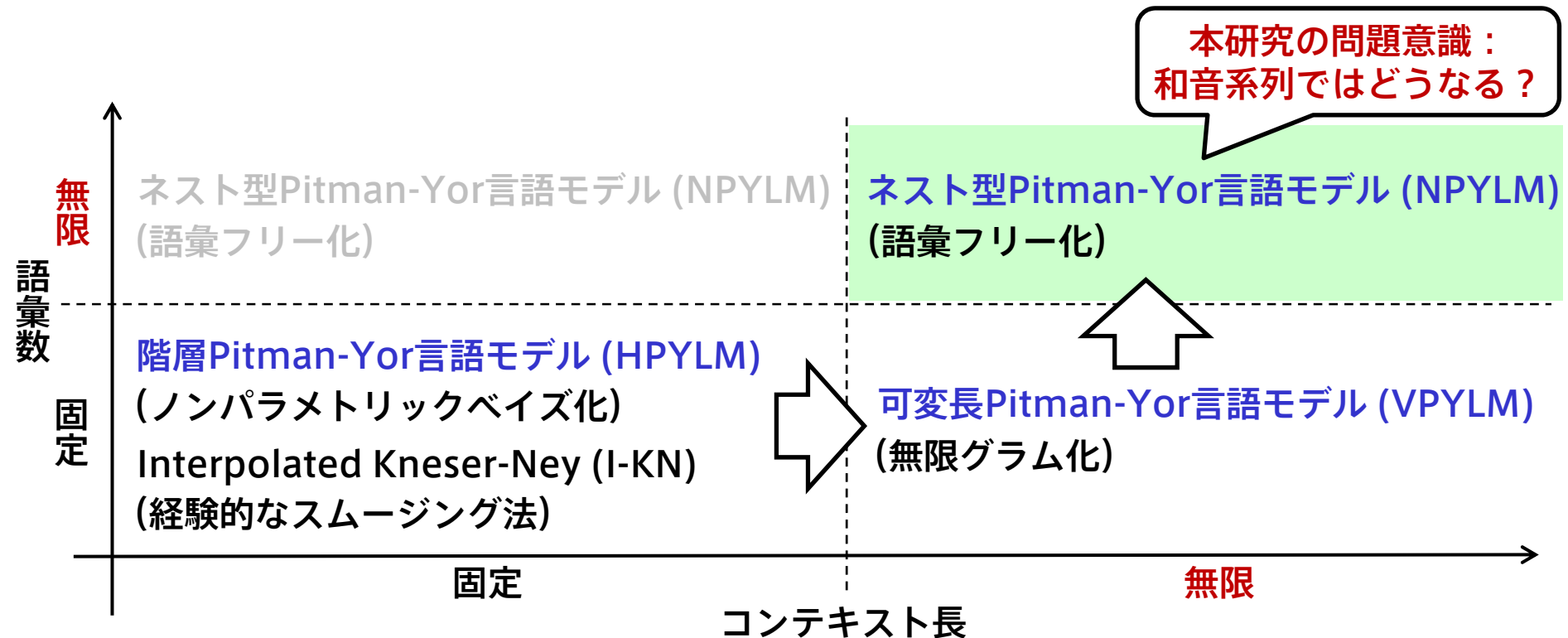


# 自然言語処理分野における発展

## • 単語系列に対する $n$ グラムモデル

### – ヒューリスティックな方法 から 理論的な方法 へ

- I-KN: 従来から良好な性能で知られていたスムージング法
- HPYLM: 最近提案されたノンパラメトリックベイズモデル
- I-KN は HPYLM の近似であることが判明！

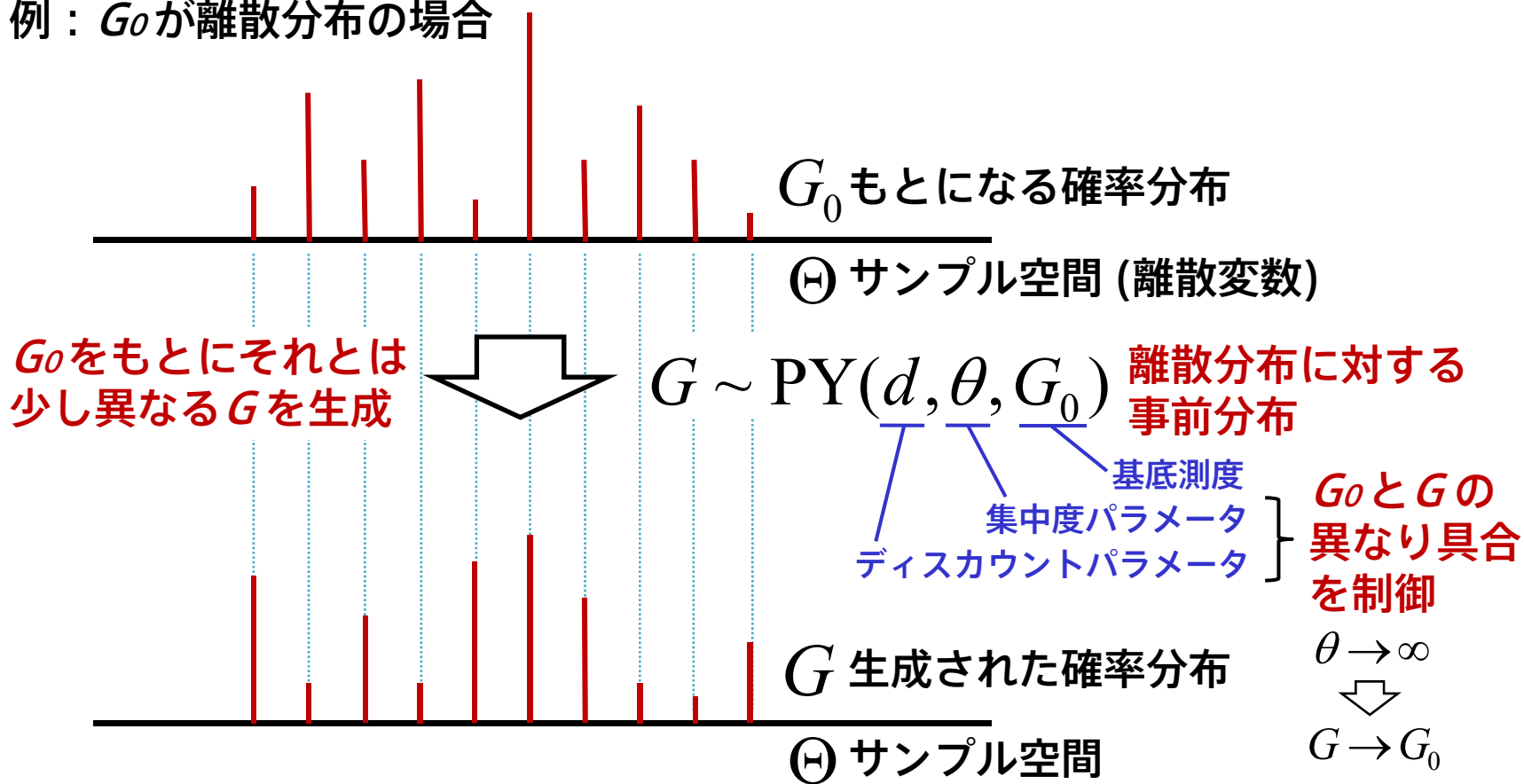




# Pitman-Yor過程

- PY: 確率分布上の確率分布
  - ディリクレ過程 (DP) の一般化
  - ある確率分布  $G$  に対する事前分布として利用可能

例:  $G_0$  が離散分布の場合

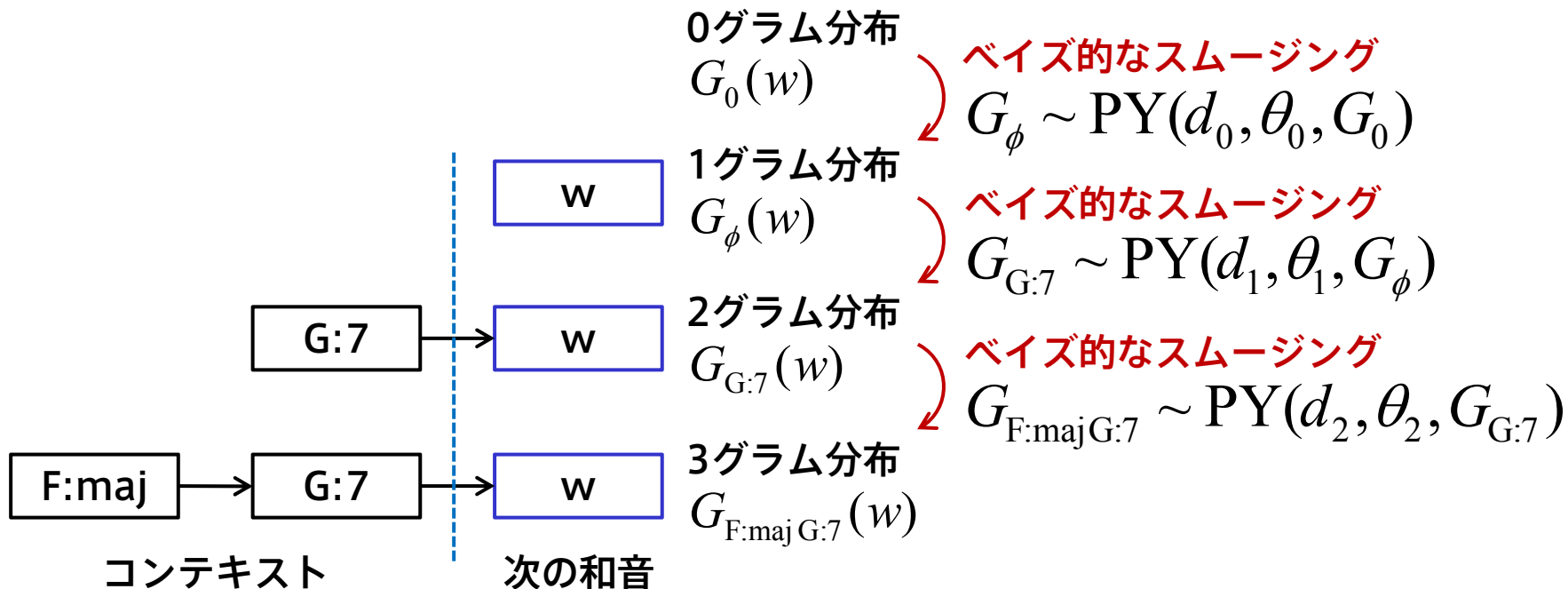


# $n$ グラムモデルへの応用

- PYを用いた $n$ グラム分布のスムージング

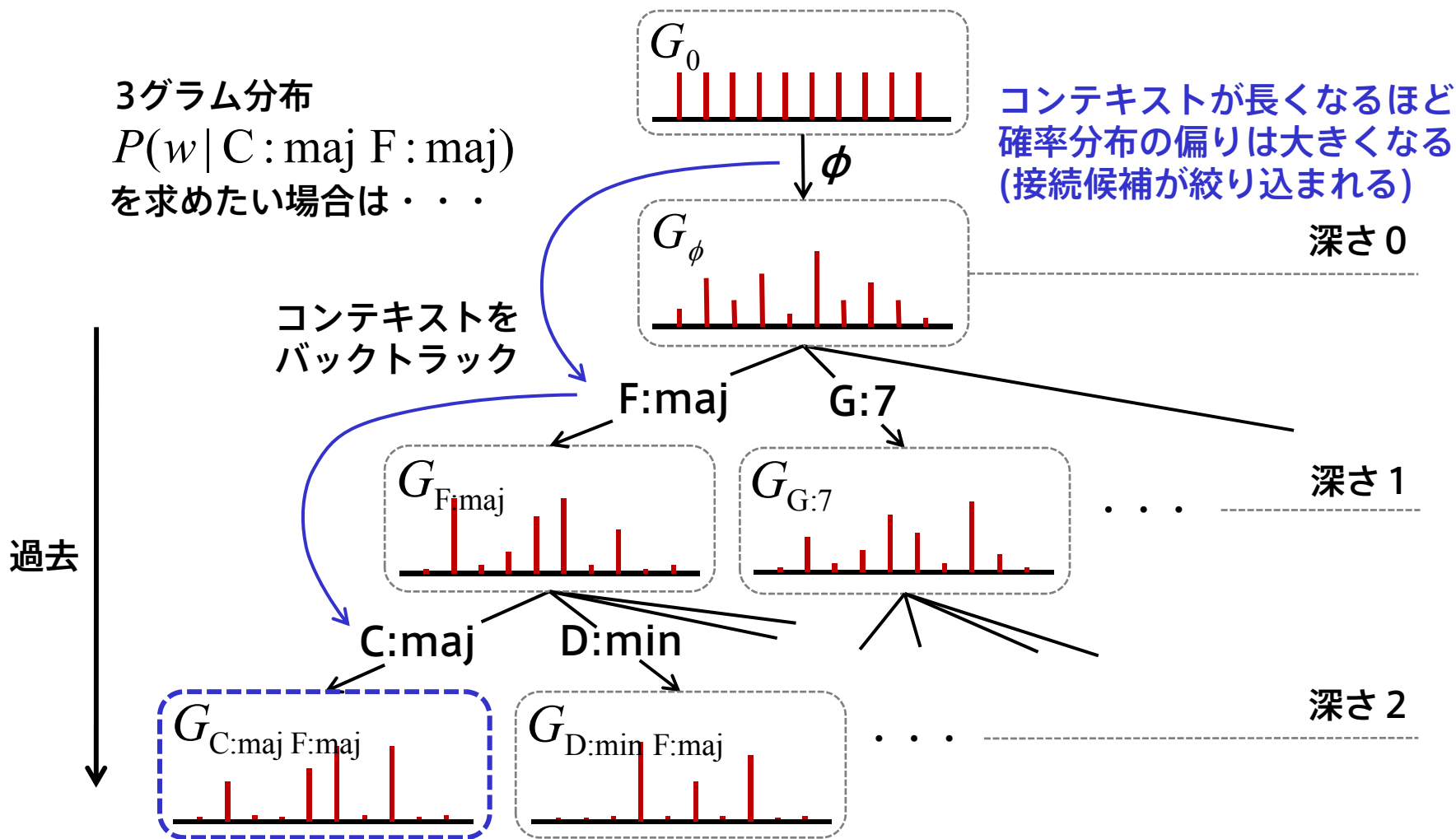
- 再帰的な定式化

- $n$ グラム分布は $n-1$ グラム分布を基底測度とするPYから生成
- $n-1$ グラム分布は $n-2$ グラム分布を基底測度とするPYから生成
- . . .
- ゼログラム分布には一様分布を仮定



# 1. ノンパラベイズ化

- 階層Pitman-Yor言語モデル (HPYLM)
  - 木構造を持つ  $n$  グラム分布群の階層的生成モデル



# 2. 無限グラム化

- 可変長Pitman-Yor言語モデル (VPYLM)

- 和音ごとに異なるコンテキスト長を許容

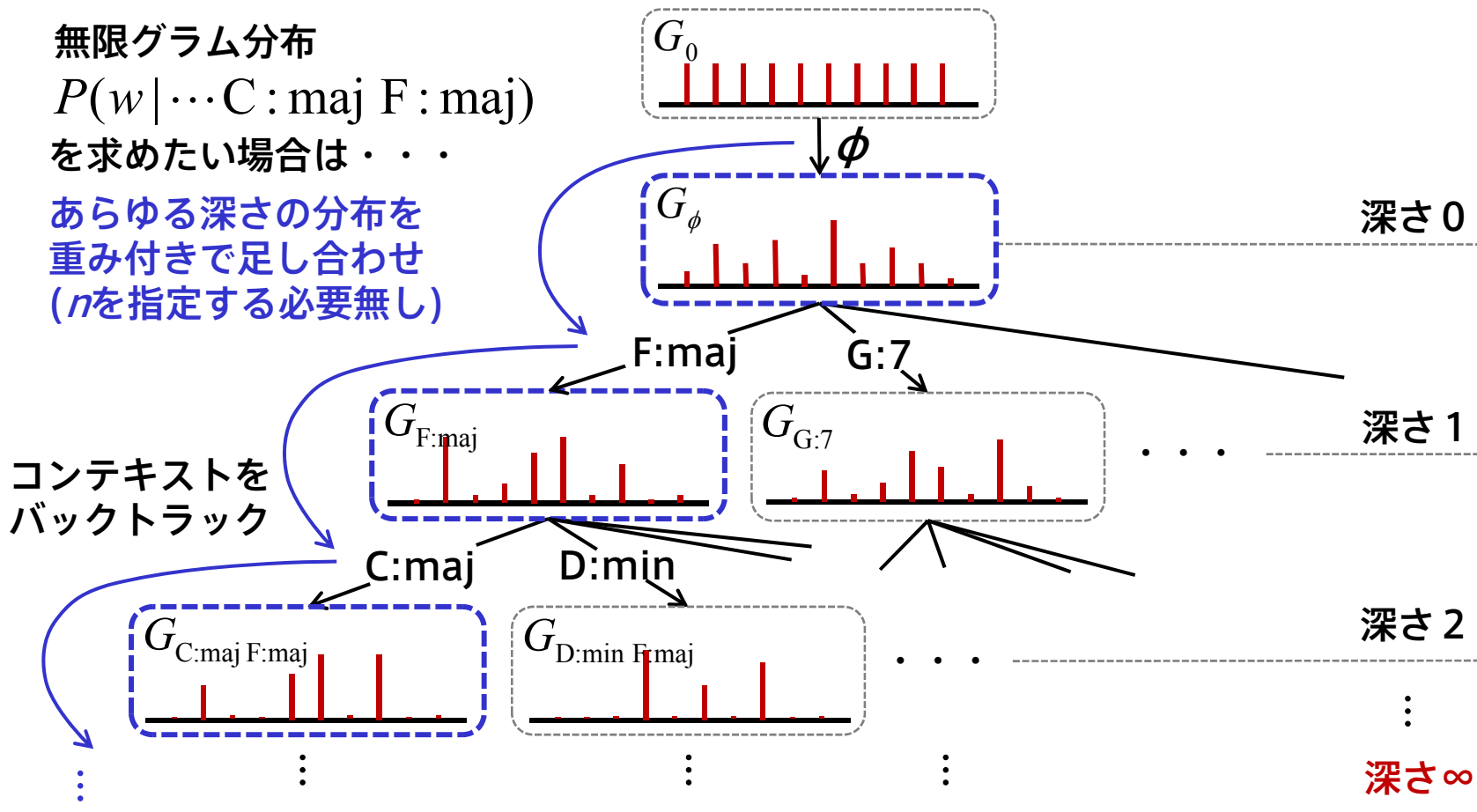
- 無限の深さ持つ木の生成モデル

無限グラム分布

$$P(w | \dots C:maj F:maj)$$

を求めたい場合は...

あらゆる深さの分布を  
重み付きで足し合わせ  
( $n$ を指定する必要無し)



# 3. 語彙フリー化

## 可変長Pitman-Yor言語モデル (NPYLM)

– あらゆる綴りを単語として許容 (無限語彙)

• 単語を構成する文字列が無限グラムモデルに従う

例: 

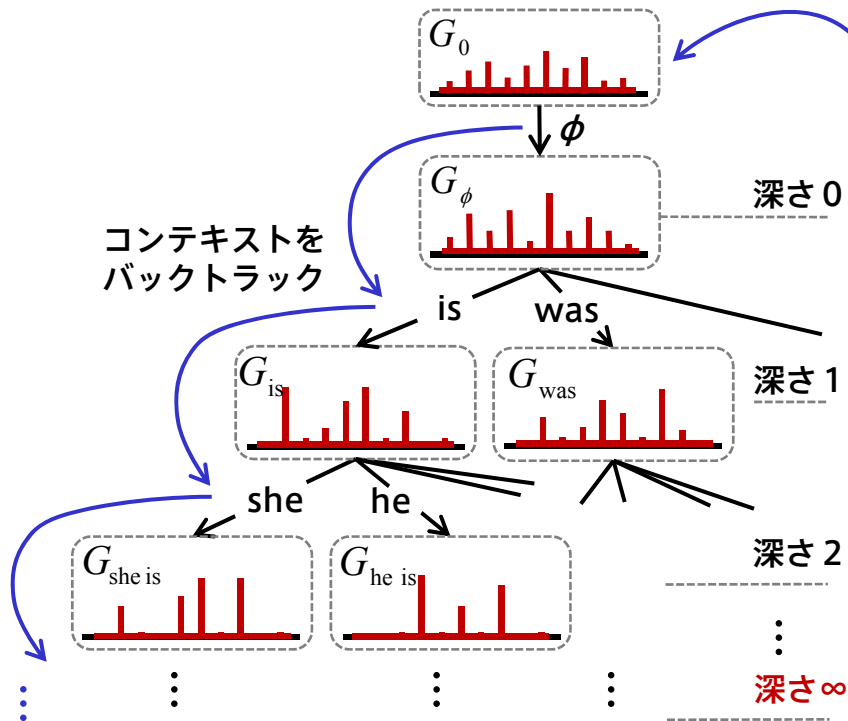
s	h	e
---	---	---

i	s
---	---

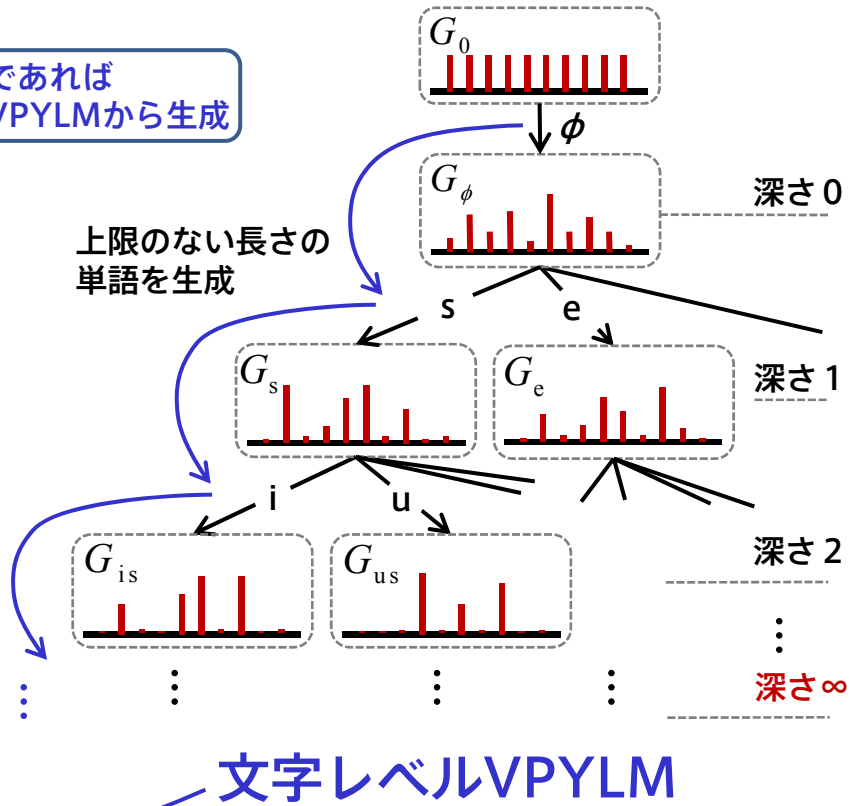
c	u	t	e
---	---	---	---

初出の単語であれば  
文字レベルVPYLMから生成

### 単語レベルVPYLM



上限のない長さの  
単語を生成



### 文字レベルVPYLM

# 本研究の目標

- 和音系列に対する語彙フリー無限グラムモデル
  - 無限グラム化
    - VPYLMを利用可能
  - 語彙フリー化 (楽譜を使えば無限語彙化も可能)
    - NPYLMは利用不可
      - 単語：文字の系列
    - 新たな基底測度が必要
      - 和音：音高の集合

本研究の問題意識：  
和音系列ではどうなる？



# 和音語彙再考

- 従来：ラベルに基づく表記法 [Harte2005]

- 語彙サイズ205

- 根音12種類 x ラベル17種類 + 非和音ラベルN

maj, min, dim, aug, maj7, min7, 7, dim7, hdim7, minmaj7,  
maj6, min6, 9, maj9, min9, sus2, sus4

- 本研究：構成音に基づく表記法

- 語彙サイズ49153

- 根音12種類 x 音高の組み合わせ  $2^{12}$  + 非和音ラベルN

カウント (非負整数値) に  
着目すれば無限語彙!

12種類のピッチクラスの存在有無

C:maj	→	C:100010010000
D:maj	→	D:100010010000
D:min	→	D:100100010000
D:maj add4	→	D:100011010000

構成音の根音に対する  
相対的な位置に着目

対応するラベルがなく  
てもあらゆる音高の  
組み合わせが表現可能

# 提案モデル

- 和音系列に対する語彙フリー無限グラムモデル

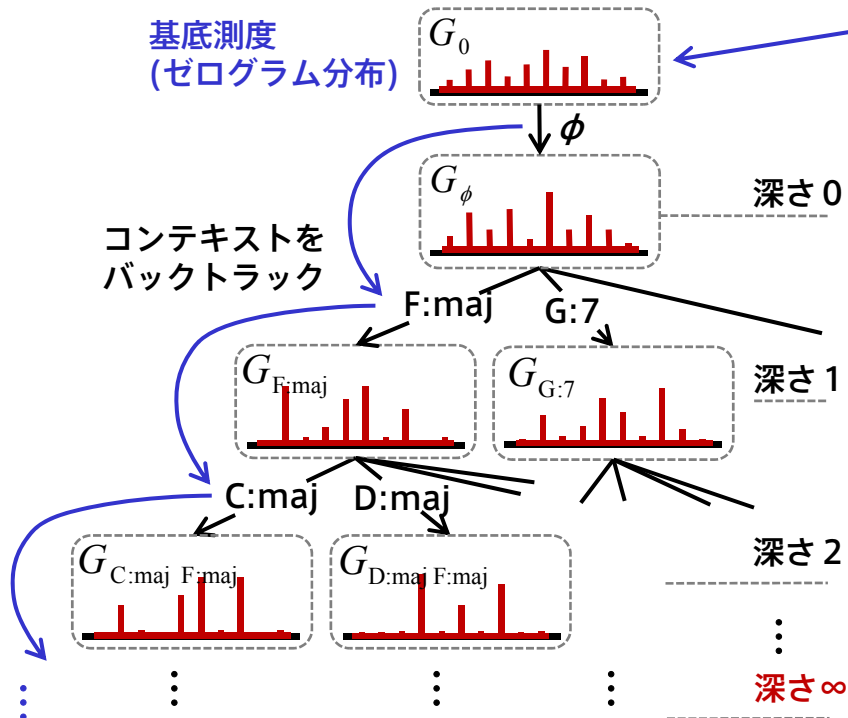
- 構成音に基づくゼログラム分布  $G_0$  の定式化

- 根音+Nの選択: 13次元離散分布 各音高の有無: ベルヌイ分布

例: C:maj F:maj G:7

初出の和音であれば和音生成モデルから生成

## 和音の無限グラムモデル (VPYLM)



## 音高の組み合わせモデル (ゼログラム分布の生成モデル)

$$G_0(C : 1000100100 00) = \pi_C \tau_1 \tilde{\tau}_2 \tilde{\tau}_3 \tilde{\tau}_4 \tau_5 \tilde{\tau}_6 \tilde{\tau}_7 \tau_8 \tilde{\tau}_9 \tilde{\tau}_{10} \tilde{\tau}_{11} \tilde{\tau}_{12}$$

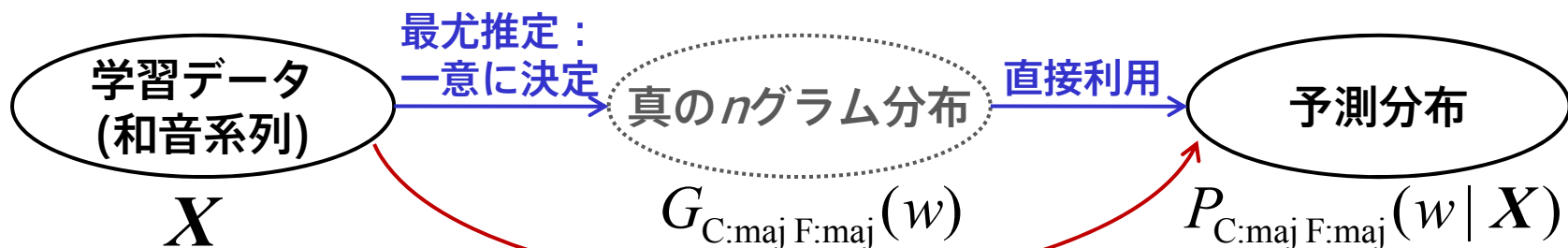
$$G_0(F : 1000100100 00) = \pi_F \tau_1 \tilde{\tau}_2 \tilde{\tau}_3 \tilde{\tau}_4 \tau_5 \tilde{\tau}_6 \tilde{\tau}_7 \tau_8 \tilde{\tau}_9 \tilde{\tau}_{10} \tilde{\tau}_{11} \tilde{\tau}_{12}$$

$$G_0(G : 1000100100 01) = \pi_G \tau_1 \tilde{\tau}_2 \tilde{\tau}_3 \tilde{\tau}_4 \tau_5 \tilde{\tau}_6 \tilde{\tau}_7 \tau_8 \tilde{\tau}_9 \tilde{\tau}_{10} \tilde{\tau}_{11} \tau_{12}$$



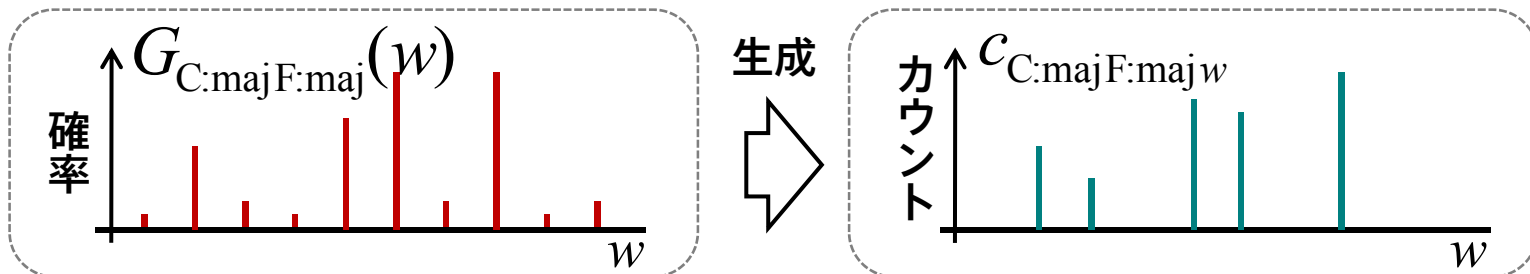
# 予測分布のベイズ推定

- あるコンテキストに続く和音確率のベイズ的計算
  - 未知の  $n$  グラム分布を点推定せずに積分消去
    - 学習データが与えられた上で、あらゆる可能性を考慮



ベイズ推定: あらゆる可能性を考慮 (積分消去)

$$P_{C:maj F:maj}(w | X) = \int G_{C:maj F:maj}(w) P(G | X) dG$$



$G$ が未知でも和音  $w$  のカウント (学習データ  $X$  の生成過程) から次和音予測が可能! → ギブスサンプリング

# 評価実験

- 実験データ

- Beatlesコードアノテーション [Harte2005]

- 長調の137曲を対象・ハ長調に転調して使用
    - 総和音数：10761

- 比較手法 (9種類)

- 古典的なスムージング法

- Good-Turing (GT), Witten-Bell (WB)
    - 低次  $n$  グラム分布を用いた補間版: I-WB

- 性能が良いことが実験的に知られている手法

- Kneser-Ney (KN), Modified Kneser-Ney (MKN)
    - 低次  $n$  グラム分布を用いた補間版: I-KN, I-MKN

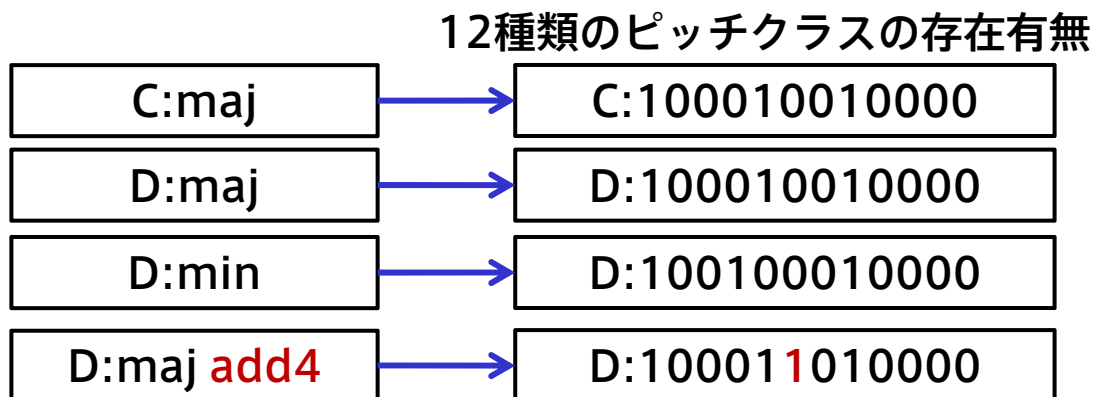
- ノンパラメリックベイズモデル

- 固定長  $n$  グラムモデル (HPYLM)
    - 可変長  $n$  ラムモデル (VPYLM)

パープレキシティで評価  
(次和音の平均候補数)  
低いほど高性能

# 評価実験

- 実験1: 無限グラムモデル (VPYLM) の性能評価
  - ラベルに基づく表記法を利用 (語彙サイズ205)
    - 根音12種類 x ラベル17種類 + 非和音ラベルN
  - 9手法を比較
- 実験2: 語彙フリーモデル (VF-) の性能を評価
  - 構成音に基づく表記法を利用 (語彙サイズ49153)
    - 根音12種類 x 音高の組み合わせ  $2^{12}$  + 非和音ラベルN
  - 9手法・VF-HPYLM・VF-VPYLMを比較



語彙フリー  
無限グラムモデル

構成音の根音に対する  
相対的な位置に着目

対応するラベルがなく  
てもあらゆる音高の  
組み合わせが表現可能

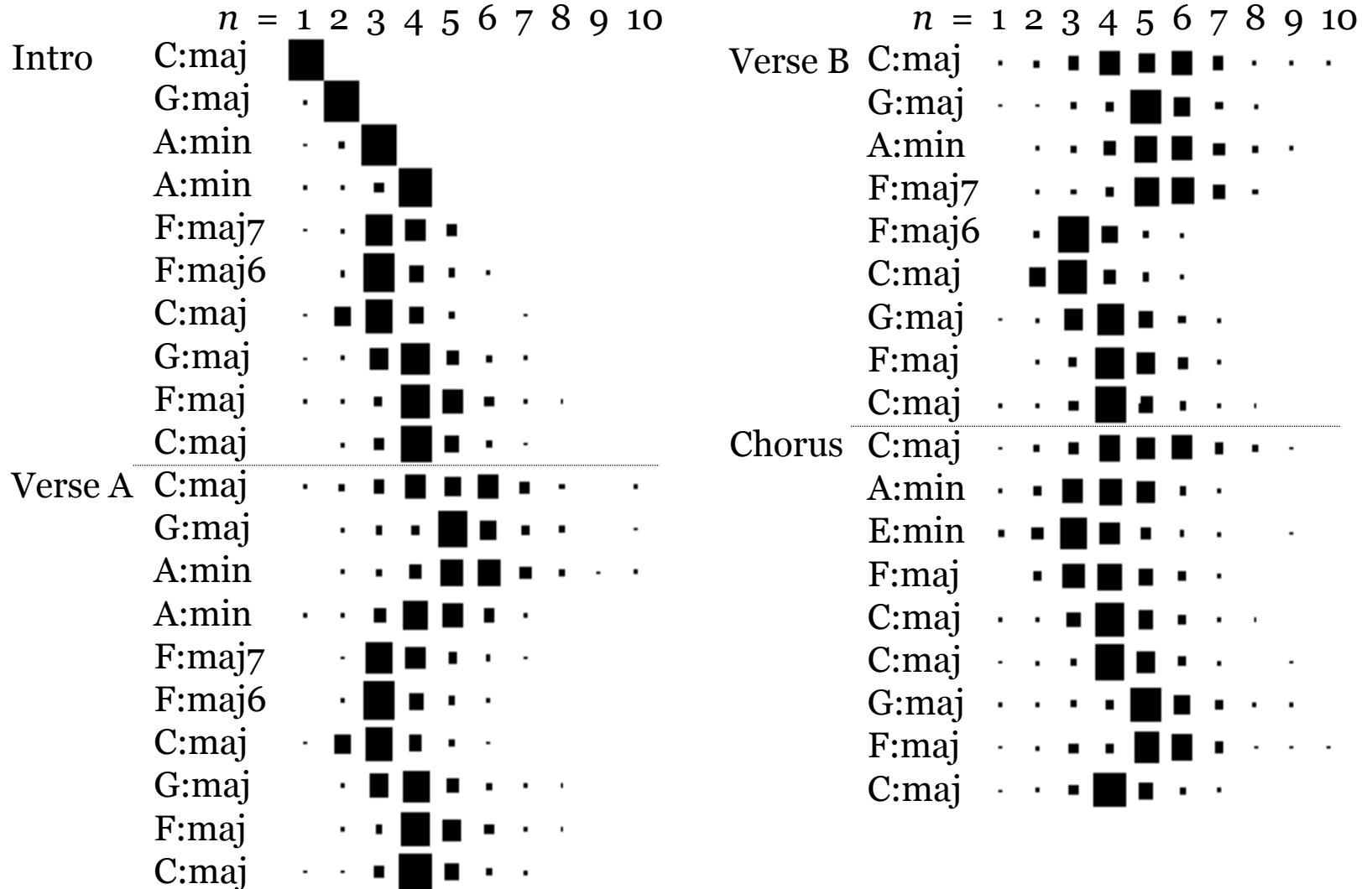
# 実験1: パープレキシティ評価

- 無限グラムモデル (VPYLM)が最高性能
  - $n$ を一意に決めずにあらゆる可能性を考える有効性

$n$	GT	WB	I-WB	KN	I-KN	MKN	I-MKN	HPYLM	VPYLM
1	16.8	15.6	15.7	15.8	16.0	15.6	15.7	15.8	
2	20.3	14.2	15.5	14.5	15.2	14.5	15.8	14.5	n: サンプル
3	23.5	15.4	18.5	16.1	16.0	16.1	16.3	16.0	13.4
4	25.5	16.8	22.5	16.8	17.7	16.5	15.5	13.9	
5	26.3	17.5	25.4	15.5	16.2	16.0	14.1	13.7	n: MAP値
6	27.0	17.8	27.2	14.6	15.1	16.2	13.5	13.6	12.9
7	27.3	18.0	28.3	14.3	14.5	16.4	13.3	13.6	
8	27.3	18.0	28.8	14.1	14.2	16.6	13.2	13.6	n: 積分消去
9	26.3	18.0	29.0	14.0	14.1	16.8	13.1	13.5	11.9
10	27.3	18.0	29.2	14.0	14.0	16.7	13.1	13.5	

# 「 $n$ 」の推定例

- Beatles 「Let It Be」 冒頭













# 実験1: 確率句の列挙

- 統計的に特徴的な和音進行パターンの発見

- $P_{\text{context}}(w, n | X)$  が大きい順にソート

- 長さ  $n$  の和音進行パターン  $context$   $w$  の生成確率  
長さ  $n-1$  の和音列 和音

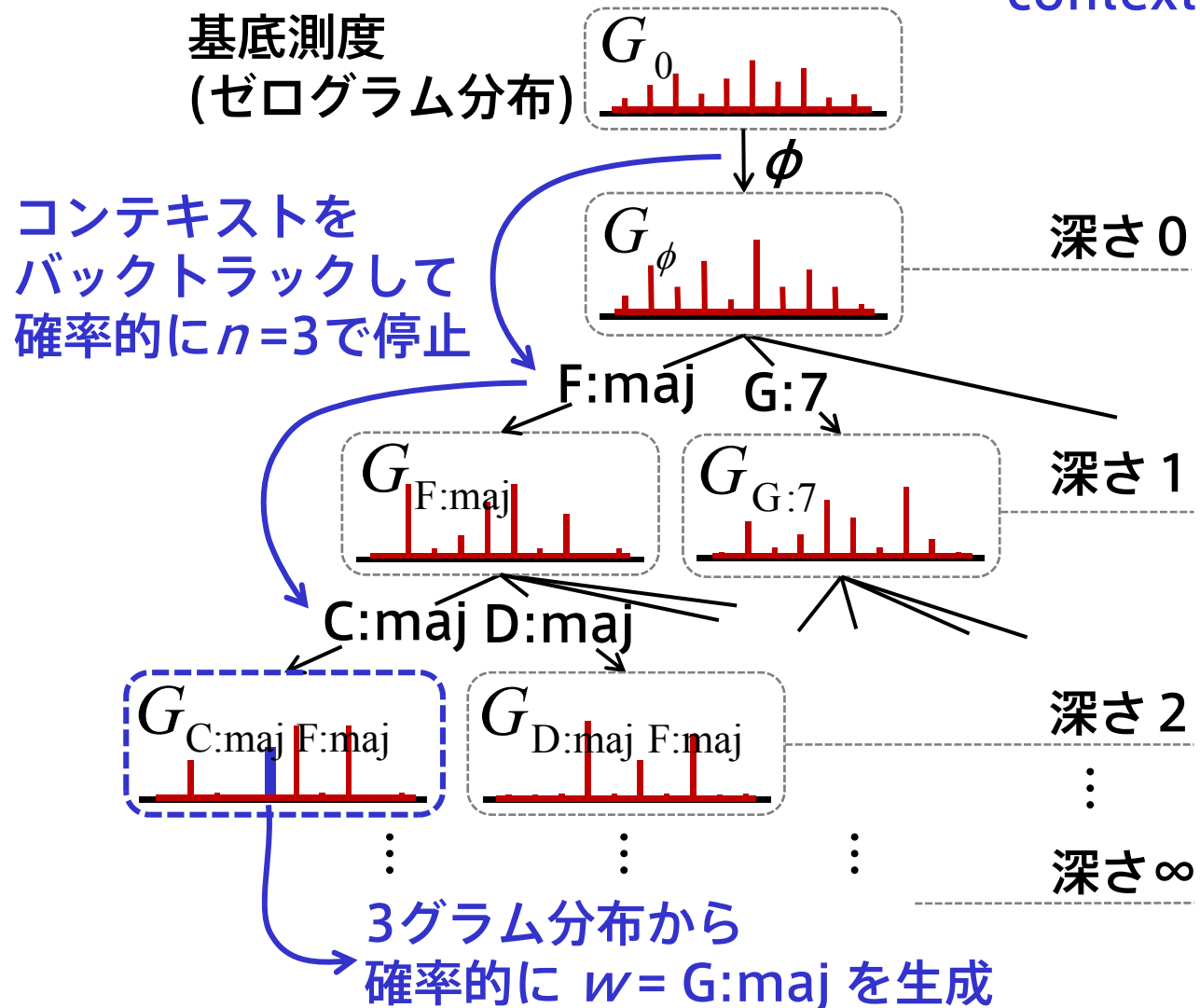
生成確率	和音進行パターン (nが3以上のものを表示)
 0.701	C:7 F:7 C:7
 0.682	B:maj F:maj G:maj
 0.656	A:min C:7 F:maj
 0.647	F:min G:maj C:maj
 0.645	F:maj F:maj G:maj C:maj
 0.632	E:min C:7 F:maj
 0.630	C:maj7 D:min7 E:min7
 0.623	B:maj F:maj G:maj C:maj
 0.622	D:min7 G:sus4 G:maj
 0.620	D:min G:maj C:maj F:maj C:maj

典型的なブルース進行

全体的にセブンスが多用  
(当時は珍しかった)

# 和音進行パターンの生成

- C:maj F:maj G:maj の生成確率  $P_{\text{C:maj F:maj}}(\text{G:maj}, 3 | \mathbf{X})$   
context  $w$   $n$



# 実験2: パープレキシティ評価

- 語彙フリー無限グラムモデル (VF-VPYLM) が最高性能
  - 構成音に基づく精緻なゼログラム分布モデルの有効性
  - 語彙サイズが大きくなると優位性が拡大
    - 学習データがスパースであっても頑健なスムージングが可能!

$n$	GT	WB	I-WB	KN	I-KN	MKN	I-MKN
10	38.3	24.4	39.1	18.4	18.5	21.7	17.5

$n$	HPYLM	VF-HPYLM
10	18.0	16.5

語彙フリー化

語彙フリー固定長  $n$  グラムモデル

$n$	VPYLM	VF-VPYLM
$\infty$	15.8	14.6

語彙フリー無限グラムモデル



# まとめと今後の課題

- 和音系列のための語彙フリー無限グラムモデル
  - 階層ノンパラメトリックベイズモデル [Teh2006]
    - 確率モデルの最適化の枠組みで $n$ グラム分布を推定
  - 無限グラム化 [持橋2007]
    - 各和音ごとに異なるコンテキスト長を持つことを許容
  - 語彙フリー化 [本研究のキーポイント]
    - 音高の組み合わせに基づくラベルフリーな和音表現
- 今後の課題：音楽音響信号解析への応用
  - 和音系列認識
    - 和音ラベルが与えられている
      - 無限グラムモデルを言語モデルとして利用
  - 複数基本周波数推定
    - 和音ラベルが与えられていない
      - 語彙フリー無限グラムモデルを自己組織化に利用