

# 8. 音声の認識：言語モデル

- 8.1 文法記述による言語モデル
- 8.2 統計的言語モデルの考え方
- 8.3 統計的言語モデルの作り方
- ニューラルネットワーク言語モデル

# 8.1 文法記述による言語モデル

- 文法記述を言語モデルとみなす
  - 規則に従う単語列であれば  $P(W) > 0$ 、そうでなければ  $P(W) = 0$  として、認識対象の単語列を限定
- 文法記述の例

\$ 文  $\rightarrow$  \$ 表示 | \$ 設定 | \$ 検索  
\$ 表示  $\rightarrow$  地図 を 表示  
\$ 設定  $\rightarrow$  \$ 登録場所 へ  
\$ 検索  $\rightarrow$  \$ 手段 で 検索  
\$ 登録場所  $\rightarrow$  自宅 | 職場  
\$ 手段  $\rightarrow$  住所 | 名称 | 履歴

\$ マークの付いている記号は  
非終端記号で, 必ず規則によって  
展開されます.

\$ マークの付いていないものは,  
終端記号で, 単語に相当します.



図 8.1 文法記述の例

# 8.1 文法記述による言語モデル

- 文法の種類
  - 文脈自由文法
    - 文法規則の左辺は非終端記号一つ
    - 右辺は「終端記号または非終端記号」の列
    - おおよそ自然言語の文法が記述可能
  - 正規文法
    - 文法規則の左辺は非終端記号一つ
    - 右辺は「終端記号」、「終端記号 + 非終端記号」、「空文字列」のいずれか
    - おおよそ文節レベルの文法が記述可能
  - 典型的な音声対話システムの文法は、正規文法の範囲内で記述可能

## 8.2 統計的言語モデルの考え方

- 統計的言語モデル

- $P(W) = P(w_1, \dots, w_n)$  の値を言語統計から求める
- 条件付き確率への展開

$$P(w_1, \dots, w_n) = P(w_1)P(w_2|w_1)P(w_3|w_1, w_2) \cdots P(w_n|w_1, \dots, w_{n-1})$$

- N-グラム言語モデル

- 長い履歴を持つ条件付き確率の値の推定は難しい
- 履歴を、過去N-1単語で近似

$$P(w_1, \dots, w_n) = P(w_1)P(w_2|w_1) \prod_{k=3}^n P(w_k|w_{k-1}, w_{k-2})$$

## 8.2 統計的言語モデルの考え方

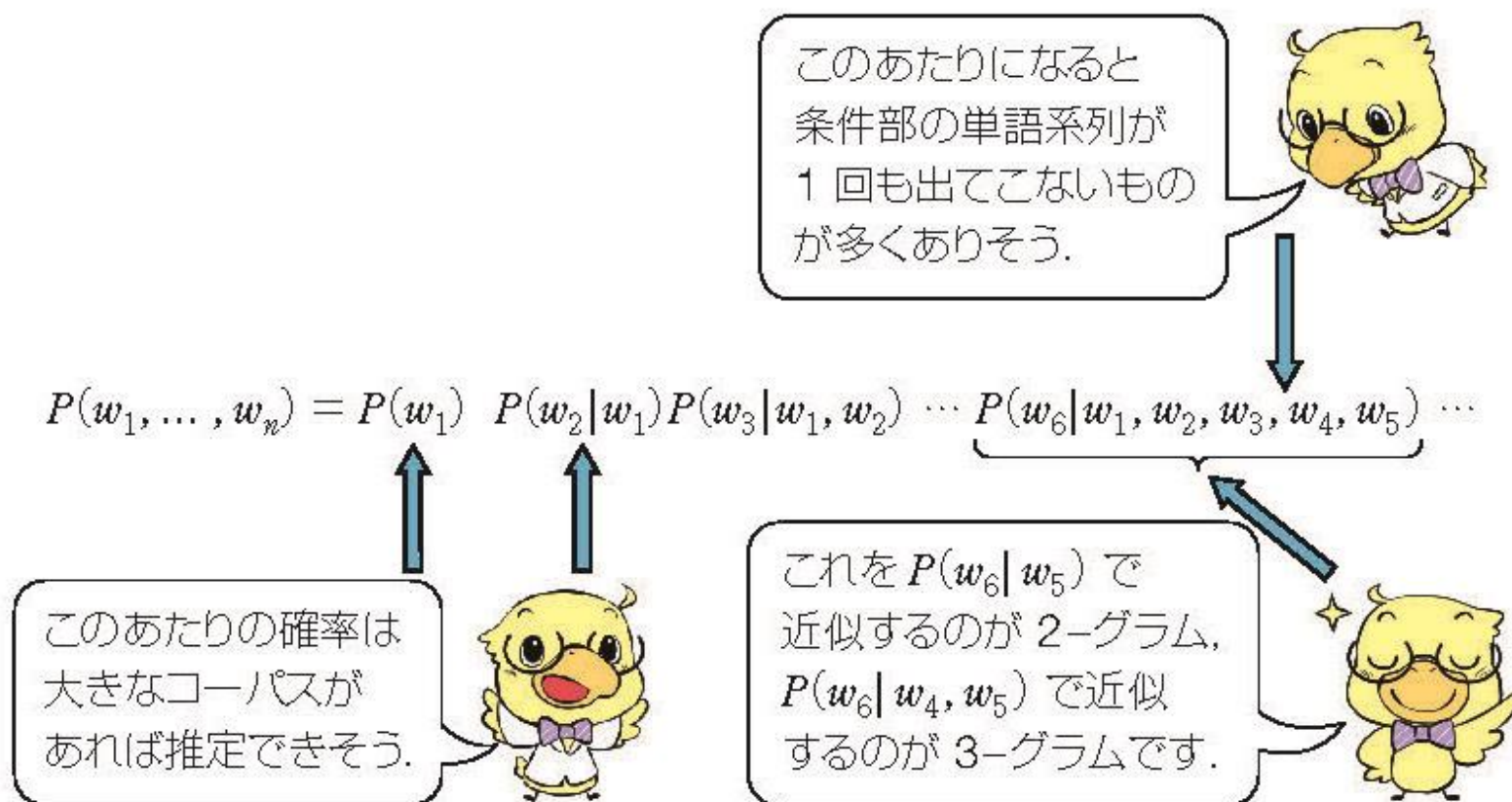


図 8.2 N-グラムによる  $P(W)$  の近似

## 8.3 統計的言語モデルの作り方

### 1. コーパスを準備する

大量の電子化された文章（新聞記事、webページなど）を集める

### 2. コーパスを単語に区切る

形態素解析処理

### 3. 条件付き確率を求める

確率の推定値が0にならないよう工夫したうえで

$P(w_k \mid w_{k-N+1}, \dots, w_{k-1})$  を求める

## 8.3 統計的言語モデルの作り方

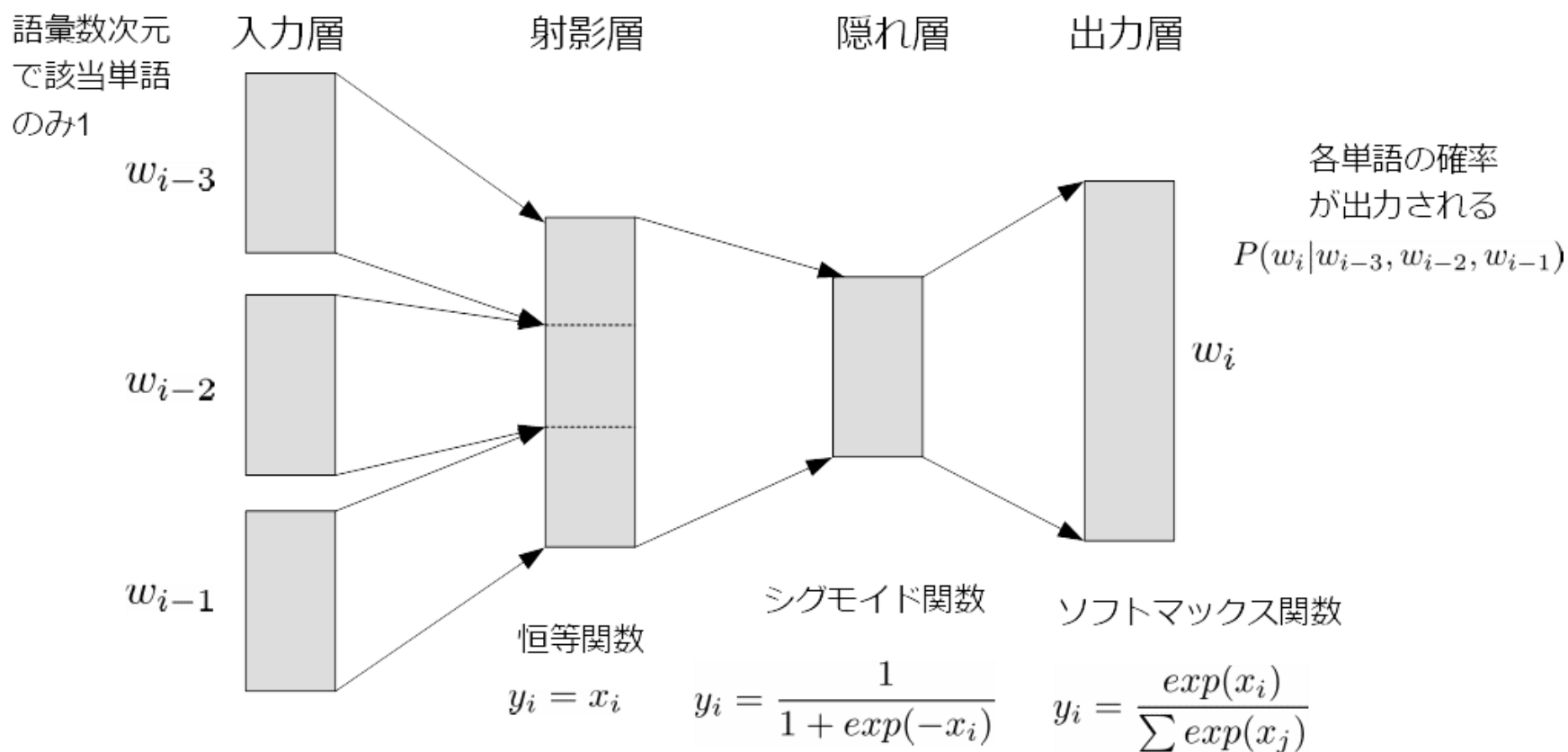
- N-グラムを最尤推定するときの問題点
  - 例) 2-グラムの単純な最尤推定

$$P_{ML}(w_i | w_{i-1}) = \frac{C(w_{i-1}, w_i)}{C(w_{i-1})} \quad C(W): W \text{の出現回数}$$

- コーパス中に  $w_{i-1} w_i$  が1度も出現しなければ、この値は0
  - 単語列中に値0の2-グラムが1つでもあれば、全体の確率が0
- バックオフスムージング
  - 最尤推定したN-グラムのうち、確率0でないものから少しずつ値を削り、確率0のものに分配する

# ニューラルネットワーク言語モデル

- フィードフォワード型
  - 過去N単語から次単語の確率分布を求める





# ニューラルネットワーク言語モデル

- リカレント型
  - フィードバックで仮想的にすべての履歴を表現

