

11. 音声の認識：WFST による音声認識

11.1 WFSTによる音声認識の概要

11.2 音響モデルをWFST に変換する

11.3 発音辞書をWFST に変換する

11.4 言語モデルをWFST に変換する

11.5 WFST の探索

11.1 WFSTによる音声認識の概要

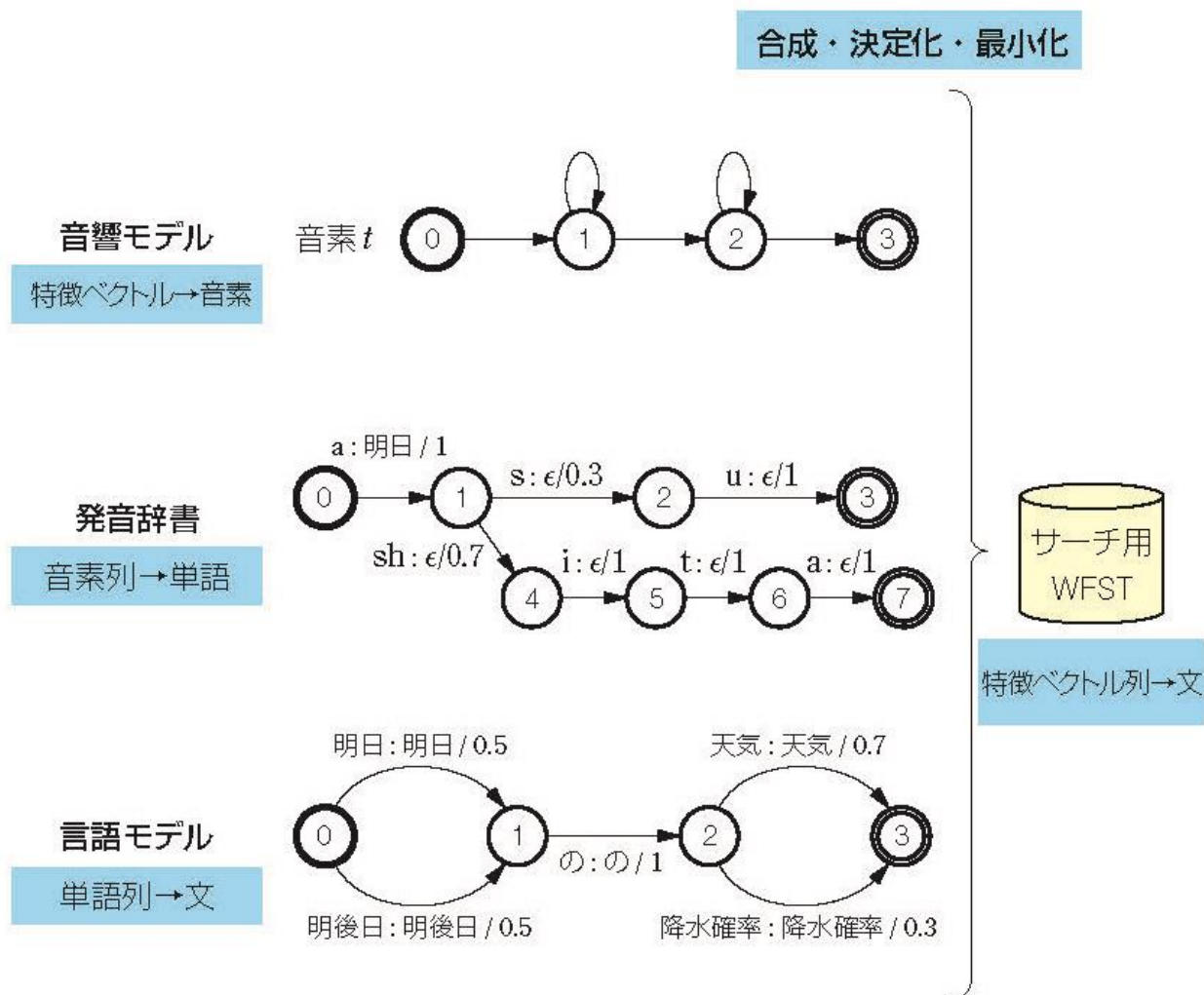
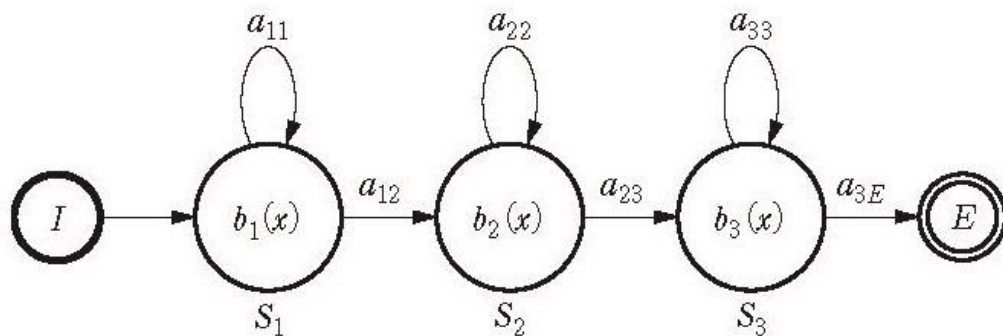


図 11.1 WFST による音声認識の概要

11.2 音響モデルをWFSTに変換する

- HMMをWFSTに変換



HMMは
入力を持たない

図 11.2 音素 HMM

- すべての特徴ベクトルを現す記号 \mathbf{x} を入力として導入

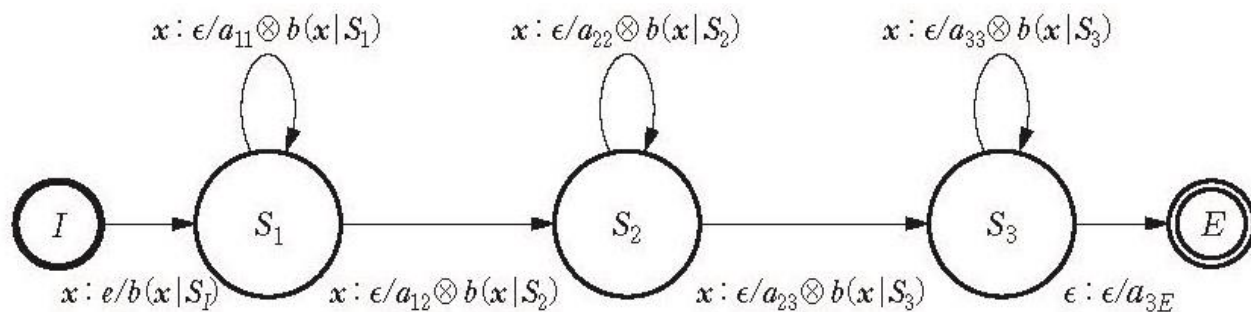


図 11.3 HMM から変換された WFST

$b(\mathbf{x} | S_i)$ の値が認識時までわからないので、事前に合成ができない

11.2 音響モデルをWFST に変換する

- WFSTの分離

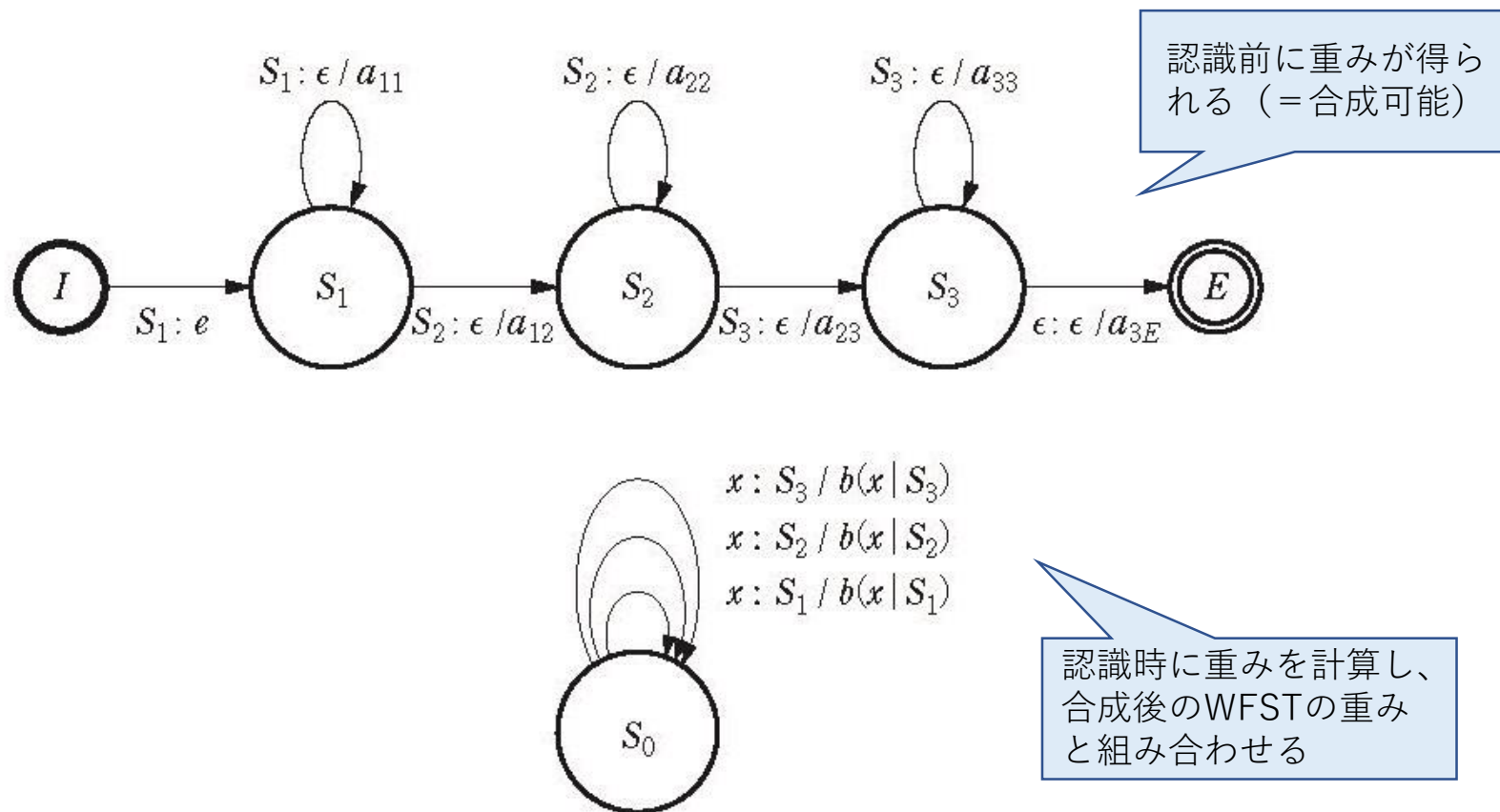


図 11.4

HMM を表す WFST の分離

11.3 発音辞書をWFSTに変換する

• 発音辞書

- 単語表記と発音の関係は、単純に列挙すれば良いので、正規表現で記述可能
- 実際は、音素列と単語列との対応の曖昧性を除去するために、各単語の最後にユニークな識別記号を付ける

明日	asu	0.3
明日	ashita	0.7

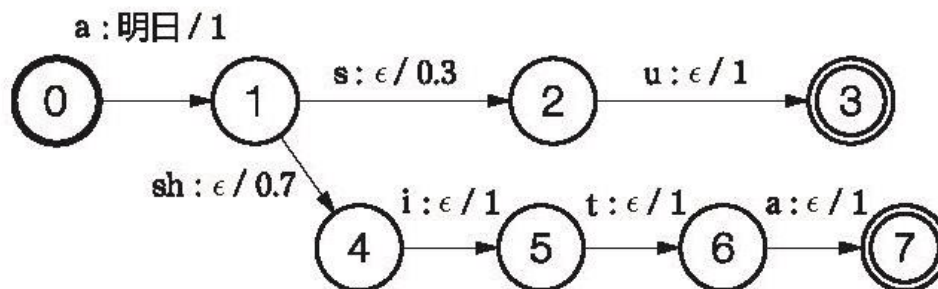


図 11.5

単語辞書から WFST へ変換

11.4 言語モデルをWFSTに変換する

- 言語モデルとして文法を用いた場合
 - 通常は正規言語なので、そのままWFSTで表現可能
- N-グラムの場合
 - N-1個の単語列を状態とし、N-グラム確率を重みとすることでWFSTで表現可能

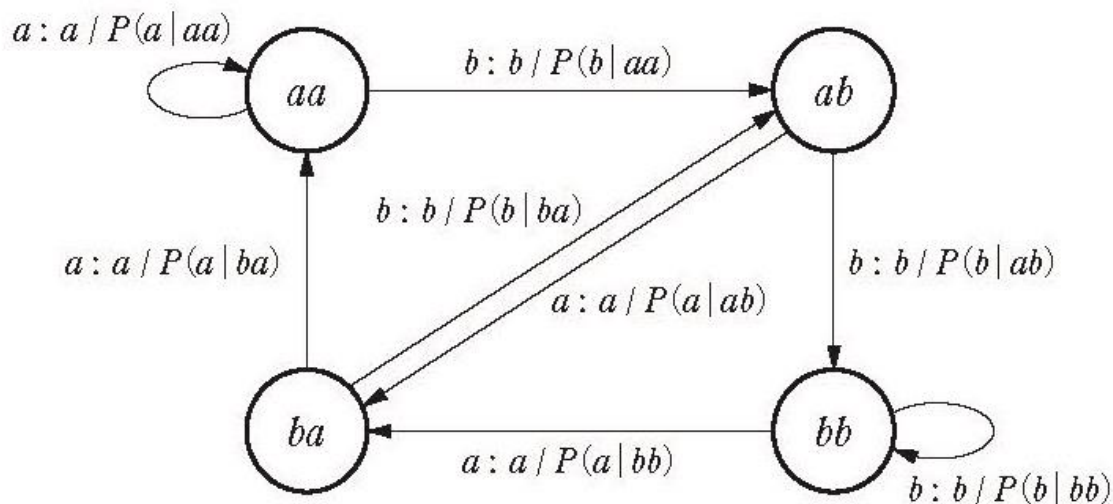


図 11.8 3-グラムを表現した WFST

11.4 言語モデルをWFST に変換する

- バックオフへの対応
 - バックオフ状態を設ける

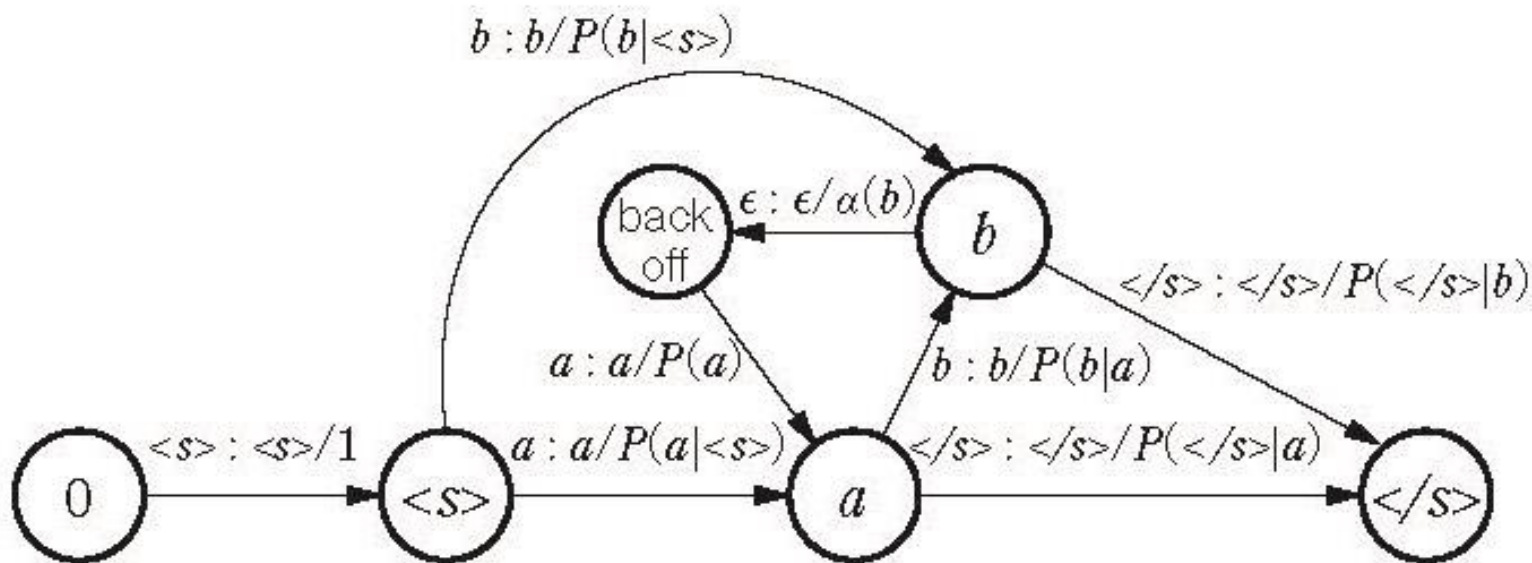


図 11.9 バックオフを利用した 2-グラムから WFST への変換

11.5 WFST の探索

- 合成・最適化後のWFSTをビタビアルゴリズムでビームサーチ



単純なビームサーチを行えばよく、
仮説の動的展開や、
スコアの近似などは不要です。

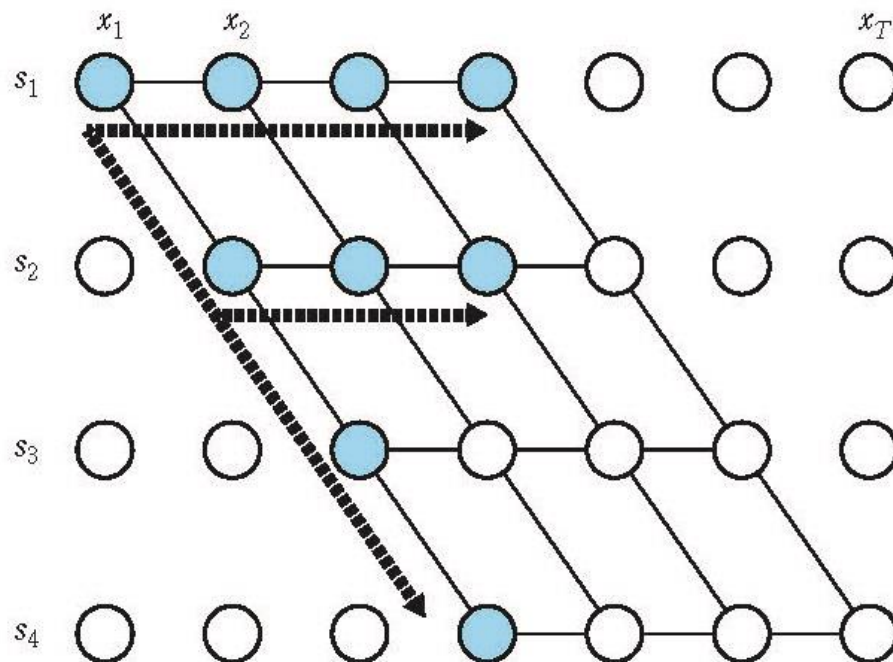


図 11.10

WFST の探索