

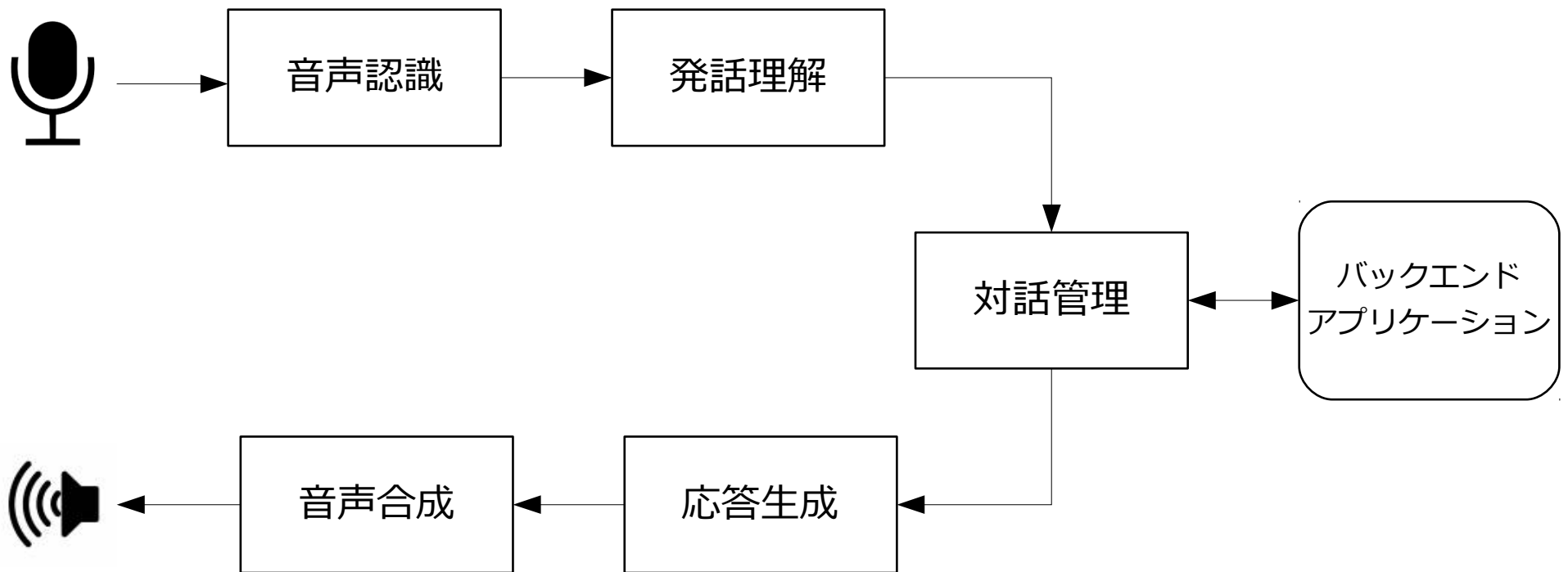
15 章 会話のできるコンピュータを目指して

- 対話管理手法
 - オートマトンによる状態の表現
- 対話管理規則の学習
 - 強化学習
- 音声対話システムの実現
 - MMDDAgent



15.1 音声対話システムの構成

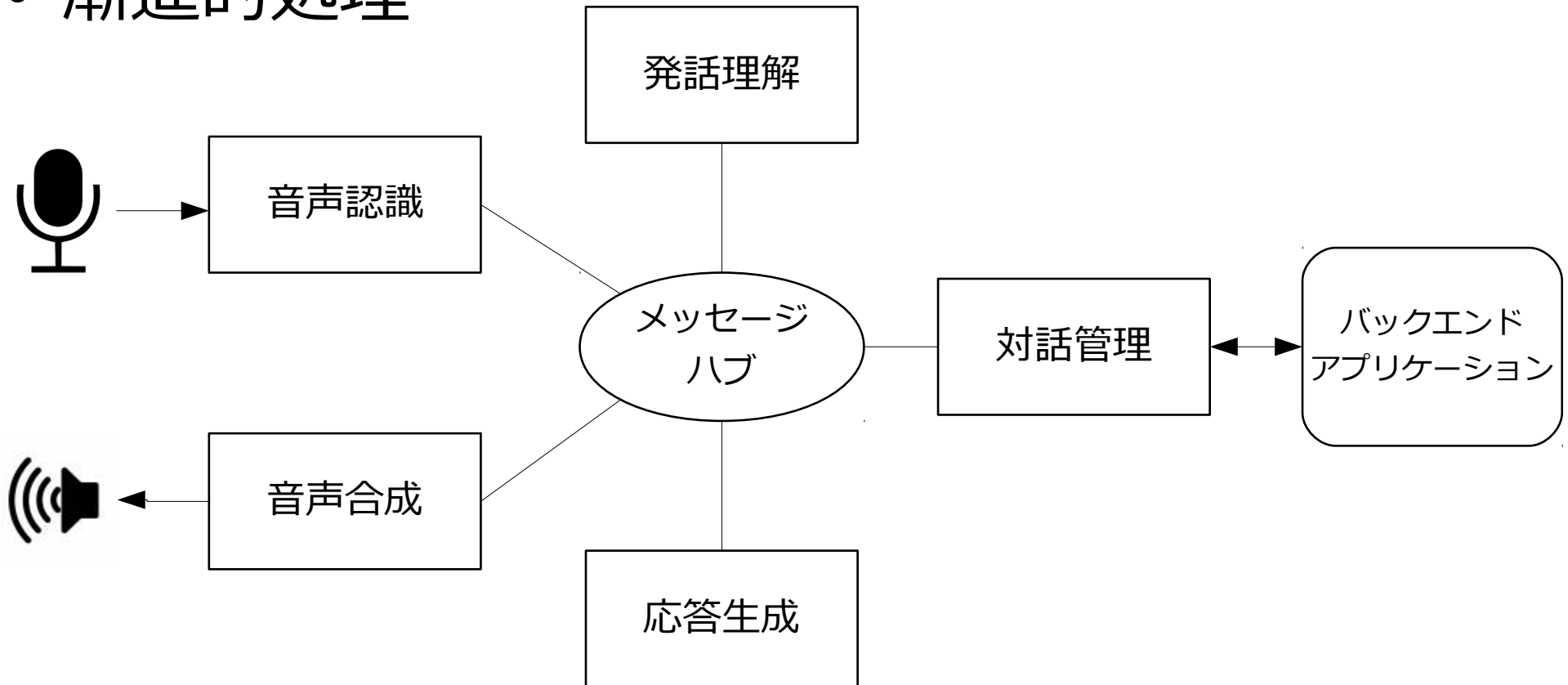
- 逐次的処理



- ひとまとまりの音声入出力を仮定
- モジュールの逐次結合で構成可能

15.1 音声対話システムの構成

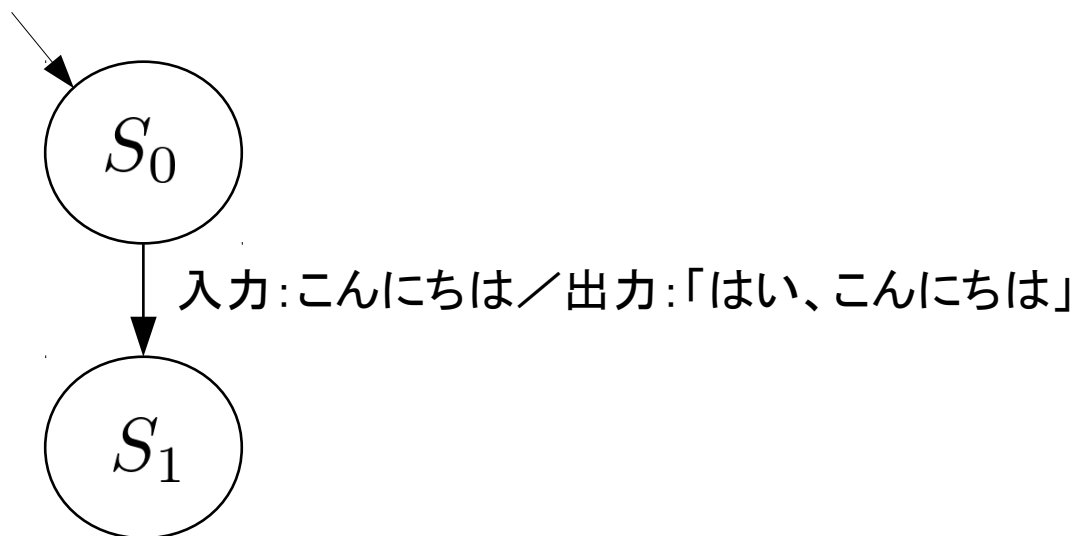
- 漸進的处理



- 小さな単位の音声入出力にも対応可能
- 非同期メッセージの交換で動作

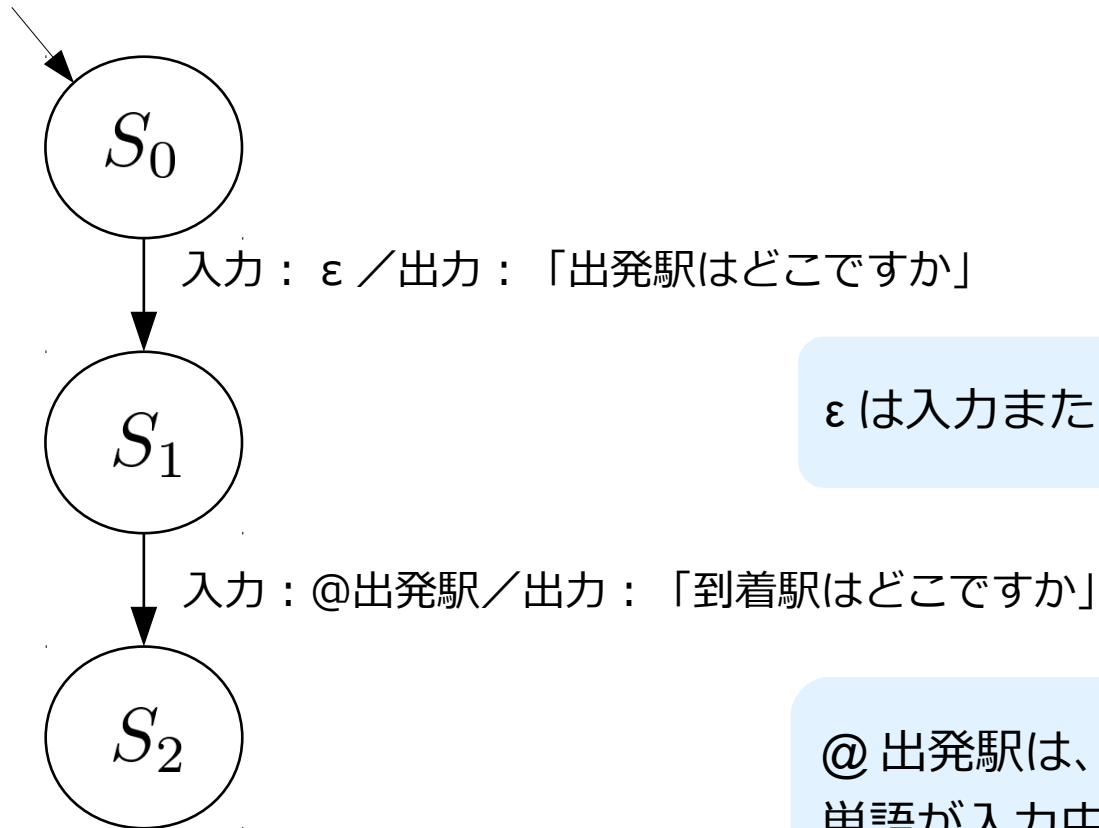
15.2 対話管理の方法

- 有限状態トランスデューサ（FST）による対話のモデル化
 - 入力：ユーザ発話 or イベント
 - 出力：システム発話／行為



15.2 対話管理の方法

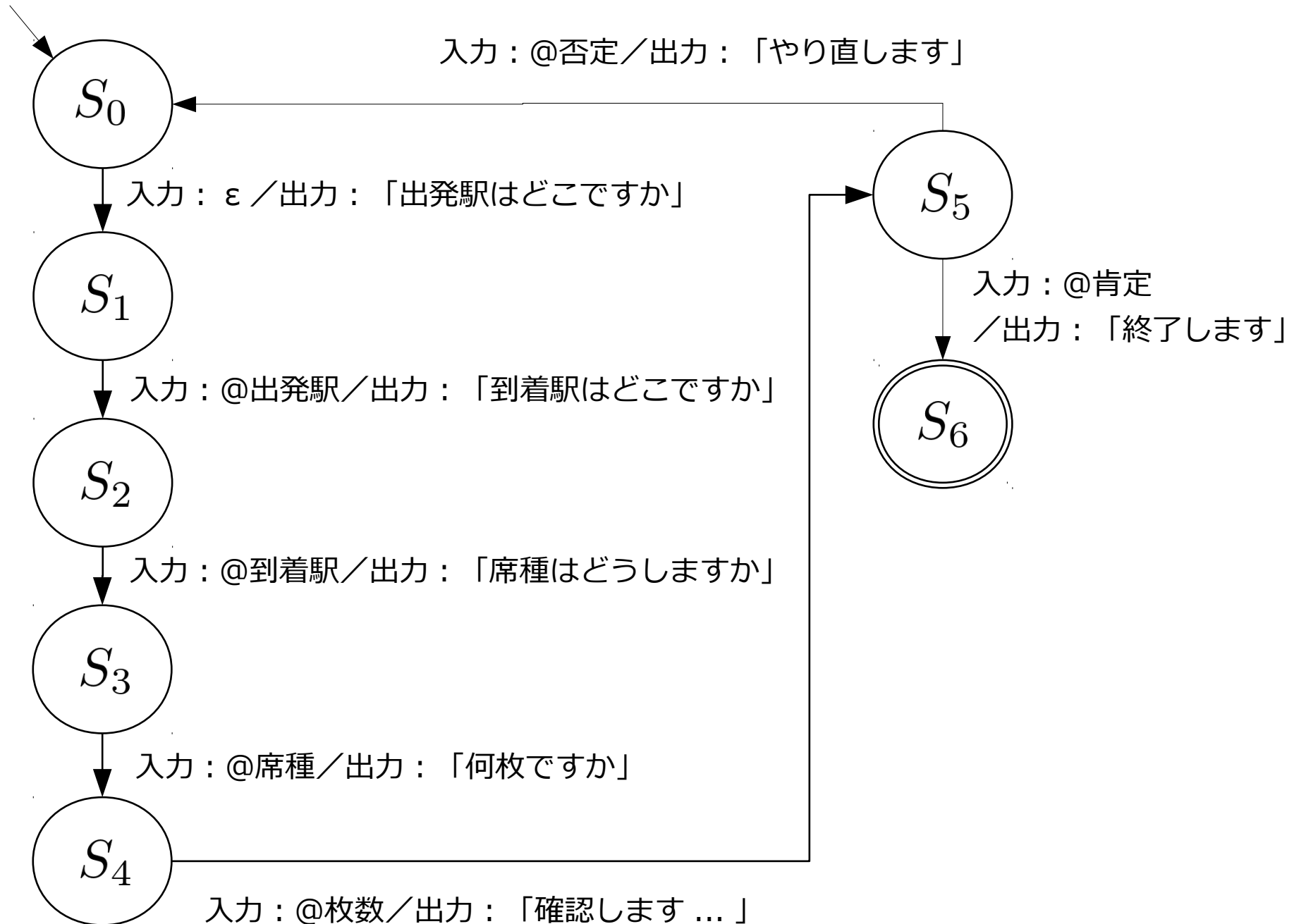
- 特急券販売対話の例



ε は入力または出力がないことを示す

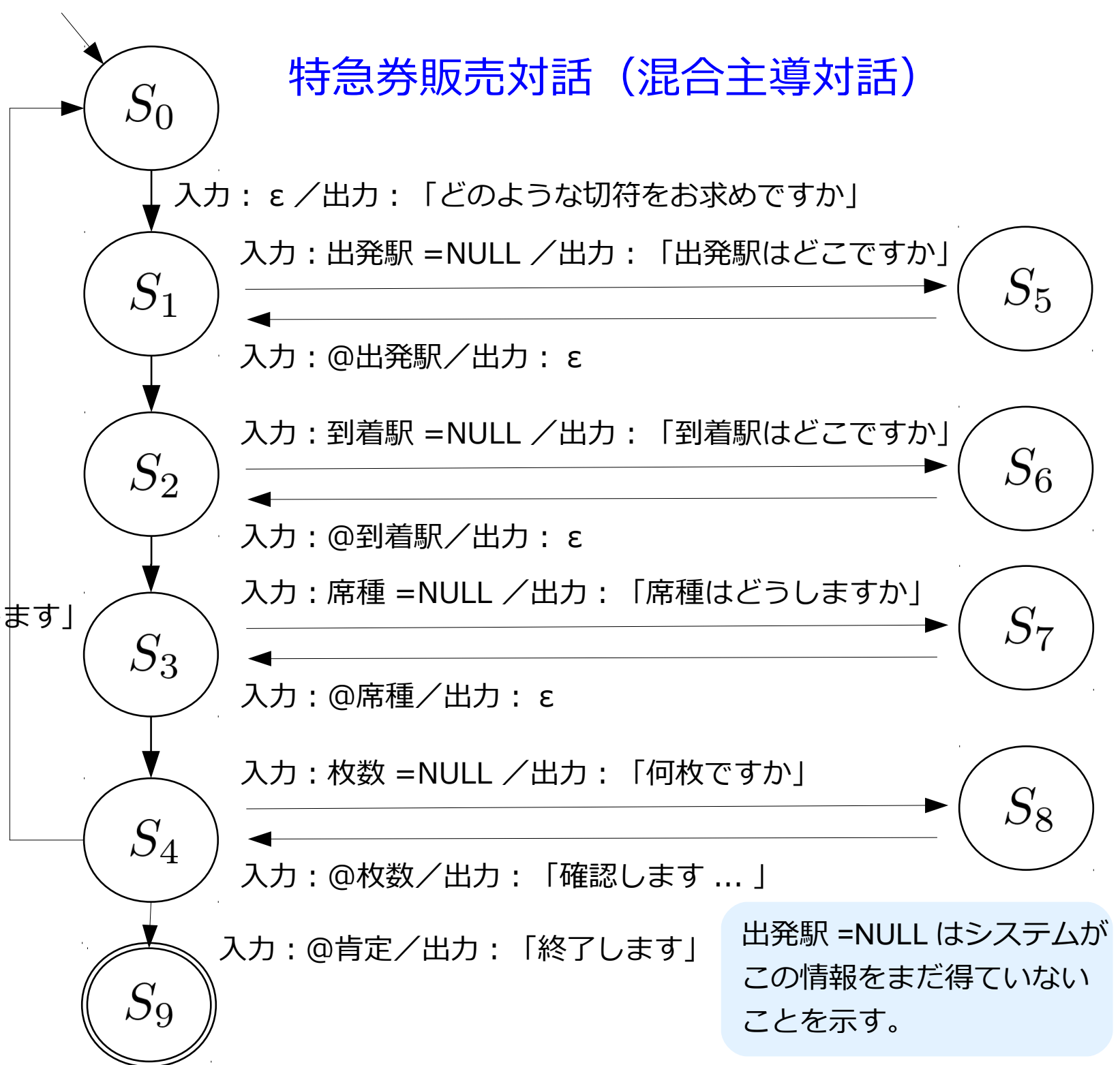
@ 出発駅は、この意味タグが付いた単語が入力中にあることを示す

特急券販売対話（エラーがない場合）



特急券販売対話（混合主導対話）

入力：@否定／
出力：「やり直します」



FST による対話モデルの問題点

- 開発段階

- 主導権・確信度・ユーザモデル等、様々な要素を組み合わせると、実用的なシステムでは状態数は1,000を超える
- 事前にユーザの行動を予測することは難しい

- 保守段階

- 振る舞いの一貫性を保持しながら、部分的な改良を行うのはほとんど不可能

機械学習による対話管理の実現

- 背景 1

- 音声認識・理解における統計モデルの成功

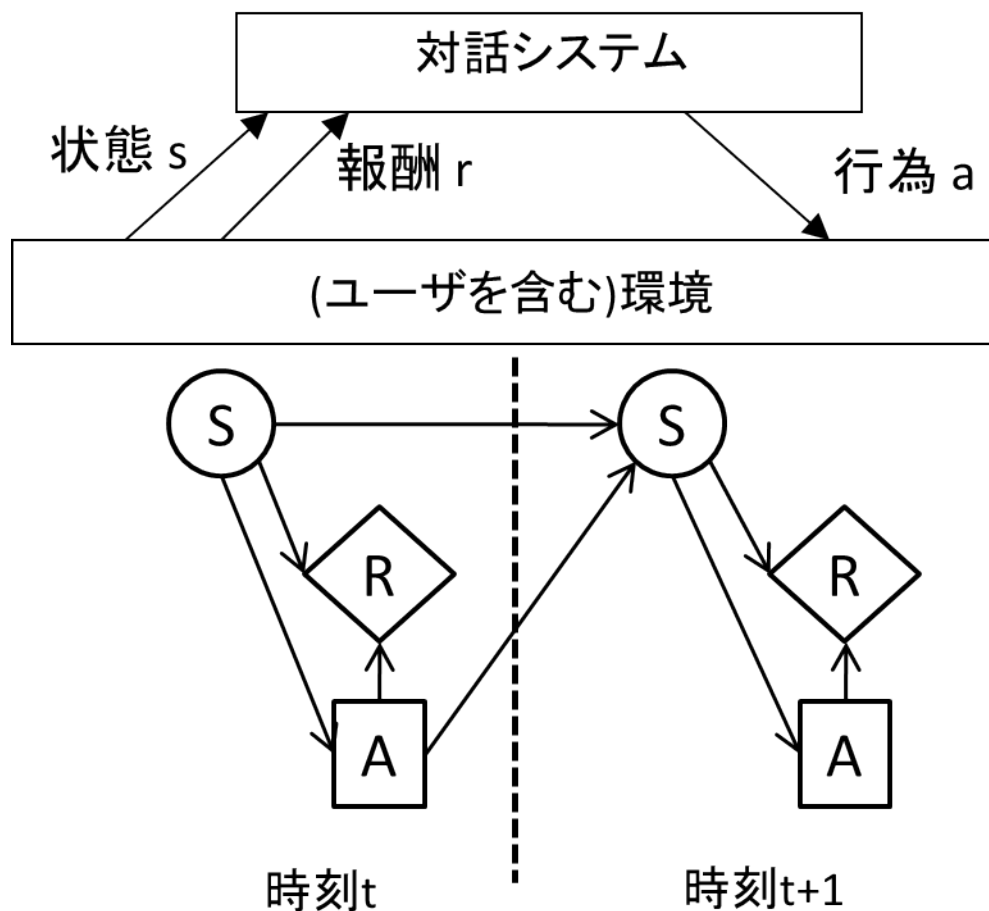
- 背景 2

- ロボットなどを対象とした強化学習の研究の進展
- 対話シミュレーションによる学習データの生成

MDP による対話過程の定式化

- MDP (Markov Decision Process) の定義
 - 時刻 t における状態 $s_t \in S$
 - 時刻 t におけるシステムの行為 $a_t \in A$
 - 報酬 $r_t = r(s_t, a_t)$
 - 状態遷移確率 $p(s_{t+1} | s_t, a_t)$
- 対話管理への適用
 - 対話状況を s_t とし、システム発話やシステム行為を a_t とする
 - 各 s_t で、期待報酬最大となる a_t を学習によって定める

MDP による対話過程の定式化



Q 値の更新アルゴリズム

1. 各 s, a に対する Q の推定値 $\hat{Q}(s, a)$ を 0 に初期化
2. 現在の状態 s を観測
3. 以下を繰り返す

i. 行為 a を選択し、実行する

ii. 報酬 r を受け取る

iii. 新しい状態 s' を観測する

iv. $\hat{Q}(s, a)$ を更新する

$$\hat{Q}(s, a) \leftarrow r + \gamma \max_{a'} \hat{Q}(s', a') \quad \gamma : \text{割引率}$$

v. $s \leftarrow s'$

強化学習による対話管理の実現

- Levin et al. 97
 - ATIS(航空情報サービス) タスク
 - 状態 : greeting, asking(origin, airline, etc.),
retrieving, output, release(origin, airline, etc.),
closing
 - 報酬 (コスト)
$$C = W1 * \text{\#interactions} \\ + W2 * \text{cost of data retrieval} \\ + W3 * \text{cost of data presentation} \\ + W4 * \text{cost of user dissatisfaction}$$

強化学習の例

学習前

S: 座席を変更しますか。

U: えっ？

S: 何時のフライトですか

U: ええと、札幌まで行きたいんだけど

S: 1758 件該当しました。

1 件目は 1 月 1 日 7:00 発 ...

U: ちょっと待って。大阪から札幌までの明日のフライトを調べてください。

S: ありがとうございます。
またのご利用をお待ちしております。



学習後

S: いらっしゃいませ。こちらは
フライト情報システムです。

U: 札幌までのフライト

S: どの空港からですか

U: 大阪

S: 何日にご出発ですか

U: 2 月 2 日

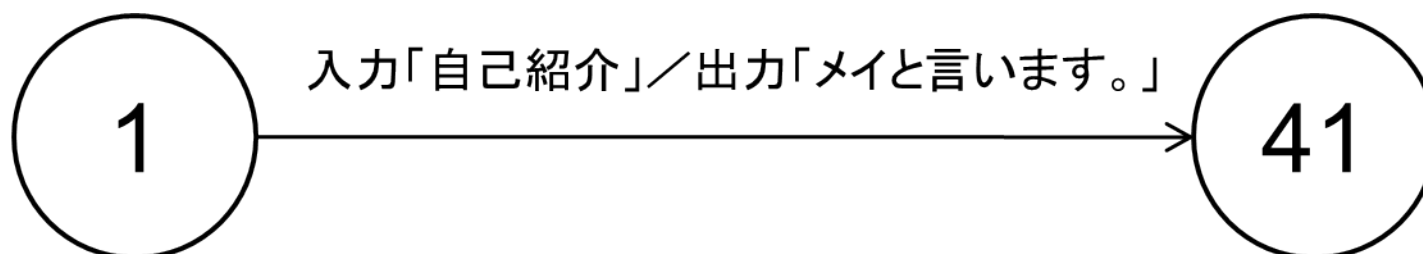
S: 該当するフライトが 4 件あります。 ...

対話システム作成ツール

- MMDDAgent
 - FST による対話記述
 - 入力：イベント
 - 音声認識結果
 - 合成音声出力終了
 - タイマー
 - 出力：コマンド
 - 合成音声、動作、画像の出力
 - web ページの表示



MMDAgent の対話記述 -FST-



1 41 RECOG_EVENT_STOP|自己紹介
1 41 RECOG_EVENT_STOP|あなた,誰
1 41 RECOG_EVENT_STOP|君,誰
41 42 <eps>
42 43 SYNTH_EVENT_STOP|mei
43 2 SYNTH_EVENT_STOP|mei

SYNTH_START|mei|mei_voice_normal|メイと言います。
SYNTH_START|mei|mei_voice_normal|メイと言います。
SYNTH_START|mei|mei_voice_normal|メイと言います。
MOTION_ADD|mei|...¥mei_self_introduction.vmd|PART|ONCE
SYNTH_START|mei|mei_voice_normal|よろしくお願いします。
<eps>

入力イベント
RECOG_EVENT_STOP
SYNTH_EVENT_STOP
TIMER_EVENT_STOP
VALUE_EVENT_EVAL

出力コマンド
SYNTH_START
MOTION_ADD
TIMER_START
VALUE_EVAL
EXECUTE