

補助シナリオを利用したマルチモーダル対話システムの検討*

福井 教順† 桂田 浩一† 入部 百合絵‡ 新田恒雄†

†豊橋技術科学大学 大学院工学研究科

‡豊橋技術科学大学 情報メディア基盤センター

1. はじめに

近年、携帯電話、PDA、タブレット PC 等のユーザ端末多様化に伴い、Web アプリケーションへのマルチモーダル対話 (MMI) 応用が活発に議論されている。筆者らの研究グループでも、web ベース MMI システムの構築を行うと共に、MMI 記述言語 XISL[1]を検討してきた。しかしながら、これまで開発してきたシステムでは、対話アプリケーションを作成する際に XISL の記述量が膨大になってしまうという問題があった。そこで、本報告では XISL の記述量を削減するために、対話内容の修正や、取り消し等を行うための対話を補助シナリオという形で MMI システムに組み込み、共有化する手法を提案する。これにより、修正、取り消し等に関する対話を XISL で記述する必要がなくなり、記述量の削減が可能になる。以下、補助シナリオの詳細とその評価について述べる。

2. MMI システム

2.1 対話システムの概要

MMI システムの構成を図 1 に示す。システムはフロントエンド、入力統合部、出力生成部、対話制御部の 4 つのモジュールから構成される。フロントエンド部は入出力を管理するモジュールである。入力統合部、出力生成部はそれぞれ、入力統合ルール、出力生成ルールの記述に従って入力解釈、出力分化を行う。対話制御部は対話シナリオ、補助シナリオに従って対話進行を行う。対話シナリオはドキュメントサーバに格納され、必要に応じてダウンロードされる。対話シナリオは特定タスクの主要な対話の流れをアプリケーション開発者が記述したものであり、我々が提案している XISL2.0 で記述されている。XISL2.0 は、VoiceXML[2]をベースにマルチモーダル化した MMI シナリオを記述する言語である。XISL2.0 にはユーザの発話を受け付ける一個以上のスロットが用意されており、全てのスロットが埋まると次の対話に遷移する FIA (Form Interpretation Algorithm) に基づいた対話進行方法を採用している。対話進行のイメージを図 2 に示す。

モジュール間のデータの受け渡しには EMMA (Extensible MultiModal Annotation markup language) というデータ構造を用いる。これは W3C[3]で検討されている規格の一つで、様々な入力モダリティの解釈結果を統一的に表現するための言語である。記述例を図 3 に示す。これは「みかん」という音声入力を受け付けた結果の例である。このように、入力結果は XML 形式で表現され、タグ名がその値の「意味」となる。

2.2 音声入力処理の流れ

音声入力の処理には、キーワード抽出と意味解釈を併用した方法を用いる。キーワード抽出部では、入力として受け付けるキーワードと、そのキーワードを格納するスロットの対応関係を予め文法として用意しておき、研究室で開発されたキーワードスポッティング法[5]により

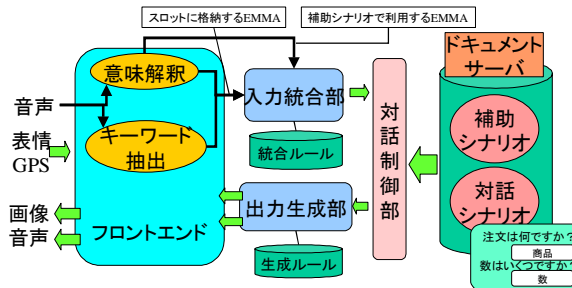


図 1 MMI システムの構成

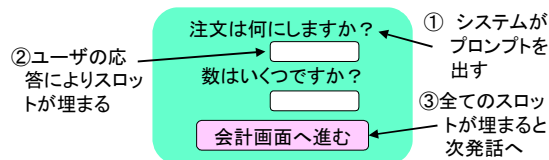


図 2 XISL2.0 を用いた対話進行

```
<emma:interpretation id="ols"emma:mode="speech">
  <商品>
    <果物>みかん</果物>
  </商品>
</emma:interpretation>
```

図 3: EMMA の記述例

スロットに埋める単語を取得する。

一方、意味解釈部は、ディクテーション結果を構文解析、係り受け解析にかけ、さらに概念辞書を参照してフレーム構造を作り上げる。その後取得した概念とスロット名とのマッチングを行い、合致したスロットに入力単語を埋める。意味解釈には文献[4]の手法を用いている。意味解釈とキーワード抽出の結果の競合を考える。意味解釈は、構文解析等の自然言語解析を行い、文全体の整合性も判断している。これより、意味解釈の結果を、キーワード単体よりも信頼性が高いと考え、優先的に用いる。

以上の処理をまとめたものを図 4 に示す。

3. 補助シナリオ

補助シナリオとして用意する対話を選別する際には、文献[6]を参考にした。文献[6]では、スロットの値の変化に着目してスロットフィリングのタイプを分類している。例えば、現在 A という値がスロットに埋まっているとして、ユーザが A を否定する入力を行った場合、スロットの変化を A to notA と表し、スロットの値の取り消しを行う。本システムでは、文献[6]で示されたスロット変化のうち、我々の対話システムで予め対話制御方式として用意されているもの（例えば、空のスロットに値を埋める等）を除いた〔修正、取り消し対話〕、〔スロット入力時の確認〕、〔システムからユーザへの聞き返し〕、〔エラー警告〕の 4 つを補助シナリオとして導入すると共に、文献[6]で示されていない〔シナリオ終了時の警告〕、〔スロットに対する値参照〕の 2 つを補助シナリオとして設定した。

*A MMI system using Side-Scenarios
 N.Fukui†, K.Katsurada†, Y.Iribe‡, T.Nitta†,
 †Graduate School of Engineering, Toyohashi Univ.of Tech
 ‡Information and Media Center, Toyohashi Univ.of Tech

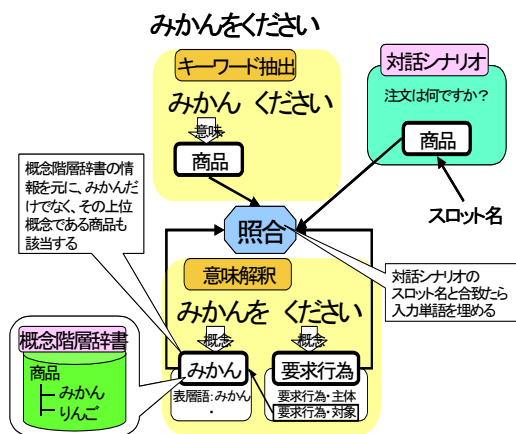


図4 概念照合の詳細

3.1. 補助シナリオの種類

以下、補助シナリオに該当する発話の種類、動作について述べる。

修正、取り消し

スロットに[A]が埋まっている状態で、ユーザが[A]を否定する入力を行った場合には[A]を取り消す。また、「B」という入力を行った場合には[A]を[B]に修正する。

例えば、[みかん]という値がスロットに入っている状態で、「みかんをやめます」という入力があれば、取り消しを実行し、「みかんのかわりにりんごにします」という入力があれば修正を実行する。

シナリオ終了時の警告

スロットが埋まってない状態で、対話を終了するユーザの入力があった場合に、警告を提示する。

スロット入力時の確認

ユーザの入力がスロットに埋まったときに、認識の信頼度に応じて、ユーザへの確認を行う。

例えば、[みかん]がスロットに埋まったときに、「みかんではよろしいですか」と問いかけ、ユーザの返答が「はい」であればスロットの内容を確定し、「いいえ」であれば取り消す。

システムからユーザへの聞き返し

ユーザ入力の認識に失敗した場合に、システムがユーザに発話を促すプロンプトを提示する。

例えば、ユーザの入力が認識できなかった場合に、「もう一度言ってください」とプロンプトを出力する。

スロットに対する値参照

スロットの値を参照する質問をユーザがした場合にその内容を提示する。

例えば、商品というスロットに[みかん]という値を埋め、数ターン対話が経過した後に「さっき入力した商品はなんですか?」と質問された場合に、「みかんです」と応答する。

エラー警告

スロットに値がない状態で、内容の取り消しが発生した場合や、[A]という内容がスロットにある状態で、[B]という内容を取り消した場合に警告プロンプトを出力する。

例えば、スロットが[空の状態]や[みかん]という値が入っている状態で「りんごをやめます」という入力が来た場合、「入力が不正です」と応答する。

今回は、上記の補助シナリオのうち、修正、取り消し

対話をシステムに実装した。

4. 実験

4.1 評価方法

補助シナリオを XISL2.0 を用いて記述し、全シナリオ中の補助シナリオの割合を XISL2.0 の要素/属性数で評価した。対話シナリオは旅行に関するもので、移動検索&決定(move)、観光地検索&決定(sightseeing)、ホテル検索&決定(hotel)、天気検索(weather)、レストラン検索&決定(restaurant)の5つのタスクを用いた。

4.2 結果

実験の結果を表1に示す。「対話シナリオ+補助シナリオ」は対話シナリオと補助シナリオを合わせた全シナリオを表しており、削減率は全シナリオ中における補助シナリオの割合を示している。

実験結果によると、補助シナリオに該当する発話は、平均で対話シナリオ全体の記述量の約28%を占めている。このことより、補助シナリオの導入が記述量削減に効果的であることがわかった。

今回、補助シナリオは修正、取り消し対話のみを対象としている。今後、他の補助シナリオも導入することにより、より一層の削減率の向上が期待できる。

5. おわりに

補助シナリオを利用することで対話シナリオの記述量を大幅に削減できた。これにより、アプリケーション開発者のシナリオ記述の負担軽減が期待できる。本稿で採用した補助シナリオは XISL という言語に依存しない。したがって XISL と同様の言語 (VoiceXML 等) を用いたシステムにおいても、同様の効果を得られると考えられる。

今後、スロット入力の確認対話等、今回導入していない補助シナリオ導入し、評価を行いたい。

表1 記述量の結果

	対話シナリオ +補助シナリオ (要素/属性数)	補助シナリオ (要素/属性数)	削減 率(%)
Hotel	695	163	23.45
Move	803	151	18.80
Restaurant	355	139	39.15
Sightseeing	351	127	36.18
Weather	96	66	68.75
平均	460	129.2	28.09

参考文献

- [1] 青木一峰 他, 「異なる端末環境から利用可能な MMI アプリケーション開発における記述負担の軽減」, 情報処理学会研究報告, 2004-SLP-54, pp.271-276, Dec.2004
- [2] <http://www.w3.org/TR/voicexml20/>
- [3] <http://www.w3.org/>
- [4] 高木朗 他, 「文脈への位置づけを重視した対話意味表現」, 人工知能学会研究報告会資料 SIG-SLUD-A202-09 (2002)
- [5] 新田 他, 「汎用 LVCSR を用いた対話音声の認識」, 情報処理学会研究報告 2002-SLP-41, pp.69-74 (2002)
- [6] 平沢純一 他, 「CTI 向け自由発話対応 音声対話システム RexDialog」, 情報処理学会研究報告, 2003-SLP-47, pp35-40(2003)