

## ホテル予約をタスクとする日韓音声翻訳処理について

鈴木 雅実 井ノ上 直己 谷戸 文廣  
KDD 研究所

### 1 はじめに

筆者らの研究グループでは、日韓音声翻訳システムのプロトタイプを試作中である[1][2]。本稿では、このシステムで仮想的なタスクとして設定している「ホテル予約」に関する処理の内容について述べる。

音声翻訳システムの研究は、これまで ATR [3] や CMU [4] をはじめ、幾つかの先駆的な研究が知られているが、その多くはどちらかというと要素技術の向上や理論の実践に主眼が置かれていた。これらと比較して我々のシステムは、実時間に近い速度で動作する音声翻訳システムを最小限の資源を用いて実装し、可搬性を高めるとともにシステム上での音声翻訳対話実験を通して、利用者インタフェースを含めたシステム要素・統合上の問題を明らかにすることをねらいとしている。以下では、まず我々の試作システムのアウトラインを簡潔に示した後、音声翻訳処理の内容について説明し、問題点の検討や今後の課題等について述べる。

### 2 日韓音声翻訳システムの概要

音声対話翻訳システムとして備えるべき幾つの特徴のうち、リアルタイム処理は大きな要素である。また、システム全体の可搬性や拡張性も重要であり、今後はシステムを実際に動作させた上で種々の音声通訳に関する実験を進めて行く必要がある。そこで、以下のような目標を設計の指針として日韓音声翻訳システムを試作した。

1. 準リアルタイム処理 リアルタイムの2倍以内の処理時間
2. コンパクトなシステム構成 最小限のリソースで構築
3. ホテル予約のタスクに必要十分なカバレッジ 拡張性を考慮した言語知識の記述
4. 高い処理精度 上位5位以内で90%の文認識(理解)率 + 十分コミュニケーション可能な翻訳結果

これらを実現するため、次の図1のようなシステム構成からなる処理を実装した。ハードウェアは、音声認識の前処理(音響分析)用に4個のDSPを使用しているほかは、汎用のワークステーション(Sparc-10)1台ですべての処理を行なっている。SUN OS 4.1.3上で、プログラムはCで記述されている。

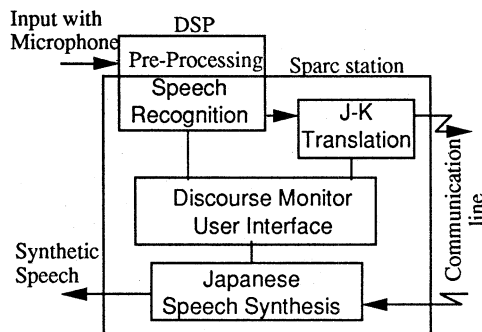


図1: システム構成図

システムの構成要素は以下の通りである。

- 音声認識前処理(音響分析)部 - DSPを使用
- 音声認識処理部(Speech Recognition)
- 発話状況監視部(Discourse Monitor)
- 日韓音声翻訳処理部(J-K Translation)
- 画面表示/発話者インタフェース部
- 日本語音声合成処理部(今回の報告では説明省略)

このシステムのタスクとして想定しているのは「ホテル予約」であり、ホテルの予約係と利用者による一種の協調的な対話の自動通訳である(語彙は約1,000語の規模)。日程や部屋タイプの希望、人数・名前等の確認が行なわれるが、発話者の役割に対して非対称な特徴がある。また、ホテル側話者はタスクの内容に習熟しているのに対して、利用者側話者はそうではないものと仮定することができる。

### 3 音声翻訳処理の内容

以下では、日韓音声翻訳システムを構成する各要素の処理内容と、これまでに行なった予備実験の結果および問題点について説明する。

#### 3.1 発話状況監視部 (Discourse Monitor)

この発話状況監視部の役割は、対話状態を表すパラメタの値とその変化を監視して、音声認識および翻訳処理に有効な情報を与えることである。処理系の一部におけるシミュレーション等を除いて、談話情報や対話に関する知識に基づく処理を音声翻訳システム上に実装した事例はまだ見られない[5][6]。そこで本システムでは、実際に音声認識率と処理効率の向上および、より適切な翻訳候補の選択に利用可能な一種の談話管理を導入した。これを発話状況監視(Discourse Monitor)と呼ぶことにする。その方式は基本的には発話状態の遷移モデルに基づくものであり、関連の研究も多く存在する([7])ほか、発話意図タイプの文間の遷移に見られる統計的な特徴を利用した誤認識候補の削除についても報告されている[8]。<sup>1</sup>

ここでねらいとしているのは、音声認識後のフィルターとして利用するのではなく、文候補の決定時に直接制約を与えるような枠組みである。そのため、音声認識の際に適用する言語知識(文法)と発話状態監視部の間で次のようなリンクを取っている。まず、話者/発話タイプ/話題情報からなる発話状態パラメタを定義する。本システムでは、ある発話状態から別の発話状態への遷移を、遷移可能性マトリクスとして記述した。現在は話者情報(ホテル側/利用者)と発話タイプ(16種)<sup>2</sup>の組み合わせを用いている。一方、CFG形式で記述された認識文法中には、発話タイプに対応する文トップレベルのノードを設けたほか、各規則について、話者/発話タイプ/話題情報の特定が可能なものに対して記号によるヘッダを付与してある。このヘッダ情報を無視すれば、すべての規則を含んだ認識文法が作られ、もし、上の発話状態遷移マトリクスと合わせて利用すれば、若干の前処理により、発話状態に応じた文法サブセットを用意することができる。このことにより、発話の状態から予想される範囲に認識候補を絞り込むことが可能であり、計算コストを削減させる効果も得られる。

#### 3.2 音声認識

音声認識部は日本語連続音声入力に対して、離散型のHMMとCFGによる単語予測を行ない、音韻照合はフレーム同期方式を採用している。また、Nベスト探索のアルゴリズムを用いており、ビーム幅1,000で、音声検出後リアルタイムの約1.5~2.0倍程度で上位5位までの文認識候補をディスプレイ画面上に出力表示する。

タスクの複雑性を文法全体の平均分岐数を尺度として推定すると、総規則数=約800に対し、平均(静的)分岐数=17.7である。8名(男女各4名)の被験者による音声認識の予備評価結果を表1に示す。各被験者は事前に用意された発声リスト中の各52文(ホテル予約2会話分に相当)をシステムの前でマイクを用いて入力した。レベル調整のためのテスト発声以外は、いずれも初めてのトライアルである。

現時点では、実際のシステム上での対話レベルの実験は未了のため、前述の談話管理を使用した場合の評価結果はまだ得られていないが、話者ロール情報だけ(限定された文法サブセットを使用)でも効果があることが分かる。しかし、「ホテル予約」のタスクにおいて重要な、日付けや利用者名の応答/確認については、発話状態監視による談話情報の利用が期待できないため、問題点として残ると思われる。今回の実験でも、これらの要素を含

<sup>1</sup>また、この種の統計的なデータを得るために、発話意図等の談話的なラベルを付与した対話コーパスの利用が提案されている[9]。

<sup>2</sup>認識文法中にも陽に記述可能なように表層表現に対応しており、その内容はタスク依存性は少なく一般性を持たせてある(文献[9]等を参考に再定義)。またこのことは、本システムと対になる韓国音声翻訳システムからの出力が、談話情報を含まない日本語テキストだけと仮定した場合でも、その表層表現から対話相手側の発話状態を推定できるようにすることも考慮したものである。

む文については認識率が低下している。この解決には、例えばカレンダーの画面表示(マウスクリック/タッチパネル)や入力フォームへのキータイプ等のマルチモーダル・インタフェースを検討する必要がある。

表 1: 音声認識予備評価実験結果

利用者側の会話文 (8名の平均)			ホテル側の会話文 (8名の平均)		
発話者情報	1位正解率	5位以内の文理解率	発話者情報	1位正解率	5位以内の文理解率
無し	68%	75%	無し	69%	75%
有り	86%	90%	有り	77%	82%

注) 理解率は同じ意味の認識結果を含む。 1文平均文節数: 利用者側=2.1, ホテル側=3.0

### 3.3 日韓翻訳処理

日韓翻訳処理部は、日本語と韓国語が語順等の文法的な特性に類似点を持つことから、意味関係を含む接続制御を主とする語レベルの逐次翻訳方式を採用しており(金 [10]), 十分な頑健性を持つ。ただし、形態素レベルでの変換ばかりでなく、話し言葉への柔軟な対応を図るため、コロケーション(連語)を重視した処理単位の設定を行なって、処理効率を高めている [11]。さらに、翻訳辞書をシステム全体のマスター辞書と位置付け、音声認識文法における認識単位(語)も翻訳の処理単位と同一とし、両者を連携させて記述およびコンパイル(実行形式を作成)することにより、言語知識全体の作成・管理効率をより向上させている。なお付録に翻訳例を記す。

これまでに行なった予備実験の範囲内では、次のような結果が得られている。すなわち、ホテル予約のタスク内で設定した6会話(各10~30文程度を含む)に対する、文単位の日韓翻訳処理内容の妥当性について、韓国語に通じた日本人による主観評価を実施した。内容的正確性・文法的整合性・自然性の3種類の尺度に基づいて3段階評価を行なった結果では、内容的正確性と文法的整合性についてはほぼ問題がなく、自然性の点で約15%程度の翻訳結果に多少の問題があることが判明した。また、今回の予備実験の範囲の周辺でも種々の翻訳上の問題点が明らかとなりつつあるので、以下にその一部を紹介する。

日韓翻訳全般について言うと、前述したような言語間の類似性から、他のヨーロッパ系言語への翻訳と比較して問題は少ない。ただし、「ホテル予約」のような話し言葉の翻訳となると、特有の問題も発生する。試みに、1) 特に分野や文のスタイル等にあまり関係しない翻訳上の問題、2) タスクや文脈に依存する度合いの大きい問題、3) 文化的なギャップ等に起因する言語運用上の問題に便宜上分類した上で、顕著な例について述べる。

#### 1. 一般的な問題: 助詞の訳し分け etc.

日本語の「の」は、これにほぼ対応する韓国語の助詞“eui”があるが、訳出しない方が自然な場合が多い。さらに「が」/「は」と“i(ga)”/“eun(neun)”との対応もバラレルではない。現状ではヒューリスティックな規則を使用しているが、この種の問題には対訳用例ベースからの類似例の検索が有効かもしれない。

#### 2. 文脈に依存する問題: 含意の曖昧性 etc.

典型的な例は、「～します」という日本語表現の含意が、現在/未来の事象を指す場合、話者の意志を表す場合でともに使用されるのに対して、韓国語ではそれぞれ異なる表現(“~hamnida”/“~hagesseumnida”)が対応する現象である。このほかにも、「お願いします」に対応する表現“putakhamnida”が使用できない(不自然な)場合等がある。現在のところはデフォルト処理で対処している。

#### 3. 文化的な差異に起因する問題: 両言語の運用上の差 etc.

例えば会話冒頭の「ありがとうございます。こちらは …」のような挨拶表現をはじめ、会話中でも「すみません」等の感動詞(相当表現)が、日本語の電話会話では多用されるが、この直訳は不自然な場合が多いことが指摘されている。上記3項目ともに、発話状態に関する情報を高度化すること等により翻訳結果の段階的な向上が期待できるが、音声通訳実験を継続しながら優先課題を探る必要がある。

## 4 今後の課題とまとめ

ホテル予約をタスクとする日韓音声翻訳システムについて、本稿ではその処理内容と予備実験結果、問題点等を報告した。目標とした処理のレベルにはほぼ到達しつつあり、現在はシステム全体の総合評価を進めようとしている。これに関して、評価方法の確立が問題となるが、個々の文の認識率等ばかりでなく、タスクのゴール達成までの発話数等、複数の尺度を組み合わせる必要があると思われる [12]。

また、システムの当面の改良ポイントとしては、音声認識の音韻モデルの連続型(音素環境依存)HMM への変更や、発話状態パラメタ等の談話知識の高度化等が考えられる。さらに自由な発声形式への対応は将来の課題であるが、その前に本システムで実現した処理のレベルによる音声通訳実験を通して、システム利用者への教示方法や適用可能なタスクの探求等のアセスメントを実行するとともに、要素技術への還元を図りたい。

**謝辞** 本研究の機会を与えて下さった KDD 研究所・浦野所長、村上次長に感謝する。また、適切な助言を頂いた音声言語処理グループの山本 前リーダーをはじめ同グループ関連の諸氏に感謝する。さらに、翻訳処理に関して貴重な情報を頂いた慶応義塾大学の金 泰錫 氏(現在 東義大 学校)と金 政仁 氏、翻訳辞書の作成でお世話になっている河井 章子さんとソフトウェア開発に貢献された桶谷氏 (ISS) に厚くお礼申し上げる。

## 参考文献

- [1] 鈴木, 井ノ上, 野垣内: “日韓音声翻訳システムの設計”, 情報処理学会第 49 回全国大会, (1994).
- [2] 鈴木, 井ノ上, 谷戸: “日韓音声翻訳システムの試作”, 人工知能学会, SIG-SLUD-9403-4, pp.21-26, (1995).
- [3] Morimoto, T., et al.: “ATR’s Speech Translation System: ASURA”, *EUROSPEECH ’93*, (1993).
- [4] Woszczyana, M., et al.: “Recent Advances in JANUS: A Speech Translation System”, *EUROSPEECH ’93*, (1993).
- [5] Yamaoka, T. and Iida, H.: “Dialogue Interpretation Model and its Application to Next Utterance Prediction for Spoken Language Processing”, *EUROSPEECH’91* (1991).
- [6] Suzuki, M., Kikui, G., et. al.: “Towards Discourse-Sensitive Dialogue Interpretation – An Empirical Study in a Speech Translation Project –”, *Proc. of PACLING-93*, (1993).
- [7] 巖寺, 石崎, 森元: “対話のインタラクション構造と話題の認識”, 情報処理学会第 49 回全国大会, 3-159, (1994).
- [8] 永田 昌明: “統計的な対話モデルの試みとその音声認識への応用”, 人工知能学会 言語・音声理解と対話処理研究会資料 SIG-SLUD-9202, (1992).
- [9] 永田 昌明, 鈴木 雅実: “日英対話コーパスへの発話行為タイプ付与の試みとその統計的対話モデルへの利用”, 人工知能学会 言語・音声理解と対話処理研究会資料 SIG-SLUD-9310, (1993).
- [10] 金 泰錫, 浦 昭二: “意味接続関係に基づく翻訳テーブルを用いた日韓機械翻訳システム”, 情報処理学会第 43 回全国大会, 3-203, (1991).
- [11] 北 研二, 他: “仕事量基準を用いたコーパスからの定型表現の自動抽出”, 情報処理学会論文誌, Vol.34 No.9, pp.1937-1943, (1993).
- [12] 浦谷 則好, 鷹見 淳一: “音声認識の良否の判定基準”, 情報処理学会第 49 回全国大会, 2-57, (1994).

**付録 (翻訳例)** 部屋を予約したいのですが。 → 방을 예약하고 싶은데요.

いつお泊まりのご予定でしょうか。 → 언제 숙박하실 예정입니까?

その日はお取りできます。 → 그날에는 예약하실 수 있습니다.

ツインを一部屋お願いします。 → 트윈을 하나 부탁드립니다.