

タスク環境を考慮した日韓自動通訳システムのインタフェース改良

鈴木 雅実 井ノ上 直己 谷戸 文廣

KDD 研究所

〒 356 埼玉県上福岡市大原 2-1-15

E-mail: {msuzuki, inoue, yato}@kddnews.nes.lab.kdd.co.jp

あらまし 本稿では、筆者らが試作した日韓自動通訳システムについて行なった、ホテル予約のタスク環境を考慮した発話者向けインタフェースの改良内容について報告する。自動通訳システムに必要な特性として、可能な限りリアルタイムに近い高速で精度の高い認識・翻訳処理が要求されるが、利用者の操作性等のヒューマンファクターを考慮したシステム設計も重要である。今回実装したシステムを用いて行なった通訳実験とその評価結果は、限定されたタスクについては対話環境下で高い成功率を得られることを示しているが、実用性を追求するためには種々の検討課題が残されており、それらの考察を行なった。

キーワード 自動通訳システム, 音声翻訳, 対話, ユーザインタフェース, タスク環境

A Design of User Interface for an Automatic Interpretation System considering the Task Environment

Masami SUZUKI Naomi INOUE Fumihiro YATO

KDD Research and Development Laboratories

2-1-15, Ohara, Kamifukuoka-shi, Saitama 356 JAPAN

E-mail: {msuzuki, inoue, yato}@kddnews.nes.lab.kdd.co.jp

Abstract This article describes about the improvement of user interface for our Japanese-Korean automatic interpretation system, considering the task environment: "Hotel Reservation". A user-friendly interface with a proper system integration is also required for an interpretation system, as well as a high-speed and accurate speech recognition and translation. It should be designed based on human factors like as handling features. Experiments on automatic interpretation and their results using our recent version of the system showed that our system has a feasibility for a limited task domain. However, we have to continue further investigation to realize a practical system as we will discuss.

key words automatic interpretation system, speech translation, dialogue, user interface, task environment

1 はじめに

筆者らは、1993年の秋から開発に着手した日韓自動通訳システムの試作結果をふまえ、このほど韓国の共同研究機関である韓国通信(KT)および韓国電子通信研究所(ETRI)との間で「日韓自動通訳共同実験」を行ない、ホテル予約をタスクとした場合の自動通訳技術の現状を示し、到達レベルと今後の課題を考察した。

利用者にとって使い易いシステムを設計することは基本性能の向上に劣らず重要な点であり、タスクの成功率にも影響する要因と考えられる。今回の試作の最終段階では、自動通訳システムの利用可能性を探る観点から、特にシステムのユーザインタフェースについて種々の改良を実施した。本稿では、ユーザインタフェースの改良を中心とした、日韓自動通訳システムの全体的な試作内容について述べるとともに、上記の共同実験とその後に行なわれたシステムの評価実験結果を通して、現状分析と問題点の検討を行なう。

2 システムの基本設計

日韓自動通訳システムの試作内容に関しては、これまで幾つかの文献([1],[2],[3])で報告したので、ここではシステム設計上の留意点について概略を述べる。

自動通訳システムとして備えるべき特徴として、リアルタイム処理は大きな要素である。また、種々の対話環境でシステムを実際に動作させながら自動通訳に関する実験を行なえることが、今後研究を進める上で重要と考えた。そこで、以下のような項目を設計の指針として日韓自動通訳システムを試作した結果、ほぼ所期の目標レベルに到達した。

- 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 最小限のハードウェアリソースで構築
- 3. 頑健性・拡張性を考慮した言語知識の記述
ホテル予約のタスクに必要な十分なカバレッジ
- 4. 高い処理精度 上位5位以内で90%の文認識(理解)率 および 十分コミュニケーション可能な翻訳結果

システム構成を図1に示す。ハードウェア資源として、汎用のワークステーション(Sparc-20)1台

と音声認識の前処理(音響分析)用の4個のDSPを用いて試作した[2]。

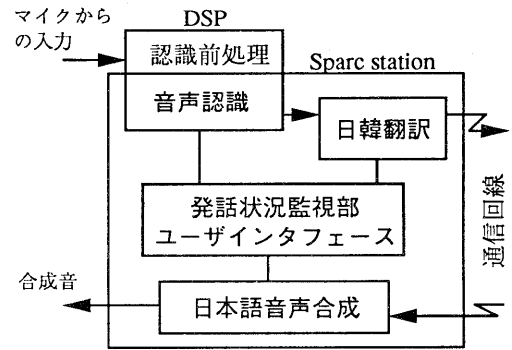


図1: システム構成図

本システムにおける音声翻訳処理上の特徴として、発話状況監視部(Discourse Monitor)を用いた対話音声認識処理がある。これは、発話状態(発話者/発話のタイプ等のパラメタの組合せとして定義)から遷移可能な次の発話状態を知識として記述しておくことにより、発話状態との関連性を示すヘッダー情報を付与した音声認識文法から、対応するサブセットのみを取り出して適用する手法である。この方法により、認識文法全体の体系的な記述の枠組を維持しつつ、計算コストの削減と対話音声の認識率の向上を図ることが可能となった。

3 タスク環境

自動通訳の対象となる「ホテル予約」のタスクは、部屋予約を希望する利用者(Client)とホテル予約係(Hotel)との間で交わされる、一種の協調的な目的指向の対話を前提としている。このことから、利用者側話者がホテル側の質問に適宜答える、ホテル主導型の対話を処理の対象として想定することができる。

また、発話スタイルは連続的な文発声を基本とするが、現段階ではシステムの基本性能に影響を与えたと考えられる冗長語や言い淀み等は扱わないこととした。将来的には、これも含めてより自然で柔軟な対話を扱えるようにする必要がある。

対話の過程において宿泊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。語彙は両発話者分を合わせて約1,000語の規模である。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できる。

4 ユーザインタフェースの改良

自動通訳システムとして備えるべき種々の要件のうち、利用者にとっての使い易さは重要な項目の一つである。筆者らが開発した日韓自動通訳システムでは、当初からワークステーションの画面を用いて、音声入力波形・音声認識結果・翻訳結果の表示を行なうとともに、発話開始の指示・複数の音声認識候補からの正解候補選択等を利用者の操作に委ねる、一種のマルチモーダル対話システムとして設計した。しかし、システムの利用者(以下ユーザ)の視点から、次のような問題点があることが判明したので、それらに対する改良を試みた。

1. 発話開始のタイミング

音声認識システム一般に言えることであるが、発話の開始時に何らかの操作を伴うシステムはユーザに負担を与えると考えられる。一方、常時システムを音声入力可能な状態にしておくことも、処理の負荷や雑音による誤動作の可能性が高くなることが予想される。そこで、日韓双方の対話の進行(すなわち発話権の移譲により交互に発声する対話環境)を考慮して、相手側からのターン信号到着後の適切なタイミングで音声入力可能表示を行なうことにより、操作無しで発声を開始できるように変更を加えた。ただし、最初の発声の認識が失敗した場合や、認識結果選択後に続けて次の発話を行なう場合は、入力可能状態に移行させるための指示が必要である。さらに、発話途中で入力をキャンセル(リセット)できる機能を追加した。

2. 発話権の交替

前記の通り、発話権の交替情報を相互にやりとりする方式を用いているが、その時点で発話権があるサイトは発話終了後に相手側の発話を促すターン(発話権交替)情報を送出する必要がある。一つの方法として考えられるのは、発話者が発声終了後にターンの送出を行なうことである。本システムでも当初はこの方法を採用していたが、発話者がターン送出を忘れてたり操作自体が煩雑さを増すことになる。そこで、認識候補選択後3秒以内に続いて発声開始の指示が無い場合、自動的にターンを送出するように変更した。

3. 音声認識結果の確認

音声対話システムの現状の技術レベルとして、各発話の音声認識候補のうち第1位のものが常に正解として得られないことから、ある程度以上のタスク規模では発声内容の確認は避けられないと思われる。そこで、複数候補を尤度順に表示して発話者による確認・選択を行なうためのインタフェースを用意した。ただし、システム性能の評価結果やユーザの操作性を考慮して、当初5位まで表示していた候補数を3位までに削減した。なお、光学式のタッチパネルも実装して、マウスクリックと同様な指示選択操作も可能となっている。

4. タスク依存型の対話内容表示

前述したホテル予約における発話者(利用者/ホテル予約係)の非対称性を考慮して、対話開始時点でロールを選択する機能を設け、これと連動して発話状況監視による対話音声認識が実行される枠組を採用した。当初は、選択された認識候補(日本語文)を対話履歴としてモニタスクリーン上に記録表示していたが、タスクの性格を考慮して、予約内容のキーワードのみ(日付/部屋タイプ・グレード・名前)を予め用意した表示欄に、決定されたものから表示するインタフェースに改良した。また、日付と連動するカレンダーをこの表示欄の下に配置した。

以上の、ユーザによる直接操作に関わる改良の結果、対話中の大部分の発話については認識結果の確認時の選択操作のみがユーザに委ねられることとなった。さらに、対話により要求・確認された内容が、視覚的により分かり易く提示されることになった。図2に改良したシステム表示画面を示す。

図2. システム表示画面

日韓自動通訳システム KDD Japanese-Korean Automatic Interpretation System					
日本側	利用者	電話	発聲	終了	韓国側対話者の 履歴表示部
韓国側	ホテル予約係	入力音声(日本語)波形の表示部			
音声認識候補	Two 領域を兼ねる				(予約内容表示) ↓
	上位3位までの認識候補(文)を表示 マウスによる正解候補選択				予約者名: <input type="text"/>
	日本語: 上の正解候補を韓国への入力として再表示				発話日: <input type="text"/> 月 <input type="text"/> 日 から <input type="text"/> 月 <input type="text"/> 日まで
日韓翻訳	韓国語: 韓国語結果をハン글表示				部屋のタイプ: <input type="text"/> <input type="checkbox"/> 室 グレード: <input type="text"/>
韓日翻訳	日本語: 韓国側システムの翻訳結果を表示				カレンダーの表示
	韓国語: 韓国側システムの音声認識結果をハン글表示				

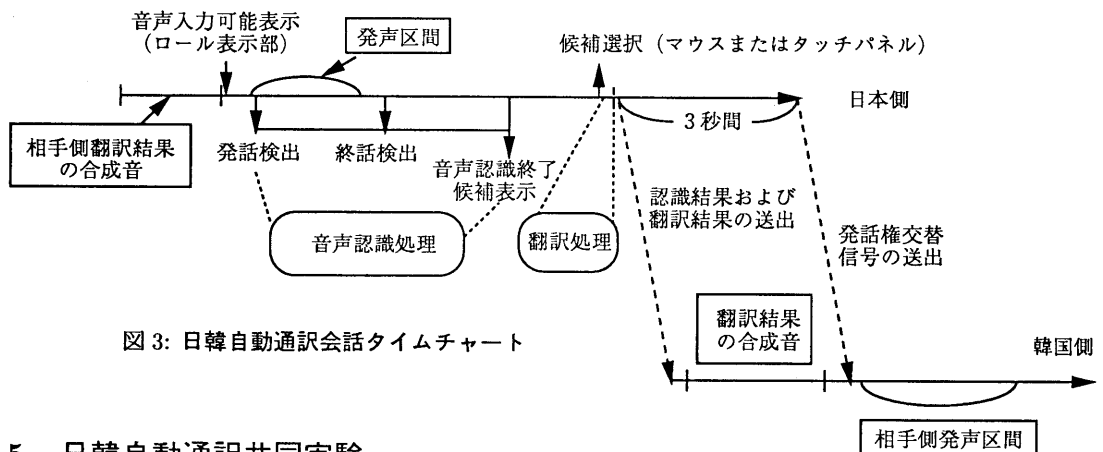


図 3: 日韓自動通訳会話タイムチャート

5 日韓自動通訳共同実験

日韓共同実験は、東京の KDD 大手町ビル内の TV 会議室とソウルの KT リサーチセンター内の会議室を結んで 95 年 5 月に実施されたものである。KDD 側は、日韓自動通訳システムにより、日本語音声認識と日韓言語翻訳処理を行なって、認識結果および翻訳結果をテキスト情報として韓国側に送信し、翻訳結果のハングル文字列は韓国側で音声合成出力された。韓国語から日本語への方も、韓国側のシステムを用いて対称に同様な処理がなされた。

このような構成の下で、システム間は国際専用回線を使用したモデム接続 (2400bps) を行ない、表 1 に示すプロトコルに従ってテキスト情報のみを送受した。なお、この接続における通訳対話の進行を図 3 に示す。

表 1. システム間通信プロトコル

パケット形式	{ 識別子 }return { 認識結果テキスト }return { 翻訳結果テキスト }return
識別子	0 結果の転送 1 発話権の交替 2 対話の終了
認識結果テキスト	<日本語文字列> S-JIS
翻訳結果テキスト	<ハングル文字列> KSC

注) 識別子 1,2 の場合テキストは null

また、ISDN 回線を使用した TV 会議システムを用いて、対話相手の顔画像と入力音声、および翻訳結果テキストからの合成音声を相互に送受信した。このうち相手顔画像については、TV 会議システムからのビデオ信号をシステムのモニタスクリーン上に表示した (図 2 のシステム表示画面を参照)。

6 システム評価実験

本章では、上記の日韓自動通訳共同実験後に行なった、日韓自動通訳システムの評価実験結果について述べる。

6.1 予備評価実験の結果

音声認識および翻訳の各要素レベルの基本性能に関する予備実験 (実施済み) の結果は以下の通りであった ([2] [3])。

・音声認識部の予備評価実験結果

非対話環境における 8 名の男女被験者の読み上げ会話文に対し、認識文法全体を使用した場合の正解率が約 75%、話者のロールを考慮した文法サブセットを用いた場合、利用者側の例文で約 90%、ホテル側の例文で約 82% の正解率を得た。

・日韓翻訳部の予備評価実験結果

ホテル予約に関する典型的な 2 会話文のサンプル文 (45 文) をシステムで翻訳した結果を対象に、了解性 (Intelligibility)、文法性 (Grammaticality)、自然性 (Naturalness) の 3 つの尺度で 3 段階の主観評価を行なった。評価結果は、了解性および文法性についてはほとんど問題がなかったが、自然性のみ約 15% の文が「やや不自然」と判断された。

6.2 対話環境におけるシステム評価実験

以下に述べる評価実験は、特にシステムの利用可能性を中心とした評価を目的としたものである。このため、韓国との共同実験で行なったような国際回線を使用した実環境ではなく、疑似的な双方向の対話環境を設定して実験を行なった。また、被験者

としては本システムについての知識を持たない一般の男女16名を選んで実験を行なった。

この実験では、双方向の対話環境を提供するため、「韓日自動通訳疑似システム」を使用した。この疑似システムでは実際の韓国語音声認識は行なわないが、利用者またはホテル側の想定会話メニューを開いて日本語文を選択することにより、それに対応する韓国語文とのペアが日韓自動通訳システム側に転送される。また、転送前に韓国語文を合成音声出力することにより、双方向通訳会話としての臨場感を与えている。なお、これに用いた韓国語合成ソフトウェアはETRIで開発されたものである。以下の実験では、疑似システム・オペレータが実験条件に従って操作を行なった。実験方法・実験結果は右表2、表3に記した。さらに、被験者に対して実験終了後アンケート調査を行なったので、その結果を記す(括弧内は人数)。

1. システムの処理速度についての印象
速い(1), 普通(11), やや遅い(4), 遅い(0)
2. マウスとタッチパネルではどちらが快適か?
マウス(6), タッチパネル(8), その他(2)
3. この種のシステムに必要な改善要素(意見の集約)
問投詞や明瞭でない表現の受容等の柔軟性(10)
語彙の大きさ等の処理範囲の拡大(5)
高速性(5), 機能的な手軽さ(4)

実験結果から得られた所見は次の通りである。

- (1) 各被験者とも、日韓自動通訳システムに対しては初めての試行であったにもかかわらず、相当に高いレベルの成功率を記録したが、このことは
 - ・事前説明が十分になされていた
 - ・タスクが比較的容易であった(ホテル側との交渉的な要素は含まない)
 ことによるものが大きいと思われる。

- (2) 今回設定したホテル予約のタスクの場合、利用者側の会話1発話の所要時間は、1回の発声で成功の場合約6秒間であり、その内訳は既ね
 - ・平均発声時間として2秒
 - ・認識処理の遅れ時間が2秒
 - ・発声者の反応時間が2秒(発話可能表示から発声までと、認識結果表示から候補選択まで)であった(図3のタイムチャートも参照のこと)。また、わずかではあるが、反応時間に被験者の試行回数に従った学習効果が現れているようである。

- (3) 反応時間には個人差が見られ、マウスとタッチパネルによる違いははっきりとは現れなかった。ま

た、日付のように視覚的にも似通った候補が複数表示される場合に、選択に要する時間が長くなる傾向が見られた。さらに、2Aおよび2Bの実験条件では、発声内容と一致しないが意味的に妥当な候補が表示された場合の選択時間も同様に長めとなった。

- (4) 被験者へのアンケート結果を見ると、本システムの処理速度はタスクの遂行に関して遅過ぎるという印象はないものの、柔軟性の向上が要求されていることが分かる。

表2. システム評価実験方法

実験環境:	実験室内(ほぼ定常的なノイズ環境)
被験者:	20歳台～40歳台の男女各8名 (予備実験とは異なる)
事前説明:	(1) 別室で日韓自動通訳実験の様相を収録したビデオによる説明(5分) (2) 実験室内で、本システムと疑似システムを用いた対話のデモと操作説明(5分)
実験条件:	次の3条件を本システムで順次試行(対話相手側は疑似システムで対向)
(1C)	・準備したシナリオを見ながらの読み上げ発声(利用者側)
(2A)	・被験者自身が記入した予約内容のメモを見ながらの対話発声
(2B)	・前記とは異なる内容の予約メモを渡された被験者による再試行
付随条件:	被験者により認識候補選択の手段をマウス/タッチパネルのいずれかを先に用いて試行
その他:	実験者から被験者への指示 ・音声入力レベルの調整 ・認識失敗時の再発声指示等

表3. 対話環境でのシステム評価実験結果(平均)

成功率の尺度	1C	2A	2B
1回の発声での成功率	95.1%	94.2%	96.5%
1発話当りの発声回数	1.05	1.13	1.07
正解時の1位認識率	88.7%	86.3%	87.2%
1ターン当りの所要時間	6.3sec.	6.3sec.	6.0sec.

注)1ターン当りの所要時間は、発声可能表示から認識候補選択までを指す(この場合1回の発声で成功した時の平均)。

7 現状分析と今後の課題

7.1 到達レベルの位置づけ

我々の日韓自動通訳システムは基本的な性能面では、従来の他のシステムと比較して相当に高いレベルに位置づけられると思われる [4]。特に処理全体の高速性、および発話状態監視に基づく高精度で効率の良い音声認識等が特徴として挙げられる。一方、本システムを実際のタスクに利用しようとする観点から利用者の操作を極力少なくする方向でユーザインタフェースの改良を行なった。この結果、初めてのユーザでもそれほど無理なく対話によるタスク遂行が可能であることが実験により示唆された。

以上のように、今回の研究開発では実用的な日韓自動通訳システムを実現するための第一歩として基盤を作ることに成功したと言える。しかし、現段階では対象となるタスクは小規模に留まっており、誰でも自由に使えるという目標にはまだ相当な道程がある。現在の基盤の上にさらに優れた先進システムの実現を目指すには、音声認識・翻訳等の要素技術の性能面のレベルアップを図るとともに、利用分野等を考慮したシステム統合技術を確立する必要がある。

7.2 ユーザインタフェース上の問題点

今回試作した自動通訳システム用のユーザインタフェースは、翻訳処理を伴わない対話型のマンマシンインタフェースにも共通する問題点を包含している。すなわち現状では、ある程度以上の規模の音声対話システムにおいては、発声内容(認識結果)の確認が避けられない。本システムでは、システムのモニタースクリーンを用いて音声認識候補の上位3位までを表示し、ユーザによる候補の選択を行なっている。しかし、本システムでは認識結果をそのまま候補として表示したため問題が生じている。すなわち、類似の候補が複数表示された場合、自分の発声した通りの表現がどこにあるのか、あるいはそのものでなくても意味的に等価な候補があるのかを判断するのに多少の時間を要する点である。この問題には次のような疑問が内在している。

・発声した通りを忠実に表示するのが良いのか？

認識文法の記述にも関連する問題である。日本語の話し言葉に顕著に見られる助詞の脱落の有無に関して、両方の候補を表示してもユーザは戸惑うだけかもしれない。また、ほぼ同一の意味として用いられる複数の表現を一つにマージして表示した場合、ユーザが直ちに確認できるかという疑問もある。

・候補の数は3個で適切か？

同一レベルの認識候補(例えば日付/部屋の種類/人名等)の分岐の数にもよると思われる。何が認識対象となっているかを判断して候補の数を可変にすることも考えられる。または、スコアの分布等を考慮して確認そのものを行なわない形態も可能かもしれないが、ユーザにとっては一貫性のあるインタフェースの方が望ましいとも考えられる。

ここで提起したような問題を解決するためには、種々の設定条件下で(通訳)対話実験を行なって、システムの性能や被験者の行動を分析する必要がある。システムと人間が対話によってタスクを遂行するような場合には、システムの応答に冗長性を持たせることにより、システムからのメッセージ内容にユーザの発話内容の確認を包含するようなアプローチも可能であるが、人間と人間が対話する自動通訳システムの場合はやや状況が異なる [5]。

今回の対象タスクではシステムに慣れたホテル側の話者による対話の誘導を前提としているが、このような場合もシステムに(特に利用者側の)ユーザモデルを持たせることにより、認識結果の確認を最小限に抑える工夫等が考えられる。

8 まとめ

試作した日韓自動通訳システムについて、特にユーザインタフェース関係の改良内容と、このシステムを用いて行なった双方向の通訳対話実験について報告した。今後、適用可能な分野の調査を含め、将来の自動通訳システムの実現のための検討課題をさらに追究する予定である。

最後に、本研究の遂行に際してご指導・ご助言を頂いたKDD村谷取締役、浦野研究所長と、鈴木次長をはじめとする研究所および関連各位に感謝いたします。

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-9403-4，1995。
- [3] 鈴木・井ノ上・谷戸：“ホテル予約をタスクとする日韓音声翻訳処理について”，言語処理学会第1回年次大会，1995。
- [4] Morimoto, T., et al.: “ATR's Speech Translation System: ASURA”, *EUROSPEECH '93*, 1993.
- [5] 小林・中島・新美：“音声模擬対話における対話制御と快適性について”，情報処理学会研究報告，95-SLP-6，1995。