# 自動翻訳システムを用いた日英対話データの収集

菊井玄一郎, 竹澤寿幸, 鈴木弥生, 西野敦士
ATR 音声言語コミュニケーション研究所
{genichiro.kikui, toshiyuki.takezawa, yayoi.suzuki, atsushi.nishino}@atr.co.jp

## 1. はじめに

音声翻訳システムを設計したり評価したりするためには想定される発話を広くカバーする音声言語コーパスが必要である.

この課題に対して, 我々は, 異なる言語を話す2人の話者に通訳者を介して対話 (模擬対話) をさせることによって将来の翻訳システムに対して利用者がどのような発話を行うかというデータの収集を行ってきた (SLDB)[i]. この方法の問題点は, 1) 広い話題を覆う大量のコーパスを作成するには大きなコストがかかること, 2) 通訳者の能力と現状 (あるいは近未来) の音声翻訳システムの能力には差があるため後者を利用した対話データとして妥当かどうか分からないこと, などがあげられる.

最初の問題に対しては旅行場面で使われることが想定される多様な言語表現を集めた「旅行会話基本表現集 (BTEC)[ii]」と呼ぶパラレルコーパスを作成した. これはこの種の会話の経験が豊かなライターがテキストとして作成するため, 模擬対話を実施して書き起こすより低いコストで広範囲の言語表現を収集することができる. しかし, 実際にシステムに入力される発話のサンプルとしての妥当性については, 音声起源でないことを考慮するとSLDB以上に検証が必要である.

第二の問題に対しては実際に音声翻訳システムを使った対話を行ってデータを取らざるを得ない. しかし, 現状の音声翻訳をそのまま使うと認識誤りや誤訳によって, 頻繁に対話が破綻してデータが得られないことが懸念される[1].

音声対話の分野でも実データの収集とシステム構築とは「鶏と卵の関係」にあることが指摘されており[iii], この問題を解決するためにシステムの機能 (の一部) を人間が代行する Wizard-of-Oz (human-in-the-loop) と呼ばれる方法が用いられている.

同様の問題意識のもと, 本研究では, 音声認識の部分のみタイピストで置き換えた処理系を作り, これを用いた対話データの収集を試みた.

本稿では, まず, 我々のデータ収集方法について説明し, 次に, 収集されたデータの特徴について述べる.

## 2. データ収集システム

### 2.1 全体構成

音声翻訳システムは音声認識・自動翻訳・音声合成という3つの部分から構成される. 今回のデータ収集では翻訳以外の部分に起因する誤りを抑えるため, 音声認識部分をタイピストに置換え, 合成音と共に翻訳結果を携帯型PC上に文字表示することとした. このデータ収集環境の全体構成を図1に示す.

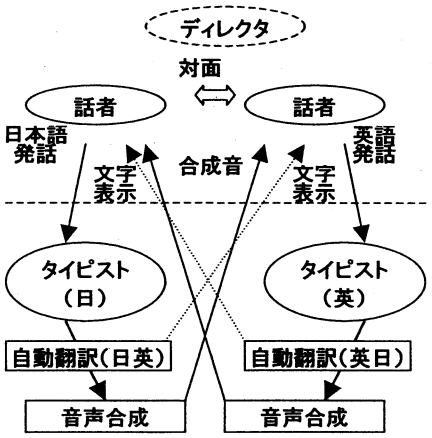


図 1: データ収集環境

日英翻訳には音声翻訳システム ATR-MATRIX [iv]の翻訳部 TDMT (Transfer Driven Machine Translation)[v]を拡張したものと DP マッチングを用いた用例に基づく機械翻訳 D3 (DP-matching Driven transDucer)[vi]の 2002 年 5 月バージョンを組み合わ

[1] たとえば現状の連続音声認識のレベルは文認識率で高々70-80%である。

せたもの[2]，英日翻訳にはTDMTを拡張したものを使用した．なお，第2回目の実験では事前にホテル名等の固有名詞を追加登録した．また，日本語音声合成にはCHATR[vii]，英語音声合成にはCHATR（第1回実験）とAT&T Labs' Natural Voices™(第2回実験)を使用した．

### 2．2　ユーザインタフェース

音声の入出力は携帯電話（第1回実験）およびヘッドフォン付き接話マイク（第2回実験）を使い，携帯電話あるいはノートPCに付いているボタンを押してから対話者に発話させる方式とした．ボタンを押すとタイピストに対して発話開始を示すビープ音が送られる．なお，あらかじめ決められた時間（後述）を超える発話はタイピストには送られないようにした．

## 3．データ収集実験

上記の収集システムを使って現在までに2回のデータ収集実験を行った．

第1回目(MAD1)はこのような設定で本当にデータ収集が可能かを検証することを目的に1日5時間ずつ12日間行った．対話のタスクは例えば「タクシー乗り場を訪ねる」といった簡単なものとした．

第2回目(MAD2)はタスクの達成度と発話の関係を調べることを目的に11日間行った．タスクは例えば「日程や人数等を勘案してパッケージツアーを選んで予約する」といった複合的なものとした．

### 3．1　対話者

対話者として理想的なのは旅行経験豊富，あるいは，旅行サービスに従事する人で，かつ，極力相手側の言語が分からない人であるが，現実的には難しく，実際の対話者には相手言語を若干理解できる人が含まれる．なお，原則として実験日ごとに異なる対話者を使った[3]．

### 3．2　タイピスト

タイピストは日英ともに選抜して訓練することにより，発話が終了してからタイプし終わるまでの時間が発話時間を越えないようにした（1ターンが30～40秒程度）[4]．なお，タイピストにはなるべく忠実に発話を書き起こすよう指示した．

### 3．3　ユーザへのインストラクション

ほとんどの実験参加者（＝対話者）にとって音声翻訳システムを使うことは未知の体験であり，また，対話の設定自体，実験担当者から与えられたものであるから，対話の進め方や発話様式は担当者の指示に大きく依存する．対話者には実験の目的を説明したあと，(単言語の)音声認識を体験させ，さらに，次の点に注意するよう指示した．

・大きめの声で明瞭に話す．

・1回の発話は10秒以内とする[5]

・時々誤りが発生するが，確認や再発話をすることにより対話を続ける．

・時々処理に時間がかかることがあるが，その場合は少し待つ．

なお，2回目の実験では相手との間で了解した（と思った）ことをチェックリストに記入させた

## 4．対話データの分析

### 4．1　語彙・構文的な特徴

以上のようにして得られたデータの基本的な統計量を表1に示す．比較のために，1）通訳者を介した対話(SLDB)，旅行会話基本表現集(BTEC)のデータを併記する．なおSLDBにおいても発話時間は10秒に制限している．

発話あたりの単語数や単文の割合等が発話の複雑さの最も粗い指標であると考えると，MADはBTECとSLDBの中間でSLDBの方にやや近い複雑度と考えることができる．

---

[2] D3で利用する用例距離計算の値が0.2より小さければD3の結果を採用し，それ以外はTDMTを拡張したものの結果を採用する．

[3] 1回目の実験では午前と午後で日本語話者を変える一方，同じ日本語話者を異なる英語話者と組み合わせて使った。

[4] 10秒の音声を速やかに通訳しても音声で伝えるのに10秒要し，相手が10秒で答えたものを再度通訳して音声で伝えるのに10秒要するとすれば，合計40秒となる．

[5] SLDB収集時に経験的に得られた適切な値を採用した．

表 1：発話の長さ

| | MAD1 | MAD2 | BTEC | SLDB |
|---|---|---|---|---|
| 発話数(日英計) | 3568 | 3404 | 17万 | 32,168 |
| 対話数 | 445 | 69 | − | 618 |
| 発話あたり平均単語数（日／英） | 10.0/ 10.3 | 12.6/ 11.1 | 6.9/ 5.9 | 13.3/ 11.3 |
| 発話あたり平均文数（日／英） | 1.29/ 1.61 | 1.44/ 1.54 | 1.07/ 1.08 | 1.35/ 1.38 |
| 単文の割合（日） | 68.3% | 72.0% | 82.8% | 65.9% |

次に日本語についてBTEC，SLDB各々の言語表現がMAD1で対話者の発話した言語表現をどの程度カバーしているかをn-gramの被覆率で表す図2とのようになる．

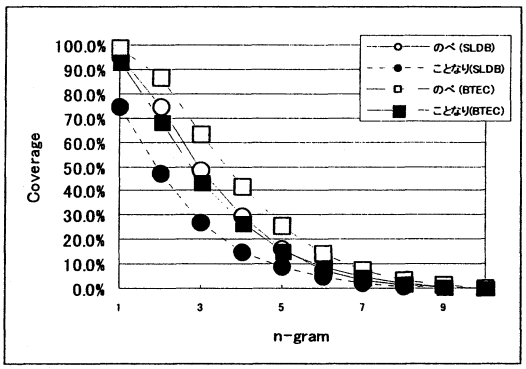


図 2：収集データに対する既存データのNgram被覆率

BTECとSLDBの絶対的なデータ量の違いを考慮する必要があるが，前者の方が後者よりもMADデータを広くカバーしている．のべ数でみるとBTECはMAD1のtrigramの60%以上の高いカバー率を示している．

一方，文レベルのデータの差を見るために同じデータについて内容語列，および，付属語列の一致度を調べると表2のようになり，MADはBTECより会話起源のSLDBに似ていることが分かる．

表 2：文全体のカバー率(延べ数／異なり数)

| | BTEC | SLDB |
|---|---|---|
| 内容語列 | 21.0% / 18.6% | 28.2% / 20.2% |
| 付属語列 | 46.9% / 44.1% | 52.6% / 50.8% |

## 4．2　対話レベルの特徴

今回の対話データ収集の目的の一つは翻訳誤りが対話の進行やタスクの達成にどういう影響を与えるかを超去することである．以下では現在までに得られた分析結果について述べる．

### 対話の進行パターン

翻訳誤りが発生した場合に対話がどう進行するかは大きく次の5種類に分類できる．なお，括弧内の数字は第1回目実験における個数で，2つの分類に属するもの27個を含む．

1）再発話要求(51)

「もう一度言ってください」のような単純な再発話要求．相手の発話の内容が推論不可能な場合に使われる．再発話で回復できたものは32例あった．

2）言い換え質問(60)

誤りが応答的発話で生じた場合に質問（要求的発話）の仕方を変えて発話する，あるいは，訳文に対する推測結果を確認する表現を埋め込むもの．うまく回復できたものは30例．

3）特異的反応なし（135）

対話の流れを見る限り，翻訳誤りがない場合と変わらない．原文の意味が正しく推論できている場合にはこうなるが，後述するように逆は成り立たない．

4）ディレクターによる介入(37)

下記のような場面ではディレクターが介入し，必要であれば代替表現を教示して再発話させた．

A)タイピストのスペリングミス，表記ミス(5)

B)未登録語（固有名詞）の発話（別の固有名詞を提示した）(14)

C)対話者が代替表現を思いつかない場合[6](18)

数量表現の誤訳，情報の脱落や過剰，語順ミス

5）対話の中止(17)

次のような場面では対話そのものを中止した．

A)何回言い換えても理解可能な訳が出ない

B)対話の脱線(脱線部分の誤訳)

C)対話参加者からの中止要請

上記1，2で誤りから回復できたものは，相手か

[6] 本来は介入すべきではないが，ある程度は対話を進行させないとデータ量が確保できないのに加え，対話者のモチベーションが下がるのであえて介入した。

らの再発話要求や言い換え質問を契機として，対話者が複文を単文の列に言い換えたり，イディオムの少ない文にしたりするなど，翻訳システムにとってより「平易な」文に言い換えた場合である（末尾に対話例を添付）．以上から，翻訳誤りの一部（２割程度）が対話を通して検出・訂正されうることが実対話において確認できた．

### 対話の進行パターンとタスク達成度

翻訳装置を使った対話タスク達成の可能性という観点から考えると，上記の４，５は失敗したものであり，１，２はその時点で致命的になっていないものである．問題は３で，外部からの観測だけでは対話が正常に進んでいるのかどうか区別できない場合が多い．例えば，単に「分かりました」と言うだけで相手側の発話内容の復唱等がない場合には翻訳結果からどれだけの情報を汲み取っているのかが分からない．

そこで，第２回の実験ではチェックリストによって理解度が外部から判断できるようにした．しかし，このことは対話者が確認発話を多発する[viii]など対話の流れに対して影響を与えたものと思われる．

その他，対話者によっては相手側発話の一部を直接理解して対話を進めるケースも見られたため，各対話者の相手側言語の能力を測っておくことも必要であると思われる．

## ５．おわりに

翻訳システムを介して目的志向の対話を行わせることにより，翻訳システムを使った場合の対話行動に関するデータ収集を行った．発話の長さや表現などは既存のデータを組み合わせたものである程度カバーできることが分かった．また，翻訳誤りの一部は再発話や言い換えによって回復できることが分かった．なお，翻訳品質と対話の進行やタスク達成との間の定量的な関係等，さらなる分析は今後の課題である．

### 謝辞

翻訳エンジンを提供して下さった隅田英一郎氏ほか翻訳グループの皆様，データ収集実験の実作業を手伝って下さった高嶌浩司，松井孝典，染川智子各氏，単語列一致度の分析ツールを作成して頂いた柏岡秀紀氏に心より感謝申し上げる．

本研究は通信・放送機構の研究委託「大規模コーパスベース音声対話翻訳技術の研究開発」により実施したものである．

### 参考文献

[i] Morimoto, T. et al.: A speech and language database for speech translation research, Proc. 3rd ICSLP (1994).

[ii] Takezawa, T. et al.: Toward a broad-coverage bilingualcorpus for speech translation of travel conversations, Proc. 3rd LREC (2002).

[iii] Zue, V. et al.: Conversational Interfaces: Advances and Challenges, Proc. IEEE, Vol.88, No.8 pp.1166-1180 (2000).

[iv] 菅谷他：音声翻訳システム ATR-MATRIX の開発と評価，情処論，Vol.43, No.7, pp.2230-2241 (2002).

[v] 古瀬他：構成素境界解析を用いた多言語話し言葉翻訳,自然言語処理，Vol.6, No.5, pp.63-91 (1999).

[vi] Sumita, E.: Example-based machine translation using DP-matching between word sequences, Proc. ACL-2001 Workshop on Data-Driven Methods in Machine Translation, pp.1-8 (2001).

[vii] Campbell, N.: CHATR: A high-definition speech re-sequencing system, Proc. ASA/ASJ Joint Meeting, pp.1223-1228 (1996).

[viii] 竹澤他:コーパスベース音声翻訳研究のための対話データ収集,情処研報，Vol.2003, No.14, pp.71-76 (2003).

(対話サンプル)言い換え発話によって回復した例
英: What kind of wine did you have in mind?
訳:どのようなワインが前もってありましたか？
日: もう一度言っていただけますか
訳: Could you say that again?
英: What kind of wine would you like?
訳: どのようなワインが良いですか？