

TR-IT-0136

言語データベースの概要
Overview of Linguistic Database

古瀬 蔵 増井 淳子*
Osamu Furuse Atsuko Masui

1995年10月

概要

ATR音声翻訳通信研究所で構築している日英・日韓のバイリンガル会話、日英韓旅行会話表現集などの言語データについて、収集方法、ファイルの種類、ファイル名の命名法、ディレクトリ構成など概要を説明する。

エイ・ティ・アール音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories
©(株)エイ・ティ・アール音声翻訳通信研究所
1995 by ATR Interpreting Telecommunications Research Laboratories

目次

1 はじめに	1
2 言語データの種類	2
3 会話の収集方式	6
4 ファイルの種類	9
5 英語の修正	19
6 ディレクトリの構成	21
7 ファイル名	23
8 課題遂行型会話	29
9 今後の予定	32
付録A ディレクトリ構成とファイル数	34
付録B 課題遂行型会話の基本設定と会話 I D	39

1 はじめに

ATR音声翻訳通信研究所は、音声データと言語データをリンクさせた音声言語データベース [1] を構築している。しかし、翻訳処理研究には、より多くの語彙や表現や場面例や、英語だけでなく韓国語などを含めた多言語のデータが必要であるし、対話処理研究には、より実会話に近い言語データがあることが望ましい。

このような観点から音声言語データベースを補完するために、ATR音声翻訳通信研究所は、翻訳処理や対話処理などの自然言語処理に利用することに重点を置いた言語データベースの構築も行なっている。言語データベースの主な特徴は、(1) 日日モノリンガル会話、英英モノリンガル会話、日英バイリンガル会話、日韓バイリンガル会話、日英韓対応表現集など様々な形態のデータがある、(2) 対象トピックがホテル予約とホテル内サービスだけでなく旅行会話全般に渡っている、(3) 電話会話だけでなく対面会話も含む、(4) 複数の会話をシリーズにした課題遂行型会話がある、(5) 音声データとリンクしておらず、模擬会話はオフィス環境で収集されている、などである。

言語データベースは、音声言語データベース同様、ATR音声翻訳通信研究所のファイルサーバーにあり、ATR音声翻訳通信研究所内で利用可能である。

本報告書は、1995年9月時点での言語データベースの概要について説明する。内容は、2節で言語データの種類について、3節で会話を収集した際の方式について、4節で言語データベースのファイルの種類について、5節で英語テキストファイルの英語修正について、6節でファイルのディレクトリ構成について、7節で各ファイル名称の決め方について、8節で日英バイリンガルの課題遂行型会話の設定内容について、9節で言語データベースの今後の拡張予定について説明している。

尚、マルチモーダル、N人会話に関する言語データは、本報告書の範囲外とする。

2 言語データの種類

現在の言語データベースのデータは、収集著作方法によって次の3種類に大別できる。

- (1) モノリンガル会話
- (2) バイリンガル会話
- (3) 旅行会話表現集

(1) と (2) は、模擬会話の書き起こしによって収集されている。
(3) は、机上で著作されている。

以下に、各々のデータの内容について説明し、(1) と (2) の収集方法については、3節で説明する。

2.1 モノリンガル会話

モノリンガル会話は、通訳者を介さず1言語で会話を行い、後から別言語への翻訳結果を付与した形式で収集された会話である*1。現在は、英英モノリンガル会話と日日モノリンガル会話がある。

英英モノリンガル会話は、英語話者のみによる会話に、後から日本語の翻訳を加えた会話であり、日日モノリンガル会話は、日本語話者のみによる会話に、後から英語の翻訳を加えた会話である。図2-1、2-2にそれぞれ英英モノリンガル会話、日日モノリンガル会話のテキストファイルを示す。

図2-1 英英モノリンガル会話のテキストファイル

```
ファイル名:tas11001.euc
E*2|Hello, New Washington Hotel.*3
J|はい、ニューワシントンホテルでございます。*4
*5
E|Yes, I'd like to make a reservation, please.
J|もしもし、予約をお願いしたいんですけど。
E|Okay.*6|Can you tell me when you'll be staying and how many persons?
J|かしこまりました。|いつのお泊まりで、何名様でしょうか。
```

図2-2 日日モノリンガル会話のテキストファイル

```
ファイル名:tas10001.euc
J|もしもし。|京都観光ホテルでございます。
E|Hello.|This is the Kyoto Tourist Hotel.|May I help you?
J|部屋の予約をお願いしたいんですが、一泊いくらぐらいから有りますか。
E|I'd like to reserve a room.|What is the rate for one night?
J|はい、一万円から二万五千円まででございます。|いつのお泊まりですか。
E|[well] We have rooms ranging from ten thousand yen to twenty-five thousand yen.|When would you like to stay?
```

*1 ただし、今後、すべてのモノリンガル会話に後づけ翻訳を行なわなうわけではない。

*2 言語タグとして日本語: 半角の"J", 英語: 半角の"E"を記述する。

*3 1発話ごとに話者の書き起こしを記述する。

*4 1発話ごとに別言語の翻訳結果を記述する。

*5 話者の交代を改行で示す。

*6 言語タグと発話の開始の区切り、1発声内の文の区切りに"|"(半角)を挿入する。

2.2 バイリンガル会話

バイリンガル会話は、1発話ごとに通訳者を介して、2言語間で会話を行う形式で収集された会話である。現在は、日英バイリンガル会話と日韓バイリンガル会話がある。

日英バイリンガル会話は、日本語話者、英語話者、日英通訳者、英日通訳者の4者で行った会話であり、日韓バイリンガル会話は、日本語話者、韓国語話者、日韓通訳者、韓日通訳者の4者で行った会話である。

日英バイリンガル会話は、93年下期、94年上期にシナリオ方式またはチェックシート方式により収集を行っており、94年下期は課題遂行型方式により収集を行っている。日韓バイリンガル会話は、すべてシナリオ方式により収集を行っている。図2-3に日英バイリンガル会話のテキストファイルを示す。

図2-3 日英バイリンガル会話のテキストファイル

ファイル名: A23001.euc

J|ありがとうございます、京都<kyoto>クイーンズホテルでございます。

E|Good morning, Kyoto Queen's Hotel, may I help you?

E|Yes, hello. |[ummh] My name is Carol Simmons and I've just arrived in Kyoto.

J| [あ] すいません。| キャロル・シモンズと申しますが、今京都<kyoto>に着いたところなんです。

E|And I was hoping to book a single room beginning today if that's possible.

J|それですね、きょうからシングルルームをお願いしたいんですが、空<a>いていますか。

J|はい、ありがとうございます。|きょうからシングルルーム、いつまで、ご宿泊のご予定でしょうか。

E|Yes, [ummh] a single room starting today. |And, how long will you be staying?

2.3 旅行会話表現集

旅行会話表現集は、模擬会話を収集したものではなく、会話場면을想定して著作した文を集めたものである。現在は、日英韓旅行会話表現集を収集している。

想定する会話場面は、以下の通りである。

- (1) 日本場面での、米国人旅行者と現地日本人の会話
- (2) 米国場面での、日本人旅行者と現地米国人の会話
- (3) 日本場面での、韓国人旅行者と現地日本人の会話
- (4) 韓国場面での、日本人旅行者と現地韓国人の会話
- (5) 特定の国場面に依存しない普遍的な旅行会話

すべての会話場面に対して、日本語、英語、韓国語の表現を与えており、例えば(1)、(2)に対する韓国語文、(3)、(4)に対する英語文も用意している。著作の基本となるオリジナル言語を仮定して文を作成した上で、他の言語の対応文が与えられる。例えば、日本語がオリジナル言語であれば、日本語で旅行会話表現を作成し、その日本語表現に対して、英語、韓国語の翻訳を与える。同じ会話、トピックとオリジナル言語で著作された表現を集めて、ファイル単位としている。1つのファイルは、会話を構成しているのではなく、そのトピックで使われる表現を列挙している。図2-4は、ホテルの予約を会話トピックとし、オリジナル言語を日本語とする旅行会話表現集のテキストファイルである。

図2-4 日英韓旅行会話表現集のテキストファイル

```
ファイル名 : A4JP
J|XX|AU4JP001|今晚のホテルを予約したいのですが。
E|XX|AU4JP001|I would like to make a reservation for tonight.
K|XX|AU4JP001|오늘밤 호텔을 예약하고 싶은데요.

J|XX|AU4JP002|何名さまでですか。
E|XX|AU4JP002|How many people?
K|XX|AU4JP002|몇 분이십니까?

J|XX|AU4JP003|一名です。|シングルを一つお願いします。
E|XX|AU4JP003|One. |I would like to have a single room.
K|XX|AU4JP003|싱글을 하나 부탁드립니다.
```

3 会話の収集方式

この節では、モノリンガル会話、バイリンガル会話の収集方法について述べる。モノリンガル会話、バイリンガル会話の収集は、以下の3種類の方式により行なわれている。

- (1) シナリオ方式
- (2) チェックシート方式
- (3) 課題遂行型方式

3.1 シナリオ方式

シナリオ方式では、(1) あらかじめトピックについての標準会話を作成し、それを話者が予習により会話の内容を把握し、収録時には、その標準会話を見ないで、会話を行なう方式、または、(2) 標準会話を用いずに、トピックについての設定内容(名前、日時、話者の目標、状況など)を示した資料に基づいて会話を行なう方式、により会話収集を行なう。この方式は、モノリンガル会話、日韓バイリンガル会話、日英バイリンガル会話の収集に用いられている。図3-1に設定内容の例を示す。

3.2 チェックシート方式

チェックシート方式は、標準会話を用いずに、あらかじめトピックについての会話を行なう上で必要な事項について、選択肢を設けたチェックシート型の資料を作成し、その資料を用いて会話を行なう方式により会話収集を行なう。この方式は、音声言語データの会話の一部の収集に用いられている。図3-2でチェックシートの例を示す。

3.3 課題遂行型方式

この方式は、日英バイリンガル会話の一部の収集に用いられている。課題遂行型方式については、8節「課題遂行型会話」で詳しく述べる。

図3-1 設定内容例

ホテルの予約

スピーカーAの資料

氏名／年齢：鈴木和夫 30歳（日本人・男性）

日時：1994年10月1日午後3時

立場：ニューヨーク滞在中の日本人

要望：10月10日1泊で予約をとりたい。

確認：ツインからトリプルへの変更は可能か。

スピーカーBの資料

氏名／年齢：ジョン・フィリップス 32歳（アメリカ人・男性）

日時：1994年10月1日午後3時

立場：ホテルデュポンプラザ（ワシントンDCアメリカ）のフロント係

ツイン料金：\$100から

エクストラベッド料金：\$30

図3-2 チェックシート

<1> 非滞在者 (I) - 予約 (1)

ID: A

I-1-A (1)

■話者Aの設定ノート

(基本設定)

日時: 1994年7月20日 (水) 午後4時

話者: ジョン/メアリ・フィリップス (会社員 旅行者)

年齢: 20代 30代 40代 50代

場所: 神戸 (自宅 会社 宿泊先)

電話先: 都ホテル (京都)

目的: 直接、ホテルに電話してシングルまたはツインの部屋を予約する。

(詳細設定)

宿泊予定者:

本人 家族 友人 会社同僚 その他

希望宿泊日:

1泊 8/10 (水) チェックアウト 8/11

2泊 8/10 (水) ~11 (木) チェックアウト 8/12

3泊 8/10 (水) ~12 (金) チェックアウト 8/13

その他 (泊 8/ () ~ () チェックアウト 8/)

人数:

1人 2人 3人 その他

部屋の種類:

第一希望 シングル 第二希望 ツイン

チェックイン:

午後3時 午後5時 午後7時 その他 (時)

朝食:

注文する - アメリカン コンチネンタル その他 ()
(6時 7時 8時 9時 10時)

注文しない

説明を聞いてから決定

注文する - アメリカン コンチネンタル その他 ()
(6時 7時 8時 9時 10時)

注文しない

支払方法:

現金

トラベラーズチェック

クレジット・カード (番号: VISA 1598-2145-3914-0863 有効期限: 1995年4月)

不明

連絡先:

自宅 (078-331-8251)

会社 (CLE 078-871-5101)

宿泊先 (ホテル神戸 078-221-5431)

その他 (- -)

4 ファイルの種類

モノリンガル、バイリンガルの各会話、表現集の各トピック、オリジナル言語ファイルは、表4-1に示すように複数の種類のファイルより構成されている。これらをファイル形式により分類すると、次のようになる。

- (1) 会話属性ファイル
- (2) 多言語対応テキストファイル
- (3) 単一言語テキストファイル
- (4) 形態素タギングファイル

以下に、各々のファイルの内容について説明する。

表4-1 ファイルの種類

	モノリンガル会話		バイリンガル会話		旅行会話 基本表現集
	日日	英英	日英	日韓	日英韓
会話属性 (.ENV)			○	○	
多言語対応					
日英 (.EJTEXT)	○	○	○		
日韓 (.JKTEXT)				○	
日英韓 (.JEKTEXT)					○
単一言語					
日本語 (.JTEXT)	○	○	○	○	○
英語 (.ETEXT)	○	○	○		○
韓国語 (.KTEXT)				○	○
形態素タギング					
日本語 (.JMOR)	△	△	○	○	○
英語 (.EMORT)			○		○
韓国語 (.KMOR)				○	○

○: 現在作成済み、または、作成中を表わしている。

△: 作成予定を表わしている。

() 内は、ファイル名の拡張子 (7節参照) を表わしている。

4.1 会話属性ファイル

会話属性ファイルは、図4-1に示すように、モノリンガルとバイリンガルの模擬会話が収集された際の諸条件(収録日時、総会話時間など)が記載されている。

図4-1 会話属性ファイル

```
[AT120011.ENV]
会話ID:AT120011
収録日時:1994年06月14日
収録社名:日本アイアール
総会話時間:04分43秒
有効会話時間:NIL
ノイズレベル:<オフィスルーム>
マイクロホン:ECM-T140<SONY>
DATの機種:NIL
ミキサーの機種:AT-MX50<AUDIO_TECHNICA>
サンプリング周波数:NIL
発話形態:自由会話
ドメイン:旅行
トピック:ホテルの部屋の予約
言語パターン:日本語-通訳者-英語
発声方法:文発声
対面・非対面:非対面
発呼者:山田・太郎__男性__30__埼玉__会社員
被呼者:メアリー・フィリップス__女性__35__ブリティッシュコロンビア__アナウンサー
通訳者1:スーザン・トンプソン__女性__26__オーストラリアンキャピタルテリトリー__通訳
通訳者2:鈴木・花子__女性__26__神奈川__通訳
翻訳者:NIL
コメント:NIL
```

4.2 多言語対応テキストファイル

多言語対応テキストは、話者の会話とそれに対応する通訳、翻訳を含めて書き起こしたテキストのことである。現在、(1) 日日、英英のモノリンガル会話と日英バイリンガル会話から作成した、日英対応テキスト、(2) 日韓バイリンガル会話から作成した、日韓対応テキスト、(3) 日英韓の旅行会話表現集から作成した、日英韓対応テキストの多言語対応テキストがある。

日英バイリンガル会話では、英語の書き起こしに対して修正を行ない、修正結果の入力を行なっている。英語の修正に関しては、5節で説明する。

ファイルの文字コードは基本的にはEUCコードであるが、韓国語を含むテキストについては、日本語とハングル文字が共存できるよう、Junetコードを使用している。

多言語対応テキストファイルの形式は、モノリンガル会話とバイリンガル会話では同一で、旅行会話基本表現集だけは、別の形式になっている。図4-2にモノリンガル会話、図4-3にバイリンガル会話、図4-4に旅行会話表現集の多言語対応テキストファイルを示す。

図4-2 モノリンガル会話の多言語対応テキストファイル

```
ファイル名 : tas11001.euc
E|Hello, New Washington Hotel.
J|はい、ニューワシントンホテルでございます。

E|Yes, I'd like to make a reservation, please.
J|もしもし、予約をお願いしたいんですけど。

E|Okay. |Can you tell me when you'll be staying and how many persons?
J|かしこまりました。|いつのお泊まりで、何名様でしょうか。
```

図4-3 バイリンガル会話の多言語対応テキストファイル

ファイル名:AT120011.euc
E|Good afternoon, the Peninsula New York.|May I help you?
J|はい、ペニンシュラニューヨークでございます。
*4
J|そちらでやっている、週末の宿泊割引を利用したいんですが。
E|I'd like to know about the weekend discount.
*5J|二人でディナー込みで一泊三百ドルでお願いしたいんですけども。
E|We have a budget around three hundred dollars for two persons including dinner.

E|All right.|We have a weekend special.
J|はい。|週末特別割引をご用意しております。

図4-4 旅行会話表現集の多言語対応テキストファイル

ファイル名:A4JP
J|XX|AU4JP001|今晚のホテルを予約したいのですが。
E|XX|AU4JP001|I would like to make a reservation for tonight.
K|XX|AU4JP001|오늘밤 호텔을 예약하고 싶은데요.

J|XX|AU4JP002|何名様ですか。
E|XX|AU4JP002|How many people?
K|XX|AU4JP002|몇 분이십니까?

J|XX|AU4JP003|一名です。|シングルを一つお願いします。
E|XX|AU4JP003|One.|I would like to have a single room.
K|XX|AU4JP003|싱글을 하나 부탁드립니다..

J|XX|AU4JP004|はい、ご用意できます。
E|XX|AU4JP004|Certainly, we can arrange that for you.
K|XX|AU4JP004|예, 준비됩니다.

4.3 単一言語テキストファイル

単一言語テキストファイルは、多言語対応テキストからある1言語の部分を抜き出し作成したテキストである。現在、バイリンガル会話と会話表現集から作成した、日本語、英語、韓国語のテキストがある。

単一言語テキストファイルの形式は、多言語対応テキストと同じくバイリンガル会話と会話表現集から作ったものとで異なっている。日英バイリンガル会話の日本語テキストファイル、英語テキストファイルをそれぞれ図4-5、図4-6に示す。また、旅行会話表現集の日本語テキストファイル、英語テキストファイル、韓国語テキストファイルをそれぞれ図4-7、図4-8、図4-9に示す。

図4-5 バイリンガル会話の日本語テキストファイル

```
ファイル名 : AT120011.JTEXT
通訳者*1 : *2はい、ペニンシュラニューヨークでございます。
申込者 : そちらでやっている、週末の宿泊割引を利用したいんですが。
          *3二人でディナー込みで一泊三百ドルでお願いしたいんですけども。
通訳者 : はい。週末特別割引をご用意しております。
          割引は土曜日の宿泊で、ツインのお部屋のみとなっております。
```

図4-6 バイリンガル会話の英語テキストファイル

```
ファイル名 : AT130022.ETEXT
interpreter: Yes. We can reserve for you seven twin rooms and six
single rooms for the twenty-fifth and twenty-sixth.
customer: Can I book ten rooms for the twenty-seventh and the
twenty-eighth?
interpreter: Yes, that can be so arranged.
customer: Okay. That would be wonderful.
```

-
- *1 話者の役割を示すラベルとして、日本語は「申込者」、「担当者」、「通訳者」、英語は、「customer」、「clerk」、「interpreter」で示している。
 - *2 話者ラベルの後に、日本語は「:(全角)」を、英語は「:(半角)」を記述する。
 - *3 1発声ごとに改行し、2発声目からは、日本語は8 byte、英語は13 byteの字下げを行う。

図4-7 旅行会話表現集の日本語テキストファイル

ファイル名 : J-A4JP
J|XX|AU4JP001|今晚のホテルを予約したいのですが。
J|XX|AU4JP002|何名様ですか。
J|XX|AU4JP003|一名です。|シングルを一つお願いします。
J|XX|AU4JP004|はい、ご用意できます。
J|XX|AU4JP005|部屋代はいくらですか。
J|JK|AU4JP006|シングル一部屋十九万ウォンになります。
J|XX|AU4JP007|あれ、ガイドブックで見た料金と違うようですが。
J|XX|AU4JP008|今日は週末なので料金は高くなります。
J|XX|AU4JP009|エアコン付きの部屋をお願いします。
J|XX|AU4JP010|バス付きシングルはいくらになりますか。

図4-8 旅行会話表現集の英語テキストファイル

ファイル名 : E-A4JP
E|XX|AU4JP001|I would like to make a reservation for tonight.
E|XX|AU4JP002|How many people?
E|XX|AU4JP003|One. |I would like to have a single room.
E|XX|AU4JP004|Certainly, we can arrange that for you.
E|XX|AU4JP005|What is the price of the room?
E|JK|AU4JP006|It's one hundred and ninety won for a single room.
E|XX|AU4JP007|Really?|It seems to be different from the rate that I saw in the guidebook.
E|XX|AU4JP008|Since it is the weekend today, the rates are higher.
E|XX|AU4JP009|I would like a room with an air conditioner.
E|XX|AU4JP010|How much is a single room with bath?

図4-9 旅行会話表現集の韓国語テキストファイル

ファイル名 : K-A4JP
K|XX|AU4JP001|오늘밤 호텔을 예약하고 싶은데요.
K|XX|AU4JP002|몇 분이십니까?
K|XX|AU4JP003|싱글을 하나 부탁드립니다.
K|XX|AU4JP004|예, 준비됩니다.
K|XX|AU4JP005|방 값은 얼마입니까?
K|JK|AU4JP006|싱글 룸 하나에 십구만원입니다.
K|XX|AU4JP007|아?|가이드 북에서 본 요금과 다른 것 같은데요.
K|XX|AU4JP008|오늘은 주말이기 때문에 요금은 비싸집니다.
K|XX|AU4JP009|에어콘이 있는 방으로 부탁드립니다.
K|XX|AU4JP010|욕실이 있는 싱글은 얼마나 됩니까?

4.4 形態素タギングファイル

形態素解析タギングファイルは、文を形態素単位に分割し、それぞれに品詞や活用などの形態素情報を付与したデータである。現在は、日本語、英語、韓国語の形態素タギングファイルがある。

日本語には、ATR音声翻訳通信研究所での標準体系 [2] と変換主導翻訳システム用の2種類の体系がある。変換主導翻訳システム用の日本語形態素解析については、音声言語データベース、言語データベースのすべてのテキストに対して形態素タギングファイルの作成を行う予定である。図4-10に標準体系の日本語形態素タギングファイル、図4-11に変換主導翻訳システム用の日本語形態素タギングファイルを、図4-12に英語形態素タギングファイルを、図4-13に韓国語形態素タギングファイルを示す。

図4-10 標準体系日本語形態素タギングファイル

```
ファイル名 : AT120012.JMOR
10:通訳者 :
10:0010:10:10:はい|ハイ|はい|感動詞
10:0010:10:20:、|。|記号
10:0010:20:30:ペニンシュラニューヨーク|ペニンシュラニューヨーク|ペニンシュラニューヨーク|
固有名詞
10:0010:20:40:で|デ|で|助動詞|特殊サ|語幹
10:0010:20:50:す|ス|す|語尾|特殊サ|終止
10:0010:20:60:。|。|記号
20:申込者 :
20:0020:30:70:週末|シュウマツ|週末|普通名詞
20:0020:30:80:の|ノ|の|連体助詞
20:0020:40:90:宿泊|シュクハク|宿泊|サ変名詞
20:0020:40:100:割引|ワリビキ|割引|サ変名詞
```

図4-11 変換主導翻訳システム用日本語形態素タギングファイル

```

ファイル名 : AT120012.JMOR
10:0010:10:10:はい|ハイ|はい|感動詞|
10:0010:20:20:ペニンシュラ|ニュー|ヨーク|ペニンシュラ|ニュー|ヨーク|ペニンシュラ|ニュー|ヨーク|
普通名詞|
10:0010:20:30:です|デス|です|判定詞|特殊サ|基本|
10:0010:20:40:。|。|記号|
20:0020:30:50:週末|シュウマツ|週末|普通名詞|
20:0020:30:60:の|ノ|の|連体助詞|
20:0020:40:70:宿泊|シユクハク|宿泊|サ変名詞|
20:0020:40:80:割引|ワリビキ|割引|サ変名詞|

```

図4-12 英語形態素タギングファイル

```

ファイル名 : AT120012.EMORT
;;; Hello
|10|10|hello|hello| ||INTERJ||
;;; The Peninsula New York
|20|10|the|the| ||DET||
|20|20|peninsula new york|peninsula new york| ||PROPN||
;;; |20|30|new york|new york| ||PROPN||
;;; May I help you
|30|10|may i help you|may i help you| ||INTERJ||
;;; We're offering a discount for a stay on Saturday in a twin room
for two people
|40|10|we|we| |1P|PRON||
|40|20|'re|be| |1P|BEV||
|40|30|offering|offer|ING| |V||
|40|40|a|a| ||DET||
|40|50|discount|discount| ||CN||
|40|60|for|for| ||PREP||
|40|70|a|a| ||DET||
|40|80|stay|stay| ||CN||
|40|90|on|on| ||PREP||
|40|100|saturday|saturday| ||CN||

```

図4-13 韓国語形態素タギングファイル

ファイル名 : AF440011.KMOR

#1

0.어서 오세요

1.

a.어서 오세요

c.AsA oseyo

d.어서 오세요

f.(감동사)

#2

/./기호

#1

0.예약

1.

a.예약

c.yeyaG

d.예약

f.(동작명사)

0.을

1.

a.을

c.UL

d.를

f.(목적격조사)

5 英語の修正

日英バイリンガル会話には、英語ネイティブでない日本人の日英通訳者が、限られた時間で逐次通訳を行なっているために、英語として不自然な表現が含まれていることがある。そこで、書き起こしの英語部分の修正が必要な箇所に、書き起こしに加えて修正結果の入力を行っている。修正の方針としては、(1) 原文の構造をできるだけ維持して、英語処理の入力データとして使用可能な水準の文となるような修正、(2) 原文の構造をできるだけ維持して、適切な英文を得ることが困難な場合に限り、構造を変更した修正、(3) 文法的に不適格でも、話し言葉として自然な場合は修正を行わない、こととしている。修正の対象となるのは、以下のような表現である。

- ・ 語句、表現の不適切な選択による不自然な表現
- ・ 日本語原文に対して情報が不足している表現
- ・ 日本語原文に対して余分な情報が付加されている表現
- ・ 文脈上の不適切な表現
- ・ 過度の文脈推論、過度の意識がされている表現

1993年下期と1994年上期のテキストは、会話を書き起こしただけの対応テキストと、修正結果を加えた英語修正情報付の対応テキストの2種類に分けて作成している。また、それ以外の1994年下期のテキストは、修正結果を加えた英語修正情報付の対応テキストのみである。

図5-1に英語修正情報付日英対応テキストを、図5-2に英語修正情報付日英対応テキストから作成した英語テキストを示す。

図5-1 英語修正情報付日英対応テキスト

ファイル名: AT130022.EJTEXT
J|はい、二十五日と二十六日はツインルームを七部屋<nanaheya>、シングルルームを六部屋、大丈夫でございます。
E-^{*1}|Yes.|We can reserve for you seven twin rooms and six single rooms for twenty-fifth and twenty-sixth.
E+^{*2}|Yes.|We can reserve for you seven twin rooms and six single rooms for the twenty-fifth and twenty-sixth.
E|Can I book ten rooms for the twenty-seventh and the twenty-eighth?
J|二十七日<nijuushichinichi>と二十八日はツイン十部屋予約できますか。
J|はい、大丈夫でございます。
E|Yes, that can be so arranged.
E|Okay.|That would be wonderful.
J|では、それをお願いします。

図5-2 英語修正情報付日英対応テキストから作成した英語テキスト

ファイル名: AT130022.ETEXT
interpreter: Yes. We can reserve for you seven twin rooms and six single rooms for the twenty-fifth and twenty-sixth.
customer: Can I book ten rooms for the twenty-seventh and the twenty-eighth?
interpreter: Yes, that can be so arranged.
customer: Okay. That would be wonderful.

- *1 修正を必要とする発話の言語タグの直後に"-"(半角)を入れ、英文修正の発話を表わす。
- *2 修正した結果の発話の言語タグの直後に"+"(半角)を入れ、修正結果の発話を表わす。

6 ディレクトリの構成

言語データベースのディレクトリ構成は、以下のようになっている。
現在の各ファイル数などに関しては、付録Aに示している。

/DB/LDB/ — | ;英英モノリンガル会話

| — EE/ — | — DELIVERY/ — | — 9312/

|

| ;日日モノリンガル会話

| — JJ/ — | — DELIVERY/ — | — 9312/

|

| ;日英バイリンガル会話

| — JE/ — | — ENV/ ;会話属性

| | — EJTEXT/ ;日英対応テキスト

| | — EJTEXT+/ ;英語修正情報付日英対応テキスト

| | — ETEXT/ ;英語テキスト

| | — ETEXT+/ ;英語修正版テキスト

| | — JTEXT/ ;日本語テキスト

| | — EMOR/ ;英語形態素解析データ

| | — JMOR/ ;標準体系日本語形態素解析データ

|

以上のディレクトリは、収集時期によって分類した以下のような
4つの下位ディレクトリを持つ。

— | — SOBA93/ ;93年下期収集

| — SOBA94A/ ;94年上期収集

| — FURU94B/ ;94年下期収集

| — DEN94B/ ;94年下期収集 (課題遂行型会話)

|

| ;日韓バイリンガル会話

| — JK/ — | — ENV/

| | — JKTEXT/

| | — JTEXT/

| | — KTEXT/

| | — KMOR/

|

以上のディレクトリは、収集時期と収集業者によって分類した以
下のような4つの下位ディレクトリを持つ。

— | — C9410/ ;94年10月収集

| — C9502/ ;95年2月収集

| — S9502/ ;95年2月収集

(次項へ)

/DB/LDB/ — | ;日英韓会話表現集

| — JEK/ — | — JEKTEXT/
| | — JTEXT/
| | — ETEXT/
| | — KTEXT/

以上のディレクトリは、収集時期によって分類した以下のような
下位ディレクトリを持つ。

— | — S9503/

| ;変換主導翻訳システム用形態素タギングデータ

| — TDMT/ — | ;英語形態素
| | — EMOR/ — | — LDB-JE/
| | | — LDB-JEK/
| | | — SLDB/
| |
| | ;日本語形態素
| | — JMOR/ — | — LDB-JE/
| | | — LDB-JEK/
| | | — LDB-JK/
| | | — SLDB/

7 ファイル名

言語データのファイル名は、「会話ID.拡張子」からなっている。^{*1}
7.1節で会話IDについて、7.2節で拡張子について説明する。

7.1 会話ID

言語データベースの会話IDはデータの種類、作成時期により異なり、音声言語データベースの会話ID同様、トピックなどいくつかの特徴要素を並べて定義されている。特徴要素は略号で表現されており、表7-1でトピックとその略号を、表7-2でその他の特徴要素とその略号を示す。次に示すのは、音声言語データベースの会話IDである。

●音声言語データベース

T	A	S	1	2	0	0	4
旅行	トピック	発話種別	会社番号	会話形態	会話番号		

すべての言語データベースの会話IDは、音声言語データベースのものと、区別されている。以下に言語データベースのデータの会話IDを示す。

●英英モノリンガル会話・日日モノリンガル会話

音声言語データベースと同じ会話IDであるが、会話形態の部分がモノリンガル会話では、「0」か「1」で、音声言語データベースは「2」か「3」であることで区別される。

T	A	S	1	0	0	0	1
旅行	トピック	発話種別	会社番号	会話形態	会話番号		

*1 ただし、いくつかのファイルには拡張子がない。

●日英バイリンガル会話

93年下期収集(SOBA93)

音声言語データベースの会話IDから、旅行会話を表わす先頭の'T'と発話種別の特徴要素を削除した形式である。

A	2	3	0	0	1
---	---	---	---	---	---

トピック 会社番号 会話形態 会話番号

94年上期収集(SOBA94A)

音声言語データベースの会話IDから、旅行会話を表わす先頭の'T'と発話種別の特徴要素を削除し、トピック、会話種別、プロット番号の特徴要素を加えた形である。
プロット番号を使うことにより、同じ設定で行われた会話の判別ができるようになっている。

A	T	1	2	0	0	1	1
---	---	---	---	---	---	---	---

トピック 会話種別 会社番号 会話形態 プロット番号 会話番号

94年下期収集(FURU94B)

94年上期収集のデータの会話IDと同じである。

A	F	4	3	0	0	1	1
---	---	---	---	---	---	---	---

トピック 会話種別 会社番号 会話形態 プロット番号 会話番号

94年下期収集/課題遂行型会話(DEN94B)

1	3	0	1	0	4	1	1	.	A	T
---	---	---	---	---	---	---	---	---	---	---

会社番号 会話形態 設定番号 話者番号 場面番号 トピック 会話種別

●日韓バイリンガル会話

94年上期収集の日英バイリンガル会話と同じ会話IDである。会話形態が、日英では"1"か"2"、日韓では"3"か"4"であることで区別される。

A	F	3	4	0	0	2	1
トピック	会話種別	会社番号	会話形態	プロット番号		会話番号	

●日英韓旅行会話表現集

A	4	J	P
トピック	会社番号	オリジナル言語	

日英韓旅行会話表現集の1ファイルの内容は、まとまった会話ではなく、同一トピックの表現の列挙であるため、各文に対して独立した以下のような文IDを付与している。

A	U	4	J	P	0	0	1
トピック	会話種別	会社番号	オリジナル言語		通し番号		

表7-1 トピック

略号	トピック
A	ホテルの部屋の予約、変更、解約（ホテルフロントとの会話）
B	ホテルの紹介（観光案内所との会話）
C	ホテルでのサービス、トラブル（ホテルフロントとの会話）
D	ホテル内施設の利用（ホテルフロントとの会話）
E	現在、使用されていない。
F	空港や港での出入国
G	フライトチケット
H	列車、バス、船などの切符
I	交通手段の問い合わせ
J	道案内
K	レンタカー
L	タクシー乗車、タクシー内での会話
M	現在、使用されていない。
N	現在、使用されていない。
O	ツアー、観光案内
P	観光スポットや娯楽施設での会話、記念写真
Q	現在、使用されていない。
R	演劇、コンサート、スポーツ観戦のチケット
S	レストランの予約
T	料理の注文
U	ホテル以外のトラブル
V	ショッピング
W	郵便
X	現在、使用されていない。
Y	出入金、両替
Z	あいさつ、電話応対

表7-2 会話IDに使われる項目

項目	略号
	モノリンガル・バイリンガル会話、 旅行会話表現集で共通に使用
会社番号	1: 日本アイアール, 2: インターグループ, 3: コングレ, 4: サイマル
	モノリンガル・バイリンガル会話のみで使用
発話種別	C: チェックシート S: チェックシート以外
会話形態	0: 日日, 1: 英英, 2: 日英(米国場面), 3: 英日(日本場面), 4: 日韓(韓国場面), 5: 韓日(日本場面)
プロット番号	会話設定番号で、同一のプロット設定で複数の話者が会 話収録を行っている。
会話種別	F: 対面, T: 非対面(電話)
	バイリンガル会話の課題遂行型のみで使用
設定番号	基本設定の番号(01~04) 詳細は、8節で説明。
話者番号	課題遂行型の主人公話者の番号(01~12)
場面番号	場面の番号(1~5)に通常設定(1)かトラブル設定(6)かの区 別をつけたもの。 詳細は、8節で説明。
	旅行会話表現集のみで使用
オリジナル言語	JP: 日本語, US: 英語, KR: 韓国語 表現を考案する基点となった言語を表わし、例えば、JPであれば、 日本語を基点として表現された文である。
会話種別	F: 対面, T: 非対面(電話), U: 対面、非対面のどちらでも起こりうる表現

7.2 拡張子

拡張子は、ファイルの種類ごとに表7-3のように定義されている。

表7-3 拡張子

ファイルの種類	拡張子	ファイル名の例
日英対応テキスト	euc, TXT, EJTEXT	A23001.euc, AF430011.TXT, 12030116.VF.EJTEXT
日韓対応テキスト	JKTEXT	AF340021.JKTEXT
日英韓対応テキスト	拡張子なし	A4JP
英語テキスト(模擬会話集)	ETEXT	A23001.ETEXT
(旅行会話基本表現集)	拡張子なし	E-A4JP
日本語テキスト(模擬会話集)	JTEXT	A23001.JTEXT
(旅行会話基本表現集)	拡張子なし	J-A4JP
韓国語テキスト(模擬会話集)	KTEXT	AF340021.KTEXT
(旅行会話基本表現集)	拡張子なし	K-A4JP
日本語形態素タギング	JMOR	A23001.JMOR
英語形態素タギング	EMORT	AT120012.EMORT
韓国語形態素タギング	KMOR	AF340021.KMOR

8 課題遂行型会話

課題遂行型会話は、主たる手段として達成されるようなタスク(ホテルの予約など)を話者に与え、自発的に会話を構成しながらタスクを遂行する過程を会話収集したものであり、日英バイリンガル会話の一部が担当する。

課題遂行型会話では、以下の表8-1、表8-2に示すようないくつかの場面からなる基本設定のもとで旅行するような状況を考え、旅行者と対応者(ホテルのフロントなど)に必要な情報(ホテルの対応方法など)を図8-1のような資料として与えた。

1つの基本設定(ブロードウェイでミュージカルを見るなど)に対して、5場面設ける。1人の話者(旅行者)は、5つの場면을必ず経験する。ただし、それぞれの場面はトラブル場面(チケット売り切れなど)となる可能性がある。1つの基本設定に12人の話者が旅行者として会話を行なう。対応者は、それぞれの場面で固定されている。各場面に対し、12人の旅行者の内、2人がトラブル場面に遭遇し、各旅行者は、1つの基本設定の中で、トラブル場面は0か1回しか経験しないように設定している。従って、まったくトラブル場面に遭遇しない旅行者は2人であり、10人は5つの場面の内、1つのトラブル場面に遭遇する。この方式により、12基本設定×5場面×12話者で720会話の収集を行なった。ただし、13人以上で会話を行なった場合があるため、749ファイルの会話を収集している。

以下、課題遂行型会話の会話IDについて説明する。

例えば、「ブロードウェイでミュージカルを見る」という基本設定の中での会話には、次のような会話IDが与えられる。話者、場面番号の"#"は、数字1文字を、トピック、会話種別は、アルファベット1文字を示す。

1	2	0	3	#	#	#	#	.	#	#
会社番号	会話形態	設定番号		話者番号		場面番号			トピック	会話種別

表8-1は、トラブルでない通常場面を、表8-2は、通常場面对応するトラブル場面を示す。場面番号の数字には、トラブルでない通常場面对しては、"#1"を、トラブル場面对しては、"#6"を与えている。話者番号は、"01~12"の数字により話者IDを表わす。従って、各場面の会話IDは、以下のように与えられる。

場面番号	通常場面	トラブル場面
場面1	1203##11.VF	1203##16.VF
場面2	1203##21.RT	1203##26.RF
場面3	1203##31.JT	1203##36.UF
場面4	1203##41.TF	1203##46.TF
場面5	1203##51.CF	1203##56.CF

例えば、場面1でトラブルに遭遇する旅行者01が経験する場面の会話IDは、以下のようになる。

場面1 (トラブル) 12030116.VF
 場面2 12030121.RT
 場面3 12030131.JT
 場面4 12030141.TF
 場面5 12030151.CF

表8-1 通常場面

	目的	行動	場所	対応者	種別
1	公演の確認	公演情報雑誌の購入	ニューススタンド	店員	対面
2	ミュージカルを見る	チケットを予約する	ホテル	劇場窓口担当者	電話
3	劇場へ行く	劇場への行き方を聞く	ホテル	劇場窓口担当者	電話
4	夕食をとる	料理をオーダーする	レストラン	ウェイター	対面
5	帰る	チェックアウトの手続きをする	ホテル	フロント	対面

表8-2 トラブル場面

	目的	行動	場所	対応者	種別
1	公演の確認	買った情報雑誌に落丁があり、欲しい情報が得られない	ニューススタンド	店員	対面
2	ミュージカルを見る	チケットの予約番号の控えを忘れてきた	劇場	劇場窓口担当者	対面
3	劇場へ行く	劇場に手帳を忘れてきた	劇場	劇場窓口担当者	対面
4	夕食をとる	注文したものでない料理が届いた	レストラン	ウェイター	対面
5	帰る	出発日の朝8時に仕上がっているはずの写真がまだできていない	ホテル	フロント	対面

その他の基本設定の内容、会話IDについては、付録Bに示す。

図8-1 旅行者の資料

目的：ディズニーランドへ行く

行動：

1. 1995年1月20日（金）～1月27日（金）にかけてロサンゼルスへ友人と3人で訪れる。
 2. 宿泊については、ディズニーランド・パーク・ホテルに6泊する。
 3. いつから、何日間ディズニーランドで遊ぶか決めておく。
 4. どうやってディズニーランドへ行くか確認する。
 5. ディズニーランドの入場チケットを買う。
 6. ディズニーランドでの食事はあらかじめ何を食べるか決めておく。
 7. ディズニーランドではぐれたときの集合場所を決める。
 8. ディズニーキャラクターの縫いぐるみを、お土産として買う。
- * 3、6、7については、旅行に行く前に決めておいてください。

条件：

1. 旅行について
行き = 1月20日（金） JAL26便 午後大阪発、午前ロス着
帰り = 1月26日（木） JAL35便 午後ロス発、27日（金）午後大阪着
に乗らなければならない。
2. 旅行費用について
旅行費用は、航空運賃、宿泊費を除いて1人400米ドルとする。超過は不可。
3. 旅行者について
友人と3人でロサンゼルスを訪れるということに関しては、変更不可。
4. ホテルについて
ディズニーランド・パークホテルに宿泊する。滞在中変更は不可。
チェックイン日 1月20日（金） 3:00PM
チェックアウト日 1月26日（木） 12:00PM
5. ディズニーランド・パークについて
はぐれた時の待ち合わせ場所は、アトラクションのなかから選ぶ。
6. お土産について
・ 家族にディズニーキャラクターの縫いぐるみを買ってきて欲しいと頼まれている。
・ お土産は手で持って帰る。
7. その他、日本から旅行会社を通して、次のオプションツアーに申し込んである。
・ 1月23日（月）ユニバーサルスタジオ見学ツアー参加。
・ 1月25日（水）午後8時、ホテル内レストラン（MARIE ET EMILE）にて食事。

9 今後の予定

今後、拡張を予定している言語データは現在次のとおりである。

- (1) 日独バイリンガル会話の著作 [1995年12月頃]
- (2) 旅行会話表現集の拡張
 - a. 日英韓旅行会話表現集の拡張 [1995年11月頃]
 - b. ドイツ語を含めた旅行会話表現集の作成
- (3) 日本語話者同士の、実対話及び相互主導の課題遂行型会話の収集
[1996年3月頃]

9.1. 日独バイリンガル会話

9.1.1 会話著作の方針

言語データベースの日英バイリンガル会話、72会話（1会話平均30文）を編集して日独バイリンガル会話を作成する。日本語部分は原文をできるだけ維持し、英語部分をドイツ語に置き換えることを基本にして、日独バイリンガル会話の編集を行なう。ただし、できあがった会話が全体として、日本人の会話とドイツ人の会話として自然なものとなるかをチェックし、元の会話の内容では日独バイリンガル会話として不自然になる場合は、会話構成の変更を行なう。

9.1.2 会話ID

原作となる日英バイリンガル会話のIDに、拡張子として"JG"を加えたものを、日独バイリンガル会話のIDとする。例えば、オリジナルの日英会話が"TAS12008"であれば、編集作成した日独会話は"TAS12008.JG"とする。

9.2 旅行会話表現集の拡張

1995年11月頃、日英韓旅行会話表現集を6000文追加する予定である。また、ドイツ語を含めた（日英独または日英韓独）旅行会話表現集の作成を計画している。

9.3 日本語話者同士の実対話及び相互主導の課題遂行型対話の収集

1995年度下期に、対話処理の研究に利用することに重点をおいた日本語話者同士の実対話及び相互主導の課題遂行型対話の収集を計画している。これらのデータの収集に際しては、発声時間、割り込み禁止、通訳の介在など、これまでの模擬会話収集で課した音声認識処理および言語翻訳処理のための制約のほとんどを排し、話者がより自由にかつ自然に会話できるようにする。

トピックは現在まだ確定していないが、実対話の対話課題に関しては、例えば、旅行・観光案内や受付案内の対話のように、場面が固定されていて、現実的に収集可能なものを調査している。

相互主導の課題遂行型対話に関しては、経路探索課題などの課題について小規模な実験を行っており、必然的に相互主導になる課題を検討している。

謝辞

言語データベースの収集、作成、整理にご協力いただきました、浦谷主幹研究員(現在、NHK)、隅田主任研究員、石崎主任研究員、竹澤主任研究員、側嶋研究員(現在、CSK)、赤峯研究員(現在、日本電気)、伝研究員、柏岡研究員、巖寺研究員、田代研究員、東洋情報システムの森田さんに感謝いたします。

参考文献

- [1] 浦谷則好, 竹沢寿幸, 松尾秀彦, 森田千帆 : "音声言語データベースの構成" ATR Technical Report TR-IT-0056
- [2] 浦谷則好, 田代敏久, 山田久子, 松本香 : "音声言語データベースにおける日本語形態素解析マニュアル" ATR Technical Report TR-IT-0009

付録A ディレクトリ構成とファイル数

形態素タギングファイルに関しては、現在拡張中であるため、ファイル数をここでは記述していない。

●モノリンガル会話

英英モノリンガル会話

/DB/LDB/EE/DELIVERY/9312/ 会話 I D.euc 91ファイル

日日モノリンガル会話

/DB/LDB/JJ/DELIVERY/9312/ 会話 I D.euc 90ファイル

●バイリンガル会話

日英バイリンガル会話

/DB/LDB/JE/—	— ENV/ —				
		—SOBA93/			
		—SOBA94A/	会話 I D.ENV	360	ファイル
		—FURU94B/	会話 I D.ENV	60	ファイル
		—DEN94B/	会話 I D.ENV	749	ファイル ;課題遂行型
	— EJTEXT/—				
		—SOBA93/	会話 I D.euc	104	ファイル
		—SOBA94A/	会話 I D.euc	360	ファイル
		—FURU94B/	会話 I D.TXT	60	ファイル
		—DEN94B/	会話 I D.EJTEXT	749	ファイル ;課題遂行型
	— EJTEXT+/—				
		—SOBA93/	会話 I D.EJTEXT	104	ファイル
		—SOBA94A/	会話 I D.EJTEXT	283	ファイル
			(360ファイルに拡張予定)		
		—FURU94B/	会話 I D.ETEXT	60	ファイル
		—DEN94B/	会話 I D.EJTEXT	749	ファイル ;課題遂行型

(次項へ)

(前項より)

/DB/LDB/JE/—	— ETEXT/ —			
		— SOBA93/	会話 I D.ETEXT	1 0 4 ファイル
		— SOBA94A/	会話 I D.ETEXT	3 6 0 ファイル
		— FURU94B/	会話 I D.TXT	6 0 ファイル
		— DEN94B/	会話 I D.ETEXT	7 4 9 ファイル ;課題遂行型
		— ETEXT+/ —		
		— SOBA93/	会話 I D.ETEXT	1 0 4 ファイル
		— SOBA94A/	会話 I D.ETEXT	2 8 3 ファイル (3 6 0 ファイルに拡張予定)
		— FURU94B/	会話 I D.ETEXT	6 0 ファイル
		— DEN94B/	会話 I D.ETEXT	7 4 9 ファイル ;課題遂行型
		— JTEXT/ —		
		— SOBA93/	会話 I D.JTEXT	1 0 4 ファイル
		— SOBA94A/	会話 I D.JTEXT	3 6 0 ファイル
		— FURU94B/	会話 I D.JTEXT	6 0 ファイル
		— DEN94B/	会話 I D.JTEXT	7 4 9 ファイル ;課題遂行型
		— EMOR/ —		
		— SOBA93/	会話 I D.EMOR	
		— SOBA94A/	会話 I D.EMOR	
		— FURU94B/	会話 I D.EMOR	
		— DEN94B/	会話 I D.EMOR	
		— JMOR/		

日韓バイリンガル会話

/DB/LDB/JK/—	— ENV/ —			
		— C9410/	会話 I D.ENV	6 4 ファイル
		— C9502/	会話 I D.ENV	6 0 ファイル
		— S9502/	会話 I D.ENV	6 0 ファイル
		— JKTEXT/ —		
		— C9410/	会話 I D.JKTEXT	6 4 ファイル
		— C9502/	会話 I D.JKTEXT	6 0 ファイル
		— S9502/	会話 I D.JKTEXT	6 0 ファイル
		— KTEXT/ —		
		— C9410/	会話 I D.KTEXT	6 4 ファイル
		— C9502/	会話 I D.KTEXT	6 0 ファイル
		— S9502/	会話 I D.KTEXT	6 0 ファイル

(次項へ)

(前項より)

/DB/LDB/JK/—	—	JTEXT/ —			
		—	C9410/	会話 I D.JTEXT	64 ファイル
		—	C9502/	会話 I D.JTEXT	60 ファイル
		—	S9502/	会話 I D.JTEXT	60 ファイル

●旅行会話表現集

日英韓旅行会話表現集

/DB/LDB/JEK/ —	—	JEKTEXT/ —	—	S9503/	会話 I D	63 ファイル
	—	JTEXT/ —	—	S9503/	J-会話 I D	63 ファイル
	—	ETEXT/ —	—	S9503/	E-会話 I D	63 ファイル
	—	KTEXT/ —	—	S9503/	K-会話 I D	63 ファイル

●変換主導翻訳システム体系形態素タギングデータ

このデータは現在拡張中でファイル数は増加している。

/DB/LDB/TDMT/—	—	EMOR/ —			
		—	LDB-JE/	会話 I D.EMORT	
		—	LDB-JEK/	会話 I D.EMORT	
		—	SLDB/	会話 I D.EMORT	
	—	JMOR/ —			
		—	LDB-JE/	会話 I D.JMOR	
		—	LDB-JEK/	会話 I D.JMOR	
		—	LDB-JK/	会話 I D.JMOR	
		—	SLDB/	会話 I D.JMOR	
	—	KMOR/ —			
		—	C9410/	会話 I D.KMOR	
		—	C9502/	会話 I D.KMOR	
		—	S9502/	会話 I D.KMOR	

●関連データベースの種類とそのファイル構成

/DB/SLDB/SPH/ ;以下は音声言語データベースにおける音声関係データが
収められている。

/DB/SLDB/LNG/ ;以下は音声言語データベースにおける言語関係データが
収められている。

音声言語データベースにおける言語関係データの詳細は以下のとおりである。

日英対応テキスト

/DB/SLDB/LNG/EJTEXT/	—	—	9312/	会話 I D.euc	1 4 4	ファイル
			9403/	会話 I D.euc	1 5 8	ファイル
			9405/	会話 I D.euc	6	ファイル
			9406/	会話 I D.euc	1 0 0	ファイル
			9409/	会話 I D.euc	2 1 0	ファイル

英文修正情報付日英対応テキスト

/DB/SLDB/LNG/EJTEXT+/	—	—	9312/	会話 I D.EJTEXT	1 4 4	ファイル
			9403/	会話 I D.EJTEXT	1 5 8	ファイル
			9405/	会話 I D.EJTEXT	6	ファイル
			9406/	会話 I D.EJTEXT	1 0 0	ファイル
			9409/	会話 I D.EJTEXT	2 1 0	ファイル

英語テキスト

/DB/SLDB/LNG/ETEXT/ 会話 I D.ETEXT 6 1 8 ファイル

英語修正版テキスト

/DB/SLDB/LNG/ETEXT+/	—	—	9312/	会話 I D.ETEXT	1 4 4	ファイル
			9403/	会話 I D.ETEXT	1 5 8	ファイル
			9405/	会話 I D.ETEXT	6	ファイル
			9406/	会話 I D.ETEXT	1 0 0	ファイル
			9409/	会話 I D.ETEXT	2 1 0	ファイル

日本語テキスト

/DB/SLDB/LNG/JTEXT/ 会話 I D.JTEXT 618ファイル

英語形態素タギング

/DB/SLDB/LNG/EMOR/ 会話 I D.EMOR

日本語形態素タギング

/DB/SLDB/LNG/JMOR/ 会話 I D.JMOR

付録B 課題遂行型会話の基本設定と会話ID

会話の設定は、作成会社ごとに異なっている。

基本設定

会話IDの"#"は、数字かアルファベット1文字を示す。

1つの基本設定に対して60個の会話IDが与えられている。

日本アイアール

設定内容	会話ID
出張先でランチミーティングを開く	1301####.##
三十三間堂に行く	1302####.##
ブロードウェイでミュージカルを見る	1203####.##
ニューヨーク近代美術館に行く	1204####.##

インターグループ

設定内容	会話ID
寺社巡りバスツアーに参加する	2301####.##
精進料理を食べる	2302####.##
ゴルフをする	2203####.##
ディズニーランドへ行く	2204####.##

コングレ

設定内容	会話ID
京都で歌舞伎をみる	3301####.##
嵐山で紅葉を楽しむ	3302####.##
ハイキング旅行	3203####.##
友人訪問	3204####.##

日本アイアール

基本設定1 出張先でランチミーティングを開く

[通常場面]

	目的	行動	場所	対応者	種別
1	宿泊する	チェックインの手続きをする	ホテル	フロント	対面
2	ミーティング資料を作成する	ビジネスセンターの利用日時を変更する	ホテル	クラーク	対面
3	支社からの連絡を受ける	連絡手段を確保する	ホテル	フロント	対面
4	ランチミーティング準備をする	レストランに席を予約する	ホテル内	予約係	電話
5	ファックス送信の確認をする	取引先の会社に電話をする	ホテル内	会社社員	電話

[トラブル場面]

	目的	行動	場所	対応者	種別
1	宿泊する	部屋のタイプが予約したものと違う	ホテル	フロント	対面
2	ミーティング資料を作成する	コピーの色が薄い	ビジネスセンター	クラーク	対面
3	支社からの連絡を受ける	ポケットベルを落として壊してしまった	ホテル	フロント	対面
4	ランチミーティングをする	予約した席でない席が用意されていた	レストラン	案内係	対面
5	ファックス送信の確認をする	送ったファックスが薄く、読みにくいというメッセージを受けた	ホテル内	会社社員	電話

会話 I D

	通常場面	トラブル場面
場面1	1301##11.CF	1301##16.CF
場面2	1301##21.ZF	1301##26.ZF
場面3	1301##31.CF	1301##36.CF
場面4	1301##41.ST	1301##46.SF
場面5	1301##51.ZT	1301##56.ZT

基本設定2 三十三間堂に行く

[通常場面]

	目的	行動	場所	応対者	種別
1	宿泊する	バック旅行の予約をする	滞在中の ホテル	予約係	電話
2	自転車を借りる	レンタルの手続きを行う	レンタサイ クル店	店員	対面
3	三十三間堂に行く	行き方を尋ねる	ホテル	寺院職員	電話
4	土産を買う	品物を選ぶ	寺院売店	店員	対面
5	葉書を出す	切手を買う	郵便局	郵便局員	対面

[トラブル場面]

	目的	行動	場所	応対者	種別
1	宿泊する	予約したバックのホテルが客 の希望したものと違っていた	旅行会社	予約係	対面
2	自転車を借りる	自転車がパンクした	レンタサイ クル店	店員	対面
3	三十三間堂に行く	途中で道に迷う	東山 郵便局前	寺院職員	電話
4	土産を買う	買った絵葉書が破れていた	寺院売店	店員	対面
5	葉書を出す	切手を張らずに投函してしま った	郵便局	郵便局員	対面

会話 I D

	通常場面	トラブル場面
場面 1	1302##11.OF	1302##16.OT
場面 2	1302##21.ZF	1302##26.ZF
場面 3	1302##31.JT	1302##36.JT
場面 4	1302##41.VF	1302##46.VF
場面 5	1302##51.ZF	1302##56.ZF

基本設定3 ブロードウェイでミュージカルを見る

[通常場面]

	目的	行動	場所	対応者	種別
1	公演の確認	公演情報雑誌の購入	ニューススタンド	店員	対面
2	ミュージカルを見る	チケットを予約する	ホテル	劇場窓口担当者	電話
3	劇場へ行く	劇場への行き方を聞く	ホテル	劇場窓口担当者	電話
4	夕食をとる	料理をオーダーする	レストラン	ウェイター	対面
5	帰る	チェックアウトの手続きをする	ホテル	フロント	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	公演の確認	買った情報雑誌に落丁があり、欲しい情報が得られない	ニューススタンド	店員	対面
2	ミュージカルを見る	チケットの予約番号の控えを忘れてきた	劇場	劇場窓口担当者	対面
3	劇場へ行く	劇場に手帳を忘れてきた	劇場	劇場窓口担当者	対面
4	夕食をとる	注文したものでない料理が届いた	レストラン	ウェイター	対面
5	帰る	出発日の朝8時に仕上がっているはずの写真がまだできていない	ホテル	フロント	対面

会話ID

	通常場面	トラブル場面
場面1	1203##11.VF	1203##16.VF
場面2	1203##21.RT	1203##26.RF
場面3	1203##31.JT	1203##36.UF
場面4	1203##41.TF	1203##46.TF
場面5	1203##51.CF	1203##56.CF

基本設定 4 ニューヨーク近代美術館に行く

[通常場面]

	目的	行動	場所	対応者	種別
1	宿泊する	宿泊の予約をする	飛行場	ホテル 宿泊係	電話
2	美術館で絵を見る	館内案内図を入手する	美術館内	インフォメ ーション 対応者	対面
3	昼食をとる	軽食を注文する	美術館内 カフェ	ウェイター	対面
4	美術書を買う	品物を選ぶ	美術館売店	販売係	対面
5	書籍を日本に送る	郵送を依頼する	美術館売店	配送係	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	宿泊する	部屋のタイプが電話予約した ものと違う	ホテル	ホテル 宿泊係	対面
2	美術館で絵を見る	美術館で迷う	美術館内	係員	対面
3	昼食をとる	食事の中に何か固い異物が入 っている	美術館内 カフェ	ウェイター	対面
4	美術書を買う	買うはずの本が売り切れてい た	美術館売店	店員	対面
5	書籍を日本に送る	一部の書籍の送り先を間違え て書いてしまった	美術館売店	配送係	対面

会話 I D

	通常場面	トラブル場面
場面 1	1204##11.AT	1204##16.CF
場面 2	1204##21.ZF	1204##26.ZF
場面 3	1204##31.TF	1204##36.ZF
場面 4	1204##41.VF	1204##46.VF
場面 5	1204##51.ZF	1204##56.ZF

インターグループ

基本設定 1 寺社巡りバスツアーに参加する

[通常場面]

	目的	行動	場所	対応者	種別
1	宿泊する	チェックインする	ホテル	フロント	対面
2	寺社巡りをする	ツアーに申し込む	旅行会社	予約係	対面
3	夕食をとる	寿司屋への行き方を尋ねる	寿司屋	店員	電話
4	お土産を買う	図録について問う	本屋	店員	対面
5	帰国する	空港への交通手段を確保する	ホテル	フロント	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	宿泊する	希望のタイプでない	ホテル	フロント	対面
2	寺社巡りをする	予約した日時に行けなくなった	旅行会社	予約係	対面
3	夕食をとる	時間に間に合わない	寿司屋	店員	電話
4	お土産を買う	メモの内容が間違っている	本屋	店員	対面
5	帰国する	電車の予約がとれない	ホテル	フロント	対面

会話 I D

	通常場面	トラブル場面
場面 1	2301##11.CF	2301##16.CF
場面 2	2301##21.OF	2301##26.OF
場面 3	2301##31.JT	2301##36.JT
場面 4	2301##41.VF	2301##46.VF
場面 5	2301##51.IF	2301##56.IF

基本設定2 精進料理を食べる

[通常場面]

	目的	行動	場所	対応者	種別
1	宿泊する	ホテルを予約する	ホテル	フロント	電話
2	精進料理を食べる	店に予約を入れる	お店	店員	電話
3	お店へ移動する	交通手段を確認する	ホテル	フロント	対面
4	近辺を散策する	観光名所の紹介	お店	店員	対面
5	お土産を買う	品物を選ぶ	お店	店員	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	宿泊する	その日の予約リストに名前がない	ホテル	フロント	対面
2	精進料理を食べる	予約人数を二人で取ってしまったので二人席しか確保できない	お店	店員	対面
3	お店へ移動する	京都マラソンで交通規制されておりバスが動いていない	ホテル	フロント	対面
4	近辺を散策する	お寺が改築中で入れない	お店	店員	対面
5	お土産を買う	購入した品物の製造年月日が古い	お店	店員	電話

会話 | D

	通常場面	トラブル場面
場面 1	2302##11.AT	2302##16.AF
場面 2	2302##21.ST	2302##26.SF
場面 3	2302##31.IF	2302##36.IF
場面 4	2302##41.OF	2302##46.OF
場面 5	2302##51.VF	2302##56.VT

基本設定3 ゴルフをする

[通常場面]

	目的	行動	場所	対応者	種別
1	ゴルフをする	予約する	日本	ゴルフ場プロショップ 受付	電話
2	ゴルフ場に行く	レンタカーを借りる	レンタカー店	店員	対面
3	ゴルフをする	ゴルフ道具を借りる	ゴルフ場 レンタル カウンター	受付	対面
4	食事をする	料理を注文する	ゴルフ場 レストラン	ウエイター	対面
5	お土産を購入する	品物を選ぶ	ゴルフ場 売店	店員	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	ゴルフをする	予約時間が間違っていた	ゴルフ場プロショップ	受付	対面
2	車に乗る	エンジンの調子がおかしい	レンタカー店	店員	対面
3	ゴルフをする	借りたセットが合わない	ゴルフ場 レンタル カウンター	受付	対面
4	食事をする	注文と違うものがでてきた	ゴルフ場 レストラン	ウエイター	対面
5	お土産を購入する	購入した品物が入っていない	ゴルフ場 売店	店員	対面

会話 I D

	通常場面	トラブル場面
場面 1	2203##11.ZT	2203##16.ZF
場面 2	2203##21.KF	2203##26.KF
場面 3	2203##31.ZF	2203##36.ZF
場面 4	2203##41.TF	2203##46.TF
場面 5	2203##51.VF	2203##56.VF

基本設定4 ディズニーランドへ行く

[通常場面]

	目的	行動	場所	対応者	種別
1	ホテルへ行く	チェックインする	ホテル	フロント	対面
2	ディズニーランドへ移動する	交通手段を確認する	部屋	フロント	電話
3	ディズニーランドへ入場する	チケットを購入する	パーク	チケット販売員	対面
4	食事をとる	飲食店のお問い合わせ	パーク	インフォメーション係	対面
5	お土産を買う	品物を選ぶ	売店	店員	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	ホテルへ行く	チェックアウト日が間違っていた	ホテル	フロント	対面
2	ディズニーランドへ移動する	整備中でモノレールが動かない	ホテル	フロント	電話
3	ディズニーランドへ入場する	入場規制されておりチケットを買えない	パーク	チケット販売員	対面
4	食事をとる	友人とはぐれてしまった	パーク	インフォメーション係	対面
5	お土産を買う	買った品物が汚れていた	売店	店員	対面

会話 I D

	通常場面	トラブル場面
場面1	2204##11.CF	2204##16.CF
場面2	2204##21.IT	2204##26.IT
場面3	2204##31.RF	2204##36.RF
場面4	2204##41.ZF	2204##46.ZF
場面5	2204##51.VF	2204##56.VF

コングレ

基本設定1 京都で歌舞伎をみる

[通常場面]

	目的	行動	場所	対応者	種別
1	歌舞伎を観る	劇場窓口に予約を入れる	ホテル	チケット係	電話
2	宿泊する	部屋を予約する	ホテル	宿泊予約係	電話
3	歌舞伎を観る	通訳イヤホンを借りる	受付	受付・案内係	対面
4	食事をする	食堂の席を選ぶ	食堂	受付係	対面
5	お土産を買う	品物を選ぶ	売店	店員	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	歌舞伎を観る	チケットの枚数が違う	劇場	チケット係	対面
2	宿泊する	予約した部屋とタイプが違っていた	ホテル	フロント係	対面
3	歌舞伎を観る	通訳イヤホンの音がでない	劇場	受付・案内係	対面
4	食事をする	料理がなかなか運ばれてこない	食堂	店員	対面
5	お土産を買う	買った品物を店に置き忘れた	ホテル	劇場売店店員	電話

会話ID

	通常場面	トラブル場面
場面1	3301##11.RT	3301##16.RF
場面2	3301##21.AT	3301##26.AF
場面3	3301##31.PF	3301##36.PF
場面4	3301##41.SF	3301##46.SF
場面5	3301##51.VF	3301##56.VT

基本設定2 京都で歌舞伎をみる

[通常場面]

	目的	行動	場所	対応者	種別
1	宿泊する	宿泊先情報を入手	ツーリスト インフォメ ーション	係員	電話
2	紅葉を楽しむ	地図を入手	旅館/民宿	フロント係	対面
3	昼食を食べる	注文をする	料亭	接客係	対面
4	保津峡の渓谷美を楽しむ	トロッコ列車の切符を買う	トロッコ 嵐山駅 切符売り場	窓口係	対面
5	宿泊先に戻る	トロッコ亀岡駅からJR馬堀駅 へ行き方を聞く	トロッコ 亀岡駅	駅にいる人	対面

[トラブル場面]

	目的	行動	場所	対応者	種別
1	宿泊する	和室を依頼したが洋室だった	旅館/民宿	フロント係	対面
2	紅葉を楽しむ	地図が品切れ	旅館/民宿	フロント係	対面
3	昼食を食べる	オーダーと違うものが配膳さ れた	料亭	接客係	対面
4	保津峡の渓谷美を楽しむ	列車の時間が違う	トロッコ 嵐山駅 切符売り場	窓口係	対面
5	宿泊先に戻る	道に迷う	路上	道で会っ た人、 タバコ屋	対面

会話 I D

	通常場面	トラブル場面
場面 1	3302##11.BT	3302##16.CF
場面 2	3302##21.CF	3302##26.CF
場面 3	3302##31.TF	3302##36.UF
場面 4	3302##41.HF	3302##46.UF
場面 5	3302##51.JF	3302##56.JF

基本設定 3 ハイキング旅行

[通常場面]

	目的	行動	場所	応対者	種別
1	宿泊する	予約をする	自宅	宿泊予約係	電話
2	国立公園へ行く	レンタカーを借りる	レンタカー店	受付	対面
3	アウトドア用品店へ行く	店を教えてもらう	ホテル	フロント係	対面
4	必要品を購入する	品物を選ぶ	アウトドア用品店	店員	対面
5	ハイキングをする	ルートを尋ねる	国立公園 ロッジ	フロント係	対面

[トラブル場面]

	目的	行動	場所	応対者	種別
1	宿泊する	予約内容が違う	国立公園 ロッジ	フロント係	対面
2	国立公園へ行く	レンタカーを返却する	サンフランシスコ	受付	対面
3	アウトドア用品店へ行く	店の住所が変わっていた	公衆電話	店員	電話
4	必要品を購入する	水筒の中身が漏れる	アウトドア用品店	店員	電話
5	ハイキングをする	植物にかぶれた	国立公園 ロッジ	フロント係	対面

会話 I D

	通常場面	トラブル場面
場面 1	3203##11.AT	3203##16.CF
場面 2	3203##21.KF	3203##26.KF
場面 3	3203##31.CF	3203##36.JT
場面 4	3203##41.VF	3203##46.UT
場面 5	3203##51.CF	3203##56.CF