

地図課題コーパス (中間報告)

The Japanese map task corpus: an interim report

青野元子, 市川薫, 小磯花絵, 佐藤伸二, 仲真紀子, 土屋俊,
 八木健司, 渡部直也 (千葉大学), 石崎雅人 (ATR 音声翻訳通信研究所)
 岡田美智男 (NTT 基礎研究所), 鈴木浩之 (松下電器)
 中野有紀子 (NTT 情報通信研究所), 野中慶子 (大東文化大学)
 Motoko Aono, Akira Ichikawa, Hanae Koiso, Shinji Sato, Makiko Naka,
 Syun Tutiya, Kenji Yagi, Naoya Watanabe (Chiba University)
 Masato Ishizaki (ATR Interpreting Telecommunications Research Laboratories)
 Michio Okada (NTT Basic Research Laboratories)
 Hiroyuki Suzuki (Multimedia Systems Laboratory, Matsushita Electric Co.)
 Yukiko Nakano (NTT Information and Communication Systems Laboratories)
 Keiko Nonaka (Daito Bunka University)

This paper describes the Japanese Map Task Corpus project at Chiba University. The project aims to build a linguistically and acoustically reliable and cognitively significant corpus of dialogs conducted in Japanese. The basic design follows that of the Edinburgh map task corpus with respect to map and route designs and situational parameters such as familiarity and eye contact. The feature names are carefully selected to highlight linguistically and phonetically interesting phenomena in spontaneous oral discourses in Japanese. Enhancement is also made in recording non-verbal behavior of participants, including the recording of the drawing hand on the map. The corpus will consist of the separately taped 2 channel audio data coupled with a fully tagged orthographical transcription.

1 序論

信頼できる音声対話資料の作成は、言語学、音声学、心理学、会話分析、文化人類学などの基礎研究的側面、および、音声処理、インターフェイス設計、言語処理などの工学的側面のいずれからも求められているが、すくなくとも日本語音声対話については、依然として十分に努力がなされていない。現在までに試みられたものは、品質のために自然さを欠かざるを得なくなったり、また、自然さのために品質を犠牲にせざるを得ないというような制約を受けている。とくに、自然で自発的な対話の収録については本格的な試みがなかった。

本論文で報告するプロジェクトは、可能なかぎり「自然で自発的な」条件のもので、可能なかぎり品質のよい音声の収録を目的として、エディンバラ大学で行なわれた地図課題コーパス作成 [Anderson 1991] を踏襲しつつ、いくつかの点について改良した音声対話収録を実現するものである。今回の報告では、基本的な理念および収録計画と収録の実際について記述する。この収録の結果のついての基本的分析は別の機会に報告する。また、今回の収録では、エディンバラ大学における課題の設定を考慮して、日本語

と英語の音声対話について比較研究が可能であるように配慮した。

2 目的と意義

自然で自発的な発話を収録するための方法として地図課題を採用したのは以下の理由による。

1. 実験参加者が課題の解決に集中することによって、発話自体を目的しないので、課題としては人工的であるが、自然な発話が得られる。
2. 一定の時間継続して、一定の時間内に終了することによって異なる条件を設定して反復収録することができる。
3. 特徴名を適当に選べば、実験参加者に要求される知的背景の影響を少なくすることができる。

以上のような特徴をもつ地図課題による対話の収録によって、少なくとも以下のような研究関心に応えることが可能となるであろう。

1. 自然な発話における日本語音声の変異について十分な資料を得ることができる。
2. 親近性および相互視認の有無が協調行動において果す役割について分析することが可能になる。
3. コミュニケーションの方略を厳密に分析できる。
4. 文脈的な情報が発話およびその理解にたいして持つ役割を分析できる。

しかし、そのような関心に限定される必要はなく、この収録の成果が公開されることによってさらに広範囲の研究関心に応えることが可能になるであろう。またとくに、日本語の言語コーパス自体が、書き言葉を含めても今後の課題であるという現状を踏まえるならば、品詞タグのカテゴリ、統語論タグの組織などの検討のためにこのような音声対話コーパスの作成の貢献は大きいであろう。さらに、本プロジェクトでは、ビデオによる顔および地図(手の動作)の収録をとくに重視して設計したので、非言語的手段による情報伝達の様態を従来よりも詳細に分析することが可能になるであろう。

地図課題は、音声対話のひとつのカテゴリにすぎないので、さらに他の種類の課題(たとえば、共同達成、相談、質疑応答など)について同様の統制を行なった収録を続け、比較研究することが望ましい。しかし、統制の条件などについても本プロジェクトの試みは参考になると思われる。

3 今回の収録計画

地図課題は、二人の実験参加者がそれぞれ地図を持ち、情報提供者に描かれているルートを情報追隨者の地図に再現するというものである(地図については、図2参照)。情報提供者の地図には、出発地点と目標地点、そこに至るまでの経路、そして幾つかの目標物が描かれている。一方、情報追隨者の地図には、出発地点と目標物のみ描かれており、目標地点と経路は描かれていない。ここでは、まず地図や目標物といった材料に関して、次に被験者条件に関して、最後に実験デザインに関して概観する。

3.1 今回の収録計画の特徴

今回の収録においては、比較のためにエディンバラ大学の計画を考慮したという点に加えて、とくに非言語的コミュニケーション手段の使用、対話理解における動作の役割などを記録するために、音声だけでなく、顔の表情、向き、地図上の手の動きを音声と同期させて記録した。また、若干の音韻的特徴に注目して、同一単語をさまざまな条件で発話するように構成されている。

3.2 材料

3.2.1 音韻的特徴

目標物の中には、主要目標物と呼ばれるものは必ずある特定の音韻的特徴を有する。またその他の目標物に関して

も、音韻的特徴、とくに簡略化、脱落などの傾向を有する場合もある。ここでは、まずこれらの音韻的特徴に関して触れる。

主要目標物の音韻的特徴 主要目標物は、次の4つの音韻的变化のうち、いずれかの特徴を備えており、それぞれ二つずつ存在する。これらの四つの音韻的变化を「母音脱落」「拗音脱落」「鼻音化」「長母音化」としたが、それぞれ次の目標名が割り当てられている。(この変化の分類は、かならずしも網羅的なものではない。)

表-1

	音韻的变化	主要目標物
1	母音脱落	北の沼 / 南の沼
2	拗音脱落	魔術山 / 呪術山
3	鼻音化	銀鉢 / 金鉢
4	長母音化	西北の湾 / 東南の湾

その他の目標物に関する音韻的、韻律的特徴 その他の目標物に関しては、音韻的变化に関わらず、音韻的に関心のあるものを選んだ。例えば「母音脱落」「拗音脱落」「鼻音化」「長母音化」「二重母音化(母音連鎖)」「撥音化」「無声化」「連濁現象」などが関心の対象となった。また、複合語におけるアクセント移動が、自発的な発話においてどのように現われるかを見るために、複合語も含まれている。また、外来語も含まれている。

3.2.2 目標名の読み上げ一覧

実験参加者は地図課題を終えた後、使用した地図の目標名の一覧を二回読み上げるという、読み上げ課題が課された。目標名は、用紙の表と裏にそれぞれランダムに並べられたが、読み上げの順序による発音に対する影響を考慮して、その順序は異なっている。

3.2.3 目標物のタイプ

地図とは、A3の紙に名称の付いた目標物が組織的に配置されたものである。目標物は、「主要目標物」「共有目標物」「名称変更」「有り/無し」「2:1」「場違いな目標物(Odd-Man-Out)」の6タイプから成っている。このうち、主要目標物は対比条件に従って、2~5の目標物は、共有性条件に従って配置された。

対比条件 四種類の主要目標物には、それぞれ1~4の番号が付けられている(表-1参照)。一枚の地図には、この内一種類の主要目標物が対比条件に従って配置された。

- ++ : 情報提供者, 情報追隨者に二つ存在
- + - : 情報提供者に二つ, 情報追隨者に一つ存在
- + : 情報提供者に一つ, 情報追隨者に二つ存在
- : 情報提供者, 情報追隨者両方共一つ存在

この対比条件と4種類の主要目標物を組み合わせることによって、16組の地図が作成され、対比条件と主要目標物の番号とを組み合わせた名称が付けられた。

共有性条件 目標物はその名前、イラスト、数が一致しているかどうかで、次の四つのカテゴリーに分けることができる。

表 - 2

共有目標物	目標名・イラスト・数が一致
名称変更	イラスト・数は一致、目標名は異なる
有り/無し	片方の地図に存在
2:1	情報提供者に二つ、情報追隨者に一つ存在

場違いな目標物 (Odd-Man-Out) 地図にはそれぞれ場面設定がなされており¹、目標物はその場面に合わせて決められているが、意外性の導入により会話の展開や表現の多様性を誘うという目的のために、場面に合わない目標物が地図にそれぞれ一つずつ設定してある。このような目標物を場違いな目標物 (Odd-Man-Out) と呼ぶ。

3.2.4 経路

地図に描かれている経路の形は四種類あり、主要目標物毎にそれぞれの経路が配置されている。それゆえ、主要目標物と同じ番号が経路にも与えられた。経路は、次の三つの規則に従って描かれている。

- ・経路は、情報提供者も情報追隨者も共通して持つ目標物から始まる。
- ・経路は必ず共有目標物で終る。
- ・経路は必ず共有目標物とその他の目標物との間を交互に進む。

3.2.5 地図セット

全ての主要目標物と対比条件を含めるように、地図を次のように組み合わせた。それぞれの組み合わせを地図セットと呼び、1~4の番号が付けられている。

表 - 3

地図セット1	++1	+2	-3	-4
地図セット2	++4	+1	-2	-3
地図セット3	++3	+4	-1	-2
地図セット4	++2	+3	-4	-1

3.3 被験者条件

3.3.1 親近性

実験参加者は、まず親近性のある同性の二人組を構成し、更に親近性のない同性のペア2組で四人組が構成された²。その結果、一つの四人組は全て同性で構成されることになる。この四人組の中から、2人が課題に取り組むが、その組み合わせで、親近性が統制された。

¹実験参加者にどういった場面であるかの説明はしなかった。

²この4人には便宜上、a1,a2,b1,b2というコードが付けられている。a1とa2、b1とb2のペアがそれぞれ親近性があるが、a1,a2とb1,b2は互いに親近性はない。

3.3.2 視認条件

情報提供者と情報追隨者は、互いの地図が見えないように、窓の付いた壁によって隔てられた実験室にそれぞれ向かい合って座った。視認可能条件では窓越しに相手の顔が見えるが、視認不可能条件では、窓を通して相手の顔を見ることができなかった。

3.4 実験デザイン

3.4.1 諸条件に対する実験参加者の配置

実験は、視認可能条件、視認不可能条件共に、次の表-5、表-6に従って行われた。

表 - 5

四人組				非親近性開始条件			
1	2	3	4				
地図セット				会話番号	親近性	提供者	追隨者
1	2	3	4	1	-	a1	b1
++1	++4	++3	++2	2	-	b2	a2
+2	+1	+4	+3	3	+	a2	a1
-3	-2	-1	-4	4	+	b1	b2
-4	-3	-2	-1	5	-	a2	b2
-3	-2	-1	-4	6	-	b1	a1
-4	-3	-2	-1	7	+	a1	a2
++1	++4	++3	++2	8	+	b2	b1
+2	+1	+4	+3				

表 - 6

四人組				親近性開始条件			
5	6	7	8				
地図セット				会話番号	親近性	提供者	追隨者
1	2	3	4	1	+	a1	a2
++1	++4	++3	++2	2	+	b2	b1
+2	+1	+4	+3	3	-	a2	b2
-3	-2	-1	-4	4	-	b1	a1
-4	-3	-2	-1	5	+	a2	a1
-3	-2	-1	-4	6	+	b1	b2
-4	-3	-2	-1	7	-	a1	b1
++1	++4	++3	++2	8	-	b2	a2
+2	+1	+4	+3				

男子4組、女子4組、計8組の四人組がそれぞれ視認可能条件、残りの計8組が視認不可能条件で実験に取り組んだ。四人組には、表にあるように、両視認条件共に1-8の番号が付いており、この番号と視認条件とで、それぞれの四人組を同定することができる³。四人組の番号は、

³カウンターバランスをとるために、男性と女性は次の表-4のように振り分けられている。

表 - 4

地図セット番号に対応しており⁴、下に並んだ八組の地図にそって、四人組は地図課題を行った。その際の情報提供者、情報追随者、親近性条件については、右の欄に対応している。実験参加者は一人四会話、一つの四人組で八会話行われた。この八話を1セッションとする。

実験は非親近性開始条件と親近性開始条件とに分かれているが、非親近性開始条件では会話を親近性の無いペアから、親近性開始条件では会話を親近性のあるペアから始める。1～4の四人組は非親近性開始条件で、5～8の四人組は親近性開始条件で行う。

3.4.2 会話の同定方法

セッションを同定するコードとして $j[1-8][en]$ が、会話を同定するコードとして、 $j[1-8][en][1-8][fu]$ が用いられた⁵。セッションコードの最初の j は日本語の会話であることを表す。二番目の $[1-8]$ は、四人組の番号を指しており、三番目の $[en]$ は、“e” が視認可能条件を、“n” が視認不可能条件を示している。条件を表現する、以上のコードに加えて、会話番号を表す $[1-8]$ 、親近性の有無を表す $[fu]$ を添えて会話を同定した。f が親近性有り条件、u が親近性無し条件を示している。

4 収録の実施

4.1 実験収録の日時、場所および設備

今回の実験は、1994年7月25日から8月5日の二週間をかけ、千葉大学工学部1号棟5階、505号室内の防音室で行なわれ、DAT 1台、ビデオカメラ4台を使用し、情報提供者と情報追随者の音声、行動を記録した。

4.1.1 防音室

防音室は、それぞれ防音された、隣接した2つの部屋から構成され、一つの部屋が約3×3mの底面を持つ。防音設計は2つの防音室間の壁も含めてD40以上、室内の空調による騒音はNC20以下を満たしている。

防音室と防音室との間には、60(W)×46(H)cmの窓があり、二人の会話者は、窓越しに向かいあって座った。この窓は相手の顔のみが見える高さに設定されており、地図や手許は見えないようになっている。また視認不可能条件の場合は、この窓を塞ぎ、視覚的な情報交換が一切行なわれないようにした。

四人組番号	視認不可能条件	視認可能条件
1	男	女
2	男	女
3	女	男
4	女	男
5	女	男
6	女	男
7	男	女
8	男	女

⁴例えば、四人組番号5は地図セット1の4組の地図を用いることになる。

⁵これらのコードは、収録の際にヘダーとして利用された。

各防音室内には、地図を置く机が窓をはさまり向い合せて設置されている。台を用いて約30度傾斜させることで、視認可能条件の際に自分の地図を覗き込むことを避け、相手ができるだけ自然に視界に入るように工夫した。

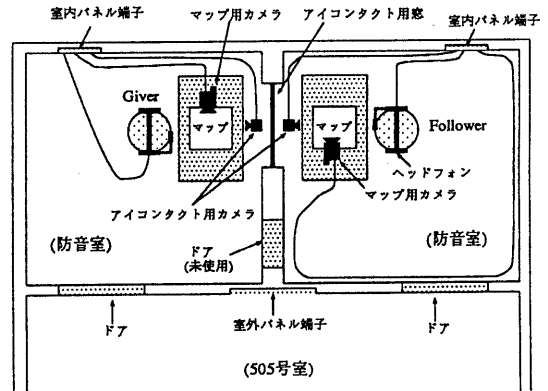


図1: 実験室の見とり図

4.1.2 音声の入出力

今回の実験では、対話者の音声を分離録音するために情報提供者/情報追随者間も遮音条件としたので、二人の音声対話を可能にするため、ヘッドホンマイク HMD 25-1 (SENNHEISER 製) を使用した。これは、ヘッドホン用のステレオ入力と、マイク用のモノラル出力を持つものである。マイクは、机に据え置き型の物なども検討したが、発話者の口からの距離が一定で、音質的に安定した音声記録できるという理由でヘッドホンマイクを用いることに決定した。

マイクの入力はモノラルで行なわれ、実験室内のパネル端子から室外へ伝えられ、各々をL/Rとするステレオ音声として2チャンネルで録音するとともに、平行して実験室内のヘッドホンに伝えられた。室外から室内に指示を与える場合は、この信号に外部からの音声信号をミキシングしたものを、室内のヘッドホンに伝えた。録音には、DTC-2000ES (SONY 製) を使用した。実験では48KHz サンプリングを使用した。

今回の実験では、録音される音声に含まれるノイズを極力押えた。また、録音レベルが低くなる傾向があったので、マイク位置を装着度に調節し、よりレベルの大きい、クリアな音を録音できるようにした。

4.1.3 ビデオ

今回の収録計画の重要な特徴はビデオによる記録の重視である。ビデオカメラは情報提供者と情報追随者各々に対して2台、合計4台を準備した。とくに、ビデオカメラの存在が、発話の自然さを損なわないような配慮が必要であり、今回は、カメラの機種を選定などによってそれを実現した。また、資料分析の観点からは、複数画面の同期が

重要であり、同一画面に4つの小画面を同期させる資料が作成されている。また、ビデオ映像に関する肖像権の問題については十分に注意した。

情報追隨者がどのようにルートを描くか、情報提供者が地図をどのように使うかを見るために、各会話者の左上から地図を写した。この地図撮影用のビデオカメラとして、CCD-V90(SONY製)の8mmビデオカメラが使用された。

表情をみるために、各人の正面、アイコンタクト用窓の下部にFLEXCAM(VIDEOLABS製)の卓上CCDカメラを設置した。これは、台座部分とカメラ部分に分かれており、台座部分から伸びている約40cmのフレキシブルチューブの先に非常に小さなカメラ部があり、カメラ位置を自由に交えることができた。これを使用することで、会話やアイコンタクトに与える影響を極力少なくするように試みた。

実験室外への出力は、4台のディスプレイでモニターすると同時に、4台のビデオデッキで8ミリテープに録画した。使用したビデオデッキは、CVD-1000(SONY製)で、VISCA(SONY video system control architecture)、Hi8に対応したものであった。

4台のビデオの同期を取るために、同じ機種 of デッキ4台を1つのリモートコントローラーによって操作した。DATとビデオの間の同期は、手動で行なった。音声信号は、DATと同じものを記録したので、各実験の最初にベルを鳴らすことによって、記録を処理する際に同期を取りやすいように工夫した。

4.2 参加者

実験参加者は、18歳から24歳の千葉大学の学生で、男性32人、女性32人であった。参加者は、親近性のある同性の二人組を構成し、互いに親近性のない同性のペアー2組で、四人組(四人とも同性)を構成した。

出身地は、南関東を中心に集めた。地域別に見ると、64人の参加者のうち、関東地方出身者が43人、特に南関東出身者は32人であった。

表 - 7

地方	北海道	東北	関東	中部	近畿	四国	九州
人数	2	3	43	12	1	1	2

参加者には、地図を使った簡単なゲームをしてみようということと、ビデオ撮映をすることのみ事前に伝えられ、会話の収録が目的であることは知らされていなかった。

4.3 手続き

4.3.1 手続きの統制

実験は各セッション毎に責任者(主に大学院生)、補助者、録音技師の三名が担当した。全セッションを通し、12人が交代で参加した。責任者は実験全体の進行を、補助者は参加者の誘導や控え室の管理にあたった。手順を統制するために、予備実験をもとにマニュアルを作成し、それに従って実験が行われた。

4.3.2 被験者の統制

参加者の親近性を統制するために、また、参加者が課題に関する情報の交換をしないようにするために、以下の注意が払われた。

- ・ 集合時に互いに話しをしないように、補助者がその管理に当たった。
- ・ 参加者には別々の控え室で待機してもらった。
- ・ 参加者の実験室と控え室との間の誘導は、補助者が必ず行った。
- ・ 他の参加者の実験には立ち合わないようにした。

4.3.3 全体の教示

責任者は4人同時に、実験全体に関する簡単な説明、地図課題の簡単な説明、控え室に関する説明の他、他の参加者と絶対に話さないように教示した。

録音技師が参加者にヘッドホン、ビデオなどの装置に関して説明を行った。説明終了後、最初に実験に当たる二人の被験者を実験室に残し、残りの二人は、補助者がそれぞれの控え室に案内した。その際、部屋から勝手に出ないこと、他の参加者と顔を合わせても、話しをしないことを確認した。

4.3.4 課題の実行

以下が1会話の流れであり、これを8会話分繰り返した。

地図課題: 地図は予め裏にしてセットしておき、被験者には開始の合図があるまで、見えないようにしておいた。会話を行う二人の参加者のうち、責任者が情報提供者を、録音技師が情報追隨者をそれぞれの部屋に誘導し、課題のやり方に関する教示文を読んでもらい、それに関する質問を受けた。その後、次のような教示を与えた。

- ・ ペンを自由に使ってもかまわない。
- ・ ゲームを始める時と、終わった時は、情報提供者がベルを鳴らして合図する。
- ・ 地図をずらしたり、地図がみ込まないようする。

教示が終わったら、ヘッドフォンを着けてもらった。録音技師がヘッドフォンを通じて、録音レベルを確認するために、参加者に名前を言ってもらい、互いの声が聞こえるかを確認した。その際、声が小さい場合は、マイクの位置を調整した。その後、地図を表にし、情報提供者がベルを鳴らして初めるよう教示した。

読み上げ課題: 読み上げ課題は、どの四人組についても、a1,a2,b1,b2の順に行なった。4人の参加者を実験室の情報提供者用の部屋に集め、読み上げ文課題に関して簡単な教示を行ったが、その際、単語をゆっくり発音し、単語と単語の間は少し間隔を開けるように注意をした。読み上げの音は、他の被験者に聞こえないようになっていた。

基礎アンケート: 読み上げ課題が終わった参加者に、次の基礎アンケートと資料公開に関する承諾書を記入してもらった。

た。アンケートでは、年齢、性別、学部学科、両親と本人の出身地、対話の相手との関係などが尋ねられた。

4.3.5 結果

今回の実験により、128対話、約24時間の音声データと映像データが得られた。表-8は、各対話の所要時間と、セッション毎の平均所要時間である(単位:秒)。1対話の所要時間は平均約11分弱である。

表-8

j1n1u	431	j2n1u	415	j3n1u	270	j4n1u	1217
j1n2u	658	j2n2u	1004	j3n2u	360	j4n2u	257
j1n3f	651	j2n3f	1042	j3n3f	320	j4n3f	1240
j1n4f	503	j2n4f	1219	j3n4f	319	j4n4f	721
j1n5u	383	j2n5u	759	j3n5u	296	j4n5u	439
j1n6u	630	j2n6u	709	j3n6u	274	j4n6u	1159
j1n7f	539	j2n7f	491	j3n7f	339	j4n7f	887
j1n8f	417	j2n8f	1428	j3n8f	324	j4n8f	856
計	4212	計	7067	計	2502	計	6776
j5n1f	619	j6n1f	377	j7n1f	1425	j8n1f	1409
j5n2f	719	j6n2f	745	j7n2f	561	j8n2f	992
j5n3u	483	j6n3u	449	j7n3u	838	j8n3u	657
j5n4u	435	j6n4u	445	j7n4u	654	j8n4u	1098
j5n5f	676	j6n5f	329	j7n5f	856	j8n5f	1118
j5n6f	341	j6n6f	865	j7n6f	731	j8n6f	1149
j5n7u	429	j6n7u	369	j7n7u	759	j8n7u	1037
j5n8u	534	j6n8u	437	j7n8u	822	j8n8u	746
計	4236	計	4016	計	6646	計	8206
j1e1u	557	j2e1u	482	j3e1u	517	j4e1u	259
j1e2u	2113	j2e2u	766	j3e2u	1247	j4e2u	504
j1e3f	676	j2e3f	509	j3e3f	878	j4e3f	560
j1e4f	678	j2e4f	507	j3e4f	717	j4e4f	531
j1e5u	503	j2e5u	670	j3e5u	628	j4e5u	733
j1e6u	397	j2e6u	477	j3e6u	577	j4e6u	259
j1e7f	843	j2e7f	525	j3e7f	774	j4e7f	425
j1e8f	764	j2e8f	635	j3e8f	1440	j4e8f	365
計	6531	計	4571	計	6778	計	3636
j5e1f	771	j6e1f	785	j7e1f	551	j8e1f	1124
j5e2f	581	j6e2f	738	j7e2f	318	j8e2f	850
j5e3u	415	j6e3u	803	j7e3u	416	j8e3u	1476
j5e4u	410	j6e4u	563	j7e4u	260	j8e4u	456
j5e5f	718	j6e5f	647	j7e5f	320	j8e5f	1162
j5e6f	471	j6e6f	632	j7e6f	311	j8e6f	490
j5e7u	510	j6e7u	357	j7e7u	277	j8e7u	501
j5e8u	718	j6e8u	617	j7e8u	334	j8e8u	911
計	4594	計	5142	計	2787	計	7570

5 まとめ

千葉大学において収録した地図課題に関し、その目的や意義、収録計画、実験設備、収録の実施に関して概観した。現在、得られた音声データを基に、転記用ツールの開発とともに書き起こしを進めており、最終的にはSGMLのタグの付いた正書法転記(TEIP3準拠)を含む地図課題コーパスを作成する予定である。この実験で収録されたデータは、コーパスを含め、今後CD-ROMで配布することが計画されている。

参考文献

Anderson, A. H. et al.(1991), "The HCRC Map Task Corpus," *Speech and Language*, Vol. 34, No. 4, pp.351-366.

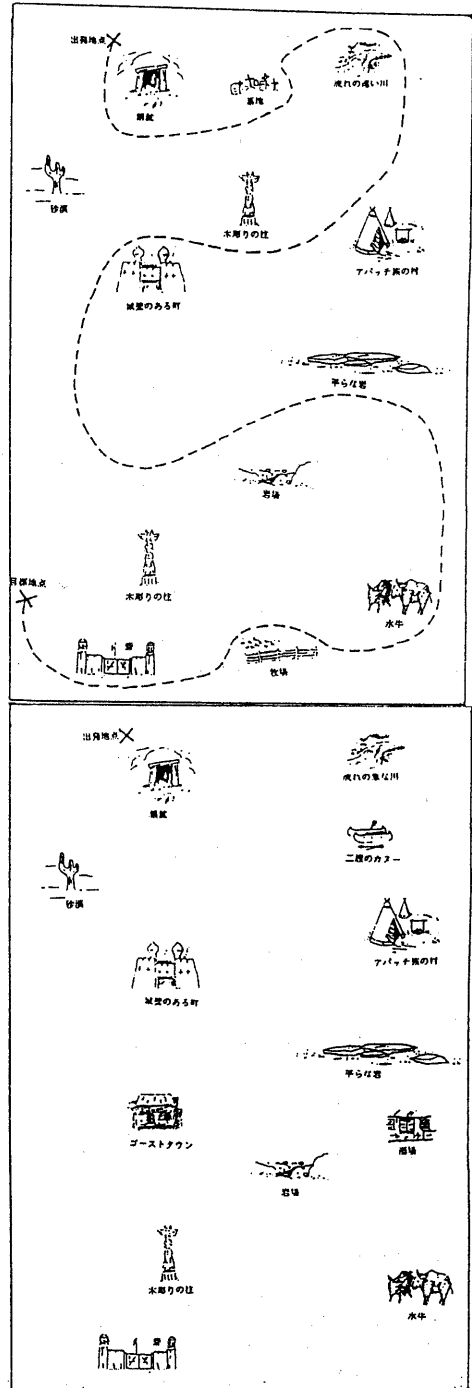


図-2