自然会話データ「偶然の初対面」の公開
～その方法論について～

http://japanese.human.metro-u.ac.jp/mic-j/kaiwa

西郡 仁朗

1．はじめに

近年、自然会話をデータとして分析した研究が盛んであるが、データ収集の方法論をみると、研究者間で統一がなく、科学的なデータとして扱うには杜撰な印象を受けるものが少なくない。自然な会話を数量的に分析するためにデータ収集を始めると、発話をどこで区切るべきか、コーディング（タグ付け）の信頼性をどうすれば確保できるかなどさまざまな問題があるのだが、多くの研究では、この点に関する詳しい記述や配慮がない。特にコーディングは、直観的、主観的判断がもととなるものもあるため、分析者がある仮説を抱き、その検証のためにコーディング作業を進めると、無意識のうちに判断に影響を与える場合がある。

本稿は WEB 上で公開している『偶然の初対面』のデータのデータ収集法及び利用法の紹介、自然会話データを量的に扱う際の方法論上の問題点や解決方法の提案が主目的である。実際の分析結果（多変量解析等を用いている）については別稿に譲る。筆者は自然会話データを今後も公開していく計画であり、その際にも同様の方法論がとられる。方法の一つ一つはさほど複雑なものではないが、それが積み重なると説明にかなりの紙幅を要するので、本稿でまとめて記すことにした。また、方法論を明示することで他の研究者にも同様の方法でのデータの公開を呼びかける。これから自然会話の分析を学びたい、行いたいという人には自学自習素材としての利用をすすめたい。

2．場面条件統制「通常の初対面」と「偶然の初対面」

自然会話のデータ収集には誰がどこでどのような状況でという条件統制が必要である。筆者が現在行っているのは初対面の大学生・大学院生 2 名の会話を対象としたものである。被験者の属性（性別・学年・母語）についてはすべての組み合わせで実験が行われ、対話の環境（部屋・話者間の距離・向き合い方等々）については同一条件となるよう統制されている（3．参照）。

初対面として二つの場合を設定している。一つは「通常の初対面」と呼んでいるもので、パーティーやコンパなどで出会ったと想定して会話を遂行してもらう。もう一つは「偶然の初対面」と呼んでいるもので、たまたまある場所に居合わせた二人の会話を採集するものである。具体的に述べると、「音声学関係の実験があり謝礼付きで被験者を求めている。やってもらうのは、被験者二人で文章を読んでもらうことで、その音声を録音する。」と被験者を募集する。やってきた二人の被験者を座らせ、実験者は手順の説明を開始しようとするが、読んでもらう予定の文章を忘れてきたと称し、その場に二人を置き去りにする。このあと 15 分間実験者は戻って来ない。その間の二人の会話をデータにしようというものである。この設定にはさまざまな目的がある。「通常の初対面」とは目的性、緊張度が明らかに異なり、両データを比較対照できるほか、例えば待遇表現のレベルを決めるのに必要な相手の属性情報をどのように入手するか、ポライトネス的見地からの談話のストラテジー、心理学的自己開示の様子、初対面印象形成等々である。また、音声を明瞭に記録するためにマイクを仕込まなければならないが、上記のような設定だと机上にマイクがあることも自然である。現在公開しているのはこの「偶然の初対面」の方である。

3．対話の環境

「偶然の初対面」のデータ収集では、音声・映像の記録がさまざまな角度から分析される可能性があり、記録が明瞭なものとなるよういろいろな配慮と工夫を行った。また、どの対話もできるだけ同一の環境条件で行われるべく、対座角度（真正面から右 30 度程度に対話相手がいる）、被験者間距離（約 120㎝）、

対座時には鞄などが身辺にないこと、携帯電話の電源が切られていることなどに留意した。さらに、この実験は密室で行われるため、万が一の事故等がないよう、隣室をモニター室として実験者が待機・監視した。実験室内の被験者・什器、機器の配置は以下の通りである。

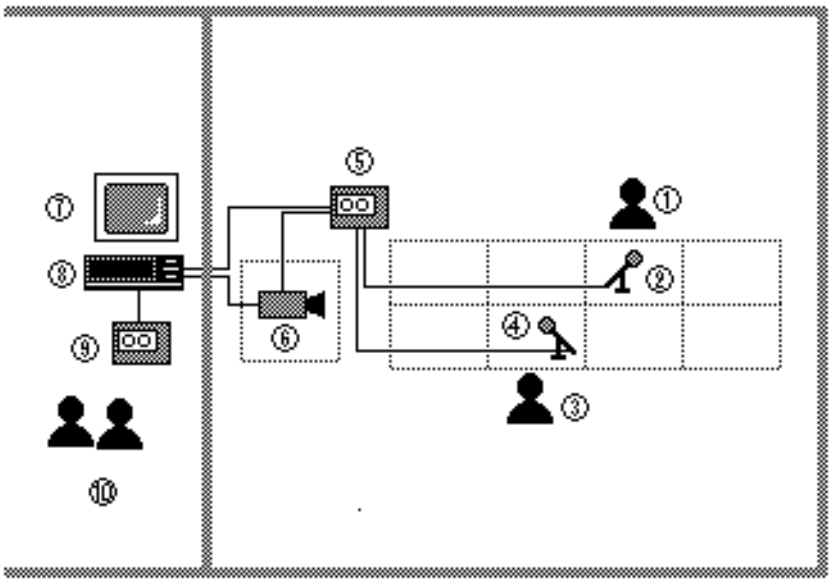

図-1．対話の環境

①被験者 1　②被験者 1 用マイク　③被験者 2　④被験者 2 用マイク
⑤テープレコーダ（マイクはこのレコーダに結線される。このレコーダはミキサー機能付きで、音声出力はモニター室の機器と⑥のビデオカメラに接続されている。）
⑥ビデオカメラ（AV ラックの上に何気なく置かれていて、被験者両名を撮影

できるアングルとなっている。実験中は録画状態が続くことになるので、それを示すカメラ前面の赤いランプにはシールを貼り、録画中であることを被験者に悟られないようにした。⑤の音声出力を受け、ビデオテープには画像・音声とも記録される。画像出力は調整室機器に接続される。
⑦映像モニタテレビ
⑧VTR：音声と映像の記録用（バックアップ）
⑨テープレコーダ：上記の出力をとって録音（バックアップ）
⑩実験者 2 名：被験者が来る前から調整室にいて、映像と音声の記録が正常に行われているか、また被験者の挙動に不審な点がないかをモニターする。

4．被験者の人権

「偶然の初対面」のデータ収集はプライバシー侵害と紙一重のものであり、実験終了後、被験者にことの次第を詫び、研究目的での会話データの利用を許諾するかを聞く。これまで 20 組以上実験を行ってきたが、幸いなことに全員が許可してくれた。これは初対面の人との会話で差し障りのない内容であったことが主因であろうが、実験者が教員で被験者が学生であるという力関係が背景にあるかもしれない。最近アメリカなどの大学では、被験者が必要な実験調査には事前に審査委員会の許可を得なければならない（日本でも医学の治験などでは当然必要となっている）。科学実験の危険性を監察すること、実験によって被験者の人権や被教育権の侵害がないかを審査することなどが目的だが、教員によるアカデミック・ハラスメント防止の目的も含まれているという。本実験はこうした審査を受けていない。この点については実験者側で十分な配慮をしているが、アメリカなどで許可されるかどうか、100%の確信はない。

実験中、室内の様子は別室でモニターしていたが、これは被験者が万一妙な挙動をした場合には、即座に部屋に飛び込んでいくためである。

本来、実験というものは被験者の事前の「インフォームド・コンセント」が必要であろうが、「偶然の初対面」の場合には、これを事後にせざるを得

ない。しかし、データ提供への強要がなく、またプライバシーが保護されるべく万全の配慮をしたと考えている。

図-2．「偶然の初対面」実際の様子

5．実験の手続き

①前置き：実験者の一人（教員）が実験室（セミナー室）で待機、被験者 2 名がそろった時点で次の台詞を言う。「今日は、どうもありがとう。これから、二人である文章を読んでもらいたいんだけど、大体 40 分ぐらいかかると思う。それで、あのう、今日は私は朝からどうかしてて、読んでもらう原稿を車の中に忘れてきたのね。それで駐車場まで取りに行って来るから、ちょっとここで待っててくれる？すみません。」

②実験：実験者が実験室を出ていった時点を開始時点として、3．に述べた方法で会話データを記録する。

③実験終了：15 分経過時点で実験者が入場し、実験の本当の目的を述べ詫びる。対話者名を含めて固有名詞をすべて匿名にすること、その上で

文字化されたデータは公開もあり得ること、音声・静止画・動画などを公開する場合には事前に被験者の許諾を得ることをことわり、研究目的のデータとしての利用への諾否を尋ねる。

④フェイス・シートと事後アンケートへの記入：実験終了後、各被験者にフェイスシートと事後アンケートへの記入を求める。事後アンケートには対話相手に関するいくつかの評価、印象の評定が含まれるため、被験者２人が実験室とモニター室に別れて記入する。フェイスシートと事後アンケートの内容は表-1 及び 2 の通りである（西郡・樋口, 1997)。事後アンケート項目は、1〜4 の四段階で評価を求める項目である。

表-1 フェイス・シートの項目

1．氏名・国籍・所属・年齢などの個人情報
2．使用可能な言語に関する情報（運用レベルについては自己申告）
3．相手の年齢についての推測（自分より上か下か同じくらいか）
4．相手の社会的地位についての推測（自分より上か下か同じくらいか）

6．被験者

日本語の初対面の会話では、相手の年齢・性別・学年・帰属集団・社会的地位に関する情報の交換が重要な要素となって会話が展開する（樋口，1997)。また、待遇表現など言語形式に着目した分析を行う場合には、被験者対の年齢・社会的地位の関係に統制がとられていないと有効なデータを収集することはできない。会話の推進者などを分析する場合も同様である。本稿の実験でも、こうした被験者の属性的要素に注目し、被験者対の間でカウンターバランスが整うように配慮した。

具体的には、以下の三つの属性について、被験者の組み合わせが均等となるように対を選んだ。

① 被験者の母語・・・本実験では日本人同士の対話だけではなく、外国人と日本人の日本語による対話も分析対象となる。すべての外国人被験者は日本

表-2 事後アンケート項目

Ⅰ．相手の言語能力やパラ言語に関する項目 （評定尺度法による）

1．日本語文法をよく知っていて、ことば使いが正しい。
2．語彙が豊富である。
3．意見・説明等の内容が明確で分かりやすい。
4．言葉の丁寧さの程度が適当である。
5．一つ一つの言葉の発音がはっきりしている。
6．アクセント・イントネーションが正しく分かりやすい。
7．話し方の速度が適当である。
8．相槌の打ち方が適当である。
9．間の取り方が適当である。
10．二人で対話をすることに協力的な話し方である。
11．話の内容が興味深い。
12．声の大きさが適当である。
13．視線の合わせ方が適当である。
14．表情の変化が適当である。
15．ジェスチャーが適当である。
16．外見が魅力的な人だ。

Ⅱ．相手の対人印象（評定尺度法による）

1．好きになれそうな人だ。
2．暖かそうな人だ。
3．話しやすい人だ。
4．知的な人だ。
5．明るい人だ。
6．礼儀をよくわきまえた人だ。
7．協調性のある人だ。
8．信用できそうな人だ。
9．近しい感じがする人だ。
10．開放的・オープンな人だ。
11．積極的な人だ。
12．まじめな人だ。
13．誠実な人だ。
14．親切そうな人だ。

語については「超」上級者である（日本語能力試験の１級を越えるレベルで一年以上日本に滞在している者）。目下、公開データでは、日本人と日本人、中国人と日本人の対話だけが示されているが、その他の母語話者（当面は韓国語）についても実験を行っていく予定である。

②年齢と学年・・・公開データでは大学の学部１・２年生と大学院生という年齢・学年にかなり開きのある被験者対が会話を行っている。現在、同年齢・同学年の対についても実験を行っている。
③性・・・上記①②の全条件が確保された上で、男性-男性、男性-女性、女性-女性の対が均等に配置されている。

以上の対の例を略号で表すと

COM1-JYF1

（中国人：Chinese，年上の院生：Older，男性：Male，この組み合わせで１番目の人）と（日本人：Japanese，年下の学部生：Younger，女性：Female，この組み合わせで１番目の人）ということになる。

表-3 に被験者の組み合わせを列挙する。

表-3．被験者の組み合わせ

1．JOM-JYM	2.JOM-JYF	3．JOM-CYM	4．JOM-CYF
5．JOF-JYM	6．JOF -JYF	7．JOF -CYM	8．JOF -CYF
9．COM-JYM	10.COM-JYF		
11．COF-JYM	12．COF -JYF		

表-3 で右下のセルが空白となっているのは、これらが中国人同士の対で中国語での会話となるため除外したからである。

数量的分析のためには、同一セルの内容が多ければ多いほどよいが、現在までに収集されているのは 24 対、公開データでは 12 対である。

7．文字化の手順

データの文字化に当たっては、BTSJ (Basic Transcription System for Japanese；宇佐美，1997)を採用した。今回の実験では多方面にわたる分析

が行われる。例えば、非言語行動などもその対象になるが、詳細な分析のためには、文字おこしデータばかりでなく記録ビデオを詳細に検討していく必要がある。文字おこしデータは、その前段階の全体論的な検討、また参照データとして活用されることになる。そうした性質を考えると、限定的な目的に利用されるプログラムではなく、いろいろな汎用ソフトでも使用できる比較的単純なデータ構造が求められると考えた。また、量的な分析ばかりでなく、質的な面を全体論的、内省的、直感的にとらえられる必要もある。自然会話の分析、特に発達心理学での分析ツールとして CHILDES SYSTEM (Oshima-Takane and MacWhinney, 1995) が広く知られており、本実験でも採用が検討された。CHILDES SYSTEM は会話データを形態素に分けて入力すれば、さまざま集計・検索・分析などを行うことができるものである。もともと、分かち書きをするアルファベット系言語用であったため、かつては日本語もローマ字化する必要があったが、現在は日本語での入力も可能であり、相当数の利用者がいる。しかし、日本語のすべての形態素について研究者間で完全な一致がある訳ではないし、今回の実験では形態素では表せない会話ならではの言い淀みや不明瞭な発話も興味の対象となっている。また、日本語学習者の会話データでは形態素よりも、よりマクロな「文型」での分析の方が有効な場合が多い。集計や検索機能についても、他のプログラム言語や汎用ソフトで十分対応可能であるため CHILDES SYSTEM の導入は見送った。

既述の通り、本実験では単純なデータ構造が望ましく、また、読みやすいこと(readability) も重要な要素で、漢字仮名交じり表記でなければならないであろう。しかし、だからといって音声を単純に文字おこしするだけでは、会話の重要な要素である声のオーバーラップ、イントネーションなどの情報が失われてしまう。これを補うのが BTSJ である。BTSJ では表-4 に示す記号を用いてこうした言語関連情報を記述する。どの言語関連情報にどの記号を用いるかは任意に定めることが可能であり、他の研究では別種の記号や方法がとられているものもあるが、BTSJ にはすでに数年間の試用と改良の蓄積があり、本実験の趣旨にもっとも適した記号体系及び方法論であると思われ

た。

実際の文字化に際しては以下の手順を踏んだ。

①研究補助者[i]による一次文字おこし・・・この段階から BTSJ に従った文字おこしを行うが改行や発話文の問題（後述）にはそれほど配慮せず、音

表-4．BTSJ の記号（宇佐美，1997 より抜粋。）

・・・	言い淀んで中途で終了する発話
？？	いわゆる「半疑問文」
< >{>} < >{<}	オーバーラップしている発話。同時発話されたものは、重なった部分双方を< >でくくり、重ねられた発話には、< >の後に、{<}をつける。また重ねた方の発話には、< >の後に、{>}をつける。
( )	短く、特別の意味を持たない聞き手の「あいづち」。話し手側の台詞の中に入れる。
< >	笑いながら発話したものや、笑い等は、< >の中に、<笑いながら>、<二人で笑い>などのように説明を記す。
(<笑い>)	相手の発話の途中に、相手の発話と重なって笑いが入った場合の表記法
/沈黙 秒数/	沈黙の表示
##	聞き取り不能の際の表記法。マーク数は推測された拍数
[ ]	非言語行動などの周辺言語情報の略述

声をできる限り正確に再現することに集中する。入力に際には Microsoft Excel を用いたが、この段階では表計算やデータベース機能を用いるわけではないので、テキストデータを扱うどのソフトでも構わない。

②研究者１と研究者２によるチェック・・・上のデータを研究者１と研究者２が分担し、BTSJ の記号利用の面から点検した[ii]。

③発話文の分割に関する一致度確認と分担作業・・・発話文[iii]の分割について、同じ部分（全体の６分の１）を研究者１、研究者２がともに認定を

行い一致度（Cohen’s κ）を測定した。基準値(0.8)以上ならば分担して作業を行った（詳細後述）。

④コーディングの一致度確認と分担作業・・・発話文ごとのスタイル（言語形式上の丁寧さのレベル）と機能（情報要求発話と自発的情報提供発話のみ）についてコーディングを行った。その際にも一致度を測定し、基準値(0.8)以上ならば分担して作業を行った（スタイルと機能については後述する）。

8．発話文の分割と一致度（Cohen’s κ）の測定

本稿のような自然会話データの収集では、発話文というまとまりに分割する際、及び発話文のスタイルや機能などのコーディングを行う際に研究者の主観的判断がどうしても入り込んでしまう。勿論、各コーディングの方法については詳細な定義があり、誰が行っても同じ結果になるような工夫はなされているが、偶然の一致（恒常誤差に近い）や調査者の個性・思いこみ・特定の結果への予測や期待によってある種の「偏り」や「ズレ」（個体誤差に近い）が生じる場合がある。また、曖昧性などコーディングの定義自体に問題がある場合もある。特に、本稿の実験では、二人の研究者が手分けしてデータの分割やコーディングに当たるため、どちらが個々のデータを分割、コーディングしても同じ結果になるという何らかの保証がないとデータの科学性が損なわれる。この保証のためには様々な方法があるが、実験の性質から考え、Cohen'sκ(kappa) という指標[iv]を採用し、二人の研究者の一致度と定義の明確性を見ることにした(Bakeman and Gottman, 1986)。一致度の客観性と科学性の確保については、ここまでやれば絶対に問題ないという明確な基準はない。調査者の訓練や試行錯誤、点検者の確保等、いろいろな面での洗練と配慮が必要である。今回とった方法は、様々な方法からあるものを採用しただけであり、たとえ一致度が高くてもそれで大丈夫だということではない。

発話文分割の Cohen'sκ測定については独自の方法がとられているので以下

に記す。

点検済みの一次文字おこしデータでは、句読点が打たれ、また、話し手が交代する際に改行もなされている。これを図-3 のような「ベタ打ち」に変換した。発話文分割について Cohen'sκを測定する場合、その基礎となる単位（ここでは区切れる、ここでは区切れないと判断するデータの基本長とでもいうべきもの）が必要だが、今回は、改行及び句読点による区切り（○）を採用した。文節なども単位の候補として考えられたが、文節の場合は基礎となる単位（区切り）の数が非常に多くなり、Cohen'sκの精度が低くなると思われたので改行・句読点による区切りを用いることにしたのである。図-3 では基礎となる区切りの箇所に○が記されている。なお、○が重なるような場合や記号を挟んで連続している場合は、一つに統一してある（例：「ああ○<そうですか○>{<}○」は「ああ○<そうですか>{<}○」としてある）。

【鈴木】あ○あの工学部の建築学科で○（はい）勉強してます○鈴木と申します○【田中】工学部の建築科・・・（建築科）あ○そうですか○（はい）○【田中】建築科○だったら・・・<笑い>○【鈴木】え？○【田中】いや○あの○建築科だったら○ほとんど男性じゃないかな○と思って○（ああ○ああ○ああ）それで・・・○【鈴木】いや○最近は○でも○女子も多くって○【田中】ああ○そうですか○【鈴木】ええ○あの○私の学年なんかは半々ずつです○男女は○【田中】半々ずつですか？○【鈴木】はい○【田中】ああ○<そうですか>{<}○【鈴木】<結構○>{>}うん○あの○ね○理系のなかでも○結構文系よりだから○【田中】そうですか○（うん）○【田中】でも・・・理系のなかでも文系より？そうすると設計者とかに<なるんですか○>{<}○【鈴木】<そうですね○>{>}○【田中】あ○そうですか○【鈴木】あの○設計者だけではないですけど○（うん）まあ

図-3．一次文字おこしデータの「べた打ち」

研究者二人は、別個にベタ打ちのどこで発話文が区切られるかを、原稿の文

字列、ビデオ、テープを参考に斜線を引いて記していく。改行の○と句読点の○が単純な相槌を挟んで連続しているような場合（例：そうですか○（うん）○）には、相槌が発話文と認められない限り後ろの○に斜線を引くこととした。その結果の例を図-4 に示す。

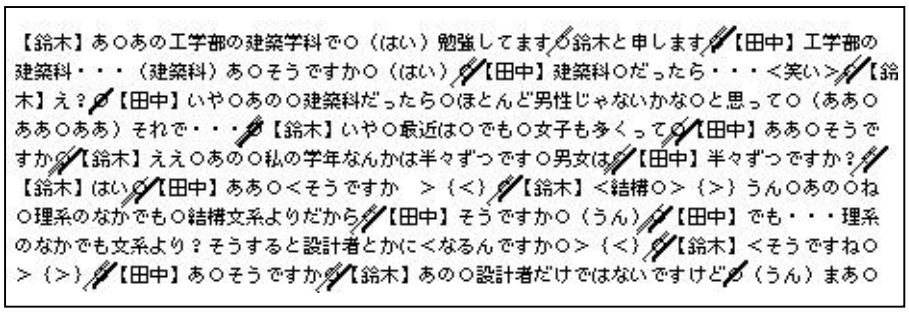

図-4．発話文の区切り-一致度測定の様子

こうした作業をもとに○を基本単位数とした一致・不一致を数え、Cohen's κを算出し、その値が 0.8 以上であれば、研究者１と２が作業を分担することとした。

Cohen's κがどの値以上であれば、一致度が安定しており作業を分担してよいという明確な基準はないが、経験的には直感的判断が伴う難しいものでは 0.7 以上とされている(Bakeman and Gottman, 1986)。しかし、今回の実験での発話文分割やコーディングは比較的機械的な作業であったため、基準をやや厳しくし、0.8 とした。

なお､発話文分割及び下記のコーディングともに研究者１は常に筆者であり、研究者２は複数の者が当たっている（註 II 参照)。

9．公開データでのコーディング

公開データでは２種のコーディングの結果が示されている。

スタイルは、各発話文を以下の基準に従って分類・コーディングした。

PH：相手に対する尊敬語、謙譲語が含まれる発話。(Polite Honorific/Humble)

PO：敬体（です、ます)、改まったことばを含む発話。「はい」など、よりぞんざいな「うん」や「ええ」などの同義語があるのにていねいなことばが含まれる発話文もこの範疇とする。(POlite)

NP：常体（だ、である)、くだけたことば等を含む発話。(Non-Polite)

？：上記のような文体を示すマーカーがない発話。

スタイルを厳密に分類するには、発話文内容を点検し、「詞」と「辞」の観点や聞き手に対する敬意か話題の中の人物に対する敬意かなど、さまざまな面からの検討が必要であろうが、公開データでのコーディングは表面的な言語形式としての分類に留まっている。スタイルの大まかな傾向を見るものにすぎない。

機能は、対話のイニシアチブをどちらがとっているかを検討する参考として各発話文を分類・コーディングしたもので、以下の基準に従っている。

GI：自発的情報提示。質問されてもいないのに自分から進んで自分に関すること、その場の状況に関すること等について情報を提示している発話文。Giving Information

RI：情報要求。相手からの情報を要求する発話で、多くの場合質問の形をとる。Requesting Information

空白：上記のどちらでもない発話。

以上のコーディングについても Cohen'sκを測定しその値が 0.8 以上であれば研究者１と２が作業を分担することとした。

10．データ形式と利用可能なアプリケーション・ソフト

公開データは CSV (Comma Separated Value) ファイルとしてダウンロード

可能になっている（図-5．参照)。周知の通り、CSV はカンマとリターンで

対話者対,行#,発#,*?,分,TB,C1,C2,発言者,発言内容
3.COM2-JYF1,111,97,*,5,2,PO,RI,COM2,あの今、学校のサークルの、卓球サークルの、まあそのサークルの人、知っていますか、あの、あなたは?。
3.COM2-JYF1,112,98,*,5,2,PO,,JYF1,卓球サークルの人ですか?(はいはい)。
3.COM2-JYF1,113,99,*,5,2,PO,,JYF1,んー、て言うか、一応その、所属している方は(はい)、わかりますけども、(はい)んー。
3.COM2-JYF1,114,100,*,5,2,PH,RI,COM2,あの、あの、人の電話番号をちょうだいしていただけますか?。
3.COM2-JYF1,115,101,*,6,2,PO,,COM2,もしこの試合前に、(はい)僕の方からこのサークルの、誘いたいです。
3.COM2-JYF1,116,102,*,6,2,PO,,COM2,一緒にあのー、遊んでー、それは楽しいと思うんですけど。<二人で笑い>

図-5．公開データの一部　CSV

データを区切ったテクストデータであり、どのデータベースソフト、表計算ソフトにも取り込めるものである。分析する人がどのソフトウェアを使っていても使用できる形式としてこの CSV を選んだ。また、CSV のもう一つの利点は記憶容量が少なくて済むことで、特定のソフトウェアのファイルにすると、そのソフトのさまざまな機能が付随して「重く」なるが、CSV にはそうしたものが一切ない。公開データを下記 Microsoft Excel に載せると CSV の四倍の容量が必要となる。インターネット上で公開する形式としては CSV が最適であると思われる。

この CSV ファイルをダウンロードし、使い慣れたデータベースや表計算ソフ

トを使えば分析が可能となる。例として Microsoft Excel に載せ、書式等を整えたものを図-6．に示す。

図-6．Excel に変換された公開データ

A 列：対話者対略号　B 列：行番号　C 列：発話文番号
D 列：発話文の終了する行を表す記号*　E 列：経過分
F 列：タイムブロック（序盤：１，中盤：２，終盤：３）
G 列：コーディング１（スタイル）　H 列：コーディング２（機能）
I 列：発言者　J 列：発言内容

11．終わりにかえて

自然会話データを収集し、処理可能なデータ形式にするのには大変な労力と時間がかかる。しかも、統計的に安定した知見を得るには大規模データが必要になる。本データ公開が、自然会話を研究する人の一助になればと願うとともに、他の研究者のデータ公開を呼びかけたい。

註

[i] 1996～1998 年度の東京都立大学人文学部『言語表現法』受講者有志が協力してくれた。記して感謝する。

[ii] 研究者１は筆者であるが、研究者２は張元哉氏、李栄華氏、及び注 i の有志の一部である。

[iii] 改行と発話文について BTSJ（宇佐美,1997）では次のように定義されている。（原文の抜粋）。

実際の会話における話し言葉では、文中にいわゆるあいづちが入ったり、文末が省略されたりすることが多い。また、文法的には一単語に相当するものが、実質的機能を担っている場合もある。よって、ここでは、いわゆる「一語文」や、述部が省略されているもの、或いは、最後まで言い切られていないものなども、「一発話文」として扱い、発話文ごとに改行する。ただし、何かを思い出そうとする時などに用いられる、「フィラー」としての「そうですねえ」などは、「あのう」などと同じく、後続部とまとめて一発話文と数える。

また、逆に、途中で相手のあいづちなどが入って、話者が一旦交替してラインが変わっても、同一話者によって発せられた、構造的に「文」を成していると捉えられるものは、それら数行をまとめて「一発話文」と数える。例のＢ２とＢ３で、「１発話文」を成していると捉え、この２行に同じ発話文番号をつける。

例　Ａ１：目白ですと、バス・・・、ですか？。
　　Ｂ１：え、あの、バスもありますけど、（ええ）大概はちょっと歩いて、
　　　　あのー、行っています。
　　Ｂ２：そうですねー、歩くとですねー、１２、３分、
　　Ａ２：はい。
　　Ｂ３：もう、ちょっとかかるかもしれませんけれども。

[iv] Cohen's κ は次式で得られる。$\kappa = (Po - Pc) / (1 - Pc)$
ただし、Po: 観察された単純一致率。Pc: 偶然一致する確率。

下表のように、研究者１，２があるデータについて A, B, C の三者択一のコーディングを行った場合、灰色部分が一致しているので Po（単純一致率）は 0.87 となる。研究者１，２ともに偶然 A を選ぶ確率は、各人の選択傾向から、(42/100)*(40/100) = 0.168 で、偶然 B, C を選ぶ確率を加算して Pc = 0.342。従って κ = 0.802 となる。

		研究者 1			
		A	B	C	計
研	A	38	0	2	40
究	B	1	26	4	31
者	C	3	3	23	29
2	計	42	29	29	100

引用文献

Oshima-Takane, Y. and MacWhinney, B. (1995) “ CHILDES Manual for Japanese “

Bakeman, R. and Gottman, J. M. (1986) " Observing interaction: an introduction to sequential analysis " Cambridge University Press

西郡仁朗・樋口斉子(1997) 「研究概要と実験の方法について」平成 7 年度～平成 8 年度文部省科学研究費-基盤研究(C)(2)-研究成果報告書『日本人の談話行動のスクリプト・ストラテジーの研究とマルチメディア教材の試作』（研究代表者：西郡仁朗）50-57

樋口斉子(1997)「初対面会話での話題の展開」同上書. 75-109

宇佐美まゆみ(1997)「基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）の開発について」同上書. 12-26