



The ICNALE Spoken Dialogueの設計 : 対話における L2口頭産出研究のために

石川, 慎一郎

(Citation)

Learner Corpus Studies in Asia and the World, 3:9-26

(Issue Date)

2018-03-12

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81010115>

(URL)

<https://hdl.handle.net/20.500.14094/81010115>



The ICNALE Spoken Dialogue の設計
 —対話における L2 口頭産出研究のために—
 石川 慎一郎 (神戸大学)
 iskwshin@gmail.com

Design of the ICNALE Spoken Dialogue
 —For Studies of L2 Oral Production in Dialogues—
 ISHIKAWA Shin'ichiro (Kobe University)

概要

アジア圏国際英語学習者コーパス ICNALE (International Corpus Network of Asian Learners of English) の最新モジュールとして、現在、ICNALE Spoken Dialogue の開発が進行中である。本稿では、既存の主な対話型学習者発話コーパスについて概観した後、ICNALE Spoken Dialogue の基本理念とインタビュー構成を示し、あわせて、先行して構築が進められている日本モジュールの概要を紹介する。

キーワード

学習者コーパス, 対話, OPI

1. はじめに

アジア圏英語学習者の L2 における書き言葉・話し言葉を体系的に収集する ICNALE (International Corpus Network of Asian Learners of English) は、2007 年のプロジェクト開始以降、10 年を経て、世界最大級の学習者コーパスの 1 つとなり、各国で研究や教育に広く使用されている (石川, 2012; 石川, 2015b)。

ICNALE には、作文を集めた ICNALE Written Essays (Ishikawa, 2013)、校閲作文と作文評価データを集めた ICNALE Edited Essays (Ishikawa, in press)、モノローグ発話を集めた ICNALE Spoken Monologue (Ishikawa, 2014) の 3 つのモジュールがあり、これまでに公開された作文・発話サンプル数は 10,000 件以上、総語数は約 195 万語に及ぶ。

ICNALE は、(1) アジア圏への特化 (EFL 圏/ESL 圏併せて 10 か国・地域でデータを収集)、(2) 産出条件の統制 (トピック・時間・テキスト長等を統一)、(3) 学習者の習熟度調査 (習熟度テストや語彙テストのスコア情報を収集)、(4) 学習者の背景情報調査 (性別・年齢・専攻・学習動機・L2 使用経験・L2 学習歴等の情報を収集)、(5) 母語話者データ収集 (比較基準となる母語話者の同一条件下での産出データを収集)、(6) データの全面公開 (ダウンロード版およびオンライン版の提供) といった特徴を持つコーパスで、とくに、学習者・母語話者間、異なる母語を持つ学習者間、異なる習熟度の学習者間において、精緻で信頼性の高い中間言語対照分析 (contrastive

interlanguage analysis: CIA) (Granger, 1996)が可能になるよう設計されている。

当初, ICNALE は学習者作文コーパスとしてスタートしたが, 2016 年に ICNALE Spoken Monologue が公開されたことで, 書き言葉と話し言葉という産出モード間での新たな対照研究の可能性も広がっている (Ishikawa, 2015a)。しかし, 話し言葉に内在する対話性やコミュニケーション性を考えた場合, 独話データだけで話し言葉を概括するのは必ずしも適切ではない。そこで, プロジェクトチームでは, 2017 年度から 2019 年度までの 3 か年計画で, 新たに, ICNALE Spoken Dialogue の構築を開始することとなった。

本稿では, 以下, 既存の対話型学習者発話コーパスを概観し, その後, ICNALE Spoken Dialogue の基本理念, データ収集プロトコル, および, 先行して構築が進められている日本モジュールの概要について解説する。

2. 対話型学習者発話コーパス

2.1 対話収集手段としてのオーラルインタビュー

初期の英語学習者コーパスの多くは書き言葉を対象としたもので, 話し言葉のコーパスは一般的ではなかった。これは, 発話の収集や書き起こしに大きなコストがかかるためである (De Cock, 2010; Ishikawa, 2014)。しかし, 2000 年代以降, 対話型の発話コーパスについても, 一定の規模を持つものがいくつか開発・公開されるに至っている。

もともと, 対話型のコーパスと言っても, 自由な雑談を無計画に集めただけでは研究資料としての価値は低い。そこで, 先行するコーパスプロジェクトでは, 学習者の発話を収集する手段として, 一定のタスクと手順を定めたインタビューが行われた。こうしたオーラルインタビューには, 口頭発表能力試験の目的で実施されるもの (Oral Proficiency Interview: OPI) と, 評価や診断を行わないものが存在する。本稿では, 前者をテスト型オーラルインタビュー, 後者を非テスト型オーラルインタビューと呼ぶ。

学習者コーパスの開発に必要となるのは, (1)発話者の習熟度情報が付与された, (2)自発的で, かつ, (3)統制的な発話データである。(1)は分析の妥当性を, (2)は研究資料としての正統性を, (3)は比較研究の妥当性を, それぞれ担保する要素となりうる。ここで注目すべきは, 発話データの収集手段としてのテスト型, 非テスト型という 2 種類のオーラルインタビューが, 3 つの観点について相補的な関係になっていることである。

平たく言えば, テスト型オーラルインタビューには, 1 回の調査で発話データ収集と習熟度データ収集を同時に行えるというメリットがあるものの, 発話の自発性や統制度は保ちにくい。一方, 非テスト型オーラルインタビューは, 発話の自発性や統制度を保ちやすいというメリットがあるものの, 別途, 習熟度テストを受験させなければならないというデメリットがある。こうしたことから, 以下で具体的に示すように, 先行する対話型学習者発話コーパスの中にも, テスト型オーラルインタビューでデータを取るものと, 非テスト型オーラルインタビューでデータを取るものが混在している。

表 1 テスト型／非テスト型オーラルインタビューを用いたコーパスデータ収集

	テスト型	非テスト型
話者 習熟度	○ 試験官が、下限を測るレベルチェック(level check)や、上限を測る突き上げ(probe/ pushing)を行うことで、インタビューを通して個々の被験者の口頭能力を正確に診断することができる	△ インタビューでは習熟度情報が取れないため、別途、習熟度テストを受験させる時間的・物理的コストが発生する
発話 自発性	△ 評価対象となるため、間違いを恐れて、受け身的に反応しがちである	○ 評価対象とならないので、リラックスして積極的に話しやすい
発話 統制性	△ 評価のため、個々の被験者のレベルにあった質問やタスクを行うので、発話内容がばらつきやすい	○ 必要であれば、全員に同じ質問やタスクを与えることができ、発話内容の統一性を高めやすい

なお、テスト型オーラルインタビューは、パフォーマンス評価の典型例として広く実践されている一方、(1)発話の内容がスタイル(インタビュー／報告)やトピックの影響を受けやすい、(2)受験者は受け身の応答に終始しがちで、会話の主導権を取ることがない(発問、トピック選定、ターン取りなど)、(3)コミュニケーションを特徴付ける無計画性や話者間の対等性が欠如している、といった制約が指摘されており、自然な会話能力(conversational ability)を診断できていないという批判も存在する(Csépes, 2009, pp. 32-36 のレビューを参照)。

2.2 LINDSEI と NICT JLE Corpus

ここでは、一般に利用が可能な大型の対話型学習者発話コーパスのうち、テスト型オーラルインタビューのデータを集めた NICT JLE Corpus (National Institute of Communication and Information Technology, Japanese Learners of English Corpus) (和泉・内元・井佐原, 2004) と、非テスト型オーラルインタビューのデータを集めた LINDSEI (Louvain International Database of Spoken English Interlanguage) (Gilquin, De Cock, & Granger, 2010)を取り上げ、それぞれの概要を比較しておこう。

表 2 NICT JLE Corpus および LINDSEI の概要

	NICT JLE Corpus	LINDSEI
インタビュー	テスト型、約 15 分	非テスト型、約 15 分
学習者国籍	1 か国 (日本)	11 か国 (ブルガリア・オランダ・フランス・ドイツ・ギリシア・イタリア・ポーランド・スペイン・日本・中国)
学習者人数	1,281 人	554 人 (1 か国約 50 人)

母語話者人数	20 人	なし
総語数	200 万語（学習者 130 万語＋調査者 70 万語）	108 万語（学習者 79 万語＋調査者 29 万語）

すでに述べたように、非テスト型であっても、インタビューは一定の手順に則って行われる。以下、は、それぞれのインタビューのタスク設計である。タスクの数は、NICT JLE Corpus が 5 種類、LINDSEI が 3 種類であるが、NICT JLE Corpus におけるウォームアップとwindダウンは付加的なものなので、正味のタスクに限れば、いずれも 3 種類であると言える。

表 3 LINDSEI および NICT JLE Corpus におけるオーラルインタビューのデザイン

課題	NICT JLE Corpus	LINDSEI
1	ウォームアップ：簡単な質問に回答する	トピック発話 (set topic) : 3 つのトピック（自分にとっての重要な出来事・行ったことのある国・見た映画／芝居）から 1 つを選び、2～3 分の準備の後で自由に話す
2	イラスト描写：全 7 種から指定された 1 枚の絵を描写する	自由議論 (free discussion) : 上記トピックに関する質問、および、その他の質問 (大学生活・趣味・海外旅行等) に回答する
3	ロールプレイ：全 14 種（状況 5 種×レベル 2～3 種）から指定された 1 枚のカードに基づき役割会話を行う	絵描写 (picture description) : 4 コマイラスト (絵のモデルになった女性が画家に注文して絵の髪型を変えさせる) の内容を描写する
4	ストーリーテリング：全 9 種から指定された 1 枚のイラストカード（4 コマないし 6 コマ）を見て、それらについて語る	N/A
5	windダウン：簡単な質問に回答する	N/A

NICT JLE Corpus と LINDSEI に共通するのは、イラストを描写するタスクである。イラスト描写は、学習者にとって、(1) 発話内容を考える負担がなく、L2 発話そのものに集中できる、(2) 言えない内容は言い換えたり飛ばしたりするなど、一定の方略を使用して、破綻なく最後まで話しきることができる、(3) 発話中に質問者とのやりとりが不要なため、自分のペースで話せる、といった利点がある。また、コーパスデータとしても、(1') イラストが同一であれば相互に比較可能なデータが取れる、(2') 複数のイラストを時系列に沿って物語るタスク (※コーパスにより、この種のタスクを「絵描写」と呼ぶ場合と「ストーリーテリング」と呼ぶ場合がある) では、過去形や完了形などの時制使用を引き出すことができ、学習者の文法力や全般的な習熟度を表すデータが取れる、(3) 英語での表現が難しい内容を盛り込んだイラストの場合は、回避・言い換えといった主体的な方略使用を引き

出すことができ、学習者のコミュニケーション能力を表すデータが取れる、といった利点がある。

一方、NICT JLE Corpus にのみ含まれるのは、ロールプレイである。ロールプレイでは、通例、何らかの課題の遂行を行わせる。たとえば、駅での切符購入を場面としたロールプレイでは、初級者には行き先を告げて切符を買うことが、中級者には目的地に到着する最速の列車を問い合わせてその切符を買うことが、上級者には乗り遅れた列車の切符の払い戻し交渉を行うことがそれぞれ課される。こうしたタスクでは、細かい英語の間違ひの有無よりも、様々な方略や交渉術を駆使し、相手と意味のあるやり取りを積み上げ、最終的に指示された課題を達成できたかどうか問われる。学習者には負荷の大きいタスクであるが、コーパスデータとして見た場合、狭義の文法力や語彙力に留まらず、学習者の実践的なコミュニケーション力を表すデータが取れるという点で価値が高い。

3. ICNALE Spoken Dialogue の設計

ICNALE Spoken Dialogue の開発に先立ち、(1)対話性の確保、(2)発話量の確保、(3)統制性の確保、(4)相互比較性の確保、(5)検証性の確保、(6)ELF 視点の確保、(7)匿名性の確保、(8)マルチモーダル性の確保、という 8 つの基本理念を定めた。

3.1 対話性の確保

言語産出のタイプについては、一般に、書き言葉と話し言葉というモードの違いが意識される。しかし、それとともに重要になるのは、独話か対話かという、産出の方向性の違いである。これらは、以下のように整理することができる。

表 4 言語産出の下位分類

	書き言葉	話し言葉
独話型	作文, エッセイ	モノローグ
対話型	手紙, 電子メール	ダイアログ

ICNALE プロジェクトでは、Written Essays によって独話型の書き言葉を、Spoken Monologue によって独話型の話し言葉を集めてきたが、Spoken Dialogue で新たに収集しようとするのは、これまで欠けていた対話型の話し言葉である。

この点をふまえれば、学習者が、緊張した雰囲気の中で調査者の質問に受け身的に答えるだけでなく、リラックスした雰囲気の中で主体的に主張したり反論したりできるインタビューの形態が望ましい。そこで、ICNALE Spoken Dialogue では、非テスト型のオーラルインタビューを採用し、あわせて、ロールプレイを導入することとした。

もともと、ロールプレイの体裁を取るだけで真の対話性が引き出せるわけではない。実際、行き先を告げて切符を買うような依頼型タスクや、複数の経路を尋ねて最速の経路を探すような情報確認型タスクでは、やりとりは定型の域を出ない。

そこで、ICNALE Spoken Dialogue では、一般の OPI ならば上級用とみなされる難度の高い

課題をあえて全員に課すこととした。ロールプレイは 2 種類用意されているが、一方の課題では、学習者は、アルバイトを控えて勉強に励むよう忠告してくれる教授に反論する学生役を、他方の課題では、喫煙可を公言しているレストラン店主に抗議して返金を求める顧客役を担当する。ここで指摘しておきたいのは、いずれの場合も、学習者の側が道義的に不利な立場になるよう設定されているということである。不利な状況に置かれた学習者は、言葉を尽くして自説を展開し、相手を説得するほかなくなる。こうした特殊な設定は、主体的に意見を述べるのが苦手な学習者に対して、話す動機と必然性を与える仕掛けとなっている。

こうした設定はまた、インタビューをする側とされる側に潜在する社会的関係性とロールプレイにおける話者の社会的関係性を一致させるためのものでもある。インタビューという形態を取る以上、いかに配慮したとしても、質問者の側が強い立場に立つことは避けようがない。面識の有無によらず、インタビューをする側が教師、される側が学生であればなおさらである。この点をふまえると、「友人同士の会話」といった対等な話者関係を無理に設定してその枠の中で話させるよりも、実際の話者関係に抵触しない設定を使うほうが、状況的に自然なやりとりが引き出せると考えられる。

ここで言う「自然さ」は、学習者コーパス構築におけるキーワードの 1 つである。Gilquin, De Cock, & Granger(2010)は、自身が構築した LINDSEI を「コーパス」ではなく「データベース」と呼んでいるが、この理由として、いわゆるコーパスが自然言語データ(natural language data)の集積と定義されるのに対し、LINDSEI のデータは厳密な意味で自然言語と言えない可能性があるからだ、としている(p.5)。一方で、彼女らは、LINDSEI が採用した 3 種のタスクに関して、少なくとも、トピック発話とそれに続く自由議論は自然データに近いものであると述べている。

None of these data qualify as fully natural as they were not produced for real communicative purposes. However, both the warming-up part and the discussion come close to the natural end of the continuums as the learners were free to choose both the ideas they wanted to express and the language in which to express them. (p.6)

話す内容とそれを伝える言い方の両方に自由度があることがデータの自然性と真正性の根拠であるとするとすれば、ロールプレイは、トピック発話や自由議論以上に「自然さの連続的段階の極点」に近接したものと言えるだろう。

なお、発話における「自然さ」は多面的な概念であるが、発話にかかる状況そのものの自然さ(状況的一般性)と、特定の状況の中での言語使用の自然さ(言語的適切性)は切り分けて理解されるべきである。ICNALE Spoken Dialogue のロールプレイは、後者の自然さを志向するものだが、前者の自然さを主張するものではない。

3.2 発話量の確保

既存の対話型学習者発話コーパスにおけるインタビュー時間は 15 分程度のものが多い。しかし、

LINDSEI の報告書では、同じ 15 分間のインタビューを行ったにもかかわらず、インタビューにおける学習者の平均発話量は、国により、728 語(日本)～1,862 語(ポーランド)まで、2 倍以上の開きが存在したとされる(p. 25)。

上記にも示されるように、各国学習者の中で、日本人の発話量の少なさは際立っている(独話における日本人学習者の流暢性については石川, 2016 参照)。LINDSEI では 1 か国につき原則 50 人分のデータが集められているわけであるが、1 人の発話量が 700 語しかないとなると、「日本人学習者」の対話について概括的な結論を導くことはきわめて難しくなる。

この点をふまえると、相対的に高い習熟度と流暢性が期待できる欧州圏学習者でなく、アジア圏学習者をターゲットとして発話データを収集する場合、15 分間のインタビューでは必ずしも十分でないということになる。そこで、ICNALE Spoken Dialogue では、1 つのインタビューの時間を 30 ～40 分と定めた。これにより、流暢性の低い学習者からも一定量の発話を引き出すことが可能になると考えられる。

3.3 統制性の確保

学習者コーパスのデータとなることを考えた場合、インタビューは、可能な限り、同一の手順・内容で行われることが望ましい。そこで、ICNALE Spoken Dialogue では、各国でインタビューを行う調査者の数を絞り(日本モジュールでは 1 名)、かつ、「導入→絵描写→ロールプレイ→絵描写→ロールプレイ→終結」という共通のプロトコルの下、学習者の習熟度レベルに関わらず、原則として全員に同様の指示・発問・タスクを与えることとした。

すでに述べたように、LINDSEI では、3 つのトピックから 1 つを選んで学習者に発話させ、NICT JLE Corpus では、イラスト 7 種、ロールプレイ 14 種、ストーリーテリング 9 種の中から、それぞれ 1 種を試験官が選んで学習者に発話させていた。つまり、同じインタビューデータであっても、その内容は人によりまちまちであったことになる。Ishikawa (2013) でも述べたように、内容の異なりは、語彙の異なりや使用される文法項目の異なりに直結するため、内容が統制されていないデータでは、学習者と母語話者間や、異なる母語を持つ学習者間で、信頼できる対照研究を行うことはきわめて困難になるだろう。この点をふまえ、ICNALE Spoken Dialogue は、ICNALE プロジェクト全体の基本理念を踏襲しつつ、すべての学習者にあえて同じ内容について話させることで、統制性の高いデータの収集を目指す。

3.4 相互比較性の確保

ICNALE プロジェクトのユニークな特徴は、トピックが「大学生にとってのアルバイトの是非」、「レストラン全面禁煙の是非」という 2 つに絞られていることである。これにより、ICNALE のモジュール間で多様な対照研究を行うことが可能になっている。

こうしたモジュール間での相互比較性をできる限り維持すべく、ICNALE Spoken Dialogue においても、「アルバイト」と「禁煙」という 2 つのテーマを基軸として、それぞれ、絵描写課題とロールプレイ課題を独自に設計することとした。

「アルバイト」について言うと、学習者は、まず、「海水浴に行きたいがお金がない大学生がアルバイトで資金をためて海水浴に行く」という一連の流れを示した 4 コマイラストの絵描写課題を行い、次に、「アルバイトをやめて勉強に励むよう忠告する指導教授にアルバイトの継続を認めてもらう」というロールプレイ課題を行う。さらに、その後のやりとりで「大学生にとってのアルバイトの是非」についての意見を述べる。また、「禁煙」については、はじめに、「公園でタバコを吸っている男性を女性が注意してやめさせる」という流れを示した 4 コマイラストの絵描写課題を行い、続いて、「喫煙を認めているレストランの店主に抗議して食事代金の返金を訴える」というロールプレイ課題を行う。さらに、その後のやりとりで「レストラン全面禁煙の是非」について意見を述べる。

こうしたタスク構成により、学習者の発話で使用される語彙や表現は、少なくとも部分的に、既存の ICNALE の各モジュールと重複するものとなり、同一条件下での相互比較が可能になる。

なお、モジュール間での相互比較を行おうとする場合、加えて重要なことは、学習者の習熟度や属性を揃えることである (Ishikawa, 2015b 参照)。この点をふまえ、ICNALE Spoken Dialogue では、ICNALE の他のモジュールと完全に同等の学習者属性調査を実施している。これにより、属性を統制したモジュール間比較 (たとえば、女子学生に限定した独話・対話比較など) も可能になる。

3.5 検証性の確保

学習者の発話分析において、しばしば議論のテーマとなるのが、(1)L2 発話の流暢性と、(2)L2 発話内容の妥当性である。ただ、これらの点を L2 発話コーパスのデータだけで厳密に議論することは予想以上に難しい。というのも、(1)に関しては、L1 では普通に話せるのに L2 においてのみ流暢性が阻害されているのか、あるいは、L1, L2 問わず、単に「無口」であるのかが判断できず、(2)に関しても、L1 ならきちんとした内容が話せるのに L2 だから内容が伝わらないのか、あるいは、L1, L2 問わず、そもそも論理的な主張ができていないのかが判断できないからである。

これらの問題を考えるヒントを得るべく、ICNALE Spoken Dialogue では、インタビューの最後に、5 分程度の母語による振り返り (reflection) を行うこととした。振り返りの中で、調査者は、個々のタスクに対する学習者自身の手ごたえを問い、また、学習者が誤りたり意味が曖昧になったりした点について、その理由や意図を学習者に確認する。さらに、学習者がインタビュー中に英語で述べた内容や、ロールプレイでの発言に関して日本語での補足を求める (例:「もし日本語なら他にどんなことを言えばよかったと思いますか」)。こうした一連の L1 のやりとりを通して、学習者の L1 における流暢性の程度や、主張の論理性や一貫性について確認することができる。

このうち、前者はとくに重要である。というのも、30～40 分にわたる英語インタビューの後で行われる L1 振り返りでは、通例、学習者は非常にリラックスして回答を行い、L2 流暢性の高低を議論する際のベースラインとなる L1 における流暢性がかなり正確に把握できるからである。筆者自身の過去の研究 (石川, 2016) も含め、ICNALE Spoken Monologue を用いた L2 流暢性の研究では、L1 流暢性を組み込んだ分析ができなかった。しかし、ICNALE Spoken Dialogue では、L1 振り返りのデータを使うことで、L1 と L2 の流暢性を比較し、学習者の流暢性をより厳密に分析す

ることが可能となる。

下記は、L1 と L2 における流暢性の現れ方のパタンを類型化したイメージである。

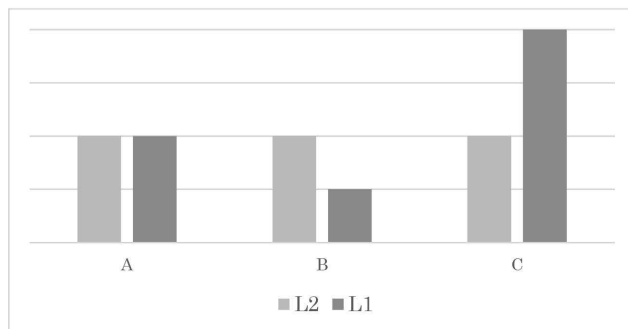


図 1 学習者の L1 および L2 における流暢性のパタン

上記において、A～C の学習者の L2 流暢性(発話量)はまったく同じであるが、その意味するところは大きく異なっている。たとえば、A の学習者は L1 と L2 の流暢性が同程度であり、L2 能力の不足による流暢性の阻害はほとんどない。B の学習者はむしろ L2 において流暢性が高まっている。一方、C の学習者は L2 において流暢性が大きく低下しており、L2 能力の不足による阻害が起こっている可能性が高い。このように、ICNALE Spoken Dialogue の L1 振り返りは、L2 流暢性を新たな視点から検証するための基礎データを提供するものとなる。

3.6 ELF 視点の確保

従来、オーラルインタビューを実施するにあたっては、目標言語の母語話者が調査者(インタビュワー)を務めるのが一般的であった。LINDSEI では 11 か国中 7 か国で母語話者調査者率が 75%以上である(p.37)。しかし、ICNALE Spoken Dialogue では、一部の例外を除き、学習者と同じ母語を持つ L2 英語教員が調査者を務めることとしている。これは、理想的には、英語を母語話者の専有物とみなすのではなく、非母語話者間の国際コミュニケーションツールとして位置付ける English as a Lingua Franca (ELF) の考え方に則ったものである(Seidhofer, 2011)。アジア圏における英語使用の現状をふまえれば、英語を ELF として位置付けることはより現実に近いと言える。

非母語話者教員が調査者となることで、学習者は過度の緊張感を持たずにやりとりに臨むことができ、質問の英語が聞き取れないので話せないといった事態も減少する。また、調査者が学習者の母語を使えることで、目標言語によるインタビュー部分と学習者母語による振り返り部分をシームレスに行うことが可能になる。

ICNALE Spoken Dialogue の日本モジュールの構築に関しては、1 名の非母語話者 L2 教

員が100名すべてのインタビューを行った。一般に、調査者の数が増えれば、調査者の性別・L1・外国語知識・立場の違いにより、それぞれのインタビューが異なったものになってしまう可能性がある (Gilquin, De Cock, & Granger, 2010, p.7)。この意味で、調査者の数を減らすことは、データの統制性の向上にも寄与すると考えられる。

3.7 匿名性の確保

コーパスを構築する場合、協力者の匿名性を守るとはさきわめて重要である (McEnery & Hardie, 2012, pp. 60-64)。この点をふまえ、ICNALE プロジェクトでは、参加者の匿名性の確保に可能な限りの注意を払っており、過去の ICNALE Spoken Monologue の開発では発話音声のピッチやフォルマントの人為的変更を行った。

今回の収集対象となる対話では、フィラーや相槌を含め、微妙な音声的差異が重要な談話的意味を持つことがあるため、音声変換は行っていないが、参加者には指定のマスクをつけてインタビューに参加することを求め、動画を公開しても個人が特定されないような配慮を行った。

従来、動画をコーパスデータとして公開する場合には、画像にシャドウをかけたり (shadowing)、ぶれを加えたり (blurring)、モザイク化 (pixellating) したりして発話者の顔が特定されないように加工してきたわけであるが、大量のデータに対してこうした処理を後から加えていくことは困難であり、また、表情や顔の動きといった重要な分析データを棄損することともなる (Adolphs & Carter, 2013, p. 11)。この意味において、はじめからマスクをつけてインタビューを受けさせることは、匿名性を確保する合理的な手段であると言える。

すべての参加者には以上の詳細を説明した上で承諾書に署名を求め、将来のデータ公開に際して問題が起こらないよう適切な対応を取った。

3.8 マルチモーダル性の確保

近年の発話コーパス開発では、音声データ (およびそれを書き起こし文字データ) だけでなく、動画データをあわせて記録することで、言語手段と非言語手段を組み合わせたマルチモーダル分析を行う可能性が模索され始めている。

Adolphs & Carter (2013) は、話し言葉のやりとりをビデオに録画することが、音声のみの記録に代わる「重要な代替法」になりつつあるとした上で、そうしたデータを集めることでデータ分析の範囲が拡張すると述べている (p. 8)。また、発話中に見られる顔の動き (傾きなど) や手の動きは、それ自身で、コミュニケーション上の意味機能を持ちうるものであるとして、それらを“headtalk”および“handtalk”と呼んでいる (p. 158)。

こうした近年の研究の潮流をふまえ、ICNALE Spoken Dialogue の日本モジュールでは、すべてのインタビューを音声と動画で記録しており、その両方を公開予定である。動画データについては、行動科学やスポーツ科学の分野で普及しつつある運動分析ソフトウェアを

使用して解析を行うことで、書き起しの文字データに加え、“headtalk”や“handtalk”の回数や大きさを計量化して分析することが可能になる。

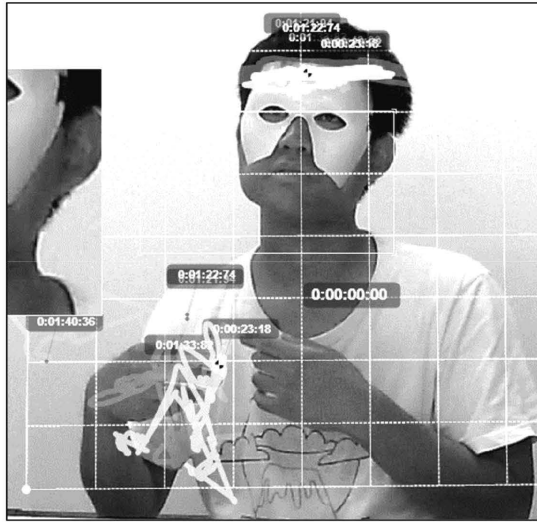


図 2 ICNALE Spoken Dialogue 収録データの運動解析

図 2 は、額の中央と手の先にトラッキングポイントを設定し、約 90 秒間の動きをそれぞれ記録したものである（図中の白い線の部分が運動の軌跡を示す）。現在、数名の学習者データをサンプルとして同様の分析を行っているが、学習者により、発話時の顔や手の運動量には驚くほど大きな差があることが明らかになった。こうした「非言語流暢性」のデータが入手できれば、前述の L1 流暢性および L2 流暢性のデータと合わせ、従来にない精緻な流暢性研究が行えるだろう。

4. ICNALE Spoken Dialogue のインタビュープロトコル

以上の基本理念に沿って、インタビュープロトコルの開発を行った。以下はその概要を日本語でまとめたものである（実際のプロトコルでは調査者が使用する英語表現についても具体的な指定がある）。なお、全体の時間とステージごとの時間について目安は設定しているが、とくに習熟度の低い学習者の場合、発話にきわめて長い時間がかかることが少なくないため、時間配分については、その都度のインタビューの実態をふまえて、調査者の裁量を最大限に認めることとしている。

表 5 ICNALE Spoken Dialogue におけるオーラルインタビューの流れ

ステージ	概要
1 導入会話	1-1 英語で話すのが好きか、それはなぜか 1-2 一週間にどの程度英語で話す機会があるか 1-3 英語で話す機会を増やしたいか、その場合どんな内容を話したいか 1-4 英語の発話力を高めるためにどんな工夫をしているか 1-5 英語で話す際、一対一の対話とグループ対話ではどちらが好きか、なぜか 1-6 英語の 4 技能の中でどれが一番大事だと思うか、それはなぜか
2 絵描写 A	2-0 課題指示 2-1 一連のイラスト(大学生が電気店でアルバイトして資金を貯めて海水浴に行く)を描写する 2-2 泳ぐのは好きか、なぜか 2-3 海水浴の経験はあるか、どんな思い出があるか 2-4 Windows と Mac のどちらを持っているか、なぜそれを選んだのか 2-5 Windows と Mac の最大の違いは何だと思うか 2-6 「最近の学生がスマートフォンばかり使用してパソコンが使えないのは問題だ」という意見についてどう思うか
3 ロールプレイ A	3_0 課題指示 3-1 バイトを辞めろという教授に反論し、教授を説得する 3-2 今やっている(今後やりたい)バイトはどんなものか、多くのバイトの中でその仕事を選んだ理由は 3-3 家庭教師と塾講師ならどちらが好きか、なぜか 3-4 「大学生にとってアルバイト重要だ」という意見に賛成か反対か
4 絵描写 B	4-1 一連のイラスト(子どもと公園に来ている母親が近くでタバコを吸っている男性に注意して止めさせる)を描写する 4-2 公園に行くのは好きか、なぜか、公園では何をするか 4-3 一か月にどのぐらい公園に行くか 4-4 好きな公園を一つ選んで紹介する 4-5 イラストのような女性は好きか、なぜそう思うか 4-6 「公園の砂場は不潔なので子供を遊ばせるべきでない」という意見についてどう思うか

5 ロールプレイ B	5-1 喫煙を認めているレストラン店主に対して食事代の返金を要求する
	5-2 レストランでの良い体験としてどんなことがあるか
	5-3 日本料理レストラン、中華レストラン等で好きなものはどれか、なぜか
	5-4 「レストランでは全面禁煙すべきだ」という意見に賛成か反対か
6 総括発話	6-1 今日の出来栄えはどうだったか
	6-2 緊張したか、緊張をほぐすにはどうすればよいか
	6-3 アルバイトと禁煙とどちらのトピックが話しやすかったか、なぜか
	6-4 絵描写とロールプレイのどちらが好きか、なぜか
7 L1 振り返り	7-1 絵描写 A について難しかったコマはどれか、なぜか、過去形を使うことを意識していたか
	7-2 ロールプレイ A について出来栄えはどうか、日本語ならどう説得するか
	7-3 「大学生にとってアルバイトは重要だ」というテーマについての意見として日本語で補足することはあるか
	7-4 絵描写 B について難しかったコマはどれか、なぜか
	7-5 ロールプレイ B について出来栄えはどうか、日本語ならどう交渉するか
	7-6 「レストランでは全面禁煙すべきだ」というテーマについての意見として日本語で補足することはあるか
	7-7 全体を振り返ってどんな感想を持ったか

5. ICNALE Spoken Dialogue の日本モジュールの構築

プロジェクトの初年度にあたる 2017 年度においては、日本語母語話者 100 名のデータ収集を行い、インタビュー本体部分、L1 振り返り部分ともに、書き起こしを実施した。

書き起こされたテキストデータは、表計算ソフト上に、センテンス単位で記録されている。これにより、学習者発話と調査者発話を区別して取り出したり、タスクごとにデータを取り出したり、また、必要に応じてデータを並べ替えたりすることが可能となっている。

下記は、B1 低位レベルの日本人学習者 (No_001) によるロールプレイ A の発話の一部である。冒頭の数値コードは、タスク番号 (3_1) とタスク内でのセンテンスの連番 (_6, _7...) を示す。発話文の前に付けられた [T] と [E] は、それぞれ教師 (teacher) と学習者 (examinee) を示す。

表 6 ロールプレイ A の書き起こしデータの一部

コード	発話者と発話文
3_1_6	[E] I have to earn money to play with my friend.
3_1_7	[E] So, I have to work.
3_1_8	[T] All right, I understand.
3_1_9	[T] But maybe, you have entered university to study, not to play.

3_1_10	[T] So, that's why I tell you not to work.
3_1_11	[T] You have to study, because you are a college student.
3_1_12	[E] But I think we have to do something in university, not only study.
3_1_13	[E] If I do a part-time job, I can experience a lot of things.
3_1_14	[T] But maybe I – I think that many students have part-time jobs and then, they don't have enough time to do, to study, right.
3_1_15	[T] I think this is a very bad thing.
3_1_16	[T] So, that's why actually, I want you to be a good scholar, right, a good researcher.
3_1_17	[T] So, that's why I want you to spend more time for studying.
3_1_18	[T] Please stop working now.
3_1_19	[E] But I have – to study, we need notebook.
3_1_20	[T] Notebooks?
3_1_21	[T] It costs just only ¥100.
3_1_22	[T] Maybe you don't need to work for so, you know, long time.
3_1_23	[E] But if I only study, I am so tired.
3_1_24	[E] But we – thanks to doing part-time job, I can enjoy and concentrate more on study.
3_1_25	[T] Oh, I see, all right.
3_1_26	[T] So, what – what part-time job do you do now?
3_1_27	[E] I am teaching children.
3_1_28	[T] Okay, so do you think it contributes to your study, to your research?
3_1_29	[E] Yes.
3_1_30	[T] Okay, all right.

やり取りを見ると、この学習者が、アルバイトを続けたい理由を自分なりに考え(19/23-24/27 など)、最後まであきらめず、なんとか教授を説得しようとしている様子がうかがえる。コーパス研究では、データを集約して量的に処理する傾向が強いが、上記のようなデータは質的な分析にも応用できるものである。

一般に、オーラルインタビューでは、学習者が調査者の質問を理解できなかったり(理解の障害)、言いたいことをうまく言えなかったりする(伝達の障害)ことがしばしば起こる。理解の障害が起こると、調査者の側では、自分で気づいて質問を繰り返したり(自己開始修復)、学習者の無言に促されて再質問を行ったりして(他者開始修復)(Kasper & Ross, 2007; NICT JLE Corpus を使った分析例として谷村, 2013 も参照)、また、学習者の側では、主体的に質問を行って相手から情報を引き出す交渉的意味理解(negotiation of meaning)を行ったりして、問題の解消が図られる。また、伝達の障害が起こった場合も、調査者の側で助け舟を出したり、学習者の側で各種の方

略使用を行ったりして(婉曲法, 言い換え, 上位語・同位語・下位語による置き換え) (Ellis & Barkhuizen, 2005, pp. 165-228), 同様に問題の解決が図られる。

こうしたコミュニケーション障害の場面において, 学習者の側で主体的にどのような障害解決の手立てが取られているかを見ることは, 学習者の L2 運用能力を議論する上できわめて重要である。ICNALE Spoken Dialogue で収集される対話データは, マクロ的な観点に基づく計量分析だけでなく, 従来, 対話分析 (interactional analysis) や会話分析 (conversational analysis) の枠組みで行われてきたミクロ的な質的分析にも有用である。

ところで, 上記の引用部にも見られるように, この学習者は, 他の学習者と比べて, 流暢性がきわめて低い。インタビュー全体における学習者および質問者の発話量は以下の通りである。

表 7 No_001 のオーラルインタビューにおける学習者・質問者の発話量

発話者	センテンス数	トークン数
学習者	147	901
調査者	341	2968

学習者の発話量は約 900 語で, 質問者の発話量の約 3 分の 1 にとどまっているが, それでも, LINDSEI の 15 分間インタビューにおける日本人の平均発話量(728 語)よりは多い。従来の対話型学習者発話コーパスに比べてより長いインタビュー時間を確保することで, 流暢性の低い学習者からも一定量の発話を引き出せていることがわかる。

下記は, 同じ学習者による, ロールプレイ A に関する L1 振り返り時の発話の一部である。

表 8 L1 振り返りの書き起こしデータの一部

コード	発話者と発話文
7_2_1	[T] ぜ, 全体的に。はい。OK, OK。はい。じゃあ次, あの, えっとその, アルパイトのロールプレイのカード出していただいて, それだね。教授, 自分のまあなんか指導教授との会話だったんだけど, あのー, まあ私, 教授役をやったんだけど, やってみて, なんかもその, これ, 言いたかったけどうまく言えなかったこととか, 本当はこういうことを言うつもりだったけど, フフ, 英語が思い付かなかったこととかありました?
7_2_2	[E] なんか, 社会勉強にもなる。
7_2_3	[T] それは難しいな。
7_2_4	[E] フフ, フフ。
7_2_5	[T] 私も言えないわ。
7_2_6	[E] 言えない。フフフ。
7_2_7	[T] 言えないな。他は? うん。
7_2_8	[E] あ, あとはなんか, 一人暮らしをしてて, その, 自分の使うお金を, 自分で稼がなきゃいけないみたいなことを, 言おうと, フフフ。

7_2_9	[T] ああ。あ、そう言ってくれたら良かった、アハハハ。
7_2_10	[E] うまく、言えません、フフフ。
7_2_11	[T] ハハハハ。だよな。だから play とか言ったんだね。遊ぶお金が要るとか言っちゃったんだよな。はい、分かりました。えっと、じゃあそのロールプレイ、もし、その、本当にこれ日本語でさ、教授とやり合って、教授に何とか納得してもらおうというシチュエーションだとすると、日本語だったらまあ、こう言ったら、教授、納得してくれたんじゃないかなと思えることとかあった？
	[E] あ、やっぱり、勉強だけじゃなくて、その、アルバイトすることで、いろんな人と会ったりして。
7_2_12	[T] うん、ハハハ。
7_2_13	[E] (unclear) 勉強にもなると。フフフ。
7_2_14	[T] ハハハ。ああ、なるほど。まあ教授としてアドバイスすると、その説得は効かないと思うな。ハハ。多分、そのなんか、家、私生活が苦しいとかって言ってくれたほうが、仕方ないなと思うけど。

コーパスに収録された他の学習者の L1 発話と比較してみれば一目瞭然であるが、この学習者は、L1 においても発話がきわめて少ない。3.5 節で挙げたパターンで言うと、この学習者は、一見、C 型 (L2 でのみ流暢性が低い) に見えるが、L1 データを見ることで、実際は A 型 (L1, L2 ともに同程度の流暢性) であった可能性が高いことがわかる。

6. まとめ

ICNALE Spoken Dialogue は、2020 年中の全面公開を目指し、今後、海外学習者のデータや母語話者データの収集を段階的に進め、あわせて、収集したデータの保存・公開方法の検討や、オンライン検索システムの開発にも着手する予定である。

LINDSEI や NICT JLE Corpus をはじめとする各種の先駆的な学習者コーパス研究の成果を継承しつつ、新たな視点と観点を盛り込んで設計された ICNALE Spoken Dialogue の完成により、ICNALE の既存モジュールとの連携分析ができるようになれば、アジア圏学習者の母語 (国籍) や習熟度、また、L2 の産出モード (書き言葉 / 話し言葉) や発信モード (独話 / 対話) といった多様なパラメタを組み合わせた、従来にない精緻な中間言語対照分析の可能性が一気に広がることとなる。

今後、ICNALE Spoken Dialogue で収集されたデータを基盤として、対話場面での学習者の L2 使用に関する新たな知見の蓄積や、学習者の L2 流暢性を論じる新たな理論的モデルの構築が進んでゆくことを強く期待したい。

謝辞

本研究は科学研究費基盤(B)「アジア圏英語学習者自然対話コーパス ICNALE-Dialogue 開発と分析(17H02360)」(代表者:石川慎一郎, 研究期間:2017~2019 年度)によるプロジェクトの成果の一部である。本研究への継続的な支援に深謝したい。

引用文献

- Adolphs, S., & Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal*. London, UK: Routledge.
- Csépes I. (2009). *Measuring oral proficiency through paired-task performance*. Frankfurt am Main, Germany: Peter Lang.
- De Cock, S. (2010). Spoken learner corpora and EFL teaching. In M. C. Campoy-Cubillo, B. Belles-Fortuno, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 123-137). London, UK: Continuum.
- Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford, UK: Oxford University Press.
- Gilquin, G., De Cock, S., & Granger, S. (2010). *Louvain international database of spoken English interlanguage*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast* (pp. 37-51). Lund, Sweden: Lund University Press.
- 石川慎一郎 (2012).『ベーシックコーパス言語学』東京:ひつじ書房.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1* (pp. 91-118). Kobe, Japan: Kobe University.
- Ishikawa, S. (2014). Design of the ICNALE-Spoken: A new database for multi-modal contrastive interlanguage analysis. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 2* (pp. 63-76). Kobe, Japan: Kobe University.
- Ishikawa, S. (2015a). A consideration of the difference between the spoken and written English of native speakers and Japanese learners: A corpus-based study. *Discourse and Interaction, 8*(1), 37-52.
- Ishikawa, S. (2015b). The influences of learners' basic attributes and learning histories on L2 speech fluency: A case study of Japanese and Chinese learners of English. *Procedia: Social and Behavioral Sciences, 192*, 516-525.
- 石川慎一郎 (2015a). 「学習者コーパスⅡ:国内における英語学習者コーパスの開発と研究」. 投野由紀夫(編)『コーパスと英語教育』(pp.99-129). 東京:ひつじ書房.

- 石川慎一郎 (2015b). 「The ICNALE: 国際中間言語対照分析研究のための新たな学習者コーパスの開発」『電子情報通信学会技術研究報告: 信学技報』115 (361), 13-18.
- 石川慎一郎 (2016). 「日本人学習者の L2 英語の発話量: 母語話者及びアジア圏学習者との比較」『日英言語文化研究』(日英言語文化学会) 5, 15-26.
- Ishikawa, S. (in press). The ICNALE edited essays: A dataset for analysis of L2 English learner essays from a new integrative viewpoint. *English Corpus Studies*, 25.
- 和泉絵美・内元清貴・井佐原均 (2004). 『日本人 1200 人の英語スピーキングコーパス』東京: アルク.
- Kasper, G., & Ross, S. (2007). Multiple questions in oral proficiency interviews. *Journal of Pragmatics*, 39, 2045-2070.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge, UK: Cambridge University Press.
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford, UK: Oxford University Press.
- 谷村緑 (2013). 「研究例・1: スピーキングテストに見られる対話の修復」 投野由紀夫・金子朝子・杉浦正利・和泉絵美(編)『英語学習者コーパス活用ハンドブック』(pp. 169-177). 東京: 大修館書店.