

データベース準備班の活動について

遠藤藍子・鎌田美千子・菅谷奈津恵

要旨

データベース準備班は、OPI テープをデータベース化する際に留意すべき点について整理・検討することを目的として発足した。2004 年 3 月の立ち上げ後、約 2 年にわたり、技術面、倫理面から討議を行った。その結果、データベース化にあたっては、インフォームド・コンセントの保証、プライバシー尊重、秘密保持といった倫理面を考慮することが非常に重要であることがわかった。

【キーワード】 インフォームド・コンセント、同意書、個人情報、書き出し用フォーム、学習者発話コーパス

1. 活動の概要

1-1 発足の経緯と会合記録

データベース準備班(以下、DB 班)は、OPI 研究会で蓄積されてきたインタビュー・テープを有効活用すべく、2004 年 3 月に発足した。データベースの作成そのものを担当するのではなく、データベース化する際に留意すべき点を整理・検討することを目的としている。主な活動内容は、以下の通りである。

・ 2004 年 3 月 6 日

OPI 研究会定例会において、DB 班の立ち上げが承認された。

・ 2004 年 3 月 26 日 第 1 回 DB 班会合

データベースの作成目的について検討し、今後の作業分担を決定した。

・ 2004 年 9 月 6 日 第 2 回 DB 班会合

既存コーパスの特徴について情報交換を行った。インタビュー内容を整理する方法として、トピックの書き出し用フォームの試作も行った。

・ 2004 年 11 月 27 日 第 3 回 DB 班会合

コーパスの文字データ、音声データの処理方法について検討した。

・ 2005 年 2 月 19 日 第 4 回 DB 班会合

データベース化にあたって、著作権や匿名性、秘密の保持の問題をどうクリアするか討議を行った。

・ 2005 年 3 月 21 日 第 5 回 DB 班会合

国際シンポジウムでの発表内容と、今後の作業について、討議を行った。

・ 2005 年 7 月 31 日 第 4 回 OPI 国際シンポジウム

調査結果をまとめ、口頭発表を行った。

上記の他に、メールを活用して、3 人のメンバーで意見交換を行ってきた。また、国際シンポジウムの発表内容については、その際に受けたフィードバックを踏まえて改稿をし、

シンポジウム報告・論文集への投稿を行った。

1-2 研究成果の発表

以下のように、口頭発表と論文発表を行った。

・口頭発表

遠藤藍子・鎌田美千子・菅谷奈津恵（2005）「日本語 OPI のデータベース化をめぐる諸問題」第 4 回 OPI 国際シンポジウム、2005 年 7 月 31 日、函館国際ホテル

・研究論文

遠藤藍子・鎌田美千子・菅谷奈津恵（2005）「日本語 OPI のデータベース化をめぐる諸問題」『日本語・日本文化講座夏期セミナー20 周年記念シンポジウム&第 4 回 OPI 国際シンポジウム予稿集』54-57.

遠藤藍子・鎌田美千子・菅谷奈津恵（2006）「日本語 OPI のデータベース化をめぐる諸問題」『日本語・日本文化講座夏期セミナー20 周年記念シンポジウム&第 4 回 OPI 国際シンポジウム報告・論文集』99-106.

2. 研究成果の概要

ここでは、遠藤・鎌田・菅谷（2006）の報告から、研究成果の概要を紹介する。

当班では、OPI テスターによってそれまでに蓄積されてきた、あるいは将来にわたって蓄積されていくインタビューのテープ（以下、OPI テープ）のデータベース化を検討するに当たり、その基本方針として、研究者だけでなく、テスターや現場の日本語教師にも資する方向を考えて活動を行ってきた。そのためにも、音声資料と文字資料の両面からのデータ化が必要であると考え、既存コーパスのデータ処理方法、文字化データの検索方法、コーパスを活用した言語研究事例などの検討を行った。

そこで、まずデータベース化を検討するに当たり、話し言葉コーパスに焦点を当て、その方式、公開方法をリストアップし検討した⁽¹⁾。その結果、既存の方式の多くは文字起こしデータであること、発話内容の正確な理解には音声データが重要であるが、その公開には個人特定化の問題が絡むことが示された。

一方、OPI テープのデータベース化された時の利用者と利用目的を考えると、研究者が資料とする場合でも、データ資料提供者である OPI テスターがインタビュー技術の向上を目指す場合でも、日本語教師が学習者のスピーチレベル向上のための話題や指導方法の検討をする場合でも、データベースから必要情報が取り出し易いことが望まれる。そこで、それら利用者の便を考慮した「書き出し用フォーム」を提案した。フォームは、簡便かつ視覚的にインタビューの全体像を俯瞰できるよう、データ内容の概略を整理して提示する形になっている。内容は、OPI の 4 つの評価指標の内、①タスク・機能、②場面・話題の 2 事項がベースで、インタビューの質問をテスト全体の流れの中に位置付けながら概観し検索できる⁽²⁾。

しかしながら、調査を進める過程で、技術面以上に倫理面において考慮すべき重要な課

題があることが明らかになった。柳町(2006)の主張によれば、調査者は、調査研究を行う際、データ提供者であるインフォーマント（データ提供者）の権利を尊重し、利益の保護に努める義務を負っているという。すなわち、データ収集に当たっては、インフォーマントのプライバシーやその他の権利を侵害したり、尊厳の喪失をしたりしないようにすべきだということである。柳町は、日本語教育の研究者にとって最も身近で切実な問題として次の2点への配慮の必要性を指摘している。

- 1) 匿名性の問題。すなわち研究を進める上で得られた個人情報の取り扱い
- 2) インフォームド・コンセントに関する問題。すなわちインフォーマントに対し、十分な説明と自発的な同意を得ること

OPI のデータベース化において検討してみると、1) については、文字化資料ならば個人が特定される部分は削除すればよいので、匿名性の保持が可能である。だが、音声資料については、声によって人物の同定が可能であるため、匿名性の保持は困難である。2) については、発話データを一般に公開するのであれば、プライバシーの保護を尊重して、テスト前にインフォームド・コンセントを得ることが必要である。しかし、OPI は、そのテストの性格上、身の回りのことや個人的なこと、例えば趣味、嗜好から出発し、個人的なこと、すなわち体験、信念、意見などを問うため、どうしてもインタビューイーにプライバシーを語らせざるを得ない。それが、インフォームド・コンセントによってインタビューイーが発話内容を差し控えてしまった場合、レベル判定に影響を与える恐れが生じる。

以上に加え、OPI 独自の問題として、秘密保持の問題が絡んでくる。それは、OPI は本来「能力テスト」であるという点である。そこから、インタビューの過程や判定結果は、テスターが職務遂行上知り得た秘密の情報と考えられる。したがって、当然テスターには、その情報を守秘する義務が生じてくるはずである。データベースによる情報公開は、テスターの守秘義務違反という道徳的な責任問題につながる恐れがあり、慎重な取り扱いが必要である。

以上、倫理的課題という面から見た場合、OPI のデータベース化には種々困難な問題が存在していることが明らかになった。ただし、当班では次のような措置をとることによって、問題点を克服することは可能と考えた。

- 1) 利用目的をテスターの養成や技術向上の訓練に限定する
- 2) データベースは一般には非公開とする(つまり、OPI 研究会会員のみへの公開とする)
- 3) 上記の1、2を盛り込んだ「同意書」を事前に取り交わす

以上から、当班では、この3点を、OPI テープのデータベース化の実現に向けての対応策として提案した。

3. まとめ

本稿では、約2年間のDB班の活動とその成果について報告した。

OPI テープのデータベース化においては、インフォームド・コンセントの保証、プライバシー尊重と秘密保持といった倫理面を考慮し、慎重に進めなければならないことを述べた。さらに、その対応策として、利用目的や公開の範囲を限定することと、その旨を盛り込んだ「同意書」を事前に取り交わすことを提案した。日本語 OPI 研究会としてデータベースを作成する際には、具体的な作業に入る前に、こうした倫理的な問題を解決することが必要である。

注

- (1) 筆者らが検討を行った際に公開されていた日本語学習者の発話コーパスには、文字起こし方式では「KY コーパス」(カッケンブッシュ 1999)、「就労ブラジル人の発話文字化資料」(尾崎 2001)、「日本語研修生追跡調査」(村上 2001)、「BTS による多言語話し言葉コーパス：日本語会話 2 (日本人と学習者の会話)」(宇佐美 2005)、音声データと文字起こしの併用方式では「上村コーパス」(上村 1997)、公開予定のものとして「発話対照 DB」(宇佐美・籠宮・楢本 2004)があった。
- (2) OPI の評価基準の 4 つの柱である ①タスク・機能、②場面・話題、③テキストの型、④正確さの内、①②はテストのコントロール下にある事項で、インタビューの方向性の決め手となるが、③④の中身はテスト次第で、インタビューの進め方や流れを検索する上では不用との判断から「書き出しフォーム」からは除いてある。「書き出しフォーム」は省略。

参考文献

- 上村隆一(1997)「データベースで調べる」『日本語学』11月号, 60-68.
- 宇佐美まゆみ(1997)「基本的な文字化の原則 (Basic Transcription System for Japanese : BTSJ)」西郡仁朗 (研究代表)『日本人の談話行動のスク립ト・ストラテジーの研究とマルチメディア教材の試作』平成7-8年度科学研究費研究成果報告書, 12-26.
- 宇佐美まゆみ(2003)『多文化共生社会における異文化コミュニケーション教育のための基礎的研究』13-14年度科学研究費研究成果報告書, 4-21.
- 宇佐美まゆみ監修(2005)『BTS による多言語話し言葉コーパス：日本語会話 2 (日本人と学習者の会話)』CD-ROM
- 宇佐美洋・籠宮隆之・楢本総子(2004)「日本語学習者による日本語/母語発話の対照言語データベース」の設計『電子情報通信学会技術研究報告』104, 29-34.
- 遠藤藍子・鎌田美千子・菅谷奈津恵(2006)「日本語 OPI のデータベース化をめぐる諸問題」『日本語・日本文化講座夏期セミナー20周年記念シンポジウム&第4回 OPI 国際シンポジウム報告・論文集』99-106.
- 尾崎明人(2001)「就労ブラジル人の発話文字化資料」『日本語電子化資料収集・作成：コーパスに基づく日本語研究と日本語教育への応用を目指して』名古屋大学教育研究改革・改善プロジェクト報告書 (<http://prairie.lang.nagoya-u.ac.jp/J-corpus> より電子的に入手)
- カッケンブッシュ寛子 (研究代表) (1999)『第二言語としての日本語の習得に関する総合

- 研究』平成 8・10 年度科学研究費研究成果報告書
- 小磯花絵(2003)「コーパスによる音声談話の研究」『日本語学』22, 200-209.
- 阪本昌成(1986)『プライバシー権論』日本評論社
- 西郡仁朗(2002)「自然会話データ『偶然の初対面の会話』その方法論について」『人文学報』330, 東京都立大学人文学, 1-18.
- 藤原静雄(1997)「個人データの保護」桂木隆夫他『岩波講座 現代の法 10 情報と法』岩波書店
- 村上京子(2001)「日本語研修生追跡調査概要」『日本語電子化資料収集・作成：コーパスに基づく日本語研究と日本語教育への応用を目指して』名古屋大学教育研究改革・改善プロジェクト報告書 (<http://prairie.lang.nagoya-u.ac.jp/J-corpora> より電子的に入手)
- 柳町智治(2006)「言語コーパスの構築と使用をめぐる倫理的問題：インフォームド・コンセントを中心に」『日本語教育』130, 32-41.
- ルース, R. フェイドン・トム, L. ビーチャム(1994)『インフォームド・コンセント：患者の選択』酒井忠明・秦洋一訳, みすず書房
- American Psychological Association (1992) *Ethical principles of psychologists and code of conduct*. (富田正利・深澤道子訳 1996 年 日本心理学会)
- Mari Korkeaa-aho (1999) *Anonymity and Privacy in the Electronic World*, Department of Computer Science, Helsinki University of Technology.
(<http://users.tkk.fi/~mkorkeaa/doc/anonpriv.html#chap2.1> より電子的に入手)