

「日本・韓国・台湾の大学生による日本語意見文データベース」ご使用に際して

本データベースは、日本語を母語とする大学生（134名）と日本語を学ぶ大学生（台湾57名、韓国55名）が日本語で執筆した意見文を収録したもので、教育及び研究目的で利用可能な資料として公開しました。日本語学習者データには、作文執筆者本人による母語への翻訳（意識）も含まれます。データの収集に際しては、国立国語研究所（2001/2004）「日本語学習者による日本語作文と、その母語訳との対訳データベース ver.2. CD-ROM版/同オンライン版」を参照しました。

1 データベースのご利用に際して

(1) 本データベースは、平成19～22年度文部科学省科学研究費若手研究（B）（研究代表者：伊集院郁子、課題番号：19720119）による研究成果の一部です。本データベースを使用して研究を発表される場合は、「日本・韓国・台湾の大学生による日本語意見文データベース」（2011年3月）を利用した旨、明記してください。

(2) データをお使いの上、研究発表等を行われた際は、以下にお知らせください。

連絡先：ijuin@tufs.ac.jp

（東京外国語大学留学生日本語教育センター 伊集院郁子）

(3) データは2011年3月30日時点のもので、今後、入力の際の誤りなどが発見された際は、随時訂正が加えられることをご了承ください。訂正が加えられた場合は、ホームページで訂正箇所をお知らせします。

2 データベースの構成

本データベースは以下の4つのデータで構成されています。

- ①検索ファイル（エクセル形式）
- ②全作文1文1行ファイル（エクセル形式）
- ③作文pdfファイル（執筆者の母語別に分類）
- ④作文テキストファイル（執筆者の母語別に分類）

①検索ファイル

このファイルから、③のpdfファイル及び④のテキストファイルにリンクが張られていて、手書きのままの作文（pdf形式）及びテキスト化された作文（txt形式）を見ることができます。検索ファイルは、次の4つのシートで構成されています。

- a JP（日本語母語話者）の作文へのリンク及び執筆者情報
- b KR（韓国学習者）の作文へのリンク及び執筆者情報
- c TM（台湾学習者）の作文へのリンク及び執筆者情報
- d 凡例

a から c の執筆者情報として、以下の情報が記載されています。

JP の執筆者情報

専門：文系・理系の別（及び専攻）／学年／性別／年齢／出身地／3年以上の海外滞在歴／データ収集時期

KR 及び TM の執筆者情報

性別／出身地／SPOT（日本語能力測定簡易テスト） ver. 2 の結果／日本語能力試験の合格級／年齢／その他の言語（KR は母語・日本語以外で日常会話レベルの使用ができる言語、TM は第一言語と第二言語）／日本語学習歴／日本滞在歴／日本以外の3年以上の海外滞在歴／データ収集時期

d の凡例には、以下のような記号の説明があります。

- 日本語作文の段落開始を意味する（あるいは原稿用紙上、空白とされたマスを表す）
- 日本語作文の段落終了を意味する
- L 日本語作文の作文終了を意味する
- # 原文にかかわる特記事項（例：ルビ有り）
- * 入力不可能な文字（画数が足りない、多すぎる、中国語の繁体字や簡体字など）が書かれていたことを意味する。（原稿に書かれていた文字に最も近い文字が入力されている。）
- ※ *の付された漢字に関する説明、またはJPデータの漢字や文法の誤用に関するコメント（KR/TMデータの誤用に関してはコメントなし）
- ◆ 全く入力不可能、判読不可能な文字を表す
- 母語訳作文の段落開始を意味する
- 母語訳作文の段落終了を意味する

②全作文1文1行ファイル

本データベースの全作文を1文レベルに分解し、1文を1行に入れて入力し直したものです。エクセルのオートフィルタ機能などを利用した研究に適しています。研究の便宜上、多少の加工がなされています。詳細は、「資料」にある「全文1文1行ファイル作成時の覚書」をお読みください。

③作文 pdf ファイル

執筆者による手書きの作文をそのままスキャナーで取り込んだ画像データです。

④作文テキストファイル

③の作文 pdf ファイルを「入力マニュアル」に従い、テキスト化したものです。詳細は、「資料」にある「入力マニュアル」をお読みください。

3. 執筆者及びデータ収集に関する情報

2011年3月末時点での作文執筆者の人数、性別、平均年齢は、表1のとおりです。JPは東京都内の大学に通う日本人大学生、KRは韓国、TMは台湾の大学に通う日本語学習者で、JPは3つの大学、TMは1大学、KRは2大学から調査対象者を募り、全ての応募者を調査対象としました。執筆者は、同一の課題文を読み、辞書などは使用せずに約60分で原稿用紙1枚に800字程度で執筆しました。

KRおよびTMに関しては、日本語能力試験2級（学習時間600時間相当）以上の日本語学習者を対象とし、実際の日本語能力を測る参考データとしてSPOTも実施しました。（SPOTは、筑波大学留学生センターで開発された日本語能力簡易試験です。日本語能力試験の1級レベルと2級レベルを識別するために、音声の聞き取りが難しいver.2を使用しました。）また、母語の影響が論じられるよう、日本語で執筆した後で、同じ内容の意見文を母語でも執筆してもらいました。

表1 執筆者情報

執筆者	人数	性別		平均年齢
		男性	女性	
JP(日本)	134名	90名	44名	19.4歳
KR(韓国)	57名	14名	43名	21.1歳
TM(台湾)	55名	9名	46名	22.2歳

4. 収録データに関する情報

作文収録数及びタイトル数、本文数（本文中の文の数、タイトル文は含まない）、段落数は、表2のとおりです。

表2 データベースの概要

	作文数		タイトル数		本文数		段落数	
	日	母	日	母	日	母	日	母
JP(日本)	134		133		2176		553	
KR(韓国)	55	55	55	52	918	—	244	256
TM(台湾)	57	50	57	49	1050	—	252	223

表2の「日」は日本語作文を、「母」は母語による作文を表しています。母語による作文の本文数は、韓国語及び中国語の文をどう定義するかによって異なるため、言及していません。段落は、執筆者自身が原稿用紙上に設けた空欄マスや改行によって認定しました。

5. テキストファイルの入力について

テキストファイルの入力は、以下の原則にしたがって行いました。詳細は、「入力マニュアル」に記載してあります。

- テキストファイルは、誤字脱字等があっても訂正せず、原文通りに入力する。
- 挿入や削除など推敲のあとがある場合には、推敲後の本文を入力する。
- 句読点は、「、」「。」とする。（原文で「,」「.」が用いられていても、「、」「。」で統一する。）
- 段落開始で1（あるいは複数）マス空けられている箇所は、その数だけ、「□」を入力する。単なる空白（スペース）にはしない。下げられていない場合には、そのまま、1番左のマスから入力する。
- 段落の最後には「■」を「。」の前に入力する。改行マークを入れる。最終段落の最後には「■L」を入力する。
- 入力不可能な文字（画数が足りない、多すぎるなど）があった場合、その文字にできるだけ近い文字を入力しその文字の右に半角で「*」を入力する。
- 中国語の繁体字や簡体字が用いられている場合には、その文字が入力可能であればそのまま入力する。入力が不可能な場合には、日本で用いられている漢字、あるいはその文字にできるだけ近い文字を入力し、その文字の右に半角で「*」を入力する。
- 全く入力不可能、あるいは判読不可能な文字の場合は、その文字の数だけ、「◆」を入力する。

6. 謝辞

本データベースのデータ収集に際しては、研究協力者の高橋圭子さん（東洋大学）、林淑璋さん（台湾・元智大学）、盧（女主）鉉さん（韓国・東国大学校）をはじめ、データ収集先の大学の先生方から多大なご協力をいただきました。また、高橋圭子さんには、データベースの整備・作成の段階でもご尽力いただきました。

作文執筆者の皆様はもとより、データ収集の準備や当日の現場でご協力いただきました多くの方々に心よりお礼申し上げます。

本データベースが、日本語教育研究に関心を寄せる方々に少しでもお役に立てば幸いです。

東京外国語大学留学生日本語教育センター
伊集院郁子

●追記事項

2011年10月4日

検索ファイルに、doc ファイルを追加しました。これまでは pdf ファイルとテキストファイルのみでしたが、テキストファイルに、アップロードの際に文字化けする現象が見られたため、doc ファイルも追加しました。doc ファイルの中身はテキストファイルの中身と全く同じものです。

2012年6月5日

表2「データベースの概要」のKRの母語のタイトル数を53から52に修正しました。