**北京日本語学習者縦断コーパス**

**B-JAS（Beijing Corpus of Japanese as a Second Language）**

**データ概要・タスク説明・文字化概要**

2023.8.22.

本コーパスは、日本語を勉強している中国語母語の大学生の発話及び作文データを、4年間にわたって

収集した縦断コーパスである。大規模横断コーパス「多言語母語の日本語学習者横断コーパス：I-JAS」と対応するコーパスで、対話だけでなく、ロールプレイやストーリーテリングなど、様々なタスクのデータを収集している。

以下にB-JASのデータ概要と、データ収集で実施したタスクの説明を明記する。

# **１．データの概要**

1. 学習者の概要　　＊学習者の情報の詳細は、別紙：「B-JAS背景調査」を参照のこと

・中国語母語の日本語学習者（大学生　日本語主専攻）　17名

　　・入学時、日本語学習歴無し・・・16名　　学習歴1年・・・1名

　　・3年次に日本との交換留学を経験

交換留学1年・・・14名（3年次：8月／9月～翌年8月／9月まで）

交換留学半年・・・2名（3年次：9月～翌年2月まで）

留学経験なし・・・1名

注）3年次の調査について

第5回調査は留学前に中国で実施

（一部、8月来日の学生は、来日直後に日本で実施）

第6回調査は、留学中の14名は日本で実施　それ以外の3名は中国で実施

1. データ収集時期

　　データの収集時期は、20\*\*年1月～20\*\*年4月の4年間

（個人情報保護のためデータ収集年は非公開）

　　1学年期間中に2回　調査を実施してデータ収集を行った（1年次は多少変則的）

　表１　対面調査データ収集時期

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  学年 | 1年次 | 2年次 | 3年次 | 4年次 |
| 調査回 | 第1回 | 第2回 | 第3回 | 第4回 | 第5回 | 第6回 | 第7回 | 第8回 |
| 時期 | 1月 | 5月 | 10月 | 4月 | 9月 | 4月 | 10月 | 4月 |

　表２　作文調査データ収集時期

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  学年 | 1年次 | 2年次 | 3年次 | 4年次 |
| 調査回 | 第1回 | 第2回 | 第3回 | 第4回 | 第5回 | 第6回 | 第7回 | 第8回 | 第9回 | 第10回 | 第11回 |
| 時期 | 1月 | 5月 | 10月 | 1月 | 5月 | 10月 | 1月 | 5月 | 10月 | 1月 | 5月 |

1. データ収集について

　　調査は、国語研の関係者（日本語母語話者）が現地に赴き実施した。1対1の対面で行う「対面調査」、および、オンライン上で学習者に課題を提示し期限を設けて提出してもらう「作文調査」を行い、データを収集した。実施したタスクの詳細は、２．タスク説明　にて述べる。また、対面調査における音声データについては、文字化を行っており、文字化の概要については　３．文字化の概要　で述べる。

　　調査では、上記のデータに加え、1年に一度、学習者に2種類の習熟度テストを実施しており、その結果も公開している。＊各学習者のテストの得点については、別紙：「B-JAS背景調査」に記載

【習熟度テスト】

　　　●　J-CAT：Japanese Computerized Adaptive Test　[[1]](#footnote-1)

　　　●　SPOT：Simple Performance- Oriented Test　[[2]](#footnote-2)

1. 公開しているデータ（2023年3月時点）

下記のデータを公開している。

文字化データはtextファイル　音声データはmp3ファイルである。

【対面調査データ】

　●　文字化データ（textファイル）

・絵描写（D）　・・D

・対話（I）　　・・・I1　I2

・ロールプレイ（RP）　・・・RP1　RP2　RP3　RP4

・ストーリーテリング（ST）　・・・ST1　ST2　ST3　ST4

　　●　音声データ（mp3ファイル）

・絵描写（D）　・・D

・対話（I）　　・・・I1　I2

・ロールプレイ（RP）　・・・RP1　RP2　RP3　RP4

・ストーリーテリング（ST）　・・・ST1　ST2　ST3　ST4

●　ストーリーライティングデータ（textファイル）

　・ストーリーライティング（SW）　・・・SW1　SW2　SW3　SW4

【作文調査データ】

●　作文データ（F）　・・・F1　F2　F3　F4　F5　F6　F7　F8　F9　F10　F11

1. ファイル名について

各データのファイル名は、下記のルールのもとに付与している

学習者ID – タスク名–タスク実施回数–調査実施回

例）　CCB001　–　RP1　　–　　03　　–　　05

　（CCB001というIDの学習者の　ロールプレイ１のデータ。

　　　　3回目のロールプレイ１であり、第5回調査の時に実施されたもの）

　注意：I-JASは、学習者IDの数字が2桁であり、B-JASは3桁である。学習者IDが付与されるファイル名・発話番号なども同様に異なるため注意すること

　　　　I-JASのファイル名　：　CCH01-RP1　　　B-JASのファイル名　：　CCB001-RP1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  学年 | 1年次 | 2年次 | 3年次 | 4年次 |
| 調査回 | 第1回 | 第2回 | 第3回 | 第4回 | 第5回 | 第6回 | 第7回 | 第8回 |
| ST1 | ST1-01-01 | × | ST1-02-03 | × | ST1-03-05 | × | ST1-04-07 | × |
| ST2 | ST2-01-01 | × | ST2-02-03 | × | ST2-03-05 | × | ST2-04-07 | × |
| ST3 | × | ST3-01-02 | × | ST3-02-04 | × | ST3-03-06 | × | ST3-04-08 |
| ST4 | × | ST4-01-02 | × | ST4-02-04 | × | ST4-03-06 | × | ST4-04-08 |
| I1 | I1-01-01 | × | I1-02-03 | × | I1-03-05 | × | I1-04-07 | × |
| I2 | × | I2-01-02 | × | I2-02-04 | × | I2-03-06 | × | I2-04-08 |
| RP1 | RP1-01-01 | × | RP1-02-03 | × | RP1-03-05 | × | RP1-04-07 | × |
| RP2 | RP2-01-01 | × | RP2-02-03 | × | RP2-03-05 | × | RP2-04-07 | × |
| RP3 | × | RP3-01-02 | × | RP3-02-04 | × | RP3-03-06 | × | RP3-04-08 |
| RP4 | × | RP4-01-02 | × | RP4-02-04 | × | RP4-03-06 | × | RP4-04-08 |
| D | D-01-01 | D-02-02 | D-03-03 | D-04-04 | D-05-05 | D-06-06 | D-07-07 | D-08-08 |
| SW1 | SW1-01-01 | × | SW1-02-03 | × | SW1-03-05 | × | SW1-04-07 | × |
| SW2 | SW2-01-01 | × | SW2-02-03 | × | SW2-03-05 | × | SW2-04-07 | × |
| SW3 | × | SW3-01-02 | × | SW3-02-04 | × | SW3-03-06 | × | SW3-04-08 |
| SW4 | × | SW4-01-02 | × | SW4-02-04 | × | SW4-03-06 | × | SW4-04-08 |

　表３　タスク実施時期とファイル名対応表

# **２．各タスクの説明**

　調査で行ったタスクは下記である。

【対面調査】

1. ストーリーテリング（ ST ）

　提示された4コマ・5コマのコマ割り漫画を見て、そのストーリーを話すというタスクである。

奇数調査回／偶数調査回でそれぞれ2種類ずつ、毎年次同じタスクを実施した。（図1・2は奇数回　図3・4は偶数回に実施）

　タスクを始める前に、漫画を見てストーリーを理解するための時間を与え（1分程度）、その後、指定された言い出し文に続けて、ストーリーを話すよう指示した。漫画には、主人公の名前や日本語能力試験の旧2級以上の名詞語彙には日本語と英語／中国語を付与した。





図1　　ストーリーテリング１（ST1）　　　　図2　　ストーリーテリング２（ST2）



図3　　ストーリーテリング３（ST3）　　　　図4　　ストーリーテリング４（ST4）

＜言い出し文＞

ST1　「ケンとマリはサンドイッチを作りました」

ST2　「ケンはうちの鍵を持っていませんでした」

ST3　「ケンは自分の部屋の中で遊んでいました」

ST4　「ケンとお母さんは一緒にごはんを食べていました。」

1. 対話　（ I ）

学習者と調査者が1対1で30分程度の会話を行うタスクである。ある程度の会話内容の統一をはかるため、14項目程度の話題を設定しており（半構造化インタビュー）、調査者は基本的にその流れに沿って進めるが、一問一答とならないよう、できる限り自然な会話となるよう心掛けながら会話を行っている。

下記は、設定した話題である。概ね、下記の流れで会話を進めた。前半は毎回同じ話題で会話をし、後半の自身の意見を述べる話題に関してのみ、奇数調査回と偶数調査回で異なる内容とした。

　＜各回共通話題＞

・ウォームアップ（天候・食事の話題など）

・昨日の１日の出来事

・日本に関する関心事

・日本の本や映画、ドラマなどについて（ストーリーについても話す）

・出身地について

・出身地の料理や食べ物・産物などについて

・出身地のお勧めしたい観光スポット

・誕生日などの行事について

・子供の頃の話

・学校の先生との思い出

・怖かった経験や思い出について

・将来の夢や仕事について

＜奇数回のみ提示した話題＞

　　・将来住みたいのは、都会と田舎、どちらがよいか。それはなぜか。

　　・お金と時間のどちらが大事か。それはなぜか。

＜偶数回のみ提示した話題＞

　・忙しくても給料が安い仕事、給料が安くても自分によって良い仕事、どちらがしたいか。なぜか。

　・子供がいたら、その子を自由に遊ばせるか、または、習い事や学習塾などに通わせるか。なぜか。

1. ロールプレイ （ RP ）

　場面を設定し、与えられた役柄を場面に沿った状況で演じるタスクである。ロールプレイについても、ストーリーライティングと同様、奇数調査回／偶数調査回でそれぞれ「依頼」「断り」という2種類の場面について、毎年次同じ内容で実施した。

　タスクを始める前に、場面や設定が書かれたロールカードを提示し、準備ができたら、カードを回収した上でタスクを行った。なお、ロールカードは指示が正確に理解できるよう中国語に翻訳されたものを使用した。

ロールプレイの内容

＜奇数回実施＞

①依頼：アルバイトの日数を3日から2日に減らしてほしいと店長に依頼する

②断り：アルバイト先で接客から調理に変更してほしいと店長に頼まれるが断る

＜偶数回実施＞

①依頼：推薦状の作成を恩師の大学の先生に依頼する

②断り：北京の観光案内を恩師から頼まれるが、それを断る



1. 絵描写（ D ）

イラストを見ながら、そこに書かれている内容を描写するというタスクである。イラストは、I-JAS同様、許（1997）[[3]](#footnote-3)の研究で使用されたものを使用した。

　 このタスクは、各調査すべて同じものを使用した。

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　 図５　絵描写イラスト

1. ストーリーライティング　（ SW ）

ストーリーテリング（（1）で述べたもの）と同じ漫画を用い、同様に漫画のストーリーについて記述するというタスクである。対面調査の最後に行い、対面調査を行った場所で調査者立ち合いの元PCに日本語で記入をしてもらった。

【非対面調査】

 対面調査とは別に、テーマと文字数を指定した作文課題について、学習者各自が自宅等でPCを用いて期限までに作成し、web上で提出するという調査を行った。1年に3回（1年次は2回）実施した。

各回の作文テーマは表4のとおりである。

　表４　作文調査課題内容

|  |  |  |
| --- | --- | --- |
| **学年** | **調査回** | **作文テーマ** |
| 1年次 | 第1回 | 私の１日―朝起きてから夜寝るまで（200字以上） |
|  | 第2回 | 私のもらって嬉しかった贈り物（400字以上） |
| 2年次 | 第3回 | 私の休日の過ごし方（600字以上） |
|  | 第4回 | 私の今一番ほしいもの（700字以上） |
|  | 第5回 | 私の思い出の旅行（800字以上） |
| 3年次 | 第6回 | 私の得意料理の作り方（900字以上） |
|  | 第7回 | 私の好きな日本の言葉（1000字以上） |
|  | 第8回 | 私の将来の夢（1200字以上） |
| 4年次 | 第9回 | 私のふるさとの四季（1300字以上） |
|  | 第10回 | 私の好きな有名人（1500字以上） |
|  | 第11回 | 私の趣味の変遷（1600字以上） |

# **３．文字化の概要**

　音声データは、以下の方針に従って文字化を行った。

＜文字化における基本方針＞

1. 1文の単位、改行

発話データにおいて1文の単位を定めるのは困難であるため、本データにおいては文の単位については考慮せず、句点の付与はしていない。

●　対話（I）、ロールプレイ（RP）については、話者の交代をもって改行を行い、発話番号を付与している

●　絵描写（D）、ストーリーテリング（ST）については、文字化作業者（日本語母語話者）の判断で、1文と思われる単位で改行をしている。

1. 発話番号

改行ごとに下記のような発話番号を付与している。（①～④まではファイル名と同じ）

　例）CCB001 - ST1 – 03 – 05 – 00010 – K

 ①　　　 ② ③ ④　　⑤　　⑥

①　学習者のID

②　タスク名

③　タスク実施回数

④　調査実施回

⑤　発話の番号（00010は１番目の発話　以下、00020　00030・・・と続く）

⑥　発話者の記号　（　K：学習者（日本語学習者）　C：調査者（日本語母語話者）　）

1. あいづち

あいづちとみなされる発話は〈　〉（やまかっこ）で示し、相手の発話の中のおおよその位置に挿入する。相手の発話と完全に重なるあいづちは、その発話の区切りにまとめて示すか、別の発話として立てる。

1. 発話の重なり

発話が重なっている場合は表記が困難なため、別の発話として扱うか、もしくはあいづち同様〈　〉を使用して相手の発話中に挿入する。基本的に短いものであれば挿入し、長いものは次の発話として扱う。

＜表記の方針＞

1. 長音

前の音節が長く伸ばされていることを表す。長さに関わらず「ー」１つで示す。ただし、ひらがなで表記されることが一般的な長音はこの限りではない。

1. 間・ポーズ

発話が途切れることを表し、長短に関わらず「、」１つで示す。

1. 上昇イントネーション

発話末のイントネーションが上昇調である場合、疑問符「？」を付ける。

1. 非言語行動等

笑いや咳などの非言語行動は、ターンや発話に関わるものを中心に｛　｝で簡単に示している。

1. 数字・アルファベット

発話内に出てきた数字は基本的に漢数字とする。また、アルファベット表記については、その通り発音するものは全角アルファベットとする。 例）　DVD　USB　など

＜個人情報保護について＞

　本データにおいては、個人を特定する情報にあたるものは伏せ字にて表記している。

　個人情報にあたる箇所は、【　】で囲み、一般的な名詞で表記している。

　なお、付与している数字が同じ場合は、同一の語彙を示す。

　　例）　【人名１】　【大学名１】　など

以上

1. 一般社団法人日本語教育支援協会　日本語テストシステムJ-CAT　 https://j-cat.jalesa.org/ [↑](#footnote-ref-1)
2. 筑波日本語テスト集 https://ttbj.cegloc.tsukuba.ac.jp/p1.html [↑](#footnote-ref-2)
3. 許夏珮（1997）「中・上級台湾人日本語学習者による『テイル』の習得に関する横断研究」『日本語教育』95:37-48． [↑](#footnote-ref-3)