

# I-JAS 第二次データ 一般公開のお知らせ

このたび、I-JAS の第二次公開データを一般公開する運びとなりましたので、お知らせいたします。特に、「3. 第一次公開からの変更点」は重要です。必ずお読みください。

## 内容

I-JAS 第二次公開データ 一般公開のお知らせ .....	1
1. 第二次データ一般公開の言語話者グループ .....	2
1-1 第二次公開で追加されるデータの内訳 .....	2
1-2 ID について.....	2
1-2-1 調査地 ID .....	2
1-2-2 調査 ID .....	2
2. 一般公開の内容 .....	3
2-1 調査内容.....	3
2-2 第二次公開で利用できるデータ .....	3
3. 第一次公開からの変更点【重要必読】 .....	4
3-1 作文課題（任意調査） .....	4
3-2 ストーリーテリング・ストーリーライティング .....	4
3-3 検索画面の拡張機能.....	4
3-4 フェイスシートの参照.....	5
3-5 関連データ配布のフォルダ修正 .....	5
4. お願い.....	5

## 1. 第二次データ一般公開の言語話者グループ

### 1-1 第二次公開で追加されるデータの内訳

	言語話者グループ	調査地 ID	公開人数
1	中国語母語話者	CCM	35
2	韓国語母語話者	KKD KKR	35
3	英語話者	EAU EUS	35
4	トルコ語母語話者	TTR	35
5	国内：教室環境学習者	JJC JJE	25
6	国内：自然環境学習者	JJN	25
7	日本語母語話者	JJJ	35
合計			225

### 1-2 IDについて

#### 1-2-1 調査地 ID

調査地ごとの ID は、以下のようになっています。

- ・ 1 つの国で調査地が 1 つの場合

言語の略称 1 文字 + 国の略称 2 文字

例) ベトナム語の「V」+ベトナムの「VN」

- ・ 1 つの地域で複数の調査地がある場合

言語の略称 1 文字 + 国名の略称や大学名の略称 2 文字

例) 中国語の「C」+地名や大学名などの略称「CM」

※一部例外もある。

#### 1-2-2 調査 ID

学習者一人ずつに識別番号を付与しています。

調査 ID+数字 2 ケタ 例) RRS01

対面調査では、基本的に、「調査地の ID の下 2 ケタ」を呼称として使用しています。

例) 学習者 RRS01 の場合、調査地 ID は RRS なので、「RS さん」

## 2. 一般公開の内容

### 2-1 調査内容

調査では日本語を使った 6 種類 12 のタスク、日本語能力テスト、背景情報の収集を行いました。詳細は以下の通りです。

#### (1) 調査内容

	データの 種類	種 類	タスク 番号	タスクの名称	略称
対面調査	発話 データ	1	1	ストーリーテリング 1 「ピクニック」	ST1
			2	ストーリーテリング 2 「鍵」	ST2
		2	3	インタビュー	I
		3	4	ロールプレイ 1 (依頼)	RP1
			5	ロールプレイ 2 (断り)	RP2
	4	6	絵描写	D	
	作文 データ	5	7	ストーリーライティング 1 「ピクニック」	SW1
			8	ストーリーライティング 2 「鍵」	SW2
作文調査 (任意参加)	6	6	9	メール 1	m 1
			10	メール 2	m 2
			11	メール 3	m 3
			12	エッセイ	e

注) 作文調査は任意参加のため、対面調査参加者全員は参加していません (有志のみ)。

#### (2) 日本語能力テスト結果

- ・ J-CAT (Japanese Computerized Adaptive Test)
- ・ SPOT (Simple Performance- Oriented Test)

#### (3) 調査協力者の背景情報 (FS : フェイスシート)

注 1 調査実施の詳しい情報は、添付の以下の資料をご覧ください。

プロジェクトレビュー : 迫田他 (2016) 「多言語母語の日本語学習者横断コーパス  
International Corpus of Japanese as a Second Language」

### 2-2 第二次公開で利用できるデータ

#### (1) 検索システム『I-JAS 中納言 第二次公開版』

- ・ 対面調査のタスク (タスク 1~8、一部はタスク 6 を除く) <sup>注1</sup>
- ・ 日本語能力測定の 2 つのテスト結果
- ・ 学習者の背景情報

(2) 発話プレインテキスト

検索システムおよび関連データ配布からダウンロードできます。

(3) 発話の音声データ

検索システムおよび関連データ配布からダウンロードできます。

注1 第二次公開のデータのうち、中国語・英語の調査地では「絵描写」タスク（タスク6）は含まれていません。

### 3. 第一次公開からの変更点【重要必読】

#### 3-1 作文課題（任意調査）

エッセイについては剽窃したと思われるデータの報告があったため、公開方法の変更について国内の科研メンバーで協議されました。その結果、「エッセイ」と「メール3種」はI-JAS中納言(検索システム)の検索対象からは除外することとしました。それに伴い、I-JAS中納言では検索できなくなり、検索対象を絞り込むチェックボックスが変更されます。

また、作文課題の「エッセイ」と「メール3種」のデータはI-JAS中納言の「関連データ配布」のみでの公開となりました。データは、剽窃したと思われるデータを含め、すべてのデータを公開します。「関連データ配布」の「非対面調査」フォルダをご確認ください。

#### 3-2 ストーリーテリング・ストーリーライティング

ストーリーテリングおよびストーリーライティングは調査時のイラストに、以下のような課題のタイトルとリード文（初めの一文）が付けられていました。学習者はこのタイトルとリード文を言ってからタスクに入っていました。

ピクニック：『ピクニック』、朝、ケンとマリはサンドイッチを作りました。

鍵：『鍵』、ケンはずちの鍵を持っていませんでした。

しかし、昨年、I-JASを使った研究内容の相談があった際、上記の課題のタイトルおよびリード文が全ての学習者に出現するため、それが分析結果に影響すると判断し、データから削除することといたしました。この箇所については、検索システムおよびプレインテキストから削除いたします。

#### 3-3 検索画面の拡張機能

この度の第二次公開で225人分のデータを追加いたしますが、それに伴い、検索対象として、第一次データ、第二次データを指定するチェックボックスを追加いたしました。検索画面の「検索対象」の「データセット」欄でデータ群を指定することができます。

注1. 検索対象では指定した条件すべてを満たすデータを対象として検索を行います。

例) 「第一次データ」にチェックし、CCMのST1、ST2を選んだ場合。

→第一次データのうちの、CCMのST1、ST2のみが検索対象となります。

CCMのST1、ST2以外の第一次データは対象となりません。

### 3-4 フェイスシートの参照

これまで、フェイスシートは検索結果からしか参照できませんでしたが、以下のURLより、直接フェイスシートを閲覧することができるようになりました。赤字の調査IDの部分を変更し、検索するとご参照いただけます。

例) <https://chunagon.ninjal.ac.jp/ijas/facesheet/CCM02>

### 3-5 関連データ配布のフォルダ修正

これまではデータ収集の方法でフォルダを区別しておりませんでした。第二次公開より、「対面調査データ」と「非対面調査データ」でフォルダを分けることにいたしました。

「対面調査データ」フォルダには、「対話」「ロールプレイ (2種)」「ストーリーテリング (2種)」「ストーリーライティング (2種)」、 「非対面調査データ」フォルダには、「エッセイ」「メール (3種)」が載っております。

## 4. お願い

データの不備や不具合などがあった場合は、速やかにお知らせくださいますよう、お願いいたします。

**【連絡先】** 必ず2つを同報でお願いします  
迫田久美子 sakodak@ninjal.ac.jp  
事務局 lsaj@ninjal.ac.jp

I-JAS プロジェクト事務局  
統括責任者 迫田久美子  
研究員 佐々木藍子  
須賀和香子  
細井 陽子  
2017年5月20日