

I-JAS 第三次データ 一般公開のお知らせ

I-JAS に関心を持ってご利用くださり、ありがとうございます。このたび、I-JAS の第三次データを内部公開する運びとなりましたので、お知らせいたします。

内容

I-JAS 第三次データ 一般公開のお知らせ	1
1. 第三次データの言語話者グループ	2
1-1 第三次データの内訳.....	2
1-2 第一次公開から第三次公開までのデータの内訳.....	2
1-3 ID について.....	3
1-3-1 調査地 ID	3
1-3-2 調査協力者 ID	3
2. 内部公開の内容	3
2-1 調査内容.....	3
2-2 第三次公開で利用できるデータ	5
3. 第二次公開からの変更点【重要必読】	5
3-1 I-JAS 中納言のインターフェイスの変更	5
3-2 第一次データのミスの修正.....	5
3-3 検索結果および検索結果のダウンロードファイルについて.....	5
4. お願い.....	6

1. 第三次データの言語話者グループ

1-1 第三次データの内訳

	言語話者グループ	調査地 ID	公開人数
1	中国語話者	CCT	50
2	ベトナム語母語話者	VVN	35
3	ハンガリー語母語話者	HHG	35
4	スペイン語母語話者	SES	35
5	ロシア語母語話者	RRS	35
6	教室環境学習者	JJC/JJE	10
7	自然環境学習者	JJN	10
合計			210

1-2 第一次公開から第三次公開までのデータの内訳

最終公開の予定人数、第一次公開から第三次公開までのデータの内訳は以下の通りです。

	環境／母語		最終公開 予定人数	第一次公開	第二次公開	第三次公開
				2016年5月	2017年5月	2018年5月
1	海外 （教室環境）	インドネシア語（IID）	50	15	—	—
2		スペイン語（SES）	50	15	—	35
3		タイ語（TTH）	50	15	—	—
4		ドイツ語（GAT/GDE）	50	15	—	—
5		ハンガリー語（HHG）	50	15	—	35
6		トルコ語（TTR）	50	15	35	—
7		フランス語（FFR）	50	15	—	—
8		ベトナム語（VVN）	50	15	—	35
9		ロシア語（RRS）	50	15	—	35
10		英語（EAU/ENZ/EUS/EGB）	100	15	35	—
11		韓国語（KKR/KKR）	100	15	35	—
12		中国語（CCM/CCT/CCH/CCS）	200	15	35	50
13	国内	教室環境（JJC/JJE）	100	15	25	10
14		自然環境（JJN）	50	15	25	10
15		日本語母語話者（JJJ）	50	15	35	—
合計			1050	225	225	210

1-3 IDについて

1-3-1 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」 →VVN

- ・ 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」+大学や地域から検討したもの「CM」 →CCM

※一部例外もある。

- ・ 日本国内の調査の場合

日本での調査を表す 1 文字 + 学習環境・時期を表す 2 文字

例) 日本の「J」+国内の教室環境学習者を表す「JC・JE」 →JJC/JJE

例) 日本の「J」+日本語母語話者を表す「JJ」 →JJJ

- ・ また、同じ母語の学習者でも、調査地 ID が異なる場合があります。これは、調査地あるいは調査時期の違いによるものです。以下の母語については、調査 ID の一部が異なっていて同じ母語の学習者を表します。

中国語：CCM、CCT、CCH、CCS

英語：EAU、ENZ、EUS、EGB

ドイツ語：GAT、GDE

韓国語：KKR、KKD

1-3-2 調査協力者 ID

調査協力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. 内部公開の内容

2-1 調査内容

I-JAS の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。お使いいただけるデータの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	対話	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
		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)	作文 データ	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)

学習者の背景情報が確認できる資料です。フェイスシートは検索結果から確認できますが、それに加え以下の URL より直接、個別のフェイスシートを閲覧することができます。赤字の調査 ID の部分を変更し、お使いのブラウザで検索していただくとご参照いただけます。

<https://chunagon.ninjal.ac.jp/ijas-for-members/facesheet/CCM02>

注) 調査実施の詳しい情報は、以下の HP の I-JAS 研究詳細の「I-JAS 関連資料」より、以下の資料をご覧ください。

HP : <http://lsaj.ninjal.ac.jp/>

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」

(4) 語数表

第三次公開分までの学習者、調査者の発話およびライティングの語数を各学習者のタスクごとに集計した一覧です。公開は6月初旬を予定しております。

2-2 第三次公開で利用できるデータ

(1) 検索システム『I-JAS 中納言 第三次公開版』

- ・対面調査のタスク（上記 2—1（2）のタスク 1～8、一部はタスク 6 を除く）
- ・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

I-JAS 中納言の検索結果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

I-JAS 中納言の検索結果および関連データ配布からダウンロードできます。

3. 第二次公開からの変更点【重要必読】

3-1 I-JAS 中納言のインターフェイスの変更

- (1) 検索条件の「言語環境・調査地およびタスク」から「調査 ID」の行を削除しました。
- (2) 検索条件の「データセット」のうち、第三次データのチェックボックスを有効化しました。
- (3) 研究詳細、マニュアル、バージョン情報の場所を右上に移動しました。また、マニュアルについては内容を更新いたしました。

3-2 第一次データのミスの修正

第二次データ公開より、文字化のデータの最終的なチェック方法を変更し、文字化やタグなどのミスを減らしてきました。しかし、第一次データは第二次データ以降とは異なる方法でのチェックだったため、文字化や形態論情報の不備、誤解析が残ってしまっていました。この度、第一次データもチェックし、修正しました。

第一次データで研究を進めてこられた方は、今一度、確認をしていただいたほうがいい場合がございます。ご了承くださいませよう、お願いいたします。

注) 修正した箇所は LSAJ の HP (LSAJ>I-JAS>研究詳細>I-JAS 関連資料>(6) データ修正箇所一覧) からご確認いただけます。

http://lsaj.ninjal.ac.jp/?page_id=364

3-3 検索結果および検索結果のダウンロードファイルについて

これまでは検索結果および検索結果をダウンロードしたファイルには第何次のデータか、どの調査地のデータかの情報は載せていませんでしたが、第三次公開から、「データセット」列（何次公開か）、「調査地」列（どこの調査地か）を追加しました。

4. お願い

4.1 ぜひ、第三次公開版の検索システムにアクセスしていただき、利用していただければと思います。その際、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。

4.2 I-JAS を利用して研究をなさったり、論文を書いたりした場合は、I-JAS 利用の記載をお願いします。また、そのタイトル等の情報を事務局までお知らせくださいますようお願いいたします。差し支えなければ HP でご紹介させていただきます。

【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 learner_corpus_jimu@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

国立国語研究所プロジェクトリーダー 野山 広

2018 年 5 月 21 日