

「JLPTUFS 作文コーパス」のご使用にあたって

1. 使用にあたってのお願い

- (1) CD は複製・譲渡を禁じます。
- (2) 本データをお使いの上、研究発表等を行われる場合には、使用資料として以下の出典を明記してください。

「JLPTUFS 作文コーパス」東京外国語大学留学生日本語教育センター

2011 年 3 月

- (3) 本データをお使いの上、研究発表等を行われた場合には、以下にご一報ください。

sakubun-c@tufs. ac. jp

2. 「JLPTUFS 作文コーパス」の構成

- (1) 「JLPTUFS 作文コーパス」は、以下の 3 つのデータから構成されています。
 - a. 情報一覧ファイル (EXCEL 形式)
 - b. 作文テキストファイル
 - c. 作文 PDF ファイル
- (2) 情報一覧ファイルの各作文番号から、テキストファイルおよび PDF ファイルにリンクが張られています。当該の番号からそれぞれの作文ファイルを見ることができます。

(※イメージ図については、別添の報告書 4～6 ページをご覧ください。)

3. 収録作文数

(1) 収録作文データ数は、以下の表の通りです。

(※報告書 6～7 ページに同じ表があります。)

表 1 「JLPTUFS 作文コーパス」収録作文データ数（レベル・学期別）

レベル	クラス	2009 年度 春学期	2009 年度 秋学期	2010 年度 春学期	レベル別計
100 (入門～初級)	集中	22	37	36	95
200 (初級後半～初中級)	集中	35	40	13	88
300 (初中級)	総合	28	38	36	390
	文章	129	81	78	
400 (中級前半)	総合	25	67	37	270
	文章	42	69	30	
500 (中級後半)	総合	43	49	23	335
	文章	53	52	115	
600 (上級前半)	総合	0	8	17	51
	文章	3	8	15	
700 (上級後半)	総合	15	59	18	183
	文章	31	51	9	
800 (超級)	アカデミック・ ライティング	72	15	16	103
小 計		498	574	443	1,515
計					

表 2 「JLPTUFS 作文コーパス」収録作文データ数（国・地域別）

国	データ数	国	データ数	国	データ数
中国	223	ラオス	26	メキシコ	8
イタリア	107	中国（香港）	26	モロッコ	8
アメリカ合衆国	89	ロシア	22	シリア	7
イギリス	86	カナダ	21	ミャンマー	7
ドイツ	78	チリ	21	イラン	6
台湾	59	オーストリア	20	ウクライナ	6
スペイン	58	ポーランド	17	グアテマラ	6
韓国	56	オーストラリア	15	エストニア	5
カンボジア	49	エジプト	14	クロアチア	5
インドネシア	44	スロベニア	14	ニカラグア	5
タイ	44	フィリピン	14	アイルランド	3
ベトナム	36	スウェーデン	13	アゼルバイジャン	3
トルコ	35	ペルー	13	オランダ	2
モンゴル	31	マレーシア	13	スロバキア	2
ブラジル	30	ハンガリー	12	ヨルダン	2
フランス	30	インド	10	レバノン	2
スイス	29	コロンビア	10	ニュージーランド	1
シンガポール	27	チェコ	10		
ウズベキスタン	26	ブルガリア	9	計	1,515

4. 「JLPTUFS」（東京外国語大学 全学日本語プログラム）のレベル・クラス構成

(1) 「JLPTUFS」は、100～800 の 8 段階のレベル設定になっています。

100 [入門～初級]	→	ゼロスタート～1 学期（半年）間で初級を終了 （終了後 300 レベルに進級）
200 [初級後半～初中級]	→	（終了後 400 レベルに進級）
300 [初中級]		
400 [中級前半]		
500 [中級後半]	→	日本語能力試験 2 級（2009 年度までの級設定） 合格がおよその到達目標
600 [上級前半]		
700 [上級後半]		
800 [超級]	→	既に日本語能力試験 1 級（同上）合格済み

(2) 各レベルの開講授業は以下の通りです。(※2010 年度春学期の例)

表 3 2010 年度春学期「全学日本語プログラム」(JLPTUFS) レベル別開講授業の構成

a. 100～200 レベル

レベル	クラス
100 (入門～初級)	集中 (10 コマ)
200 (初級後半～初中級)	集中 (10 コマ)

(注：作文データ収集の対象クラスは **太字**)

b. 300～700 レベル

	クラス						
	総合	技能別 (各1コマ)					
300 (初中級)	総合 (5コマ)	文法	読解	聴解	文章表現	口頭表現	
400 (中級前半)	総合 (5コマ)	文法	読解	聴解	文章表現	口頭表現	
500 (中級後半)	総合 (5コマ)	文法	読解	聴解	文章表現	口頭表現	
600 (上級前半)	総合 (3コマ)	文法	読解	聴解	文章表現	口頭表現	時事日本語
700 (上級後半)	総合 (2コマ)	文法	読解	聴解	文章表現	口頭表現	時事日本語

c. 800 レベル

	クラス					
	技能・テーマ別 (各 1 コマ)					
800 (超級)	読解	ドラマ・ ドキュメンタリー	アカデミック・ ライティング	ビジネス 日本語	時事 日本語	文学 日本語

・ 100、200 レベル :

「集中」コース → 90 分 (1 コマ) の授業×週 10 回 (毎日 2 コマ)

・ 300、400、500 レベル :

「総合」クラス → 90 分 (1 コマ) ×週 5 回 (毎日 1 コマ)

技能別クラス → 「文法・語彙」「読解」「聴解」「文章表現」「口頭表現」
週 1 回ずつ

- ・ 600、700 レベル：

「総合」クラス → 600 レベルは 90 分（1 コマ）×週 3 回

700 レベルは 90 分（1 コマ）×週 2 回

技能別クラス → 上記 5 技能＋「時事日本語」週 1 回ずつ

- ・ 800 レベル：

技能別・テーマ別 → 「読解」「時事日本語」「ビジネス日本語」

「ドラマ・ドキュメンタリー」「文学日本語」

「アカデミック・ライティング」週 1 回ずつ

その他

- ・「漢字」クラス（300 レベル以上対象 4 レベル各 1 クラス）週 1 回

- ・「発音」クラス（500 レベル以上対象 1 クラス）週 1 回

5. テキストファイルの入力について

- (1) 本データのテキストファイルは、学習者が執筆した通りに入力されています。
誤用の修正や、誤用タグの付与は行っていません。

- (2) 表記の誤りがあっても、入力可能な文字（存在する文字）であれば、原則として誤表記のまま入力しています。

- (3) ひらがな、カタカナ、漢字、ローマ字以外の**入力記号**は以下の通りです。

●● → 固有名詞等を削除したところ

→ 作文原文と、テキスト入力とが一致していないところ

(例 1) 入力不可能な表記の誤りのうち、文脈から正用が想定可能なものは、正用の文字に置き換えて入力し、当該文字の右（後ろ）に # を付しています。

(例 2) 中国語の繁体字・簡体字が使われている場合、文脈から日本語で使用されている漢字が想定可能な場合は、その漢字に置き換えて入力し、当該文字の右（後ろ）に # を付しています。

◆◆ → 判読不可能・入力不可能な文字

(例 1) 判読不可能、かつ文脈から正用が想定不可能な文字の場合は、当該文字を◆に置き換えて入力しています。

(例 2) 日本語に対応する漢字がなく、入力不可能な中国語の繁体字・簡体字は◆に置き換えて入力しています。

◇◇ → ローマ字以外の外国語文字（中国語の繁体字・簡体字を除く）

(※入力マニュアルの詳細は、報告書 69～77 ページをご覧ください。)

以 上

「JLPTUFS 作文コーパス」プロジェクト