

辞書の使用が引き起こす学習者の不自然な表現
- 「JLPTUFS 作文コーパス」の作文から見えてくること-

東京外国語大学留学生日本語教育センター

鈴木 智美

tmsuzuki@tufs.ac.jp

[キーワード] JLPTUFS 作文コーパス、文章表現、辞書、初中級レベル、直訳の壁

1 はじめに

本稿では、現在東京外国語大学留学生日本語教育センター（以下、「センター」とする）において作成が進められている「JLPTUFS 作文コーパス」（東京外国語大学「全学日本語プログラム」を受講する日本語学習者の作文データベース）¹の概要を説明するとともに、2009年度（春・秋学期）に収集された作文データ（約1,070件）をもとに、日本語学習者の作文に見られる不自然な表現（日本語として慣習的に定着しているとは言えない表現）²を、特にその初中級レベルの作文に注目して、「辞書」使用の側面から質的に分析することを目的とする。

本研究は、日本語学習者が文章を産出する際、辞書を使用するにあたって求められる留意点は何か、また、さらには日本語学習者の文章産出を効果的に支援するために辞書のできることは何かを探ることを目指すものである。

初中級レベルの学習者は、初級の文法・語彙等を一通り学習し終えた段階であり、その表現したい内容や母語による知的な思考力のレベルと、身に付けている日本語の知識および運用力との間にギャップがあることも多く、辞書を頼りに難しい表現を試みがちな傾向も見られる。そこには、語の意味の理解や、類義表現の選択、適切なコロケーション（連語）の形成などに関わる諸問題とともに、特にその辞書使用に関係すると思われる点として、不自然な漢語表現や句単位の表現が観察され、母語等からのいわば“直訳”の発想が壁となっていると思われる点が浮かび上がってきた。

2 JLPTUFS 作文コーパスについて

2.1 コーパス構築の目的と概要

まず、「JLPTUFS 作文コーパス」の概要を以下に記す。2008年度にはコーパスの設計と作成準備を行い³、2009年度よりデータの収集を行っている。

(1) 「JLPTUFS 作文コーパス」

目的：日本語学習者の作文を電子データ化し、数多く蓄積していくことで、作文における文法項目、語彙、漢字等の用いられ方、および学習している人の母語や学習レベルと作文との間にどのような関係があるか等を分析し、日本語教育に資することを目的とする。そのための基礎資料をコーパスとして提供する。

内容：東京外国語大学「全学日本語プログラム」(JLPTUFS)⁴の教育課程において書かれた作文のうち、執筆者によるデータ提供の同意が得られた作文をデータ化する。執筆者氏名や、作文中の個人名は削除した上でデータ化を行う。

対象期間：2009年度(春・秋学期)～2010年度(春学期)の計3学期分。

2.2 コーパスのイメージサンプル

「JLPTUFS 作文コーパス」は、「情報一覧ファイル」、「作文テキストファイル」、「作文 PDF ファイル」の3つのファイルから構成される。図1にイメージサンプルを示す。

a. 情報一覧ファイル

A	B	C	D	E	F	G	H	I	J	K	L	M
作文番号	PDFファイル	執筆者ID	レベル	クラス	国籍	母語	専門	作文のテーマ	作文タイトル	実施形態	制限時間	字数指定
2009110010051901	2009110010051901	10005	100	集中	フィリピン	フィリピン語	国際関係	わたしの1日	わたしの一日	宿題	特になし	約400字
2009110010062307	2009110010062307	10016	100	集中	アイルランド	英語	美術	日本に来る前と日本に来てから	日本のせいかつ	宿題	特になし	500-600字
2009120010072106	2009120010072106	10010	200	集中	ブラジル	ポルトガル語	史学	留学と将来の夢	わたしの将来のゆめ	授業時間内	90分	約600字
2009130030052110	2009130030052110	10024	300	文章	中国	中国語	教育学	自国と日本の習慣の違い	中国と日本の習慣の違い	宿題	特になし	約400字
2009140020071305	2009140020071305	10013	400	総合	アメリカ合衆国	英語	文化人類学	自国の教育制度	アメリカペンシルベニア州の教育制度について	宿題	特になし	特になし
2009170020050814	2009170020050814	10135	700	総合	ウズベキスタン	ウズベク語	言語学	「機械との共存」を読んで考えたことについて書きなさい	人間の未来	授業時間内	45分	800字以内
2009180040061601	2009180040061601	10120	800	アカデミックライティング	ロシア	ロシア語	日本語・日本文化	因果関係を述べる	ロシアにおける頭脳流出	宿題	特になし	400字

b. 作文テキストファイル

sample

c. 作文 PDF ファイル

図1 「JLPTUFS 作文コーパス」イメージサンプル

a. の「情報一覧ファイル」には、以下のような情報が記載される⁵。

(2) 「情報一覧ファイル」記載項目

- ①基本情報：作文番号⁶、執筆者 ID 番号、受講レベル、クラス
- ②執筆者情報：性別、年齢、専門、国籍、国籍以外の3年以上の居住地、母語、母語以外の使用言語、日本語能力試験の合格級および合格年
- ③作文情報：実施日、出題テーマ、作文タイトル、実施形態(授業時間内、宿題など)、制限時間、字数指定、文体指定、筆記形態(手書き、ワープロの別)、その他条件(辞書使用可、教科書・ノート参照可など)⁷

作文執筆の際に「辞書」が使用可であったかどうかは、③の「その他条件」の欄にチェックされる。特にチェックがない場合でも、「宿題」の形態で実施された場合には、

本稿では「辞書使用可」の状況であったものと考え、分析対象に含めている。

2.3 2009 年度に収集された作文データ数

2009 年度には、以下の表 1 に示すように計 1,072 件の作文データが収集された⁸。

表 1 2009 年度の収集作文データ数（レベル別）⁹

レベル	クラス	春学期	秋学期	レベル別計
100 (入門～初級)	集中	22	37	59
200 (初級後半～初中級)	集中	35	40	75
300 (初中級)	総合	28	38	276
	文章	129	81	
400 (中級前半)	総合	25	67	203
	文章	42	69	
500 (中級後半)	総合	43	49	197
	文章	53	52	
600 (上級前半)	総合	0	8	19
	文章	3	8	
700 (上級後半)	総合	15	59	156
	文章	31	51	
800 (超級)	アカデミック・ライティング	72	15	87
小計		498	574	1,072
計				

「全学日本語プログラム」(JLPTUFS)では、日本語レベルは100～800の8段階のレベルコードで示されており、本コーパスではその全レベルをデータ収集の対象としている。表1は2009年度の収集データ数をそのレベルごとに示したもののだが、これを学習者の国籍別に示すと、次頁の表2のようになる。

「レベル」および「クラス」の構成を2010年度春学期を例として説明すると、以下の通りである。(以下に掲げる表3を参照されたい。)

「全学日本語プログラム」(JLPTUFS)では、100(入門～初級)および200(初級後半～初中級)レベルは、90分の授業が週に10回(毎日2コマずつ)行われる「集中」コースの形態をとる。100レベルは1学期間(半年)に初級の内容を一通り終え、200レベルは初級後半からスタートし、1学期間に初中級レベルまでをカバーする。

300(初中級)、400(中級前半)、500(中級後半)レベルは、日本語の力を総合的にバランスよく身につけることを目標とする「総合」クラスと、「文法・語彙」「読解」「聴解」「文章表現」「口頭表現」の各種技能別クラスから構成される。「総合」クラスは90分の授業が週5回(毎日1コマずつ)行われ、技能別クラスは週1回である。500(中級後半)レベル終了段階で、日本語能力試験の(2009年度までの級設定で言う)2級取得がおよその到達目安となる。

600(上級前半)、700(上級後半)レベルも同様の構成だが、「総合」クラスは600

レベルでは週に3回、700レベルでは週に2回となる。技能別クラスには「時事日本語」が加わる。なお、400～700レベルでは、学生数に応じて、同一の授業を2つ開講する学期もある。

表2 2009年度の収集作文データ数（国別）

国	データ数	国	データ数	国	データ数
中国	141	中国（香港）	21	ブルガリア	7
イギリス	67	オーストリア	18	エジプト	6
ドイツ	66	フランス	18	グアテマラ	6
イタリア	65	ベトナム	18	チェコ	6
アメリカ合衆国	64	シンガポール	17	マレーシア	6
韓国	48	ポーランド	17	クロアチア	5
スペイン	47	ブラジル	16	スウェーデン	5
台湾	44	チリ	15	メキシコ	5
インドネシア	40	フィリピン	14	ウクライナ	4
タイ	40	カナダ	13	アイルランド	3
トルコ	31	ペルー	13	アゼルバイジャン	3
モンゴル	30	ラオス	11	ハンガリー	3
スイス	28	インド	10	ミャンマー	3
ウズベキスタン	26	スロベニア	8	スロバキア	2
カンボジア	25	オーストラリア	7	ニュージーランド	1
ロシア	22	シリア	7	計	1,072

800（超級）レベルは、既に日本語能力試験の（2009年度までの級設定による）1級を取得済みが目安となるレベルで、「読解」「時事日本語」の他に「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」「文学日本語」のように、技能別・テーマ別のクラスが開講される。授業はいずれも週に1回である。

この他に300レベル以上を対象とした「漢字」クラス（4レベル各1クラス）、500レベル以上を対象とした「発音」クラス（1クラス）も週に1回開講されている。

3 辞書使用を可とする作文に見られる不自然な表現

3.1 初中級レベル

本稿の執筆者は、本作文コーパス作成プロジェクトと平行し、2009年度には「全学日本語プログラム」の100（入門～初級）および200（初級後半～初中級）レベルのコーディネーターおよび授業を担当した。なお、この2つのレベルでは、2008年度秋学期より本センターで開発された新しい初級テキスト『大学生の日本語』（Elementary Japanese for Academic Purposes）を使用して授業を行っている¹⁰。

初級レベルにおいては、習った文法項目・語彙等を使用して身近なトピックについて文章表現の練習をすることが多く、学習者が教科書以外に辞書を調べつつ文章を作成するという状況はさほど頻繁には発生しない。しかし、初級を一通り終え初中級レ

レベルに入ると、文章表現の課題を遂行中、学習者が自分の考えや描写したい状況をより詳細にかつ正確に表現しようと、辞書を調べる姿にしばしば接するようになる。

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¹¹

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

ここでは、本稿執筆者が実際に初中級レベル（「全学日本語プログラム」における「200」レベルの後半）で授業を担当した際の経験を踏まえつつ、2009年度に収集された初中級レベルの作文データ（同「200」および「300」レベルの作文）から、「辞書使用可」という状況のもとに書かれた作文（2009年度春学期 192、同秋学期 159、計 351 件）を対象に、そこに見られる不自然な表現を、辞書の使用という観点から考えてみたい。

3.2 語彙・意味的な問題点

鈴木（1999, 2002）では、主として中級レベルの学習者の作文を語彙・意味的な観点から分析し、その不自然な表現には「漢字熟語」「類義表現」「コロケーション（連語）」「その他慣用的な言い回し」¹²に関わるものが観察されることを述べた。

本稿において分析の対象とした「辞書使用」を可とする初中級レベルの作文におい

ても、その特徴的な点を大きくまとめれば、文法的な誤りの他に、やはり「不適切な語の使用」という語彙・意味的な問題が見られる。当該の語の意味が的確にとらえられずに使用されているものや、類義表現の選択、コロケーション（連語）の形成などに関わる問題が同様に観察される¹³。

しかし、語彙・意味的に不自然な表現を「辞書使用」の観点からさらに見直してみると、学習者の母語あるいは第二言語等からの辞書を介したいわば「直訳」により引き起こされていると考えられるもの、また既存の辞書からは句単位の的確な表現を探し出すことができないという問題から生じていると思われるものが見られる。

以下の表4に、そのような「辞書使用」の問題により引き起こされていると思われる例を示す¹⁴。

表4 辞書使用の観点から見た学習者の語彙・意味的な不自然な表現

学習者の表現（不自然な箇所を下線） 〔 〕内は本稿執筆者の補足	想定される自然な表現	辞書使用が影響していると思われる点
せかいのかんきょうの <u>会話</u> がありました	～についての会議/会談/協議	“talk”の訳か
百年前の <u>場面</u> はとても悪かった	状況	“situation”の訳か
[現代社会の交通について] 今日（こんにち）の <u>運動</u> はとても便利だ	移動	“movement”の訳か
人生の中で彼にはいろいろな <u>ししょう</u> がありました	問題がありました/ 困難にぶつかりました	“difficulty”あるいは “trouble”等の訳か
チーズの <u>混合</u> をなべて入れて	～を混ぜたもの	“mix”の訳か
コンビニで寮の <u>つけ</u> が払えます	(寮の) お金/電気代など	“bill”の訳か
いろいろな <u>えらい</u> 建物を建てました	立派な	“great”の訳か
日本の都市は <u>ちゅうみつ</u> です	込んでいる/人口密度が高い	“density”から 探した表現か
かれらの生活は <u>あんたいだ</u>	かれらは平和に暮らしている	“safe”あるいは “peaceful”の訳か
夜に町は <u>活発</u> になります	にぎやかに	“lively”の訳か
コピー機がよく <u>せんりょう</u> かにある	誰かに使われている	“occupied”の訳か
交通の <u>頻発</u> が高いので	(電車や地下鉄の) 本数が多いので	“frequency”の訳からか
[DVDは] わからなかったら <u>後退</u> することができる	戻って見ることができる	“backward”から探した 表現か
学校へ行くのを <u>停止</u> する	行くのをやめる	“stop”の訳か
科学と芸術は <u>開発</u> した	発展した	“develop”の訳か
[地震で] 多くの人は <u>失踪</u> した	行方不明になった	“missing”の訳か
たまごを <u>こわ</u> してまぜる	たまごを割って	“break”の訳か
日本に来た時にホテルに <u>とど</u> まりました	泊まりました	“stay”の訳か
[切符を買う時] 線に立たなければなりません	(一列に) 並ばなければ なりません	“stand in a line”の 逐語訳か
水の上できれいな <u>はんえい</u> があった※	水の上できれいに映っていた	“reflection”の訳か
きれいな見通しを <u>きょうじゅ</u> した※	とても見晴らしがよかった	“enjoy”の訳か

3.3 考えるべきと思われる問題

学習者には日本語で文章を書く際、母語あるいは第二言語（英語等）において頭に浮かんだ語を辞書で引き、その日本語訳を調べ、文を作成するということがしばしば見られる。その文において中心的な概念を表す語が日本語で思い浮かばない場合に、

まず辞書で訳語を調べ、それを文中にあてはめて文を構成するという手順を踏むことがあるようである。

しかし、通常辞書では当該の語の多義的な意味にしたがって複数の訳語が並べられている。その際、学習者がその中のどれを選択すればよいか、的確な判断ができないことが、上記のような不自然な表現を生み出している一因としてあると思われる。その結果、適切でない類義表現の選択や不自然な漢語表現の使用が現れることになる。

的確な選択ができないのは、辞書に挙げられている例文が十分ではないか、あるいは辞書が日本語学習者の目的や日本語のレベルに合っていないなどの理由が考えられる。学習者がよく使用している電子辞書や、携帯電話のアプリケーションとして搭載されている辞書などが、日本語学習者を対象に作成されたものではないことも大きいだろう。少なくとも学習者のレベルに合った適切な例文が十分に示されれば、上記のような不自然な表現の中には避けられるものもあるのではないだろうか。

また、学習者は辞書を調べることによって、1語1語については目指す言葉の日本語訳を見つけることができるかもしれない。しかし、コロケーションおよび構文までを含め、語を超えたレベルにおいてそれを的確に使用することにつまずいている様子が見られる。表4において※を付した表現などは、むしろ異なる構文を用いることでより自然な表現が得られると思われるものである。日本語学習者の文章表現を効果的に支援するためには、辞書には、句単位あるいは構文単位での的確な表現へと、学習者を導くような何らかの工夫が必要ではないかと思われる¹⁵。

4 今後の課題

4.1 辞書使用の実態調査

以上のような観察を踏まえた上で、今後考察を進めていかなければならない点は2点ある。まず第1点は、学習者の「辞書」使用の実態について、より詳細に明らかにすることである。

2009年度、上記「全学日本語プログラム」の受講者による授業評価アンケートでは、回答者の75%（春学期）および79%（秋学期）という多くの者が、学習リソースとして「辞書をよく使う」と答えていることがわかった（「全学日本語プログラム学生アンケート」集計結果（2009年度春学期・秋学期）。「辞書」は、学習者にとって非常に身近な学習リソースとなっていることがうかがわれる。

しかし、学習者が使用する「辞書」とは、一体どのようなものなのだろうか。学習者はそれをいつ、どのように使っているのだろうか。日本語学習者の「辞書」使用の実態については、実は各教員が授業その他で目にするような個別の場合を除き、その詳細は包括的には明らかにされていない。本稿での観察結果も踏まえ、学習者の辞書使用の実態について、今後まとまった調査を行うことを課題として考えている。

4.2 学習者辞書の可能性

第2点は、言うまでもなく学習者辞書の改善すべき点を検討することである。日本

語学習者が文章表現を行う際に“言いたい”日本語の表現を的確に見つけるために「辞書」のできることは何なのだろうか。

日本語学の分野では、日本語学習者も視野に入れた日本語辞書の充実が言われるようになって久しい。日本語教育の分野においても、姫野監修（2004）などコロケーション（連語）情報を豊富に載せた特色ある辞書も刊行されるようになってきている。

しかしながら、本稿で「辞書使用可」の作文における不自然な表現を見たように、学習者が文章を産出する際、辞書を使用してもなお「言いたい表現」が的確に探し出せない現状があることがわかる。辞書はもちろん万能のツールではないが、少なくともより良い学習リソースとして、辞書にできることには何があるのだろうか。辞書の改善点について再検討が必要ではないだろうか。その際には、初中級レベルだけではなく、中級～上級レベルの学習者も視野に入れて考える必要がある。

さらに日本国内の第二言語としての学習環境のみならず、日本国外での外国語としての学習環境において、学習者がどのような辞書をどのように使用しているかという点についても、次段階では同様に調査の課題となってくるだろう¹⁶。

付記

「JLPTUFS 作文コーパス」作成プロジェクトは、本稿の執筆者ならびに本センターの中村彰准教授が協同で作成を進めているものである。2008 年度に同プロジェクトメンバーであった本センターの伊集院郁子講師にも、随時力を貸していただいている。また、本コーパス作成に向けての準備段階においては、滝沢直宏先生、宇佐美洋先生、大曾美恵子先生、家田章子先生にも講演会・研究会等でヒントや助言をいただいております。ここに記して感謝申し上げます。

なお、本稿におけるコーパスの概要についての説明部分（2.1 節、2.2 節）は、鈴木・中村・韓（2010）における報告内容（鈴木執筆部分）より抜粋・修正を行った。また、本稿で報告した収集作文データ数については、本プロジェクト補佐である本学大学院後期課程在籍の韓金柱さんに集計作業をお願いしている。

また、コーパス構築の実際の作業には、韓金柱さんほか本学大学院生（および大学院修了生）の黄慧さん、上久保明子さん、モンコンチャイ・アッカラチャイさん、徐承希さん、左嵯遥香さん、高橋希美さんに分担してあたってもらっている。皆さんにも、ここに記して感謝します。

なお、本論文の内容は、日本学術振興会科学研究費補助金（平成 22 年度挑戦的萌芽研究「留学生の文章産出時における辞書使用の実態調査—言いたい日本語はどう見つけるか」研究代表者：鈴木智美、課題番号：22652047）による研究へと発展的につながっている。

引用文献

鈴木智美（1999）「意味的な誤用に見られる主な傾向—慣習的に定着した表現および類似の表現に関わる誤り—」平成8年度～平成10年度科学研究費補助金（基盤研究（A）（1））研究成果報告書（研究課題番号 08558020）研究代表者：大曾美恵子（名古屋大学大学院国際言語文化研究科教授）『日本語学習者の作文コーパス：電子化による共有資源化』pp. 131-145（研究協力者）

- 鈴木智美 (2002) 「2000年度中級作文に見られる語彙・意味に関わる誤用—初中級レベルにおける語彙・意味教育の充実を目指して—」東京外国語大学留学生日本語教育センター『留学生日本語教育センター論集』第28号 pp. 27-42
- 鈴木智美・中村 彰・韓 金柱 (2010) 「JLPTUFS作文コーパスの構築について—全学日本語プログラムで学ぶ日本語学習者の作文データベース化」『東京外国語大学センター論集』第36号 pp. 123-133
- 東京外国語大学留学生日本語教育センター (編著) (2010) 『大学生の日本語 I・II』 (Elementary Japanese for Academic Purposes. Vol. 1, Vol. 2) (試用版)
- 東京外国語大学留学生日本語教育センター (2010) 『全学日本語プログラム履修案内』 (2010 年度春学期)
- 東京外国語大学留学生日本語教育センター「全学日本語プログラム」運営委員会 (2009) 「全学日本語プログラム学生アンケート」集計結果 (2009 年度春学期・秋学期)
- 姫野昌子 (監修) (2004) 『日本語表現活用辞典』研究社

注

- 1 「JLPTUFS 作文コーパス」構築プロジェクトは2007 年秋に計画がスタートし、2008 年度に本センター内における「教育研究開発プロジェクト」の1 つとして始動した。2008 年度秋からは文部科学省「質の高い大学教育推進プログラム」(教育 GP) 平成 20 年度選定取組「世界的基準となる日本語スタンダードの構築」(東京外国語大学) の予算を得て行われている。
- 2 ここでは、学習者の作文上に見られる、日本語として慣習的に定着しているとは言えない表現を一括して「不自然な表現」と呼ぶこととする。ここで言う「不自然な表現」には、文法的な規範から明らかに逸脱するものだけではなく、「日本語の表現として自然さを欠くもの」を広く含めるものとする。
- 3 2008 年度の準備段階における本プロジェクトのメンバーは、本稿執筆者のほか、本センター准教授中村彰氏、同じく本センター専任講師伊集院郁子氏の計 3 名であった。2009 年度からは本稿執筆者と中村彰准教授の 2 名で進めている。
- 4 東京外国語大学「全学日本語プログラム」(JLPTUFS: Japanese Language Program, Tokyo University of Foreign Studies) は、世界各国・地域の交流協定校からの交換留学生 (「ISEP-TUFS」(International Student Exchange Program, TUS) 等で学ぶ短期留学生) のほか、日本語・日本文化研修留学生、教員研修留学生、国費および私費の研究生、予備教育課程の国費研究留学生などを対象とする全学向けの日本語教育プログラムである。2004 年 4 月より開始された。入門から超級まで 8 段階のレベル設定がなされ、2010 年度春学期現在、計 45 に及ぶ科目が開講され、1 週間の延べ開講コマ数は計 87 に上る (1 コマ=90 分の授業)。受講者総数 (履修登録者数) は 2009 年度春学期 192 名、同秋学期 182 名、2010 年度春学期は 206 名である。学習者 1 名につき平均週 6~7 コマの授業を受講している。学習者の出身国・地域は、アジア、中東、北米、中南米、欧州、太平洋、アフリカと世界各国・地域にわたっている。
- 5 図 1 はサンプルであり、実際の「情報一覧ファイル」の記載項目・記載順などは図 1 の通りではない。
- 6 データの収集年度、学期、レベル、クラス、収集日などがわかるようにした 16 桁の番号である。
- 7 「その他条件」としては、この他に「提示されたモデル文にしたがって書く」「指定された段落構成にしたがって書く」「学習した機能文型・表現を使って書く」等の条件が挙げられている。

- 8 表1に示すのはコーパスに収録するデータ数であり、各クラス・コースで実際に行っている文章表現タスクの総数はこの数を上回る。複数回の書き直しを行った同一作文については原則として初稿のみをデータとして収集し、また、図表の説明を中心としたものや、文献からの引用を伴う長文の論文形式のものなどテキストデータ化の難しいものについては、コーパス化の対象から除くこととしている。
- 9 表1に示す各レベルの呼称（「初級」「初中級」「中級前半」など）は、100～800のレベルコードを本稿においてわかりやすく呼び換えることを目的としており、現行の「全学日本語プログラム」の履修案内に記されている正規の呼称とは完全には一致しない。
- 10 本センターでは、各教育コース・プログラムの運営のほか、教育研究開発に関わる各種プロジェクトを毎年複数実施しており、教材開発もその1つである。本センター「初級総合教材開発」グループでは、アカデミックな日本語力を身につけることを目的に大学等の日本語教育機関で学ぶ、初級レベルの学生のための初級教科書を2003年より開発してきた。2008年秋学期より「全学日本語プログラム」の初級レベルで実際にこの教科書を使用して教育を行い、2010年3月末にはその試用版が完成した。大学等の学期制に合わせ15週単位の日本語教育プログラムで使用することを念頭に、全26課上下2分冊の構成となっている。第1分冊（1課～13課）は初級前半レベル、第2分冊（14課～26課）は初級後半レベルの内容である。各課に対応した聴解練習、読解練習、および漢字学習も巻末に整備し、総合的な日本語の力の養成を目指す内容としている。
- 11 表3は、『全学日本語プログラム履修案内』（2010年度春学期版）を参考に記したもののだが、ここでの「初級」「初中級」などの呼称も、表1と同様にそのレベル内容をわかりやすく呼び換えることを目的としており、履修案内におけるレベルの正式名称とは必ずしも一致していない。
- 12 慣用的な言い回しに関する不自然な表現としては、例えば「この悲しい感じは短い間だけした」「レストランでは多すぎる客が食べていた」「事故でたくさんの自動車がその道をあまり良くなく通った」「(だめだという返事で)自分の気持ちが悪くなった」などのようなものが見られた。
- 13 「人気が増える (→高くなる)」「健康が悪くなる (→を害する)」「手数料を使う (→払う)」「生活が困った (→苦しかった)」などのコロケーションに関する不自然な表現、「風邪をひいた時便利だ (→役に立つ/よく効く)」「いつも他の人を手伝っている (→助けている)」「[田舎で] 私はいつも多忙じゃない (→少し退屈だ)」「責任がいい (→責任感がある)」「失意に生きた (→がっかりした)」など、類義表現を含めた語の意味的確な理解、および慣用的な言い回しに係る不自然な表現が見られる。
- 14 ここには文法的な誤りは含めていない。
- 15 学習者の母語と日本語との対訳辞書が電子辞書等に搭載されていない場合、学習者は英語・日本語の対訳辞書を便宜的に使用するという方策をとることがしばしば観察される。このことも不自然な語の選択に影響を与えていないか、今後の調査で確認する必要があると思われる。
- 16 また、本作文コーパスの構築は、将来的にはより広く日本語教育に資する総合的データベースの作成を行うものとして、日本語教育アーカイブ化構想にもつながっていく。「全学日本語プログラム」を例に考えれば、文章表現のみならず口頭表現における音声データなども含めて、学習者の産出する日本語を総合的・体系的にデータベース化していくこと、また本センターで開発された種々の教材を使用した各レベル・各技能の日本語教育の授業の実際を「日本語教授法アーカイブ」としてデータベース化していくということなども、日本語教育における資料としての価値だけでなく、日本語教師養成に資するという点から見ても検討課題の一つとなる。