

上級日本語学習者の文体把握に関するケーススタディ 「BCCWJ 図書館サブコーパス文体情報」を用いた読解調査

著者	小西 円
雑誌名	言語資源ワークショップ発表論文集
巻	1
ページ	233-242
発行年	2023
URL	http://doi.org/10.15084/00003739

上級日本語学習者の文体把握に関するケーススタディ —「BCCWJ 図書館サブコーパス文体情報」を用いた読解調査—

小西 円 (東京学芸大学) †

A Case Study of Advanced Japanese Language Learners' Understanding of Styles: Reading Comprehension Survey Using the BCCWJ Library Subcorpus Stylistic Information

Madoka Konishi (Tokyo Gakugei University)

要旨

本研究では、日本語学習者のための類義表現の記述方法に関する研究の一環として、上級日本語学習者が日本語の文章を読む際に文体をどのように把握しているかを探るために調査を行った。本研究は、その調査をケーススタディとして、分析を行うものである。調査には「BCCWJ 図書館サブコーパス文体情報」を用いて、そこで示された指標と本研究の調査との差を分析した。その結果、文体的な指標では特に「硬度」「くだけ度」の理解に難しさが見られたが、その理解を生む過程が、学習者によって異なっていることがわかった。丁寧体・普通体の理解が学習者によって異なり、それが影響している側面があった。また、自分が知らない語や漢字語を硬く感じる、オノマトペを軟らかく感じるなどの特徴が見られた。また、「客観的で硬い」のに「くだけている」文章の文体的な理解が難しいこともわかった。

1. はじめに

日本語学習者が多様な類義表現を使い分けるためには、それぞれの表現をいつ使うかに関する情報が必要である。その違いは、「話しことば」「書きことば」といった用語で説明されたり、「レポートや論文で使う」というように文章・談話のジャンルを用いて記述されることが多い。しかし、「話しことば」「書きことば」という二項対立はあいまいさを含んでいる。

通常、「話しことば」「書きことば」は媒体が音声か文字かという点によって区別される(日本語記述文法研究会編 2009)。それと同時に「話しことば」は、場面依存性、縮約性、即興性、冗長性、対面性を持つ(日本語記述文法研究会編 2009)。「なければ」が「なきゃ」になるなどの音声化の負担を軽減する縮約性や、その場で考えたことを瞬時に発話するために起こる倒置や主述の不一致(即興性)、また、発話したその瞬間から消えていく音声的な情報の伝達性を高めるために同じことを何度も繰り返す等することによる文の整合性の欠如(冗長性)等によって、「話しことば」は必然的に「くだけ度」の高さを兼ね備えてしまうと言えるだろう。「くだけている」「改まっている」というような文体的な特徴は、本来は「媒体が音声であること」と区別されるべきであろうが、「話しことば」という用語ではそれらを区別することが難しい。

また、「レポートや論文」というジャンルは、そのようなジャンルの文章を産出する必要の学習者にとっては必須で有用な提示方法であるが、それ以外のジャンルでの対応が難し

† kmadoka@u-gakugei.ac.jp

い。

そのため、類義表現の使い分けは「話しことば」「書きことば」といった二項対立にとどまらない多様な情報によって示される必要がある。その一つとして、「硬い／軟らかい」「くだけている」などの文体に関する情報がある。文体情報を用いて類義表現を記述するためには、日本語学習者が文体をどのように把握しているかを知る必要があるが、現状では十分に明らかになっているとは言えない。そのため、本発表では、日本語学習者が日本語の文章の文体をどのように把握しているかに関する調査をケーススタディとして行い、その結果と考察を行う。このケーススタディは、12名の日本語学習者を対象とした読解調査を通して彼らの文体把握の様子を分析するものである。以下、詳細を示す。

2. 調査について

2.1 「BCCWJ 図書館サブコーパス文体情報」

本調査では使用データとして「BCCWJ 図書館サブコーパス文体情報」（以下、「図書館サブコーパス文体情報」）を用いる。「図書館サブコーパス文体情報」は、『現代日本語書き言葉均衡コーパス』（以下、BCCWJ）の「図書館サブコーパス」に収録されている書籍サンプルに対して、人手で文体情報を付与したデータである（柏野 2013、国立国語研究所 2015）。付与されている文体情報は、(A) 対象読者に想定される読解レベル、(B) テキストの作成意図、(C) さまざまな文体情報の3種である。具体的には、(A) として「専門度」、(B) として「客観度」、(C) として「硬度」「くだけ度」「語りかけ性度」の5つの指標がある。それぞれの指標は表1のような段階があり、各書籍サンプルに以下のような数値が付与されている。

表1 「図書館サブコーパス文体情報」の各指標

専門度	1 専門家向き, 2 やや専門的な一般向き, 3 一般向き, 4 中高生向き, 5 小学生・幼児向き
客観度	1 とても客観的, 2 どちらかといえば客観的, 3 どちらかといえば主観的, 4 とても主観的
硬度	1 とても硬い, 2 どちらかといえば硬い, 3 どちらかといえば軟らかい, 4 とても軟らかい
くだけ度	1 とてもくだけている, 2 どちらかといえばくだけている, 3 くだけていない
語りかけ性度	1 とても語りかけ性がある, 2 どちらかといえば語りかけ性がある, 3 特に語りかけ性はない

「図書館サブコーパス文体情報」の文体研究における有用性は馬場（2018）によっても示されている。馬場（2018）は BCCWJ の図書館サブコーパス内の語（語彙素）に対して「図書館サブコーパス文体情報」の5つの指標別の平均値を算出し、それらの平均値が語の文体差に関する内省判断と強い相関があることや、和語・漢語・外来語の特徴の違いが平均値の違いに現れていることなどを示した。そのため、「図書館サブコーパス文体情報」のアノテーションデータは文体情報に関する一定の信頼性があると判断し、本調査のデータとして使用する。

2.2 使用データ

本調査では、「図書館サブコーパス文体情報」の5指標を参考にして、さまざまな文体的特徴を持った以下の5つのサンプルを選出し、これを調査使用データとした。各サンプル

とタイトルと 5 指標の詳細を表 2 に示す。各サンプルは「j-Readability」¹によって難易度を調査し、調査協力者となる日本語学習者にとって難易度が高すぎないように考慮した。表 2 に示した難易度は、「j-Readability」のサイトで判定されたものである。

サンプルの具体的な文章は表 3 に示す。実際の調査では、表 3 の文章を含む 500 文字前後を使用した。また、中級以上の漢字語にはフリガナをつけた。

表 2 調査使用データ

ID	タイトル (略称)	文体情報の指標	難易度	
LBi4_00028	岩波講座現代の物理学 (物理)	1 専門家向き 1 とても硬い 3 特に語りかけ性はない	1 とても客観的 3 くだけていない	中級後半
LBt1_00037	松岡修造のカッコいい大人になるための 7 つの約束 (約束)	4 中高生向き 4 とても軟らかい 1 とても語りかけ性がある	4 とても主観的 1 とてもくくくくく	中級前半
LBon_00004	キラキラ恋うらない (占い)	5 小学生・幼児向き とても軟らかい とても語りかけ性がある	4 とても主観的 1 とてもくくくくく 1	中級前半
LBf9_00067	男はオイ！女はハイ… (男女)	3 一般向き 4 とても軟らかい 2 どちらかといえば語りかけ性がある	4 とても主観的 1 とてもくくくくく	中級前半
LBd4_00005	新史・動物行動記 (動物)	3 一般向き 2 どちらかといえば硬い くくくくく 3 特に語りかけ性はない	2 どちらかといえば客観的 1 とても	上級前半

表 3 調査使用データの文章 (一部)

物理	われわれは、なにゆえに、自由な Dirac 場が Fermi 統計をみだし、自由な Klein-Gordon 場が Bose 統計に従うかをみてきた。相対論的な自由場はこれ以外にもさまざまなスピンを記述するものが知られている。これらを逐一述べるのは本書の目的ではない。ただスピンと統計の関係として次の定理 (W.Pauli, 1940) はよく知られている。 * 証明法はいろいろ考えられるが、前節の方法の一般化としては、例えば文献 [8] を参照。
約束	僕はスポーツが大好きです。自分の頭、身体、そして心、持てる力のすべてをありったけぶつけることができますからです。 結果がはっきりと出るのもいいですね。記録や勝ち負けという形で、そのときの能力やそれまで積み重ねてきた毎日の努力が、具体的に表れてきます。「目標を立てる」→「努力する」→「(残念な) 結果が出る」→「目標を立て直す」→「また努力する」→「(いい) 結果が出る！」という流れが、スポーツをやっていると自然に身についてくるのです。 これは、将来スポーツと関係のない仕事に就いたとしても役立つことだと思うんです。だから、みんなにも、何かひとつはスポーツを経験してもらえたら、と思います。
占い	コインやトランプ…etc。さあ、今度はいろ～んなタイプのうらないで、恋のことさぐっちゃお。 1 絶対、モテモテ少女になりたあ～い！！

¹ 日本語の学習者と教師のための Web システム「jReadability」の中の「日本語文章何度判定システム」を使用した。日本語の文章テキストを入力すると、その難易度を 6 段階で判定するシステムである。(李 2016) <https://jreadability.net/>

	<p>恋の五百円玉うらない</p> <p>■うらない方 五百円玉を1個用意してね。そして、イラストのように、左手の人さし指のツメで五百円玉を立て右手の人さし指のツメで、チョンと五百円玉をはじいて回転させるの。クルクルと回転した五百円玉が止まったとき、どっちの面が上になっていたかを確認するのを忘れずにね。</p>
男女	<p>つい先日の話だ。</p> <p>最近流行りの通信販売。例の新聞の日曜版の裏面などに、克明にズラリと商品が写真などで広告してあるやつ。あれをば何となく眺めているうちに、どうしても欲しくなった商品があった。</p> <p>よし、こいつひとつについてやれとばかりすぐ電話にとびついた。</p> <p>「ハイ、こちら一です」と出たのは、耳ざわりだけでわかるアルバイトギャルの声。</p> <p>「商品番号をおっしゃって下さい」といわれて答える。</p> <p>さらに「御住所と御名前、電話番号を郵便番号からどうぞ」ってんで、こいつにも律儀に返事をする。</p> <p>が、その次がいけない。</p> <p>「では生年月日とお年をどうぞ」</p> <p>「え！」</p>
動物	<p>5 動物たちの愛は血縁淘汰では解けない 血縁者でない仲間を多く助けたことこそ大切</p> <p>私が「警戒音を発して仲間を逃がす利他行動」を遺伝子レベルで考え、その犠牲になった個体の遺伝子が生きのこった兄弟の中に伝わるからいいのだという「血縁淘汰」の考え方にハテおかしいぞと思ったのは、すぐにいくつかの疑いが生まれたからである。</p> <p>だいいち、犠牲者の兄弟はそれで助かって、大へん結構だったが、多くの血縁者でない仲間たちはどう評価する？一彼らの方がはるかに多く助かっていることも多いだろうに。それは犠牲者の遺伝子とかかわりがないから、「派生的なもの」で、行動学的には無視するのか、考えなくていいのか、ソナバカナ、ということが一つ。</p>

各調査使用データの文章は、次のような文体的特徴を持っている。『物理』（以降、サンプルのタイトルは表2に掲載した略称で示す）は「専門家向きで、客観的で、硬く、くだけておらず、語りかけ性がない」文体である。『約束』は『物理』とはおおよそ逆の文体的特徴をもっており、「中高生向きで、主観的で、軟らかく、くだけており、語りかけ性がある」文体である。『約束』では丁寧体が用いられている。また、『占い』は『約束』と類似の傾向を持っているが、対象読者がより幼くなり、「小学生・幼児向きで、主観的で、軟らかく、くだけており、語りかけ性がある」文体である。『男女』は『約束』とほぼ同じ特徴を持ち、対象読者のみが異なっている。対象読者が異なることに起因して、『男女』は使用されている語に漢語や難度の高い語が増え、また、文体が普通体である。一方、『動物』は、「客観的で硬い」のに「くだけている」という点で、そのほかの4サンプルと異なっている²。馬場（2018）は、「専門度」「客観度」「硬度」「くだけ度」の4指標が相互に強い相関があることを明らかにしているが、その点から見ると『動物』はやや異質なサンプルである。『動物』は他の4サンプルより難易度がやや高めであるが、日本語学習者がこの種の文体をどのように把握するかを確認するために、調査対象とした。

² 一般的には、「客観的で硬い」文体は「くだけていない」場合が多く、『図書館サブコーパス文体情報』でも、「とても／どちらかといえば客観的」で「とても／どちらかといえば硬い」文体の文章2370サンプルのうち、「とてもくだけている」のは『動物』を含む2サンプルだけであったしかし、メイナード（2008）が「マルチジャンルという現象こそがごく頻繁に見られる言語のバリエーションの一種である」（p.14）と指摘するように、ひとつのディスコースの中に他のジャンルの特徴とされる表現が混用される例は少なくない。『動物』もそのような捉え方ができる可能性がある。

2.3 調査協力者

調査協力者は表4に示す12名で、2022年6月の調査時点で日本の大学または大学院に在籍している留学生である。日本語の習熟度を把握するために、SPOT90(小林 2015)を受験してもらった。SPOT90の成績目安は、「上級が81~90点」「中級が56~80点」となるため³、中級と判定された協力者H~Lも、中級の上位であると考えられる⁴。

表4 調査協力者一覧

協力者	母語	日本語 学習期間	日本 滞在期間	SPOT レベル	SPOT 点数
A	中国語	8年7か月	3年7か月	上級	88
B	トルコ語	5年6か月	0年9か月	上級	86
C	中国語	8年8か月	3年5か月	上級	85
D	中国語	6年3か月	5年8か月	上級	84
E	中国語	6年8か月	2年8か月	上級	83
F	韓国語	5年6か月	1年2か月	上級	83
G	ベトナム語	5年0か月	0年9か月	上級	81
H	中国語	6年3か月	2年8か月	中級	80
I	中国語	6年6か月	4年0か月	中級	80
J	韓国語	6年10か月	0年8か月	中級	80
K	中国語	6年0か月	5年9か月	中級	77
L	中国語	8年7か月	5年2か月	中級	72

2.4 調査実施手順

調査は、①背景調査シートの記入、②読解調査5種、③読解調査の回答に対するインタビュー、④SPOTの受験、という流れで行った。5種の読解は、調査協力者によって読む読解文の順序を変え、読解順序の影響が偏って出ないようにした。調査はテストではないので、わからない語や表現があってもかまわないことや、趣味で読書をするときと同じような読み方をしてもらうことを調査協力者に伝えた。そのため、辞書を使いたい場合は使ってもよいと伝えたが、辞書を使った者はほとんどおらず、また、使った場合も文体理解に影響を及ぼす語ではなかった。また、SPOTの受験は自宅で都合の良い時間に行ってもらった。

読解調査の設問は表5のとおりである。「専門的」「客観」「主観」「硬い」「軟らかい」「くだけている」「語りかける」という重要語については、調査前に各調査協力者に理解を確認した。

「◆内容がどれくらい理解できましたか。」という設問は、日本語学習者である調査協力者の文章理解の程度を知ることが目的としたものである。また、「◆この文章はどれに当てはまると思いますか。」の各設問は、「図書館サブコーパス文体情報」の指標を問うものである。「◆(3)~(5)について、どうしてそう思いましたか。」は、そのような指標の判断に至った語や表現、文法項目などを把握することを目的とした設問である。柏野(2013)によると、(1)で問う専門度は対象読者に想定される読解レベル(難易度)、(2)で問う客観度はテキスト作成意図を示しており、文体情報を表すのは(3)「硬度」、(4)「くだけ度」、(5)「語りかけ性度」である。そのため、文体情報の把握について知るために、(3)~(4)について詳細を問う。また、読解文は、BCCWJの図書館サブコーパスのサンプルとなる段階で、レイアウトやフォントの情報などが取り除かれている。現物の書籍であればイラストがついていたり、フォントやレイアウト等に工夫があったりすると思われるが、今回の調査は、日本語学

³ 筑波日本語テスト集 TTBJ のサイトに掲載がある。 <https://ttbj.cegloc.tsukuba.ac.jp/p1.html>

⁴ そのため、本稿のタイトルは両者をまとめて「上級」と仮称している。

習者が日本語の文章からどのような文体情報を取り出すかを知ることが目的としているため、レイアウト等の情報は調査から除外された状態で行った。

表5 読解調査の設問

<p>◆内容がどれくらい理解できましたか。○をつけてください。 とてもよく理解できた ・ だいたい理解できた ・ あまり理解できなかった</p> <p>◆この文章はどれに当てはまると思いますか。○をつけてください。 (1) とても専門的 ・ やや専門的 ・ 一般の人向き ・ 中学や高校生向き ・ 子供向き (2) とても客観的 ・ どちらかといえば客観的 ・ どちらかといえば主観的 ・ とても主観的 (3) とても硬い ・ どちらかといえば硬い ・ どちらかといえば軟らかい ・ とても軟らかい (4) とてもくだけている ・ どちらかといえばくだけている ・ くだけていない (5) とても語りかけられている感じがする ・ 少し語りかけられている感じがする ・ 語りかけられていない感じがしない</p> <p>◆(3)～(5)について、どうしてそう思いましたか。(あとでインタビューでも聞きますから、線を引くのが難しい場合はひかなくてもいいです) (3)について(硬い・軟らかい) そのように思った部分に_____を引いてください。 (4)について(くだけている) そのように思った部分に_____を引いてください。 (5)について(語りかけ) そのように思った部分に <input type="checkbox"/> をつけてください。</p>
--

2.5 指標の計算手順

各調査協力者が「◆この文章はどれに当てはまると思いますか。○をつけてください。」という設問で回答した結果から、各指標の数値を得ることとした。指標を数値に表す場合は、各項目の一番左端を1点とし、右に進むにつれて1点加算されることとした。たとえば、「客観度」において「とても客観的」は1点、「どちらかといえば客観的」は2点、「どちらかといえば主観的」は3点、「とても主観的」は4点である。

このように計算した各調査協力者の結果を基に、各サンプルの各指標における平均値を算出した。

3. 調査結果と分析

3.1 調査結果の全体像

まず、調査から得られた各指標の値を表6に示す。「図書館サブコーパス文体情報」の指標の値を「指標」列の「コーパス」の行に示す。また、「指標」列の「調査結果」の行に、各調査協力者の選択した指標の平均値を示す。また「差」の行には、「図書館サブコーパス文体情報」の指標と本調査結果で得られた平均値の差を絶対値で示す。

表6 各指標の調査結果と「図書館サブコーパス文体情報」の指標との差

サンプル	指標	理解度	専門度	客観度	硬度	くだけ度	語りかけ度	差の合計
物理	調査結果	2.8	1.2	1.2	1.3	3.0	2.9	
	コーパス		1.0	1.0	1.0	3.0	3.0	
	差		0.2	0.2	0.3	0.0	0.1	0.8
約束	調査結果	1.0	3.8	3.3	3.2	2.3	1.5	
	コーパス		4.0	4.0	4.0	1.0	1.0	
	差		0.3	0.8	0.8	1.3	0.5	3.7
占い	調査結果	1.1	3.7	3.3	3.9	1.3	1.3	
	コーパス		5.0	4.0	4.0	1.0	1.0	
	差		1.3	0.7	0.1	0.3	0.3	2.7
男女	調査結果	1.5	3.1	3.8	3.2	1.5	1.8	
	コーパス		3.0	4.0	4.0	1.0	2.0	
	差		0.1	0.3	0.8	0.5	0.2	1.9
動物	調査結果	1.7	2.4	3.3	2.3	2.6	2.5	
	コーパス		3.0	2.0	2.0	1.0	3.0	
	差		0.6	1.3	0.3	1.6	0.5	4.3

※理解度：とてもよく理解できた1、あまり理解できなかった3

専門度：とても専門的1、子供向き5 客観度：とても客観的1 とても主観的4

硬度：とても硬い1、とても軟らかい4 くだけ度：とてもくだけている1、くだけていない3

語りかけ度：とても語りかけられている感じがする1、語りかけられている感じがしない3

3.2 読解サンプルからみた分析

表6をもとに、読解サンプルごとに分析を行う。

5種の読解サンプルの差の合計を比較すると、「図書館サブコーパス文体情報」の指標と最も調査結果が近いものは、差が0.8の『物理』である。『物理』は、理解度は最も低いが、文体情報は母語話者とほぼ同様に把握できている。調査の回答およびインタビューからは、「Dirac場」や「Fermi統計」などの専門用語がわからないため、内容を理解することはできないが、定理が示されていたり文献情報が掲載されていたりしていることから、「論文っぽいもの」として理解したと述べる調査協力者が多く、「硬い」「くだけていない」「語りかけていない」と判断することが可能だったとみられる。

一方、差の合計が大きい読解サンプルは差が3.7の『約束』と4.3の『動物』である。『約束』は「客観度」「硬度」「くだけ度」で、『動物』は「客観度」「くだけ度」で差が大きい。『約束』は口頭で行われたスピーチを文字に落としたような文章で、丁寧体で書かれている。また、『動物』は扱う内容が生物学的でやや専門性の高いものであるが、「ソナバカナ」のようなくだけた表現がちりばめられており、「客観的で硬い」のに「くだけている」という面で文体を理解するのが難しいと思われるサンプルである。差の合計は『動物』が最も高く、調査協力者がこの種の文体の理解に苦労したことがわかる。

差が2.7の『占い』と1.9の『男女』は上記2タイプの間位置する。『占い』で差が大きいのは「専門度」と「客観度」である。「専門度」では「中高生向き」と答えた調査協力者が多く、インタビューで「日本の女子高生には占いが人気であるから」という趣旨の発言が複数あった。「また、「客観度」では「客観的」と答えた調査協力者が多く、インタビューには「占いを科学的だと考えるかどうかは人による。占いを科学的だと考える人は、この文章は客観的だと思うだろう」という趣旨の発言が複数見られた。対象読者やテキストの作成意図などに関する部分は、それが読まれる社会文化的な理解とも関連することがわかる。また、『占い』には「いろ～んな」や「なりたあ～い」のようなくだけた表現や「モテモテ」

のようなカタカナ表記、オノマトペが多用されており、これらから「軟らかい」「くだけている」と判定した調査協力者が多かった。その結果、「硬度」「くだけ度」に関しては「図書館サブコーパス文体情報」の指標とほぼ同様の結果となっている。『男女』で差が大きいのは「硬度」で、文体情報に関わる指標である。

『約束』『動物』『男女』では文体に関わる指標で偏差が大きい。次の節では、これら3つの読解サンプルの文体情報に関わる「硬度」「くだけ度」の分析を行う。

3.3 指標からみた分析—「硬度」

本節では、文体情報に関わる指標を扱う。その中でも、3.2 節で差が大きかった「硬度」「くだけ度」を中心に分析する。

まず、「硬度」について分析する。『約束』も『男女』も、「図書館サブコーパス文体情報」では「とても軟らかい」となっている。しかし、本調査では『約束』は「どちらかといえば軟らかい」を選んだ調査協力者がもっとも多く8名、次いで「とても軟らかい」が3名、「どちらかといえば硬い」が1名であった。また、『男女』では「どちらかといえば硬い」が4名、「どちらかといえば軟らかい」が2名、「とても軟らかい」が4名で、ばらつきが大きかった。

インタビュー等を通してみると、「軟らかい」と判断するに至った語や表現は、『約束』では「でしょう」や終助詞の「よ」を伴う表現が目立ち、「話しことばなので軟らかく感じる」

『です・ます体』は軟らかく感じる」といった意見が多かった。『男女』では「やつ」「よし」のようなくだけた語や、「ガッチャーン」のようなオノマトペが「軟らかい」と判断する根拠となっており、『約束』と同様に「話しことばなので軟らかく感じる」という意見が見られた。また、『男女』は「克明」「凄みのこもった詰問調」など、日本語学習者にとって難度の高い語が複数あり、「漢字語が多いので硬く感じる」「会話で使わない語／自分の知らない語は硬く感じる」と答えた調査協力者が見られ、それが「どちらかといえば硬い」という指標の回答が多かった理由であろう。

また、「硬度」を判断する際の文体の捉え方には、異なる意見が見られた。『約束』は丁寧体で書かれているが、『です・ます体』は敬語なので硬い」という意見と『です・ます体』は軟らかく、硬いのは『である体』という意見があった。また、『男女』は普通体で書かれているが、「普通体を軟らかく感じる」という意見があった。このような異なる意見は、異なる調査協力者間に見られ、同一の調査協力者内では読解サンプルごとに異なる見解に揺れることはなかった。

調査協力者によって異なるこのような文体理解は、媒体が音声である音声言語における「丁寧体／普通体」の役割と、媒体が文字である文字言語における「丁寧体／普通体」の役割の混同が関係しているように思われる⁵。音声言語においては、丁寧体は対人的な丁寧さを表し、普通体が非丁寧さを表す。つまり、丁寧体は形式的でくだけていないことを表し、普通体はぞんざいさや親密な対応であることを表す。一方、文字言語においては、丁寧体が読み手を意識した丁寧さを表す点は同じであるが、普通体はぞんざいさや親密さを示さず、文字言語における無標の文体である。つまり、丁寧体が形式的でくだけていないことから硬さを表すのは音声言語のみであり、普通体が軟らかさ（ぞんざいさ）を表すのも音声言語においてのみである。日本語学習者である調査協力者には、音声言語におけるその知識が、部分的に文字言語に適応されているのではないだろうか。

⁵ ここでは、音声・文字という媒体の違いを指し示すため、「話しことば」「書きことば」という用語を用いない。音声という媒体で表される言語を「音声言語」、文字という媒体で表される言語を「文字言語」と呼ぶこととする。

3.4 指標からみた分析—「くだけ度」

次に「くだけ度」について分析する。「図書館サブコーパス文体情報」の指標では、『約束』『動物』ともに「とてもくだけている」となっているが、本調査では『約束』は「くだけていない」が5名、「どちらかといえばくだけている」が6名、「とてもくだけている」が1名であった。『動物』では「くだけていない」が7名、「どちらかといえばくだけている」が5名であった。特に『約束』で判断が二分している。

「くだけているかどうか」という判断基準は、調査協力者によって大きく異なっており、いくつかのタイプに分けられることがわかった。1つ目のタイプは、「文が完結していないこと」や「何かが省略されていること」を「くだけている」と捉えるタイプである。そのような調査協力者の場合、『動物』の「～が一つ。」「ということ。」のような体言止めの文にコンピュータがないために「くだけている」と判断したり、『約束』の「もらえたらと思います」のように「たらしい」の「いい」が省かれていることを「くだけている」と判断したりした。2つ目のタイプは、文章の書式や構成に着目するタイプである。そのような調査協力者は、『男女』の「ソナバカナ」のように普段ひらがなで書かれるものがカタカナで書かれていることや、同じ内容を何度も繰り返すことを「くだけている」と捉えた⁶。3つ目のタイプは、表現を判断基準とするもので、「日常的な話しことばはくだけている」と捉えたり、「漢字語が多いのでくだけていない」と捉えたりした。また、「くだけ度」においても、調査協力者によって異なる文体の理解が影響を与えており、『です・ます体』は硬いと考える者は「くだけ度」でも『です・ます体』はくだけていないと判断した。

4. おわりに

本研究は、「図書館サブコーパス文体情報」の指標を用いて、上級日本語学習者が日本語の文章の文体をどのようにとらえるかをケーススタディとして調査し、その結果を考察したものである。調査の結果を、「図書館サブコーパス文体情報」の指標との差という数量的な側面と、インタビュー調査による質的な側面から分析した。その結果、文体的な指標では特に「硬度」「くだけ度」に理解の難しさが見られたが、その理解を生む過程が、調査協力者によって異なっていることが示された。丁寧体・普通体の理解に調査協力者による異なりがあることは注目し得る。ここから、日本語教育において「音声言語における丁寧体・普通体」と「文字言語における丁寧体・普通体」の違いが重要視されていない可能性と、そのことが文体の理解の違いを生じさせることが示唆された。

調査協力者はインタビューで「話しことば」「書きことば」という用語を何度も用いたが、それが何を表すと理解しているのか、捉えることが難しかった。ここには、「話しことば」という用語が、音声という媒体を指すだけでなく、くだけていることや改まっていないことを必然的に含むことによる難しさも関連している。また、自分が知らない語や漢字語を硬く感じる、オノマトペを軟らかく感じるなどの特徴が見られた。また、「客観的で硬い」のに「くだけている」文章の文体的な理解が難しいこともわかった。

日本語学習者に類義表現や文法的な表現を説明・記述する際に、文体情報は重要な役割を果たすと思われる。よりよい記述につなげるために、引き続き、学習者の文体把握に対する理解を深めていく必要があると思われる。

謝 辞

本研究は JSPS 科研費 18K12420 の助成を受けたものである。

⁶ 1つ目のタイプや2つ目のタイプの調査協力者は「くだけている」という語の理解が不十分であった可能性もある。調査前に各語の確認をしたが、調査の中身に関わるため、「くだけている」表現がどんなものかについては具体的に説明しなかった。しかし、調査協力者が上級～中級の後半であることを考えると、「くだけている」ことを十分学ばないことが現状の日本語教育であるとも考えることができる。

文 献

- 柏野和佳子 (2013) 「書籍サンプルの文体を分類する」『国語研プロジェクトレビュー』 4 (1), pp.43-53.
- 国立国語研究所 (2015) 『BCCWJ 図書館サブコーパスの文体情報』 (2015 年公開第 1 版) <http://id.nii.ac.jp/1328/00003109/>
- 小林典子 (2015) 「SPOT」李在鎬編『日本語教育のためのテストガイドブック』くろしお出版, pp.110 -126
- 日本語記述文法研究会編 (2009) 『現代日本語文法 7 第 12 部談話 第 13 部待遇表現』くろしお出版
- 馬場俊臣 (2022) 『語の文体値データ』 (2022 年 2 月公開第 1 版) <http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/12130>
- 馬場俊臣 (2018) 「『BCCWJ 図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性」『言語資源活用ワークショップ発表論文集』 3, 国立国語研究所, pp.241-256.
- 李在鎬 (2016) 「日本語教育のための文章難易度研究」『早稲田日本語教育学』 Vol. 21, pp.1-16.