

上級～超級日本語学習者の作文から見た言語産出実態

趙 海城 (明星大学 人文学部)

Language Production Reflected in the Composition by Advanced and Super-Advanced Japanese Language Learner

ZHAO Haicheng (Meisei University)

要旨：

本稿は「YNU 書き言葉コーパス」を使い、中韓の留学生と日本人学生の作文を分析し、上級～超級日本語学習者の言語産出実態、日本人との違いの一端を明らかにすることを目的とする。分析した結果、留学生の作文には挨拶語、人称代名詞、指示代名詞、連体詞「此の、其の」、助動詞「たい」が過剰使用され、様態・推量助動詞、「てる」の過剰使用が見られる。中韓留学生間にも、「此の、其の、此れ、其れ」の産出数、受身・尊敬、使役助動詞の産出数の違いが見られる。留学生はレベルが上がるにつれ、作文語数が増え、品詞の大半は異なり語数、延べ語数が増え、相手との交渉表現、文構造が複雑化する。また留学生は感動詞の産出が減り、「僕・俺」が「私」に取って代われ、形状詞・助動詞・終助詞が増加するなど、日本人の使用実態に近づいている。ただし、人称代名詞の多用、様態・推量助動詞の過剰使用など、超級になっても日本人学生の産出と違う特徴が見られる。

1. はじめに

本稿は「YNU 書き言葉コーパス(以下「YNU コーパス」と略称する)」を用いて、中韓両国からの留学生と同年代の日本人学生(学部生、大学院生、研究生を一括して学生と呼ぶ)の課題作文の品詞別産出状況を考察するものである。分析する過程で、日本人学生の産出状況を参照基準値とするが、むしろ日本人学生でも日本語習熟度の面においては、個体差もあり、社会人、言語熟練者と比べれば、未熟さが残る者もある(小野他 2007)ため、この参照基準値は留学生が目指す絶対基準ではないことを断っておく。

山内(2009)は、日本語学習者の OPI データ(KY コーパス)から、学習者のレベル別言語特徴を明らかにした。具体的には「レベル判定に寄与する形態素を探す」ことを試みた結果、「だ(助動詞)」、「よ(終助詞)」、「から(接続助詞)」、「やっぱり」、「と思います」は「上級以上であることを決定する形態素」とし、「こう(フィラー)」、「けれども(接続助詞)」、「っていう(複合助詞)」、「んです(ど)」は「超級であることを決定する形態素」としている。橋本(2011)は、山内の研究に習い、KY コーパスの上級話者 12 人、超級話者 15 分の発話データから「抽象的關係」を表す名詞を抽出して考察した。その結果、「面、風、辺、自身、状況、互い、逆」などが超級を示す実質的形態素であること、「そういう面」「そういうふう」「その辺」のように、機能形態素の超級マーカーと実質形態素の超級マーカーが密接に関連していることを明らかにした。毛(2013)は、中国日本語学習者コーパス(CJLC、4 級作文 1200 篇、8 級作文 1200 篇からなるもの)を母国語話者コーパス(会話、小説、論説からなるもの)と比較し、中国日本語学習者の高頻度語産出の特徴を考察した。その結果、中国語日本語学習者は名詞、形容詞、形容動詞、副詞などの自立語及び複合辞の産出が多く、カバー率も高く、過剰使用する傾向にあるが、接続詞、格助詞、係助詞に関しては、産出量の面においては母

語話者コーパスと多く違わないが、学習者コーパスのほうはカバー率が高いことが分かった。また、学習者が推量助動詞「そうだ、みたいだ、らしい」、準体助詞「の」、終助詞の産出が少ないことが分かった。これらのことから、学習者は文の基本的な成分を構成する単語を多用するが、ムードを表す語彙の使用が少なく、表現が単調であると指摘している。

山内(2009)、橋本(2011)の考察対象はインタビューによる学習者の発話データで、書き言葉ではない。毛(2013)は学習者の作文データを対象に分析しているが、作文は中国教育部高等学校外語専業教学指導委員会日語分委員会の主催で実施された2007年～2009年度試験作文の一部であり、比較対象とする母語話者コーパスもやや古い小説、論説文、会話文となり、文体、話題等も違うため、それをもって、外国語学習環境にある中国語日本語学習者が日本語を過剰使用、過少使用と言っても統一性がないように考えられる。そこで、本稿では、第二言語学習環境にある中韓両国からの上級～超級留学生と同年代の日本人学生の同じテーマの課題作文を比較することにした。

2. 調査データと調査概要

YNU コーパスは日本人学生30名と、日本国内にいる中国人・韓国人留学生30名¹ずつに対し、状況や難易度の異なる12種類の作文タスクを課し、各国の学生よりそれぞれ360編、三カ国合わせて計1080編の作文データを収集したものである。12種類のタスクは、手紙、PCメール、携帯メール、投書、レポートなどのスタイルのものとなるように配慮され、また自発型か頼まれ型か、読み手は特定の相手か不特定の相手か、読み手は特定の相手の場合、目上なのかそれとも同僚・友人なのかに分かれるように設定された。さらに、中韓両国の留学生が書かれた作文は独自の評価基準(タスクの達成、タスクの詳細さ・正確さ、読み手配慮、体裁・文体の四項目)で評価され、その達成度に応じて、下位群(10名)、中位群(10名)、上位群(10名)という三つのグループに分けられる。本稿もこのグループ分けに基づく。

YNU コーパスにはオリジナルデータとオリジナルデータを補正した補正データ²があるが、本稿では形態素解析の利便性を考え、補正データを分析対象とした。「茶まめ」を用いて、YNU コーパスの補正データを対象に、形態素解析処理を行った。形態素解析器は「MeCab 0.996」、解析用辞書は「UniDic-mecab 2.1.2」を使用している。基本的に形態素解析して得た解析結果を使うが、「形状詞-助動詞語幹:(そうだ(様態)、ようだ、みたいだ)」は従来「らしい」などと同じく助動詞とみなされることが多いため、助動詞に分類し直した。

3. 調査結果

3. 1 全体の傾向

表1はYNU コーパスにおける延べ語数、異なり語数、文の数、【 】の数を示したものであり、表2はこれらの項目の10万語あたりの調整頻度を示すものである。

¹ 両国の留学生は日本の大学、大学院で講義を受けられるレベルで、一般的に言えば、上級レベル及びそれ以上のものである。旧日本語能力試験、2010に改定された新日本語能力試験の受験結果を見ると、韓国人留学生の内訳は1級、N1合わせて19名、2級、N2合わせて3名、未記入8名で、中国人留学生は1級、N1合わせて26名、2級、N2合わせて3名、未記入1名である。本稿ではこれに基づいて上級～超級日本語学習者とした。

² 補正の主なポイントとしては、一文一行とし、不要な改行・空欄を削除する；誤漢字と送り仮名は適宜修正する；すべて平仮名書きで読みにくいものは漢字に変換して修正するというものだった(金澤 2014: 16)。

表1 YNU コーパスの延べ語数・異なり語数・文数・読点数・【 】数(産出実数)

	中国				韓国				日本
	下位群	中位群	上位群	総計	下位群	中位群	上位群	総計	総計
延べ語数	26176	30818	31892	88886	21833	28288	31344	81465	79337
異なり語数	2090	2451	2453	4024	1521	2109	2260	3358	3633
文(句点)の数	1230	1492	1339	4061	1048	1242	1278	3568	3165
読点数の数	1384	1723	1500	4607	668	1136	1326	3130	3724
【 】の数	112	120	115	347	87	85	115	287	330

表1の産出実数の延べ語数で言うと、中国人留学生>韓国人留学生>日本人学生となっている。中韓の留学生と比べれば、日本人学生はより少ない語数でタスクを達成させていることが分かる。産出実数の異なり語数(厳密には異なり形態素数)で見ると、中国人留学生の作文全体が長い分、異なり語数ももっとも多い。それに対し、韓国人留学生の作文の延べ語数は日本人学生より多いにもかかわらず、異なり語数は日本人学生より少ない。

表2 YNU コーパスの延べ語数・異なり語数・文数・読点数・【 】数(調整頻度)

	中国				韓国				日本
	下位群	中位群	上位群	総計	下位群	中位群	上位群	総計	総計
延べ語数	100000	100000	100000	100000	100000	100000	100000	100000	100000
異なり語数	7984	7953	7692	4527	6967	7455	7210	4122	4579
文(句点)の数	4699	4841	4199	4569	4800	4391	4077	4380	3989
読点数の数	5287	5591	4703	5183	3060	4016	4230	3842	4694
【 】の数	428	389	361	390	398	300	367	352	416

表2の異なり語数の10万語あたりの調整頻度を見ると、日本人学生>中国人留学生>韓国人留学生の順で、日本人学生の異なり語数をもっとも多くなる。日本人学生のほうは語彙量が豊富であると予測できるため、短い作文の中でより多くの種類の語彙を産出していることが分かる。それに対し、韓国人留学生は異なり語数が少なく、同じ語が繰り返し使用されていることが示されている。

また、表1、表2から、中国人留学生は句読点をたくさん打っており、韓国人留学生は中国人留学生・日本人学生と比べれば、句点を打つわりに、読点をさほどたくさん打っていないことが分かる。日本人学生は中国人留学生ほど句読点をたくさん打っていないが、読点が句点より多いという状況は両者が似ている。表2の10万語あたりの調整頻度を見ると、日本人学生の作文には句点が一番少なく、言い換えれば文が長いことがうかがえる。文が長くなるということは、連体修飾表現をたくさん使うなど、文の構造が複雑になり、より難易度の高い文を産出していることが予測される。また、句読点の出現数については、それぞれの母語における句読点の重要さの違い、思考過程においてつい打ってしまうということも関わっている可能性がある。

【 】は「氏名」「住所」「電話番号」「メールアドレス」といった個人情報が入っている部分である。調整頻度で見ると、日本人学生(416)>中国人留学生(390)>韓国人留学生(352)の順となっており、日本人学生がタスクを達成させるために、一番よく個人情報を開示していることが分かる。

表1に示されたデータには形態素解析辞書「UniDic-mecab 2.1.2」の品詞分類による「記号(一般、文字)」「空白」「補助記号(一般、句点、読点、括弧等)」が含まれるが、品詞別の産出状況を分析するにあたり、これらのものを削除した。

留学生の産出された作文と日本人学生の作文との難易度を測るため、語彙密度(語彙のバラエティ)を分析する。語彙密度は文章の難易度や内容の豊富性を示す指標である。語彙密度を測定する指標として TTR、R 値が使用されることが多く、TTR、R 値が高いほど文章がバラエティに富むと言える。TTR は異なり語数(Type)を延べ語数(Token)で割る (Type/Token Ratio) ものである。R 値(Guiraud 値)は Type を Token の平方根で割った値で、データ間のサイズに差がある場合にも安定的に語彙密度を測定し、比較できると言われる(石川：2012)。表3に YNU コーパス(記号・補助記号・空白削除後)の延べ語数、異なり語数、語彙密度(TTR,R 値)を示している。

表3 YNU コーパス(記号・補助記号・空白削除)の延べ語数・異なり語数、語彙密度(TTR,R)

	中国				韓国				日本
	下位群	中位群	上位群	総計	下位群	中位群	上位群	総計	総計
延べ語数	22140	26339	27732	76211	19065	24677	27365	71107	68107
異なり語数	2026	2394	2400	3939	1455	2040	2197	3249	3555
TTR	0.092	0.091	0.087	0.052	0.076	0.083	0.080	0.046	0.052
R 値	13.62	14.75	14.41	14.27	10.54	12.99	13.28	12.18	13.62

表3から、中韓両国留学生の作文はともに、レベルが上がるにつれ、産出作文の異なり語数と延べ語数が増え、作文が長くなっていることが読み取れる。語彙密度(TTR、R 値)を見ると、中国人留学生の書かれた作文はどのレベルにおいても韓国人留学生より高く、より多様な語彙が使われていることが分かる。このことは注1に示したように、調査対象者の韓国人留学生に比べ、中国人留学生のほうが1級、N1の合格者が多く(韓国人留学生は未記入が8人)、レベル的に相対的に高いことによる可能性がある。ただし、タスク完成に向け、取り組み態度といった外部要素も関わってくるため、ここでは中国人留学生のほうは語彙量が豊富と断言できない。また、R 値を見ると、中国人留学生の中位群が一番高い。一方、中韓の留学生と比べれば、日本人学生は相対的に短い作文でタスクを完成させている。

表4 品詞別産出数、品詞構成比(中国人留学生のを降順基準に)

	中国		韓国		日本	
	産出数	構成比	産出数	構成比	産出数	構成比
名詞	23501	30.84%	20852	29.32%	20498	30.10%
助詞	23033	30.22%	21695	30.51%	21010	30.85%
動詞	10984	14.41%	10312	14.50%	10146	14.90%
助動詞	8798	11.54%	8823	12.41%	8370	12.29%
接尾辞	2300	3.02%	1915	2.69%	1758	2.58%
副詞	2062	2.71%	1740	2.45%	1555	2.28%
形容詞	1565	2.05%	1437	2.02%	1246	1.83%
代名詞	1062	1.39%	1226	1.72%	984	1.44%
接頭辞	896	1.18%	794	1.12%	873	1.28%
連体詞	757	0.99%	748	1.05%	568	0.83%
形状詞	619	0.81%	616	0.87%	640	0.94%
接続詞	371	0.49%	346	0.49%	295	0.43%
感動詞	168	0.22%	224	0.32%	117	0.17%
未知語	95	0.12%	379	0.53%	47	0.07%
総計	76211	100.00%	71107	100.00%	68107	100.00%

表4に示すように、日本人学生、中韓両国留学生が書かれた作文では、品詞別産出数を見ると、中国人留学生は名詞、助詞、動詞、接尾辞などの品詞9種類において多く、韓国人留学生は助動詞、代名詞、感動詞の3種類で最も多い。日本人学生は形状詞の産出数が中韓両国留学生より多いことが分かる。一方、品詞構成比に大きな開きは見られないが、中韓両国の留学生は接尾辞、副詞、形容詞、接続詞、感動詞の構成比が日本人学生より高く、日本人

学生は接頭辞、形状詞の構成比が高いことが見て取れる。なお、未知語とはアラビア数字、英語、中国語簡体字、韓国語の固有名詞の片仮名表記したものなどを指す。

表5 中韓両国留学生の作文レベル別の品詞出現数

行ラベル	中国			韓国		
	下位群	中位群	上位群	下位群	中位群	上位群
名詞	7088	8107	8306	5879	7542	8274
助詞	6610	7892	8531	5435	7036	8381
動詞	3063	3838	4083	2717	3632	3963
助動詞	2477	3119	3202	2401	3130	3292
接尾辞	716	760	824	453	681	781
副詞	605	673	784	481	632	627
形容詞	459	547	559	398	528	511
代名詞	292	366	404	399	430	397
接頭辞	233	326	337	168	267	359
連体詞	224	275	258	258	254	236
形状詞	181	205	233	183	212	221
接続詞	93	146	132	117	144	118
感動詞	64	55	49	97	117	132
未知語	35	30	30	79	72	73
総計	22140	26339	27732	19065	24677	27365

表5は中韓両国留学生の作文レベル別の品詞出現数を示したものである。両言語とも、レベルが上がるにつれ、名詞、助詞、動詞、助動詞、接尾辞、接頭辞、形状詞の産出が増えていく。副詞、形容詞、代名詞に関して、中国人留学生の作文はレベルが上がるにつれ、産出数が増えているが、韓国人留学生の作文は下位群から中位群にかけては増えるが、上位群ではやや下がっている。連体詞については中国人留学生の中位群が一番多く、上位群になるとまた下がっており、韓国人留学生の作文は下位群でもっとも多く、レベルが上がるにつれ産出数が減っている。接続詞については、両言語とも中位群が一番産出しており、上位群は下がっている。感動詞については、韓国人留学生の作文はレベルが上がるにつれ、産出数が増えるのに対し、中国人留学生の作文はかえって産出数が減っている。中韓両国留学生の作文における語彙は、品詞により違う産出様態を呈していることが分かる。

3. 2 品詞別の使用傾向

以下、特徴的な使用傾向を示す品詞について、品詞別に見る。産出した異なり語数が20語を超える場合、上位20語に限定して提示する。

下表6は感動詞の産出状況を示すものである。中韓両国留学生の作文とも、日本人学生より感動詞を多く産出している。詳細を見ると、70%前後は「今日は」「有り難う」「御早う」といった初級で習う挨拶語であった。特に、中国人留学生は「有り難う」を、韓国人留学生は「今日は」「御早う」「あの」をよく使っており、日本人学生は「はい」「うん」といった応答用の感動詞を留学生より多く産出していることが分かる³。

³ 感動詞と解析されたものの中で、誤解析されたものが一部含まれる。例えば、「あの」は感動詞「あのさ」として使われるものがある一方で、「あの本」のような連体詞の誤解析が見られた。また、「あっ」は感動詞の「あ」以外に、「あっという間」が1例見られた。

表6 感動詞の産出状況

感動詞(37個)	中国			中国集計	韓国			韓国集計	日本	総計
	下位群	中位群	上位群		下位群	中位群	上位群			
レベル										
今日は	22	25	20	67	35	37	27	99	36	202
有り難う	23	11	12	46	5	10	18	33	15	94
御早う			4	4	14	3	4	21	7	32
あの	3			3	10	6	3	19	10	32
はい	1	1	1	3			2	2	10	15
おー	1	7	1	9		1	1	2	2	13
うー	2	1	1	4	1	3	3	7	1	12
まー	2			2	1	1	2	4	4	10
うん	1			1	1		1	2	6	9
あー	3	1		4		1		1	1	6
いざ		3		3			1	1	2	6
さあ		2		2		1	1	2	1	5
ほら	2		1	3	1			1		4
初めまして							1	1	3	4
否							1	2	3	4
あっ		1	1	2					2	4
ううん							1	1	2	3
ああ	1			1			1	1	1	3
ねえ	1			1	1			1		2
今晚は			2	2						2
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
総計	64	55	49	168	79	72	73	224	117	509

表7 形状詞の産出状況

形状詞(205個)	中国			中国集計	韓国			韓国集計	日本集計	総計
	下位群	中位群	上位群		下位群	中位群	上位群			
レベル										
大事	9	7	21	37	20	16	20	56	20	113
奇麗	15	9	12	36	2	10	15	27	39	102
有名	16	11	16	43	7	8	16	31	21	95
懸命	6	4	8	18	5	5	8	18	57	93
好き	5	9	11	25	10	10	8	28	28	81
簡単	11	14	13	38	10	7	9	26	16	80
大変	4	4	7	15	14	10	15	39	21	75
大丈夫	10	5	3	18	9	3	6	18	32	68
色々	12	3	3	18	15	6	3	24	10	52
大切	7	7	4	18	10	4	1	15	19	52
可哀想	6		3	9	7	9	12	28	12	49
様々		4	2	6	1	6	11	18	19	43
真面目	1	2	4	7	3	8	2	13	23	43
重要	5	9	3	17	9	3	2	14	11	42
非常	2	12	16	30	3	4	1	8	4	42
可能	4	8	7	19	4	2	4	10	10	39
ぼろぼろ	1	6	2	9					28	37
沢山	4	2	10	16	4		2	6	9	31
如何		4	3	7	3	3	6	12	10	29
確か	2	4	3	9	3	5	3	11	9	29
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
総計	181	205	233	619	183	212	221	616	640	1875

表7は形状詞の産出状況を示すものである。形状詞の中には、副詞として使われることが多いと思われるもの(例：非常(に)、(一生)懸命)が混じっている。本稿ではこれを分類しなおしていない。中国人留学生は「有名」「簡単」「非常」「可能」のような二字漢語のものをたくさん産出している。韓国人留学生は「大事」「大変」「可哀想」のような感情を表す形状詞をたくさん産出している。日本人学生は「(一生)懸命」「大丈夫」「真面目」のような形状詞を多く産出している。「ぼろぼろ」に関して、タスク「七夕の物語」紹介に使われている語彙であるが、韓国人留学生の作文には1例も見られなかった。

表8 代名詞の産出状況

代名詞(29個)	中国			中国 集計	韓国			韓国 集計	日本 集計	総計
	下位群	中位群	上位群		下位群	中位群	上位群			
私-代名詞	98	125	131	354	96	137	113	346	322	1022
其れ	45	69	99	213	61	106	87	254	165	632
何	28	21	40	89	19	32	34	85	137	311
此れ	27	30	27	84	37	29	48	114	80	278
其処	4	23	17	44	18	17	14	49	63	156
何時	9	16	19	44	28	18	24	70	41	155
僕	21	13	4	38	46	6	10	62	10	110
俺	8		2	10	34	18	9	61	32	103
君-代名詞	12	19	1	32	3	13	12	28	5	65
彼	9	6	13	28	3	4	8	15	16	59
此処	5	5	13	23	8	10	7	25	9	57
貴方	2	2	4	8	19	7	8	34	6	48
彼女	8	11	4	23	5	2	8	15	4	42
何処	6	1	5	12	5	3	2	10	20	42
御前			1	1	11	7		18	20	39
誰	2	8	4	14	1	7	2	10	14	38
我々	5	2	9	16	2	8	4	14	5	35
私			5	5	2			2	9	16
此方		1	1	2	1		1	2	10	14
何れ		2	2	4			3	3	6	13
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
総計	292	366	404	1062	399	430	397	1226	984	3272

表8は代名詞の産出状況を示している。人称代名詞を見ると、中国人留学生は「僕」をよく産出するが、「俺」はあまり使わない。また、「我々」、「君」、「彼女」「彼」もよく産出する。韓国人留学生は「僕」、「俺」、「我々」、「君」、「貴方」をよく産出する。中韓両国の留学生はレベルが上がるにつれ、「僕」、「俺」の代わりに、「私」を使う場合が多くなる。日本人学生は親しい友人に対して使うと思われるが、「僕」より「俺」をよく産出している。人称代名詞の使用を控える日本人学生に比べ、留学生、特に韓国人留学生は上級～超級になっても人称代名詞を過剰に使用していることが分かる(韓:595>中:515>日:429)。一方で、中国人留学生は「俺」「御前」のようなややぞんざいな言い方を控える傾向にあることが分かる。指示代名詞「此れ」「其れ」「彼れ」「何れ」の産出状況を見ると、韓国人留学生には多く、日本人学生には少ないという、人称代名詞と同じ傾向が見られる(韓:371>中:301>日:251)。場所を表す指示代名詞「此処」「其処」「彼処」「何処」に関しては、日本人学生の使用は多いが、あまり差が見られない(日:92>中:84=韓:84)。

表9 連体詞の産出状況

連体詞(23個)	中国			中国 集計	韓国			韓国 集計	日本 集計	総計
	下位群	中位群	上位群		下位群	中位群	上位群			
其の	66	105	116	287	112	137	110	359	206	852
此の	101	105	97	303	92	75	77	244	206	753
同じ	10	13	8	31	10	10	10	30	30	91
或る	12	13	9	34	11	3	8	22	11	67
そんな	4	2	7	13	7	2	2	11	42	66
こんな	12	4	2	18	5	5	3	13	19	50
大きな	5	8	6	19	3	2	10	15	16	50
色んな	4	4	2	10	11	6	7	24	4	38
どんな	1	3	1	5		8		8	15	28
彼の	5	1	1	7	3	1	2	6	4	17
何の	1	2	4	7		1	1	2	4	13
我が		3		3	1	2		3	3	9
更なる		2		2			1	1	3	6
主な	1		1	2	2	1	1	4		6
小さな	2		1	3					2	5
本の		2	1	3			1	1	1	5
所謂		3		3			1	1		4
大した		2	1	3			1	1		4
単なる		3		3					1	4
あらゆる							1	1	1	2
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
総計	224	275	258	757	258	254	236	748	568	2073

表9は連体詞の産出状況を示すものである。中韓両国留学生の作文とも、「此の」「其の」「或る」「色んな」が日本人学生より多く産出している。特に、中国人留学生は「此の」、韓国人留学生は「其の」が多く、日本人学生は「そんな」「どんな」の使用数が多いことが分かる。産出する連体詞の中で、「此の」「其の」は三カ国とも70%~80%と高い割合を占めている(韓:603>中:590>日:412)。また、頻度こそ低い「大した」「所謂」「主な」は中韓両国の留学生の作文にもに見られたが、日本人学生の作文には見られなかった。

表10 助動詞の産出状況

助動詞(38個)	中国			中国 集計	韓国			韓国 集計	日本 集計	総計
	下位群	中位群	上位群		下位群	中位群	上位群			
だ	533	643	782	1958	514	721	755	1990	2084	6032
た	497	653	758	1908	606	723	730	2059	1857	5824
ます	571	766	681	2018	522	699	735	1956	1645	5619
です	367	388	284	1039	311	350	324	985	824	2848
ない	159	198	187	544	157	183	178	518	478	1540
れる	69	88	102	259	35	67	86	188	269	716
ず	41	104	77	222	42	85	79	206	236	664
様	45	52	90	187	51	86	106	243	221	651
たい	45	54	64	163	50	63	75	188	128	479
てる	36	34	31	101	35	53	70	158	205	464
せる	30	50	64	144	20	26	46	92	124	360
られる	32	31	34	97	11	23	26	60	84	241
ちやう	14	12	5	31	12	10	17	39	40	110
らしい	1	2	11	14	10	9	18	37	54	105
べし	10	10	5	25	4	9	11	24	34	83
みたい	6	5	9	20	6	4	8	18	28	66
させる	7	14	8	29		4	8	12	11	52
そう-様態	2		7	9	3	6	8	17	23	49
り	5	1	1	7	4	1	3	8	4	19
なり-断定		3		3	1	1	1	3	4	10
	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
総計	2477	3119	3202	8798	2401	3130	3292	8823	8370	25991

表10は助動詞の産出状況を示すものである。中国人留学生は「です」、「ます」、使役助動詞「せる」・「させる」を多く産出している。韓国人留学生は「れる」・「られる」⁴、推量助動詞「様だ」、希望助動詞「たい」の使用が多い。日本人学生は「れる」「られる」、様態・推量助動詞「らしい」「みたいだ」「そうだ」、継続・持続の意味を表す「てる」、断定助動詞「だ」を多く産出しているが、「です」「ます」「たい」が少ない。また、留学生はレベルが上がるにつれ、「れる」・「られる」、様態助動詞、推量助動詞、希望助動詞の産出数が増えている。

次の表11は終助詞の産出状況を示している。全体的に言えば、日本人学生は終助詞を多く産出するが、中国人留学生の産出は少ない。また、レベルが上がるにつれ、終助詞の産出数が増えている。中国人留学生が「の」「わ」を多く産出し、韓国人留学生は「ね、さ、の、じゃん、(「よ」「さ」は下位群で多いが、上・中位群で減る)」を多く産出している。日本人学生は「か、よ、な、ぞ」を多く産出している。また、「よ」「ね」の産出数の中に、「よね」は99例で、内訳として日本人学生が多く産出している(日:60>韓:30>中:9)。

⁴ 受け身、可能、尊敬、自発を表す用例すべてが含まれる。

表 1 1 終助詞の産出状況

終助詞(19個)	中国			中国 集計	韓国			韓国 集計	日本 集計	総計
	下位群	中位群	上位群		下位群	中位群	上位群			
か	68	77	100	245	51	74	87	212	299	756
よ	56	69	61	186	77	68	62	207	232	625
ね	45	52	70	167	70	66	83	219	207	593
な	16	16	21	53	15	22	39	76	117	246
さ		8	2	10	19	11	12	42	25	77
の	10	6	11	27	6	17	14	37	12	76
じゃん	1	1	2	4	4	7	5	16	8	28
わ		4	5	9	2	3		5	4	18
ぞ	1		3	4					9	13
もの	1	1	1	3	1	1		2	5	10
ぜ						5		5	2	7
い	1		1	2	2	1		3	1	6
け	2			2	1			1	1	4
べい			3	3						3
のう									2	2
や						1	1	2		2
ねん							1	1	1	2
もが									1	1
ばや		1		1						1
総計	201	235	280	716	248	276	304	828	926	2470

表 1 2 準体助詞の産出状況

準体助詞(1個)	中国			中国 集計	韓国			韓国 集計	日本 集計	総計
	下位群	中位群	上位群		下位群	中位群	上位群			
の	196	174	319	689	200	241	234	675	792	2156
総計	196	174	319	689	200	241	234	675	792	2156

表 1 2 は準体助詞の産出状況を示している。日本人学生が準体助詞を多く産出している。

4. 考察とまとめ

以上では、中韓両国の上級～超級日本語学習者が書かれた作文を品詞の使用実態を基に量的に使用傾向を見た。考察した結果、留学生は上位群へレベルが上がるにつれ、作文語数が増え、品詞の大半は異なり語数、延べ語数が増え、語彙量が豊富になり、相手との交渉表現、文構造が複雑化していることが読み取れる。

中韓両国の留学生の作文とも、日本人学生より感動詞を多く産出している。初級の挨拶語「今日は」「有り難う」が多く見られた。特に、中国人留学生は「有り難う」を、韓国人留学生は「今日は」「御早う」「あの」をよく使っており、日本人学生は「はい」「うん」といった応答用の感動詞を留学生より多く産出している。

形状詞に関しては、中国人留学生は「有名」のような二字漢語のものをたくさん産出している。中国語母語の影響があると思われる。韓国人留学生は「大変」「可哀想」のような感情を表す形状詞を、日本人学生は「(一生)懸命」「大丈夫」「真面目」「ぼろぼろ」のような形状詞をたくさん産出している。

人称代名詞に関しては、中国人留学生は「僕」をよく産出するが、「俺」はあまり使っていない。また、「我々」「君」「彼女」「彼」もよく産出している。韓国人留学生は「僕」、「俺」、「我々」、「君」、「貴方」をよく産出している。中韓両国の留学生ともレベルが上がるにつれ、「僕」、「俺」の代わりに、「私」を使う場合が多くなっている。日本人学生は親しい友人に対して使うと思われる「俺」が、「僕」より多く産出している。人称代名詞の使用を控える日本人学生に比べ、留学生、特に韓国人留学生は上級～超級になっても人称代名詞を過剰に使用する。一方で、中国人留学生は「俺」「御前」のようなややぞんざいな言い方を控える傾向にある。指示代名詞「此れ」「其れ」「彼れ」「何れ」に関しては、

韓国人留学生には多く、日本人学生には少ないという、人称代名詞と同じ傾向が見られる。

連体詞を見ると、中韓両国留学生とも、「此の」「其の」「或る」「色んな」が日本人学生より多く産出している。特に、中国人留学生は「此の」、韓国人留学生は「其の」が多く、日本人学生は「そんな」「どんな」の使用数が多いことが分かる。産出する連体詞の中で、「此の」「其の」は三カ国とも、70%~80%と高い割合を占めている。

中国人留学生は「です」、「ます」、使役助動詞を多く産出している。韓国人留学生は「れる・られる」、「様だ」、「たい」の使用が多い。日本人学生は「れる・られる」、様態・推量助動詞、「てる」、断定助動詞「だ」を多く産出するが、「です」「ます」「たい」が少ない。また、留学生はレベルが上がるにつれ、「れる・られる」、様態助動詞、推量助動詞、希望助動詞の産出数が増える。

終助詞に関しては、全体的に日本人学生は終助詞を多く産出するが、中国人留学生の産出は少ない。また、レベルが上がるにつれ、留学生の終助詞の産出数が増えている。中国人留学生は「の」「わ」を多く産出し、韓国人留学生は「ね、さ、の、じゃん」を多く産出している。日本人学生は「か、よ、な、ぞ」を多く産出し、また、「よね」も多く産出している。

日本人学生は準体助詞を多く産出している。準体助詞を使うことにより、文が複雑になり、視点固定にも寄与できるという特徴を持っている。中韓両国の留学生もレベルが上がるにつれ、準体助詞の使用数が増えており、日本人学生の使用状況に近づいている。

全体的に言えば、留学生の作文には挨拶語、人称代名詞、指示代名詞、連体詞「此の、其の」、助動詞「たい」が過剰に使用され、様態・推量助動詞、継続・持続助動詞の過少使用が見られた。中韓留学生の間にも、「此の、其の、此れ、其れ」の産出数、受身・尊敬、使役助動詞の産出数などの違いが見られる。中国人留学生は「です、ます」の多用、「俺、御前」の過少使用から、中国人留学生は相手が「親」でも、相対的に改まり度の高い文体で書いていることがうかがえる。日本人学生の作文は句点が少なく、指示代名詞、連体詞「此の、其の」が少なく、また準体助詞が多いことから、文が長く、文構造が複雑だと推測される。

レベルが上がるにつれ、留学生は感動詞の産出が減り、「僕・俺」が「私」に取って代わられ、形状詞・助動詞・終助詞が増加傾向にあるなど、日本人の使用実態に近づいている。

以上の考察により、留学生が日本語のレベルが上がるにつれ、全体的傾向として日本人学生の言語産出に近づいていることが分かる。ただし、人称代名詞の多用、様態・推量助動詞の使用が少ないなど、上級~超級になっても日本人学生の産出と違う特徴が見られる。

参考文献

- 石川慎一郎(2012)『ベーシックコーパス言語学』ひつじ書房
 小野望・田中省作・持尾弘司(2007)「母語学習者コーパスの基礎調査」『筑紫女学園大学・短期大学部人間文化研究所年報』18, 27-36, 筑紫女学園大学・短期大学部人間文化研究所
 金澤裕之(編)(2014)『日本語教育のためのタスク別書き言葉コーパス』ひつじ書房
 橋本直幸(2011)「学習者コーパスから見る超級日本語学習者の言語特徴—2つの観点から—」『日本語教育文法研究のための多様なアプローチ』ひつじ書房
 毛文偉(2013)「中国日本語学習者作文詞彙量及高頻詞目研究」『外語電化教学』152, 9-15
 山内博之(2009)『プロフィシェンシーから見た日本語教育文法』ひつじ書房

使用データと形態素解析ツール

金澤裕之(編)(2014)『日本語教育のためのタスク別書き言葉コーパス』付属CDのデータ
 形態素解析処理ソフト「茶まめ」、解析器「MeCab 0.996」、解析用辞書「UniDic-mecab 2.1.2」