NICT JLE vs. JEFLL: *n*-gramを用いた語彙・品詞使用の発達

投野由紀夫

2008
英語コーパス研究　第15号　抜刷
英　語　コ　ー　パ　ス　学　会

NICT JLE vs. JEFLL: *n*-gram を用いた 語彙・品詞使用の発達

投野由紀夫[1]

1. はじめに

本論では2つの学習者コーパス，NICT JLE および JEFLL の全般的な特徴を把握するために，n-gram 分析を通した語彙・品詞使用の発達傾向を概観する。まず各コーパスの大まかな設計基準を解説し，続いて単語 n-gram の全体傾向，特に n-gram 比較で顕著だった言語特徴，そして品詞 n-gram の全体傾向と特徴的なパターンに関してまとめる。

2. 分析対象のコーパス

2.1 NICT JLE Corpus

NICT JLE Corpus は，アルクが開発した英会話能力試験 Standard Speaking Test（SST）を書き起こした話し言葉コーパスである。SST は 15 分間の口頭能力インタビューテスト（Oral Proficiency Interview：以下 OPI）で，世界的に普及している ACTFL OPI に準拠し，日本人英語学習者用に開発されたものである。テストは前後のウォームアップ，ワインドダウンを除くと，イラスト描写，ロールプレイ，ストーリー作りの3つのタスクがメインになっており，15 分の面接の後，2名の訓練された判定者により，9段階のレベル判定がなされる。

NICT JLE Corpus はこのアルクに所蔵されていた膨大なインタビュー音源を情報通信機構（NICT）がアルクと共同で書き起こしたもので，1281 名分の書き起こしデータを含んでいる。全体のコーパスのサイズは約 200 万語。各被験者のデータには SST レベル（1〜9）が付与されているので，会話能力と実際のインタビュー内容をコーパス分析して，語彙・文法などの特徴とレベルの関係などを研究するのに有益なデータとなっている。詳細は和泉他（2004），さらに通信総研（NICT の前身）の報告書などを参照されたい。

『英語コーパス研究』第 15 号（2008），pp. 119-133

2.2 JEFLL Corpus

JEFLL Corpusは，日本人英語学習者（中学1年生～高校3年生）約1万人を対象に自由英作文データを書かせたものをコーパス化したものである。6種類の英作文タスク（叙述文・論説文各3種類）から1種類を選択して，授業時間内で20分間，辞書なしで書かせる。タスク，学年，学校種，学校レベル別にデータを検索でき，サイズは2007年9月現在で約70万語である。JEFLL Corpusはインターネット上でweb検索の形で公開されており，[2] JEFLL全体の情報も研究用ホームページで公開されている。[3]

3. 分析方法

2つのコーパスのレベル別の語彙・品詞の全般的傾向を見るために，n-gram分析を行った。分析上，問題となると思われる点を中心に下記のような前処理を施した：

(a) JEFLLに関しては，日本語部分は#JP#というタグに置き換えて1まとまりとした。
(b) NICT JLEについては〈B〉（被験者）のみの発話に限定し，Fillerは#F#，JPは#JP#とした。
(c) 品詞のn-gramを得るために，CLAWSで品詞タグを付与。タグ修正などは一切行わなかった。

N-gram抽出については，Michael Barlow氏の開発したコロケーション分析ソフトCollocateを使用した。

4. 結果(1)：単語n-gram

4.1 単語n-gramのヴァリエーション

まず，それぞれのコーパス全体における単語n-gramの種類がどのように増えているかを見てみる。図1，2はJEFLLの結果である。図1を見ると，中学1年（JH1）から高校3年（SH3）まで，全体のコーパス・サイズの違いがあるので，かなりパターンにばらつきがあるように見えるが，図2のようにそれぞれの学年におけるtrigram（3語の連鎖）の組み合わせの種類と全体のtrigram総数の比をとると，ほぼ学年を追うごとに伸びが観察される。図2に見られるような伸びは，学年を追うごとに，trigramに現れる単語の連鎖のヴァリエーションが豊富になっていることを示している。それだけ，定型表現を使いつつも，生産的な表現を多くするようになっている証拠ではないかと思われる。

一方，同様の分析を話し言葉に対して行ってみると，図3，4のようになる。図3はやはりレベル別のコーパス・サイズの影響を受けてレベル4，5を中心に特に多くなっているが，図4のtrigramのヴァリエーションと総頻度の割合

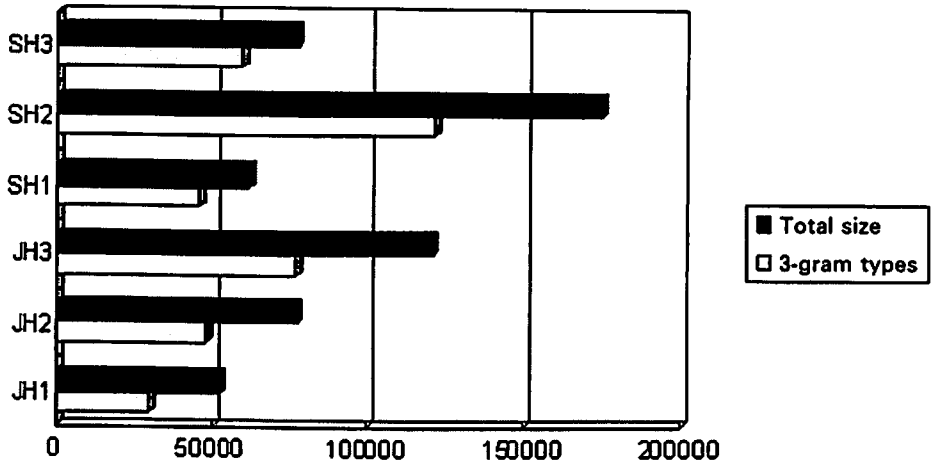

図1　JEFLL における学年ごとの trigram の総量と種類の変化

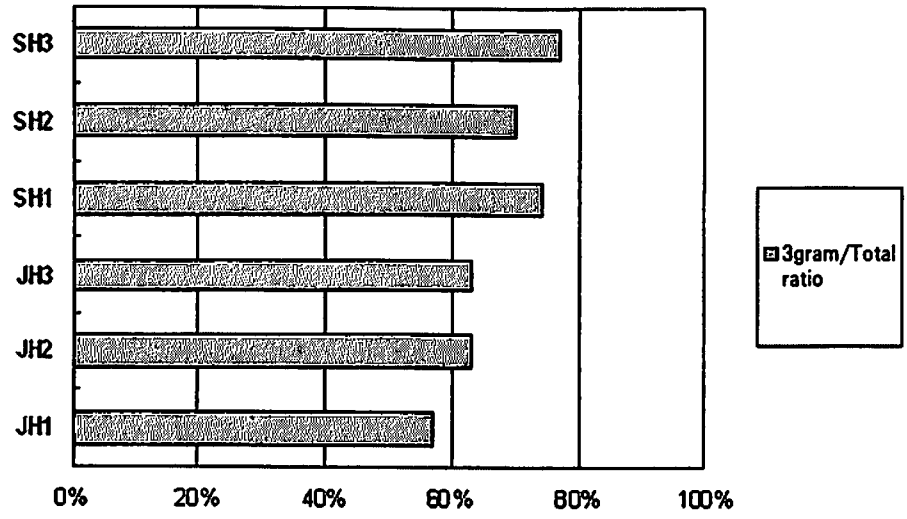

図2　JEFLL における Trigram の種類と総量の割合の変化

を見ると，レベル 3，4 を境目にレベル 9 まで伸びを示している。レベル 1 の割合が極めて高いのは日本語（#JP#）の含まれた trigram が頻出しているからで，これはレベル 3 くらいまでに大部分が消滅する。そのために中間レベルで割合は低くなるが，そこからレベルが上がるに連れて trigram のパターンも豊富になる。

4.2　Filler

表 1 は NICT JLE Corpus における filler を含む trigram 連鎖がレベルを経て消失していく様子を示したものである。表 1 では，図の左側から右側に向かっ

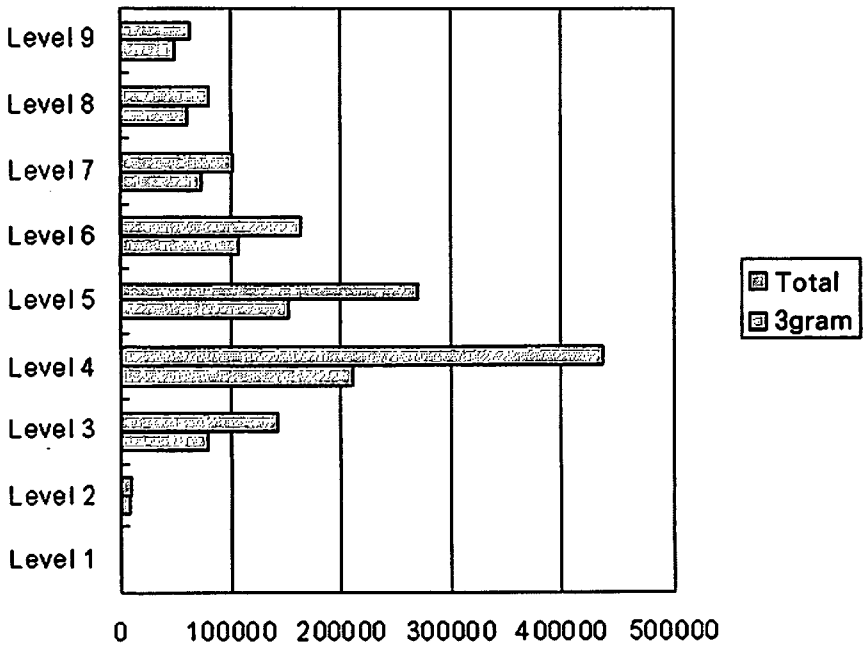

図3 NICT JLE Corpus におけるレベルごとの trigram の変化

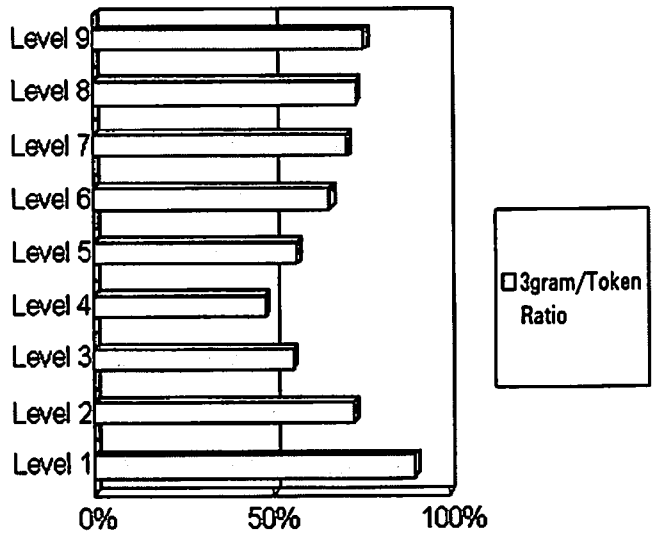

図4 NICT JLE Corpus における trigram 種類と総数の割合の変化

て SST レベルが上がるに従って，filler を含んだ網掛けのパターンの数が減少していることがわかる。これは filler を含むいわゆる言いよどみ（dysfluency）の現象が，レベルが上がるに連れて解消されていることを示している。

表1　Filler を含む trigram（網掛け部分）のレベルごとの推移

Level 1		Level 2		Level 3		Level 4		Level 5		Level 6		Level 7		Level 8		Level 9	
Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq
64	. #F# .	617	. #F# .	4124	. #F# .	8302	. #F# .	3750	. #F# .	1868	. #F# .	1157	. #F# .	603	. #F# .	382	. #F# .
21	#F# . #F#	305	#F# . #F#	1585	. Yes .	3572	. Yes .	1978	. Yes .	1031	. Yeah .	559	. Yeah .	440	. Yeah .	254	I do n't
21	. Yes .	192	. Yes .	1538	#F# . #F#	2669	. Yeah .	1372	. Yeah .	674	. Yes .	386	. Yes .	264	you know ,	251	. Yeah .
19	. #J# .	180	. #F# #F#	1385	. #F# #F#	2603	#F# . #F#	1013	#F# . #F#	496	. And #F#	363	I do n't	261	. And #F#	230	. #F# I
13	. #F# #J#?	90	. Yeah .	930	#F# #F# #F#	2552	. #F# #F#	998	#F# #SC# #F#	495	I do n't	310	. And #F#	259	I do n't	198	. Yes .
11	Yes . #F#	89	. #J#? .	823	#F# #SC# #F#	2334	#F# #SC# #F#	963	. And #F#	451	. O K.	278	you know ,	242	. Yes .	167	. O K.
10	. #F# #F#	84	Yes . #F#	823	. Yeah .	1931	#F# #R# #F#	833	. #F# #F#	380	#F# . #F#	252	. O K.	216	. O K.	163	. And #F#
9	#J?# . #F#	80	#F# #F# #F#	711	Yes . #F#	1871	#F# #F# #F#	740	Yes . #F#	351	you know ,	246	. #F# I	188	. #F# I	138	. And I
7	#F# . #J?#	64	. #F# #R#	651	#F# #F# #R#	1623	Yes . #F#	656	#F# #R# #F#	334	#F# #SC# #F#	238	. So I	186	, you know	134	. #F# #F#
5	#F# #F# .	61	#F# #F# .	640	#F# #R# #F#	1555	. And #F#	651	. O K.	324	. #F# #F#	232	#F# . #F#	183	. So I	134	. So I
5	#F# . Yes	55	#F# . #R#	566	. #F# #R#	1376	. #F# I	609	. #F# I	295	. #F# I	212	. #F# #F#	166	. And I	104	#F# . #F#
5	#J?# . #J?#	51	. #F# #J?#	526	. #F# I	1355	#F# #F# #R#	598	. Thank you	285	. Thank you	200	. And I	157	. #F# #F#	98	you know ,
4	#F# #J?# .	49	#F# . Yes	524	#F# . #R#	1131	. #F# #R#	567	I do n't	280	. So I	179	. Thank you	150	#F# . #F#	97	do n't know
4	#F# . .	49	#J?# #F#	484	#F# #F# .	1045	. Thank you	527	#F# #F# #F#	267	Yeah . #F#	161	, you know	135	. Thank you	88	. And then
4	. #J?# #J?#	49	Yeah . #F#	448	. Thank you	1043	I do n't	513	. So ,	246	Yes . #F#	159	#F# . I	110	Yeah . #F#	83	. It 's
4	Yes . Yes	44	#F# #F# #R#	394	#F# #R# #R#	1007	Yeah . #F#	476	. So I	245	. So ,	150	. It 's	109	. But #F#	81	. Thank you
3	#F# #J?# #F#	44	. . #F#	393	#F# #R# I	1002	#F# #R# I	469	. So #F#	241	. But #F#	149	I have to	109	O K. #F#	78	. But #F#
3	#F# . #R#	43	#F# . #J?#	375	Yeah . #F#	988	. O K.	424	Yeah . #F#	228	O.K. #F#	143	. But #F#	106	#F# #SC# #F#	78	Yes . #F#
3	#J?# #F# .	43	. No .	346	Thank you .	944	I want to	420	#F# #R# I	206	. It 's	143	do n't know	105	#F# it 's	74	, I think
3	. #F# #R#	40	. O K.	341	#F# #F# #SC#	921	#F# #F# .	411	. #F# #R#	200	#F# . I	141	Yes . #F#	102	. I 'm	70	#F# it 's
3	. #J?# #F#	40	Thank you .	328	. And #F#	910	#F# . #R#	397	O K. #F#	196	. And I	140	Yeah . #F#	96	. So #F#	70	Yeah . #F#
3	. Thank you	39	#F# yes .	326	. #F# #SC#	879	. So #F#	388	#F# #F# #R#	191	. Yeah ,	138	O K. #F#	96	do n't know	69	. Mh-hmm .
3	. Yeah .	35	#F# #SC# #F#	324	#SC# #F# #R#	844	#SC# #F# #F#	381	Thank you .	189	. I 'm	120	. I 'm	95	. Mh-hmm .	68	, you know
3	Thank you .	35	#F#... I	324	I do n't	816	Thank you .	371	I want to	185	I have to	116	Thank you .	94	. It 's	68	do n't have
3	Yes . #J?#	35	. #R# I	321	#SC# #F# #F#	805	#F# #R# #R#	371	very much .	184	, #F# I	112	#F# #SC# #F#	89	Thank you .	67	a lot of
3	old . #F#	35	. Thank you	321	I want to	781	#SC# #F# #R#	360	. But #F#	182	#F# #F# #F#	111	. So ,	85	I have to	65	. I 'm
2	#F# #F# #J?#	34	#F# #R# #F#	316	. O K.	780	#F# #F# #SC#	342	#F# . I	182	Thank you .	109	. And ,	82	. And then	65	O K. #F#
2	#F# #J?# #J?#	33	. #F# yes	313	. #SC# #F#	779	#F# #F# I	338	. #F# #SC#	182	very much .	109	. So #F#	81	, #F# I	65	Thank you .
2	#F# #R# #F#	32	. #F# I	301	#F# #F# I	734	#F# . I	335	#F# #F# #SC#	181	#F# #R# #F#	107	. Yes ,	80	Yes . #F#	62	. So #F#
2	#F# . Saturd	30	#F# #R# #R#	291	#F# . #SC#	730	O K. #F#	335	, #F# I	173	#F# I 'm	106	#F# #F# .	76	#F# . I	61	O K .
2	#F# . This	29	I do n't	290	#R# #SC# #F#	713	#R# #F# I	333	. #SC# #F#	173	do n't know	104	, I think	74	. And the	59	#F# I do
2	#F# . Where	28	My name is	282	#F# . I	709	. #F# #SC#	323	#SC# #F# #F#	171	. Yes ,	101	#F# it 's	73	. it 's	59	. But I
2	#F# . XXX03	28	name is XXX0	274	. #R# I	655	. #R# #F#	318	Yes . Yes	168	. So #F#	99	very much .	72	a lot of	57	#F# I think
2	#J?# #J?# #F	27	is XXX02 .	269	#F# . Yes	652	. #SC# #F#	312	#SC# #F# #SC#	164	. #SC# #F#	97	. But I	71	very much .	55	#F# #F# I
2	#J?# #J?# .	26	#F# #F# #J?#	257	Yes . Yes	640	#R# #F# #F#	311	#F# #F# .	162	, you know	96	. And #SC#	70	, I think	55	#F# I 'm
2	#J?# No .	26	#F# . #SC#	256	#R# #F# I	633	. So ,	310	I have to	155	you very much	95	#F# . Yes	69	. And #SC#	55	. Mm-hm .
2	#J?# Yes .	25	. #F# #SC#	255	#SC# #F# #SC	622	#SC# #F# #SC#	308	#F# it 's	153	#F# #F# .	95	I did n't	69	I did n't	50	. And the
2	. #F# XXX05	25	. My name	249	O K. #F#	570	Yes . Yes	303	#SC# #F# #R#	150	, and #F#	94	#F# #F# #F#	67	. So ,	49	. And we
2	. #J?# Yes	24	#F# #F# I	245	#F# #R# #SC#	567	#F# . Yes	302	. It 's	148	#F# #R# I	94	. And #R#	67	O K .	48	, it 's
2	. . #F#	24	#F# . .	241	#R# #F# #F#	566	, #F# #R#	297	#F# #F# I	148	#F# . Yes	89	#F# I 'm	66	#F# I 'm	48	, too .
2	. Best .	24	I go to	240	#R# #R# #F#	565	#F# . #SC#	296	, #F# #R#	144	. And #SC#	87	do n't have	65	you very much	48	. So .
2	. Consulting	24	Yes . Yes	240	. #R# #F#	564	very much .	292	#F# . #R#	141	. And ,	85	. So .	64	#F# #F# I	47	I have to
2	. Do you	23	#R# #SC# #F#	225	#F# I like	557	#R# #SC# #F#	283	. #R# #F#	140	Yeah . Yeah	85	you very much	61	#F# #F# #F#	46	, I do
2	. Friend .	23	. #J?# #J?#	221	. #J?# .	556	. So I	277	you very much	139	#F# it 's	84	. And then	61	. Yes ,	46	. I do
2	. No ,	23	I live in	213	#R# #F# #R#	554	. And #R#	270	. Uh-huh .	138	#SC# #F# I	82	. #F# #SC#	60	#F# #R# I	45	. #SC# #F#
2	. No .	23	O K. #F#	208	is XXX02 .	540	#SC# #F# I	266	#F# I 'm	138	a lot of	82	. #SC# #F#	60	I mean ,	45	. O K
2	. Today .	23	yes . #F#	202	. And #R#	531	#F# yes .	263	a lot of	137	. And then	81	, I do	59	. #SC# #F#	45	. Uh-huh .
2	Ha . #F#	22	#F# #J?# .	201	#J?# . #F#	523	, #F# I	262	to go to	134	, I think	81	. I do	59	. But I	44	#F# #SC# #F#

4.3 モデル文からの解放

JEFLL は作文タスクの指示文に，モデルとして簡単な英作文のサンプルを載せている。コーパス作成に当たり，このようなモデル文の影響がどの程度見られるかが注目された。太田 (2007) の分析によると，4-gram 上位 10 位に現れるモデル文と同一構造のパターンの割合は中 1 では 9 割近くに上っているのに対し，その後 50% (中 2)，30% (中 3)，高校 1 年以上は 10%以下，と着実に減少している。すなわち，中 1 の段階ではかなりの生徒がモデル文をもとに作文を書こうとしているが，中 2 以降ではモデル文そのままの表現を使って書く，という生徒は徐々に減少する。

4.4 動詞の時制

JEFLL では中 1 では上位 100 の trigram 中，2 例しか過去形を含む連鎖が出現しないが，中 2 以降そのパターンの出現率が増加する。be 動詞が多いが，had や came, went, などの移動動詞も早い時期から過去形で使用される。教科書で指導をされた順序とコーパスの出現頻度の相関が比較的高い項目である。

一方，NICT JLE では，レベル 1〜4 付近までは，基本動詞 (be, have, want, go など) による短文の羅列が多いが，レベル 5 から上に過去形 had が出現する。

4.5 接続詞

接続詞は JEFLL, NICT JLE ともに発達の重要な指標になる。談話マーカーの観点からは本シンポジウムの小林・山田が重点的に調査しているが，それ以外を概観しておく。

まず JEFLL では文頭の but の多用が特に低学年に顕著である。中 1 では上位 100 位の trigram 中に 8 パターン見られる。それが中 2〜3 年では 5，6 パターン，高 1〜3 では 2〜3 パターンと半減していく。代わって，When I was ..., If there is a ..., If I don't have ... などの条件節を導く接続詞が学年を上がるに連れて頻度が高くなっていく。

NICT JLE の場合は，レベル 4 から so が，レベル 5，6 から but, or などの等位接続詞の多用が観察される。レベル 7〜9 では，I think that ... などの従属節を導く that の使用が顕著になる。

4.6 冠詞

冠詞の獲得は日本人にとって最も苦手な項目の 1 つである (Tono, 2000)。表 2-1，2-2 は冠詞を含むパターンの出現推移を JEFLL と NICT JLE で比較した図である。JEFLL では the, a 共に比較的低学年から出現しているが，これらは学年が低いほど，in the morning とか a lot of のような定型パターンでの使用が顕著である。冠詞の用法は高校になると着実に増加していることがわかる。一方，NICT JLE の場合は，話し言葉であるため，冠詞の補充は一層少ない。全般

表 2-1　Trigram における冠詞を含むパターンの出現推移（JEFLL）

J1		J2		J3		S1		S2		S3	
Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq
576	#JP# . I	426	. So I	644	I do n't	224	I do n't	600	I do n't	235	I do n't
441	. I like	425	I do n't	528	. So I	170	#JP# . I	450	I want to	188	I want to
417	. But I	408	#JP# . I	527	. But I	166	. But I	348	. But I	166	. So I
405	I do n't	378	I will bring	425	#JP# . I	164	. So I	343	. So I	146	. But I
296	. So I	347	. But I	371	. I was	153	. I was	295	. It is	131	. It was
236	#JP# . #JP#	335	. I like	345	I will bring	147	I want to	291	I ca n't	126	. It is
208	I will bring	275	. I was	336	I want to	139	school festival .	281	. I think	122	in the morning
200	#JP# . But	255	. I will	327	. I like	130	. It was	274	. I was	118	and so on
195	in the morning	233	I want to	283	. I will	114	our school festival	273	. When I	116	I ca n't
195	is #JP# .	217	in the morning	283	in the morning	112	. Our class	269	. I like	106	. Our class
189	very much .	193	the morning .	264	. And I	109	in the morning	239	a lot of	105	school festival .
178	the morning .	170	very much .	263	I ca n't	103	a lot of	237	and so on	105	so on .
177	. Our class	160	I was very	254	. I do	99	. So ,	235	#JP# . I	103	. Our school
175	Our school festival	159	#JP# . But	228	. I think	98	I was very	229	. And I	101	Our school festival
174	. It 's	152	very #JP# .	218	. I usually	90	I will bring	224	. But ,	98	a lot of
173	very #JP# .	151	. So ,	212	the morning .	88	. And I	221	. If I	95	. I think
171	. I will	149	. It is	209	. Because I	88	. I think	216	so on .	94	. I was
161	. Our school	148	. It was	196	do n't have	87	. Because I	215	. I do	92	. #JP# is
158	. I do	143	#JP# . So	195	I was very	87	Our school festival	210	. I have	92	I will bring
154	and #JP# .	142	. I want	188	very much .	86	. I like	210	our school festiva	92	our school festival
145	. I usually	138	#JP# . #JP#	185	. It is	84	. Our school	208	school festival .	91	the morning .
143	do n't have	138	#JP# . He	179	I usually have	83	I did n't	204	, I will	89	. I do
142	. It is	137	do n't like	171	. I have	82	. I will	204	. So ,	88	do n't have
140	But I do	134	. Because I	158	#JP# . But	81	. I do	197	in the morning	87	#JP# . I
138	#JP# . Our	134	. I do	158	. It was	79	I ca n't	191	do n't have	86	. So ,
137	. I eat	130	. Our class	158	. When I	77	the morning .	182	I will take	85	. I like
134	#JP# . It	128	do n't have	150	. I want	70	. If I	181	. It was	85	. I will
131	. #JP# I	127	. I usually	148	. He was	68	So , I	177	. I want	84	. If I
129	morning . I	126	. Urashima Taro	147	. If I	67	a big earthquake	170	a big earthquake	83	. I want
125	. I 'm	126	I usually have	147	. So ,	66	do n't have	168	. Because I	80	. And I
124	. I have	123	. Our school	144	#JP# . So	65	, too .	167	, but I	78	. I have
120	, too .	122	. And I	139	I did n't	65	school festival ,	164	, too .	71	, I will
117	. I #JP#	122	. But ,	136	. One day	61	, so I	159	, so I	68	. Urashima Taro
115	. Urashima Taro	118	Our school festiva	134	a lot of	61	the school festival	152	I think that	65	I could n't
112	. I want	117	. It 's	133	. And he	60	. I have	148	Our school festiva	63	. I usually
109	#JP# . So	115	Urashima Taro was	126	, too .	60	. I want	143	, and I	63	I usually have
108	is very #JP#	110	a lot of	125	. I 'm	59	#JP# . But	143	. Our school	63	very much .
107	I usually have	106	#JP# . It	125	. So he	58	I could n't	142	. I will	59	. He was
102	#JP# and #JP#	106	I ca n't	123	#JP# . He	58	and #JP# .	137	for breakfast .	59	. It 's
102	every day .	103	is #JP# .	120	and #JP# .	57	, I 'll	136	. For example	58	. Because I
101	. I love	102	morning . I	117	. It 's	57	. But ,	132	school festival is	58	. For example
98	. And I	100	. I 'm	114	dream . I	55	, but I	131	very much .	58	. I 'm
98	rice and #JP#	98	. I think	111	morning . I	55	. It is	127	. It 's	57	. When I
98	school festival is	97	it . I	109	and so on	54	#JP# . It	127	I was very	56	, so I
97	#JP# . He	97	very happy .	107	do n't like	54	I went to	126	. I 'm	56	I was very

（表 2-1 つづき）

95	#JP# is #JP#	95	I went to	106	, I will	54	very much .	124	So , I	55	for me .
95	do n't like	95	was #JP# .	106	me . I	53	, I was	121	, I was	55	morning . I
88	#JP# in the	93	. I have	104	, so I	53	, I will	118	For example ,	51	, #JP# ,
84	I like #JP#	92	#JP# and #JP#	102	breakfast . I	53	. Urashima Taro	115	the morning .	50	. And he
82	bread and milk	92	. I #JP#	101	. Then I	52	very happy .	112	there is a	50	. One day
81	#JP# #JP# .	92	So I will	100	Our school festi	50	school festival was	109	. He was	49	For example ,
81	much . I	92	first . I	98	. But he	49	#JP# #JP# #JP#	109	I went to	46	. But ,
78	will bring a	90	, too .	98	. Our class	49	, and I	109	it . I	46	. But he
75	#JP# . My	89	is very #JP#	96	do n't want	49	. When I	108	, I have	45	. So he
75	It 's very	87	. One day	95	I could n't	48	#JP# . So	108	, I think	45	I did n't
75	Urashima Taro was	86	. #JP# I	95	so on .	48	. I usually	108	I did n't	45	to go to
74	. But he	86	. He was	94	, but I	47	there is a	108	When I was	44	, but I
73	#JP# is very	86	had a #JP#	92	was #JP# .	46	. I 'm	104	breakfast . I	44	. However ,
73	I like rice	86	will bring a	91	do n't know	46	Our class had	103	will take out	44	Our class had
71	. He was	85	#JP# in the	90	I went to	46	class had a	98	I have to	44	and #JP# .
70	. But #JP#	85	I did n't	90	have breakfast	46	festival . I	98	is a big	43	, I was
70	. But ,	85	So , I	89	. I always	45	#JP# and #JP#	96	I could n't	43	, I would
70	So I gave	85	and #JP# .	88	. Urashima Taro	45	n't want to	95	, I would	43	So , I
70	festival is very	84	Taro was very	88	n't want to	44	, #JP# ,	94	I had a	43	school festival ,
69	. I am	83	But I do	88	will bring my	44	I 'll take	93	, I 'll	43	school festival was
68	Our class #JP#	82	rice and #JP#	87	. Our school	44	It was very	93	, I can	42	Urashima Taro was
67	. #JP# .	80	I like it	87	So , I	44	Urashima Taro was	93	for me .	41	#JP# and #JP#
67	I ca n't	78	. But he	86	Urashima Taro wa	44	do n't want	93	take out my	41	, and I
67	I gave #JP#	77	. We were	86	is #JP# .	43	. It 's	92	this year .	41	. This is
66	I like bread	76	One day I	85	#JP# . And	42	I usually have	91	. Our class	41	I think that
66	a lot of	73	#JP# . We	85	. But ,	42	morning . I	91	do n't like	41	It was very
66	gave #JP# to	72	, I will	84	I 'll bring	41	I think that	90	, #JP# ,	40	, I have
66	very happy .	72	. But we	84	One day I	40	. If there	90	rice and #JP#	40	class had a
65	#JP# to buy	72	the #JP#	83	it . I	39	, I think	89	I have a	40	do n't want
64	morning . But	71	and so on	82	#JP# in the	39	. Then ,	88	I like rice	40	it . I

（注：冠詞の含まれた trigram を網掛けにしてある）

表 2-2　Trigram における冠詞を含むパターンの出現推移（NICT JLE）

Level 1		Level 2		Level 3		Level 4		Level 5		Level 6		Level 7		Level 8		Level 9	
Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq
64	. #F# .	617	. #F# .	4124	. #F# .	8382	. #F# .	3750	. #F# .	1868	. #F# .	1157	. #F# .	603	. #F# .	382	. #F# .
21	#F# . #F#	305	#F# . #F#	1585	. Yes .	3572	. Yes .	1978	. Yes .	1031	. Yeah .	559	. Yeah .	440	. Yeah .	254	I do n't
21	. Yes .	192	. Yes .	1538	#F# . #F#	2669	. Yeah .	1372	. Yeah .	674	. Yes .	386	. Yes .	264	you know ,	251	. Yeah .
19	. #JP# .	180	. #F# #F#	1385	. #F# #F#	2603	#F# . #F#	1013	#F# . #F#	496	. And #F#	363	I do n't	261	. And #F#	230	. #F# I
13	. #F# #JP#	90	. Yeah .	930	#F# #F# #F#	2552	. #F# #F#	998	#F# #SC# #F#	495	I do n't	310	. And #F#	259	I do n't	198	. Yes .
11	Yes . #F#	89	. #JP# .	823	#F# #SC# #F#	2334	#F# #SC# #F#	963	. And #F#	451	. O K.	278	you know ,	242	. Yes .	167	. O K.
10	. #F# #F#	84	Yes . #F#	823	. Yeah .	1931	#F# #R# #F#	833	. #F# #F#	380	#F# . #F#	252	. O K.	216	. O K.	163	. And #F#
9	#JP# . #F#	80	#F# #F# #F#	711	Yes . #F#	1871	#F# #F# #F#	740	Yes . #F#	351	you know ,	246	. #F# I	188	. #F# I	138	. And I
7	#F# . #JP#	64	. #F# #R#	651	#F# #F# #R#	1623	Yes . #F#	656	#F# #R# #F#	334	#F# #SC# #F#	230	. So I	186	, you know	134	. #F# #F#
5	#F# #F# .	61	#F# #F# .	640	#F# #R# #F#	1555	. And #F#	651	. O K.	324	. #F# #F#	232	#F# . #F#	183	. So I	134	. So I
5	#F# . Yes	55	#F# . #R#	566	. #F# #R#	1376	. #F# I	609	. #F# I	295	. #F# I	212	. #F# #F#	166	. And I	104	#F# . #F#
5	#JP# . #JP#	51	. #F# #JP#	526	. #F# I	1355	#F# #F# #R#	598	. Thank you	285	. Thank you	200	. And I	157	. #F# #F#	98	you know ,
4	#F# #JP# .	49	#F# . Yes	524	#F# . #R#	1131	. #F# #R#	567	I do n't	280	. So I	179	. Thank you	150	#F# . #F#	97	do n't know
4	#F# . .	49	#JP# . #F#	484	#F# #F# .	1045	. Thank you	527	#F# #F# #F#	267	Yeah . #F#	161	, you know	135	. Thank you	88	. And then
4	. #JP# #JP#	49	Yeah . #F#	448	. Thank you	1043	I do n't	513	. So ,	246	Yes . #F#	159	#F# . I	110	Yeah . #F#	83	. It 's
4	Yes . Yes	44	#F# #F# #R#	394	#F# #R# #R#	1007	Yeah . #F#	476	. So I	245	. So ,	158	. It 's	109	. But #F#	81	. Thank you
3	#F# #JP# #F#	44	. . #F#	393	#F# #R# I	1002	#F# #R# I	469	. So #F#	241	. But #F#	149	I have to	109	O K. #F#	78	. But #F#
3	#F# . #R#	43	#F# . #JP#	375	Yeah . #F#	988	. O K.	424	Yeah . #F#	228	O K. #F#	143	. But #F#	106	#F# #SC# #F#	78	Yes . #F#
3	#JP# #F# .	43	. No .	346	Thank you .	944	I want to	420	#F# #R# I	206	. It 's	143	do n't know	105	#F# it 's	74	, I think
3	. #F# #R#	40	. O K.	341	#F# #F# #SC#	921	#F# #F# .	411	. #F# #R#	200	#F# . I	141	Yes . #F#	102	. I 'm	70	#F# it 's
3	. #JP# #F#	40	Thank you .	328	. And #F#	910	#F# . #R#	397	O K. #F#	196	. And I	140	Yeah . #F#	96	. So #F#	70	Yeah . #F#
3	. Thank you	39	#F# yes .	326	. #F# #SC#	879	. So #F#	388	#F# #F# #R#	191	. Yeah ,	138	O K. #F#	96	do n't know	69	. Mh-hmm .
3	. Yeah .	35	#F# #SC# #F#	324	#SC# #F# #R#	844	#SC# #F# #F#	381	Thank you .	189	. I 'm	120	. I 'm	95	. Mh-hmm .	68	, you know
3	Thank you .	35	#F# . I	324	I do n't	816	Thank you .	371	I want to	185	I have to	116	Thank you .	94	. It 's	68	do n't have
3	Yes . #JP#	35	. #R# I	321	#SC# #F# #F#	805	#F# #R# #R#	371	very much .	184	, #F# I	112	#F# #SC# #F#	89	Thank you .	67	a lot of
3	old . #F#	35	. Thank you	321	I want to	781	#SC# #F# #R#	360	. But #F#	182	#F# #F# #F#	111	. So ,	85	I have to	65	. I 'm
2	#F# #F# #JP#	34	#F# #R# #F#	316	. O K.	780	#F# #F# #SC#	342	#F# . I	182	Thank you .	109	. And ,	82	. And then	65	O K. #F#
2	#F# #JP# #JP#	33	. #F# yes	313	. #SC# #F#	779	#F# #F# I	338	. #F# #SC#	182	very much .	109	. So #F#	81	, #F# I	65	Thank you .
2	#F# #R# #F#	32	. #F# I	301	#F# #F# I	734	#F# . I	335	#F# #F# #SC#	181	#F# #R# #F#	107	. Yes ,	80	Yes . #F#	62	. So #F#
2	#F# . Saturday	30	#F# #R# #R#	291	#F# . #SC#	730	O K. #F#	335	, #F# I	173	#F# I 'm	106	#F# #F# .	76	#F# . I	61	O K .
2	#F# . This	29	I do n't	290	#R# #SC# #F#	713	#R# #F# I	333	. #SC# #F#	173	do n't know	104	, I think	74	. And the	59	#F# I do
2	#F# . Where	28	My name is	282	#F# . I	709	. #F# #SC#	323	#SC# #F# #F#	171	. Yes ,	101	#F# it 's	73	, it 's	59	. But I
2	#F# . XXX03	28	name is XXX0	274	. #R# I	655	. #R# #F#	318	Yes . Yes	168	. So #F#	99	very much .	72	a lot of	57	#F# I think
2	#JP# #JP# #F#	27	is XXX02 .	269	#F# . Yes	652	. #SC# #F#	312	#SC# #F# #SC#	164	. #SC# #F#	97	. But I	71	very much .	55	#F# #F# I
2	#JP# #JP# .	26	#F# #F# #JP#	257	Yes . Yes	640	#R# #F# #F#	311	#F# #F# .	162	, you know	96	. And #SC#	70	. I think	55	#F# I 'm
2	#JP# No .	26	#F# . #SC#	256	#R# #F# I	633	. So ,	310	I have to	155	you very much	95	#F# . Yes	69	. And #SC#	55	. Mm-hm .
2	#JP# Yes .	25	. #F# #SC#	255	#SC# #F# #SC	622	#SC# #F# #SC#	308	#F# it 's	153	#F# #F# .	95	I did n't	69	I did n't	50	. And the
2	. #F# XXX05	25	. My name	249	O K. #F#	570	Yes . Yes	303	#SC# #F# #R#	150	, and #F#	94	#F# #F# #F#	67	. So ,	49	. And we
2	. #JP# Yes	24	#F# #F# I	245	#F# #R# #SC#	567	#F# . Yes	302	. It 's	148	#F# #R# I	94	. And #R#	67	O K .	48	, it 's
2	. . #F#	24	#F# . .	241	#R# #F# #F#	566	. #F# #R#	297	#F# #F# I	148	#F# . Yes	89	#F# I 'm	66	#F# I 'm	48	, too .
2	. Best .	24	I go to	240	#R# #R# #F#	565	#F# . #SC#	296	, #F# #R#	144	. And #SC#	87	do n't have	65	you very much	48	. So .
2	. Consulting .	24	Yes . Yes	240	. #R# #F#	564	very much .	292	#F# . #R#	141	. And ,	85	. So .	64	#F# #F# I	47	I have to
2	. Do you	23	#R# #SC# #F#	225	#F# I like	557	#R# #SC# #F#	283	. #R# #F#	140	Yeah . Yeah	85	you very much	61	#F# #F# #F#	46	, I do
2	. Friend .	23	. #JP# #JP#	221	. #JP# .	556	. So I	277	you very much	139	#F# it 's	84	. And then	61	. Yes ,	46	. I do
2	. No ,	23	I live in	213	#R# #F# #R#	554	. And #R#	270	. Uh-huh .	138	#SC# #F# I	82	. #F# #SC#	60	#F# #R# I	45	. #SC# #F#
2	. No .	23	O K. #F#	208	is XXX02 .	540	#SC# #F# I	266	#F# I 'm	136	a lot of	82	. #SC# #F#	60	I mean ,	45	. O K
2	. Today .	23	yes . #F#	202	. And #R#	531	#F# yes .	263	a lot of	137	. And then	81	, I do	59	. #SC# #F#	45	. Uh-huh .
2	No . #F#	22	#F# #JP# .	201	#JP# . #F#	523	, #F# I	262	to go to	134	, I think	81	. I do	59	. But I	44	#F# #SC# #F#
2	O K. #F#	22	#F# . My	197	very much .	522	, #F# #F#	259	#F# . Yes	133	#F# . #R#	80	, #F# I	59	. So .	44	. I think
2	Too . #F#	22	. #SC# #F#	184	name is XXX0	515	#F# #R# #SC#	258	#F# . So	130	do n't have	80	a lot of	59	. That 's	44	I think it
2	Yes . My	22	. And #F#	182	#F# yes .	512	. #F# yes	255	. And ,	127	Thank you very	80	to go to	59	it 's a	44	it 's a

（表 2-2 つづき）

2	Yes . No	21	#F# #JP# #JP	181	#SC# #F# I	503	#R# #F# #R#	252	. And #R#	126	#F# #F# I	79	, and I	58	I went to	43	#F# I have
2	and #SC# #F#	21	#JP# . #JP#	175	you . #F#	499	#F# I 'm	252	. And I	125	#SC# #F# #SC#	78	Thank you ver	58	Thank you very	43	. #F# it
2	boy and girl	21	#R# #R# #R#	170	. #R# #R#	497	#R# #R# #F#	251	#R# #F# I	125	day last week	77	#F# . O	57	. #F# #SC#	42	. And #SC#
2	is XXX02 .	20	#JP# #JP# #J	169	. No .	492	#F# I like	250	, #F# #F#	124	#F# . O	77	, yeah ,	57	. I do	42	O K. O
2	my name is	20	. I 'm	167	#R# #R# I	485	. But #F#	245	I went to	123	. But I	76	, so I	57	I could n't	41	I ca n't
2	name is XXX02	20	thank you .	167	I went to	471	#F# . And	245	Thank you ver	123	it 's very	76	I could n't	57	is XXX02 .	40	. And it
2	system . #F#	19	#F# #R# #SC#	165	#F# I 'm	457	to go to	242	, #F# #SC#	122	there is a	75	. That 's	56	. So it	39	it 's really
2	train . #F#	19	#SC# #F# #R#	164	#F# . And	451	#F# I have	242	. I 'm	122	to go to	75	it 's a	56	. Yeah ,	39	to go to
2	years old .	19	. #JP# #F#	161	, #F# #R#	451	I went to	241	. And #SC#	121	, #F# #SC#	74	#F# yeah .	56	to go to	38	, #F# I
1	#F# #F# #F#	19	. #R# #F#	159	you very muc	446	#R# #F# #SC#	241	. So #R#	121	. #F# #R#	74	, it 's	55	. And also	38	, I guess
1	#F# #F# #R#	19	. #R# #R#	157	#F# #JP# #F#	446	is XXX02 .	241	. Yes ,	121	I went to	74	I went to	54	#F# yeah ,	37	#F# . I
1	#F# #F# Going	19	I 'm fine	157	#R# #F# #SC#	443	. And #SC#	240	#SC# #F# I	119	I ca n't	73	. Uh-huh .	54	. O K	37	. So ,
1	#F# #F# His	19	movie . #F#	155	I live in	416	#F# it 's	238	#R# #SC# #F#	118	#SC# #F# #R#	72	. And the	53	#F# I think	36	I could n't
1	#F# #F# XXX05	18	#F# #JP# #R#	154	#F# #F# #JP#	409	#F# I want	231	#F# #R# #R#	117	. That 's	72	. Yeah ,	52	, and #F#	36	I did n't
1	#F# #F# my	18	#JP# #F# .	154	#R# #R# #R#	405	. Yes ,	228	#F# yes .	116	, #R# I	72	I ca n't	52	day last week	36	I have a
1	#F# #F# one	18	#SC# #F# #F#	153	. #F# yes	396	you very much	224	#R# #F# #F#	116	. Uh-huh .	72	I think .	51	do n't have	36	I used to
1	#F# #JP# #R#	18	I want to	153	. . #F#	392	Yeah . Yeah	223	there is a	114	, #F# #R#	71	. Mh-hmm .	50	#SC# #F# I	36	think it 's
1	#F# #JP# Americ	18	Yes . I	147	My name is	390	#R# #R# I	216	, and #F#	114	, it 's	71	is XXX02 .	50	. And #R#	35	. That 's
1	#F# #JP# Good	17	#F# #JP# #F#	145	Thank you ve	384	you . #F#	216	. #F# yes	114	. And #R#	70	I have a	50	. Mm-hm .	35	day last week
1	#F# #JP# I	17	#F# #R# I	142	yes . #F#	383	. It 's	216	is XXX02 .	113	#R# #F# I	70	day last week	50	. Uh-huh .	35	when I was
1	#F# #JP# O	17	#F# . No	141	. #F# #JP#	382	, #F# #SC#	215	#F# . And	113	. So .	68	#F# #R# I	50	I think it	34	. #F# #SC#
1	#F# #JP# School	17	#F# I like	141	. I like	381	name is XXX02	211	#F# #R# #SC#	112	#F# . Yeah	67	#F# I do	49	. #F# it	34	. #F# and

（注：冠詞の含まれた trigram を網掛けにしてある）

に脱落エラーが顕著であるため n-gram にも冠詞の連鎖が全般に少ない。

5. 結果(2)：品詞 n-gram

5.1 NICT JLE の全体傾向

表3および図5は NICT JLE Corpus における品詞連鎖の推移を上位 100 の trigram パターンをアルファベット順にソートして，同一品詞で始める系列をまとめたものである。

表3 レベル別の trigram に現れる品詞パターン（NICT JLE）

NICT JLE	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
#F#	34	27	24	25	22	15	9	8	6
AT	0	2	2	5	8	9	11	9	10
ADJ	2	3	2	3	3	5	5	6	6
NOUN	7	6	10	6	4	5	4	4	6
PRON	0	5	7	6	8	10	12	12	14
VERB	2	6	6	5	8	8	11	9	10

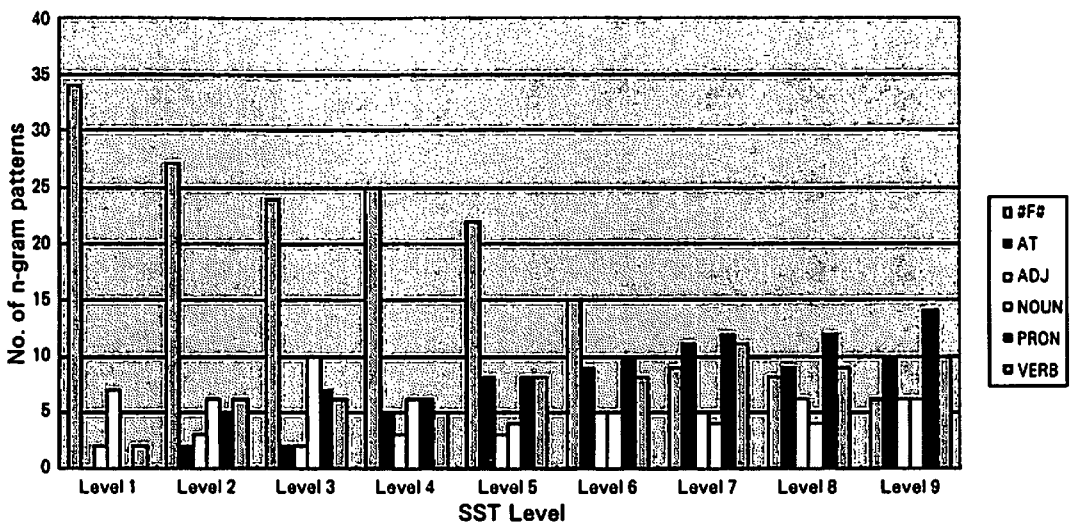

図5 レベル別の trigram に現れる品詞パターン（NICT JLE）

レベル1から9にわたる変化でいくつか顕著なものをあげると，まず最も特徴的なのが，filler（#F#）で始まる連鎖の急激な減少である。これは前述の言いよどみ（dysfluency）現象が解消されていくことを示している。次にコンスタントに増加している項目として，冠詞，形容詞，代名詞，動詞を含む連鎖が挙げら

れる。学習者言語では一般的に冠詞の脱落が顕著だが，上級になるほど冠詞の補充が的確になされてくる。もっともこれは冠詞の補充が出来ている，ということであって，正しい使用法かどうかはまた別問題である。形容詞は修飾関係のパターンの複雑化と関係があり，形容詞の増加は文（特に名詞句）の構造の複雑化を示しているといえる。代名詞で始まるパターンの複雑化も代名詞に続く助動詞などの連鎖の多様化と関係している。

5.2 JEFLL の全体傾向

表 4 および図 6 は JEFLL Corpus における品詞連鎖の推移を上位 100 の trigram パターンをアルファベット順にソートして，同一品詞で始める系列をまとめたものである。

表 4 レベル別の trigram に現れる品詞パターン（JEFLL）

EFLL	J1	J2	J3	S1	S2	S3
#JP#	14	6	3	1	1	0
AT	5	7	7	9	9	12
ADJ	5	5	4	5	5	5
NOUN	20	18	15	23	23	21
PRON	8	12	12	11	9	9
VERB	13	12	11	12	9	12

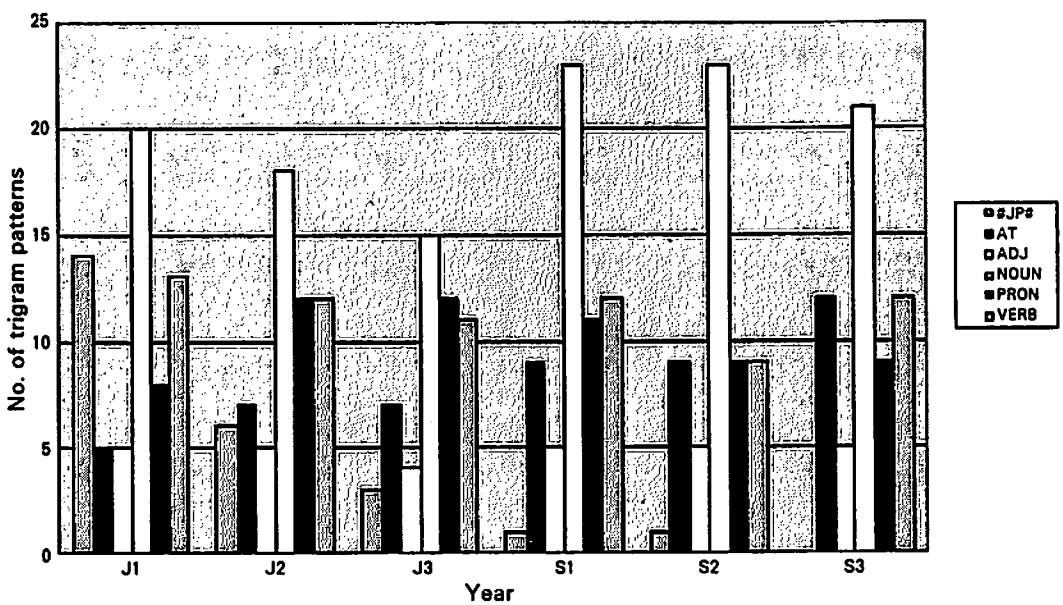

図 6 レベル別の trigram に現れる品詞パターン（JEFLL）

#JP# は日本語が含まれた部分を一括置換した #JP# でスタートした連鎖が示されており，それが中学～高校と行くにしたがってパターンが減少していることがわかる。全般的な傾向として，学年を追って連鎖パターンの数が増加していくものとしては冠詞，名詞で，それぞれ始まる連鎖が顕著な増加傾向を示している。さらに代名詞のパターンは名詞のパターンが増加するのと反比例して減少していく。この部分は学年が上がるに連れて，会話調の文体から，より書き言葉らしい代名詞を抑えて名詞の連鎖を活用して表現するような傾向が出てきているといえよう。NICT JLE に比べると，全体的に学年の初めからある程度安定した連鎖が出現しており，やはり話し言葉のモードとは異なり，自分のパフォーマンスを客観的にモニターできるような情報処理的な余裕があるから，このような安定した使用が可能になるのではないかと推察される。

6．考察

話し言葉と書き言葉の 2 つの異なるモードの学習者英語をレベル別に比較してみるという新しい試みはいろいろな可能性を秘めている。母語話者コーパスにおける比較においても，話し言葉は書き言葉とかなり異なる独自の傾向を示す。例えば，British National Corpus における最も頻度の高い単語 10 個程度を比較しても，書き言葉ではこの 10 位の中に前置詞が 3 つ（of, to, in）入っているのに対して，話し言葉では前置詞は 1 つだけ（to のみ）になってしまう。これは書き言葉でいかに前置詞句を多用するかを表している。さらに代名詞にも特筆すべき変化がある。書き言葉では it と he が 10 位以内に入っているが，話し言葉では he は脱落し，代わりに I, you が加わる。これも，話し言葉では対話する話者同士（I, you）と話題になっているもの（it）の関係で会話が進み，書き言葉では詳細に述べる必要のある名詞関係の表現が会話では両者の暗黙の合意の下，代名詞で省略されてしまうといった「会話の文法」が見え隠れする。

こういった特徴をとらえておくと，学習者コーパスの比較においても，一般的な「話し言葉 vs. 書き言葉」という次元でとらえるべき現象と，学習者の文法獲得上の変化と捉えるべき現象を注意深く見極めることが可能になる。

JEFLL Corpus と NICT JLE Corpus の品詞タグ連鎖の分布を比較してみると，この産出モードの違いによる言語使用の差異が観察できる。話し言葉と書き言葉では即時性が異なるので，脳内で言語処理をする時間的余裕がある書き言葉データの方が，冠詞，形容詞，動詞といった項目のパターンのヴァリエーションが比較的初期の段階から発現している。これに対して，話し言葉ではそういった表現のヴァリエーションを使う言語処理上の時間的余裕がないために，英語力の低い学習者では特にレパートリーがせまい定型句を多用する傾向があり，それが上級に行くにしたがって，脳内処理が自動化（automatization）するのと並行してレパートリーも増加していくのが顕著に見受けられる。つま

り，産出モードによる言語の自動化の度合いとの関連で，産出できる言語表現が規制されるという現象が話し言葉では顕著に見られるのである。

もう1つ興味深いのは，冠詞の補充に見られるような現象である。冠詞に関しては，両者のコーパスともレベルが上がると冠詞を含むパターンが増加する傾向を示した。これは話し言葉，書き言葉というモードの影響よりも，言語内の冠詞の習得度レベルが直接パフォーマンスに反映していることを示唆している。そういった意味で，モードの影響を比較的受けにくい項目であるといえる。こういったことがある程度わかってくると，話し言葉，書き言葉といったモードの違いに注目して観察すべき言語特徴と，それらに共通に観察すべき項目を選別することが可能になってくる。今後の研究が待たれる領域である。

その他の項目でも，NICT JLE Corpus に見られるレベルが上がるに連れて上昇する代名詞の使用などは，より母語話者の会話モードの特徴に近づく様子を表しており興味深い。これらの言語特徴の観察を十分に行いつつ，将来はレベルの変化を説明できる言語特徴の選定や重み付け，それらが話し言葉，書き言葉といった産出モードとどういう相関にあるかなどを，習得モデルとの関連で説明していけるような説得力のある言語モデル構築を目指していく必要がある。

7．まとめ

本稿では2つの英語学習者コーパス，NICT JLE Corpus と JEFLL Corpus の全体的な比較を単語・品詞 n-gram 統計をもとに行った。話し言葉と書き言葉という産出モードの違いを比較するには，厳密には同一タスクで同一被験者の行う言語使用データを採取する必要がある。しかし，大量データによる今回のような研究は，その予備的な観察として多くの示唆を得ることができ，また重要な観察すべき語彙・文法特徴を選定するためにも意義があると言えよう。

今後は，各レベルの品詞付きデータから機械学習による確率的言語モデルや n-gram モデルの構築が可能か，その後レベル別モデルの推移を説明すべき因子を統計的に抽出可能か，といった点が言語処理的には興味あるテーマと言えよう。また，このような記述的な習得モデルと合致した習得理論として，たとえば Nick Ellis らの提唱する Associative-Cognitive CREED などの認知言語学や emergentism の理論との融合，また Rens Bod らの Data Oriented Parsing の考え方を LFG の枠組みに組み込んだプロセスモデルとの融合など，学習者データをコンピューター処理する際に親和性の高い「説明モデル」を構築していくことも重要な課題となろう。

それらの基礎となるコーパス・データの整備充実はいうまでもない。JEFLL Corpus が研究用に公開された今，ますますいろいろな記述研究がなされ，データの補充もされていき，日本人英語学習者の習得の全体像が浮き彫りにされる

日が来ることを願ってやまない。

注

＊本稿は，第 29 回大会シンポジウム「英語学習者コーパスの新展開」において，口頭発表したものに加筆修正したものである。本稿をまとめるに当たり，編集事務局および匿名論文査読者から非常に有益なご指摘とご意見を賜り，重ねて感謝を申し上げる次第である。

1．N-gram 統計の処理は元明海大学大学院修士課程の上村崇氏が担当した。ここに謝意を表したい。

2．小学館コーパス・ネットワークの無料コンテンツとして公開（http://scn02.corpora.jp/~jefll03/jefll_top.html）

3．JEFLL 研究用サイト（http://jefll.corpuscobo.net/）を参照

参考文献

Tono, Y. (2000). "A computer learner corpus-based analysis of the acquisition order of English grammatical morphemes." Burnard, L. and T. McEnery, (eds.) *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the third international conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, pp. 123-132.

朝尾幸次郎 (2000)『第二言語習得研究のための英語学習者コーパスの構築とその利用』平成 9 年度～平成 11 年度科学研究費補助金（基盤研究（B）(1)）研究課題番号 09558018 研究成果報告書，pp. i-iiv, 1-120.

和泉絵美・井佐原均・内元清貴（編著）(2004)『日本人 1200 人の英語スピーキング・コーパス』アルク.

太田　洋 (2007)「Lexical collocation の発達」投野（編著）(2007)，pp. 20-32.

投野由紀夫（編著）(2007)『日本人中高生一万人の英語コーパス：JEFLL Corpus』小学館.

（投野由紀夫　東京外国語大学　E-mail: y.tono@tufs.ac.jp）