

多変量解析による世界の英語学習者の談話分析

小林雄一郎

大阪大学大学院言語文化研究科 日本学術振興会

世界の様々な言語を背景とする英語学習者の中で、日本人学習者の英作文はどのような談話的特徴を持っているのだろうか。本研究の目的は、談話分析の観点からメタ談話標識を説明変数とし、対応分析、階層型クラスター分析、頻度比分析などを用いて、日本を含む17ヶ国の言語を背景とする書き手の英作文データを統計的に比較していくことである。

A Multivariate Discourse Analysis of Non-Native Englishes

Yuichiro Kobayashi

Graduate School of Language and Culture / JSPS Research Fellow
University of Osaka

What are discursual features characteristic of English essays written by Japanese EFL learners? The purpose of the present study is to shed a light on the interrelationships among various non-native Englishes. The method is based on correspondence analysis, hierarchical cluster analysis, and frequency ratio analysis.

1. はじめに

母語話者 (Native Speakers, NS) は、どうして非母語話者 (Non-native Speakers, NNS) の英語が母語話者によるものではないと分かるのであろうか。日本の英語教育では語彙や文法は重要視されてきたため、大学生レベルの作文であればスペリングや文法に関する稚拙な誤りは比較的少ない。では、母語話者の作文と非母語話者の作文との差は何か。後者に見られる「不自然さ」とは何か。それらを部分的にでも明らかにしていくことは、英語教育において非常に重要な課題の1つである。

[1] は、NS (アメリカ人大学生) と NNS (日本人大学生) の英作文データに対して、メタ談話標識を説明変数とする判別分析を行った。その結果、NS と NNS の作文は 90% の精度で判別され、self-mentions, hedges, boosters, frame markers に関する談話表現が判別に大きく寄与していることが分かった。

では、世界の様々な言語を背景とする英語学習者の中で、日本人学習者の英作文はどのような談話的特徴を持っているのだろうか。母語話者と異なる特徴を持つてはいるものの、他の言語を背景とする学習者よりは母語話者に近いの

であらうか。それとも、日本人学習者の談話は、世界の英語学習者の中でもかなり特殊な位置を占めているのであろうか。

2. データと方法論

分析データとして、ICLE (International Corpus of Learner English) の version 2 に含まれている 16 のサブコーパス (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italy, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana, Turkish) に加えて、[1] で用いた母語話者データ LOCNESS (Louvain Corpus of Native English Essays) と ICLE 日本人データのプレリリース版を用いる。以下の表 1 は、サブコーパスの総語数をまとめたものである (分析データの総語数は約 60 万語)。表中の American, Japanese_1, Japanese_2 はそれぞれ母語話者データ (LOCNESS), 日本人データのプレリリース版 (ICLE-JP), ICLE version 2 に収録された日本人データを表している。

分析手法としては、対応分析、クラスター分析、頻度比分析を用いる。また、分析にあたっては、Hyland のメタ談話標識 (metadiscourse markers) のリスト (Hyland list) [2] を用いる。

表 1: 分析データの概要

	American	Bulgarian	Chinese	Czech	Dutch	Finnish	French	German	Italy
Tokens	18804	32540	62985	31270	32041	41763	31702	37125	31555
	Japanese_1	Japanese_2	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Tokens	30223	36013	34420	31739	34425	26422	31829	28172	29326

表 2: メタ談話標識のカテゴリー [3]

Category	Function	Examples
<i>Interactive resources</i>	<i>Help to guide reader through the text</i>	
Transitions (TRA)	Express semantic relation between main clauses	in addition / but / thus / and
Frame markers (FRM)	Refer to discourse acts, sequences, or text stages	finally / to conclude / my purpose here is to
Endophoric markers (END)	Refer to information in other parts of the text	notes above / see Fig / in section 2
Evidentials (EVI)	Refer to source of information from other texts	according to X / (Y, 1990) / Z states
Code glosses (COD)	Help readers grasp functions of ideational material	namely / e.g. / such as / in other words
<i>Interactional resources</i>	<i>Involve the reader in the argument</i>	
Hedges (HED)	Without writer's full commitment to proposition	might / perhaps / possible / about
Boosters (BOO)	Emphasize force or writer's certainty in proposition	in fact / definitely / it is clear that
Attitude markers (ATM)	Express writer's attitude to proposition	unfortunately / I agree / surprisingly
Engagement markers (ENG)	Explicitly refer to or build relationship with reader	consider / note that / you can see that
Self-mentions (SEM)	Explicit reference to author(s)	I / we / my / our

このリストは、様々な談話分析に関する先行研究をベースとして、10種類のカテゴリー（表2）に分類される約400種類の談話表現を網羅的に収録したものである。また、このリストは、コーパスに基づく統計的研究を想定して作成されたものであり、これまでにアカデミック・ライティングを始め、教科書、学位論文、ビジネスレターなど、様々な言語データの分析で成果を上げている。

3. 結果と考察

3.1 頻度集計

一部の記号（括弧、疑問符、ハイフンなど）を除いて、分析データにおける全てのメタ談話標識の頻度を調査したところ、437種類のメタ談話標識が延べ602354回使用されていた。表3は、437種類のメタ談話標識の粗頻度 (raw frequency) を調べて、10種類のカテゴリー（表2）別に集計したものである。

表 3: メタ談話標識のカテゴリー別の頻度集計表

	American	Bulgarian	Chinese	Czech	Dutch	Finnish	French	German	Italy
TRA	6424	11217	23334	10410	11937	13998	10805	11715	11180
FRM	912	1342	3372	1399	1861	1642	1916	1765	1638
END	387	530	992	471	654	758	817	565	600
EVI	82	84	1142	80	94	158	96	68	114
COD	1175	2033	3601	2028	2038	2645	2085	2083	2027
HED	2655	3227	8947	3617	4111	5737	3711	4131	4146
BOO	1674	2681	4574	2782	3032	3719	3114	3201	2961
ATM	504	1089	1752	1076	1176	1635	1083	1132	1246
ENG	2973	4501	9926	4159	4453	5506	4092	5635	4351
SEM	2018	5836	5345	5248	2685	5965	3983	6830	3292
	Japanese 1	Japanese 2	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
TRA	8435	9873	10696	11780	11621	9889	10195	10371	10150
FRM	2301	2482	1384	1394	1669	1392	1453	1119	1552
END	409	392	634	490	580	453	519	282	433
EVI	44	62	42	87	88	72	38	58	144
COD	1464	2003	2035	2543	2130	1881	1706	1898	2147
HED	2301	2901	4448	4350	3882	2606	3958	2273	2442
BOO	1951	3320	2902	2677	3258	2422	2930	2344	2637
ATM	897	1054	1074	1158	1192	917	1034	723	1021
ENG	3552	5154	5318	4127	4462	3277	4099	5833	5228
SEM	8869	8772	5887	3133	5543	3513	5897	3271	3572

3.2 対応分析

まず、分析データにおけるメタ談話標識の全体的な使用傾向を俯瞰するために、対応分析を行う。図1~2は、データ中に生起する全て（437種類）のメタ談話標識を変数として、対応分析を実行した結果である。なお、視認性を重視し、ケースと変数を別々の図に出力している。これらの図は、解析の結果として得られる得点のうち、最も寄与率の高い第1次元（28.28%）と第2次元（22.30%）を2次元散布図に布置したものである（第2次元までの累積寄与率は50.58%）。図中で近接する項目は類似した性質があることを示し、図中の項目間を隔てる距離が大きいほど異質性が高いことを示す。但し、解析結果を

解釈するにあたっては、各次元が直交している（無相関である）ことに注意しなければならない。図1を見ると、母語話者を含む殆ど全てのケースが図中の下側に布置されている一方で、Japanese_1 (ICLE version 2), Japanese_2 (ICLE prerelease version), Chinese がそれ以外のケースから離れた位置に布置されている。また、同じ言語的背景を持つ Japanese_1 と Japanese_2 は、やはり近い位置に布置されている。

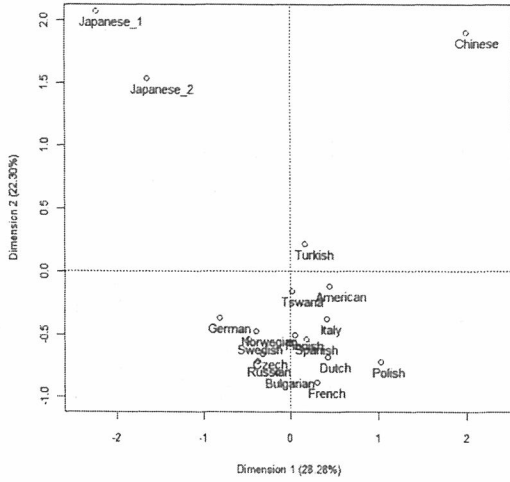


図 1: 437 種類のメタ談話標識を変数とする
対応分析 (ケースの分布)

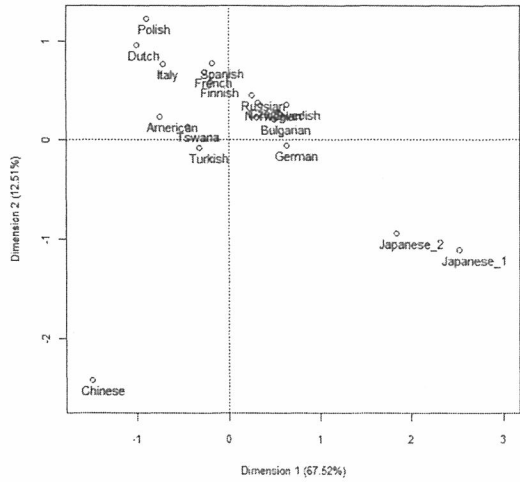


図 3: 10 種類のメタ談話標識カテゴリーを
変数とする対応分析 (ケースの分布)

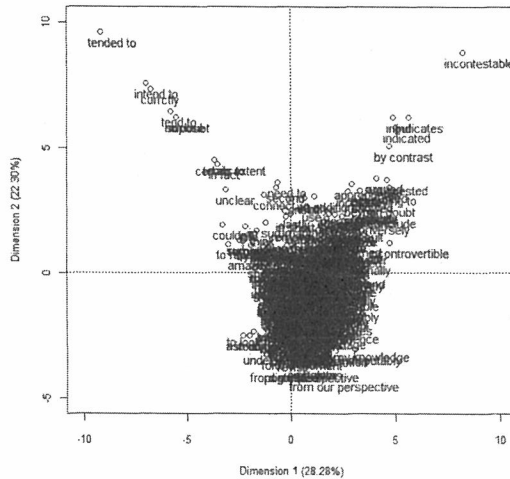


図 2: 437 種類のメタ談話標識を変数とする
対応分析 (変数の分布)

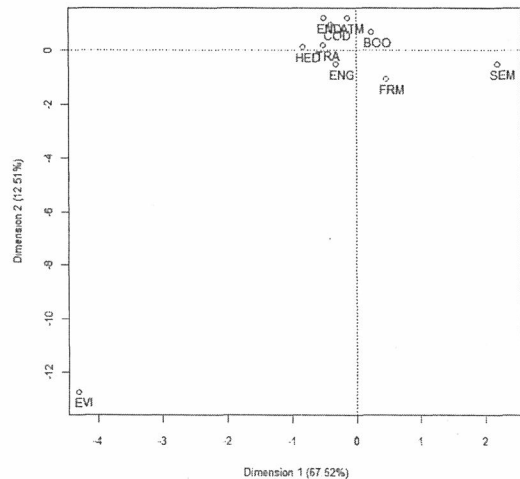


図 4: 10 種類のメタ談話標識カテゴリーを
変数とする対応分析 (変数の分布)

そして、図 3~4 は、10 種類のカテゴリー別の頻度行列 (表 3) に対して、対応分析を実行した結果である。これらの図は、解析の結果として得られる得点のうち、最も寄与率の高い第 1 次元 (67.52%) と第 2 次元 (12.51%) を 2 次元散布図に布置したものである (第 2 次元までの累積寄与率は 80.03%)。図 3 を 180 度回転してみると、先ほどの図 1 とかなり近い形で各ケースが布置されていることが分かる。

3.3 クラスタ分析

図 3 は、階層型クラスタ分析を用いて、図 1 に布置された 18 ケースを分類した結果である。階層型クラスタ分析とは、ケース間の類似度あるいは非類似度 (距離) に基づいて、最も似ているケースから順次に結合してクラスタを作っていく解析法である。分析にあたって、ケース間の距離は標準化ユークリッド距離で測り、クラスタ間距離をワード法で測った。

図中の破線の位置を Cutting・ポイントとし

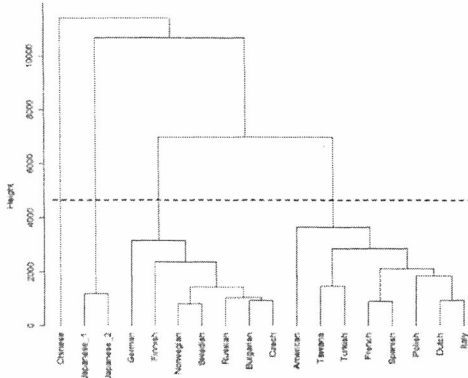


図3: クラスター分析 (ケースの分類)

た場合、4つのクラスターが形成される。左から、第1クラスター (Chinese)、第2クラスター (Japanese_1, Japanese_2)、第3クラスター (German, Finnish, Norwegian, Swedish, Russian, Bulgarian, Czech)、第4クラスター (American, Tswana, Turkish, French, Spanish, Polish, Dutch, Italy) となる。中国人学習者 (Chinese) と日本人学習者 (Japanese_1, Japanese_2) が単独でクラスターを形成していることは、中国人学習者と日本人学習者の英作文が極めて特異な談話的特徴を持っていることを意味している。

3.4 頻度比分析

ここまで、多変量解析を用いて、ケース (サブコーパス) 間の相互関係、変数 (メタ談話標識) 間の相互関係、ケースと変数の相互関係を俯瞰してきた。以下、個々のメタ談話標識を対象とし、ケース間に顕著な差が見られる表現を精査していく。

3.4.1 Transitions (TRA)

学習者が非常に好む TRA として、*moreover*, *nevertheless*, *on the contrary* がある。まず、*moreover* は、全ての学習者グループ¹が母語話者よりも多く用いており、*nevertheless* と *on the contrary* もツワナ人以外の全ての学習者グループが母語話者よりも多く用いている。以下の表 4~6 は、これらの表現をよく使うグループとあまり使わないグループをまとめたものである。表中の“NF”と“Ratio”は、10万語あたりの相対頻度と、母語話者の頻度との比率 (母語話者の値は常に1.00) をそれぞれ表している。

表4: *moreover* の使用頻度 (一部)

Heavy users			Light users				
Rk	Subcorpus	NF	Ratio	Rk	Subcorpus	NF	Ratio
1	German	62.35	18.65	1	American	3.34	1.00
2	French	61.24	18.32	2	Tswana	5.95	1.78
3	Turkish	48.87	14.62	3	Norwegian	6.08	1.82
4	Chinese	46.67	14.26	4	Finnish	9.49	1.94
5	Poland	44.10	13.19	5	Czech	9.88	2.95

¹ 本稿では、母語話者以外のサブコーパスを学習者グループと呼ぶことがある。

表5: *nevertheless* の使用頻度 (一部)

Heavy users			Light users				
Rk	Subcorpus	NF	Ratio	Rk	Subcorpus	NF	Ratio
1	German	28.75	14.33	1	Tswana	0.99	0.49
2	French	26.88	13.40	2	American	2.01	1.00
3	Poland	22.47	11.20	3	Japanese_1	3.35	1.77
4	Italy	19.90	9.92	4	Turkish	5.98	2.98
5	Czech	19.26	9.60	5	Swedish	6.93	3.45

表6: *on the contrary* の使用頻度 (一部)

Heavy users			Light users				
Rk	Subcorpus	NF	Ratio	Rk	Subcorpus	NF	Ratio
1	French	29.08	21.75	1	Tswana	0.50	0.37
2	Italy	24.32	18.19	2	American	1.34	1.00
3	Bulgarian	17.31	12.95	3	Japanese_1	4.15	3.10
4	Russian	13.42	10.04	4	Japanese_2	5.47	4.09
5	Czech	10.37	7.75	5	Turkish	5.48	4.10

以下は、学習者による *moreover*, *nevertheless*, *on the contrary* の使用例である。

- (1) **Moreover**, any kind of advertisement of drugs, including alcohol and tobacco, should be completely banned. (German)
- (2) **Nevertheless**, the best way to keep and perfect a language is to go abroad. (German)
- (3) **On the contrary**, they often used to live on the fringe of society when they were sent in prison. (French)

これ以外にも、多くの学習者グループが *as a consequence* や *though* を過剰使用している。逆に、全ての学習者グループが母語話者よりも低い頻度で用いている表現として、*even though* と *likewise* が挙げられる。また、トルコ人以外の学習者グループが *while* を、オランダ人とドイツ人以外の学習者グループが *yet* を、それぞれ母語話者よりも低い頻度で用いている。

TRA のような接続語の主な働きは、談話の根底にある様々な関係の解釈を容易にして、曖昧さをなくすことにある。だが、[4] などが指摘するように、接続語の使用数が増えても結束性の質が上がる訳ではなく、むしろ接続語の過剰使用は、読み手に不自然な印象を与え、テキストの理解度を下げる恐れがある。

3.4.2 Frame markers (FRM)

多くの学習者グループが最も好む FRM は、*firstly* と *secondly* であり、全ての学習者グループが母語話者よりも多く用いている (表 7~8)。

表 7: *firstly* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Turkish	28.43	42.51	1	American	0.67	1.00
2	Chinese	23.34	36.41	2	Norwegian	2.34	3.50
3	Tswana	18.35	27.44	3	Bulgarian	2.97	4.44
4	Spanish	10.52	15.74	4	German	3.00	4.49
5	Japanese_2	10.45	15.63	5	Czech	4.44	6.65

表 8: *secondly* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Dutch	22.40	33.50	1	American	0.67	1.00
2	Japanese_2	20.40	30.51	2	German	3.43	5.13
3	Chinese	19.07	28.52	3	Russian	3.90	5.83
4	Italy	18.13	27.11	4	Spanish	6.01	8.99
5	Turkish	17.95	26.85	5	Bulgarian	6.43	9.62

以下は、学習者による *firstly* と *secondly* の使用例である。

- (4) **Firstly**, most of the university students study only for passing the exams. (Turkish)
- (5) **Secondly**, art is part of our culture and part of the history of mankind. (Dutch)

これ以外にも、多くの学習者グループは、*finally*, *first*, *first of all*などを母語話者よりも多く用いている。その中でも、日本人学習者は、*last*, *lastly*, *next*, *second*, *then*, *third*といった sequencing markersも好んで使っている。これは学習者が論理的な文章を構成しようという意識から生じたものではある。だが、TRA の場合と同様に、接続語の過剰使用は、母語話者の目には非自然なものに映る。また、[1] が指摘するように、日本人学習者が sequencing markers を過剰使用する原因の 1 つは、それまでに受けてきたパラグラフ・ライティング指導である。何故ならば、殆どの英作文の教科書で接続語の重要性が強調されているものの、使い過ぎてはいけないことはあまり書かれていないからである。

また、*in short*, *all in all*, *aim*なども学習者が高頻度で用いている表現である。逆に、多くの学習者グループが母語話者よりも低い頻度で用いている表現として、*overall*, *focus*, *return to*, *turn to*などが挙げられる。

3.4.3 Endophoric markers (END)

END は、そこに含まれる表現の種類も頻度も少ないため、あまり顕著な差は見られない。しかし、中国人学習者が *above* を好んで使っていることは注目に値する (表 9)。

表 9: *above* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Chinese	50.10	3.26	1	Dutch	13.95	0.91
2	Polish	36.47	2.37	2	America	15.38	1.00
3	Bulgarian	31.66	2.06	3	French	17.18	1.12
4	Italy	29.62	1.93	4	Russian	18.19	1.18
5	Czech	26.17	1.70	5	Swedish	19.30	1.25

以下は、学習者による *above* の使用例である。

- (6) As mentioned **above**, I have describe the pros and cons of importing professionals from Mainland China into Hong Kong. (Chinese)

また、中国人学習者は、*above* の対義語である *below* も、母語話者の 3.64 倍使用している。このことは、彼らがテキストの他の箇所述べていることを適宜参照しながら、議論を進めていくことを好むことを意味している。

3.4.4 Evidentials (EVI)

END と同様に、EVI も、そこに含まれる表現の種類も頻度も少ないため、あまり顕著な差は見られない。だが、中国人学習者が *according to* を多用している (表 10)。前掲の図 4 で、EVI が外れ値となっている原因の 1 つは、この中国人学習者による *according to* の過剰使用である。

表 10: *according to* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Chinese	227.59	6.30	1	Norwegian	15.43	0.43
2	Turkish	71.80	1.99	2	Swedish	16.33	0.45
3	Finnish	54.45	1.51	3	Japanese_1	24.29	0.67
4	Italy	45.54	1.26	4	German	26.17	0.72
5	French	38.77	1.07	5	Japanese_2	26.87	0.74

以下は、学習者による *according to* の使用例である。

- (7) **According to** the Hong Kong Trade Development Council, the requirement of huge capital input for collection and sorting hesitates the businessmen to investigate on recycling industry. (Chinese)

中国人学習者による *according to* の過剰使用は、彼らが他者による意見や文献を援用しつつ、議論を進めていくことを好むことを意味している。END と EVI の多用は、中国人学習者による英作文の大きな特徴である。

companies, and drink juice or alcohol. (Japanese_1)

3.4.5 Code glosses (COD)

全ての学習者グループが母語話者よりも高頻度で用いている COD として, *in other words* と *for instance* がある。これらの表現は, フランス人, ノルウェー人, オランダ人などの学習者が特に好んで用いている (表 11~12)。

以下は, 学習者による *in other words* と *for instance* の使用例である。

- (8) **In other words**, television is a kind of drug which keeps people in lethargy, and it prevents them from developing a sharp critical taste. (French)
- (9) **For instance**, cancer, allergies or hereditary diseases are due to defective genetic information. (French)

表 11: *in other words* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	French	22.91	8.57	1	American	2.67	1.00
2	Norwegian	14.03	5.25	2	Czech	3.46	1.29
3	Bulgarian	13.85	5.18	3	Tswana	4.96	1.85
4	Dutch	13.52	5.06	4	German	6.86	2.57
5	Japanese_2	11.94	4.46	5	Chinese	6.90	2.58

表 12: *for instance* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	French	48.91	6.65	1	American	7.35	1.00
2	Turkish	40.39	5.49	2	Japanese_1	9.48	1.29
3	Norwegian	39.76	5.41	3	Chinese	9.74	1.32
4	Dutch	39.30	5.34	4	Czech	10.37	1.41
5	Finnish	33.90	4.61	5	German	11.16	1.52

また, *for instance* とほぼ同じ意味を持つ表現に *for example* がある。これら 2 つの表現を比較すると, 日本人学習者は, *for instance* を母語話者の 1.29~1.62 倍しか用いていないにもかかわらず, *for example* を母語話者の 3.78~3.92 倍 (この倍率は全ての学習者グループの中で最も高い) も用いていることが分かる。

表 13: *for example* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Japanese_1	136.26	3.92	1	Norwegian	34.61	1.00
2	Japanese_2	131.34	3.78	2	German	34.75	1.00
3	Turkish	100.22	2.88	3	American	34.77	1.00
4	Finnish	97.00	2.79	4	Tswana	48.59	1.40
5	Italy	84.45	2.43	5	Czech	52.83	1.52

以下は, 学習者による *for example* の使用例である。

- (10) **For example**, we can eat foods serviced by airline

逆に, オランダ人, フランス人, ノルウェー人, スペイン人の学習者は, *for example* を母語話者の 1.00~1.83 倍しか用いていないにもかかわらず, *for instance* を母語話者の 4.42~6.55 倍も用いている。

3.4.6 Hedges (HED)

学習者グループは, 母語話者よりも多くの HED を低い頻度で産出している。とりわけ, *argue*, *assume*, *claim* といった動詞の HED を過少使用している (表 14~16)。

表 14: *argue* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	American	35.77	1.00	1	Tswana	0.99	0.03
2	Chinese	32.66	0.89	2	Japanese_1	3.55	0.10
3	Norwegian	25.26	0.69	3	Czech	3.95	0.11
4	Finnish	12.62	0.34	4	Bulgarian	4.45	0.12
5	Turkish	11.47	0.31	5	Spanish	5.01	0.14

表 15: *assume* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Polish	8.48	1.06	1	Chinese	0.41	0.05
2	American	8.02	1.00	2	Czech	0.99	0.12
3	Bulgarian	7.42	0.92	3	Turkish	1.00	0.12
4	Italy	6.19	0.77	4	Tswana	1.49	0.19
5	German	5.58	0.70	5	Dutch	1.69	0.21

表 16: *claim* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	American	58.17	1.00	1	Japanese_1	1.78	0.03
2	Polish	39.86	0.69	2	Japanese_2	2.49	0.04
3	Norwegian	34.14	0.59	3	Spanish	4.01	0.07
4	Turkish	24.93	0.43	4	Tswana	5.45	0.09
5	Bulgarian	24.24	0.42	5	Czech	7.41	0.13

以下は, 母語話者による *argue*, *assume*, *claim* の使用例である。

- (11) Some may **argue** that the women in the magazines are better to look at than those on the court. (American)
- (12) One may **assume** that the NFL and its tax subsidized teams no longer owe any loyalty to the fans. (American)
- (13) An opponent of my argument could **claim** that if one lived in a negative community with parents that did not teach proper morals, then they would still have some time to pray to themselves each day

in public school to make them better, more thoughtful persons. (American)

日本人学習者が HED の意味合いを持つ動詞を過少使用することは [1] などで報告されている。しかし、本研究の分析結果を見る限り、動詞の HED を過少使用するの、学習者の母語とはあまり関係がなく、様々な言語的背景を持つ学習者に共通する現象のように思われる。

逆に、多くの学習者グループが母語話者よりも低い頻度で用いている表現として、*in general* や *in my opinion* などがある。

3.4.7 Boosters (BOO)

学習者グループは、母語話者よりも多くの BOO を低い頻度で産出している。学習者による HED や BOO の過少使用は、目標言語で自らの心的態度を適切に表現することの難しさを反映している。そして、多くの学習者グループが母語話者よりも高い頻度で用いている表現として、*surely* や *of course* などがある (表 17~18)。

表 17: *surely* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Italy	22.99	34.39	1	American	0.67	1.00
2	Japanese_1	16.00	23.92	2	Dutch	2.11	3.16
3	Japanese_2	13.43	20.09	3	Turkish	2.49	3.73
4	Czech	12.33	19.94	4	Tswana	2.98	4.45
5	German	12.44	18.16	5	Spanish	4.01	5.99

表 18: *of course* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Swedish	88.57	5.30	1	Tswana	4.46	0.27
2	Russian	82.28	4.92	2	Chinese	10.14	0.61
3	Finnish	72.48	4.34	3	American	16.71	1.00
4	Japanese_1	71.68	4.29	4	Polish	35.62	2.13
5	German	69.08	4.13	5	Spanish	37.06	2.22

以下は、学習者による *surely* と *of course* の使用例である。

(14) It is **surely** more difficult for a single woman to face alone this situation. (Italy)

(15) **Of course**, it is very important that their parents do not ignore the environment. (Swedish)

また、日本人学習者は、*think* を母語話者の 9.43~12.16 倍も用いており、学習者グループの中でずば抜けて高頻度である (その次に母語話者との頻度比

が大きいトルコ人学習者は 3.98 倍)。

3.4.8 Attitude markers (ATM)

学習者が好む ATM には *fortunately* と *surprisingly* であり、ツワナ人を除く全ての学習者グループが母語話者よりも多く用いている (表 19~20)。

表 19: *fortunately* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	French	10.13	15.16	1	Tswana	0.00	0.00
2	Polish	9.33	13.95	2	American	0.67	1.00
3	Finnish	6.49	9.71	3	Turkish	1.00	1.49
4	Czech	5.93	8.86	4	Chinese	1.43	2.12
5	German	5.58	8.34	5	Norwegian	2.81	4.20

表 20: *surprisingly* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Polish	6.36	9.51	1	Tswana	0.00	0.00
2	Dutch	5.92	8.85	2	American	0.67	1.00
3	Spanish	5.51	6.74	3	Norwegian	1.40	2.10
4	French	4.41	6.59	4	Chinese	1.62	2.43
5	Japanese_2	3.98	5.95	5	Swedish	1.98	2.96

以下は、学習者による *fortunately* と *surprisingly* の使用例である。

(16) **Fortunately** our soldiers could suppress his war machine before he could kill too many innocent civilians. (French)

(17) **Surprisingly**, the story contains no "character", no traditional protagonist, who is always the pride of any "normal" fiction and whom an average reader may expect. (Polish)

このような *-ly* 副詞は、文中に 1 語を挿入するだけで書き手の様々な感情を表すことができるため、多くの学習者が好んで用いている。

3.4.9 Engagement markers (ENG)

多くの学習者グループが最も好む ENG は、*imagine* と *think of* などがあり、全ての学習者グループが母語話者よりも多く用いている (表 21~22)。

表 21: *imagine* の使用頻度 (一部)

Heavy users				Light users			
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	NF	Ratio
1	Bulgarian	62.82	15.66	1	American	4.01	1.00
2	Russian	55.43	13.82	2	Chinese	5.68	1.42
3	Czech	43.94	10.95	3	Japanese_1	7.11	1.77
4	Norwegian	42.10	10.49	4	Tswana	8.43	2.10
5	Spanish	25.04	6.24	5	Italy	9.73	2.42

表 22: *think of* の使用頻度 (一部)

Heavy users				Light users		
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	Ratio
1	Bulgarian	29.63	22.20	1	American	1.34
2	Turkish	23.44	17.53	2	Chinese	4.87
3	German	20.17	15.08	3	Spanish	7.51
4	Norwegian	18.71	13.99	4	Japanese_1	10.07
5	Finnish	17.31	12.95	5	Japanese_2	10.45

以下は、学習者による *imagine* と *think of* の使用例である。

- (18) **Imagine** a world without electricity and communication devices. (Bulgarian)
- (19) Just **think of** the horrible situation of the disabled in Bulgaria. (Bulgarian)

このような命令文は話し言葉に顕著な表現であり [5], 学習者の書き言葉が話し言葉に近い特徴を持っていることの 1 例となっている。

3.4.10 Self-mentions (SEM)

学習者グループ全体を見た場合, SEM にはそれほど大きな差は見られない。しかしながら, 日本人学習者は, *I* や *we* などの 1 人称代名詞を非常に高い頻度を用いている (表 23~24)。また, 前掲の図 3 で, 日本人学習者 (Japanese_1, Japanese_2) が外れ値となっている原因の 1 つは, このような SEM の過剰使用である。

表 23: *I* の使用頻度 (一部)

Heavy users				Light users		
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	Ratio
1	Japanese_1	2813.98	6.22	1	Polish	142.05
2	Japanese_2	2126.86	4.70	2	Spanish	353.62
3	German	1393.14	3.08	3	French	371.43
4	Swedish	968.83	2.14	4	Dutch	373.60
5	Norwegian	906.91	2.00	5	Chinese	389.47

表 24: *we* の使用頻度 (一部)

Heavy users				Light users		
Rk.	Subcorpus	NF	Ratio	Rk.	Subcorpus	Ratio
1	Japanese_1	1484.60	5.12	1	American	290.16
2	Japanese_2	1254.72	4.32	2	Chinese	300.62
3	Swedish	930.24	3.21	3	Dutch	379.10
4	Norwegian	916.74	3.16	4	German	388.72
5	Bulgarian	905.17	3.12	5	Tswana	468.58

以下は、学習者による *I* と *we* の使用例である。

- (20) **I** think **I** have to improve my English. (Japanese_1)
- (21) If **we** can speak English, **we** can communicate with many people all over the world. (Japanese_1)

様々な言語的背景を持つ学習者が *I* や *we* のような SEM を好むことは知られているが [6], その中でも日本人学習者は SEM を極めて高い頻度で用いている。

4. おわりに

本研究では、談話分析の観点からメタ談話標識を説明変数とし、対応分析、階層型クラスター分析、頻度比分析などを用いて、日本を含む 17 ヶ国の言語を背景とする書き手の英作文データを統計的に比較した。その結果、中国人学習者と日本人学習者の英作文が極めて特異な談話的特徴を持っていることが分かった。

本研究で扱った談話表現の多くは、文法というよりは文体に関わるものであり、構造的なルールというよりは確率的なパターンに関わるものである。その限りで、コーパスから得られる客観的データ (頻度や統計量) は極めて有用な情報である。

今後は、母語や教科書の影響などを調査し、学習者による過剰使用/過少使用の原因を探っていくたい。

註

本研究の一部は、科学研究費補助金 (特別研究員奨励費) 「テキストマイニングを用いた学習者作文における談話標識の研究」によって行われたものである。

参考文献

- [1] 小林雄一郎 (2009). 「NS/NNS テキスト分類モデルに基づく日本人英作文の特徴抽出」 『人文科学とコンピュータシンポジウム論文集—デジタル・ヒューマニティーズの可能性』 (pp. 261-268). 東京: 情報処理学会.
- [2] Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. New York: Continuum.
- [3] Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25, 156-177.
- [4] Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In Granger, S., (Ed.), *Learner English on computer* (pp. 80-93). London: Longman.
- [5] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- [6] Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In Granger, S., (Ed.), *Learner English on computer* (pp. 107-118). London: Longman.