

論述構造を支える文型の基礎的研究

—多変量解析によるジャンル判別に有効な文型の抽出—

村田 年

慶應義塾大学国際センター 108-8345 東京都港区三田2-15-45

E-mail: minorim@ic.keio.ac.jp

A Study on the Sentence Patterns for Logic Structure of Texts

— Using Selected Sentence Patterns to identify a Text's Genre by
Multivariate Analysis—

MURATA, Minori

International Center, Keio University, 2-15-45, Mita, Minato-ku, Tokyo, 108-8345

専門分野における学習・研究を目的とする多くの日本語学習者にとって、学習の最終目標は論文作成だと言えよう。そのためには、論文に代表される論述文の論理構造の理解が不可欠であり、その理解には文章中の文型が指標として役立つと考えられる。本論文では、文章の論理構造を明らかにし、論述的形式を持つ文章に特有の文型群を特徴づける研究の一環として、4ジャンル（経済学教科書、物理学論文、文学作品、新聞社説）105編（計7096文）の文章における98文型項目の出現率を調査し、多変量解析の一手法である判別分析法を用いて分析を行った。その結果、文章の所属するジャンルが、7つの文型項目（機能的には「理由・帰結」「定義」「例示」「場所」の4文型群に入る）のみによって、正判別率77%の高率で判別されることが判明した。この結果の分析から、限られた資料ではあるが、異なるジャンルの文章を判別するために文型という指標が有効であることが明らかとなり、同時に論述的形式を持つ文章に共通する文型項目が抽出された。

キーワード：論述文、論理構造、文型、出現率、多変量解析、ジャンルの判別分析

1. はじめに

専門日本語教育では、専門分野における日本語の学習効果をどのように上げていくかが大きな課題であると考えられる。専門分野の日本語教育と言っても、外国人学習者に対する一般の日本語教育と同様、語彙教育と文型教育は基本的な大きな柱であることには変わりない。語彙教育の面から見れば、各専門分野における多くの専門用語の存在から分野の数だけその教育方法の開発が必要だと考えられる。しかし、文型教育の面では果たしてどうなのか。この疑問が筆者の文型研究の出発点である。大学における一般日本語教育のほか、専門分野である経済学、理工学を専門とする留学生への教育ならびに教材開発を通じて、専門分野が異なっても各分野における論述的文章の論理展開にはかなり共通点があるのではないかという考えを強め、それをより客観的な方法で、換言すれば追実験可能な科学的な方法で確認したいと考えた。文型に関する先行

研究の中には、文中における文型の出現数に着目したものもあるが、相対出現頻度に変換していないため、真の意味での文章間の比較は難しいと言える。今後、専門分野における多様な教材開発が推進されることを考えるとき、その根拠の一翼を担う基礎的な研究として、統計的手法を用いた文章分析は大きな意義を持つと考えられる。

文章の分析に統計的手法を用いた統計的文体研究の歴史は、一般的に19世紀半ばに遡るとされている。この分野ではこれまで著者推定や執筆年の決定をはじめとする多くの研究が行われてきている^{注1}。特定の作家の文体を数学的に研究するために、文章中の明確に数量化できる特徴を用いて、その作家あるいはそのジャンルに固有の規則性を見出していくという "Stylometry (計量文体論)"^{注2} は、文章研究にとって重要な方法論の一つであると考えられる。本論文は、この計量的手法を日本語教育分野の文型研究に応用したものである。

2. 研究の目的

専門分野における学習・研究を目的とする多くの日本語学習者にとって、学習の最終目標は専門分野の論文の読解および作成だと言えよう。その目標のためにには、論文に用いられる論述文の論理構造の理解が不可欠であり、その理解には文章中の文型がその指標として役立つと考えられる。この文型研究の目標は、文型を「表層表現における論理構造を明示する型」として考え^{注3}、文型が持つ種々の情報を用いることによって、文章の論理構造を明らかにし、論述的形式を持つ文章に特有の文型群を特徴づけることである。その研究の一つの段階として、本論文では、多変量解析を用いて、ジャンルの異なる文章を判別するために有効な文型群の抽出が可能であることを示すと同時に、論述的形式を持つ文章に特有な文型群を抽出する。

3. 分析に用いた資料

専門分野の論述的形式の文章の資料として、学習者の多い分野の一つである経済学の入門教科書と、論理的構成がより強いと思われる物理学論文を選んだ。比較のための資料としては、経済学教科書と物理学論文より論理的構成が弱いと思われる文学作品と新聞社説を選んだ。

(1) 経済学教科書

『はじめての経済学』岡田泰男・野澤素子・村田年編慶應義塾大学出版会（1995）

この教科書は、16名の経済学者がそれぞれ自分の専門について一つの章を執筆する形で構成された入門教科書で、理論と実践に関する記述の割合がほぼ半々になっている。16編総文数1124文。

(2) 物理学論文

『日本物理学会誌』の1997年第52巻のNo. 1～12までの「最近の研究から」に掲載された論文を各号から2編ずつ選び、合計24編を資料とした。総文数2243文。

(3) 文学作品

近代文学の文豪3人、森鷗外、夏目漱石、芥川龍之介の短編作品から、14編（新潮文庫）を資料とした^{注4}。総文数2528文。

森鷗外：

『余興』『杯』『普請中』『百物語』『二人の友』

夏目漱石：

『初秋の一日』『三山居士』『子規の画』
『日記』『手紙』

芥川龍之介：

『仙人』『蜃氣樓』『トロッコ』『好色』

(4) 新聞社説

日本経済新聞の1996年12月1日～31日までの朝刊、夕刊の社説51編を資料とした。12月を選んだのは、他の月に比べて、その年1年の主な出来事に言及した記事が多く、総括的傾向が強かったためである。総文数1201文。

4. 分析方法

2で述べた目的のために、以下の方法で文型の抽出と分析を行った。

- 異なる4つのジャンル（経済学教科書、物理学論文、文学作品、社説）の105編を資料として、中・上級文型を中心に接続詞等の語彙レベルまでを含めた98の文型項目^{注5}について、それぞれの出現頻度を調べた。そして、一文当たりの頻度に換算し直した相対出現頻度を求めて、各文型項目の出現率とした。文型項目については、形態が同じで意味機能が複数あるものは、細分化して別の項目とした（例：「～ために（目的）／～ために（理由）」「～を通じて（媒介）／～を通じて（範囲）」）。このように、各文型項目の持つ一つまたは複数の意味機能の同定を並行して行い、文型の基本的意味機能として30種の機能を採用した。なお、出現頻度を調べるにあたっては、用字の差異（例：～にともなって／に伴って）、文型項目の活用変化の形（例：～によって／により／によるN）、文末表現の文中での変化（例：～にすぎない／にすぎず／にすぎぬ）の形を同一視して同じ文型項目として扱い、全数調査を行った。
- 上の1で、各文型項目の意味機能の同定に用いた30種の機能に従って、98の文型項目を30の文型群（グループ）に分類した。

3. 各文型群ごとに統計的手法（判別分析）を用いて分析を行い、上記の4つのジャンルの判別に対しで有効な4つの文型群を抽出した。判別分析とは、基準変数がグループへの所属を表す質的データで、説明変数が量的データであるとき、各ケースの観測された特性に基づいて、所属グループの予測モデルを構築する際に有効な多変量解析の手法である^{注6}。この手法を採用した理由は、質的であるジャンルを量的である文型項目の出現率によって判別しようという本論文の目的に適合しているからである。なお、分析に際しては、各ジャンルの資料の数が異なるため、資料の大きさに基づく事前確率を考慮に入れて、判別規則を構成する方法を用いた。

4. 3で抽出した4つの文型群のうち、ジャンルの判

別に特に効いている文型項目を確定するために、ステップワイズ法^{注7}を用いた判別分析をさらに行い、上位7つの文型項目を選択した。

5. 分析結果とその考察

上の1、2の段階を経て、98の文型項目を30の文型群にグループ化^{注8}したものが表1である。「理由・帰結」文型群の文型項目数が8で最も多く、次が文型項目数6の「定義」「根拠」「添加」「義務・当然・必然」の各文型群であり、文型項目数5の「対象・関連」文型群がそれに続く。

次に、上の3の段階で、30の文型群の各群に含まれる項目を説明変数とし、ジャンルを基準変数（判別目的であるグループ）とする判別分析を行った。判別分析においては、ジャンルの判別の可否は、普通、正判

表1. 98文型項目とそのグループ化（文型群化）

	文型群名	文型項目
1	意志	～[よ]う、～[よ]うとする、ようをする
2	可能（古語）	得る／得ない
3	きっかけ	～にあたって／にあたり、～を機に
4	義務・当然・必然	～ばいけない、～ばならない、～べきだ／べきN、ざるを得ない、わけだ、わけにはいかない
5	逆接	～にもかかわらず、～ものの、ながら[も]
6	禁止	～てはいけない、～てはならない
7	限定・非限定	～にすぎない、とは限らない、わけではない
8	根拠	～に従って／従い、～に基づいて／基づき、～によって／により／によるN、～によると、～によれば、～をもとに
9	資格・立場	～として、～にとって
10	条件付け	～うえ[で]、～としても／とすれば／としたら、V限り、ただし
11	推測	はずだ
12	推量	～[よ]う
13	対象・関連	～について／についてのN、～に関して／に関するN、～に対して／に対するN、～をめぐって／めぐり、～を中心にして
14	対比	～[の]に対して、一方[で]、他方[で]、反面
15	定義	～というのは、～とは、AをBと言う、AをBと呼ぶ、すなわち、つまり
16	添加	～うえ[に]、～だけでなく、～に限らず、～のみならず、～ばかりでなく、なお
17	伝聞	～によると、～によれば、そうだ
18	同時性	～と同時に、～とともに、～に従って／従い、～に伴って／伴い
19	場所	～[の]もとで、～において／におけるN
20	範囲	～から～にかけて、～にわたって／にわたり、～を通じて／通じたN
21	比較	に／と比べ[て]、ほど、むしろ
22	比況	～ようだ／ようなN
23	付帯状況	～ず[に]、～まま、～つつ、ながら
24	変化	つつある、ようになる
25	方法	～によって／により／によるN、～を通じて／通じたN
26	目的	～ため[に]、～よう[に]V、～ように
27	様態	そうだ、ようだ
28	理由・帰結	～[た]結果/Nの結果、～ため[に]、～によって／により／によるN、から、したがって、それゆえに、ので、よって
29	例示	～ようによくなN、～をはじめ、AはBと並んで、たとえば
30	列挙	一つは～もう一つは、第（一）は～第（二）は～

別率（判別的中率）によって評価され、正判別率が高いほど、その文型群がジャンルの判別に有効に働くことを意味する。正判別率は、判別分析を行って判別規則を作成したその同じサンプルに対して判別規則を適用した場合に、サンプルが帰属する群がどの程度正しく判別されたかという割合を理論的に示す「見かけの的中率」によって簡便に評価することができる。本研究においては、文章資料の所属するジャンルは4つなので、判別分析の手法に従って最大3つの判別関数が算出され、その関数の統計的有意性はp値によって判定されることになる（判別関数の有意性の検定）。p値は、5%（あるいは1%）を境界として、 $p < .05$ （あるいは $p < .01$ ）の場合に統計的に意味のある関数であると判断される。表2は、30の文型群について判別分析を行った結果について、見かけの的中率の値と判別関数のp値を示したものである。表2において、見かけの的中率が60%以上のものを、ここではジャンルの判別に有効な文型群と考えた。上位4つの文型群は、順番に①「理由・帰結」②「定義」③「例示」

④「場所」である。

ここで、この4つの各文型群において、どの文型項目がどのジャンルに多く用いられているかを見ていいく。この考察のためには、構造係数と判別関数空間における各グループの重心の関係を見る必要がある。構造係数とは、判別関数と各文型項目との間の相関係数であり、各文型項目の判別関数との関係の強さを表すものである（構造係数の利用法については参考文献6参照）。

「理由・帰結」文型群は、正判別率が75%で、4ジャンルを明確に分離している。論述形式に不可欠な「理由・帰結」文型群は、物理学論文で最も多く用いられ、経済学教科書がそれに続く。文型項目としては、「～によって／により／によるN」「～ため（に）」「したがって」が特に物理学論文で多用されている。文学作品に特徴的な項目は「～から」である。

「定義」文型群は、正判別率69%で、経済学教科書を他の3ジャンルから明確に分離している。しかし、

表2. 30文型群の見かけの的中率とp値

順位	群番号	文型群	項目数	見かけの的中率	p値		
					関数1	関数2	関数3
1	1	理由・帰結	8	75%	0.000	0.000	0.083
2	2	定義	6	69%	0.000	0.123	0.962
3	3	例示	4	66%	0.000	0.047	0.563
4	4	場所	2	62%	0.000	0.002	NA
5	5	伝聞	3	55%	0.000	0.643	NA
5	6	方法	2	55%	0.000	0.001	NA
5	6	付帯状況	4	55%	0.000	0.041	0.096
5	8	条件付け	4	55%	0.002	0.729	0.857
5	9	根拠	6	55%	0.014	0.626	0.847
5	10	添加	6	55%	0.064	0.588	0.550
6	11	同時性	4	54%	0.014	0.071	0.281
7	12	比況	1	52%	0.000	NA	NA
7	13	義務・当然・必然	5(6)	52%	0.003	0.980	0.998
8	14	資格・立場	2	51%	0.000	0.052	NA
8	15	対象・関連	5	51%	0.001	0.229	0.942
9	16	対比	4	50%	0.003	0.440	0.869
9	17	きっかけ	2	50%	0.009	0.080	NA
9	18	範囲	3	50%	0.000	0.910	0.759
9	19	列挙	2	50%	0.105	0.390	NA
9	20	意志	3	50%	0.000	0.742	0.887
10	21	逆接	3	49%	0.256	0.981	0.868
10	22	目的	3	49%	0.146	0.943	0.993
10	23	様態	2	49%	0.000	0.763	NA
10	24	禁止	2	49%	0.470	0.964	NA
10	25	推測	1	49%	0.225	NA	NA
10	26	推量	1	49%	0.061	NA	NA
11	27	限定・非限定	3	48%	0.346	0.858	0.970
11	28	可能（古語）	1	48%	0.143	NA	NA
12	29	変化	2	47%	0.941	0.953	NA
13	30	比較	3	46%	0.081	0.463	0.506

* $p < .05$

文学作品と社説の判別にはあまり有効でないと言える。論述形式の文章には用語の定義付けが欠かせないと考えられるが、特に用語の説明が多い経済学教科書では、予想通り、「定義」文型群の出現頻度が高く、文型項目としては、「すなわち」「AをBと呼ぶ」「つまり」「～というは（～ことだ）」が多く用いられている。

「例示」文型群は、正判別率66%で、経済学教科書を他の3ジャンルから明確に分離し、物理学論文を文学作品ならびに社説から分離するのにも有効である。しかし、「定義」文型群と同様、文学作品と社説の判別にはあまり有効でないことがわかる。「例示」文型群は、経済学教科書で最も多く用いられ、物理学論文がそれに続く。文型項目としては、「～ように／ようなN」が圧倒的に多く、次が「たとえば」である。

「場所」文型群は、正判別率62%で、「例示」文型群と同様、経済学教科書を他の3ジャンルから明確に分離し、物理学論文を文学作品ならびに社説から分離するのに有効である。2つの文型項目のうち、「～において／におけるN」が半別に大きく寄与し、経済学教科書と物理学論文で多く用いられていることがわかる。

次に、4つのジャンルの判別に有効な文型項目を抽出するために、4つの文型群に含まれる20の文型項目

について、分析方法4の4で、ステップワイズ法を用いた判別分析をさらにを行い、ジャンルの判別に特に有効な文型項目を求めるに、7つの文型項目が選択された。なお、20の文型項目については、kruskal-Wallis検定（以後KW検定）も行った。KW検定とは、母集団における分布の位置の差を検出する統計的検定方法の一つで、各資料の全資料中の順位（ランク）に基づいて検定統計量が算出される。分布の位置の差を確認するためには、各群の平均ランクを参照すればよい。KW検定における検定統計量と平均ランクを求めたものが表3で、表の中の*の文型項目がステップワイズ法による判別分析によって選択された7文型項目である。その判別分析の結果を示したものが表4、表5である。選択された7文型項目による判別分析の結果から求められる判別関数平面での各資料とジャンルの重心をプロットしたものが図1である。図1を見ると、論述的形式を持つと考えられる経済学教科書と物理学論文の2ジャンルが、文学作品ならびに社説の2ジャンルからきれいに分離され、文学作品と社説が近い位置ではあるが分離されていることが読み取れる。表6のクロス集計の結果からも分かるように、正判別率も77%（4+11+45+21/105）という高い値となって、選択された7文型項目によって、4ジャンルの判別が可能であることが検証された。また、誤判別は、文学

表3. 4文型群（20文型項目）のKW検定結果

	文型項目	検定統計量	p値	平均ランク			
				経済学	文学	社説	物理学論文
1	したがって	62.340	0.000	77.13	35.50	38.15	78.69
*	～によって／により／によるN（理由）	59.988	0.000	68.34	28.00	39.58	85.88
*	～ように／ようなN（例示）	56.542	0.000	85.94	39.50	34.99	77.19
*	～において	54.104	0.000	80.91	37.71	37.78	75.65
5	すなわち	52.104	0.000	76.63	40.50	40.50	71.10
*	～とは（定義）	48.684	0.000	84.09	43.00	44.18	56.85
*	ので（理由）	46.842	0.000	49.66	66.89	38.14	78.71
8	AをBと呼ぶ	45.291	0.000	78.56	38.79	40.45	70.92
9	～ために（理由）	39.953	0.000	68.88	45.39	37.75	79.27
10	たとえば	31.608	0.000	76.22	38.50	44.49	64.06
11	つまり	26.839	0.000	69.81	45.21	44.29	64.83
12	～[た]結果／Nの結果	25.738	0.000	69.28	39.50	45.31	66.35
13	～[の]もとで	24.687	0.000	74.94	43.50	52.22	45.58
14	～というは（定義）	19.007	0.000	65.63	49.00	49.00	55.42
15	AをBと言う	14.494	0.002	65.97	47.00	49.25	55.81
16	AはBと並んで	11.233	0.011	58.56	52.00	52.00	52.00
*	よって	11.232	0.011	58.56	52.00	52.00	52.00
*	から（理由）	10.154	0.017	52.75	73.47	52.21	42.92
19	それゆえに	7.236	0.065	58.00	51.50	51.50	53.73
20	～をはじめ	5.078	0.166	60.31	47.50	52.90	51.54

作品と社説間で起きている識別の誤りが主な原因であることもわかった。

得られた判別関数に基づいて各ジャンル間の近さを総合的に評価するために、図1の各ジャンルのグループ間の重心の距離を求める表7のようになる。この距離の差により、「理由・帰結」「定義」「例示」

表4. ステップワイズ法による
7文型項目選択結果

ステップ	文型項目	Wilksのラムダ	p 値
1	～ように／ ような（例示）	0.469	0.000
2	～によって／により／ によるN（理由）	0.293	0.000
3	ので（理由）	0.216	0.000
4	～とは（定義）	0.175	0.000
5	よって	0.150	0.000
6	～において	0.126	0.000
7	から（理由）	0.109	0.000

表5. 7文型項目の正準判別関数の集計

Wilksのラムダ

関数の検定	Wilksのラムダ	χ^2 乗	p 値
1	0.109	218.365	0.000
2	0.500	68.253	0.000
3	0.865	14.304	0.014

グループ重心の関数

ジャンル	関数1	関数2	関数3
文学作品	-1.796	-0.297	0.904
経済学教科書	2.779	1.478	0.178
社説	-1.346	0.191	-0.269
物理学論文	2.056	-1.218	-0.074

「場所」の4文型群の文型項目を基準とした場合、文学作品と経済学教科書に使用される文型項目が最も大きく異なり、最も類似しているのが文学作品と社説ということがわかる。また、この4文型群を多用する経済学教科書と物理学論文の間でも、使用頻度が高い文型項目に違いがあるということもわかる。

ここで、判別分析結果の構造係数の値とKW検定の平均ランクから、7つの各文型項目がどのジャンルの判別に有効かということを見ていく。

文型項目「～ように／のようなN（例示）」「～において／におけるN」は、経済学教科書と物理学論文を、文学作品ならびに社説から分離するのに非常に効いている。また、「～によって／により／によるN（理由）」「～とは（定義）」「よって」は、経済学教科書を物理学論文と分離するのに有効な項目である。「～によって／により／によるN（理由）」は、

構造係数

文型項目	関数1	関数2	関数3
～のように／のような（例示）	*0.554	0.166	0.256
～によって／により／ によるN（理由）	0.453	*-0.470	-0.394
ので（理由）	0.172	-0.636	*0.647
～とは（定義）	0.267	*0.578	0.257
よって	0.103	*0.269	0.151
～において	*0.440	0.069	0.165
から（理由）	-0.146	0.118	*0.611

* 判別に重要な係数

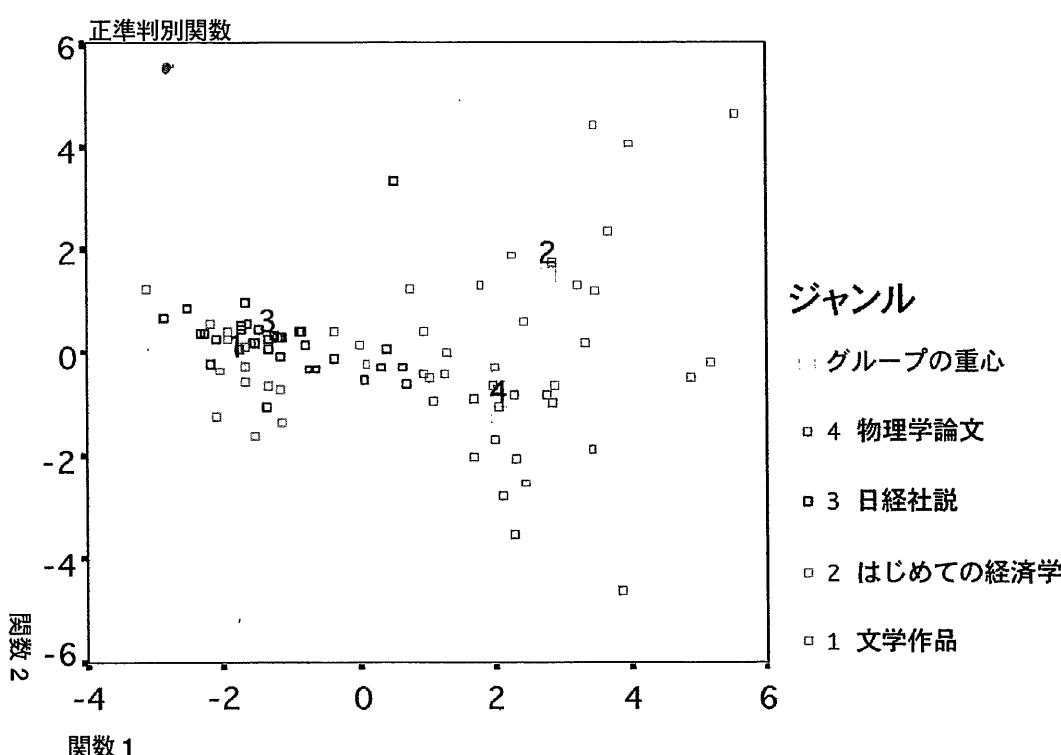


図1. 判別分析による各資料とジャンルの重心のプロット

表6. 正判別率を算出するためのクロス集計表

ジャンル	判別分析に基づく予測グループ				
	文学作品	経済学教科書	社説	物理学論文	合計
文学作品	4		10		14
経済学教科書		11	2	3	16
社説	4	1	45	1	51
物理学論文		1	2	21	24
合計	8	13	59	25	105

7文型項目を用いた判別関数による予測グループと実際のグループ（ジャンル）のクロス集計

表7. グループ重心間のユークリッド距離

\	文学作品	経済学教科書	社説	物理学論文
文学作品				
経済学教科書	4.9541			
社説	2.8317	3.6021		
物理学論文	4.3685	2.3372	3.4016	

\を境に対称である。

物理学論文で非常に多く用いられ、「～とは（定義）」「～って」は、物理学論文より経済学教科書に多いことがわかる。ただし、「～って」については、実際の出現回数を見ると、経済学教科書で2名の経済学者によって3回使用されているのみで、他の3つのジャンルの文章には1回も出現していないため、判別分析には有効な項目ではあるが、使用頻度の高い項目というわけではない。「から（理由）」「ので（理由）」は文学作品を社説から分離するのに有効な項目である。「から（理由）」は、文学作品に多く用いられ、他の3ジャンルとの分離に非常に有効であるが、「ので（理由）」は、文学作品より物理学論文で多く用いられている。

最後に、4ジャンルの判別に有効な7文型項目についての考察を離れ、各ジャンルにおける使用頻度の高い文型項目にはどのような項目があるかを見ていく。表3で4文型群20項目のうちKW検定の結果が有意な18項目の平均ランク70以上のものを比較すると、まず、経済学教科書に特徴的と考えられる文型項目は、「したがって」「～ように／ようなN（例示）」「～において／におけるN」「すなわち」「～とは（定義）」「AをBと呼ぶ」「たとえば」「～[の]もとで」である。物理学論文に特徴的と考えられる文型項目は、「したがって」「～によって／により／によるN（理由）」「～のように／ようなN（例示）」「～において／におけるN」「すなわち」「ので（理由）」「AをBと呼ぶ」「～ために（理由）」である。論述

的形式を持つ文章資料である経済学教科書と物理学論文に共通して頻度が非常に高い文型項目としては、「したがって」「～ように／のようなN（例示）」「～において／におけるN」「すなわち」「AをBと呼ぶ」の5つが抽出できる。文学作品では、70を超えるものは「～から（理由）」1項目のみである。社説は、他の資料との比較において特徴的と言える項目はないことがわかる。

以上の結果により、今回分析した4つのジャンル（経済学教科書、文学作品、社説、物理学論文）の限定された文章資料内では、文学作品や社説より論理構成が強いと考えられる経済学教科書と物理学論文に共通する特徴的な文型項目が抽出できることが判明し、また同時に、文型の出現率を基準として、文章のジャンルを判別することの可能性が示されたと考えられる。

6. おわりに

本論文で抽出された高頻度の文型は、大学における専門分野の日本語教育あるいは一般日本語教育の中・上級の教育経験者にとって、どれも馴染みの深い文型かと想像される。今回の結果は、教育現場の直観と矛盾するものではないが、新しい方法論の枠組みを提示し、そうした直観をより科学的な手法によって検証していくという手続きは非常に重要だと考える。

今回分析した資料は非常に限られた範囲のものではあるが、文章のジャンルによって多用される文型が異なるという本論文の結果から、学習者の目的によっ

て、どのジャンルの文章を教材として選択するかということに慎重である必要があると言えよう。また、短期間に論文作成能力の習得を目指す学習者には、論文作成のために文型項目をグループ化したり、文型提示の順番を考えることによって教育の効率化が図れると考えられる。さらに、経済学教科書、物理学論文というように、専門分野が違っても、論理を展開していく場合に必要な文型に共通性があることがある程度明らかになったことによって、より広い範囲の論述文（論述形式を持つ文章）に共通する文型項目が抽出できるのではないかという期待が強く持たれる。今後の課題としては、文学の論文や物理学の教科書というように、分析対象を広げて検証を行うとともに、「ジャンルの判別に有効な文型群」を踏まえて、論理展開に必要な論述文の文型の確定とその教育への応用を考えていきたい。

本論文は、平成10年度日本語教育学会秋季大会で行った報告をもとに、物理学論文の資料を加えて新たに分析を行った研究成果である。なお、本研究は文部省科学研究費基盤研究C（2）（課題番号11680317 研究代表者村田年）の補助を受けている。

注

注1 A. Kenny (1982) 参照。

注2 A. Kennyらが用い、吉岡が訳した用語。

注3 黒橋・長尾 (1992) 参照。

注4 現代文学の作品は資料の全文入力が必要なため、著作権について出版社に問い合わせたところ、出版社によっては明確な回答が得られなかつたため、後日、資料提示が必要になったときのことを考えて、今回は資料としなかった。ただ、「ジャンル」という用語は近代、現代を通じて使用されていること、また、今回取り上げた作品はどれも現代語として十分通用すると判断したもので、資料選択の際に古語が多く難解なものは外したという経緯がある。

注5 論文の読解、作成を目標とする学習段階は上級レベルであるため、中・上級文型を中心に文型項目を選定した。選定に際しては、専門分野の表現に関する先行研究、松岡(1995)、今村他(1997)、村田(1995他)を参考にした。

注6 柳井他(1986) 参照。

注7 ステップワイズ法とは、各変数毎にその変数が判別に寄与するかどうかをチェックしながら逐次的に変数を投与したり除去したりする手続きを繰り返して最終的な判別関数を求める方法である。変

数の投与または除去に関する判定基準には、Wilks のラムダの統計量と呼ばれる指標を用いた。

注8 文型のグループ化ならびにグループの名称については検討の余地があるが、森田他(1989)を参考に、便宜上表1のように整理した。

参考文献

- 1) Anthony Kenny: *The Computation of Style*, Oxford: Pergamon Press(1982) (邦訳吉岡健一:文章の計量—文学研究のための計量文体学入門, 南雲堂, (1996))
- 2) 黒橋禎夫・長尾眞: 表層表現中の情報に基づく文章構造の自動抽出, 自然言語処理, vol. 1, No. 1, pp.3~20 (1992)
- 3) 松岡弘: 論述文における「論型」の指導について, 言語文化, vol.32, pp.87~108(1995)
- 4) 今村和宏他: 専門基礎文献の語彙・表現調査から教材開発へ, 平成9年度日本語教育学会秋季大会予稿集, pp. 147~154 (1997)
- 5) 森田良行他: 日本語表現文型, アルク (1989)
- 6) 柳井晴夫他: 多変量解析ハンドブック, 現代数学社 (1986)
- 7) 村田 年: 上級日本語教育の方法を探る2—文型教育から進める教材開発ー, 日本語と日本語教育, vol.24, pp.17~37 (1995)
- 8) 村田 年: 異なるジャンルの文章における文型の出現傾向の相違—論述文を支える文型の確定を目指してー, 平成10年度日本語教育学会秋季大会予稿集, pp.165~171 (1998)
- 9) 村田 年: 文章と文型-ジャンルの異なる文章における文型の使用頻度調査-, 日本語と日本語教育, vol.27, pp.21-42 (1998)
- 10) 村田 年: 接続語句・助詞相当句による文章の所属ジャンルの判別—多変量解析法を用いてー, 言語処理学会第5回年次大会予稿集, pp.213~216 (1999)

謝辞: 統計分析については、統計数理研究所の共同利用登録制度によって、同研究所の前田忠彦助手からご助言をいただきました。深く感謝いたします。また、平成10年度日本語教育学会秋季大会の発表会場で貴重なご意見を下さった方々に感謝を申し上げます。

著者紹介

村田 年: 慶應義塾大学国際センター助教授

〔経歴〕 慶應義塾大学文学部中国文学科卒業、同大学院修士課程修了、アメリカ・カナダ11大学連合日本研究センター教務助手、東京外国语大学外国语学部附属日本語学校専任講師を経て、現職

〔専門〕 日本語教育