

専門日本語教育における論述文指導のための 接続語句・助詞相当句の研究

村田 年[†]

(受付 2007年1月31日; 改訂 2007年5月31日)

要 旨

専門日本語教育においては、外国人日本語学習者にとって論文に代表される論述文の理解は不可欠であり、その理解には接続語句・助詞相当句が役立つと考えられる。本論文では、7ジャンル(経済学教科書、経済学論文、工学論文、物理学論文、文学論文、新聞社説、文学作品)計370編の文章における指標65項目(接続語句・助詞相当句)の出現率を調査し、以下の分析を行う。(1)7ジャンル計370編の資料を対象に単変量的解析を行った後、正準判別分析(多変量解析の一手法)を用いて分析を行う。(2)論述文を代表する論文ジャンルの4つの資料を対象に(1)と同様の分析を行う。

上記の結果、文章の所属ジャンルが19の語句項目によって正判別率84.6%という高率で判別されるとともに、各ジャンルを分離する語句項目ならびに論述的形式を持つ文章に共通する語句項目が選択された。

以上、限定された資料内ではあるが、異なるジャンルの文章を判別するために、(i)接続語句、(ii)助詞相当句が有効な指標であることが再確認された。

キーワード：専門日本語教育(JSP)、接続語句、助詞相当句、ジャンル、判別分析、論述文。

1. はじめに

日本語の母語話者がある日本語の文章を読んだとき、それが小説か新聞か論文かということを知覚するのがさほど難しいということは一体何を意味するのだろうか。この問いに対する一つの答として、我々が事前知識としてそれぞれの形式(パターン)に関して学習を行っているからだという金(1999)の記述を引くことができよう。この形式(パターン)は、ジャンルの概念にもつながるものだと言え、母語の学習のみならず外国語の学習の場合にも非常に重要だと考えられる。私たちが普段使っている「ジャンル」という語の定義は一様ではない。ある場合には「小説」「新聞社説」「論文」という分類に、ある場合にはその下位分類とも言える「日記」「紀行文」「手紙」というような分類にも使われている。本論文では「ジャンル」という語を、個々人の持つ文体的特徴を超えたところに存在する、ある特徴パターンを持った文章グループという意味で用いることにする。筆者が現在携わる日本語教育学分野では、専門分野での学習・研究を目的とする留学生は、短期間に論述的な文章の理解や作成能力を修得することが要求される。そのため、文章指導の際に学習者に対して、文章がジャンルによって異なる表現形式を

[†] 慶應義塾大学 日本語・日本文化教育センター：〒108-8345 東京都港区三田 2-15-45

持つことを客観的かつ具体的に示すことが不可欠だと考えられる。具体的には、文章のジャンルによって異なる特徴パターンを選択し、提示することが重要となる。

本研究の目的は、村田(2000, 2002)に引き続き、論述的な文章の文脈展開において、表層表現¹⁾としての接続語句と助詞相当句が重要な役割を果たすと考え、ジャンルによって異なる文章の特徴が、これらの接続語句・助詞相当句の使用傾向の違いに反映されることを実証することである。また、異なるジャンルの文章に特徴的な接続語句と助詞相当句を選択することによって、その知見を専門分野の日本語教育の文章指導に応用することを目的とする。本論文では、村田(2000, 2002)で残された課題²⁾を解決すべく、対象とする文章資料のうち論文ジャンルについては経済学分野の論文を加えるとともに、文学作品ジャンルについても資料数を増やし文章資料を再編した。また、従来の指標 62 語句(接続語句と助詞相当句)についても見直しを行い、新たに 65 語句とした。本論文では、文章資料再編後の 370 編を対象に新たに定めた 65 語句を指標として従来と同じ多変量解析の方法論を用いて分析を行ったので報告する。

2. 分析資料

専門分野の学習・研究を目的とする留学生は、一般的に日本語学習に割ける時間は短い。そのため、短期間で日本語による論文の読解力、論文の作成力を身につける必要がある。その学習者にとって、専門家による教科書あるいは論文の文章は専門分野の論述的な文章作成の際の重要な文章モデルになると考えられる。そこで、村田(2000, 2002)では分析資料として、学習者の非常に多い分野の一つである経済学分野から入門教科書を、同じく学習者の多い理工学分野から科学技術論文(物理学論文, 工学論文)を、また学習者は少ないが日本研究の一環として日本文学を研究する学習者が触れる文学論文を選んだ。一方、比較のための資料としては、専門分野の教科書や論文より論理が非明示的だと考えられる新聞社説と文学作品を選んだ。本論文ではこれらの文章資料に加え、さらに経済学分野の論文 19 編、文学作品についても新たに 25 編の現代文学作品を収集した。また近代作品についても作品の一部を見直し作品数を 26 編に増やし、文学作品資料全体を 51 編とした。対象資料は下記の 370 編である³⁾。

- A. 経済学入門書：『はじめての経済学』、岡田泰男ほか共編、慶応義塾大学出版会(1995)。16 編総文数 1124 文。
- B. 経済学論文：『経済研究』、一橋大学経済研究所編(2000)。VOL. 51, No. 1~No. 4 掲載のすべての論文 19 編。総文数 3889 文。ただし、数式は除く。
- C. 工学論文：学術雑誌論文 14 編。専攻分野別に電気工学 6 編、機械工学 4 編、計算機科学 2 編、管理工学 2 編。ただし、数式・記号は除く。14 編総文数 1725 文。
- D. 物理学論文：『日本物理学会誌』の 1997 年第 52 巻の No. 1~12 の「最近の研究から」に掲載された論文 24 編総文数 2243 文。ただし、数式・記号は除く。
- E. 文学論文：文学分野の学術雑誌 4 誌『日本近代文学』『近世文藝』『中世文学』『中古文学』の合計 239 編の中から単純無作為抽出法で選んだ 24 編。ただし、引用部分は除く。総文数 4184 文。
- F. 新聞社説：1996 年 12 月一ヶ月間の四大紙『日本経済新聞』『朝日新聞』『読売新聞』『毎日新聞』の朝刊、夕刊の社説総計 222 編総文数 6514 文。
- G. 文学作品⁴⁾：近代文学作家 12 名による近代文学作品から 26 編(新字新仮名遣い)ならびに星新一編集の短編小説(ショートショート)と泡坂妻夫の短編小説から 25 編の合計 51 編、総文数 4398 文。

A から G までの 7 つの文章資料のうち、論述的な文脈展開を持つと考えられる文章資料 A か

ら E までについては、経済学入門書は Ong (1991) を踏まえ専門分野の教科書として 1 つのジャンルに所属すると仮定し⁵⁾、論文は工学論文、物理学論文、文学論文、経済学論文の各論文を 1 つずつ 1 つのジャンルに所属すると仮定する。新聞社説については四大紙を合わせて 1 つの社説ジャンルと仮定し、文学作品についても近代文学作品と現代文学作品を合わせて 1 つの文学作品ジャンルとする。このように文章資料は合計 7 つのジャンルと考える。

3. 分析

3.1 指標としての接続語句・助詞相当句とその相対出現率

接続語句・助詞相当句は村田(2000, 2002)では 62 語句を指標としたが、そこで残された課題を解決すべく、本論文では一部指標語句の見直しを行った⁶⁾。接続語句・助詞相当句の使用頻度については、一文当たりの頻度に換算しなおした相対使用頻度を求めて各語句の出現率とし、語句のうち、形態が同じで意味機能を 2 つ以上持つものについては、細分化して別の指標語句とした(例:「～ため [に] <目的>」と「～ため [に] <理由>」)。意味機能は語句の後に <> 付きで記した。このように各語句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、意味機能別に頻度を調べた。その結果、語句の総数は最終的に下記の 65 語句となった。

一方、～うえ [で]、～うえ [に]、～から、～から～にかけて、しかしながら、したがって、すなわち、～ず [に]、～ [た] 結果、ただし、たとえば、他方、～ため [に] <目的>、～ため [に] <理由>、～だけでなく、～つつ<同時>、～つつ<逆接>、つまり、～というのは、～として、～とすれば/とすると/としたら/としても、～と同時に、～とともに<同時>、～とともに<継起>、～とは、なお、～ながら<同時>、～ながら [も] <逆接>、～にあたって、～において、～に限らず、～に関して、～にしたがって<根拠>、～にしたがって<継起>、～に対して<対象>、～のに対して<対比>、～について、～にとって、～にともなって、～にもかかわらず、～にもとづいて、～によって<根拠>、～によって<動作主体>、～によって<方法>、～によって<理由>、～によって<対応>、～によると<根拠>、～によると<伝聞>、～によれば<根拠>、～によれば<伝聞>、～にわたって、～ので<理由>、～のみならず、～反面、～ばかりでなく、～まま、むしろ、～ものの、[それ] ゆえ [に]、よって、～を通じて<範囲>、～を通じて<媒介>、～をはじめ [に/として]、～をめぐって

3.2 分析方法

上記第 2 章で挙げた A から G までの 370 編、合計 7 ジャンルの文章資料について分析を行う⁷⁾。

- (1) まず単変量的に個々の変数である 65 語句のジャンルごとの分布の違いを検討するために、Kruskal-Wallis 検定(以後 KW 検定)を用いた。
- (2) 次に 65 語句の中からジャンルの判別に特に有効な語句を選択するために、多変量解析の一手法である正準判別分析⁸⁾のステップワイズ法を用いて分析を行い、判別に寄与する語句を選択した⁹⁾。

本論文の主目的は(2)の判別の可能性の検討および、論述文を特徴付けるのに有用な情報の選択である。

4. 分析結果

4.1 65 全変数の単変量的分布の比較

3.2 節分析方法(1)により、各ジャンルで使用頻度が高いのはどのような語句かを調べるため

表 1. 56 語句の KW 検定結果.

語句<代表項目>	検定統計量		経済学入門書		経済学論文		工学論文		物理学論文		文学論文		新聞社説		文学作品	
	P値	中央値	平均	中央値	平均	中央値	平均	中央値	平均	中央値	平均	中央値	平均	中央値	平均	中央値
1) において	237.072	0.000	293.47	0.045	325.32	0.065	327.50	0.063	280.83	0.035	302.83	0.033	144.97	0.000	136.91	0.000
2) しからず	229.947	0.000	295.97	0.000	296.38	0.014	291.61	0.013	309.48	0.019	226.77	0.002	151.68	0.000	151.15	0.000
3) すなわち	239.696	0.000	272.16	0.035	325.47	0.024	276.75	0.011	260.88	0.019	274.10	0.019	149.72	0.000	158.38	0.000
4) によって/によるN<理由>	183.003	0.000	277.16	0.035	303.37	0.051	280.64	0.033	337.65	0.083	249.38	0.016	154.75	0.000	182.80	0.000
5) なお	174.643	0.000	198.25	0.000	278.92	0.009	309.46	0.016	214.67	0.000	254.08	0.005	163.40	0.000	162.80	0.000
6) によって/によるN<方法>	170.889	0.000	248.06	0.018	319.24	0.027	349.68	0.135	320.04	0.038	213.13	0.010	159.03	0.000	124.15	0.000
7) ため	140.586	0.000	215.66	0.005	307.60	0.033	306.46	0.037	215.21	0.008	283.33	0.015	163.02	0.000	135.32	0.000
8) に関して	131.206	0.000	282.72	0.007	288.92	0.007	391.00	0.000	236.15	0.000	286.77	0.013	161.32	0.000	169.33	0.000
9) つまり	128.010	0.000	301.41	0.103	246.76	0.057	272.29	0.067	231.15	0.049	313.17	0.113	176.14	0.032	37.78	0.000
10) として	123.787	0.000	185.19	0.000	265.37	0.014	263.61	0.020	285.54	0.024	240.42	0.007	149.20	0.000	219.51	0.000
11) の<理由>	121.249	0.000	276.50	0.029	274.05	0.031	295.54	0.044	308.25	0.053	201.27	0.009	158.69	0.000	145.27	0.000
12) ため<理由>	115.972	0.000	300.50	0.016	211.03	0.000	196.25	0.000	206.94	0.000	300.04	0.009	167.49	0.000	164.15	0.000
13) とは<主観>	110.681	0.000	178.84	0.000	255.66	0.007	229.75	0.000	177.35	0.000	300.04	0.012	172.57	0.000	155.50	0.000
14) によって/によるN<根拠>	110.007	0.000	180.91	0.000	283.34	0.009	181.29	0.008	198.94	0.000	244.96	0.004	172.95	0.000	172.01	0.000
15) によれば<根拠>	100.146	0.000	269.72	0.016	279.71	0.014	253.11	0.008	223.23	0.010	261.96	0.012	163.76	0.000	146.32	0.000
16) ため	99.208	0.000	238.56	0.042	322.97	0.092	229.39	0.038	217.71	0.038	249.29	0.046	179.35	0.026	87.17	0.000
17) について	94.195	0.000	206.69	0.000	200.86	0.000	184.71	0.000	271.84	0.010	186.83	0.000	176.33	0.000	172.00	0.000
18) とともに<継続>	92.159	0.000	203.88	0.000	137.61	0.000	211.21	0.003	223.54	0.009	193.83	0.000	160.00	0.000	279.71	0.029
19) なから<同時>	86.538	0.000	240.63	0.000	247.05	0.005	244.75	0.003	220.15	0.000	240.21	0.002	168.95	0.000	159.00	0.000
20) の対して<対比>	83.949	0.000	207.06	0.000	229.76	0.000	225.54	0.000	225.19	0.000	226.60	0.000	172.00	0.000	172.00	0.000
21) しからず	78.946	0.000	235.50	0.000	234.26	0.004	176.50	0.000	184.06	0.000	214.58	0.000	177.37	0.000	176.50	0.000
22) ~から~にかけて	73.649	0.000	258.84	0.013	251.24	0.005	216.57	0.000	250.90	0.000	180.25	0.000	173.59	0.000	153.00	0.000
23) ~は<結果>	71.240	0.000	244.38	0.013	232.89	0.011	209.18	0.003	205.29	0.000	288.02	0.017	174.16	0.000	131.66	0.000
24) によって/によるN<動作主体>	65.293	0.000	219.03	0.000	270.50	0.008	244.00	0.003	209.06	0.000	194.81	0.000	174.77	0.000	158.50	0.000
25) によって/によるN<対成>	65.247	0.000	167.03	0.000	245.26	0.006	277.75	0.007	210.00	0.000	242.23	0.006	176.19	0.000	146.00	0.000
26) にもよって	65.221	0.000	245.28	0.043	323.53	0.022	273.43	0.037	272.02	0.032	209.44	0.015	174.58	0.000	120.85	0.000
27) に対して<対象>	63.777	0.000	222.75	0.000	243.92	0.000	189.86	0.000	198.38	0.000	183.46	0.000	178.65	0.000	176.00	0.000
28) 地方	63.287	0.000	181.94	0.000	273.55	0.006	169.50	0.000	174.79	0.000	241.90	0.001	176.67	0.000	177.61	0.000
29) 場合	60.777	0.000	254.53	0.007	263.82	0.008	186.54	0.000	183.67	0.000	186.71	0.000	178.26	0.000	161.50	0.000
30) を通じて<媒介>	59.064	0.000	236.91	0.000	188.13	0.000	178.50	0.000	201.38	0.000	209.00	0.000	178.50	0.000	182.25	0.000
31) というのは	58.689	0.000	238.19	0.000	237.89	0.005	243.89	0.003	234.98	0.000	190.40	0.000	174.58	0.000	155.50	0.000
32) ともなう	57.204	0.000	217.00	0.000	257.79	0.011	191.21	0.000	225.27	0.004	258.71	0.008	164.95	0.000	183.42	0.000
33) とすれば/とするど/としたら/としても	56.573	0.000	196.44	0.000	212.50	0.000	187.36	0.000	181.48	0.000	243.83	0.000	177.44	0.000	181.04	0.000
34) [それゆえに]	55.703	0.000	195.38	0.000	231.34	0.000	183.75	0.000	215.60	0.010	272.50	0.005	176.69	0.000	149.04	0.000
35) うえで	53.448	0.000	221.03	0.000	174.50	0.000	227.29	0.000	119.63	0.000	204.94	0.000	177.98	0.000	174.50	0.000
36) したからって<根拠>	53.431	0.000	248.72	0.048	235.97	0.036	288.61	0.067	173.10	0.024	146.08	0.014	192.42	0.032	112.81	0.000
37) 同時に	52.801	0.000	237.78	0.000	185.68	0.000	179.68	0.000	205.42	0.000	241.79	0.000	173.23	0.000	188.17	0.000
38) 同時に	52.275	0.000	205.75	0.000	182.50	0.000	222.04	0.000	182.50	0.000	190.19	0.000	182.50	0.000	182.50	0.000
39) よって	48.875	0.000	232.75	0.013	226.74	0.008	200.32	0.004	235.58	0.011	244.81	0.012	179.20	0.000	127.19	0.000
40) 一方	48.226	0.000	180.50	0.000	180.50	0.000	180.50	0.000	180.50	0.000	226.25	0.000	183.89	0.000	180.50	0.000
41) つつ<連続>	47.621	0.000	213.00	0.000	226.95	0.000	178.50	0.000	178.50	0.000	201.38	0.000	181.06	0.000	178.50	0.000
42) を通じて<範囲>	42.647	0.000	229.03	0.000	237.16	0.004	210.79	0.000	203.50	0.000	217.83	0.000	176.65	0.000	160.50	0.000
43) にわたって	40.326	0.000	181.84	0.000	224.34	0.000	209.04	0.000	181.55	0.000	250.21	0.005	178.67	0.000	166.85	0.000
44) もの	39.969	0.000	260.84	0.012	225.18	0.005	147.50	0.000	160.67	0.000	230.67	0.005	181.31	0.000	166.18	0.000
45) にとつて	38.856	0.000	210.53	0.000	185.08	0.000	203.50	0.000	180.58	0.000	247.25	0.003	180.14	0.000	169.43	0.000
46) にもかわらず	38.807	0.000	173.03	0.018	208.45	0.019	116.25	0.000	132.73	0.005	193.65	0.019	177.57	0.016	255.40	0.048
47) から<理由>	35.897	0.000	193.19	0.000	226.87	0.000	171.00	0.000	208.77	0.000	201.33	0.000	180.57	0.000	171.00	0.000
48) による<根拠>	33.840	0.000	197.19	0.000	180.34	0.000	200.43	0.000	169.56	0.000	250.08	0.006	182.60	0.000	169.38	0.000
49) つつ同時	30.014	0.000	213.81	0.000	191.13	0.000	222.79	0.000	160.63	0.000	251.06	0.007	176.74	0.000	169.16	0.000
50) とともに<同時>	23.189	0.000	162.22	0.000	187.13	0.004	175.36	0.000	171.94	0.000	246.79	0.008	174.52	0.000	222.79	0.005
51) まま	26.200	0.000	191.47	0.000	164.11	0.000	170.21	0.000	161.08	0.000	258.35	0.007	180.12	0.000	190.30	0.000
52) なから<連続>	25.912	0.000	192.72	0.000	190.95	0.000	207.50	0.000	203.94	0.000	185.65	0.000	181.00	0.000	184.69	0.000
53) したからって<継続>	24.399	0.000	168.47	0.000	176.55	0.000	158.50	0.000	158.50	0.000	201.13	0.000	196.71	0.000	158.50	0.000
54) めくって	19.045	0.004	205.97	0.013	207.42	0.014	203.36	0.008	224.31	0.013	239.15	0.016	169.86	0.000	190.57	0.005
55) すい<]	17.267	0.008	200.94	0.000	187.66	0.000	191.11	0.000	185.65	0.000	208.58	0.000	180.58	0.000	188.79	0.005
56) のみならず																

P<0.01. なお、中央値とはジャンルの中での出現率の中央値である。

表 2. 56 語句のジャンルによる出現分布の違い.

語句<代表項目>	経済学入門書	経済学論文	工学論文	物理学論文	文学論文	新聞社説	文学作品
1 において	◎	◎	◎	◎	◎		
2 したがって	◎	◎	◎	◎	◎		
3 すなわち	◎	◎	◎	◎	◎		
4 によって<理由>	◎	◎	◎	◎	◎		
5 なお		◎	◎		◎		
6 によって<方法>	◎	◎	◎	◎			
7 ただし	◎	◎	◎				
8 に関して		◎	◎		◎		
9 つまり	◎	◎			◎		
10 として	◎	◎	◎		◎		
11 ので<理由>		◎	◎	◎	◎		
12 ため[に]<理由>	◎	◎	◎	◎			
13 とは	◎				◎		
14 によって<根拠>		◎			◎		
15 によれば<根拠>		◎			◎		
16 たとえば	◎	◎	◎		◎		
17 について		◎			◎		
18 とともに<継起>				◎			
19 ながら<同時>							◎
20 のに対して<対比>	◎	◎	◎		◎		
21 しかしながら							
22 ~から~にかけて							
23 ~[た]結果	◎	◎		◎			
24 によって<動作主体>	◎				◎		
25 によって<対応>		◎	◎				
26 にもとづいて		◎	◎		◎		
27 に対して<対象>	◎		◎	◎			
28 他方		◎					
29 むしろ		◎			◎		
30 を通じて<媒介>	◎	◎					
31 というのは							
32 にもなって			◎				
33 とすれば*		◎			◎		
34 [それ]ゆえ[に]					◎		
35 うえ[で]					◎		
36 にしたがって<根拠>							
37 ため[に]<目的>	◎		◎				
38 と同時に					◎		
39 よって							
40 一方					◎		
41 つつ<逆接>							
42 を通じて<範囲>							
43 にわたって							
44 ものの					◎		
45 にとって	◎						
46 にもかかわらず					◎		
47 から<理由>							◎
48 によると<根拠>							
49 つつ<同時>					◎		
50 とともに<同時>					◎		
51 まま					◎		
52 ながらも<逆接>					◎		
53 にしたがって<継起>							
54 をめぐって							
55 ず[に]					◎		
56 のみならず							

◎ : 平均ランク239以上

*この項目には「とすると/としたら/としても」が含まれる.

に、65全語句の単変量的分布をKW検定の平均ランクによって見ていく。65語句のうちKW検定結果が有意であった56語句について、その結果を表1に示す。また同じ56語句のジャンル別の出現分布については表2にまとめる。表2内には便宜上、平均ランク239以上¹⁰⁾のものを記す。なお、この結果は専門日本語教育の実践家にとって非常に有用だと考えられる。

本分析結果によって論述的形式を持つ文章間に共通して用いられる語句が多く存在することが確かめられた。表1、表2を見るとわかるように平均ランクが239を超え4つの分野の論文すべてに共通している語句は、「～において／におけるN」「すなわち」「～によって／によるN<理由>」「ので<理由>」の4つで、3つの論文に共通している語句は「したがって」「なお」「～によって／によるN<方法>」「～に関して／に関するN」「～として」「～ため[に]<理由>」「たとえば」「～のに対して<対比>」「～にもとづいて」の9つである。これら13語句のうち9語句(「～において／におけるN」「したがって」「すなわち」「～によって／によるN<理由>」「～によって／によるN<方法>」「～として」「～ため[に]<理由>」「たとえば」「～のに対して<対比>」)は、経済学入門書の平均ランク239以上の語句と共通していることがわかる。これらの語句は論述的形式を持つ文章に共通して用いられる語句だと言えるだろう。今回新たに加えた経済学論文と経済学教科書を比較してみるとその多くが共通しているものの、経済学論文では他の論文同様「なお」「～に関して／に関するN」「ので<理由>」「～にもとづいて」が多用され、教科書ではあまり用いられていないことがわかる。また反対に経済学教科書では「とは」「～によって／によるN<動作主体>」「～ため[に]<目的>」「にとって」が多用されているが、経済学論文をはじめ他の論文ではあまり用いられていない。経済学教科書と経済学論文に共通して多用されているのは「～を通じて<媒介>」のみである。この結果は専門日本語教育の現場で学習者が習得すべき「語句」の一部が実証的に示されていると言えよう。新聞社説では平均ランク239を超える語句項目はないが「～をめぐって」(196.71)が一番高く文学論文につぐランクとなっている。文学作品については平均ランク239以上のものは「～ながら<同時>」と「～から<理由>」の2項目のみであった。このほかランクが200を超える語句を見ていくと「～まま(222.79)」「～ので(219.51)」があり、理由を表す「～から<理由>」「～ので<理由>」や付帯状況を表す「～ながら<同時>」「～まま」が文学作品を特徴付けていると考えられる。このように文脈展開が明示的だと考えられる経済学入門書と論文では、多くの語句が共通して用いられているのに対して、比較のための資料として選んだ新聞社説と文学作品では指標の65語句があまり使われていないという結果となっている。

4.2 判別分析の結果とその有効性

3.2節分析方法(2)により、論述文の特徴を記述するのに有用な情報の抽出を行うため、多変量による分析を行った。65語句の出現率を説明変数とし、文章資料グループ(以下ジャンル)を基準変数(判別目的であるグループ)としてステップワイズ法を用いた判別分析を行った。その結果、逐次的に19個の説明変数が予測的に組み込まれ、その手続き内で削除された変数もなく、7つのジャンルの判別に有効な、以下の19語句が選択された。

- ①～において／におけるN ②したがって ③すなわち ④しかしながら ⑤～によって／によるN<方法> ⑥～とは<定義> ⑦～から～にかけて ⑧なお ⑨～について／についてのN ⑩～ながら<同時> ⑪～ので<理由> ⑫～から<理由> ⑬～ため[に]<目的> ⑭～によって／によるN<理由> ⑮よって ⑯～とともに<継起> ⑰ただし ⑱～として ⑲[それ]ゆえ[に]

本論文では文章資料の所属ジャンルは7つなので6つの判別関数が算出された¹¹⁾。判別関数によるグループの分離の程度を示す記述的指標としてウィルクスの Λ ¹²⁾を用いることができ

表 3. 判別関数の固有値等.

判別関数	固有値	寄与率	p 値	Λ	χ^2 乗
関数 1	3.792	55.3	0.000	0.020	1390.581
関数 2	0.956	13.9	0.000	0.096	832.749
関数 3	0.754	11.0	0.000	0.189	593.984
関数 4	0.586	8.5	0.000	0.331	394.018
関数 5	0.462	6.7	0.000	0.524	229.919
関数 6	0.304	4.4	0.000	0.767	94.606

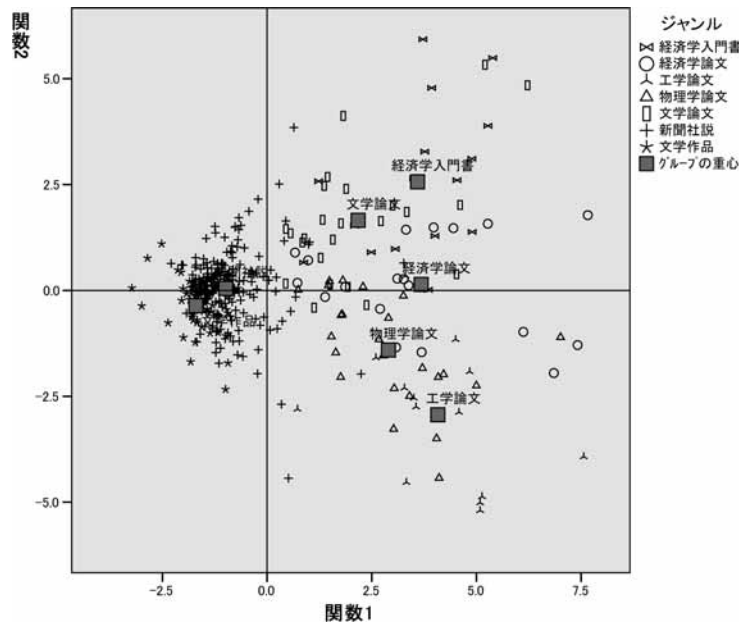


図 1. 判別分析による各文章資料とジャンルごとの重心のプロット.

る。また推測統計的立場からは、 Λ に基づく χ^2 検定で関数の有意性の検定を行うことができる。本研究では正規性の仮定が満たされないため、推測統計的指標は目安としてのみ用いる。これらに関する指標を示すと表 3 のようになる。表 3 の Λ の値は、関数 1 と関数 2 の判別に、相対的に大きく寄与する情報が含まれていることを示している。

次に、判別空間におけるジャンル間の関係について検討する。選択された 19 語句による判別関数平面での各文章資料(個体)の判別得点とジャンルの重心をプロットしたものが図 1 である。ここでは関数 2 までの情報量が多いことから、関数 1 と関数 2 の平面におけるプロットを示す。また表 4 には各ジャンルの判別空間における重心を示す。

関数 1 × 関数 2 の平面では、文学作品と社説が相互に近い位置にあり、4 つの分野の論文と経済学入門書はそれぞれ文学作品と新聞社説から離れ、原点 0 を境に正の領域である右の第一象限と第四象限とに位置している。4 つの分野の論文のうち、文学論文は経済学入門書に相対的に近い位置にあり、経済学論文と物理学論文も近い位置にあることが読み取れる。

表 4 を検討すると、関数 1 によって経済学入門書、経済学論文、工学論文、物理学論文、文学論文が、新聞社説と文学作品から大きく分離されていることがわかる。関数 2 では、経済学入

表 4. 判別空間における各ジャンルの重心.

ジャンル	関数1	関数2	関数3	関数4	関数5	関数6
経済学入門書	3.594	2.566	1.988	-0.052	1.244	0.925
経済学論文	3.680	0.141	-2.161	0.472	-1.104	1.272
工学論文	4.078	-2.937	-0.759	-0.391	2.177	-0.280
物理学論文	2.897	-1.407	1.922	0.076	-1.439	-0.376
文学論文	2.174	1.657	-0.940	0.585	-0.051	-1.586
新聞社説	-0.989	0.044	-0.068	-0.475	-0.032	0.020
文学作品	-1.701	-0.360	0.224	1.702	0.263	0.149

表 5. 19 語句を用いた判別関数による予測グループと実際のジャンルのクロス集計.

ジャンル	判別分析に基づく予測グループ							合計
	経済学入門書	経済学論文	工学論文	物理学論文	文学論文	新聞社説	文学作品	
経済学入門書	13	0	0	1	0	2	0	16
経済学論文	0	12	0	1	2	4	0	19
工学論文	0	2	10	0	1	1	0	14
物理学論文	0	0	1	20	0	3	0	24
文学論文	1	1	0	0	20	2	0	24
新聞社説	1	0	1	2	0	213	5	222
文学作品	0	0	0	0	0	26	25	51
合計	15	15	12	24	23	251	30	370

表 6. 各ジャンルの重心間のユークリッド距離.

	経済学入門書	経済学論文	工学論文	物理学論文	文学論文	新聞社説
経済学入門書						
経済学論文	4.929					
工学論文	6.156	4.784				
物理学論文	4.831	4.665	4.959			
文学論文	4.148	4.055	5.520	4.374		
新聞社説	5.640	5.426	6.324	4.920	4.003	
文学作品	6.490	6.269	7.021	5.596	4.817	2.377

門書と文学論文が工学論文と物理学論文から分離され、関数3では経済学入門書と物理学論文が経済学論文から、関数4では文学作品が新聞社説と工学論文から、関数5では工学論文と経済学入門書が物理学論文と経済学論文から、関数6では文学論文が経済学論文と経済学入門書から分離されている。判別分析においては、判別の可否は普通、正判別率によって評価され、正判別率が高いほど、説明変数が基準変数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、サンプルの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示す「見かけ的中率」によって簡便に評価することができる。本研究における判別の可否を評価するために、見かけ的中率を算出するためのクロス集計表を表5に示す¹³⁾。

表5のクロス集計結果から、正判別率は84.6% (13 + 10 + 20 + 20 + 12 + 213 + 25/370) という高い値となり、選択された19語句項目によって7つのジャンルの文章資料は十分判別が可能であることが検証された。また、誤判別は、新聞社説と文学作品間の識別の誤りによることもわかった。

次に、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と表6のようになる。

表6の距離の値から、用いられる語句が最も大きく異なるのが工学論文と文学作品で、最も似ているのが新聞社説と文学作品ということになる。また、4つの分野の論文資料に専門分野

表 7. 選択された 19 語句の構造係数.

語句<代表項目>	関数1	関数2	関数3	関数4	関数5	関数6
1 ~において	.494(*)	-0.019	-0.325	0.139	0.051	0.099
2 したがって	.443(*)	-0.008	0.278	0.065	-0.034	0.314
3 すなわち	.390(*)	0.239	-0.061	0.202	0.013	-0.040
4 しかしながら	.221(*)	-0.002	-0.042	0.095	-0.048	-0.188
5 ~によって<方法>	0.389	-.590(*)	0.090	-0.178	0.279	-0.127
6 ~とは<定義>	0.168	.383(*)	0.145	0.008	0.237	-0.115
7 ~から~にかけて	0.143	.290(*)	0.064	0.041	0.098	0.269
8 なお	0.311	-0.099	-.338(*)	0.101	0.131	-0.007
9 ~について	0.175	0.122	-.336(*)	-0.227	-0.262	0.208
10 ながら<同時>	-0.092	-0.115	0.166	.689(*)	0.186	0.062
11 ~ので<理由>	0.182	-0.235	0.051	.394(*)	-0.274	-0.034
12 ~から<理由>	-0.096	0.073	-0.057	.368(*)	0.033	0.115
13 ~ため[に]<目的>	0.052	-0.073	-0.062	-.357(*)	0.223	0.197
14 ~によって<理由>	0.369	-0.109	0.356	-0.060	-.449(*)	0.008
15 よって	0.114	0.049	0.094	-0.020	.371(*)	0.063
16 ~とともに<継起>	0.130	-0.112	0.306	-0.023	-.357(*)	-0.051
17 ただし	0.269	0.065	-0.432	0.059	-0.154	-.540(*)
18 ~として	0.288	0.254	-0.094	-0.155	0.117	-.469(*)
19 [それ]ゆえ[に]	0.088	0.170	-0.194	0.095	-0.039	-.375(*)

(*) 有意な係数

の教科書である経済学入門書を加え、論述的形式を持つ5つの文章資料間で比較すると、経済学入門書と最も異なるのが工学論文で、最も近いのは文学論文であることがわかる。さらに4つの分野の論文資料間で比べると、最も遠いのが工学論文と文学論文で、最も近いのが経済学論文と文学論文だと言える。この4つの分野の論文資料については4.3節で改めて分析を加えることにする。ここでは正準判別分析で選択された19語句がどのジャンルの判別に有効かを見ていく。この考察のためには構造係数と判別空間における各ジャンルの重心の関係を見ることが有効である。構造係数とは正準判別関数と個々の変量との間の相関係数である。表7として19語句の構造係数を示す。

関数1では、「~において/におけるN」「したがって」「すなわち」「しかしながら」が経済学入門書、経済学論文、工学論文、物理学論文、文学論文を、新聞社説と文学作品から大きく分離するのに有効である。関数2では、「~によって/によるN<方法>」「~とは(定義)」「~から~にかけて」が、経済学入門書と文学論文を工学論文と物理学論文から分離している。関数3では、「なお」「~について/についてのN」が、経済学入門書と物理学論文を経済学論文から分離するのに有効である。関数4では、「~ながら<同時>」「~ので<理由>」「~から<理由>」「~ため[に]<目的>」が、文学作品を工学論文、新聞社説から分離し、関数5では、「~によって/によるN<理由>」「よって」「~とともに<継起>」が、経済学入門書と工学論文を経済学論文と物理学論文から分離している。関数6では、「ただし」「~として」「[それ]ゆえ[に]」が、経済学入門書と経済学論文を文学論文から分離するのに有効であることがわかる。以上の結果を4.1節の単変量的検討の結果と対照しても、選択された19語句は単変量的検討の際にp値がすべて0.000であり、矛盾はないと言えよう。

4.3 4つの分野の論文資料の分離に関する分析結果

ここでは特に論文ジャンルの4つの分野の資料を対象として、4.2節の分析で選択された19語句以外に論文ジャンルをさらに分離するのに有効な情報が含まれているかどうかを検討するため、4.2節と同様の議論を進めることにする。

表 8. 26 語句の KW 検定結果.

語句<代表項目>	検定統計量	P値	経済学論文		工学論文		物理学論文		文学論文	
			平均	中央値	平均	中央値	平均	中央値	平均	中央値
1) によって<方法>	40.302	0.000	35.05	0.027	67.00	0.135	50.67	0.058	20.88	0.010
2) によって<理由>	33.348	0.000	42.76	0.051	32.39	0.033	61.65	0.083	23.98	0.016
3) ため[に]<目的>	31.378	0.000	52.24	0.036	64.43	0.067	33.96	0.024	25.48	0.014
4) について	28.033	0.000	64.74	0.092	32.64	0.038	28.69	0.038	39.40	0.046
5) ただし	25.873	0.000	61.24	0.025	43.11	0.006	28.35	0.000	36.40	0.003
6) に関して	25.694	0.000	55.16	0.033	52.29	0.027	22.63	0.008	41.58	0.015
7) むしろ	23.896	0.000	54.68	0.006	28.50	0.000	31.54	0.000	46.92	0.001
8) ながら[も]<逆接>	23.517	0.000	33.11	0.000	34.57	0.000	35.50	0.000	56.50	0.007
9) ため[に]<理由>	23.318	0.000	41.89	0.014	49.29	0.020	53.58	0.024	22.88	0.007
10) によって<根拠>	23.241	0.000	43.18	0.007	37.57	0.000	26.31	0.000	55.96	0.012
11) とともに<継起>	19.405	0.000	37.76	0.000	34.14	0.000	53.98	0.010	34.58	0.000
12) にとって	19.182	0.000	48.68	0.005	29.50	0.000	32.58	0.000	50.04	0.005
13) から<理由>	18.875	0.000	53.71	0.019	25.68	0.000	31.98	0.005	48.90	0.019
14) として	18.841	0.000	33.32	0.057	42.75	0.067	29.94	0.049	57.13	0.113
15) まま	17.677	0.001	38.50	0.004	35.75	0.000	31.71	0.000	55.33	0.008
16) つつ<同時>	17.010	0.001	36.00	0.000	40.86	0.000	33.58	0.000	52.46	0.005
17) によって<動作主体>	15.419	0.001	39.39	0.011	33.25	0.003	32.21	0.000	55.58	0.017
18) つつ<逆接>	15.188	0.002	38.00	0.000	38.00	0.000	38.00	0.000	48.13	0.005
19) によれば(根拠)	14.910	0.002	52.71	0.009	30.18	0.000	34.29	0.000	44.75	0.004
20) を通じて<媒介>	13.670	0.003	54.29	0.008	36.46	0.000	35.92	0.000	38.21	0.000
21) したがって	13.341	0.004	45.21	0.014	44.61	0.013	49.56	0.019	27.00	0.002
22) ながら<同時>	13.290	0.004	26.87	0.000	46.07	0.003	48.69	0.009	41.54	0.000
23) とは<定義>	12.691	0.005	36.55	0.000	34.86	0.000	35.71	0.000	53.40	0.009
24) つまり	12.182	0.007	43.26	0.007	26.75	0.000	37.25	0.000	51.27	0.013
25) なお	11.726	0.008	46.39	0.009	53.93	0.016	30.35	0.000	39.83	0.005
26) にもかかわらず	11.721	0.008	36.58	0.000	40.36	0.000	35.25	0.000	50.63	0.003

p<0.01. なお, 中央値とはジャンルの中での出現率の中央値である.

4.3.1 65 全変数の単変量的分布の比較

65 語句中, KW 検定結果が有意な 26 語句の資料別平均ランクをまとめたものを表 8 に示す. 4 つの分野の論文資料のみで単変量的に検討した表 8 の結果から, 論文という広く同一のジャンルに帰属していても, 専門分野によって語句の出現率に違いがあることがより明確となった. 次に 4.2 節の分析で選択された 19 語句とは別に, 論文ジャンルを分離するのに有効な語句があるかどうかを検討したい.

4.3.2 判別分析結果とその有効性

65 語句の出現率を説明変数としてステップワイズ法を用いた判別分析を行った結果, 逐次的に 8 個の説明変数が予測式に組み込まれ, その手続き内で削除された変数もなく, 4 つの分野の論文資料のジャンル判別に有効な, 以下の 8 つの変数(語句)が選択された.

①～によって/により/による N<方法> ②～によって/により/による N<理由> ③ただし ④～ながら[も]<逆接> ⑤～ため[に]<目的> ⑥～について/についての N ⑦むしろ ⑧～とともに<同時>

ここでは 3 つの判別関数が算出され, 検定の結果, 累積寄与率は関数 1 のみで 43.7%であった. 次に, 選択された 8 語句による判別関数平面での各資料の判別得点と文章資料グループの重心をプロットしたものを図 2 に示す.

図 2 を見るとわかるように, 上記 8 語句による分析では, 各論文資料はそれぞれ境界近くで何編かが重なっているのが見られるものの, 4 つのグループははっきり分離されていると言えよう. 2 次元の判別関数平面における各論文資料グループの重心の値は表 9 の通りである.

表 9 および図 2 から, 関数 1 は, 工学論文と物理学論文を経済学論文と文学論文から分離し, 関数 2 は, 経済学論文を物理学論文と文学論文から分離し, 関数 3 は, 工学論文を物理学

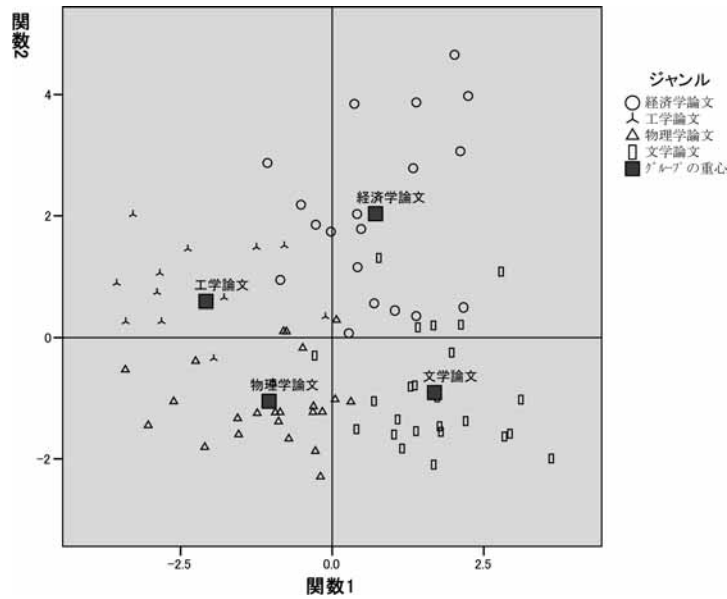


図 2. 判別分析による 4 分野の論文資料グループの重心および固体のプロット。

表 9. 判別空間における経済学・工学・物理学・文学の各論文資料の重心。

論文資料	関数 1	関数 2	関数 3
経済学論文	0.716	2.035	-0.659
工学論文	-2.081	0.590	1.579
物理学論文	-1.041	-1.049	-1.087
文学論文	1.688	-0.906	0.687

表 10. 選択された 8 語句の構造係数。

語句<代表項目>	関数1	関数2	関数3
1 ~によって<方法>	-.630(*)	0.074	0.256
2 むしろ	.320(*)	0.214	-0.048
3 ~ながら[も]<逆接>	.319(*)	-0.257	0.265
4 ただし	0.129	.574(*)	-0.091
5 ~について	0.247	.468(*)	-0.155
6 ~によって<理由>	-0.304	-0.167	-.667(*)
7 ~ため[に]<目的>	-0.378	0.379	.383(*)
8 ~とともに<同時>	0.116	-0.132	.290(*)

(*)有意な係数

論文から分離していることがわかる。

次に、構造係数と判別空間における各ジャンルの重心の関係から、選択された 8 語句がどの論文資料を分離するのに有効かを考察する。表 10 に 8 語句の構造係数を示す。

表 10 から工学論文と物理学論文を、経済学論文と文学論文から分離するのは、「~によって/による N <方法>」「むしろ」「~ながら[も]<逆接>」である。次に、経済学論文を物理学論文と文学論文から分離するのは、「ただし」「~について/についての N」である。工学論

文を物理学論文から分離するのは、「～によって/による N <理由>」「～ため [に] <目的>」「～とともに <同時>」であることがわかる。ここで、上記の分析結果を先の7つのジャンルによる分析結果と比較してみると、分析に有効だった項目のうち5項目(「～によって/による N <方法>」「ただし」「～について/についての N」「～によって/による N <理由>」「～ため [に] <目的>」)が共通しており、「むしろ」「～ながら [も] <逆接>」「～とともに <同時>」の3項目が、この分析によって新たに出てきたものであることがわかる。この結果を表1ならびに表8の単変量的な結果と比較すると、論文ジャンルの4つの分野の資料について以下のことが言えよう。

「～によって/による N <方法>」は工学論文と物理学論文でよく用いられ、とりわけ工学論文で頻度が高い。「～によって/により/による N <理由>」は、経済学論文、工学論文、物理学論文で多用され、特に物理学論文で頻度が高い。「～ため [に] <目的>」は工学論文で特によく用いられている。「～について/についての N」は経済学論文でとりわけ頻度が高く、「ただし」もよく用いられている。「むしろ」は経済学論文でよく用いられ、「～ながら [も] <逆接>」は文学論文でよく用いられている。なお、「～とともに <同時>」は単変量的な解析では $p=0.03$ であったので $p<0.01$ の語のみを示した表8には入っていないが、平均ランクは文学論文が一番高く4つの分野の論文の中では文学論文が一番用いられているのがわかる。

以上、一般的に「論文」という同一のジャンルに入れられている文章の中でも、専門分野別に見ていくと、指標語句の出現分布に違いがあることがわかった。

5. 総合考察

本研究の結果、経済学入門書、経済学論文、工学論文、物理学論文、文学論文、新聞社説、文学作品という7つのジャンルを対象に、接続語句・助詞相当句の出現率を指標として、文章の帰属ジャンルを判別することが可能であることが示された。また、論述的な文脈展開を持つジャンルの文章(経済学入門書、経済学論文、工学論文、物理学論文、文学論文)に共通する語句も抽出された。この結果は村田(2000, 2002)の結果を再確認するものであり、同時に接続語句・助詞相当句が文章のジャンル判別の指標となりうる可能性がより一層高くなったと考えられる。

ここで論述的な文章ジャンルに属すると考えられる4つの分野の論文と経済学入門書を中心に、松岡(1995)に基づき、論理展開に重要だと考えられる「並列(・同列)」「対比(・対立)」「因果(・呼応)」の表現について文章資料間で比較を行いたい¹⁴⁾。以下、KW検定結果を合わせて考察する。

「並列(・同列)」の表現としては、「～とは <定義>」は、経済学入門書と文学論文でよく用いられている。「～というのは <定義>」も経済学入門書で比較的用いられ、物理学論文、文学論文でも使われている。論述的な文章では定義づけはよく行われることを示していると言えよう。「すなわち」「つまり」は言い換え表現で、経済学入門書、経済学論文、工学論文、物理学論文、文学論文すべてにおいて、新聞社説、文学作品に比べてよく用いられていることがわかる。これも定義づけと同様、言い換え表現が論述的な文章ではよく用いられることを示していると言えよう。また、「すなわち」については村田(2000)と同様、「すなわち」の方が論術的形式の強いジャンルとそうではないジャンル間の差異をより明らかに示していて、経済学入門書、経済学論文、工学論文、物理学論文、文学論文で非常に多用されているのに対して、新聞社説、文学作品の資料中(社説222編と文学作品51編の合計273編)には新聞社説(毎日)で1回、現代文学作品で3回の合計4回しか用いられていない。例示を表す「たとえば」は、経済学入門書、経済学論文、工学論文、文学論文で多用されている。

「対比(・対立)」としては、「～のに対して<対比>」は、経済学入門書、経済学論文、工学論文、文学論文において平均ランク 239 を超え、物理学論文でも 220 を超えていて、論述的な文章グループと新聞社説、文学作品とは対照的である。「むしろ」は、特に経済学論文で多用され、次に文学論文が続く。経済学入門書、工学論文、物理学論文では、新聞社説、文学作品と同様、あまり用いられていない。「他方」は、経済学論文でよく用いられ、次に経済学入門書が続く。また、工学論文、物理学論文、文学論文のほうが新聞社説、文学作品に比べて相対的に多く用いられている。「一方[で]」は、文学論文で特に多用され、経済学入門書、経済学論文、工学論文、物理学論文でも平均ランクは 200 を超えており、論述的な文章グループと新聞社説、文学作品とは大きく異なる傾向が見られる。

「因果(・呼応)」の表現としては、「～によって/による N<理由>」「～ため[に]<理由>」が経済学入門書、経済学論文、工学論文、物理学論文、文学論文で多用されていることがわかる。「～ので<理由>」は、4つの分野の論文で多用されているほか、文学作品でも比較的よく用いられている。一方、経済学入門書では「～ので<理由>」はあまり使用されていない。文学作品を特徴付ける「理由・原因」の表現は、「～から<理由>」である。新聞社説では、「～によって/による N<理由>」「～ため[に]<理由>」「～ので<理由>」の3語句ともあまり用いられず、比較的「～から<理由>」が使用されていると言えよう。因果の中で帰結を表す「したがって」が、経済学入門書、経済学論文、工学論文、物理学論文で多用されている。「～[た]結果/Nの結果」も、経済学入門書、経済学論文、工学論文、物理学論文でよく用いられている。

このように論述的な文脈展開を持つ文章資料であると考えられる経済学入門書と論文の間でも語句の出現分布はそれぞれ異なり、すべての論文と経済学入門書に共通して多用される語句、論文に共通して多用される語句、一部の論文のみに多用される語句があることがわかった。上記の結果から、4つの分野の論文の中での語句の出現傾向の相違をまとめると、たとえば、文学論文では、条件を表す「～うえ[で]」、付帯状況を表す「～まま」「～つつ<同時>」「～とともに<同時>」、逆接を意味する「～ものの」「～にもかかわらず」「つつ<逆接>」「～ながらも<逆接>」などの語句の頻度が他の論文と比べて高くなっているが、経済学論文、工学論文、物理学論文では、付帯状況表現「～まま」「～つつ<同時>」、逆接表現「～ものの」「～にもかかわらず」の頻度は低く、逆に、帰結を表す「したがって」、方法を表す「～によって/による N<方法>」、理由・原因を表す「～ため[に]<理由>」の使用頻度が高くなっている。また、経済学論文については、上記のように文学論文に対して工学論文、物理学論文と同様の傾向を示す一方で、文学論文と共通する特徴的な語句も多く持つ。たとえば、提題を表す「～について/についての N」、根拠を表す「～によって/による N<根拠>」「～にもとづいて」、同じく出典などを示す根拠の表現「～よれば<根拠>」、例示を表す「たとえば」、比較を表す「むしろ」、条件を表す「～とすれば、～とすると、～としたら、～としても」である。一方、工学論文と物理学論文に特に共通して頻度の高い項目は、「～によって/による N<方法>」「～ため[に]<理由>」「～に対して<対象>」である。こうした結果は、各専門分野において論証を進める際の文脈展開の相違の一つの反映として捉えることもできよう。

文学作品と新聞社説のみを対象とした分析については紙幅の都合により詳細は省略するが、「～ながらも<同時>」が新聞社説と文学作品を分離するのに非常に有効な指標であることがわかった。これは村田(2000)の結果の再確認であり、今回文学作品については資料数を大幅に増やし現代作品を加えたが、接続語句・助詞相当句についてはあまり用いられていないことが改めて確認された。また「並列(・同列)」「対比(・対立)」「因果(・呼応)」の表現においては、論述的な文章グループと新聞社説、文学作品とが大きく異なる傾向を見せていることがわかった。

最後に本研究成果の日本語教育への応用という視点から教育方法の一つの可能性について触

りたい。専門日本語教育における論文読解・作成を目標とした教育方法としては、具体的に次のような方法が考えられよう。まず教材として用いる文章は学習者の専門分野の文章ジャンルに限定し(入門書、概説書、論文等)、初めに本研究で得られた高頻度の接続語句・助詞相当句を提示し、その意味用法を説明する。次に実際に学習者の専門分野の文章内でそれらの語句がどのように用いられているのかを提示し、その文章構造の中でそれらの語句の意味用法を意識化できるように指導する。さらにそれらの語句を用いた文章作成の練習をすることで、頻出する接続語句・助詞相当句語句の用法に関する記憶が強化されると言えよう。その結果、学習者が専門分野の新しい文献に触れたときに、それらの語句の用法を表現形式(パターン)として認識できるようになると考えられる。最初は読解を通じてこのようなパターンの習得を進め、最終段階として学習者自身がレポートや論文を書く際にこれらのパターンが活用できるようになれば、専門日本語教育の教育目標は達成されたと言えよう。また、ある程度学習が進んだところで、新聞社説や文学作品の文章を提示し、文章ジャンルが異なると頻出する接続語句・助詞相当句が異なることを示すことも学習者にとっては有用な情報になると推測される。今後は本研究の成果を踏まえた教材の開発を進めるとともに、分析対象をさらに広げて検討する必要があると考えられる。

本論文は学位論文(村田, 2006)の一部を加筆訂正したものである。なお、本研究は文部科学省科学研究費基盤研究C(課題番号17520354, 研究代表者 村田年)の一部として行われたものである。

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り、同研究所の前田忠彦准教授からご助言をいただきました。また、査読委員の方々には貴重なご意見を多々いただきました。深く感謝いたします。

注.

- 1) 黒橋禎夫・長尾真(1992)の表現を借りた。
- 2) 残された課題としては、対象資料について経済学分野は入門書だけで経済学論文が含まれていないこと、また文学作品についても現代文学作品が含まれていない等のコメントが論文発表後に寄せられた。指標項目についての課題は注⁶⁾参照。
- 3) 文章資料の選択方法については、新聞社説と文学論文が無作為性が高くなっているがほかは有意抽出となっている。これは現存する文章資料すべてを母集団とすることが不可能であるという理由以外に、専門日本語教育における教材としての利用を考えた時、学習者のニーズに即した文章資料を対象とすることが教育上、有意義だと考えたためである。その意味では有意抽出に意味はあるが、この調査対象から得られた結果の過度な一般化は慎まなければならない。なお、文章資料については、7つの資料を、論理が明示的だと考えられるグループ(A~E)と論理が非明示的だと考えられるグループ(F, G)の二つに大別し、各グループ内でアイウエオ順に提示した。
- 4) 今回の調査では資料を小説体のものに絞り、従来利用してきた近代文学作品から日記体、随筆、古語が多く含まれる作品は除いた。
- 5) 教科書を一つのジャンルとして立てたのは、ペトルス・ラムスがつくりだした教科書というジャンルのパラダイムに拠る。
- 6) 村田(2000, 2002)では指標として62語句を用いたが、その際に残された課題を踏まえ、「つつ」「とともに」「にしたがって」について、それぞれ「つつ」は「同時」と「逆接」、

「とともに」は「同時」と「継起」, 「にしたがって」は「根拠」と「継起」というように意味を分化させ, 指標語句を65語句とした。

- 7) 解析用プログラムとしてはSPSS14.0Jを用いた。
- 8) この用語は柳井・高木(1986)による。本研究で用いる正準判別分析は, 解の導出自体に分布の仮定は入っていない。
- 9) 本研究で用いる変数は, いずれも特定の長さの文章中における語句の出現率であり, 正規分布を仮定していない。
- 10) 平均値 $+0.5SD$ に相当する値を便宜的な基準として(個体数370に関するランクの総平均値185.5, 標準偏差106.81), 当該グループの平均ランク239以上を特徴的語句と見なす。
- 11) 判別関数は, 変数の数 p , 群の数 g としたとき, $\text{Min}(p, g-1)$ だけ算出されるので, ここでは判別関数の数は6つとなる。
- 12) ウィルクスの Λ の定義は柳井・高木(1986)による。
- 13) 各文章資料の数が異なるため, 分析の際には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。
- 14) 松岡(1995)は論理的な文章の基本的な組み立ての型(<論文>型)として「並列(・同列)」「対比(・対立)」「因果(・呼応)」の三種を提起し, これらが序論・本論・結論の組み立て, あるいは本論内部の各段落の組み立てにおいて, 書き手に明白に意識されることの必要性を述べている。

参 考 文 献

- Ong, Walter J. (1991). 『声の文化と文字の文化』(桜井直文, 林 正寛, 糟谷啓介 訳), 藤原書店, 東京.
- 金 明哲(1999). 日本現代文における書き手の特徴情報, 人文学と情報処理, 20, 64-71.
- 黒橋禎夫, 長尾 眞(1992). 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, 1(1), 3-20.
- 松岡 弘(1995). 文学的文章をモデルにした文章の論理的構成の指導について, 『阪田雪子先生古稀記念論文集 日本語と日本語教育』, 三省堂, 東京.
- 村田 年(2000). 多変量解析による文章の所属ジャンルの判別—論理展開を支える接続語句・助詞相当句を指標として—, 統計数理, 48(2), 311-326.
- 村田 年(2002). 論理展開を支える機能語句—接続語句, 助詞相当句による文章のジャンル判別を通じて—, 計量国語学, 23(4), 186-206.
- 村田 年(2006). 論述文の文脈展開を支える文型の計量分析とその日本語教育への応用, 学位論文, 慶應義塾大学 文学研究科.
- 柳井晴夫, 高木廣文(編)(1986). 『多変量解析ハンドブック』, 現代数学社, 京都.

Teaching the Reading and Writing of Technical Papers in Japanese:
A Study of Selected Conjunctive Words and Particle-phrases
in Expository Writings

Minori Murata

Center for Japanese Studies, Keio University

The ultimate goal for most students learning Japanese who intend to perform research in their own fields is to gain the ability to read and write technical papers. In order to achieve this, it is necessary to understand the contextual development of the text in a technical paper in Japanese. Conjunctive words and particle-phrases appearing at the surface level of expression in a text might be considered as providing important clues for understanding the structure of the text. This study aims to shed light on conjunctive words and particle-phrases that are commonly used in academic papers from different fields. As a step toward achieving this goal, 370 text samples from seven different sources were chosen as data. Those seven sources were: (1) pedagogical economics textbook; (2) papers from Economics; (3) papers from Science and Technology; (4) papers from the Journal of the Physical Society of Japan; (5) papers from Japanese Literature; (6) editorial articles; and (7) modern and contemporaneous novels. The frequencies of occurrences (per sentence) of the 65 selected conjunctive words and particle-phrases from each sample were analysed in the following two steps,

(a) An examination of the univariate distribution of the 65 conjunctive words and particle-phrases followed by a canonical discriminant analysis of 370 text samples.

(b) The same method repeated for the four genres of papers (2) (3) (4) and (5).

From these results, it was found that the various texts could be classified into their genres correctly at a high discriminant rate (84.6%) by using 19 conjunctive words and particle-phrases. The results also highlighted (i) the important conjunctive words and particle-phrases through which the texts could be discriminated into their genres; and (ii) the significant conjunctive words and particle-phrases that are commonly used in expository writings. These results revealed that conjunctive words and particle-phrases are an important indicator in distinguishing different textual genre. This study is expected to contribute to the improvement of teaching methods for Japanese for specific purposes, as it will enable us to provide a set of more objective basic data for the development of teaching materials.