Information Processing Society of Japan

自然言語処理 128－13
（1998. 11. 5）

新聞記事における文章末表現での名詞語彙の出現特性

中村 隆志　　本間 愛
tks@human.ge.niigata-u.ac.jp　　a-honma@dolphin.ge.niigata-u.ac.jp
新潟大学人文学部

新聞記事の文章末に着目し、文書の文字数が大きくなるほど、文章末に出現する頻度が高くなるようなタームが存在するかどうかを調査する。資料として、新聞記事データベースからランダムに選んだキーワード検索を用いて得られる文書群のうち、文書数が 300 を越える文書群を 10 個取り上げる。ある文書群において、文書の文字数が増加するにつれ、文章末での出現頻度が増加する名詞を、その文書群の”大河タ－ム”という造語で呼称する。キーワード検索により得られた文書群から大河タームを抽出した結果、一つの例外を除き、取り上げた全ての文書群から複数の大河タームが取り出される。大河タームの 80 %以上は、９個の文書群の最頻出名詞と違うものであることが確認される。

キーワード：新聞記事、テキスト分類、文章末表現

Frequency of noun usage in closing sentences in newspaper articles

Takashi Nakamura　　Ai Honma
tks@human.ge.niigata-u.ac.jp　　a-honma@dolphin.ge.niigata-u.ac.jp
Department of Humanities, Niigata University

We focus on the frequency of noun usage in the closing sentences of a group of newspaper articles. We investigate the existence of terms whose usage frequency in the closing sentences increases as the article length increases. Data are compiled from 10 groups of newspaper articles obtained from the Nikkei Newspaper Database '97 using 10 keywords randomly which include at least 300 articles. The nouns are called "taiga terms" whose usage frequency in the closing sentences of each article increases as the article length increases ("taiga" means 'big river' in Japanese). Several "taiga terms" were extracted from nine of the article groups. More than 80% of "taiga terms" differ from the most frequently used nouns which appear in the 9 groups of articles.

keywords: newspaper articles, text categorization, closing sentences

NII-Electronic Library Service

Information Processing Society of Japan

1．序に変えて

近年の情報分類、情報加工への需要の高まりに伴い、多くの研究が進んでいるが、その多くは、過剰な文書群に対する閲覧効率を上げるためのものである。自動要約、重要文抽出（例えば、文献[1,2,3,4]）は、いずれも文章の圧縮を目標としている。また、文章融合の研究[5]においては、文章の圧縮だけでなく、閲覧時の文章間の移動を減少させることにより、情報取得の効率を上げることを目指している。これらは、情報圧縮を実現するために、個別の文書あるいは非常に近縁な複数文書を操作対象とする。

また、検索時に得られる文書群を横断的に整理分類する試みも為されている。例えば、数値情報を用いた情報抽出の試み[6]、あるいは固有名詞を用いた整理分類[7]の提案などが挙げられる。これらは個別の文書や近縁な複数文書を対象とせず、文書群の全ての文書を対象として情報取得の効率化を目指す。本稿は、これら横断的整理分類の一方法を提案するための予備調査と位置づけられる。ここで整理分類のために着目するのが文章末での名詞語彙である。

では、情報分類、情報加工において、文章末表現はどう扱われているのだろうか。

新聞記事の主題推定のためにテキスト構造を利用する可能性に触れた文献[8]においては、新聞記事の末尾は主題と関連性が小さいと推察されている。ここでは、主題推定をテキスト文だけから見出し語を当てることと規定した上で調査が為された。さらに、その見出し語の大半が、テキスト文の冒頭から数十単語の間に出現することが統計的に確かめられた。この結果が、主題と新聞記事の末尾との関連性の希薄さを推察する根拠となっている。

一方で、重要文抽出において、抽出派でヒューリスティック型と位置づけされる日本語文章要約システム GREEN[9]を取り上げる。GREEN では要約文選択アルゴリズムの開始ステップにおいて、文章の最終文（及び冒頭文）を要約文として採用する。この根拠として、文章の最終文は著者が話を締め括る必要があるため、文章の他の部分と異なった、何らかの特別な意図を持って書かれた文であり、要約作成時にも欠かせない重要な文であるため、としている。

主題と関連性が小さいことと、文章において先験的に重要な文であると位置づけすることは、必ずしも矛盾することではない。しかし、最終文あるいは文章末表現については、一定の見解がないように思われる。本稿では、文章末表現に用いられる名詞語彙に着目し、文書群別に特徴ある語彙を抽出する方法を提案する。使用するのは、文書長別での、文章末における名詞語彙の出現頻度である。調査対象は新聞記事に限定する。上で述べた横断的整理分類に応用するには到らないまでも、文書群から特徴ある名詞語彙を抽出する方法の提案として本稿を位置づける。

2：文章作成指南書における文章末表現

文章作成指南書は受験参考書の小論文対策問題集に始まり、実に多くの書物が出回っているが、それらを通じて言えるのは「他人が読む価値があるものをつくる」という心構えと作法を説いていることである。これらを大きく分類すると、一つはいわゆる実用的文章、すなわち、レポート、報告書、論文の書き方を指導するものと、もう一つは実用的でない文章、すなわち、小説、エッセイなど、客観性にこだわらず、「うまい」文章の書き方を指導するものに分けられる。前者と後者を比較するならば、後者の指導を会得する方が前者のものより遙かに難しく、そもそも前者のものを会得していることが前提となっている。前者では、明瞭明晰を最善のものとし、趣旨の通じ易さを最大に重んじる。後者では「名文の書き方」という言葉に代表されるように[10]、面白く、読む人を飽きさせず、引き込み続けることを最大に重んじる。この違いは文章全体の構成の仕方にも現れる。

実用的文章指南書の一例として文献[11]を挙げる。ここではワンワード／ワンミーニング、ワンセンテンス／ワンアイディア、ワンパラグラフ／ワントピックの思想が徹底されており、細部に到るまで明晰さを重んじる。パラグラフ内では総論から各論へという作法が述べられ、パラグラフ内の文の順序は総論を先頭とし、各論については、読み手の関心をより多く引くものから順に並べていくことを指南する。２つの例外（総論の繰り返しによる説得法とパラグラフ内で因果関係を完結

NII-Electronic Library Service

Information Processing Society of Japan

的に記述する方法）を除き、パラグラフの締め括り方への言及はない。また、文章全体の構成つまりパラグラフの連結については、序論、本論、結論の構成を推奨する。結論を述べるべき最終パラグラフも、一つのパラグラフに過ぎない。もしも、パラグラフ内における順序立て、すなわち、総論から各論へ、各論は関心を引くもの順に、という指南に乗っ取るならば、文章の締め括りは、結論パラグラフ内の、相対的に最も関心が低い各論の一つが述べられることになる。

実用的でない文章の指南書の一例として、文献[12]を挙げる。ここでは、文章全体の構成として、「序論、本論、結論」という構成よりも、「起承転結」の構成を推奨している。実例を挙げて、「転」の善し悪しを論じているが、「転」だけのための議論ではない。「転」を持ち込むことは、文章構成に流れを与えるためであり、同時に「結」の部分で読者を面白がらせるためである。文章の締め括り方に具体的指南はないものの、読み手の関心を最後まで引き続け、締め括りの部分で面白がらせることを良しとする思想がある。

文献[13]では、文章の締め括り方の一例を抽象的に表した。良い終わり方は「小川が大河を突っ切る」（図１左）、悪い終わり方は「小川が大河に飲み込まれる」と表現する（図１右）。すなわち、小川が文章を締め括る前に書いてきた各の記述の意味、あるいはトピックを表し、大河はそこで書かれた内容を包括する上位概念となる。良い終わり方か悪い終わり方かは問わないとしても、ここで大河という言葉で抽象的に表される上位概念は、文章の締め括り方を考察していく中で、重要な役割を果たすと考えられる。

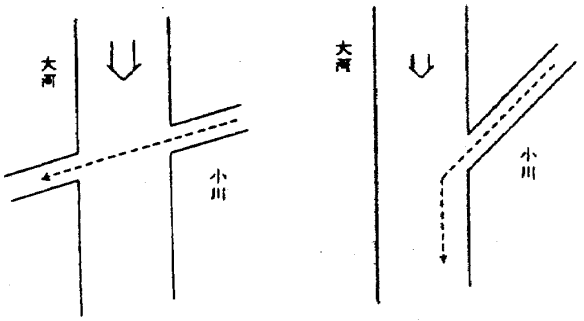

図１．「小川と大河」の概念図

以上見てきたように、実用的文章作成指南書、実用的でない文章作成指南書とも、読み手に読む価値を与えるという点では一致しているものの、具体的構成法においては明確な違いがある。文章の終わり方に統一した見解は現在のところ、見あたない。その中で、上述の「小川と大河」の対比は、実用的でない文章の終わり方を考察する上で大いに示唆的である。本研究の目論見は、この大河という言葉で抽象されるタームを文書群から如何に抽出するかの一方法を提案することにある。

なお、本章及び以下の章において実用的／実用的でないという言葉を用いて、指南書を分類しているが、役に立つ／役に立たない、あるいは意味がある／意味がないなどの分け方をしているのではない。実用的というタームはレポート、報告書、論文などを総括した分類上の用語であり、実際の文献の中で使用されていることを申し添えておく。

３．新聞記事の文書群からの抽出方針

１章で述べた、新聞記事における主題推定から得られた推察と GREEN での最終文の扱いとの対比と、２章で述べた実用的文章作成指南と実用的でない文章作成指南との対比の関連について触れておく。総論から各論へ、各論の登場順位は関心の高いものから、という方針を守って作成された文章ならば、主題となる言葉が文章の最初の方に出現することは明らかである。締め括りの部分での面白さ、あるいは大河という言葉で抽象されるような上位概念を用いて締め括られた文章であるならば、何らかの特別な意図を持って書かれた文として、最終文をアプリオリに重要文とみなすことは必然的である。

本稿では、新聞記事データベースから得られた文章群を題材として、大河という言葉で抽象される上位概念を抽出することを試みる。抽出のために一つの仮説を立てる。それは文章群をマクロ的に捉えた場合、文字数の短い文章ほど、実用的文章作成の方針に乗っ取って書かれたものが多く、逆に文字数の長い文章ほど、実用的文章の構成から外れ、締

NII-Electronic Library Service

Information Processing Society of Japan

め括りにその文章のテーマの上位概念を使用する頻度が高くなる、というものである。その根拠としては、文字数が200字程度、あるいはセンテンスの数が4～5程度の文章では起承転結の構成をとることは実質不可能であり、締め括る必要が発生するほどの情報量があるとは考えにくく、この場合は実用的文章の構成をとることが多くなると考えられるからである。一方で、文字数が数千程度になると、空間的余地ができ、文章に流れをもたらすことが出来る。また、逆に数千文字程度の長さになると、関心の低い順に各論を述べて終わるような文章では、最後まで読まれない可能性がある。よって、これらの大きさの文章では、締め括りで読者を面白がらせる文章が相対的に多くなると考えられる。起承転結の構成に関しては、今回の調査では対象とせず、文章の締め括りに使われる言葉だけに対象を絞った。

4．大河タームの抽出方法

前章の仮説に乗っ取り、文書群の中から文字数の増加に応じて、文章末に出現する頻度が高くなる名詞を取り出すことが本稿の趣旨である。以下で述べる方法で抽出された名詞を、文献[13]で用いられた言葉を用いて、"大河ターム"という造語で呼ぶことにする。以下に調査手順を述べる。

0．調査対象は新聞記事データベースを選び、本報告では日本経済新聞 CD-ROM、97年版[14]を使用した。

1．シソーラスを用いたキーワード検索を行い、検索結果の文書数が300を越えるキーワードを選び、得られた文書をまとめて、一つの文書群とする。

2．検索して得られた文書群のうち、センテンス数が4未満のもの、あるいは文章でなく箇条書きに終始するものを除外する。

3．声明で記事が終わる場合は、発言内外を問わず、句点から句点まで、あるいは鉤括弧までを1つのセンテンスとする。

4．文書群内の全ての文書を文字数順に並べ、ほぼ同数になるようにX個のブロックに分割する。平均文字数が最小のブロックを第0ブロックとし、昇順に順序付けを行い、平均文字数最大のブロックを第$(X-1)$ブロックとする。

5．各文書群において、全ての文書の末尾のkセンテンスを取り出し、これらに対して、日本語形態素解析プログラム「茶筌 1.0」を用いて形態素解析を行う[15]。この中から普通名詞、固有名詞、サ変名詞、時相名詞を取り出し、各名詞毎に集計を行う。

6．ある文書群jにおいて、第xブロック内の全ての文書の末尾kセンテンスにおける、ある名詞nの出現回数を$F(j,n,x)$とする。$F(j,n,x)$の最大値を$Fmax(j,n)$とする。文書群 j 内の全文書数をA(j)とする。この時、$Fmax(j,n') \geqq A(j)/X/10$ を満たさない場合、名詞n'を手順7,8で行う解析から除外する。

7．頻度分布$(x,\ F(j,n,x))$に対して、単純回帰分析を行い、回帰係数を求める。さらに、回帰係数の検定を行い、回帰係数の値が0である帰無仮説を有意水準 0.05 で棄却できる名詞のみを抽出する。また、得られた回帰係数Eが$E > t$を満たすことを抽出の条件とする。

8．頻度分布$(x,\ F(j,n,x))$に対して、歪み度$V(n)$を求める。手順7で抽出された名詞の内、$V(n) < 0$ を満たすもののみを抽出する。

以上の手順により、文書群から抽出された名詞の集合をその文書群の大河タームと呼び、T-Tと表す。各手順について解説する。

手順1：各の大河タームが末尾kセンテンスに出現する頻度を大きくするために、文書数が大きい文書群を選ばなければならない。300という数値はヒューリスティックな判断である。

手順2：センテンス数が4未満のものは、起承転結の構成を作りようがなく、締め括る必要があるほど多くの情報量に満ちていることは考えられない。また、今回の調査が末尾での文章の締め括りの調査であるため、箇条書きの文書は対象外とする。

手順3：鉤括弧を一つの特別なセンテンスとせず、鉤括弧内のセンテンスも、鉤括弧外のものと同様のセンテンスとする。

手順4：手順6以降の頻度分布を得るため、文書群を文字数順に並べて分割する。割り切れなくてもそのまま用いることにする。

手順5：解析対象として末尾の表現を取り出すにあたって、kセンテンスにした理由は、

NII-Electronic Library Service

Information Processing Society of Japan

KW	CT	T-T	E	?	TF	PF	A	OTF	OPF
環境問題	5 / 24 (0.35)	#企業	2.9	-1.19	538	85	618	環境 1892	環境 302
		社会	2.9	-2.70	170	28		問題 1021	問題 161
		コスト	2.6	-1.23	132	28	234.8	市 639	市 87
		経済	1.9	-1.58	251	36	387.0	処理 571	#企業 85
		開発	1.8	-1.15	255	38	555.1	#企業 538	日本 73
							858.1	日本 486	会議 71
							1491.7	地球 485	排出 68
								廃棄物 441	事業 59
								会議 398	県 57
								排出 382	地球 54
原子力	6 / 47 (0.13)	技術	2.6	-2.04	203	33	369	原子力 800	原発 111
		%問題	2.4	-1.27	217	41		原発 569	原子力 101
		#開発	2.0	-0.41	300	51	226.9	炉 386	事故 70
		政策	1.6	-2.08	107	20	327.6	事故 359	炉 60
		日本	1.5	-0.23	217	37	440.1	電力 319	電力 53
		可能性	1.4	-2.47	61	19	592.1	計画 302	#開発 51
							1213.9	#開発 300	計画 48
								発電所 242	施設 47
								施設 238	%問題 41
								核燃料 227	会 39
医療制度	11 / 54 (0.20)	#改革	5.4	-1.10	706	109	545	医療 2797	医療 374
		#負担	4.4	-0.61	992	150		病院 1185	病院 180
		企業	2.9	-1.24	214	29	252.9	#負担 992	#負担 150
		問題	2.6	-1.37	191	31	385.9	保険 967	患者 148
		国民	2.4	-1.11	309	43	537.4	患者 967	診療 124
		党	2.2	-0.95	216	38	781.2	#制度 814	保険 119
		抜本	2.2	-0.82	130	28	1523.6	診療 739	#制度 113
		健保	1.9	-0.18	234	39		#改革 706	#改革 109
		#制度	1.6	-0.45	814	113		高齢 548	報酬 81
		与党	1.2	-0.88	264	38		薬 537	医師 78
		厚生省	1.2	-0.62	328	49			
地震	3 / 30 (0.1)	%対策	2.9	-0.93	265	66	506	地震 1700	震度 268
		人	2.4	-2.51	126	23		震度 564	地震 247
		予知	1.9	-1.14	229	37	214.3	市 487	市 106
							332.0	防災 366	防災 76
							442.4	震災 363	被害 69
							611.9	被害 349	%対策 66
							1138.3	阪神 328	災害 62
								災害 324	各地 60
								観測 324	次 58
								大震災 324	計画 50
ワールド カップ	2 / 26 (0.08)	予選	2.5	-1.68	349	42	583	日本 1027	日本 188
		チーム	1.6	-0.55	343	47		杯 892	杯 115
							215.9	サッカー 553	戦 91
							297.1	戦 518	選手 76
							379.2	カップ 507	出場 62
							538.5	選手 494	大会 61
							936.1	ワールド 465	試合 59
								大会 438	監督 58
								監督 390	サッカー 53
								試合 356	世界 52

NII-Electronic Library Service

Information Processing Society of Japan

KW	CT	T-T	E	?	TF	PF	A	OTF		OPF	
汚職	9 / 56 (0.16)	部	3.6	-0.78	213	48	362	容疑者	843	被告	198
		特捜	2.7	-0.47	131	31		被告	824	容疑者	184
		医療	2.0	-1.67	147	24	269.8	事件	493	事件	72
		経営	1.7	-4.24	77	17	380.5	収賄	342	わいろ	60
		浜	1.7	-1.21	86	21	496.3	逮捕	297	工事	58
		関係者	1.7	-1.54	123	19	638.3	容疑	285	判決	55
		政治	1.6	-0.93	125	18	1076.9	汚職	284	業者	53
		病院	1.5	-1.32	128	32		会社	280	側	51
		判断	1.4	-0.91	71	19		贈賄	277	部	48
								市	275	収賄	48
就職	5 / 28 (0.18)	#学生	5.2	-0.68	1107	132	575	就職	1646	企業	236
		日本	2.7	-2.68	234	32		企業	1597	採用	204
		自分	2.7	-2.01	188	28	256.2	採用	1389	就職	185
		制度	2.3	-1.35	216	44	380.3	#学生	1107	#学生	132
		仕事	1.7	-0.72	241	40	506.9	会社	577	会	71
							804.3	活動	568	会社	60
							1565.0	大学	565	大学	56
								会	461	活動	55
								協定	418	雇用	49
								内定	403	協定	48
経済動向	6 / 58 (0.10)	#日本	5.8	-1.58	673	86	493	景気	1505	景気	193
		#経済	3.2	-0.85	954	106		#経済	954	#経済	106
		構造	2.7	-1.84	269	43	261.1	企業	787	回復	105
		改革	2.6	-2.25	254	29	449.5	回復	723	企業	102
		資金	1.9	-1.41	227	32	709.8	#日本	673	#日本	86
		減税	1.2	-1.22	177	34	957.1	消費	551	消費	65
							1760.5	金融	448	指数	62
								判断	438	マイナス	59
								消費税	422	判断	59
								市場	404	製造業	53
国際協力	4 / 35 (0.11)	中国	3.2	-0.83	417	58	511	日本	1333	日本	233
		*アジア	2.3	-1.14	420	53		協力	767	協力	137
		関係	1.1	-0.48	288	59	274.6	政府	761	政府	125
		開発	1.1	-0.51	104	35	414.3	経済	718	支援	125
							561.7	支援	614	経済	96
							767.8	国際	547	問題	96
							1397.9	援助	498	国際	91
								米	487	援助	85
								問題	484	北朝鮮	79
								*アジア	420	米	69
国際紛争	0 / 11 (0.0)						437	米	515	米	94
								戦争	452	政府	59
							256.0	日本	419	問題	56
							396.5	政府	296	大統領	54
							526.4	問題	288	日本	54
							713.6	大統領	261	戦争	49
							1326.4	中国	246	首相	31
								経済	181	経済	29
								世界	178	中国	29
								韓国	177	軍	28

表１．10個の文書群の調査結果

NII-Electronic Library Service

Information Processing Society of Japan

注意）ＫＷ：文書群の検索キーワード、ＣＴ：文書群 j において、手順６までの名詞の採否条件 $Fmax(j,n') \geq A/X/10$ を満たす名詞の個数と、文書群 j での大河タームの個数の比、T-T：大河タームの頻度分布 $(x, F(j,n,x))$ の回帰係数、？０：頻度分布 $(x, F(j,n,x))$ の歪み度、ＴＦ：大河タームの文書群での総出現回数、ＰＦ：大河タームの文章の末尾 k センテンスにおける文書群での総出現回数、Ａ：各文書群における総文書数と第０ブロックから第４ブロックまでの平均文字数、ＯＴＦ：文書群内の全名詞（但し、普通名詞、固有名詞、時相名詞、サ変名詞）の出現頻度順位表、１０位まで、ＯＰＦ：文章の末尾 k センテンスにおける文書群内の全名詞（但し、普通名詞、固有名詞、時相名詞、サ変名詞）の出現頻度順位表、１０位まで

なお、#付きの言葉は大河タームでかつ OTF、OPF の両方に、*付きの言葉は大河タームでかつ OTF のみ、%付きの言葉は大河タームでかつ OPF のみに載るもの。

１センテンスだとデータが少ないからであり、また、文章を締め括る機能が１センテンスだけに集約されていると限定されているよりは、k センテンスに拡がっていることが多いと推測するからである。名詞は具体的対象、抽象的表象が存在するもの、及び時間的表現をあらわすものに限定する。よって、形式名詞、数詞、副詞的名詞は除外する。なお、本章及び以下の章で述べる名詞とは、日本語形態素解析プログラム茶筌 1.0 の解析結果から求められた形態素を表す。

手順６：ここでは、最大値 $Fmax(j,n')$ の大きさを採否の条件とする。$A(j)/X/10$ を敷居値として採用するのはヒューリスティックな判断である。

手順７：単純回帰分析で得られた係数を意味づける指標として、回帰係数の値が０である帰無仮説を棄却することの他に、決定係数の値があるが、ここでは採用しない。頻度分布の変化は、例えば級数的変化でも良く、必ずしも、線形回帰直線で近似される必要はない。増加傾向にあるか、減少傾向にあるかの意味づけを必要とするため、回帰係数がゼロである帰無仮説を棄却するものだけを抽出する。ただし、回帰係数の値が小さい場合は、抽出から除外する。

手順８：頻度分布 $(x, F(j,n,x))$ の増加傾向を裏付けするための指標として、歪み度を用いる。歪み度が負の値をとる時、頻度分布が横軸正方向に偏っていることを示す性質を利用している。手順７での抽出を裏付けするための指標である。

５．抽出結果

今回の調査では、シソーラスを利用したキーワード検索で 10 個の文書群から大河タームの抽出を行った。10 個の文章群のためのキーワードはランダムに選んだ。４章で述べた手順の中のパラメータは X=5, k=2, t=1.0 を用いた。なお、比較対照のため、各文章群全体の最頻名詞を上位 10 単語、各文章群の全ての文書の末尾ｋセンテンスの集計したものから、最頻名詞を上位 10 単語を挙げることにする。キーワードはランダムに選んだ。表では順に、環境問題、原子力、医療制度、地震、ワールドカップ、汚職、就職、経済動向、国際協力、国際紛争である。

調査結果を表１に示す。得られた大河タームのうち、表１でのＯＴＦの１０単語に登らないもの、つまり#、*マークが付かないものの総計と大河タームの総計の比を求めると、42/51 ～ 0.82 が得られた。また、表１でのＯＰＦの１０単語に登らないもの、つまり#、%マークが付かないものの総計と大河タームの総計の比を求めると、41/51 ～ 0.80 が得られた。

６．結論

新聞記事においては、短いものほど実用的文章の構成に乗っ取りやすく、逆に長いものほど締め括るための言葉を使用しやすい、つまり、文書の文字数が大きくなるほど、文章末に大河タームが出現しやすいという仮説を取り入れて文書群を調査した。文章末尾２センテンスの出現頻度、出現頻度分布の回帰係数、歪み度を用いて、抽出を試みた結果、10個の文書群のうち、９個から大河タームが取り出された。大河タームの 80%以上は文書群全体の最頻出名詞順位表の上位の者とも、文書の末尾のみから計量した最頻出名詞順位

NII-Electronic Library Service

Information Processing Society of Japan

表のものとも違う。すなわち、長い文章の末尾に特異的に出現しやすい一方で、文章全体での最頻出の名詞と、あるいは文章末２センテンスにおける最頻出の名詞と比べれば、出現回数が少ないものが多い。今回は仮説の検証をするには到らなかった。が、あるテーマについて述べられた文章において、内容が多くなればなるほど、締め括りのために使われやすい特有のタームがそのテーマに存在することが多い、ということを今回の調査結果は示唆している。

７．課題

今後の課題を展望しておく。

１：最頻出名詞と大河タームが、シソーラス上でどのような位置関係を占めるのか、単純な上位概念なのか、それとも違った様相を占めるのかを解析する。

２：さらに多くの文書群を調査する。

３：大河タームを用いている文書の構成、あるいは論理の流れを、用いてない文書と比較する。

４：大河タームを用いて締め括る文書と使わずに締め括る文書との読後感の違いを被験者を用いた評価実験によって数値化する。

謝辞

日本語形態素解析システム茶筌を開発し、精度向上に取り組み続けている奈良先端科学技術大学院大学松本研究室の皆様に謝意を申し上げます。また、調査を許可下さった日本経済新聞社殿に感謝いたします。

参考文献

[1] 田村俊哉、田村直良：文章の表現形式に基づいた要約文章の生成について、情報処理学会報告、NL92-1,1992.

[2] 船坂貴浩、山本和英、増山繁：冗長度削減による関連新聞記事の要約、情報処理学会報告、NL114-7,1996.

[3] 野本忠司、松本裕治：人間の重要文判定に基づいた自動要約の試み、情報処理学会、NL120-11,1997.

[4] 任福継、定永靖史：統計情報と文章構造特徴に基づく重要文の自動抽出、情報処理学会報告、NL125-7,1998.

[5] 柴田昇吾、上田隆也、池田裕治：複数文章の融合、情報処理学会報告、NL120-12.

[6] 斉藤公一他：数値情報をキーとした新聞記事からの情報抽出、情報処理学会報告、NL125-6,1998.

[7] 増田恵子、梅村恭司：固有名詞に着目し記事群を整理分類し提供するシステム、情報処理学会報告、NL114-2.

[8] 野本忠司、松本裕治：テキスト構造を利用した主題の推定について、情報処理学会報告、NL114-8,1996.

[9] 山本和英、増山繁、内藤昭三：文章内構造を複合的に利用した論説文要約システムGREEN、自然言語処理, vol. 2(1), pp. 39-55.

[10] 丸谷才一：文章読本、中公文庫、1995.

[11] 篠田義明：コミュニケーション技術、中公新書 807、1986.

[12] 向井敏：文章読本、文藝春秋、1988.

[13] 加藤典洋：言語表現法講義、岩波書店、1996.

[14] 日本経済新聞社、日本経済新聞 97 年 CD-ROM 版、日本経済新聞社、1997.

[15] 松本裕治、北内啓、山下達雄、平野喜隆、今一修、今村友明：日本語形態素解析システム『茶筌』version 1.0 使用説明書、Information Science Technical Report, NAIST-IS-TR97007, 奈良先端科学技術大学、1997.

NII-Electronic Library Service