文章末の文だけが持つ特異な性質に由来している。文章中の文では，例え曖昧な表現に出会ったとしても，後続する文によって意味の補填がなされる可能性が常にある。書き手の立場からすれば，曖昧な表現を用いながら文章作成を進めていくことができる。しかし，文章末の文には後続する文がない。後続する文がない以上，文章末を作成する書き手は後続表現による意味の補填が必要な表現を可能な限り排除せねばならない。その一方で，先行する文が曖昧性を持つ場合は大いにある。従って，文章末の文では，必要に応じて先行する文の曖昧性を補填しながら，自身には補填される必要が小さい表現が使われやすいと考えられる。

言葉の意味が一般に多様であることから，言葉が多くなればなるほど，換言すれば，文章の長さが長くなればなるほど，その文章のもつ曖昧性は大きくなりやすい。従って，短い文章と長い文章を比較すれば，先行する文の曖昧性を補填しつつ自身に曖昧性の少ない表現が文章末で使われる頻度は，相対的に高くなると考えられる。さらに，文章のもつ曖昧性は単語の意味や文意に留まらず，文脈や大意に及ぶ。よって，先行する文脈や大意の曖昧性を解消しつつ，自身に曖昧性の少ない表現が文章末で使われる頻度は，長い文章の方が相対的に高いと考えられる。

野本・松本[8]は新聞記事における主題の推定について，テキスト構造を利用することで，推定の精度が向上することを見いだしている。論考の中で，本文の冒頭から特定の単語数からなるブロックを切り出して推定に用いる方法（FLM方式）の有効性を示すとともに，推定精度そのものは本文の長さに応じて下がることが付せて述べられている。このことは，文章の長さが構成，文脈，内容の展開に少なからず影響していることを示唆する。野本・松本の考え方を援用し，本研究では文章の長さに対して文章末での使用頻度に正の相関を持つ単語を抽出する方法を採り，得られた単語の頻度と意味属性を分析する。

先行研究［5－7］の結果をふまえて，分節された大河イメージに即した集計と分析を行った。本稿では，大河タームを大きく3つに分類し，その出現頻度について3年度間の比較を報告する。資料は先回と同じく，日経新聞CD-ROMの97，98，99年度版[9-11]の3年分を用いた。

2．大河タームの抽出と分類

大河タームの定義と抽出法を概掲する。

A：資料：日経新聞 CD-ROM の97,98,99年度に含まれる記事。但し，文章の形態でない記事を除いたため，使用した記事数は重複を除けば，97年度102610個，98年度104454個，99年度101450個である。

B：定義：y 年度の新聞記事データベースから，テーマＡの文章を集めたものを，「y 年度におけるテーマＡの文書グループ」と呼ぶ。各テーマの文書グループに順序をつけて，第 i 番目のグループを「y 年度における第 i テーマの文書グループ」と呼ぶ。大河タームは以下のアルゴリズムによってグループ毎に抽出される。

C：大河ターム抽出プログラム version 4

1．新聞記事 CD-ROM 内の記事の内，以下の条件を満たすものは文章の体裁をとっていないものとして除外する。

1－1：文の数が 3 以下

1－2：文字数が200以下

1－3：箇条書き

1－4：図表

1－5：スポーツの結果，書籍などの売り上げランキング，イベント告知

1－6：円相場，先物取引の相場，日銀概況

1－7：賞与，会社人事，死亡記事，家屋移転

1－8：インタビュー記事，首相の所信表明演説など口述録音の書きおこし

2．キーワード検索を用いて，同一テーマの記事を集め，それらを一つの文書グループとする。以下の条件を満たす文書グループだけを抽出に用いる。

$$200 = < \text{記事数} < 2000$$

3．得られた文書グループ内の全ての記事を文字数順に並べる。

4．記事数が同数になるように 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う。

5．文書群の全ての文書の末尾kセンテンスを取り出して，形態素解析を行う。[12]

6．各単語毎にブロック別の集計を行う。

7．第j文書グループにおける第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単語nの出現回数を$F(j, n, x)$とし，出現回数最大のブロックの出現回数を$Fmax(j, n)$とする。このとき単語nは

$$Fmax(j, n) \geqq A(j)/X/P$$

を満たさねばならない。（$A(j)$は第j文書グループの記事数，pは定数）

8．頻度分布（$x, F(n, x)$）に対して，単純回帰分析を行い，回帰係数の推定値がtより大で，かつ，回帰係数の検定において，回帰係数の値が0である帰無仮説を有意水準0.05で棄却できる単語のみを抽出する。

9．上の手続きで抽出される単語の各を「y年度の第j文書グループの大河ターム」と呼ぶ。

D：比較指標$FW[x]$：各文書グループから得られる上位x位までの頻出名詞の集合である。大河タームのグループ別平均抽出個数が7であるため，全グループについて一律$FW[7]$を用いる。

先行研究において大河タームは頻出名詞との比較の上で理念的に3つに分類されることを示した。

1：ある文書グループにおいて大河タームとして抽出されるが，同時にその当該グループの頻出名詞であるもの。文書グループのテーマに依存して頻出する非常に重要な名詞である。

2：ある文書グループにおいて大河タームとして抽出され，かつ当該グループの頻出名詞ではないが，同一年度の他の文書グループ頻出名詞であるもの。他の文書グループで頻出しており，他の文書グループの内容や関係を間接的に含むと考えられる。

3．ある文書グループにおいて大河タームとして抽出され，かつ同一年度の全ての文書グループにおいて頻出名詞とならないもの。使用されるテーマに依存せず，他のグループでの使用状況にも依存せずに長い文章の文章末に使用されやすい名詞である。

これら3パターンの頻度を集計し，3年度間の共通した傾向を見いだすために以下の定義を行う。先行研究では比較の指標として *FC*[7] を併用したが，本稿では *FW*[7] のみで行う。各大河タームは年度ごとに集計される。

*IFW 大河ターム：同一年度の大河タームの内，抽出された当該グループの *FW*[7] の要素であるもの。

*OFW 大河ターム：同一年度の大河タームの内，抽出された当該グループと別の文書グループの *FW*[7] の要素であるもの。

*NFW 大河ターム：同一年度の大河タームの内，その年度の全文書グループの *FW*[7] の要素にならないもの。

同じ名詞が別々の文書グループから大河タームとして重複抽出されることは頻繁に起こる。重複の頻度は大河タームによって異なる。よって，大河タームは重複するグループ数によって，頻度を比較可能である。以下に用語と集合を定義して，上記3分類の大河タームの出現頻度を比較する指標，及び全体との比較のための指標を導入する。

*最頻 *IFW-TT*[*x*]：重複回数の多いもの上位 *x* 位までの IFW 大河タームの集合。

*最頻 *OFW-TT*[*x*]：重複回数の多いもの上位 *x* 位までの OFW 大河タームの集合。

*最頻 *NFW-TT*[*x*]：重複回数の多いもの上位 *x* 位までの NFW 大河タームの集合。

*最頻 *TT*[*x*]：同一年度の大河タームの内，重複回数上位 *x* 位までの大河タームの集合。

この定義から，各グループから抽出される全ての大河タームはIFW大河ターム，OFW大河ターム，NFW大河タームのいずれかとなる。

OFW，NFW大河タームについて具体例を挙げよう。「影響」「業界」「資金」「通信」の4つの名詞は99年度の文書グループ「2000年問題」の大河タームである。これらは99年度の文書グループ「2000年問題」における$FW[7]$の要素ではなく，従ってIFW大河タームではない。これら4つの名詞が99年度の全文書グループから抽出される大河ターム，及び$FW[7]$の要素と，どれぐらい重複するグループを持つかを表1に示す。各の数値は重複するグループ数を示す。

4つの大河タームとも，複数の文書グループの大河タームとして重複して抽出される。1つ目の「影響」と残りの3つ「業界」「資金」「通信」は性質が異なることがわかる。「影響」は99年度の全文書グループから抽出される$FW[7]$の要素になることがない。一方，「業界」「資金」「通信」は99年度の「2000年問題」以外の多くの文書グループの$FW[7]$の要素となる。つまり，99年度におけるテーマ「2000年問題」については「影響」がNFW大河ターム「業界」「資金」「通信」がOFW大河タームとなる。

但し，99年度で抽出される全ての「業界」「資金」「通信」がOFW大河タームとは限らない。別のグループでこれらが大河タームとして抽出された場合，そのグループでの$FW[7]$の要素であることは起こり得る。つまり，別々のグループで重複して抽出された，同じ名詞の大河タームは，あるグループではIFW大河タームとして，別のグループではOFW大河タームとして分類されることが起こり得る。一方，NFW大河タームとなる名詞は，定義上，同一年度の全ての文書グループの$FW[7]$の要素にならないことから，他文書グループのIFW大河ターム，OFW大河タームのいずれにもなることはない。

表1：大河タームの重複例

	影響	業界	資金	通信
大河タームの重複回数	4	30	32	6
$FW[7]$との重複回数	0	2	23	8

3．名詞別集計結果

各年度の全ての大河ターム，及びIFW大河ターム，OFW大河ターム，NFW大河タームについて，その総数を求めた。年度別の数値を表2に示す。TTは大河ターム全体の総数，IFW-TTはIFW大河タームの総数を示す（OFW-TT，NFW-TTも同様）。IFW大河ターム，OFW大河ターム，NFW大河タームの総数のそれぞれに付く括弧内の数値は当該年度の大河タームの総数との割合である。

表2において，大河タームの総数，及び各IFW，OFW，NFW大河タームにおいて使用頻度の割合が3年度間でほぼ一定の傾向にあることが見て取れる。各年度においては，話題として取り上げられるトピックは年々変化しており（住民投票，沖縄基地問題，統一地方選挙，毒物カレー事件，2000年問題など），大河タームも入れ替わる一方で，年度単位でマクロ的に見てやれば，文章の末尾に使用されやすい名詞は一定の割合で存在していることが示される。

各年度の最頻*TT*[15]を表3に，最頻*IFW-TT*[10]を表4に，最頻*OFW-TT*[10]を表5に，最頻*NFW-TT*[10]を表6に示す。

表3の結果から，上位の重複数の大きな大河タームの出現頻度は，3年間で近い傾向にあることが見て取れる。出現頻度の傾向の近さを測るため，以下の指標z値を定義する。

$$z(i) = i - |\text{最頻}\ TT[i,97] \cap \text{最頻}\ TT[i,98] \cap \text{最頻}\ TT[i,99]|.$$

上式の最頻*TT*[i,97]，最頻*TT*[i,98]，最頻*TT*[i,99]は，それぞれ97年度，98年度，99年度の最頻*TT*[i]を示す。$|x|$は集合xの要素の数を表す。各項の最頻*TT*を入れ替えて最頻*IFW-TT*[10]，最頻*OFW-TT*[10]，最頻*NFW-TT*[10]についても同様の計量を行う。表3を例に取ると，$i=1$の時，つまり各年度の1位の大河タームについて，それぞれ「の」「こと」「企業」となるため，積集合は空集合となり，$z(1)=1$。$i=2$の時，つまり各年度の2位までの大河タームについては，97年度2位，98年度1位，99年度2位の「こと」が共通なので，積集合の要素の数は1となり，$z(1)=1$。$i=3$の時は「の」「こと」が共通なので$z(3)=1$。つまり，値は3年間の共通分からの各年度の誤差を示す。$i=12$の時，積集合の要素数が11であり，$z(12)=1$であるが，$i=13$の時，$z(13)=2$

となり，$z(14)=3$ となる。$i>12$以降，z値が大きく上昇していくことから，本稿では$z(i)<2$を満たしている順位内では3年度間の出現頻度に近い傾向があるものとする。

同様に最頻 *IFW-TT*[10]，最頻 *OFW-TT*[10]，最頻 *NFW-TT*[10] についても同様のz値の計量を行った結果（表4，表5，表6の最右列），IFW 大河タームについては第6位まで，OFW 大河タームについては第8位まで，NFW 大河タームについては第3位までが$z<2$を満たしていることから，3年度間で同様の傾向を示していることが見て取れる。特に NFW 大河タームについては，先行研究でも指摘したとおり，「今後」と「可能性」が突出して重複グループ数が多い。「今後」「可能性」以外の単語に関しては重複数も小さく，重複数の比較だけから年度間の傾向を見いだすことは難しい。

表2：年度別，各大河タームの総数

	97年度	98年度	99年度
TT	7843	8231	6445
IFW-TT	1562 (19.9%)	1597 (19.4%)	1218 (18.9%)
OFW-TT	5007 (63.8%)	5363 (65.2%)	4090 (63.5%)
NFW-TT	1274 (16.2%)	1271 (15.4%)	1137 (17.6%)

表３：年度別最頻 *TT*[15]

	97年度		98年度		99年度		z
O r	TT	No	TT	No	TT	No	
1	の	502	こ と	478	企 業	329	1
2	こ と	476	の	446	こ と	326	1
3	日 本	329	企 業	372	の	323	1
4	企 業	321	日 本	349	日 本	197	0
5	市 場	237	市 場	208	今 後	182	1
6	今 後	220	経 済	186	市 場	172	1
7	た め	160	今 後	183	た め	156	1
8	経 済	112	た め	175	事 業	113	1
9	事 業	103	事 業	121	可能性	106	1
10	米	102	金 融	120	経 済	100	1
11	改 革	100	米	118	米	80	1
12	経 営	100	経 営	107	経 営	78	1
13	競 争	90	可能性	101	地 域	60	2
14	会 社	86	競 争	98	問 題	58	3
15	問 題	86	銀 行	88	会 社	56	4

表４：年度別最頻 *IFW-TT*[10]

	97年度		98年度		99年度		z
O r	IFW-TT	No	IFW-TT	No	IFW-TT	No	
1	こ と	305	こ と	311	こ と	225	0
2	の	187	の	145	企 業	126	1
3	企 業	100	企 業	109	の	98	0
4	日 本	91	日 本	82	日 本	44	0
5	米	54	経 済	63	米	43	1
6	市 場	40	米	59	市 場	34	1
7	事 業	32	銀 行	40	銀 行	27	2
8	会 社	23	市 場	36	経 済	22	2
9	銀 行	22	事 業	35	事 業	18	1
10	開 発	18	市	25	会 社	16	2

表５：年度別最頻 *OFW-TT*[10]

	97年度		98年度		99年度		z
O r	OFW-TT	No	OFW-TT	No	OFW-TT	No	
1	の	315	の	302	の	225	0
2	日　本	238	日　本	267	企　業	203	1
3	企　業	221	企　業	263	日　本	153	0
4	市　場	197	市　場	172	た　め	145	1
5	こ　と	171	こ　と	167	市　場	138	1
6	た　め	150	た　め	157	こ　と	101	0
7	経　済	99	経　済	123	事　業	95	1
8	改　革	92	競　争	98	経　済	78	1
9	競　争	90	金　融	97	経　営	63	2
10	経　営	90	経　営	93	問　題	57	2

表６：年度別最頻 *NFW-TT*[10]

	97年度		98年度		99年度		z
O r	NFW-TT	No	NFW-TT	No	NFW-TT	No	
1	今　後	220	今　後	183	今　後	182	0
2	可能性	82	可能性	101	可能性	106	0
3	声	20	回　復	30	見　方	18	1
4	導　入	19	収　益	24	国　内	17	2
5	検　討	18	特　捜	20	声	16	3
6	見　方	18	見　方	17	考　え	16	3
7	特　捜	18	声	16	期　待	13	3
8	考　え	15	期　待	15	リストラ	13	3
9	予　想	15	地　元	15	指　摘	12	4
10	銘　柄	13	分　野	15	国	12	5

4．意味属性による集計結果

表3を見る限り，出現頻度の高い大河タームの使用状況は3年度間で一定の傾向を持つと考えられる。表3から得られるのは大河タームが抽出される重複グループ数であり，これは名詞の使用頻度を反映しているものの，文章末において，どんな概念が使われやすいか，どの意味属性を持つものが使われやすいかまでを推定することはできない。よって，全ての大河タームについて，意味属性値の分布を集計した。意味属性を与える指標として，先行研究と同様に日本語語彙大系[13]の中の単語意味辞書と単語意味属性体系を使用した。単語意味辞書から各大河タームの意味属性値を得た後，単語意味属性体系の各ノード上での個数分布を集計した。単語意味属性体系の最深12段，約2700のノードのうち，5段目までのノードを用いて，大局的に分類する。5段目以下のノードの意味属性を持つ単語については，その5段目の親ノードが意味を代表するものとして頻度を集計した。表3から表6までは，大河タームについての名詞単位の重複グループ数を示すが，以下の表7から表10では，意味属性ノード単位の重複グループ数を示している。

表7から表10まで，上位の意味属性の出現傾向は共通分が大きい。前章のz値を表7の意味属性の順位に転用するために以下の定義を行う。

*最頻 $SFTT[x, y]$：第y年度の大河タームについて意味属性ノード単位で重複回数の上位x位までの意味属性の集合。前章同様，

$$z(i) = i - |\text{最頻 } SFTT[i,97] \cap \text{最頻 } SFTT[i,98] \cap \text{最頻 } SFTT[i,99]|.$$

に従ってz値を求める。IFW 大河ターム，OFW 大河ターム，NFW 大河タームの場合も同様に行う。$z(i) < 2$を満たす最大の順位は表7で第8位，表8で第9位，表9で第5位，表10で第8位となる。各大河タームの 値の単語単位の集計と意味属性ノード単位の集計の違いを表11に示す。

表11において，意味属性ノード単位の集計を単語単位の集計と比べると，大河ターム全体，OFW 大河タームでは，共通する順位を下げている。その一方で，IFW 大河ターム，NFW 大河タームは共通と見なせる順位を上げている（太字）。NFW 大河タームについて，表6では上位2位までを除いては顕著な共通性は見いだせ

なかったが，表10の集計結果を見れば，上位８位程度まではほぼ共通した傾向を見いだすことができる。NFW 大河ターム「今後」「可能性」をそれぞれ含む「非暦日」「様相」が上位であるのは表６からの帰結といえるが，「精神」「行為」「変動」「知的生産物（思考・学習）」「制度」「人工物」が上位に共通することは注目に値する。

表７：年度別大河ターム全体についての意味属性値の集計結果

	97年度		98年度		99年度		z
	意味属性(５段目以上)	計	意味属性(５段目以上)	計	意味属性(５段目以上)	計	
1	行為	1031	行為	985	行為	918	0
2	団体・党派	650	制度	792	団体・党派	580	1
3	制度	637	団体・党派	683	制度	551	0
4	類	514	事	479	類	329	1
5	事	476	類	456	事	326	0
6	精神	365	非暦日	328	精神	288	1
7	知的生産物(思考･学習)	319	精神	289	非暦日	285	1
8	非暦日	312	人工物	285	変動	260	1
9	界	308	界	255	人工物	244	2
10	人工物	304	変動	245	知的生産物(思考･学習)	218	2

表８：年度別 IFW 大河タームについての意味属性値の集計結果

	97年度		98年度		99年度		z
	意味属性(５段目以上)	計	意味属性(５段目以上)	計	意味属性(５段目以上)	計	
1	事	305	事	311	事	225	0
2	類	187	団体・党派	196	団体・党派	189	1
3	団体・党派	179	制度	153	行為	155	1
4	行為	148	行為	149	類	98	1
5	人工物	105	類	145	制度	87	1
6	制度	73	人工物	92	人工物	75	0
7	人(職業･地位･役割)	42	人(職業･地位･役割)	47	人(職業･地位･役割)	38	0
8	界	40	行政区画	38	界	34	1
9	知的生産物(思考･学習)	32	界	37	変動	23	1
10	精神	32	公共施設	28	精神	21	2

表９：年度別OFW大河タームについての意味属性値の集計結果

	97年度		98年度		99年度		z
	意味属性(５段目以上)	計	意味属性(５段目以上)	計	意味属性(５段目以上)	計	
1	行為	783	行為	736	行為	686	0
2	制度	532	制度	594	制度	430	0
3	団体・党派	459	団体・党派	472	団体・党派	376	0
4	類	318	類	304	類	226	0
5	界	266	界	218	変動	193	1
6	知的生産物(思考･学習)	251	事	167	知的生産物(思考･学習)	175	2
7	精神	177	理由・目的等	160	界	170	2
8	事	171	人工物	157	人工物	150	3
9	人工物	169	知的生産物(思考･学習)	154	精神	150	2
10	理由・目的等	150	変動	148	理由・目的等	147	2

表10：年度別NFW大河タームについての意味属性値の集計結果

	97年度		98年度		99年度		z
	意味属性(５段目以上)	計	意味属性(５段目以上)	計	意味属性(５段目以上)	計	
1	非暦日	264	非暦日	249	非暦日	223	0
2	精神	156	様相	134	様相	118	1
3	様相	110	精神	118	精神	117	0
4	行為	100	行為	99	行為	77	0
5	変動	80	変動	73	変動	44	0
6	知的生産物(思考･学習)	36	制度	44	制度	34	1
7	制度	32	人工物	36	人〈職業･地位･役割〉	30	1
8	人工物	30	知的生産物(思考･学習)	30	知的生産物(思考･学習)	28	1
9	言語	21	人〈職業･地位･役割〉	16	因果	27	2
10	人間	18	郷里	15	機関	21	2

表11：各大河タームの集計単位と出現傾向が近いと見なせる順位

	TT	IFW	OFW	NFW
単語単位	12	6	8	3
意味属性単位	8	9	5	8

5．考　察

表7から表11の要点を述べる。

1：意味属性ノード単位で大河タームの重複グループ数を集計した結果，この場合も出現頻度の年度間の共通性が確認された。

2：単語単位の場合と同様，重複グループ数の大きい意味属性の中には，IFW 大河タームでの頻度が高いものと OFW 大河タームでの頻度が高いものがある。

3：単語単位の重複グループ数では一定の傾向が見いだしにくかった IFW 大河ターム，NFW 大河タームについて，意味属性ノード単位の集計にすることで，それぞれ上位 9 位程度，8 位程度までで共通の出現傾向が見られた。

表8における IFW 大河タームの，表10における NFW 大河タームの意味属性での重複グループ数の集計結果は，出現頻度の低い名詞を集めたことが原因で年度間での共通性を得ることができたことを示している。最も顕著な例として，表10で3年度とも出現頻度が5位になった意味属性「変動」を取り上げる。意味属性ノード「変動」あるいはその下位範疇のノードに属する NFW 大河タームの各年度で頻度の高いものを以下に挙げる。

97年度：「導入(19)」「再編(8)」「破たん(8)」「調整(7)」「抜本(5)」他15個。

98年度：「回復(30)」「介入(5)」「破たん(5)」「控除(3)」「連合(3)」他19個。

99年度：「導入(8)」「普及(8)」「再開(4)」「安定(3)」「追及(3)」他19個。

括弧内は大河タームとしての重複グループ数である。上位5つの NFW 大河タームを見る限り，3年度間で共通する傾向を見いだしにくく，また，各の大河タームの重複グループ数も小さい。これら出現頻度の小さな大河タームの重複グループ数を意味属性単位で集計した結果，共通した傾向を見いだすことができた。

6．理解の成立に向けて

「理解の成立」を考察するため，特定の論理回路の形成・習得という面に限定して，さらに文章理解の一つの手がかりとして文章末に注目した。文章の終わりを認識することは，多様な意味の連関が収束することを認識することでもある。こ

の意味で，文章末に注目することは，文章理解あるいは談話理論[たとえば,14]とされる推論，照応，連接構造，のどれとも違う観点でありながら，互いに必要とし合うことになるだろう（ここでいう推論とは論理学でいうところのそれと違い，文と文，節と節をつなぐ論理のことであり，スキーマ理論，スクリプト理論などがある)。このことはデジタルシーケンスからなる正規言語の受理装置の理論であるオートマトン理論，あるいは句構造言語の受理装置であるチューリングマシンにおいて，next move function とは独立に最終状態の定義が必要であり，お互いが完備されることで初めて受理が成立することに喩えることができる。

自然言語の文章の最終文は，見かけ上，他の文との差はない。しかし，多くの文章において，途中の文での中断は，ひとつのまとまりとしての整合性に欠ける印象を与える。「理解」には至らない。このことは，最終文あるいはそれにつながる複数の文と，他の部分の文との間に違いがあることを推測させる。特定の文字はもちろん，特定の単語，特定の意味が文章を終わらせて，整合性を持たせる機能を持つことは考えられないが，一方で全ての単語が平滑して同等に使われているとも考えられない。先行研究［例えば4］にも示したが，新聞記事の全文章を用いた調査で，文章全体の最頻名詞と最終2文だけを取り出した最頻名詞を比較したところ，顕著な差は見いだし得なかった。本報告では大河タームに焦点をあてているが，これらは最終2文での使用頻度から計量されるものではない。これらは文章の長さとの相関から導き出される。大河タームの存在が意味するところは，文章末そのものには特定の単語，特定の意味属性が偏重して使用されることはないが，文章の長さが長い場合，換言すれば，中断することでまとまりが欠けて文章の「理解」に至らない点が増える場合には，特定の単語，特定の意味属性の使用頻度が大きくなることである。大河タームそのものが文章の流れに「終わり」を与える機能を持っている，と主張するわけではない。むしろ，大河タームが使用される前の段階でなされる推論や照応，あるいは使用される連接構造が問題になるだろう。筆者はこの観点から最終文内に大河タームを導く文章の流れを意味属性のシーケンス解析という形で研究中である（準備中)。

調査内容をまとめる。大河タームの出現頻度について，重複して現れるグループ数を用いて3年度間で比較した結果，上位12位程度まで同様の傾向を得た。大河タームを IFW 大河ターム，OFW 大河ターム，NFW 大河タームに分類した。こ

れらの重複グループ数を3年度間で比較した結果，IFW 大河タームと OFW 大河タームでそれぞれ上位6位と8位程度までで同様の傾向を得た。NFW 大河タームでは上位2位の「今後」「可能性」に顕著な結果が出たが，下位の名詞では共通した傾向を見いだせなかった。さらに大河タームを，日本語語彙大系の意味属性体系を用いて意味属性ノード単位で重複グループ数を集計した結果，単語単位での比較結果と同様，上位の意味属性の出現頻度に共通した傾向を見いだすことができた。IFW 大河タームと NFW 大河タームでは，単語単位で見いだせなかった3年間での共通性が，それぞれ高い順位にまで見いだされた。

今回の意味属性の集計は単語意味属性体系シソーラスの5段目を使って行ったが，これを2段目，3段目まで上げて観るとさらに特徴が現れる。2段目まで上げると分類は2項目しかなく，「具体」「抽象」のみである。相対的な分類ではあるが，前者が具体名詞，後者が抽象名詞を指すと見なして良いだろう。前者には「団体・党派」「界」「人工物」「人〈職業・地位・役割〉」などが当たるがこれらは表7から表9においては各年度に上位3位までに「団体・党派」が入り，他の属性も上位に散見される。しかし，表10においては「団体・党派」は現れなくなり，3年度とも上位6位まで全て「抽象」で占められる。このことは大河ターム自体が抽象名詞の割合が高いことだけでなく，NFW 大河タームのほとんどが抽象名詞であることを示している。

またシソーラスの3段目に注目すると，2段目の「抽象」から3つの子ノードがリンクしており，それぞれ「抽象物」「事」「抽象的関係」の3者の意味属性がある。上記と同様，相対的ではあるが「抽象物」と「事」「抽象的関係」の意味属性を持つ名詞を比べた場合，「事」「抽象的関係」に属する物の方が抽象度は高いといえる。このうち，表10の NFW 大河タームの3年度間の上位5位の意味属性「非暦日」「精神」「様相」「行為」「変動」は全てこの2者「事」「抽象的関係」の下位ノードである。さらに NFW 大河タームの2つの特徴的名詞である「今後（5段目の意味属性は非暦日）」「可能性（5段目の意味属性は様相）」とも「抽象的関係」の下位ノードに属する。

NFW 大河タームは，定義から，長い文章の文章末で使われやすく，一般的なテーマの文章を集めても頻出単語になることもない。NFW 大河タームが抽象的関係を述べる名詞に当てはまりやすい事は，これらが使われている文章の文章末

に於いて前段の内容のある種のモデル化，あるいはモデル化を前提にして導かれる言明が表現されていると推測される。「今後」「可能性」という２つの単語は，これらが使われる文章に於いて，文章の終わりを導くターミナルとしての役割を持つと考えられる。これらがターミナルとするならば，それにつながる推論や照応の分類が続く調査となるだろう。

以上が理解の成立に向けての本研究の取り組みである。ここで得られた知見が，文章の終わりを認識し，文章を一つのまとまりとして捉えること，ひいては特定の論理回路を形成・習得すること，理解が成立することに一定の寄与があることが期待される。

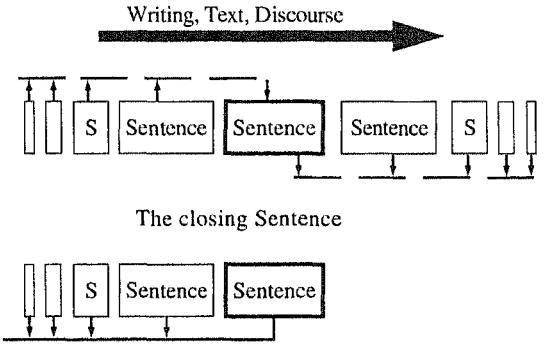

図１：最終文の他の文とのつながり

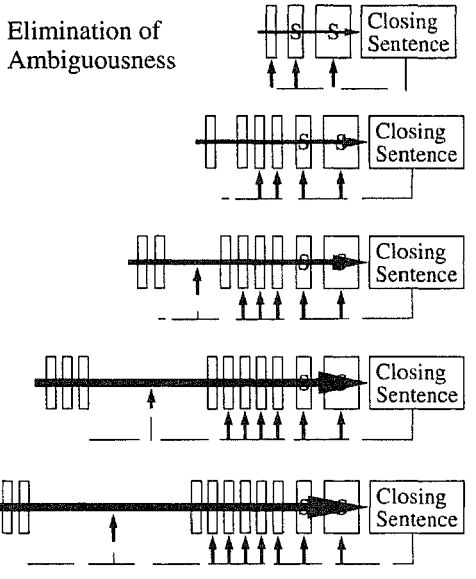

図２：文章の長さと曖昧性の増大

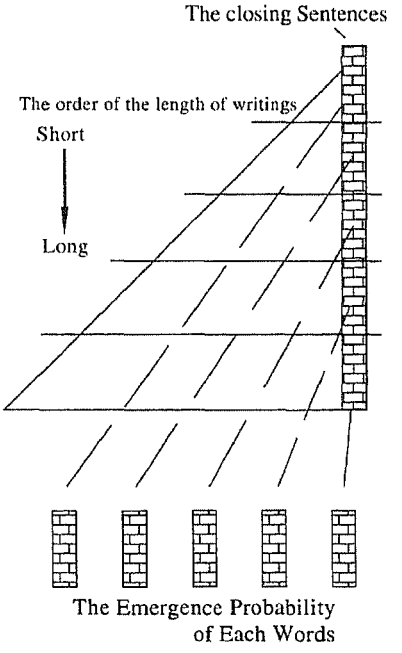

図３：文書グループのブロック分けと最終文の取り出し

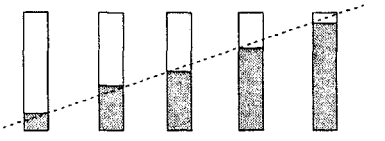

図４：ブロック別使用頻度の例

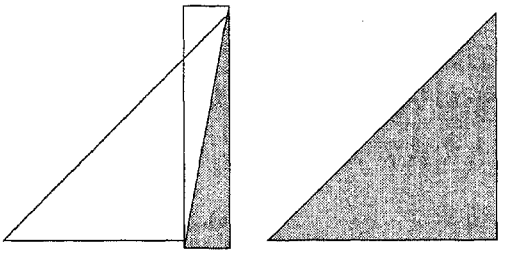

図５：文書グループ内で抽出に用いる範囲。
左が大河ターム，右がFW

参考文献

[１] 市川孝：『国語教育のための文章論概説』，教育出版，1978。

[２] 市川孝：『改訂文章表現法』，明治書院。

[３] 進藤咲子：書き終わりのタイプ，『国文学：解釈と鑑賞』臨時増刊号，1974-６。

[４] 永野賢：『文章論総説』，朝倉書店，1986。

[５] 中村隆志，小泉明日美，本間愛：日本語新聞記事の文章末における特異的名詞，情報処理学会報告，ICS116-４，1999。

[６] Takashi Nakamura, Tatsuo Hemmi & Asumi Koizumi: Semantic Features of specific words in the closing sentences of newspaper articles, Proceedings of The second Annual Conference of The Japanese Society of Language Sciences, 2000.

[７] 中村隆志，廣木真理 多国語新聞記事の大河ターム分析（その１），情報処理学会報告，CH48-５。

[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] 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，1998。

[10] 日本経済新聞社，日本経済新聞98年 CD-ROM 版，日本経済新聞社，1999。

[11] 日本経済新聞社，日本経済新聞99年 CD-ROM 版，日本経済新聞社，2000。

[12] 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。

[13] 池原悟，他編：日本語語彙大系，岩波書店，1997。

[14] 阿部純一他：『人間の言語情報処理』，サイエンス社，1994。