



Title	カラムジン『哀れなリーザ』における普通名詞表現の分析：データベースと多変量解析
Author(s)	浦井, 康男; Urai, Yasuo
Citation	スラヴ研究, 42, 101-116
Issue Date	1995
Doc URL	https://hdl.handle.net/2115/5236
Type	departmental bulletin paper
File Information	KJ00000113390.pdf



カラムジン『哀れなりーザ』における 普通名詞表現の分析

—データベースと多変量解析—

浦井康男

§1 テキストの形態処理と語彙統計

電算機は通常、個々の変化語形をそのまま一語として扱うため、例えばテキスト中の слово の個々の変化形 (слова, словом, слове 等) については、それらの出現回数を簡単に数えることができる。しかしこれらの諸形をまとめて、слово として何回出現したのかは、直接数えることが出来ない。そのため語形変化の多い言語では、簡単な語彙統計を取ることさえ、なかなか容易ではなかった。

だがロシア語の語形変化が電算機上のプログラムで実現されると、変化語形とその見出し語 (名詞では主格、形容詞では男性主格、動詞では不定形等) の間を結び付けることが可能となる⁽¹⁾。この作業を自動的に連続して行うように設計すると、テキスト中の個々の語形に、その見出し語と形態論的情報を付加したデータを作ることができる。例1は『哀れなりーザ』のテキスト⁽²⁾を縦に並び替え (WF の欄の語形を縦に読むとテキストそのものになる)、その各々の語形に文法情報 (HI, H2) と見出し語 (LEX2) を付加して作成したデータベースの一部である⁽³⁾。

例1 (Example 1)

REF	SNO	PNO	WF	HI	H2	LEX2
041_02_01	2	1	может	v	i	мочь1
041_02_02	2	1	быть	v	i	быть1
041_02_03	2	1	,	,	rup	,
041_02_04	2	1	никто	pr	prg	никто1
041_02_05	2	1	из	pre	A	из1
041_02_06	2	1	живущих	v	i	жить1
041_02_07	2	1	в	pre	A	в1
041_02_08	2	1	Москве	n	f0	Москва1

このデータをまず品詞(HI)ごとに分け、次にその中を見出し語(LEX2)のアルファベット順に並び替えて⁽⁴⁾、同じ見出し語の数を数えると、品詞ごとに分けられた見出し語の語彙統計が得られる。例2は『哀れなりーザ』で使われている名詞 (HI=n) について、上記の手続きで得られた統計を、さらに出現頻度順に並び替えたものである (頻度5まで表示)。

なお見出し語の後ろの数字は同形異義語の識別番号で、大部分は1である⁽⁵⁾。

例2 (Example 2)

頻度 見出し語

114 : Лиза1

46 : Эраст1

24 : сердце1

23 : глаз1

21 : мать1

18 : человек1

17 : день1, рука1

15 : слеза1

13 : душа1

12 : бог1, смерть1, цветок1

11 : матушка1, раз1, время1

10 : жизнь1, земля1, свет2, старушка1, друг1

9 : дочь1, любовь1

8 : Москва1, небо1, слово1, хижина1, чувство1

7 : берег1, взор1, луч1, место1, минута1

6 : девушка1, лицо1, луг1, монастырь1, мысль1, натурал1, образ1, объятие1, река1, удовольствие1

5 : голос1, город1, дело1, дуб1, окно1, отец1, поле1, роща1, рубль1, стадо1, трава1, час1

動詞のデータは注6で示すが、二つの集計結果から両者の分布には大きな違いがあるのがわかる。それは動詞で高頻度のものには、一般的な意義のものが多く(例えば、быть, мочь, сказать, хотеть, говорить等)、作品のテーマに関連しそうな語は意外に頻度が低い(плакать:7, поцеловать:6回)⁽⁷⁾。これに対して名詞では、固有名詞(Лиза, Эраст, Москва)や一般的な意義のもの(день, раз, время, место等)を除いても、具体的な意義を持ち作品のテーマと関連しそうな語が、かなり高い頻度で出現している(сердце:24, глаз:23, рука:17, слеза:15回)。

そこで本稿では、これらの普通名詞に焦点を当て⁽⁸⁾、これらが作品中でどのように使われ、どのような表現機能を担っているかを解明してみたい。そのためには各々の変化語形にその品詞や見出し語の情報が付加された、例1のデータベースを体系的に運用し、それぞれの局面に必要なデータを逐次得ていくことになる。

§2 名詞の共出現と数量化3類

電算機による語彙分析でよく使われる手法に、共出現関係の分析がある。これは「一つ

の表現域に見出し語*i*も*j*も使ってあれば、その*i*と*j*とは共出現している」と定義し⁽⁹⁾、複数個の表現域での見出し語の共出現を示すデータを形式的に処理して、これら見出し語間の内容的な関連を探ることである。例3でa1～a7は見出し語、b1～b5は表現域(例えば各センテンス)を、また1/0は見出し語の出現/非出現を示すものとする。するとこのデータは、例えば表現域b1で見出し語a1とa4が、b2でa3とa5とa7が出現することを示している。

例3

	a1	a2	a3	a4	a5	a6	a7
b1	1	0	0	1	0	0	0
b2	0	0	1	0	1	0	1
b3	1	0	0	0	0	1	0
b4	0	0	1	0	0	0	1
b5	0	1	0	1	0	0	0

例4

	a1	a4	a6	a2	a5	a3	a7
b1	1	1	0	0	0	0	0
b3	1	0	1	0	0	0	0
b5	0	1	0	1	0	0	0
b2	0	0	0	0	1	1	1
b4	0	0	0	0	0	1	1

このデータのままで、見出し語間の関連はとらえにくい。これを例4のように、出現を示す1が対角線に沿う形に、a1～a7およびb1～b5の列と行を入れ換えると、その中の内在的な関連が見えてくる。つまり例4で左上と右下に分かれた分布から、例3での見出し語はa1、a4、a6、a2と、a5、a3、a7のグループに2分され、また表現域も二つに分離できる。このような並び替えの処理は、データ量が少なければ手作業でも可能だが、データの規模が大きくなると、全体を見渡すことが困難になる。

しかしこの並び替えは、多変量解析における「林の数量化3類」という形で数学的に定式化されている。詳しい点は専門書に委ねるが⁽¹⁰⁾、この並び替えは出現を示す1を出来るだけ対角線近くに置く、つまり各データの対角線までの距離の総和が、最小になるような配置を求めるという操作の数学的な実現であり、最終的にはベクトルの固有値問題に還元される。この値を直接得ようとするとかかなり厄介な手続きが必要だが、我々がここで行うのは、複数の表現域における見出し語の共出現をチェックした、例3の形式のデータを作り、これを多変量解析プログラムに送り込むことだけである⁽¹¹⁾。

するとプログラムは固有値の近似解を計算し、共出現関係で類似したものには近い値を与え、そうでないものには離れた値を与える。そこでこれらの値に従って、個々の見出し語を空間上に配置していくと、親近性の高いもの同士が集まって、見出し語がいくつかのグループに分けられる可能性が出てくる。

なおこの作業の特徴は、語の意義内容にまったく依存しない、形式的な処理であるという点にある。そして分析の次の段階で、この結果を実際の作品の中で解釈し、それらが持つ実質的な意味を検討することになるが、まず形式的な処理から始めよう。

§3 数量化3類による分析

共出現関係を利用した普通名詞の表現機能の分析のためには、対象となる見出し語の選

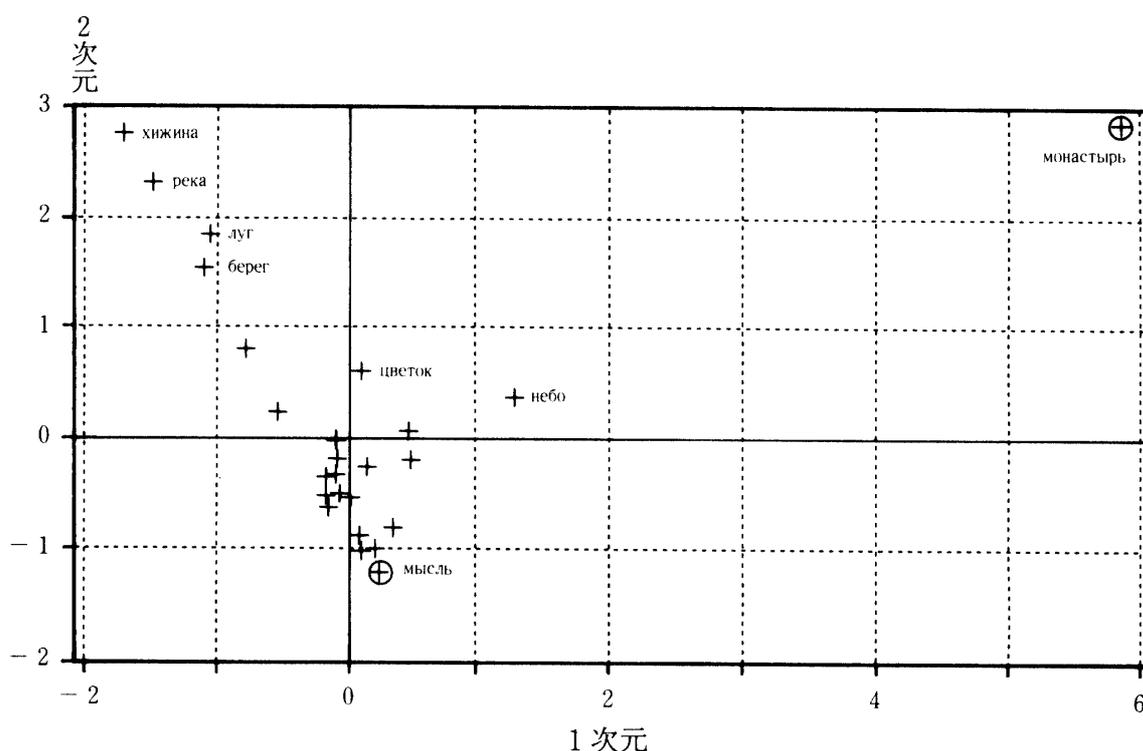


図1 (Fig.1)

平面上に配置すると図1になる。なお固有値は3次元まで計算している。

図1から монастырь が、また3次元を考慮すると мысль も⁽¹⁴⁾、原点から大きく外れた特異点をなし、一方他の見出し語は、際立ったまとまりを作らないことがわかる。これはこの作品においては、一文ずつのチェックでは「一つの表現域」の範囲が狭すぎて、データが孤立ぎみになったためと推察される。

そこで例6で示すように、「一つの表現域」の範囲を二文に拡大し、一文ずつずらしながらチェックを試みる。なおこの場合段落の終りに達したら、チェックはそこで終り、次の文には行かないようにする。このような設定をすると、見出し語は最大2回数えられ、現実とのずれが生じるが、前述より数量化3類では出現頻度は問題にしないため、このような処理は許容されると考える。

例6

段落	文番号	
3	12	┌── 12 (文番号 12 と 13 でのチェックを文番号 12 に記す) ├── 13 (13 と 14 での結果を 13 の所に記す) └── 14 (14 と 15 での結果を 14 の所に記す) └── 15 (段落末のため 15 だけでチェック)
3	13	
3	14	
3	15	
4	16	┌── 16
4	17	└──

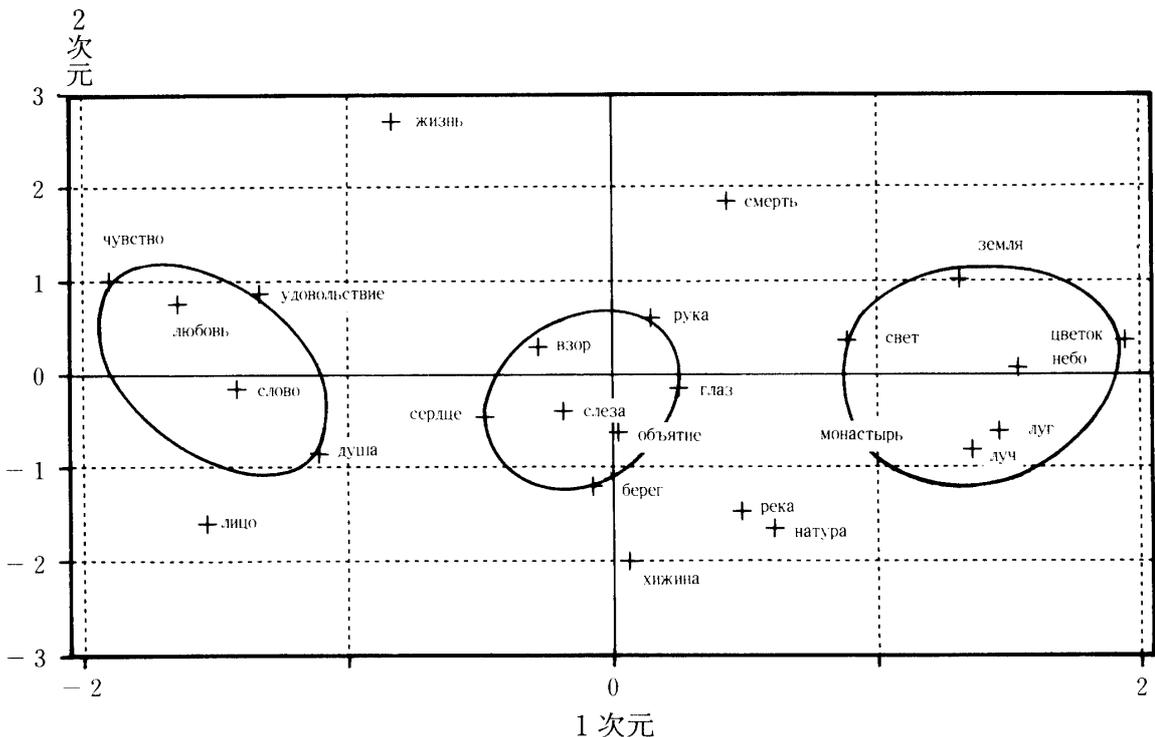


図2 (Fig.2)

チェックを二文に拡大して調べたデータの解析結果は省略するが、монастырьとмысльは相変わらず特異点をなし、他の語群には多少のまとまりができる。そこで「一つの表現域」の範囲をさらに拡大し、三文の範囲でチェックを行い、その結果を解析した。するとмысльは特異点のままであるが⁽¹⁵⁾、монастырьは他の語群に含まれ、しかもこれらはいくつかのグループに分けることが可能のように見える。ただмысльのために、データ全体が偏って表示されるので、これを除いた見出し語25個で再度解析を行った。結果を図2で示す。

図2から、これら25個の見出し語は共出現関係の親近度に従って、大きく三つのグループに分けられよう。それらを図上の右から順に示すと次のようになる。

第I群：цветок, небо, луг, луч, земля, монастырь, свет

第II群：глаз, рука, объятие, берег, слеза, взор, сердце

第III群：душа, удовольствие, слово, любовь, чувство

なおсмерть, жизнь, хижинаは少し値が離れているために、またлицоは3次元の値が高いために、どこのグループにも帰属させなかった⁽¹⁶⁾。一方рекаとнатураはI群とII群の中間に位置している。

ここで各グループに属する個々の語の意義内容を検討すると、I、III群の各々には意義に共通性があることがわかる。それはI群がいわば外界（具体的世界）に、III群が心や心情（抽象的事物）に関係していることである⁽¹⁷⁾。

これに対してII群は、一見雑多な集まりのように見える。しかしこの作品ではリーザの感情は、以下の例で示すように、言葉よりも彼女のちょっとした仕草で表現されることが多く、

またこのような場合にⅡ群の語がよく使われている。従ってⅡ群の語は、外界を示す *берег* を除き、感情を示す身体の一部や感情を引き起こすもの、つまり感情に関わるものと規定することができると思われる⁽¹⁸⁾。例7で *рука* を含む典型的な例を示すが、ここでの *рука* は、単なる「手」を越えて、彼女の感情を表現したり、感情を喚起するものとなっている。

例7 (イタリック：Ⅱ群の語)

—— Я сам по временам могу заходить к вам». Тут в *глазах* Лизиных блеснула радость, которую она тщетно сокрыть хотела; щеки ее пылали, как заря в ясный летний вечер; она смотрела на левый рукав свой и щипала его правою *рукою*./45 24-28⁽¹⁹⁾

Эраст выскочил на берег, подошел к Лизе и — мечта ее отчасти исполнилась: ибо он взглянул на нее с видом ласковым, взял ее за *руку*... А Лиза, Лиза стояла с потупленным *взором*, с огненными щеками, с трепещущим *сердцем* — не могла отнять у него *руки*, не могла отворотиться, когда он приближался к ней с розовыми губами своими.../47 09-15

一方明らかに外界を示す *берег* が、感情に関わるⅡ群の中に組み込まれているが、その理由は、リーザとエラストの恋がモスクワ河の岸辺を舞台に展開し、同じ文脈にこれらの語が共出現するためと推定される。このような点に、意味を排除した形式的な処理の弱点が現れるが、これを踏まえれば、逆にⅡ群の *берег*、Ⅰ群とⅡ群の間にある *река*、およびリーザと母親が暮らす *хижина* は、Ⅰ群の「遠い風景」に対する「近い風景」として解釈することが出来よう。

以上より、この節の初めで示した26個の見出し語は、共出現に関する数量化3類の解析によって、次のように分類される。

Ⅰ群 (外界) : *цветок*, *небо*, *луг*, *луч*, *земля*, *монастырь*, *свет*

Ⅱ群 (感情) : *глаз*, *рука*, *объятие*, *слеза*, *взор*, *сердце*

Ⅲ群 (心) : *душа*, *удовольствие*, *слово*, *любовь*, *чувство*

Ⅰ'群 (近い外界) : *берег*, *река*, *хижина*

特異点 : *мысль*⁽²⁰⁾

その他 : *смерть*, *жизнь*, *натура*, *лицо*⁽²¹⁾

このような形でのグループ化が成功したのは、最大3文の範囲内でこれらの語に共出現関係があったためであるが、これを換言すれば、『哀れなリーザ』ではこれらの名詞は、意義的に近いもの同士が互いに組み合わせられ、重ね合わされる形で使われていることになる。そしてこの結論は、第一節で問題提起した、この作品におけるこれらの名詞の使用上の特徴を示すものともなろう。例7でⅡ群の例を示したが、以下の二例でⅠ、Ⅲ群の組み合わせられた場合を示す。

例8 (イタリック：Ⅰ群の語)

«Ах, Лиза! — говорила она. — Как все хорошо у господина бога! Шестой десяток доживаю на *свете*, а все еще не могу наглядеться на дела господни, не могу наглядеться на

чистое *небо*, похожее на высокий шатер, и на *землю*, которая всякий год новою травую и новыми *цветами* покрывается. — — — /48 18-23

例9 (イタリック：Ⅲ群の語)

Все блестящие забавы большого света представлялись ему ничтожными в сравнении с теми *удовольствиями*, которыми страстная дружба невинной *души* питала сердце его. С отвращением помышлял он о презрительном сладострастии, которым прежде упивались его *чувства*. «я буду жить с Лизою, как брат с сестрою, — думал он, — не употреблю во зло *любви* ее и буду всегда счастлив!»/49 10-17

しかしながら、上記の見出し語のグループ分けは、三文の範囲での共出現関係のチェックを、一文ずつずらして行った結果であり、そのため個々の見出し語は最大3回数えられ、実体を正確に反映していない恐れがある。またこのグループ分けが、例7、8、9のような作品中の特定の箇所だけでなく、作品全体においても成功しているかどうかを検証する必要がある。

そこで次節では、これらⅠ、Ⅱ、Ⅲ群の18個の名詞を対象に、それらの実際の出現数を段落毎に集計して、その分布のデータから個々の段落の特徴を考えるという作業をしてみよう。この作業を別の観点から述べると、作品に出現する全語形5047個を約1300個の見出し語に還元し⁽²²⁾、さらにその中のわずか18個の普通名詞の出現を手がかりに、この作品の構成を探ることを意味するが、はたしてこのようなことが可能であろうか。

§4 三つの語群による段落の性格付け

これらの見出し語を使った段落の性格付けのため、例1のデータベースに対して、段落単位で該当する見出し語の出現をチェックし、これらをⅠ、Ⅱ、Ⅲ群別に集計するプログラムを作成した。出力結果に編集を加えたものを次ページの図3で示す。

なお図3では「段番」は初めから数えた段落番号、「文計」はその段落に含まれるセンテンスの合計⁽²³⁾、「語群別集計」でのアラビア数字1、2、3はそれぞれ語群Ⅰ、Ⅱ、Ⅲを、数字の後の*一つはその語群に属する名詞が1回出現したことを示す。また「コメント」はその段落の概要であるが、これを手がかりにこの作品のプロットの展開を区切ると、図3で示す一〜五(Ⅰ〜Ⅴ)部に分かれよう。

それでは第一部から第五部について、各段落の内容とそこに含まれる各語群の出現のパターンを検討しよう。

まず第一部は導入部に当り、モスクワ郊外の描写や主人公の説明等の状況設定に当てられている。そのためⅠ、Ⅱ、Ⅲ群の見出し語中、ここで基本的に使われるのが、外界に関連したⅠ群の語であることは、当然ともいえる。ただ修道院での苦行僧のイメージ(段落3)、夫を亡くしたリーザの母の嘆き(段落6)、モスクワの街頭で出会ったエラストに淡い思いをよせるリーザの様子(段落7、9)が語られるため、Ⅱ群の語もいくつかの段落でかなり現れている。

部	段番	文計	語群別集計	コメント	
I	1	2	1 * 2 *	作者はよくモスクワ郊外を散策する	
	2	6	1 * * * * * 2 *	モスクワ郊外の描写	
	3	9	1 * * * * 2 * * * * 3 * * *	修道院の廃墟、苦行僧のイメージ	
	4	3	1 * 2 * *	Лиза (以下 Л) の物語の導入	
	5	2	1 *	Л家の廃墟	
	6	17	1 * * * * * 2 * * * * * *	Лの家族の説明、夫を亡くした母の嘆き	
	7	21	1 * * * * * 2 * * *	町で花売り、Эраст (Э) に初めて出会う	
	8	10	2 * * *	Лが町に出た時の母の心配	
	9	5	1 * * 2 * *	Эのため花を摘み、町でЭを待つ	
II	10	39	1 * * * 2 * * * * * * * * * 3 *	Эが窓辺に立ち飲物を所望、Лの恥らい	
	11	6	1 * * 2 * * * * 3 *	作者によるЭの紹介、Лに一目惚れ	
	12	20	1 * * * * 2 * * * * 3 * *	Лの心にЭが焼き付く、姿を牧童に重ねる	
	13	1	0	Эがボートに乗って現れる	
	III	14	11	2 * * * * * 3 * *	Эが愛を告白、Эも愛されているのを知る
		15	26	2 * * 3 *	二人の愛の確認の対話
		16	21	1 * * * * * * * 2 * * 3 * *	Лには喜びで世界が輝く、母の自然賛歌
		17	18	1 * * * 2 * * * * * * * * 3 * * * * * *	ЛとЭの逢引、二人の充足された心
		18	9	2 * 3 * *	Лに促されЭは母にも会う
		19	23	1 * 2 * * * 3 * *	母の婿探しにЛが悲しむ
		20	6	1 * * 2 *	Эの抱擁に飛び込み、一線を越える
		21	27	2 * * * 3 * * * *	起こったことにЛはおびえる
		22	10	2 * * * * 3 * * * * * * * *	逢引は続くが全てが変わる、破局の始まり
23		5	0	Эは軍役に就くことを伝える	
IV	24	27	1 * 2 * * * 3 *	Лは嘆き、自分を忘れないように懇願	
	25	2	0	作者の注	
	26	13	2 *	Эと母の最後の会話	
	27	6	1 * 2 * * 3 *	ЛとЭの最後の逢引	
	28	1	1 * 2 * 3 *	Эが立ち去る、Лは失神	
	29	20	1 * * * 2 * * * *	Лは後を追うのをやめ、悲しみに耐える	
	30	11	2 * * * 3 *	二月後ЛはЭに再会、ЭはЛを追い出す	
	31	4	2 * *	作者の憤慨	
	32	6	0	作者によるЭの行動の説明	
	33	23	1 * * * * * 2 * * 3 * * *	Лは絶望し、池に身を投げる	
V	34	2	3 *	Лの死	
	35	2	2 *	Лの墓、作者の訪れ	
	36	4	2 *	母の死、Лの家は廃屋になる	
	37	5	0	Эの後悔、後日談	

図3 (Fig.3)

第一部と第二部を区切るのは、段落 10 の最初の部分、エラストがリーザの家の窓辺に立っているのを彼女が見つかる箇所である。

第二部ではエラストがリーザの家に入り、母親と言葉を交わす。リーザは恥らうが、彼の姿は彼女の心に深く焼き付き、翌朝モスクワ河の岸辺に出たリーザは、家畜を追う牧童に、彼の姿を重ね合わせる。ここでは外界の記述のための I 群の語と共に、リーザの恥らい (段落 10) やエラストへの思い (段落 12) を描くために、かなりの量の II 群の語が使われている。

第二部と第三部を区切るのは段落 13 の、ボートに乗ったエラストの突然の登場である。

第三部では、二人が互いに愛し合っていることを知り、リーザには喜びで世界も違って見える (段落 16)。逢引は続き二人の心は満たされるが、やがて母によるリーザの婿探しを契機に、彼らは一線を越える。ここでは上記の段落 16 を除いて、外界を示す I 群はほとんど現れず、代りに感情と心を示す II、III 群が、相当な量で用いられている。

中でも突出しているのは段落 17、22 である。前者では逢引での二人の様子や会話が、かなり長い文脈 (21 文) で述べられ、II 群 9 個、III 群 7 個が現れている。一方例 10 で示す段落 22 は、二人の逢引が一線を越えた後まったく変り、彼らの関係の破局が暗示されている部分である。この段落はこの作品で一番重みをなす部分であるが、10 文の範囲に II 群 4 個、III 群 9 個が集中して現れ、作品全体の中でも特異な分布を示している。

例 10 (II 群：イタリック、III 群：下線付きイタリック)

Свидания их продолжались; но как все переменилось! Эраст не мог уже доволен быть одними невинными ласками своей Лизы — одними ее *любви* исполненными *взорами* — одним прикосновением *руки*, одним поцелуем, одними чистыми *объятиями*. Он желал больше, больше и, наконец, ничего желать не мог, — — — — Что принадлежит до Лизы, то она, совершенно ему отдавшись, им только жила и дышала, во всем, как агнец, повиновалась его воле и в *удовольствии* его полагала свое счастье. — — — — /50 43~51 15

第三部と第四部の区切りは、軍役に就くためしばらく別れることをエラストがリーザに告げる、段落 23 である。

第四部の前半ではエラストと別れるリーザの嘆きや不安が語られ、II 群が中心に使われている。ただこれ以降の段落全体にいえるのは、I、II、III 群のどの出現もかなり抑えられ、第三部までとは異なって、これらの見出し語が、段落を性格付ける的確な指標になり難いことである。例えば段落 29 は、リーザがエラストの後を追おうとするが、母のために思いとどまり、悲しみに耐える様子が描かれている所であるが、このかなり長い文脈 (20 文) でも I 群が 3 個、II 群が 4 個しか現れない。

第四部の後半では、エラストと別れたリーザが母の用事でモスクワに行った際、偶然彼に行き会うが、彼は彼女に金を渡し屋敷から追い出してしまう。そのためリーザは絶望し池に身を投げるといふ破局が描かれている。プロット自体はこのようなダイナミックな展開を見せるが、名詞表現に関しては前半の傾向を引き継ぎ、I、II、III 群の見出し語は散発的にしか現れない。

その中で絶望したりーザが池に身を投げる段落 33 は、データの的に I、II、III 群が各々 5、2、3 個出現した特異な分布である。だがここでの I 群の 4 個は、небо と земля の繰り返しであり、これらは例 11 で示すように「天が落ちたら、地が呑み込んだら」という絶望の表現として使われている。従って実際の語群の出現は 1、2、7 個とみなされ、この段落がリーザの心の状態を示す最後の山場であることが理解される⁽²⁴⁾。

例 11

〈Мне нельзя жить, — думала Лиза, — нельзя!.. О, если бы упало на меня *небо!* Если бы *земля* поглотила бедную!.. Нет! *Небо* не падает; *земля* не колеблется! Горю мне! /54 17-20

第五部のエピローグは、段落 34 から作品の終りまでに該当するが、該当する語群は散発的にしか現れない。

以上述べたことを語群の出現の側面からまとめると、表 1 のようになろう。なお語群の出現で () 内はその出現が少ないことを示す。

表 1

	段落番号	文番号	語群の出現	作品の区分
一部	1～9	1～76	I (II)	A
二部	10～12	80～141	(I) II	
三部	14～22	143～293	II III	B
四部	24～33	299～411	(II)	C
五部	34～37	412～424	(II)	

状況設定が中心の第一部とリーザの淡い恋心を描いた第二部では、語群の出現に多少の差はあるが、この二つをまとめた序盤の A 部で基本的に用いられているのは、外界を示す I 群と感情を示す II 群である。次にリーザとエラストの恋を描く中盤の B 部では、感情を示す II 群と心を示す III 群が主体となる。この A、B 部に関しては図 3 からわかるように、I、II、III の語群はテーマの展開に積極的に関わり、各段落を性格付ける的確な指標となっている。これに対して終盤の C 部では、これらの語群の出現が急に減り、これらによる性格付けが難しくなる。また中には例 11 のように、指標の変質が生じているものもある。

従って見出し語 I、II、III 群による段落の性格付けは、A 部と B 部では成功したが C 部ではあまり有効でないことがわかった。これを評価すると成功率は約 70% になるが⁽²⁵⁾、なぜ終盤の C 部でこれらの見出し語の、指標としての有効性が下がったのであろうか。

§ 5 名詞表現と動詞表現

上記の問題を検討するため例 1 のデータベースを再度運用して、A、B 部と C 部にお

る名詞と動詞の出現数を調べてみると、表2のようになる。ここから全見出し語に対する動詞の出現比（19.8%：22.8%）でも、出現する動詞と名詞の比（0.82：1.04）からも、C部で名詞の使用が減り、逆に動詞の使用が増えていることがわかる。

しかしながらこの現象は、作品を読むと直感的に感じることでもある。つまりこの作品で導入のA部からリーザとエラストが一線を越えるB部の終わりまでは、物語の語り口はゆったりと叙情的であり、彼女の感情や心情は例7～10で見られるように、いくつもの名詞を互いに組み合わせ重ね合わす、連綿とした表現の中で示される。

だがC部に入ると、事件は急速に展開して結末を迎え、ここに動詞を表現の中心とする文が連続することになる。そして上記の統計値の差異は、この文体の変化を反映したものと考えられる。次の例12で動詞を表現の中心にした例を示すが（イタリックが動詞）、完了体動詞を畳み掛けるように使う、テンポの早いこのような表現には、名詞を組み合わせた叙情的な表現の入る余地はない。

例12 (Example 12)

В один день Лиза должна *была идти* в Москву, — — — На одной из больших улиц *встретилась* ей великолепная карета, и в сей карете *увидела* она Эраста. <Ах!— *закричала* Лиза и *бросилась* к нему, но карета *проехала* мимо и *поворотила* на двор. Эраст *вышел* и *хотел* уже *идти* на крыльцо огромного дома, как вдруг *почувствовал* себя в Лизиних объятиях. Он *побледнел* — потом, не *отвечая* ни слова на ее восклицания, *взял* ее за руку, *привел* в свой кабинет, *запер* дверь и *сказал* ей: — — — Прежде нежели Лиза *могла опомниться*, он *вывел* ее из кабинета и *сказал* слуге: <Проводи эту девушку со двора>./53 21-38

以上より、作品のC部で名詞I、II、III群による指標が効力を失うのは、この指標の取り方に誤りがあったからではなく、この作品には少なくとも二種類の表現手段が使われていることによるといえよう。その一つは名詞を組み合わせた叙情的な表現で、A部からB部にかけて多く使われ、もう一つは動詞を中心とした簡潔な表現で、主にC部で使われている。そしてC部における名詞の指標の無力化は、逆に表現手段がA、B部とC部で大きく異なっていることを示す、有力な証拠と見ることもできよう。なお表現手段のこのような二分化は、この作品だけのものか、それともカラムジンによくある手法なのか、それとも他の作家でも一般的に見られる傾向なのか等の考察は、同様の手続きで他の作品を分析し比較する過程で、今後解明されていくと思われる。

また表1で示した語群の出現のパターンは、IとII、IIとIII、あるいはII群のみ、とな

表2

	全見出し語	動詞	動詞出現比	名詞	動詞/名詞
A・B	3678	729	19.8%	884	0.82
C	1369	313	22.8%	299	1.04

る。これはこの作品が、感情を示すII群の語彙を表現の中核として用い、外界や心の表現にも常に感情に関わるII群の語彙に関連させていることを意味しよう。このことは、外界や人の心を常に感情というプリズムを通して描いたとされる、センチメンタリズムの手法そのものに対応する図式でもある。

最後にこの研究では例1で示したデータベースに、分析の目的に応じて四つの異なる側面からアクセスし、それぞれ性質の異なるデータを引き出している。それらは、

- (1) テキスト中の各出現語形のレベルで、個々の変化語形の代りに、対応する見出し語を使い語彙統計を取った。(§1)
 - (2) 語彙統計から得られた高頻度の26個普通名詞について、それらの共出現を文レベルでチェックし、数量化3類の解析のためのデータを作った。(§3)
 - (3) 共出現の解析で得た、見出し語の三つのグループ化が妥当なものかどうかを検証するため、段落レベルでこれら三つの語群の出現をチェックし、得られた語群別集計の分布から各段落の性格付けを試みた。(§4)
 - (4) 指標としての三つの語群の有効性が、作品の終盤で下がる原因を探るため、作品を序・中盤と終盤に分け、それら二つの部分での名詞と動詞の出現数をチェックした。(§5)
- 上記の作業とその結果から、このデータベースは単なる語彙統計だけでなく、文学作品における表現の分析においても、相当に有効なものであることが示せたと思われる。

本稿は1993年度の日本ロシア文学会で発表した原稿に、藤沼貴、丹辺文彦両氏の貴重な御指摘を得て、一部修正を加えたものである。ここで両先生に心からの謝意を表したい。

— 注 —

- 1 岡本哲也・山本佳代子・浦井康男「ロシア語の形態生成と Prolog 系」(I) 一名詞・形容詞・代名詞一、(II) 一動詞一『電気通信大学紀要』第6巻第1、2号、1993年、29-38頁、157-169頁。
- 2 現在のところこの作品の厳密なテキスト校正版は存在しないので、OCRによる読み取りには、比較的印刷のきれいな Н.М. Карамзин, *Записки старого московского жителя*, Москва, 1986, стр. 41-55. を使った。ただし段落の区切りについては、Н.М. Карамзин, *Сочинения*, изд. 3-е, Москва, 1820, т.6. に従った。
- 3 WF (Word Form) はテキスト中の個々の語形、HI は品詞、H2 は品詞補足情報、REF (Reference) はその語形がテキスト中の～頁～行目の前から～番目に出現したかを、SNO (Sentence Number) は作品の初めから数えた文の番号、PNO (Paragraph Number) は段落の番号を示す。なおこのデータはリレーショナル・データベース dBXL で運用されている。
- 4 実際には例1のデータベースに、HI と LEX2 フィールドで多重インデックスをかければよい。
- 5 ちなみにこのデータでは頻度10の свет2 は「世界」を、一方 свет1 は「光」を示す。

なお同形異義語の識別番号の割り当ては、A.A. Зализняк, *Грамматический словарь русского языка*, Москва, 1977, стр. 878. に準拠している。

- 6 同形異義識別番号はいずれも 1 なので省略して示す。
- | | |
|--------------|--------------------|
| 67 : быть | 19 : хотеть |
| 43 : мочь | 16 : говорить |
| 30 : сказать | 15 : думать, знать |
| 21 : любить | 12 : жить |
- 10 : взять, видеть, отвечать, смотреть
 8 : идти, пойти
 7 : бывать, взглянуть, дать, казаться, плакать, прийти, сидеть, стоять, увидеть
 6 : бояться, возвратиться, поцеловать, умереть 等
- 7 この値は動詞の体の対を考慮していないので、これを加味すればもう少し異なった様子を示すかもしれないが、現在のところ未処理である。
- 8 形容詞については、浦井康男「数量化理論によるエピテットの性格付けの試み—ロシア語形態処理の発展—」『福井大学教育学部紀要』第 I 部人文科学第 46 号、1993 年、21~33 頁で数量化 4 類による分析を試みている。
- 9 水谷静夫『語彙』朝倉日本語新講座第 2 巻、朝倉書店、1983 年、121 頁。
- 10 例えば、岩坪秀一『数量化法の基礎』朝倉書店、1987 年、254 頁、林知己夫、鮑戸弘『多次元尺度解析法』サイエンス社、昭和 57 年、290 頁等を参照。
- 11 本研究では Lotus 1-2-3 MS-DOS 版 (ロータス) と、そのアドインソフトの Lotus 1-2-3 多変量解析 (オードマン) を使用した。
- 12 現在使用しているソフトでは、データ転送に際して一行 240 バイト以下という制限があり、カテゴリ数が多すぎると見出し語のタイトルの転送ができなくなる。
- 13 同一文中で同じ語が 2 回出現したのは、чувство (文番号 14), монастырь (同 17), свет (34), жизнь (209), душа (250), луч (363) の 6 カ所であり、分析全体に影響は与えないと思われる。なお同じ語が 3 回以上出現する場合は無かった。
- 14 3 次元は他のものは +2 ~ -2 の間にあるが、мысль は +5.2 の値を取っている。
- 15 мысль を除く他の語は 1 次元、2 次元とも +2 ~ -2 の間におさまるが、мысль は 1 次元で +4 となる。なお 3 次元は全語が +2 ~ -2 の範囲内にあった。
- 16 3 次元の値は лицо が +3.5 で、他は +2 ~ -2 の範囲内にあった。
- 17 слово は意義的には若干ずれているが、共出現関係からここに評価された。
- 18 別の表現では、III 群が内的感情を示すのに対して、II 群は外的感情の表現に用いられているともいえる。
- 19 数字は注 2 のテキストでの頁・行を示す。
- 20 мысль が特異点になるのは、第 5 節と関連するが、この作品ではこの語が、動詞的な意義で使われているためと思われる。
- 21 3 次元の値が高いために除いた лицо は、意義的には II 群に近いものだが、たまたま上記 25 個の見出し語以外の語との結合が多かったため (На лице и во всех ее движениях обнаруживалась сердечная радость./48 06 等)、II 群とは少し離れた位置に

評価されたと考えられる。これをⅡ群に組み込むことも可能だが、ここではデータの客観性を保つために排除した。

- 22 見出し語が約1,300となっているのは、ни с чем、друг с другом等の2つに分離される語の処理が未確定なためである。
- 23 段落番号はタイトル部を0番とした。また会話等での文の区切りは発話内容によらず、形式的に大文字で始まる所に設定した。
- 24 藤沼貴氏はここに作品の冒頭との対応（自然の枠作り）を見ている。
- 25 A、B部の文章数／全文章数(293/424)×100。

A Study on Expressions of Common Nouns in Karamzin's "Poor Liza" — Data Base and Multivariate Analysis —

Yasuo URAI

We have performed a morphological analysis of word-forms in the text, and then produced data supplied with lexical items and morphological information. Example 1 comprises data obtained by a processing of Karamzin's "Poor Liza" in the manner described above.

On the basis of these data lexical statistics were compiled for each part of speech separately, and for the nouns (Example 2), we have selected 26 common nouns with high frequency of occurrence and concrete meaning (сердце, глаз, рука, слеза, душа, смерть, цветок, жизнь, земля, свет, любовь, небо, слово, хижина, чувство, берег, взор, луч, лицо, луг, монастырь, мысль, натура, объятие, река, удовольствие) and examined their contextual co-occurrence.

We have processed these data on the basis of Multivariate Analysis (Hayashi's Quantification Theory Type 3) and evaluated the co-occurrence proximity of these common nouns. Then the 18 nouns were classified into three sets as is shown in Figure 2. We assume that these sets of lexemes are related to objective environment (цветок, небо, луг, луч, земля, монастырь, свет), sentimental atmosphere (глаз, рука, объятие, слеза, взор, сердце) and mental phenomena (душа, удовольствие, слово, любовь, чувство).

Further, taking these data into consideration, we produced data obtained by calculating the total number of occurrences for each lexical items in the individual sets and for each paragraph. Then we tried to judge the character of each paragraph and all the work from the pattern of distribution as shown in Figure 3.

In the initial sections (I,II) and the middle section (III), the number of occurrences is sufficient to make the calculated numbers operate as parameters which reflect sensitively the content of the paragraph analyzed and the data seem sufficiently representative. On the other hand, we have found that in the final sections (IV,V) where the occurrence drops and the efficiency of these parameters decreases, different means of expression, which are essentially based on use of verbs, prevail (Example 12).

Finally, if we examine the co-occurrence pattern for each of the sets of lexemes (Figure 3), we find that, although there is more or less variability between the five sections of the work, the general tendency is as follows. Occurrence of the lexemes which are related to sentimental atmosphere is observed also in the paragraphs containing the lexemes which are related to the semantics of objective environment or to mental phenomena. This scheme is probably typical of the method of sentimentalism, which is supposed to describe the objective environment and the mental state of human heart always by means of the prism of sentiment.