

クラスター分析による イタリア語定冠詞の時系列的考察

上 野 貴 史
谷 岡 弘 二

1. はじめに

俗ラテン語における ILLE (男性)・ILLA (女性)・ILLUD (中性) といった指示詞から派生したイタリア語の定冠詞は、現代語において <表 1> のような体系として確立している²⁾。<表 1> のように現代語で確立している定冠詞の形態は、Migliorini (2000:263) の "sono molto frequenti le forme *el*, plur. *e*; un po' meno *il*, plur. *i*; *lo*, plur. *li e gli*, perde terreno" 「*el* とその複数 *e* は大変頻繁に用いられ、*il* とその複数 *i* はそれより少し劣り、*lo* とその複数 *li/ gli* は劣勢である」といった 15 世紀に関する記述からも明らかのように、通時的変遷の中で多くの揺れの形態を生じる。イタリア語史における定冠詞の形態に関して、先行研究では、Migliorini にあるように、ある時代にある形態が使用されたといったような記述は見られるが、それぞれの使用形態の計量的分布や各時代間の通時的変遷についての指摘は欠如していると思われる。

<表 1: 現代語の定冠詞体系>

性	単数形態	複数形態
男性	<i>il/ lo (l')</i>	<i>i/ gli</i>
女性	<i>la (l')</i>	<i>le</i>

そこで本稿では、各時代の文学作品をコーパスとし、各コーパスにおける定冠詞の計量的調査を行うことにより、各時代の定冠詞の量的分布を明らかにしていく。そしてさらに、数値化した定冠詞の使用割合のデータを基に、多変量解析の一つであるクラスター分析を用いてイタリア語定冠詞における時系列的な推移パターンの考察を行うことにする。

2. コーパスに関して

コーパスに関しては、地理的には、現代イタリア語の根幹を成しているトスカーナ方言で書かれたものに限定し、さらに質的には、散文の中で極端に口語調のものを避けることに留意しながら選定する。散文作品が口語調であるかどうかという判断は、D'Achille (1990) で示される語用論的な内部変数を使った質的分類において、Livello C (格調ある文語で書かれたもの) という評価を与えられたものを用いることにする。イタリア語史の時代区分としては、D'Achille (1990) で示されたものを応用することとする³⁾。

以上のことを条件として、本稿では <表 2> に示す文学作品をコーパスとして選定し、各コーパスから約 20,000 語を収集する⁴⁾。

<表2：コーパス>

時代区分	著者	作品名	コーパスの範囲
第Ⅱ期 1250～1375	Dante Alighieri (1265-1321)	<i>Convivio</i> (1304-1307)	Convivio Iより
	Giovanni Boccaccio (1313-1375)	<i>Decameron</i> (1349-1353)	Giornata Prima すべて
第Ⅲ期 1375～1525	Alberti Leon Battista (1404-1472)	<i>I Libri della famiglia</i> (1432-1443)	Libro Terzoより
	Niccolò Machiavelli (1469-1527)	<i>Il Principe</i> (1513)	すべて
第Ⅳ期 1525～1612	Giovanni Della Casa (1503-1556)	<i>Galateo</i> (1551-1555)	すべて
	Giorgio Vasari (1511-1574)	<i>Le Vite</i> (1568)	Parte Primaより
第Ⅴ期 1612～1840	Galileo Galilei (1564-1642)	<i>Dialogo sopra i due massimi sistemi del mondo</i> (1632)	Giornata Primaより
	Giacomo Leopardi (1798-1837)	<i>Operette Morali</i> (1824)	冒頭より
第Ⅵ期 1840～現在	Italo Calvino (1923-1985)	<i>Il castello dei destini incrociati</i> (1973)	すべて

3. 各コーパスにおける定冠詞の体系

各コーパスにおける定冠詞の体系と量的分布は、選定したコーパスの中に出現する定冠詞を計量的に調査していくことにより確認することができる。例えば、現代イタリア語のCalvinoコーパスからは、<表3>⁹⁾のような調査結果が得られる。

<表3：Calvinoコーパスにおける出現頻度と使用割合>

イタリア語定冠詞の形態は、後続する名詞の文法性と数、これに加えて後続語の語頭音の音韻的条件から決定される。例えばCalvinoコーパスでは、

性	後続音	単数形態	頻度数	使用割合	複数形態	頻度数	使用割合
男性	子音	il	635例	1.0000	i	207例	1.0000
		s impura	lo	41例	1.0000	gli	22例
	母音	l'	161例	1.0000	gli	40例	1.0000
					gli	7例	1.0000
女性	子音	la	579例	1.0000	le	293例	1.0000
	母音	l'	99例	1.0000	le	30例	1.0000

後続する名詞が男性単数で、後続語の語頭音がs impura⁹⁾以外の子音であるという条件で、ilという形態が635例使用されている。このように、ある条件で形態的な揺れを生じない場合の使用割合を1.0000と定める。現代語であるCalvinoコーパスでは、使用割合がすべて1.0000であることから、形態的揺れが全く存在しないことが分かる。

一方、形態的揺れが多く見られるDanteコーパスでの調査結果を<表4>⁷⁾に示す。Danteコーパスでは、一つの後続音に関する条件内で複数の形態が出現していることから、多くの形態的揺れを確認することができる。例えば、後続する名詞が男性単数で、後続語の語頭音がs impura以外の子音という条件においては、loが0.8668、l'が0.1106、ilが0.0226という使用割合で出現している。実際の使用例においても、fuoco「火」という名詞に対して、(1)のような三種類の定冠詞の使用が確認できる。

<表4：Danteコーパスにおける出現頻度と使用割合>

性	後続音	単数形態	頻度数	使用割合	複数形態	頻度数	使用割合	
男性	子音	lo	345例	0.8668	li	136例	0.9927	
		l'	44例	0.1106	i	1例	0.0073	
		il	9例	0.0226				
	s impura	lo	12例	1.0000	li	2例	1.0000	
		l'	156例	0.8211	li	82例	0.9762	
		lo	34例	0.1789	gli	2例	0.0238	
女性	母音				li	1例	1.0000	
		子音	la	742例	1.0000	le	188例	1.0000
		母音	l'	154例	0.8652	l'	36例	0.8182
		la	24例	0.1348	le	8例	0.1818	

(1) lo fuoco (I,XIII:164-4) / l' fuoco (I,XIII:164-4) / il fuoco (I,XII:162-2)

Dante コーパス内に見られるような共時的な形態的揺れに加えて、イタリア語定冠詞は、コーパス間や時代間といった通時的な使用形態の相違も多く見られる。この通時的変化について、第Ⅲ期に属する Machiavelli コーパスの調査結果（<表5>）を示しながら、後続語の語頭音が *s impura* 以外の子音である男性複数定冠詞について考察してみる。

<表5：Machiavelliコーパスにおける出現頻度と使用割合>

まず、第Ⅱ期の Dante コーパス（14世紀）においては、*li* が 0.9927、*i* が 0.0073 という使用割合を示しており、僅かな揺れがあるものの、ほぼこの条件において *li* が義務的に使用されている。これに対して、第Ⅲ期の Machiavelli コーパス（16世

性	後続音	単数形態	頻度数	使用割合	複数形態	頻度数	使用割合	
男性	子音	<i>el</i>	193例	0.5376	<i>e</i>	199例	0.7263	
		<i>il</i>	166例	0.4624	<i>li</i>	71例	0.2591	
	<i>s impura</i>		<i>lo</i>	36例	1.0000	<i>gli</i>	4例	0.0146
						<i>li</i>	18例	0.8182
		母音				<i>gli</i>	2例	0.0909
						<i>e</i>	2例	0.0909
	母音	母音	<i>l'</i>	88例	0.6241	<i>li</i>	76例	0.5352
			<i>lo</i>	53例	0.3759	<i>gli</i>	66例	0.4648
		i				<i>gli</i>	4例	0.5714
						<i>li</i>	3例	0.4286
女性	子音	<i>la</i>	326例	1.0000	<i>le</i>	183例	1.0000	
	母音	<i>l'</i>	65例	0.6019	<i>le</i>	59例	0.8806	
		<i>la</i>	43例	0.3981	<i>l'</i>	8例	0.1194	

紀)においては、*e* が 0.7263、*li* が 0.2591、*gli* が 0.0146 というような三種類の形態が使用されており、その中でも *e* という形態が高頻度で使用されている。しかしながら、第Ⅵ期の Calvino コーパス（20世紀）においては、第Ⅱ期や第Ⅲ期で優位であった *li* や *e* という形態は消滅し、*i* という形態が確立した体系として使用されている。

このようにイタリア語定冠詞は、共時的に形態的揺れが存在し、さらに通時的にも多くの形態変化を被っていることになる。このようなイタリア語定冠詞に関して、次に、各コーパスにおける定冠詞体系を数値化し、多変量解析の一つであるクラスター分析を用いて現代語で確立した体系に至る時系列的变化を考察していくことにする。

4. 多変量解析による言語の時系列的变化

複数の変数がそれぞれ複数のサンプルを取るようなデータを多変量データと言い、この多変量データを統計的に処理する手法の総称を多変量解析と呼ぶ。多変量解析には、回帰分析・判別分析・主成分分析・クラスター分析・多次元尺度法・正準相関分析など様々な手法があるが、言語研究においては、クラスター分析や主成分分析などが比較的良好に文体分析・作者の真贋判別・言語類型論などに利用されている⁸⁾。

本稿では、共時的に揺れが見られる定冠詞形態の使用割合を数値化することによって、多変量解析を行うことになるが、このような時系列的な推移に対しては、多数の変量間の関係を一覧でき、なおかつ変量間の関係をいくつかのグループに分割して、その類似度を客観的に分析することのできるクラスター分析が有効であると思われる。

分析対象間の距離に基づき分析対象を分類する手法であるクラスター分析は、非階層的方法と階層的方法に大別することができる。本稿では、変数間の距離関係が解釈しやすい凝集法による階層的クラスター分析により樹状図 (tree diagram) を作成し、クラスター

間の関係をより明確に示すことにする。分析対象間の距離の測定方法に関しては、「原データのユークリッド距離」を採用し、その距離計算の方法には、ウォード法を用いる。ウォード法は任意の2つのクラスター *A* と *B* について、これを一つのクラスターにまとめたとき（クラスター *C* とする）の残差の増分を最小にする方法である。クラスター *A* と *B* が合併された後の残差の増分を Z_c とすると、 Z_c は(2)で示す式で表すことができる⁹⁾。

$$(2) \quad Z_c = n_A \sum_{j=1}^p (\bar{y}_{j(A)} - \bar{y}_{j(C)})^2 + n_B \sum_{j=1}^p (\bar{y}_{j(B)} - \bar{y}_{j(C)})^2$$

$$= \frac{n_A \cdot n_B}{n_A + n_B} \sum_{j=1}^p (\bar{y}_{j(A)} - \bar{y}_{j(B)})^2$$

Z_c を任意の二つのクラスター *AB* 間の距離（非類似度）の指標として規定しておけば、多数のクラスターが存在している母集団において、 Z_c が最小となるクラスターを合併していけば良いことになる。この手法を繰り返すことにより、最終的には母集団全体のクラスター間の距離が計算され、それに基づいて樹状図が作成されることになる。

5. クラスター分析によるイタリア語定冠詞の解析

定冠詞の時系列的変化をクラスター分析するため、選定したコーパスにおける定冠詞の形態と使用割合を調査したものが表6¹⁰⁾となる。

<表6：各コーパスにおける定冠詞の形態と使用割合>

性	数	後続音	形態	Dante	Boccaccio	Alberti	Machiavelli	Della Casa	Vasari	Galileo	Leopardi	Calvino		
男性	単数	子音	<i>il</i>	0.0226	0.9301	0.9471	0.4624	0.9025	0.9715	0.9140	0.9717	1.0000		
			<i>'l</i>	0.1106	0.0233	0.0220	0.0000	0.0220	0.0026	0.0762	0.0000	0.0000		
			<i>el</i>	0.0000	0.0000	0.0308	0.5376	0.0000	0.0000	0.0000	0.0000	0.0000		
			<i>lo</i>	0.8668	0.0466	0.0000	0.0000	0.0755	0.0259	0.0098	0.0283	0.0000		
			<i>l'</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
			<i>lo</i>	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
		母音	<i>l'</i>	0.8211	0.8854	0.8854	0.6241	0.8062	1.0000	1.0000	1.0000	1.0000		
			<i>lo</i>	0.1789	0.1146	0.1146	0.3759	0.1938	0.0000	0.0000	0.0000	0.0000		
			複数	子音	<i>li</i>	0.9927	0.3457	0.0000	0.2591	0.0229	0.0000	0.0264	0.0288	0.2591
					<i>gli</i>	0.0000	0.0370	0.0000	0.0146	0.0000	0.0000	0.0000	0.0144	0.0000
					<i>i</i>	0.0073	0.6049	0.2917	0.0000	0.9771	1.0000	0.9736	0.9568	1.0000
					<i>e/e'</i>	0.0000	0.0123	0.7083	0.7263	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>li</i>	1.0000			0.0000	0.0909	0.8182	0.0000	0.0000	0.0000	0.0000	0.0000		
	<i>e</i>	0.0000			0.0000	0.0000	0.0909	0.0000	0.0000	0.0000	0.0000	0.0000		
	母音	<i>gli</i>		0.0000	1.0000	0.9091	0.0909	1.0000	1.0000	1.0000	1.0000	1.0000		
		<i>li</i>		0.9762	0.0000	0.0000	0.5352	0.0000	0.0000	0.0000	0.0000	0.0000		
		<i>gli</i>		0.0238	1.0000	1.0000	0.4648	1.0000	1.0000	1.0000	1.0000	1.0000		
		<i>li</i>		1.0000	0.0000	0.0000	0.4286	0.0000	0.0000	0.0000	0.0000	0.0000		
		<i>gli</i>		0.0000	0.0000	0.0000	0.5714	0.5000	0.3333	0.1667	0.1000	1.0000		
		<i>gl'</i>		0.0000	1.0000	1.0000	0.0000	0.5000	0.6667	0.8333	0.9000	0.0000		
	女性	単数	子音	<i>la</i>	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
				<i>l'</i>	0.8652	0.8605	0.7143	0.6019	0.8478	0.9697	0.9091	1.0000		
			母音	<i>la</i>	0.1348	0.1395	0.2857	0.3981	0.1522	0.0303	0.0909	0.0000		
		複数	子音	<i>le</i>	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		
<i>l'</i>				0.8182	0.9545	0.8519	0.1194	0.1600	0.9492	0.3158	0.1579			
母音			<i>le</i>	0.1818	0.0455	0.1481	0.8806	0.8400	0.0508	0.6842	0.8421			

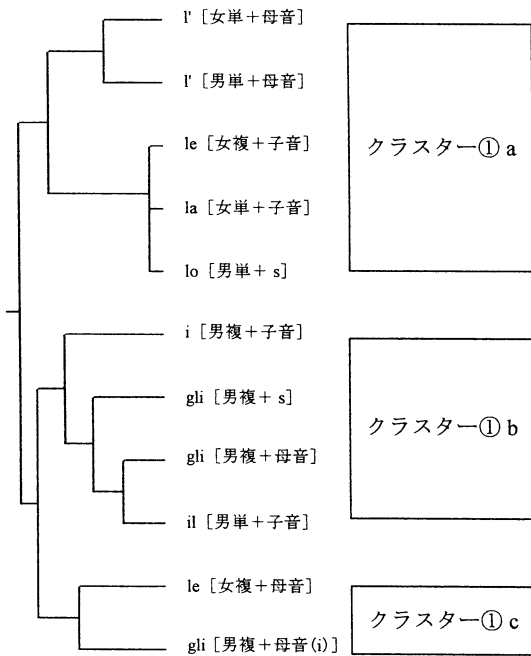
表中のゴシック（イタリック）で示したものは、現代語で確立している形態とデータである。本稿は、定冠詞が現代語の体系として確立（衰退）していく過程の時系列的推移を考

察するものである。そこで、Calvino コーパスで 1.0000 を示す現代語で確立している形態（現代語確立型）と、0.0000 を示す現代語で衰退した形態（現代語衰退型）とに区分してクラスター分析を行うことにする。

5.1. 現代語確立型のクラスター分析

現代語で確立している形態について、クラスター分析を行った結果が<図1>である。先述のように、クラスター分析は分析対象の距離に基づき、分析対象を分類する手法である。ここでの分析対象間の距離とは、変数またはサンプルの類似度を意味し、分析対象間の距離（類似度）に近いほど図の右側で分析対象同士が併合される形で示される。<図1>

<図1：現代語確立型のクラスター分析¹¹⁾>



で得られた樹状図から、現代語確立型は、3つのクラスターに分類することが可能であると考えられる。

まず、クラスター① a には、l' [女単+母音]・l' [男単+母音]・le [女複+子音]・la [女単+子音]・lo [男単+s] が含まれる。クラスター① a は、最も古い Dante コーパスから現代の Calvino コーパスまで、ほぼ安定した状態で優位に使用されているという特徴がある。l' [男単+母音] と l' [女単+母音] は、第Ⅲ期の時期に使用頻度が落ちるが、それでも使用割合は60%を超えている。

クラスター① b は、Dante コーパスでの使用がほとんど皆無であったものが、Boccaccio・Alberti コーパスでの使用

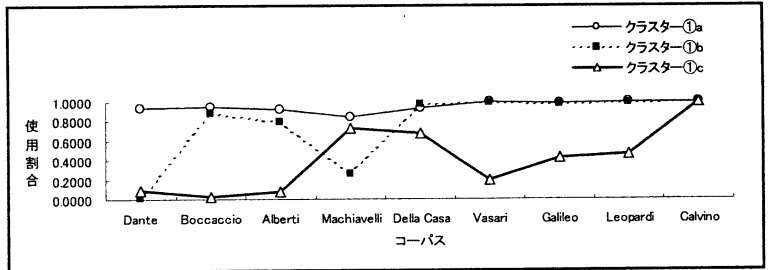
の義務化と、Machiavelli コーパスでの使用頻度の落ち込みを経て、第Ⅳ期以降に定冠詞の体系として確立するものである。これには、i [男複+子音]・gli [男複+s]・gli [男複+母音]・il [男単+子音] が含まれる。

クラスター① c は、第Ⅱ期で低い使用割合であったものが、Machiavelli・Della Casa コーパスで使用頻度が上がり、その後一旦使用率が落ち込むが、復活するようなかたちで徐々に現代語で確立していく le [女複+母音] と gli [男複+母音(i)] という形態である。ある意味、クラスター① b よりゆっくりとしたテンポで現代語として確立していくものであると言える。

この3つに分類したクラスターの時系列的推移について、各クラスターにおける各コー

パスの平均した使用割合を時系列順にグラフ化したものが<グラフ1>である。現代語で体系として確立している定冠詞は、<グラフ1>のような3つの典型的な時系列的推移を経て成立していることになる。

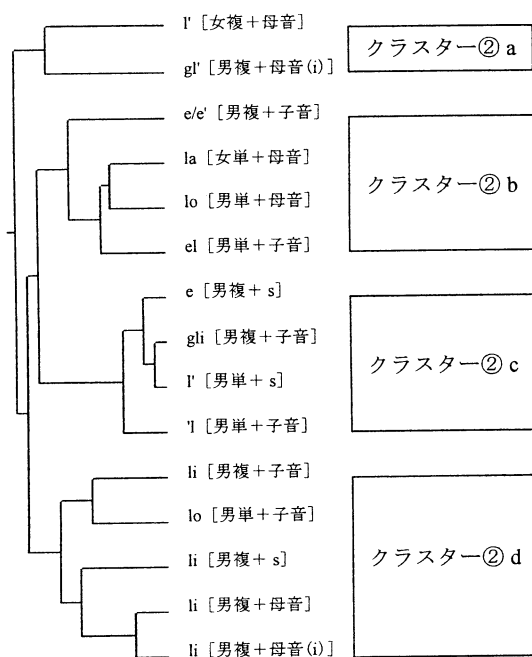
<グラフ1：現代語確立型のクラスター別時系列的変化>



5.2. 現代語衰退型のクラスター分析

現代語で衰退した形態のクラスター分析の結果は、<図2>のようになる。<図2>の樹状図より、現代語衰退型の定冠詞形態は4つに分類することが適当であると考えられる。

<図2：現代語衰退型のクラスター分析>



クラスター② aに含まれる l' [女複+母音] と gl' [男複+母音(i)] は、各コーパスにおける使用割合が極端な数値となって現れているものである。l' [女複+母音] は、Dante・Boccaccio・Alberti・Vasari コーパスでは義務的に使用されているが、それ以外のコーパスではほとんど用いられていない。また、gl' [男複+母音(i)] も時期は異なるが、同様の指摘が可能である。クラスター分析の結果においては、l' [女複+母音] と gl' [男複+母音(i)] がクラスター図のかかなり左側で併合されている。このことは、当該変数同士の類似度が低いことを示しているが、いずれの形態も使用割合が極端な数値を示すという特徴があり、他の変数や

グループとの関係からすると、同じクラスターとして見なすことができると思われる。

クラスター② bは、第Ⅲ期に高い使用割合となる e/e' [男複+子音]・la [女単+母音]・lo [男単+母音]・el [男単+子音] という形態が含まれる。

クラスター② cは、各コーパスで15%以下の低頻度で使用される形態である。これには、e [男複+s]・gli [男複+子音]・l' [男単+s]・l' [男単+子音] がある。

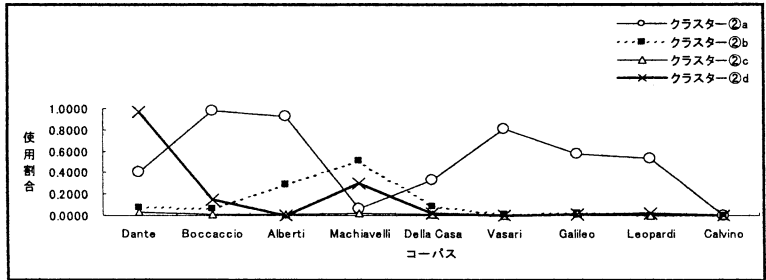
クラスター② dは、Dante コーパスでの義務的使用の後、急激に衰退し、Machiavelli コーパス

ーパスで多少回復するが、その後ほとんど使用されなくなる形態である。これには、li [男複+子音]・lo [男単+子音]・li [男複+s]・li [男複+母音]・li [男複+母音(i)] が含まれる。

現代語衰退型の4つのクラスターにおける

各コーパスの平均した使用割合を時系列的にグラフ化したものが<グラフ2>である。現代語で衰退した定冠詞は、<グラフ2>のような時系列的推移を経て現代語で消滅していることになる。

<グラフ2>：現代語衰退型のクラスター別時系列的变化>



6. 結語

最後に、本稿で行った分析の統計的妥当性を判別分析と正準判別分析を用いて検証しておく。

<表7>：判別分析>

まず、本稿では、現代語確立型（クラスター①）と現代語衰退型（クラスター②）という区分によりクラスター分析を行ったが、この前提とした区分が妥当であるかを判別分析を用いて検証することにする。目的変数をカテゴリーデータ（分類データ）として扱う判別分析は、2群に分類されるデータの識別の予測や検証に使用されるものである。

この判別分析について、有意水準 1%レベルでの検証を行った結果は、<表7>に示す通りである¹²⁾。表中の「判別得点」は、定冠詞の各項目について判別関数（ク

形態	クラスター	判別得点	マハラノビスの汎距離	確率
lo [男単+s]	①a	60.53337	2.4242	100
la [女単+子音]	①a	60.53337	2.4242	100
le [女複+子音]	①a	60.53337	2.4242	100
gli [男複+母音(i)]	①c	39.70588	9.0909	100
il [男単+子音]	①b	48.40688	9.0909	100
l' [男単+母音]	①a	34.84707	9.0909	100
i [男複+子音]	①b	76.45117	9.0909	100
gli [男複+s]	①b	47.13303	9.0909	100
gli [男複+母音]	①b	52.65998	9.0909	100
l' [女単+母音]	①a	51.26588	9.0909	100
le [女複+母音]	①c	53.22215	9.0909	100
lo [男単+子音]	②d	-50.10950	343.0506	0
e/e' [男複+子音]	②b	-58.13260	488.4406	0
li [男複+母音(i)]	②d	-52.83420	791.5997	0
li [男複+s]	②d	-46.94150	967.1282	0
li [男複+母音]	②d	-50.70660	1022.1917	0
la [女単+母音]	②b	-49.31250	1080.7005	0
lo [男単+母音]	②b	-32.89370	1133.8665	0
l' [女複+母音]	②a	-51.26880	1160.7460	0
l' [男単+s]	②c	-58.58000	1182.2976	0
e [男複+s]	②c	-56.81830	1402.4843	0
gli [男複+子音]	②c	-59.67250	1609.3375	0
l' [男単+子音]	②c	-64.78020	1633.3326	0
gl' [男複+母音(i)]	②a	-43.49840	1638.7643	0
el [男単+子音]	②b	-48.72390	2483.0248	0
li [男複+子音]	②d	-73.85280	6401.5985	0

ラスター①かクラスター②かに区分するための関数式) に代入したときの値であり、「マハラノビスの汎距離」はクラスター①に属する全データの重心からの距離を表す。分析の結果、クラスター①に分類されたものは、判別得点の値が正になり、マハラノビスの汎距離の値が 10.0000 以内に収まるのに対し、クラスター②に分類されたものは、判別得点が

負の値になり、マハラノビスの汎距離の値は約 340 ~ 6400 という非常に大きな値を取る。このような判別得点から、クラスター①に分類される場合の確からしさの度合いを示したものが「確率」である。クラスター①に属するものならば理想値として 100%となり、クラスター②に属するものならば 0%となる。〈表 7〉の数値は、クラスター①と②の分類に関して理想的な値を示している。このような判別分析の結果から、現代語確立型（クラスター①）と現代語衰退型（クラスター②）という区分により考察を行うということは、統計的に妥当性があると言える。

〈表 8：正準判別係数〉

コーパス	軸1	軸2	軸3
Dante	-0.3212	8.4563	-3.0658
Boccaccio	-0.2855	1.2231	5.0015
Alberti	3.9122	0.9978	0.5476
Machiavelli	-4.3314	-2.3513	-1.8130
Della Casa	8.0188	-11.8389	-11.4508
Vasari	-133.0863	-0.6523	0.2050
Galileo	25.9568	-0.1983	9.7345
Leopardi	-72.7950	8.5652	-3.9901
定数	73.1634	-1.7775	2.0439
正準相関	0.9999	0.9785	0.9484
Bartletの χ^2 値	265.04	106.86	51.63
自由度	48	35	24
P 値	0.0000	0.0000	0.0009

さらに、本稿で扱った 26 個の定冠詞の形態を 7 群のクラスターに区分したことの妥当性を検証するために、3 つ以上の群を判別する場合に用いられる正準判別分析を行う¹³⁾。ここでの正準判別分析は、Calvino コーパスを除く 7 つのコーパスにおける群が検証の対象となるので¹⁴⁾、6 (= 7 - 1) つの判別得点（軸）¹⁵⁾を求めることになるが、正準判別係

数に対する Bartlet の χ^2 検定の結果から、軸 1 ~ 軸 3 までの 3 軸で正準判別分析を行うこととする¹⁶⁾。

〈表 9：正準判別分析〉

定冠詞	軸1	軸2	軸3	実績群	推定群
l' [女単+母音]	-98.6253	3.6035	-1.6549	①a	1
r [男単+母音]	-100.0522	3.8344	0.0262	①a	1
le [女複+子音]	-99.7682	2.4239	-2.7872	①a	1
la [女単+子音]	-99.7682	2.4239	-2.7872	①a	1
lo [男単+s]	-99.7682	2.4239	-2.7872	①a	1
i [男複+子音]	-95.5034	-4.9026	-0.1172	①b	2
gli [男複+s]	-95.8651	-3.9855	1.8770	①b	2
gli [男複+母音]	-97.1365	-4.5726	1.1760	①b	2
il [男単+子音]	-94.4708	-3.7683	1.1916	①b	2
le [女複+母音]	26.2849	-5.0080	-6.1094	①c	3
gli [男複+母音(i)]	27.3822	-8.4346	-3.4257	①c	3
l' [女複+母音]	-52.8897	5.6543	5.3662	②a	4
gl' [男複+母音(i)]	-51.8096	1.6324	6.5253	②a	4
e/e' [男複+子音]	72.7852	-2.7634	1.1768	②b	5
la [女単+母音]	72.0205	-2.9572	0.9117	②b	5
lo [男単+母音]	73.4474	-3.1880	-0.7695	②b	5
el [男単+子音]	70.9554	-3.0109	1.0861	②b	5
e [男複+s]	72.7696	-1.9913	1.8791	②c	6
gli [男複+子音]	72.0421	-1.6433	2.1453	②c	6
r [男単+s]	73.1634	-1.7775	2.0439	②c	6
l' [男単+子音]	75.0161	-1.0696	2.3236	②c	6
li [男複+子音]	70.3980	6.4006	0.1397	②d	7
lo [男単+子音]	68.2211	4.9401	-1.2567	②d	7
li [男複+s]	69.6540	4.8456	-2.4554	②d	7
li [男複+母音]	70.5316	5.2189	-1.9192	②d	7
li [男複+母音(i)]	70.9859	5.6710	-1.7989	②d	7

まず、各定冠詞の判別得点を計算するために、各コーパスにおける正準判別係数を求めたものが〈表 8〉である。この〈表 8〉の値に基づいて、26 個の各定冠詞毎の軸別の値を算定したものが〈表 9〉となる¹⁷⁾。〈表 9〉中の「実績群」はクラスター分析で実際に区分された群を表し、「推定群」というのは、軸 1、軸 2、軸 3 という優先順に判別得点における平均の乖離の度合いから判定される群を示している。〈表 9〉に見られるように、実績群と推定群のグループ区分が一致することから、本稿で扱った定冠詞を 7 つのクラスターに区分することの統計的妥当性が検証されたことになる。

引用文献

- Dante Alighieri, *Le Opere di Dante*. Nella Sede della Società, 1960.
Giovanni Boccaccio, *Decameron, Filocolo, Ameto, Fiammetta*. Ricciardi, 1952.

- Leon Battista Alberti, *Opere volgari*. Laterza, 1960.
 Niccolò Machiavelli, *Opere*. Ricciardi, 1954.
 Giovanni della Casa, *Opere*. Ricciardi, 1960.
 Giorgio Vasari, *Le opere*. Casa Editrice le Lettere, 1998.
 Galileo Galilei, *Opere*. Ricciardi, 1953.
 Giacomo Leopardi, *Opere*. Classici UTET, 1977.
 Italo Calvino, *Il castello dei destini incrociati*. Einaudi, 1973.

註

- 1) イタリア語の定冠詞は、後続する名詞の性と数に加えて、後続語の音韻的条件により形態を使い分ける。後続語の語頭が子音である場合、基本的な男性形態は、il (単数) と i (複数) であるが、後続の名詞もしくは形容詞の語頭が ct-/ ft-/ mn-/ gn-/ pn-/ ps-/ pt-/ s + 子音/ x-/ z-/ 半母音 [j] であるとき (このような子音を s impura 「不純な s」と呼ぶ) は、lo (単数) と gli (複数) となる。また、後続語の語頭音が母音である場合、単数形態は lo の語末母音が切断された l、複数形態は gli が使用される。女性単数においても、後続語の語頭音が母音である場合は、la の語末母音を切断した l という形態が使用される。
- 2) 現代語においても、形態的揺れが指摘されることがある。例えば、Lepschy & Lepschy (1988:103) では、母音で始まる女性名詞が後続する場合に、"In modern Italian there is a tendency to use the full form of the article in front of a vowel" という指摘を行っており、la amica 「女友達」という例を挙げている。
- 3) D'Achille (1990) の時代区分は、Boström (1972) による三人称主語人称代名詞の使用形態という観点からの分類を修正したものであるが、設定されている時代がイタリア語起源 (9 世紀) から 18 世紀までであるので、19 世紀以降は Boström (1972) に従うことにする。
- 4) イタリア語の起源から 1250 年までの第 I 期に関しては、文体的に適当なものが見出せないため、本稿では扱わないこととする。また、各コーパスの範囲に見られる他文献などからの引用箇所やラテン語部分はデータとして除外している。
- 5) Calvino コーパス以外では、後続語が i- で始まる男性名詞の場合、他の母音の場合と異なり、複数形態に揺れが確認される。このため、i- を他の母音と区別して調査する (表中「i」と示す)。
- 6) 註の 1) を参照のこと。
- 7) lo 'imperatore 「皇帝」 / lo 'nquisitore 「異端裁判官」のように、名詞の語頭音 i が脱落して m- / n- などとなっている名詞は、母音で始まるものとして処理している。
- 8) 安本 (2002) は、各言語の数詞間の一致率を言語間の距離と見なして、クラスター分析により印欧語の分類を行っている。
- 9) n_A はクラスター A の要素の個数、 n_B はクラスター B の要素の個数、 $\overline{y}_{j(A)}$ は 1 から j までのクラスター A の要素の平均値、 $\overline{y}_{j(B)}$ は 1 から j までのクラスター B の要素の平均値、 $\overline{y}_{j(C)}$ は 1 から j までのクラスター C の要素の平均値とする。
- 10) 表中「後続音」の項で、s と示したものは s impura を示す。また、男性複数形態 e と e' は同じ形態として処理することにする。
- 11) 表中、定冠詞形態の後に [] で示したものは、後続語の条件を示す。例えば、il [男単+子音] は、後続語が子音で始まる男性単数名詞であるということを示す。なお、[] 中の「+s」は、後続が s impura であることを意味し、「+母音(i)」は後続語が i- で始まる語であることを示す。
- 12) 判別分析の実施に当たり、現代語確立型と現代語衰退型という分類の根拠とした Calvino コーパスのデータを含めると、識別の対象となる 2 群の母分散共分散行列がそれぞれ異なることになり、判別関数が使用できなくなる。このため、判別分析は Calvino コーパスを除外して行っている。

- 13) 3つ以上の群を判別分析を用いて識別するためには、2群の判別分析を繰り返し適用していく必要があるが、正準判別分析を適用するとこの判別を同時に行うことが可能となる。なお、本稿でのデータについて、等分散性・ボックスのM検定を行った結果、全群の分散・共分散行列は有意水準（p値）1%レベルで等しいといえるので、正準判別分析が実施可能と言える。
- 14) 判別分析のときと同じ理由で、Calvino コーパスを除外して正準判別分析を行う。
- 15) 正準判別分析において「軸」というのは、母集団をいくつかの群に区分したり、実際の区分が適切であるか否かを検証することを目的として算定されるものである。各軸を構成する重み係数は、例えば、〈表8〉に見るように、軸ごとに Dante から Leopardi について、各々の値が計算される。正準判別分析では、各軸における重み係数は区分すべき群の数に応じて、これらをできるだけうまく分離するように算定されている。
- 16) Bartlett の χ^2 検定は、正準判別係数を求めることに意味があるか、また意味があるとすればいくつの軸を用いるべきかを検定するものである。本稿では、統計学において通常用いられている有意水準5%以下という基準に基づき、軸1～軸3までの3軸としている。
- 17) 正準判別分析における判別得点は、各軸の重み係数と定数を用いて、26個の各定冠詞毎に計算される。この判別得点の計算結果が、〈表9〉に示されている。例えば、任意の定冠詞における軸1の判別得点 = $73.1634 - 0.3212 \times \text{Dante} - 0.2855 \times \text{Boccaccio} + 3.9122 \times \text{Alberti} - 4.3314 \times \text{Machiavelli} + 8.0188 \times \text{Della Casa} - 133.0863 \times \text{Vasari} + 25.9568 \times \text{Galileo} - 72.7950 \times \text{Leopardi}$ となる。例えば〈表9〉の軸1におけるl' [女単+母音] の値 (-98.5253) は、上に示した式の Dante から Leopardi に l' [女単+母音] の使用割合が代入されて計算されたものである。

参考文献

- Boström, Ingemar. 1972. *La morfossintassi dei pronomi personali soggetti della terza persona in italiano e in fiorentino*. Almqvist & Wiksell.
- D'Achille, Paolo. 1990. *Sintassi del parlato e tradizione scritta della lingua italiana: Analisi di testi dalle origini al secolo XVIII*. Bonacci Editore Roma.
- Dardano, Maurizio & Pietro Trifone. 1997. *La nuova grammatica della lingua italiana*. Zanichelli.
- Fogaras, Miklós. 1983. *Grammatica italiana del novecento: Sistemazione descrittiva*, 2nd ed. Bulzoni Editore.
- Lepschy, Anna Laura & Giulio Lepschy. 1988. *The Italian Language Today*, 2nd ed. New Amsterdam.
- Migliorini, Bruno. 2000. *Storia della lingua italiana*, 8th ed. Bompiani.
- Serianni, Luca. 1989. *Grammatica italiana: Italiano comune e lingua letteraria*. UTET.
- Ward, Joe H. Jr. 1963. *Hierarchical Grouping to Optimize an Objective Function*. American Statistical Association Journal, Vol.58, March. pp.236-244.
- 上野貴史. 2002. 『19世紀イタリア語における名詞句の研究』. 博士論文 (広島大学大学院).
- 内田治・菅民郎・高橋信. 2003. 『EXCEL アドインによる多変量解析』. 東京図書.
- 菅民郎. 1996. 『多変量統計分析』. 現代数学社.
- 古浦敏生. 1996. 『イタリア語における冠詞研究』. 文流.
- 田中豊・垂水共之. 1995. 『統計解析ハンドブック』. 共立出版.
- 谷岡弘二. 2002. 『日本の地域間所得格差に関する時系列分析』. 博士論文 (大阪商業大学大学院).
- 前川眞一. 1997. 『SASによる多変量データの解析』. 東京大学出版会.
- 安本美典. 2002. 「言語研究における多変量解析」. 『多変量解析実例ハンドブック』 (柳井晴夫他編). 朝倉書店. pp.680-692.