

最終講義 2017/02/24

# わかりやすい言語データ分析の方法

—スペイン語の地理的変異と歴史的変化を例に

上田博人

- 「わかりやすい」とは？：「わかる」→理解→納得→説明
- 「言語データ」：テキストデータ・項目データ
- 「データ分析」：観察・質的分析・量的分析
- 「地理的変異と歴史的变化」：「地理的変異 and/or 歴史的变化」

## 予定：

- はじめに
- 行列 / 頻度・和・平均・分散
- 距離・近接 / クラスタ分析 / 集中化 / 選択軸
- 分析例

\*時間・スペースの都合で確率・相関分析・多変量解析は扱いません。

次を参照してください：

『言語研究のための数値データ分析法』

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/>

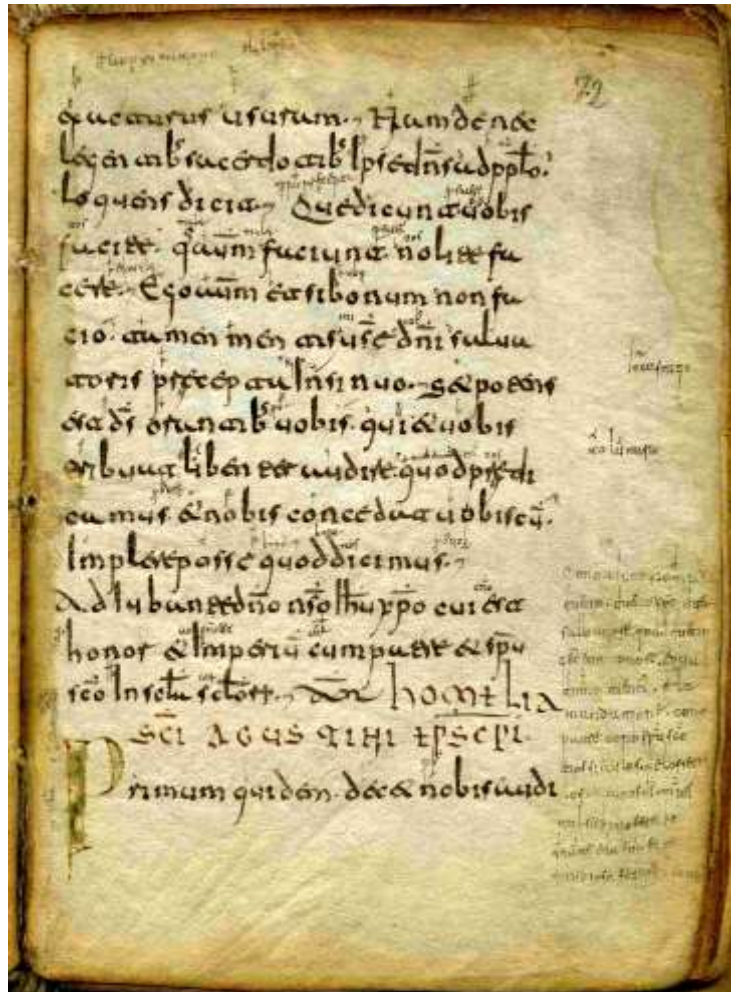
<http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/index.html>

## 0. はじめに

### 0.1. スペイン語の位置



## 0.2. 最初の文献（10世紀末）



Glosas emilianenses

### 0.3. プエルトリコ



San Juan, Puerto Rico

Si tomas, no guies. 「飲むなら運転するな。」



「(車を)運転する」:**guiar** (地図では **G**) は他にグアテマラでも記録した. スペインでは **conducir** と言う(**C**). ラテンアメリカ全域で **manejar** が優勢だが(**M**), **conducir** が併用される国もある. ボリビアのラパスでは **dirigir** も報告された(**D**)

## VARILEX-R D:013 («Variación Léxica del Español en el Mundo»)

(R) D013 [DRIVE]: Llevar el control de un automóvil. :: Ej. Ya se ve él \*\*\* también un Mercedes.

Forma	ES	GE	CU	RD	PR	MX	GU	HO	EL	NI	CR	PN	CO	VE	EC	PE	BO	CH	PA	UR	AR
guiar	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
manejar	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+
conducir	+	+	+	+	+	+	-	-	+	+	+	+	-	+	-	+	+	-	+	+	+

Proyecto VARILEX:

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/varilex/index.html>



#### 0.4. スペイン語クレオール



【写真2】フィリピン、サンボアング・ピラール Zamboanga, Pilar 砦内のごみ箱  
AQUI BUTA EL BASURA. 「ここにごみを捨ててください」

## 0.5. 調査

文献調査

現地調査

資料調査

## 0.6. 分析法

準備：行列

頻度・和・平均・分散

確率・有意度

距離・近接

分類・関係

集中化（…）

## 0.7. 研究テーマ

音声：(1) アンダルシアの子音・母音

文法：(2a) 前置詞 **de** の有無; (2b) 不規則未来形

語彙：(3a) 「(車を)運転する」; (3b) 「農夫」; (3c) 頻度と拡散

表現：(4) 「布告する・証する」 (公証文書)

表記：(5) 二重子音文字

# 1. 行列

## 1.1. ベクトルとベクトルの積

$$X_{13} Y_{31} = Z, Z = X(X_{13}, Y_{31})$$

$X_{13}$	1	2	3
1	1	2	3

 $\times$ 

$Y_{31}$	1
1	4
2	5
3	6

 $=$ 

$Z_{11}$	1
1	$1*4 + 2*5 + 3*6$

 $=$ 

$Z_{11}$	1
1	32

## 1.2. 行列と行列の積

$$X_{32} Y_{23} = Z_{33}, Z_{33} = X(X_{32}, Y_{23})$$

$X_{32}$	1	2
1	1	2
2	3	4
3	5	1

 $\times$ 

$Y_{23}$	1	2	3
1	7	9	2
2	8	1	3

$Z_{33}$	1	2	3
1	$1*7 + 2*8$	$1*9 + 2*1$	$1*2 + 2*3$
2	$3*7 + 4*8$	$3*9 + 4*1$	$3*2 + 4*3$
3	$5*7 + 1*8$	$5*9 + 1*1$	$5*2 + 1*3$

 $=$ 

$Z_{33}$	1	2	3
1	23	11	8
2	53	31	18
3	43	46	13

## 2. 頻度・和・平均・分散

行和： $S_{n1} = D_{np} I_{p1}$

$D_{np}$	1	2	3	$\times$	$I_{p1}$	1	$=$	$S_{n1}$	1
1	6	8	5		1	1		1	19
2	7	10	6		2	1		2	23
3	8	4	8		3	1		3	20
4	9	7	2					4	18
5	10	9	4					5	23

$$6*1 + 8 *1 + 5*1 = 19$$

$$\text{列和} : \mathbf{S}_{1p} = \mathbf{I}_{n1}^T \mathbf{D}_{np} = \mathbf{I}_{1n} \mathbf{D}_{np}$$

ここで  $\mathbf{I}_{n1}^T$  は単位ベクトル  $\mathbf{I}_{n1}$  を転置させた行列・ベクトル(= $\mathbf{I}_{1n}$ )。

$\mathbf{I}_{1n}$	1 2 3 4 5	$\times$	$\mathbf{D}_{np}$	1 2 3	$=$	$\mathbf{S}_{1p}$	1 2 3
1	1 1 1 1 1		1	6 8 5		1	40 38 25
			2	7 10 6			
			3	8 4 8			
			4	9 7 2			
			5	10 9 4			

$$1*6 + 1*7 + 1*8 + 1*9 + 1*10 = 40$$

$$\text{総和} : \mathbf{S} = \mathbf{I}_{1n} \mathbf{S}_{n1} = \mathbf{S}_{1p} \mathbf{I}_{p1} \quad : \mathbf{I} \text{ の左積と右積}$$

## 2.1. 平均

列平均： $M_{1p} = \mathbf{I}_{n1}^T \mathbf{D}_{np} / n$

$S_{1p}$	1	2	3
1	40	38	25

 / 5 = 

$M_{1p}$	1	2	3
平均	8.0	7.6	5.0

## 2.2. 分散・標準偏差

分散： $Vr = \sum (i) [X(i) - m]^2 / n$

標準偏差： $SD = (Vr)^{1/2}$



## 2.3. 規定分散・規定標準偏差

「規定分散」 (Regular Variance: R.Vr)

$$R.Vr = Vr / Vr.max$$

$$Vr = [(x_1 - m)^2 + (x_2 - m)^2 + \dots + (x_n - m)^2] / n$$

例: {10, 0, 0, 0, 0}, 一般化して {k, 0, 0, ..., 0}, このとき分散 Vr は最大。

$$\begin{aligned} Vr.max &= [(k - m)^2 + (n - 1)(0 - m)^2] / n \quad \leftarrow 1 \text{ 個だけ } k, \text{ 残り } 0 \\ &= [(k - m)^2 + (n - 1) m^2] / n \end{aligned}$$

このとき

$$k = \text{総和 } s = n m \quad \leftarrow \text{平均 } m = \text{総和 } s / \text{個数 } n$$

よって分散の最大値  $Vr.max$  は

$$\begin{aligned} Vr.max &= [(k - m)^2 + (n - 1) m^2] / n \\ &= [(n m - m)^2 + (n - 1) m^2] / n \leftarrow k = n m \\ &= \{[(n - 1) m]^2 + (n - 1) m^2\} / n \leftarrow m \text{ を外へ} \\ &= (n - 1) m^2 [(n - 1) + 1] / n \leftarrow m^2 (n - 1) \text{ が共通} \\ &= m^2 (n - 1) \leftarrow n \text{ を整理} \end{aligned}$$

よって規定分散( $R.Vr$ )は

$$R.Vr = Vr / Vr.max = Vr / [m^2 (n - 1)]$$

同様にして「規定標準偏差」(Regular Standard Deviation:  $R.SD$ )は

$$\begin{aligned} R.SD &= (R.Vr)^{1/2} = (Vr / Vr.max)^{1/2} = \{Vr / [m^2 (n - 1)]\}^{1/2} \\ &= SD / [m (n - 1)^{1/2}] \end{aligned}$$

S1	v1	v2	v3	v4	v5		橫軸	分散	規定分散	標準偏差	規定標準偏差
h1	10	19	14	7	12		h1	16.240	.026	4.030	.162
h2	11	7	10	0	1		h2	20.560	.153	4.534	.391
h3	0	0	1	12	1		h3	21.360	.681	4.622	.825
h4	0	1	2	3	3		h4	1.360	.105	1.166	.324

## 2.4. 分析例：語彙の頻度と拡散

García Hoz, V. 1953. *Vocabulario usual, vocabulario común y vocabulario fundamental*. Madrid. C.S.I.C.

Juilland, A. and Chang-Rodríguez, E. 1964. *Frequency dictionary of Spanish words*. Mouton. The Hague.

Ueda, H. 1987. *Frecuencia y dispersión del vocabulario español*. Universidad de Estudios Extranjeros de Tokio. Instituto de Estudios de Lenguas Extranjeras.

Diversity (Juilland, A. and Chang-Rodríguez, E. 1964, p. xv)

(1) Dramatic literature / (2) Fictional literature / (3) Essayistic literature /

(4) Technical literature / (5) Journalistic literature : 10 万語ずつ

Usage = Frequency \* Dispersion (p. LIII)

Dispersion = 1 - SD/ (2m)

ここで SD: 標準偏差、m: 平均、2 は  $(5 - 1)^{1/2}$  (?)

よって「拡散度」(Dispersion:)は

Dispersion = 1 - SD/ (2m) = 1 - SD / [m (n - 1)<sup>1/2</sup>] = 1 - R.SD

分析例 Ueda. 1987 : 10 の分野 10 万語ずつ

Orden	Frecuencia	Dispersión	Uso	Palabra	Parte de oración
1	72216	.890	64332.80	de	prep
2	36624	.907	33232.60	que	conj
3	32684	.928	30354.00	y	conj.
4	25011	.925	23147.90	en	prep
5	19403	.893	17329.00	ser	v

(...) Orden: 5056

「規定化」 : 規定標準偏差。範囲[0, 1]

- ・ 同じスケール比較が可能
- ・ 応用可能性
- ・ わかりやすい

### 3. 距離・近接

2つの数値（座標）の差二乗和(Sum of Squared Differences: SSD) :

$$\text{SSD}(x, y) = \sum_i (x_i - y_i)^2 \quad \text{ここで } i = 1, 2, \dots, n \dots(1)$$

M	X	Y	X-Y	(X-Y)^2
h1	10	19	-9	81
h2	11	7	4	16
h3	0	0	0	0
h4	0	1	-1	1
			SSD	98

「規定距離係數」 (Regular Distance: RD)

「規定近接係數」 (Regular Proximity: RP)

$$RP = 1 - RD, RD = 1 - RP$$

$$SSD(x, y) = \sum_i (x_i - y_i)^2 = \sum_i (x_i^2 + y_i^2 - 2x_i y_i) \dots(2)$$

$$\text{cf. } (x - y)^2 = x^2 + y^2 - 2xy$$

$$\begin{aligned} \text{式(2): } & \sum_i (x_i - y_i)^2 / (\sum_i x_i^2 + \sum_i y_i^2) \\ & = (\sum_i x_i^2 + \sum_i y_i^2 - 2 \sum_i x_i y_i) / (\sum_i x_i^2 + \sum_i y_i^2) \end{aligned}$$

$$\sum_i (x_i - y_i)^2 / (\sum_i x_i^2 + \sum_i y_i^2) = 1 - 2 \sum_i x_i y_i / (\sum_i x_i^2 + \sum_i y_i^2) \dots (3)$$

$$2 \sum_i x_i y_i / (\sum_i x_i^2 + \sum_i y_i^2) = 1 - \sum_i (x_i - y_i)^2 / (\sum_i x_i^2 + \sum_i y_i^2) \dots(3')$$



「規定距離係数」 (RD):

$$RD(x, y) = \sum_i (x_i - y_i)^2 / (\sum_i x_i^2 + \sum_i y_i^2) \dots(4)$$

$\sum_i (x_i - y_i)^2 = 0$  の場合、つまり  $x_i$  と  $y_i$  ( $i=1,2,\dots,n$ )が同じ場合、よって  $\sum_i x_i^2 = \sum_i y_i^2$  の場合に RD が最小値(=0)になる。

$\sum_i x_i^2 > \sum_i y_i^2$  の場合、 $\sum_i y_i^2 = 0$  の場合、つまり  $y_i = 0$  ( $i=1,2,\dots,n$ )の場合に RD が最大になる。最大値は

$$RD(x, 0) = \sum_i (x_i - 0)^2 / (\sum_i x_i^2 + \sum_i 0^2) = \sum_i x_i^2 / \sum_i x_i^2 = 1$$

$\sum_i x_i^2 < \sum_i y_i^2$  の場合、 $\sum_i x_i^2 = 0$  の場合、つまり  $x_i = 0$  ( $i=1,2,\dots,n$ )の場合に RD が最大になる。最大値は

$$RD(0, y) = \sum_i (0 - y_i)^2 / (\sum_i 0^2 + \sum_i y_i^2) = \sum_i y_i^2 / \sum_i y_i^2 = 1 \dots(4')$$

「規定近似係数」(RP)は規定距離係数(RD)の1の補数

$$RP = 1 - RD$$

RPの範囲は[0, 1]; 条件はRDとは逆転します。

$$RP(x, x) = 1 - \frac{\sum_i (x_i - x_i)^2}{(\sum_i x_i^2 + \sum_i x_i^2)} = 1 - 0 = 1$$

$$RP(x, 0) = 1 - \frac{\sum_i (x_i - 0)^2}{(\sum_i x_i^2 + \sum_i 0^2)} = 1 - 1 = 0$$

$$RP(0, y) = 1 - \frac{\sum_i (0 - y_i)^2}{(\sum_i 0^2 + \sum_i y_i^2)} = 1 - 1 = 0$$

以上は x, y が頻度などの非負データの場合

M	A	B	C	D	E
h1	10	19	14	7	12
h2	11	7	10	0	1
h3	0	0	1	12	1
h4	0	1	2	3	3

A と B の規定距離係数(RD)と規定近接係数(RP) :

A.B	X:A	Y:B	X-Y	$(X-Y)^2$	$X^2$	$Y^2$	$X^2+Y^2$		
h1	10	19	-9	81	100	361	461		
h2	11	7	4	16	121	49	170		
h3	0	0	0	0	0	0	0		
h4	0	1	-1	1	0	1	1	RD	RP
和 →				98			632	0.1551	0.8449

大きな距離を示す例：

D.E	X:D	Y:E	X-Y	$(X-Y)^2$	$X^2$	$Y^2$	$X^2+Y^2$		
h1	7	12	-5	25	49	144	193		
h2	0	1	-1	1	0	1	1		
h3	12	1	11	121	144	1	145		
h4	3	3	0	0	9	9	18	RD	RP
		計 →		147			357	0.4118	0.5882

X = Y のときの最小 RD (=0)、最大 RP (=1)の場合 :

C.C	X:C	X:C	X-Y	(X-Y) <sup>2</sup>	X <sup>2</sup>	Y <sup>2</sup>	X <sup>2</sup> +Y <sup>2</sup>		
h1	14	14	0	0	196	196	392		
h2	10	10	0	0	100	100	200		
h3	1	1	0	0	1	1	2		
h4	2	2	0	0	4	4	8	RD	RP
	和 →			0			602	0.0000	1.0000

Y = 0 のときの最大 RD (=1)、最小 RP (=0)の場合 :

C.0	X:C	Y:0	X-Y	(X-Y) <sup>2</sup>	X <sup>2</sup>	Y <sup>2</sup>	X <sup>2</sup> +Y <sup>2</sup>		
h1	14	0	14	196	196	0	196		
h2	10	0	10	100	100	0	100		
h3	1	0	1	1	1	0	1		
h4	2	0	2	4	4	0	4	RD	RP
和 →				301			301	1.0000	0.0000

## ●規定距離係数・規定近接係数の最大値の証明

先の式(3'):

$$2 \sum_i x_i y_i / (\sum_i x_i^2 + \sum_i y_i^2) = 1 - \sum_i (x_i - y_i)^2 / (\sum_i x_i^2 + \sum_i y_i^2) \quad \dots(3')$$

は  $1 -$  規定距離係数(RD)、すなわち規定近接係数(RP)を示す。この左辺をベクトルのノルムを使って表現すると

$$\begin{aligned} 2 \sum_i x_i y_i / (\sum_i x_i^2 + \sum_i y_i^2) &= 2 \mathbf{X}^T \mathbf{Y} / (\|\mathbf{X}\|^2 + \|\mathbf{Y}\|^2) \quad \dots(1) \\ &= 2 \|\mathbf{X}\| \|\mathbf{Y}\| / (\|\mathbf{X}\|^2 + \|\mathbf{Y}\|^2) \cos\theta \quad \leftarrow \mathbf{X}^T \mathbf{Y} = \|\mathbf{X}\| \|\mathbf{Y}\| \cos\theta \end{aligned}$$

上の式の一部（網掛けの部分）

$$\|\mathbf{X}\| \|\mathbf{Y}\| / (\|\mathbf{X}\|^2 + \|\mathbf{Y}\|^2) \quad \dots(2)$$

の最大値を求める。変数が  $\|\mathbf{X}\|^2$ ,  $\|\mathbf{Y}\|^2$  にあるので2つの変数の比を1つの

変数  $c$  で表し、 $c$  だけを使って(2)を表現します。

$$c = \|\mathbf{X}\|^2 / \|\mathbf{Y}\|^2 \quad \dots(3)$$

式(2)の分子と分母をそれぞれ  $\|\mathbf{Y}\|^2$  で割ると

$$(2)の分子 : \|\mathbf{X}\| \|\mathbf{Y}\| / \|\mathbf{Y}\|^2 = \|\mathbf{X}\| / \|\mathbf{Y}\| = c^{1/2}$$

$$(2)の分母 : (\|\mathbf{X}\|^2 + \|\mathbf{Y}\|^2) / \|\mathbf{Y}\|^2 = \|\mathbf{X}\|^2 / \|\mathbf{Y}\|^2 + 1 = c + 1$$

よって式(2)を  $c$  の関数で示すと

$$式(2) : f(c) = c^{1/2} / (c + 1)$$

次のような分子と分母が  $c$  の関数である式

$$f(c) = g(c) / h(c)$$

の微分  $f'(c)$  は



$$f'(c) = [g'(c) h(c) - g(c) h'(c)] / h^2(c)$$

これより

$$\begin{aligned} f'(c) &= 1 / (c+1)^2 [1/2 c^{-(1/2)} (c+1) - c^{1/2} 1] \\ &= 1 / (c+1)^2 1/2 c^{-(1/2)} [(c+1) - 2c] \\ &= 1 / (c+1)^2 1/2 c^{-(1/2)} (1 - c) \\ &= 1/2 c^{-(1/2)} (1 - c) / (c+1)^2 \\ &= c^{-(1/2)} (1 - c) / [2(c+1)^2] \end{aligned}$$

よって  $f'(c) = 0$  は  $c = 1$  のとき。増減表は

c	0	...	1	...
f'(c)	.	+	0	-
f(c)	0	↑	1/2	↓

よって

$$\text{式(2)} : \quad \|X\| \|Y\| / (\|X\|^2 + \|Y\|^2)$$

は  $c = \|X\|^2 / \|Y\|^2 = 1$ 、つまり  $\|X\|^2 = \|Y\|^2$ 、よって  $\|X\| = \|Y\|$  のとき  
最大値  $1/2$  に達します。よって式(3')の最大値は

$$\text{式(3')} \quad 2 \sum_i x_i y_i / (\sum_i x_i^2 + \sum_i y_i^2)$$

$$= 2 \|X\| \|Y\| / (\|X\|^2 + \|Y\|^2) \cos\theta = 2 * 1/2 * \cos\theta = \cos\theta$$

規定近接係数 **RP** が最大であるのは、**X** と **Y** の座標が一致するときなので角  
度( $\theta$ )はゼロ(0)、よって  $\cos\theta=1$  となる。

\* この数学的証明では東京大学大学院総合文化研究科・情報学環の倉田博史先生の全面的な支援を仰ぎました。また先生からご教示いただいた次の式が規定距離係数・規定近接係数を考える貴重な出発点になりました。

$$\|X - Y\|^2 = \|X\|^2 + \|Y\|^2 - 2 X^T Y$$

(2017/2/18)

\* 微分：高校数学 III を参照します。

(0) 微分の定義：

$$f'(x) = \lim_{h \rightarrow 0} [f(x+h) - f(x)] / h$$

(1) 関数の積の微分 :

$$\begin{aligned} [f(x)g(x)]' &= \lim_{h \rightarrow 0} [f(x+h)g(x+h) - f(x)g(x)] / h \quad \leftarrow \text{微分の定義} \\ &= \lim_{h \rightarrow 0} \{ [f(x+h)-f(x)]g(x+h) + f(x)[g(x+h)-g(x)] \} / h \\ &\quad \leftarrow f(x)g(x+h) \text{を媒介} \\ &= \lim_{h \rightarrow 0} [f(x+h)-f(x)]g(x+h)/h + \lim_{h \rightarrow 0} f(x)[g(x+h)-g(x)]/h \\ &\quad \leftarrow \lim_{h \rightarrow 0} \text{を分配} \\ &= f'(x)g(x) + f(x)g'(x) \dots(1) \quad \leftarrow \text{微分の定義} \end{aligned}$$

(2) 関数の商の微分

$$\begin{aligned} [1/g(x)]' &= \lim_{h \rightarrow 0} [1/g(x+h) - 1/g(x)] / h \quad \leftarrow \text{微分の定義} \\ &= \lim_{h \rightarrow 0} \{ [g(x) - g(x+h)] / [g(x+h)g(x)] \} / h \quad \leftarrow \text{通分} \\ &= \lim_{h \rightarrow 0} \{ -[g(x+h) - g(x)] / h * 1 / [g(x+h)g(x)] \} \end{aligned}$$

第 1 項の分子と第 2 項の分母を整理

$$= -g'(x) / [g(x)g(x)] = -g'(x) / g^2(x) \dots(2)$$

(1), (2) より

$$f'(x) = [g(x) / h(x)]' = [g(x) * 1 / h(x)]' \quad \leftarrow \text{分母を分離}$$

$$= g'(x) * 1/h(x) + g(x)*[1/h(x)]' \quad \leftarrow (1)$$

$$= g'(x) / h(x) + g(x) * -h'(x) / h^2(x) \quad \leftarrow (2)$$

$$= [g'(x)h(x) - g(x)h'(x)] / h^2(x) \quad \leftarrow \text{第 2 項の負号を前に}$$

## 4. クラスター分析

### 4.1. 最近隣法

語彙バリエーション：距離行列(ES:スペイン, GE:赤道ギニア, CI:キューバ, RD:ドミニカ共和国, PR:プエルトリコ, MX:メキシコ)

6 か国	1. ES	2. GE	3. CI	4. RD	5. PR	6. MX
1. ES	.00	.20	.25	.23	.21	.28
2. GE	.20	.00	.28	.27	.25	.33
3. CI	.25	.28	.00	.23	.22	.30
4. RD	.23	.27	.23	.00	.16	.28
5. PR	.21	.25	.22	<u>.16</u>	.00	.25
6. MX	.28	.33	.30	.28	.25	.00

最初のクラスタリング：最小値(.16)をもつ 4:RD と 5:PR が合体。

6 か国	1. ES	2. GE	3. CI	4. RD:5. PR	6. MX
1. ES	.00	.20	.25	.21	.28
2. GE	.20	.00	.28	.25	.33
3. CI	.25	.28	.00	.22	.30
4. RD: PR	.21	.25	.22	.00	.25
6. MX	.28	.33	.30	.25	.00

このとき、(1+2) と 3 の距離は、1-3 の距離(= .25), 2-3 の距離(=.28)の間のそれぞれの距離のうち小さなほうの値(.25)とします。

2 番目のクラスタリング :

6 か国	1. ES: 2. GE	3. CI	4. RD: 5. PR	6. MX
1. ES: 2. GE	.00	.25	.21	.28
3. CI	.25	.00	.22	.30
4. RD: 5. PR	.21	.22	.00	.25
6. MX	.28	.30	.25	.00

3 番目のクラスタリング :

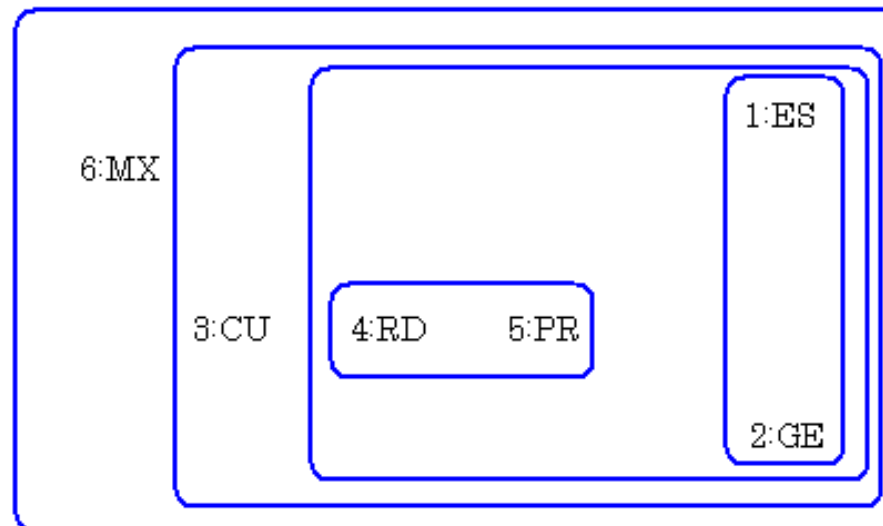
6 か国	1. ES: 2. GE:4. RD:5. PR	3. CI	6. MX
1. ES: 2. GE:4. RD:5. PR	.00	.22	.25
3. CI	.22	.00	.30
6. MX	.25	.30	.00



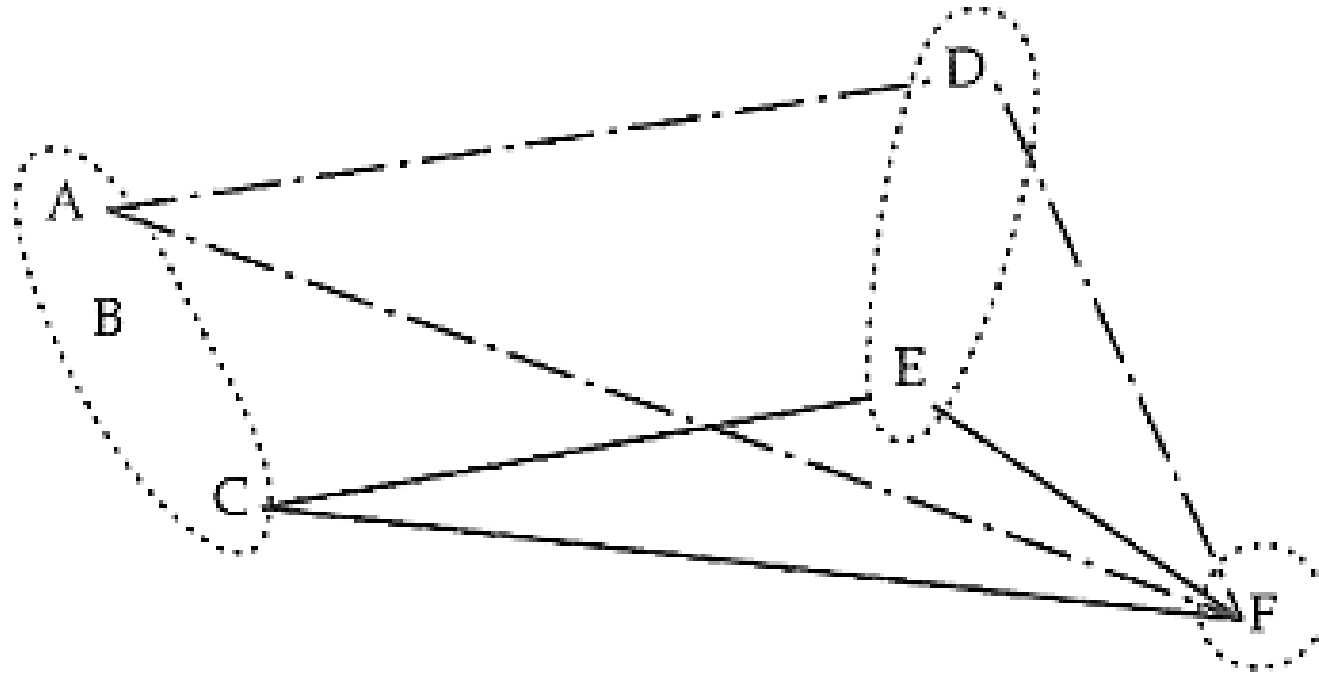
4 番目のクラスタリング :

6 か国	1. ES: 2. GE: 4. RD:5. PR: ;3. CI 6. MX
1. ES: 2. GE: 4. RD:5. PR: ;3. CI	.00 .25
6. MX	<u>.25</u> .00

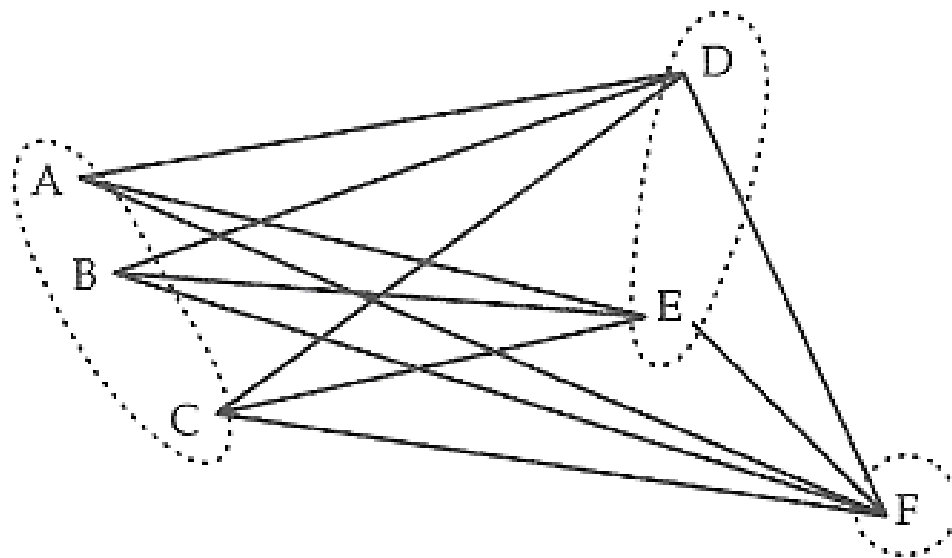
グルーピング :



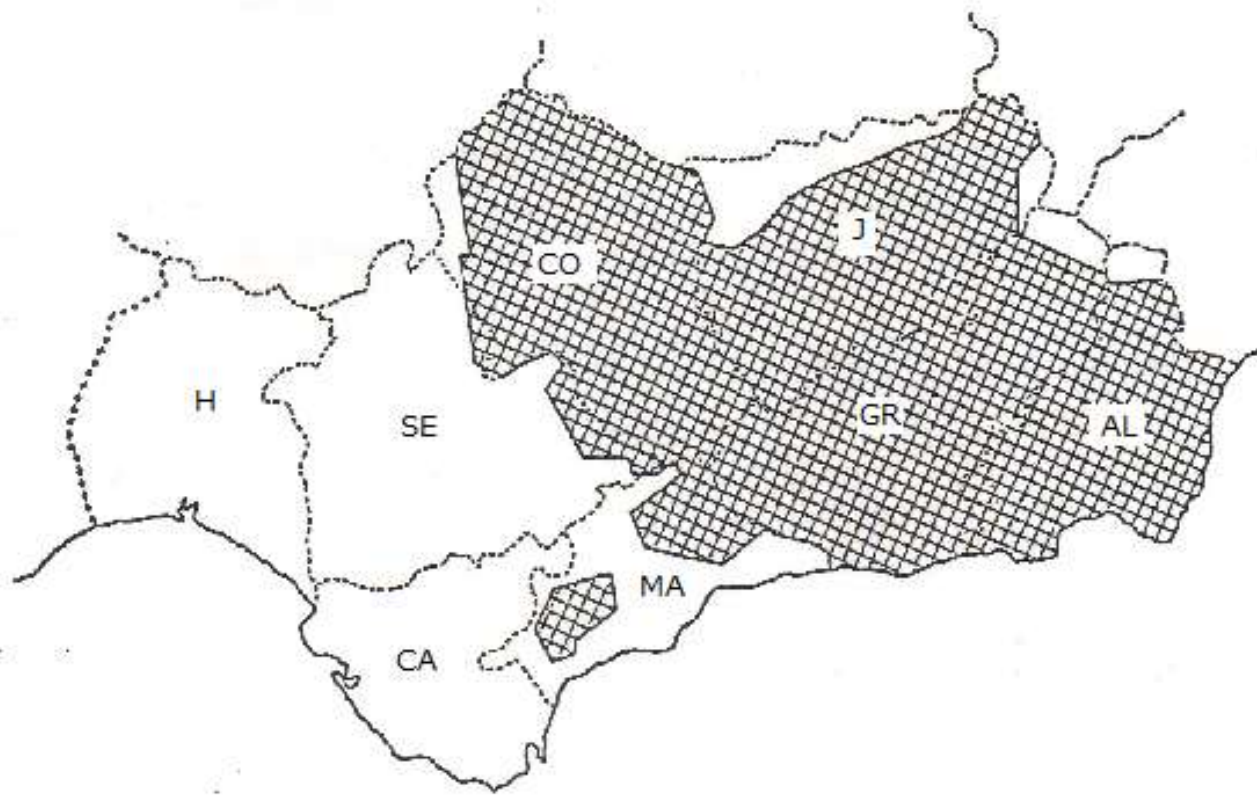
## 4.2. 最遠隣法



### 4.3. 平均法



#### 4.4. 分析例：アンダルシアの子音・母音



H: Huelva; SE: Sevilla; CA: Cadiz; MA: Malaga; CO: Cordoba      J: Jaén; GR:  
Granada; AL: Almería

網掛け部分：名詞複数形の音韻的な開母音：Alvar, Manuel. 1973. *Estructuralismo, geografía lingüística y dialectología española*. Madrid. Gredos

Cnp	H	SE	CA	MA	CO	J	GR	AL
1. miel:el>e+ (1533B)	17	10	9	15	20	29	46	30
2. miel:el>e: (1533C)	11	6	4	16	12	11	16	3
3. caracol:-ól>ó+: (1615A)	2	3	3	5	15	14	19	11
4. caracol:-ól>ó: (1615B)	18	27	15	16	3	1	6	2
5. árbol:-ol>o+ (1616A)	2	1	0	0	6	8	6	6
6. árbol:-ol>o (1616B)	23	30	17	26	18	11	23	11
7. sol:-ól>ó+: (1618A)	7	9	3	13	13	12	19	11

8. sol:-ól>ó: (1618B)	15	21	15	13	1	1	6	1
9. beber:-ér>é+l (1623A)	2	0	0	1	10	11	19	20
10. beber:-ér>é+ (1623B)	4	7	3	6	13	17	15	8
11. beber:-ér>é (1623C)	19	24	15	19	2	0	4	0
12. tos:-ós>ó+h (1626A)	6	2	2	4	7	13	17	9
13. tos:o++ (1626C)	0	0	0	2	7	10	18	12
14. tos:-ós>ó+ (1626D)	7	7	5	13	18	17	27	19
15. tos:-ós>ó (1626E)	10	22	11	9	2	1	2	0
16. nuez:-éθ>é+h (1627A)	5	2	1	0	2	3	8	3
17. nuez:-éθ>é+ (1627B)	7	13	5	17	20	25	39	26
18. nuez:e++ (1627C)	0	0	0	5	14	18	26	18
19. nuez:-éθ>é (1627D)	12	16	12	9	3	1	1	0

20. voz:-óθ>óh (1629A)	5	3	0	1	1	1	5	3
21. voz:-óθ>ó+ (1629B)	3	5	3	12	22	30	44	30
22. voz:-óθ>ó (1629C)	18	23	14	13	2	1	2	1
23. niños:-os>-o+ (1689A)	1	2	0	4	22	31	44	30
24. niños:-os>oh (1689B)	4	1	0	0	2	3	8	8
25. pared:-éd>é+ (1690A)	6	8	0	10	17	19	24	11
26. redes:redes>rede (1693A)	3	2	1	1	4	12	8	18
27. redes:redes>re+ (1693B)	14	6	4	12	3	6	16	6
28. redes:redes>reh (1693C)	1	0	2	0	1	4	7	0
29. clavel:-él>-él (1694A)	3	2	1	3	6	5	11	15
30. clavel:-él>é+ (1694B)	0	6	3	15	20	24	40	29
31. clavel:-él>ér (1694C)	0	0	0	0	0	1	5	1

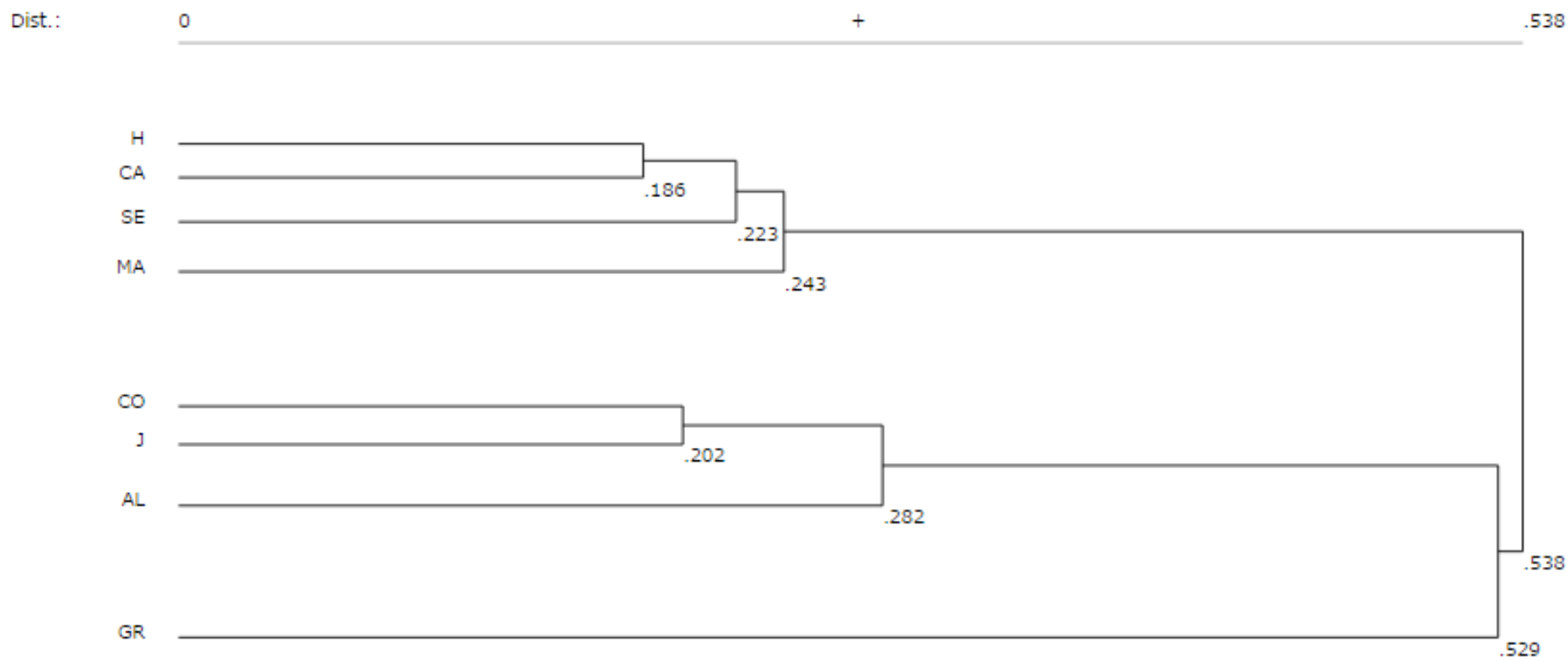
32. claveles:e-es>-e-e+ (1695A)	0	2	0	4	2	2	4	3
33. claveles:e-es>-e+-e+ (1695B)	0	1	0	7	18	24	33	21
34. claveles:-e-es>-e-e: (1695C)	0	1	0	3	1	2	1	1
35. claveles:e-es>-e-eh (1695D)	3	1	0	0	5	4	9	5

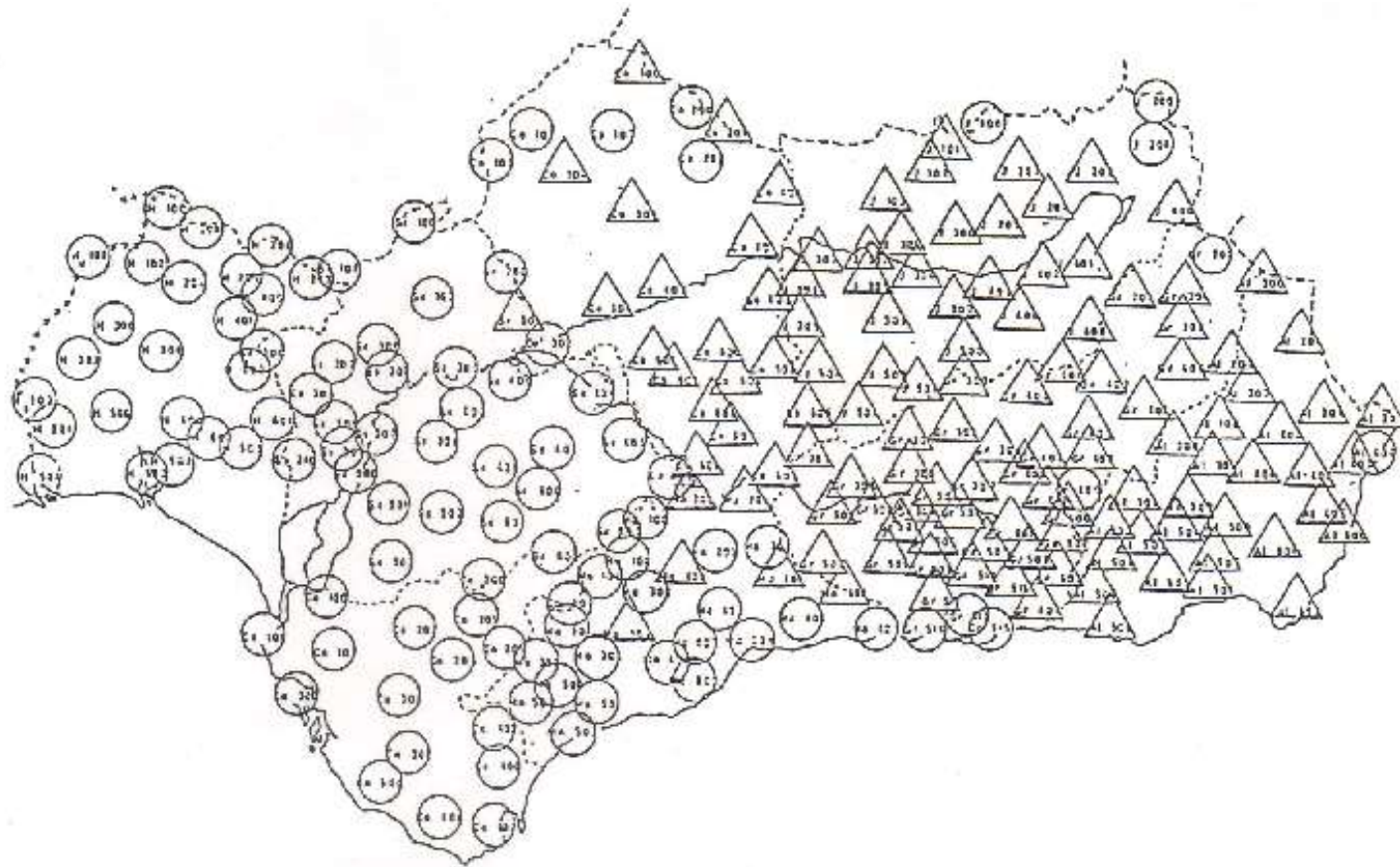
H	.000	.223	.186	.243	.538	.677	.946	.699
SE	.223	.000	.257	.275	.601	.741	.982	.763
CA	.186	.257	.000	.292	.545	.694	1.000	.710
MA	.243	.275	.292	.000	.396	.543	.792	.578
CO	.538	.601	.545	.396	.000	.202	.529	.282
J	.677	.741	.694	.543	.202	.000	.398	.204
GR	.946	.982	1.000	.792	.529	.398	.000	.421



AL	.699	.763	.710	.578	.282	.204	.421	.000
----	------	------	------	------	------	------	------	------

## デンドログラム





Ueda, Hiroto. 1993, "División dialectal de Andalucía. Análisis computacional." *Actas del III Congreso de Hispanistas de Asia*, 1993. pp. 407-419.







## 5. 集中化

### 5.1. 外的基準

Lv	v1	v2	v3	v4
h1	+	+		
h2			+	
h3		+		
h4			+	+
h5	+	+	+	

→

Lv	v1	v2	v3	v4
h1	+	+		
h3		+		
h5	+	+	+	
h2			+	
h4			+	+

$$\text{ZD}(h1) = [(1^2 + 2^2) / 2]^{1/2} = 1.581 \quad (\dots 1)$$

$$\text{cf. } \text{ZD}(h1) = [((0-1)^2 + (0-2)^2) / 2]^{1/2}$$

$$\text{ZD}(\text{h2}) = [(3^2) / 1]^{1/2} = 3.000 \quad (\dots 4)$$

$$\text{ZD}(\text{h3}) = [(2^2) / 1]^{1/2} = 2.000 \quad (\dots 2)$$

$$\text{ZD}(\text{h4}) = [(3^2 + 4^2) / 2]^{1/2} = 3.535 \quad (\dots 5)$$

$$\text{ZD}(\text{h5}) = [(1^2 + 2^2 + 3^2) / 3]^{1/2} = 2.160 \quad (\dots 3)$$

$$\text{ZD}^{\{3\}}(2) = (2^3 / 1)^{1/3} = (8 / 1)^{1/3} = 2$$

$$\text{ZD}^{\{3\}}(1, 3) = [(1^3 + 3^3) / 2]^{1/3} = (28 / 2)^{1/3} = 2.410$$

$$\text{ZD}^{\{3\}}(1, 2, 3) = [(1^3 + 2^3 + 3^3) / 3]^{1/3} = (36 / 3)^{1/3} = 2.289$$

## ■ラテンアメリカスペイン語の語彙変異: (1)外的基準

次のデータはラテンアメリカスペイン語の「農夫」を示す言語形式とその地理的分布(Cahuzac: 1980)。México, Cuba, República Dominicana, Puerto Rico, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panamá, Venezuela, Colombia, Ecuador, Perú, Bolivia, Chile, Paraguay, Uruguay, Argentina.

Cahuzac, Philippe. 1980. "La División del español de América en zonas dialectales: Solución etnolingüística o semántico-dialectal." *Lingüística Española Actual*, 10, pp. 385-461.

# 集中化前（片側：外的基準）

Cahuzac (1980)	MX	CU	RD	PR	C5	PN	VE	CO	EC	PE	BO	CH	PA	UR	AR
01 cacahuero							+								
02 cafetalista	+			+										+	+
03 camilicho													+	+	
04 campero															
05 camperuso							+								
06 campirano					+		+								
07 campiruso					+										
08 campista	+				+									+	
09 campusano						+								+	
10 campuso					+										
11 colono			+												
12 comparsa															
13 conuquero			+												
14 coquero							+								
15 chagrero								+							
16 changador															
17 chilero	+				+										
18 chuncano															
19 enmaniguado															
20 estanciero			+												
21 gaucho															
22 guajiro			+												
23 guanaco					+										
24 guaso			+												
25 huasicama															
26 huerto	+														
27 hulero	+				+										
28 invernador															
29 jbaro															
30 lampero															
31 lanudo															
32 llanero															
33 macanero															
34 manuto															
35 montero															
36 montubio															
37 paisano															
38 pajuerano															
39 partidario															
40 payazo															
41 piona															
42 rancho															
43 rondín															
44 sabanero															
45 veguero															
46 viñatero															
47 yanacón															



# 集中化後（片側：外的基準）

ZD.R	MX	CU	RD	PR	C5	PN	VE	CO	EC	PE	BO	CH	PA	UR	AR
33 macanero	+														
22 guajiro		+													
35 montero		+													
02 cafetalista	+				+										
19 enmanguado		+			+										
11 colono					+										
29 jibaro					+										
42 ranchero	+				+										
17 chilero	+				+										
27 huleo	+				+										
34 manuto					+										
45 veguero					+										
10 campuso					+										
08 campista	+				+										
07 campiruso					+										
23 guanaco					+										
13 conuquero					+				+						
06 campirano					+				+						
01 cacahuelo					+				+						
05 camperuso					+				+						
32 llanero					+				+						
40 payazo					+				+						
44 sabanero					+				+						
36 montubio					+				+						
31 lanudo					+				+						
38 pajerano					+				+						
37 paisano					+				+						
15 chagrero					+				+						
25 huasicama					+				+						
14 coquero					+				+						
43 rondín					+				+						
24 guaso					+				+						
26 huertero					+				+						
47 yanacón					+				+						
30 lampero					+				+						
09 campusano					+				+						
28 invernador					+				+						
46 viñatero					+				+						
21 gaucho					+				+						
39 partidario					+				+						
03 camilicho					+				+						
04 campero					+				+						
12 comparsa					+				+						
16 changador					+				+						
18 chuncano					+				+						
20 estanciero					+				+						
41 piona					+				+						

## 5.2. 內的基準

Lv	v1	v2	v3	v4
h1	+	+		
h2			+	
h3		+		
h4			+	+
h5	+	+	+	

→

Lv	v2	v1	v3	v4
h3	+			
h1	+	+		
h5	+	+	+	
h2			+	
h4			+	+

$$[1] \quad \text{ZD}^{\{3\}}(\text{h1}) = [(1^3 + 2^3) / 2]^{1/3} = 1.651$$

$$\text{ZD}^{\{3\}}(\text{h2}) = [(3^3) / 1]^{1/3} = 3.000$$

$$\text{ZD}^{\{3\}}(\text{h3}) = [(2^3) / 1]^{1/3} = 2.000$$

$$\text{ZD}^{\{3\}}(\text{h4}) = [(3^3 + 4^3) / 2]^{1/3} = 3.570$$

$$\text{ZD}^{\{3\}}(\text{h5}) = [(1^3 + 2^3 + 3^3) / 3]^{1/3} = 2.289$$

$X_n$ : (1.651, 3.000, 2.000, 3.570, 2.289)を最大値(Max: 3.570)と最小値  
(Min: 1.651)を使って次のように限定化。

$$X_n' = (X_n - \text{Min}) / (\text{Max} - \text{Min})$$

その結果、ベクトル  $X_n'$ の成分は

$$X_n' : (.000, .703, .182, 1.000, .333)$$

$$[2] \quad ZD^{\{3\}}(v1) = \{[(.000)^3 + (.333)^3] / 2\}^{1/3} = .264$$

同様にして

$$ZD^{\{3\}}(v2:1,3,5) = \{[(.000)^3 + (.182)^3 + (.333)^3] / 3\}^{1/3} = .243$$

$$ZD^{\{3\}}(v3:2,4,5) = \{[(.703)^3 + (1.000)^3 + (.333)^3] / 3\}^{1/3} = .773$$

$$ZD^{\{3\}}(v4:4) = \{(1.000)^3 / 1\}^{1/3} = 1.000$$

先と同様にして、このベクトル( $Y_p$ : .264, .243, .773, 1.000)を最大値(Max: 1.000)と最小値(Min: 0.243)を使って限定化する( $Y_p'$ )。

$$Y_p': (.028, .000, .700, 1.000)$$

これで第1回目の縦と横のベクトルの更新作業を終える。次に、この $Y_p'$ を使って、この[1]と[2]の作業を繰り返す。何度も繰り返すうちに、 $Y_p$ のベクトルの成分に変化がなくなれば、そこで更新作業を終了し、 $X_n$ と $Y_p$ を昇順でソートして次を出力する。

ZD.A	v2	v1	v3	v4
h3	+			
h1	+	+		
h5	+	+	+	
h2			+	
h4			+	+

Row	Xn
h3	.000
h1	.085
h5	.569
h2	.820
h4	1.000

Column	v2	v1	v3	v4
Yp	.000	.094	.724	1.000

# ■ラテンアメリカスペイン語の語彙変異:(1)両側：内的基準

ZD.A	PA	UR	AR	CH	BO	PE	MX	CU	PN	RD	PR	C5	EC	CO	VE
03 camilicho	+	+	+												
04 campero	+	+	+												
12 comparsa	+	+	+												
16 changador	+	+	+												
18 chuncano	+	+	+												
20 estanciero	+	+	+												
41 piona	+	+	+												
21 gaucho	+	+	+		+										
39 partidario	+	+	+		+										
43 rondín					+										
28 invernador	+	+	+	+											
46 viñatero	+	+	+	+											
47 yanacón			+	+											
30 lampero			+	+											
26 huerto			+	+		+									
09 campesano		+	+	+					+						
24 guaso			+	+				+					+		
33 macanero				+										+	
14 coquero															
42 rancho								+							
02 cafetalista								+		+					
22 guajiro								+							
35 montero								+							
17 chilero						+							+		
27 hulero						+							+		
34 manuto								+							
45 veguero								+							
38 pajerano						+							+		
37 paisano						+							+		
19 enmanguado								+							
11 colono										+					
29 jibaro										+					
07 campiruso										+					
23 guanaco								+					+		
10 campuso															
36 montubio						+				+					+
08 campista						+				+					+
13 conuquero								+							+
06 campirano										+					+
15 chagrero															
25 huasicama													+		+
31 lanudo													+		+
01 cacahnero															+
05 camperuso															+
32 llanero															+
40 payazo															+
44 sabanero															+

## ■ 文法バリエーション例：前置詞 de の有無：Queísmo / dequeísmo

Takagaki, Toshihiro (dir.) «Variación gramatical de Español».

[73] Estoy seguro \*que vendrá.

[74] Mi hermana está contenta \*que hayas aceptado la invitación.

[75] \*Sospecho de que me mintió

[76] Alfredo \*es feliz de vacaciones

[73]:

入力行列：

n.	Form	1:Oviedo	2:Pamplona	3:Alcalá	4:Madrid	5:Barcelona	6:Sevilla	7:Huelva	8:Tenerife	9:Las Palmas	Val.
1	3:NO	2	2	1	5	1	3	1	2	1	3.000
2	2:OIGO	19	11	5	9	13	11	6	11	17	2.000
3	1:DIGO	6	8	5	7	6	10	13	7	6	1.000
4	Val.	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000	9.000	

両側集中行列：

n.	Form	1:Oviedo	9:Las Palmas	5:Barcelona	8:Tenerife	2:Pamplona	6:Sevilla	4:Madrid	3:Alcalá	7:Huelva	Val.
1	1:DIGO	6	6	6	7	8	10	7	5	13	.314
2	3:NO	2	1	1	2	2	3	5	1	1	.195
3	2:OIGO	19	17	13	11	11	11	9	5	6	-.243
4	Val.	-.326	-.322	-.204	-.017	.039	.163	.174	.186	.526	

属性片側集中行列：

n.	Form	7:Huelva	6:Sevilla	2:Pamplona	8:Tenerife	3:Alcalá	5:Barcelona	4:Madrid	9:Las Palmas	1:Oviedo	Val.
1	3:NO	1	3	2	2	1	1	5	1	2	1.000
2	2:OIGO	6	11	11	11	5	13	9	17	19	.460
3	1:DIGO	13	10	8	7	5	6	7	6	6	.000
4	Val.	.000	.481	.677	.793	.905	.918	.927	.964	1.000	

Proyecto VARIGRAMA:

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/varigrama/index.html>

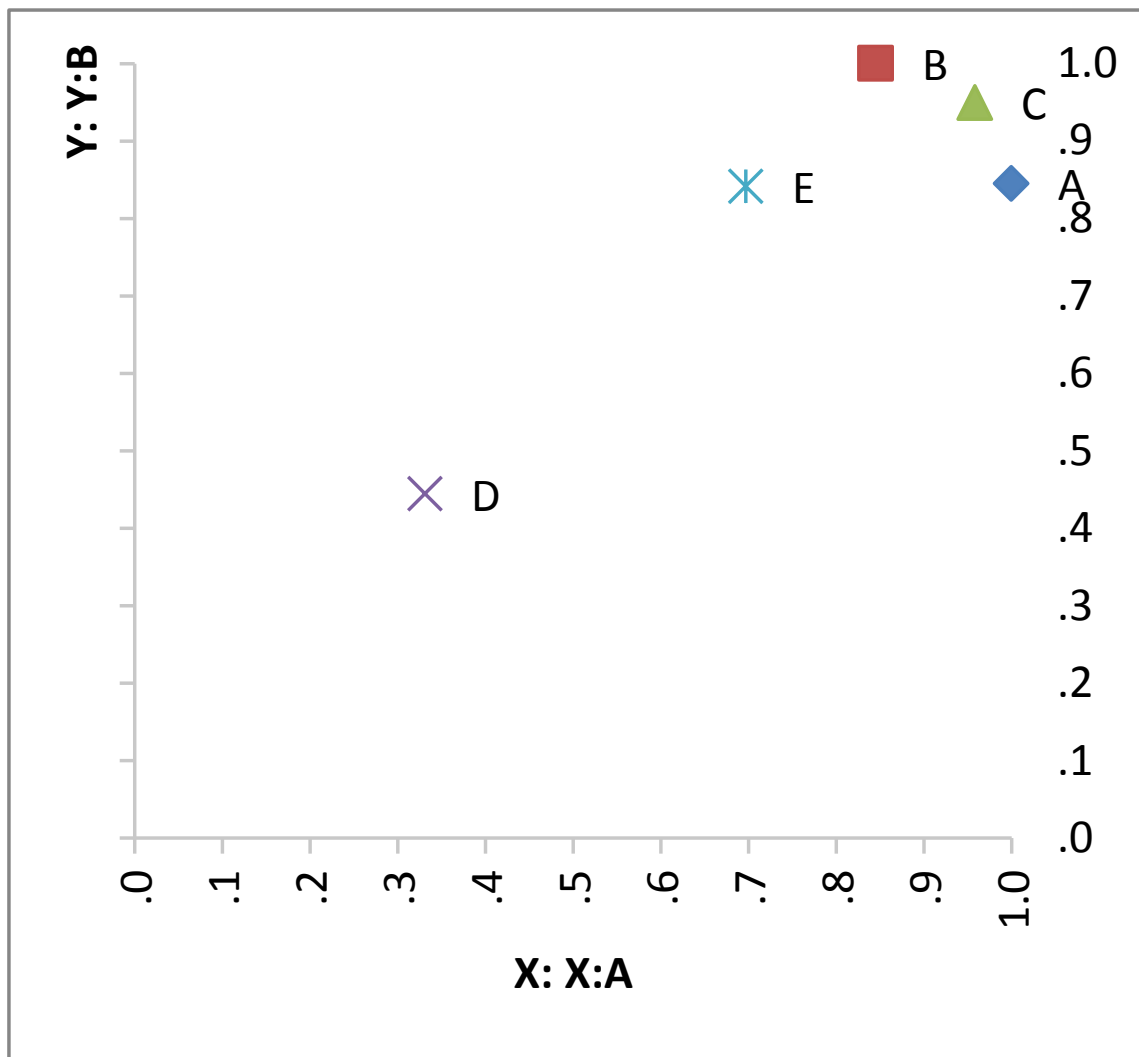


## 6. 選択軸分析

### 6.1. 属性の選択軸分析

データ行列(M)とその規定近接係数行列(Regular Proximity: RP)

M	A	B	C	D	E	RP	A	B	C	D	E
d1	10	19	14	7	12	A	1.000	.845	.958	.331	.697
d2	11	7	10	0	1	B	.845	1.000	.949	.444	.841
d3	0	0	1	12	1	C	.958	.949	1.000	.461	.811
d4	0	1	2	3	3	D	.331	.444	.461	1.000	.588
						E	.697	.841	.811	.588	1.000



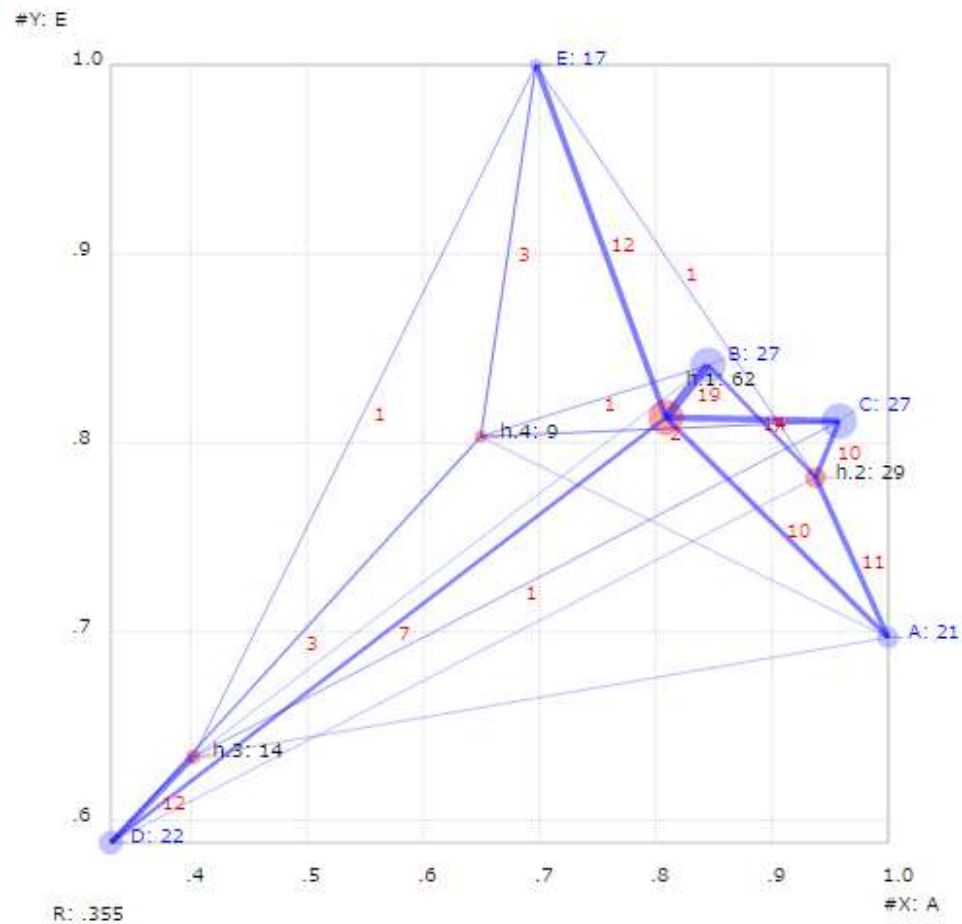
A 軸（横軸）はそれぞれの列と A の「近さ」を示し、B 軸（縦軸）はそれぞれの列と B の「近さ」を示す。

RP(A, B)	X:A	Y:B
A	1.000	.845
B	.845	1.000
C	.958	.949
D	.331	.444
E	.697	.841

【図-4】

軸の選択の可能性の1つとして相関係数の絶対値が最小になる2つの軸を考える。そこで、先の規定近接係数行列(RP)の列の相関行列  $Co(RP)$  を計算すると次のようになるので、全体の相関係数の中から最もゼロ(0)に近い値を探すと  $Co(A, E) = .355$  が見つかる。

$Co(RP)$	A	B	C	D	E
A	1.000	.876	.978	-.980	.355
B	.876	1.000	.956	-.898	.661
C	.978	.956	1.000	-.970	.498
D	-.980	-.898	-.970	1.000	-.430
E	.355	.661	.498	-.430	1.000



よって、次のように  $RP(A, E)$  のプロットがすべての組み合わせ  $Co(RP)$  の中で最もデータの分布を差異化する。

$RP(A, E)$	$x = A$	$y = E$
A	1.000	.697
B	.845	.841
C	.958	.811
D	.331	.588
E	.697	1.000

【図-5】

## 6.2. 個体の選択軸分析

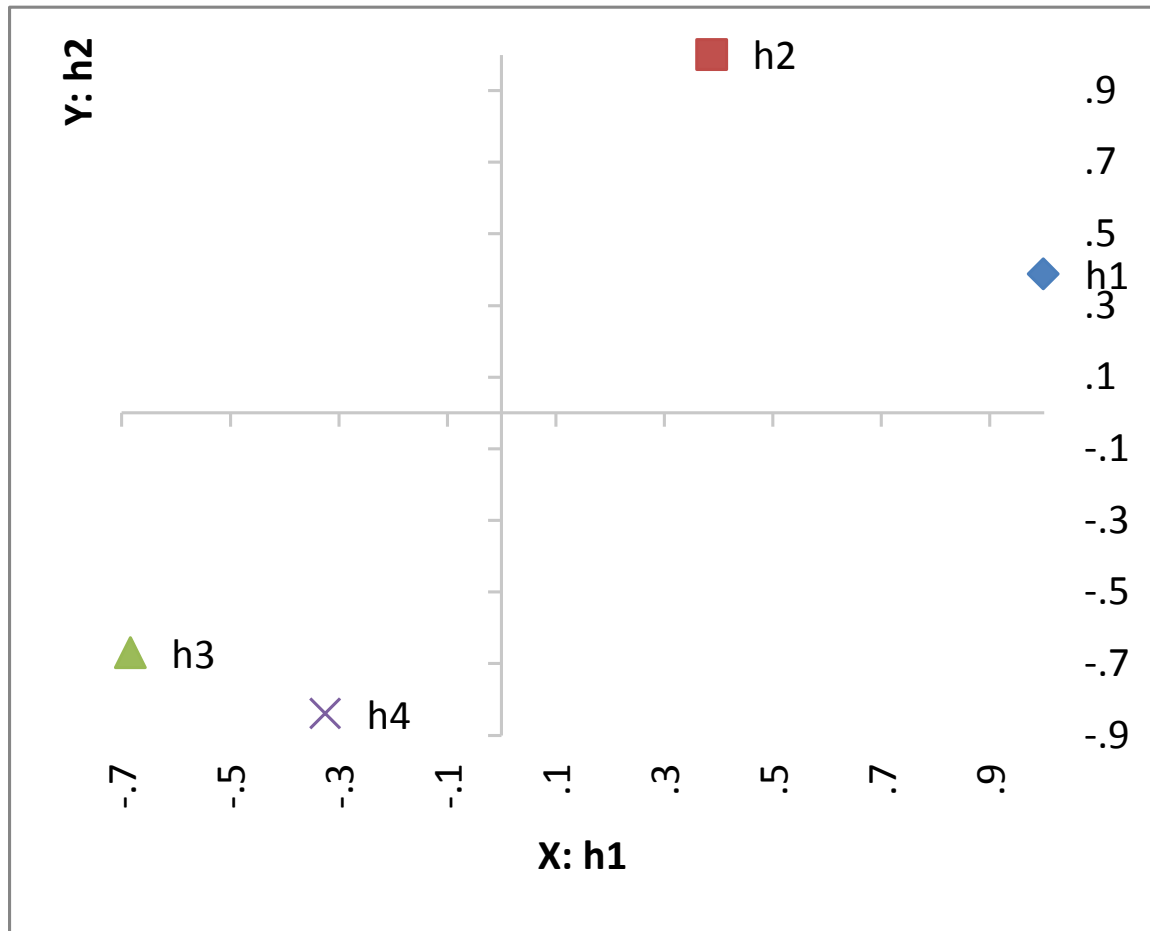
転置行列  $T(M)$  と転置行列の相関行列  $Co(T(M))$

M	A	B	C	D	E
h1	10	19	14	7	12
h2	11	7	10	0	1
h3	0	0	1	12	1
h4	0	1	2	3	3

T(M)	h1	h2	h3	h4
A	10	11	0	0
B	19	7	0	1
C	14	10	1	2
D	7	0	12	3
E	12	1	1	3

Co(T(M))	X:h1	Y:h2	h3	h4
h1	1.000	.387	-.683	-.323
h2	.387	1.000	-.670	-.840
h3	-.683	-.670	1.000	.586
h4	-.323	-.840	.586	1.000

次が個体 h1, h2 を選択した選択軸分析の結果。この結果から個体は大きく h1-h2 と h3-h4 に分けることができる。

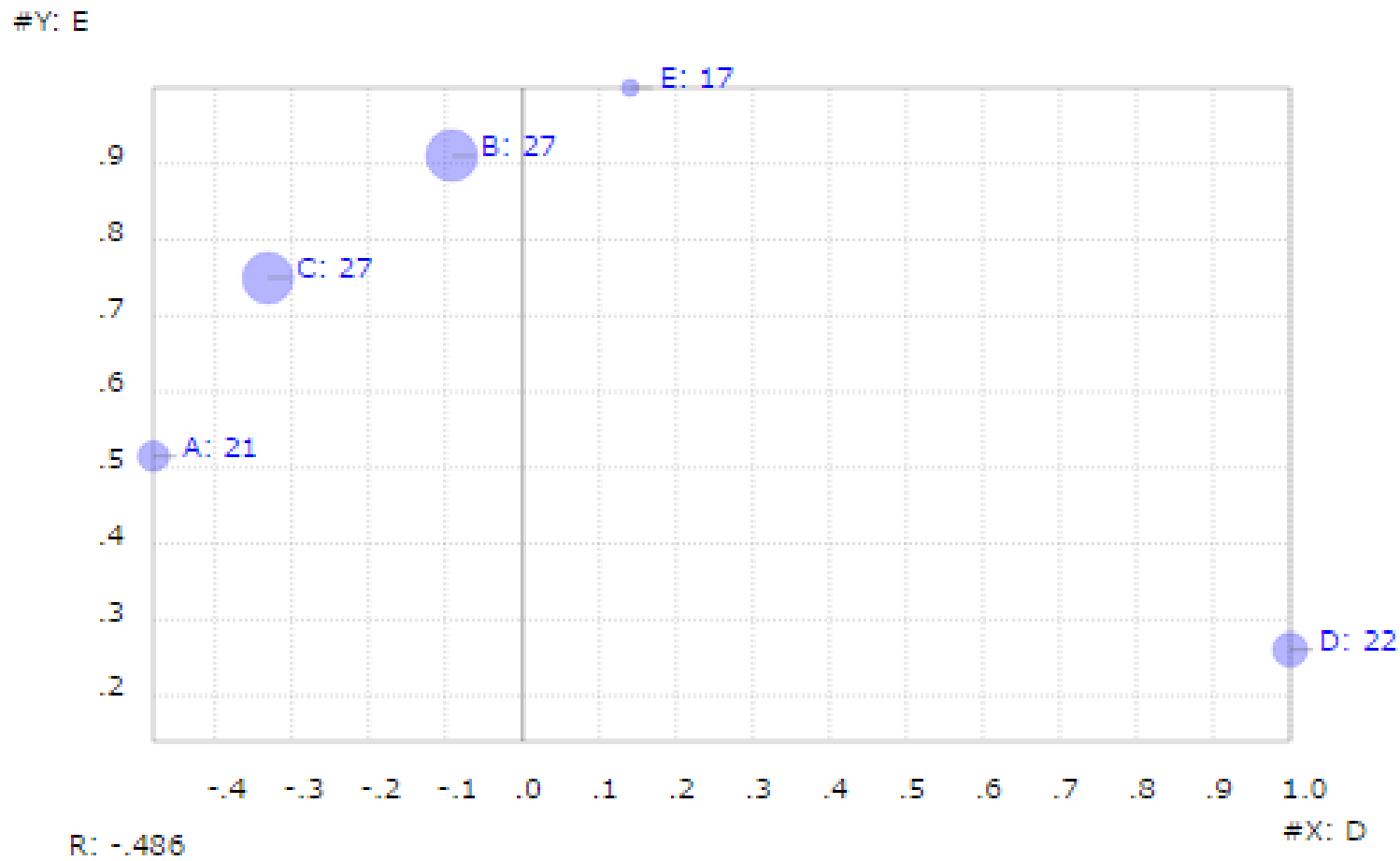


【図-6】

## 6.3. 属性・個体の選択軸分析

個体と属性を同じ平面にプロットする方法：相関係数(Co)

Mnp	A	B	C	D	E	Sn1	Co(M)	X:A	Y:B	C	D	E
h1	10	19	14	7	12	62	A	1.000	.787	.944	-.480	.436
h2	11	7	10	0	1	29	B	.787	1.000	.945	-.092	.896
h3	0	0	1	12	1	14	C	.944	.945	1.000	-.331	.709
h4	0	1	2	3	3	9	D	-.480	-.092	-.331	1.000	.140
S1p	21	27	27	22	17	114	E	.436	.896	.709	.140	1.000



【図-7】

この平面の中に個体 h1, h2, h3, h4 を合理的にプロットする方法を考える。



Mnp	A	B	C	D	E	Sn1	Co(D, E)	X:D	Y:E
h1	10	19	14	7	12	62	A	-.480	.436
h2	11	7	10	0	1	29	B	-.092	.896
h3	0	0	1	12	1	14	C	-.331	.709
h4	0	1	2	3	3	9	D	1.000	.140
S1p	21	27	27	22	17	114	E	.140	1.000

個体 h1 の重心の X 座標  $X(h1)$  と Y 座標  $Y(h1)$  :

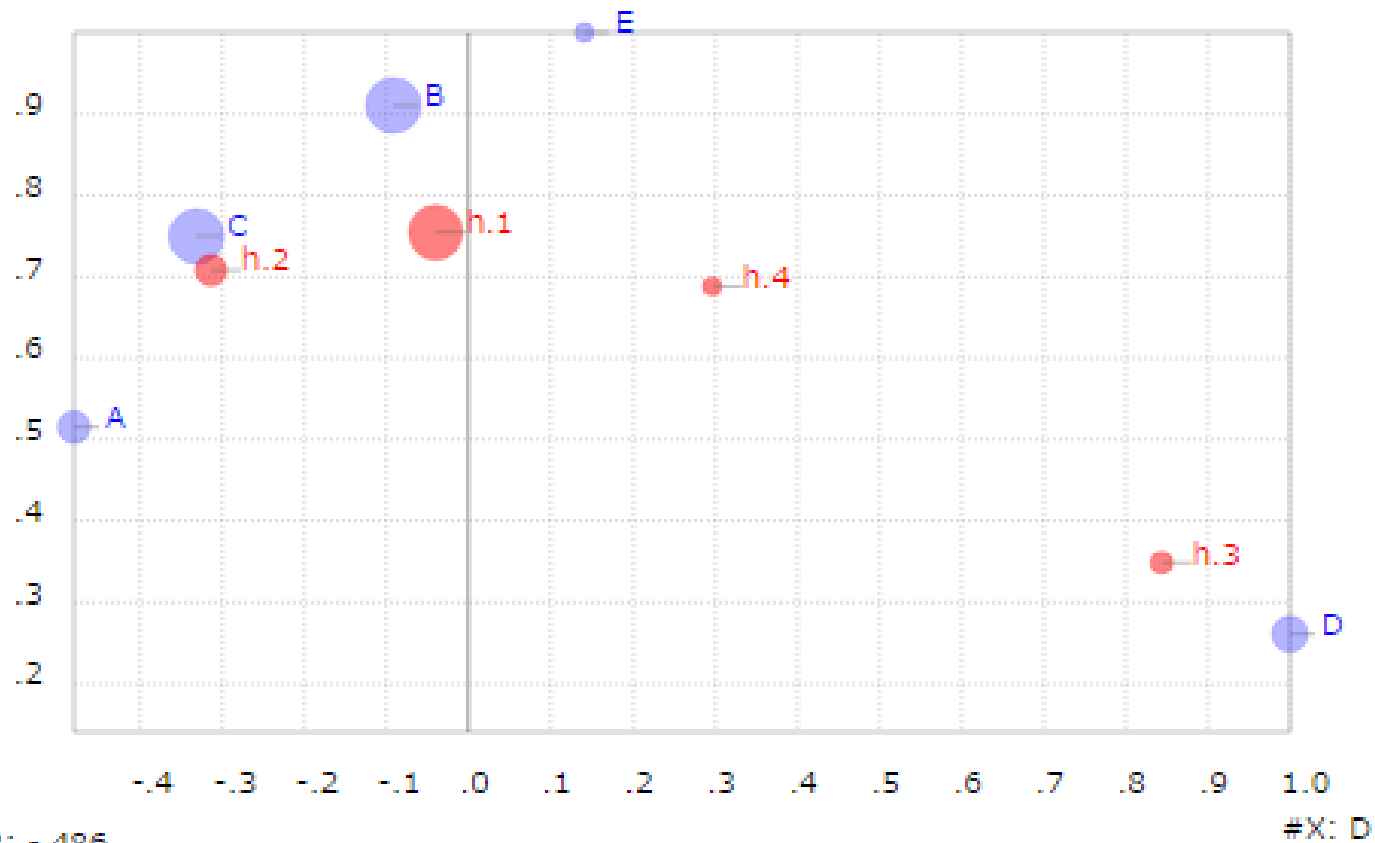
$$X(h1) = (10 * -.480 + 19 * -.092 + 14 * -.331 + 7 * 1.000 + 12 * .140) / 62 = -.040$$

$$Y(h1) = (10 * .436 + 19 * .896 + 14 * .709 + 7 * .140 + 12 * 1.000) / 62 = .714$$

Cn2	X:D	Y:E
h.1	-.040	.714
h.2	-.314	.661
h.3	.844	.242
h.4	.296	.637

Cn2 を x, y の座標を使ってプロットすると

#Y: E

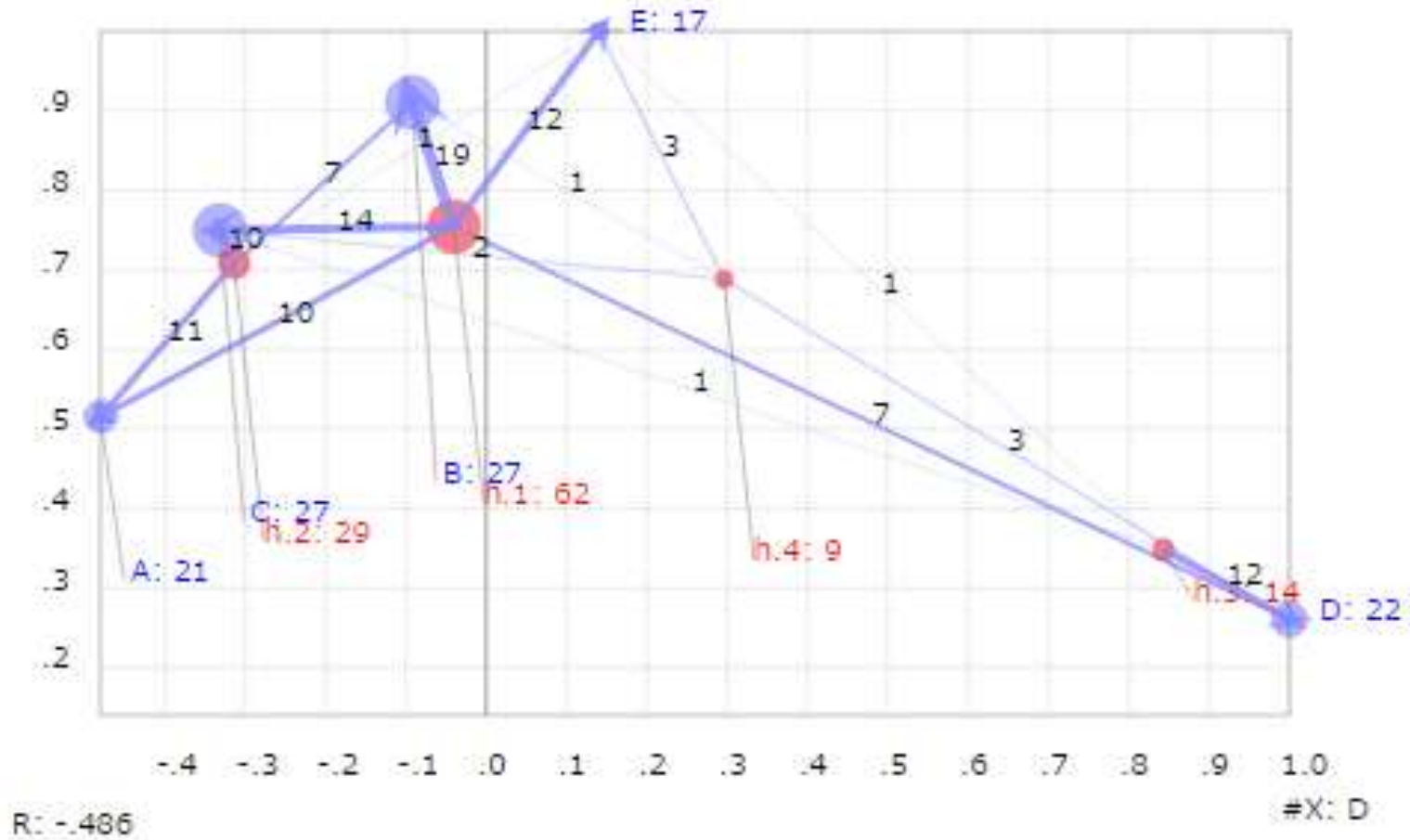


R: -.486

【図-8】

關係線：

#Y: E



【圖-9】

## ● 多変量解析と選択軸分析

### 多変量解析（主成分分析・対応分析・因子分析）

- ・ 高度な数学的準備（行列の微分、逆行列、固有値問題など）
- ・ 熟練したプログラミング技術
- ・ 多くの分析の実際的な適用の経験
- ・ プロット図で選択された 2 軸は全体の固有値・固有ベクトルの中から最大の 2 個だけを使うため情報の消失がある。
- ・ 軸の解釈が同じデータであっても分析者の観察によって異なることがある。
- ・ 多変量解析には多くの応用例がある。

## 選択軸分析

- ・ 単純な論理
- ・ 簡単なプログラム
- ・ プロットの軸には該当する列の名称がそのまま使われるので、そもそも「解釈」する必要がなく、むしろそれぞれの軸が他の軸とどのような関係にあるのかを観察するだけでよい。
- ・ 実際の分析において自由に軸を選択・合成できるので、分析者は一定の方法で出力された結果に縛られることがなく、納得がいくまで実験を重ねることも可能。
- ・ 全属性列を2分割したクラスター選択軸を使えば、選択・合成された軸によってプロットされた図は対象の相関行列・類似行列・距離行列の全情報を使っているので情報の損失はない。（一方、個別の選択

軸を使うと分析対象の視点を限定され焦点化される。)

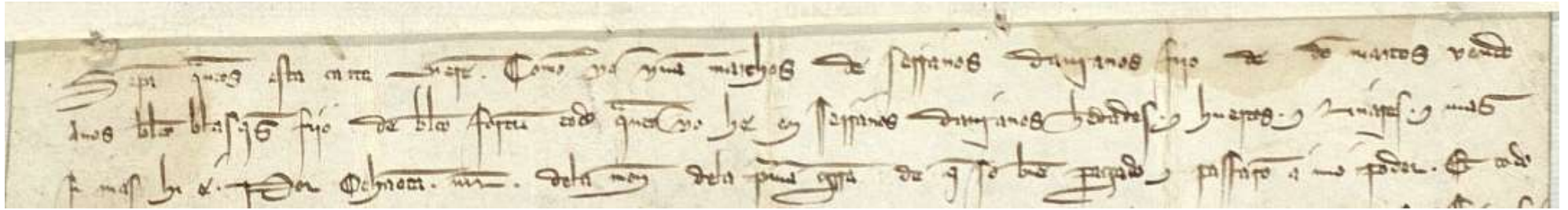
- ・しかし、選択軸分析では分析者が自由に軸を選択・合成するので、分析者の恣意（自由の裏返し）が避けられない。

- ・選択した行列の種類（相関行列・類似行列・距離行列）や選択した軸の違いによってプロット図は当然大きく異なる。よって分析者はそれらの違いを意識して使用し、発表するときにも明示化して説明できなければならない。

- ・選択軸分析の応用を開始する。

## 7. 分析例

### 7.1. 定型表現



GITHE (Grupo de Investigación Textos para la Historia del Español):  
CODEA+ 2015 (Corpus de documentos españoles anteriores a 1800) [en  
línea] Sitio web:

<http://corpuscodea.es/> (2017/2/2)

CODEA: 0065



GITHE

CODEA+ 2015

AHN, Clero, Ávila, carpeta 22, nº 2

1285 septiembre 5 (Ávila, Ávila, España)

Particular

Cartas de compraventa y contratos: compraventa

Letra gótica cursiva [albalaes]

{1} Sepan quantos esta carta vieren Como yo yuan marchos de serranos daujanos fijo de domingo marcos vendo

{2} auos blasco blasquez fijo de blasco fortun todo quanto yo he en sserranos daujanos heredades & huertos & linares. & mas

{3} si mas hi e. Por Ochaenta. marauedis dela moneda dela primera guerra de que so bien pagado & passaron a mio poder.

{1} Sepan quantos esta carta vieren cómo yo Yuan Marcos, de Serranos de Avianos, fijo de Domingo Marcos, vendo

{2} a vós, Blasco Blásquez, fijo de Blasco Fortún, todo quanto yo é en Serranos d'Avianos, heredades, e huertos, e linares e más

{3} si más y é por ochaenta maravedís de la moneda de la primera guerra, de que só bien pagado, e passaron a mio poder.

(1) 頻度：出現数（比較できない）

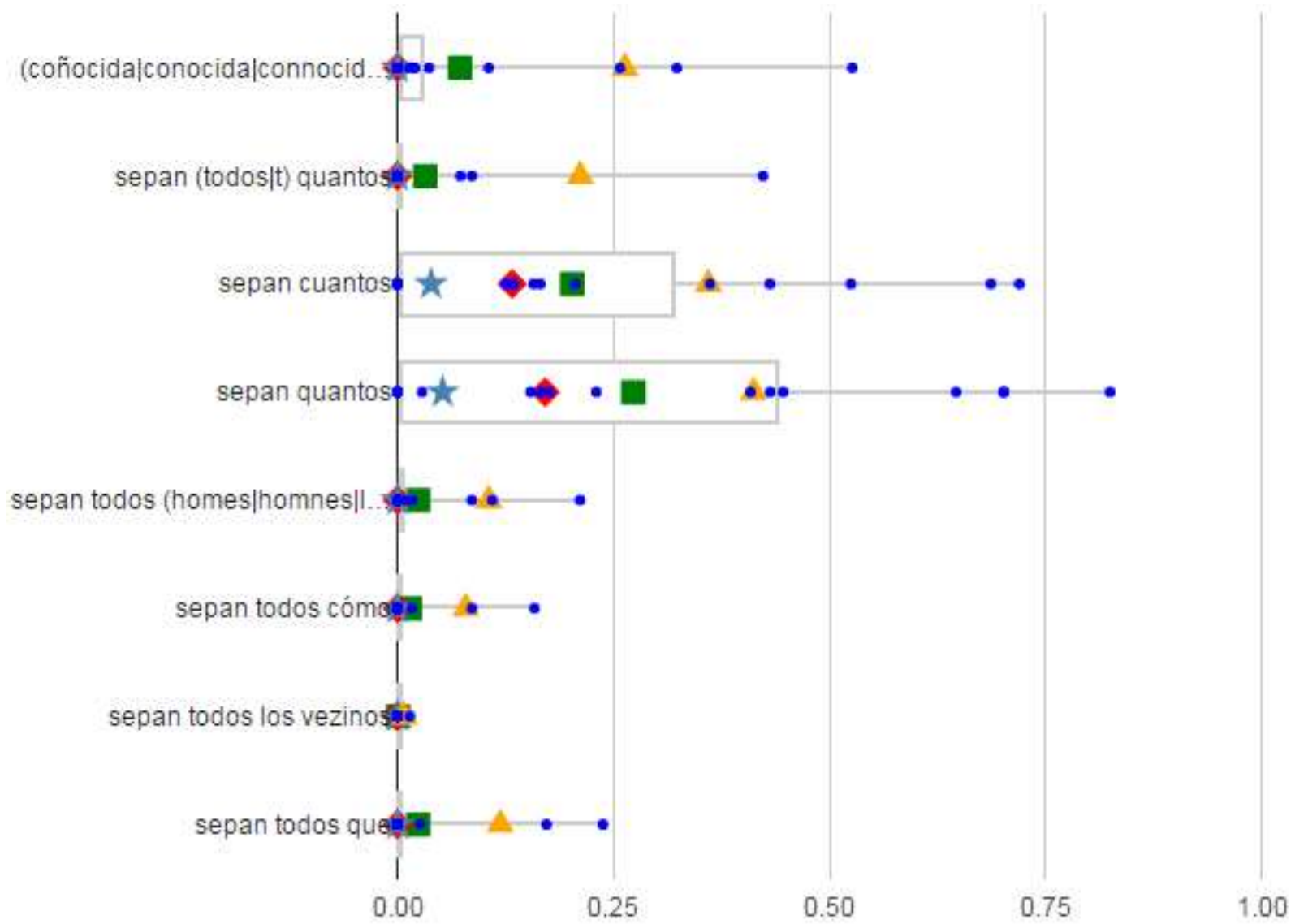
Cnp	A.1200	A.1300	A.1400	A.1500	A.1600	A.1700
conocida cosa	24	5	4			
sepan todos quantos	1	16	4			
sepan cuantos	44	62	17	22	16	1
sepan quantos	34	50	72	27		
sepan todos homnes	2	8	6	1		
sepan todos	1	6		1		
sepan todos los vezinos				2		
sepan todos que	2	1	13			
Sum	108	148	116	53	16	1
& (Words)	90493	146283	215212	232583	109063	18818

(2) 縦相対頻度 (%) : 取り出した語形だけを対象 (比較危険)

Cnp (%)	A.1200	A.1300	A.1400	A.1500	A.1600	A.1700
conocida cosa	22.2	3.4	3.4			
sepan todos quantos	.9	10.8	3.4			
sepan cuantos	40.7	41.9	14.7	41.5	100.0	100.0
sepan quantos	31.5	33.8	62.1	50.9		
sepan todos homnes	1.9	5.4	5.2	1.9		
sepan todos	.9	4.1		1.9		
sepan todos los vezinos				3.8		
sepan todos que	1.9	.7	11.2			

(3) 千語率：該当する全語形を対象（比較的安全）

Cnp (PMW)	A.1200	A.1300	A.1400	A.1500	A.1600	A.1700
conocida cosa	.265	.034	.019			
sepan todos quantos	.011	.109	.019			
sepan cuantos	.486	.424	.079	.095	.147	.053
sepan quantos	.376	.342	.335	.116		
sepan todos homnes	.022	.055	.028	.004		
sepan todos	.011	.041		.004		
sepan todos los vezinos				.009		
sepan todos que	.022	.007	.060			



LE.(西) Abs	1200	1250	1300	1350	1400	1450	1500	1550
conocida cosa	4	2				1		
sepan cuantos			14					
sepan quantos		2	9	13	21	14	12	1
Words	3786	7608	15291	11384	26885	22967	15067	3462

LE(西) PMW	1200	1250	1300	1350	1400	1450	1500	1550
conocida cosa	1.057	.263				.044		
sepan cuantos			.916					
sepan quantos		.263	.589	1.142	.781	.610	.796	.289

CV.CN. (中央) Abs	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650	1700
conocida cosa	5	13		1	1						
sepan cuantos		44	41	7	15	2	7	15	12	4	1
sepan quantos		32	14	14	26	11	12	2			
sepan todos homnes		1						1			
sepan todos cómo							1				
sepan todos los vezinos								2			
Palabras	6901	60608	43830	37963	54471	56252	81848	117990	47068	51210	8838



CV.CN.(中央) PMW	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650	1700
conocida cosa	.725	.214		.026	.018						
sepan cuantos		.726	.935	.184	.275	.036	.086	.127	.255	.078	.113
sepan quantos		.528	.319	.369	.477	.196	.147	.017			
sepan todos homnes		.016						.008			
sepan todos cómo							.012				
sepan todos los vezinos								.0			

AR.(東) Abs	1200	1250	1300	1350	1400	1450
coñocida cosa			2	2	2	
sepan todos quantos		1	4	12	3	1
sepan todos homnes		1	1	7	5	1
sepan todos cómo		1		6		
sepan todos que		2		1	8	5
Palabras	519	11071	8934	28881	35278	19359

AR.(東) PMW	1200	1250	1300	1350	1400	1450
conocida cosa			.224	.069	.057	
sepan todos quantos		.090	.448	.415	.085	.052
sepan todos homnes		.090	.112	.242	.142	.052
sepan todos cómo		.090		.208		
sepan todos que		.181		.035	.227	.258

conocida cosa のほうが sepan cuantos より 相対年代が早い  
 後出の sepan cuantos は CV.CN に 1250

conocida cosa: LE > CN.CV > AR

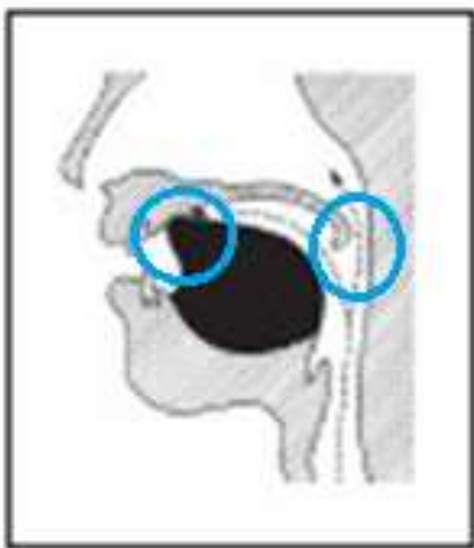
sepan cuantos: CV.CN > AR. LE

## 7.2. 不規則未来形・推量形

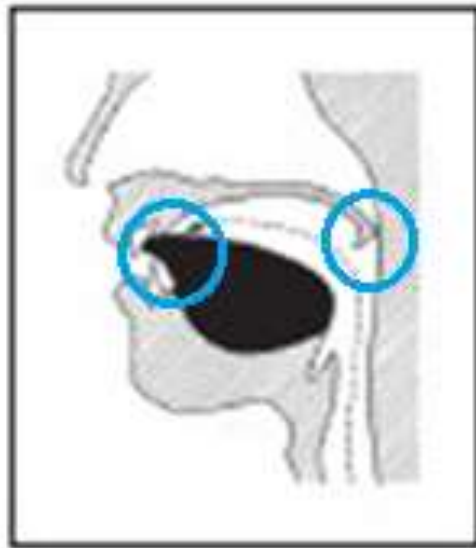
規則形：hablaré「話す」，comeré「食べる」，viviré「生きる」

不規則形：tendré「持つ」，vendré「来る」，pondré「置く」，etc.

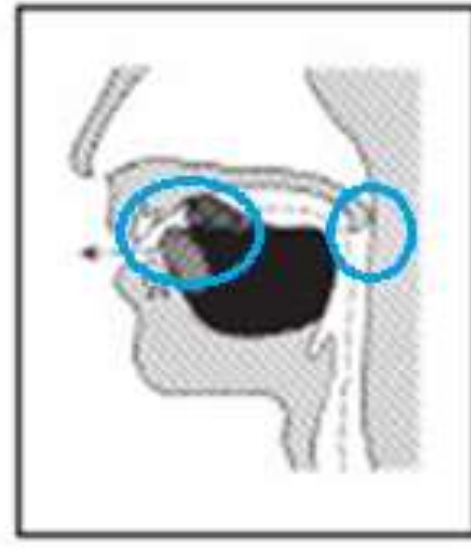
-ndr- (?)



[n]



[d]



[r]

調音	[n]	[d]	[r]
鼻腔解放	+	-	-
閉鎖	-	+	±

ten(e)re → tenre → tendre

しかし中世カスティーリャでは tendré ではなく -nr-が逆転した -rn-がある ternéが使われた。-rn- > -ndr-の音声変化は考えにくい。

ten(e)re → tenre → terne (\*?) → tendre

## 原点距離法による語形の片側集中分析 (Unilateral concentration)

A: アラゴン王国 (半島東) : 千語率片側集中分析

A	1100	1200	1250	1300	1350	1400	1450	1500
tendr			.090		.035	.397	.620	.145
vendr					.173	.283	.362	.072
pondr							.052	
vern			.271		.242	.312	.826	.361
tern			.452	.336	.069	.595	.620	

-ndr-は-rn-と共存しながらも 1250 年代から出現

C: カスティーリヤ王国（半島中央）：千語率片側集中分析

C	1100	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650	1700	1750
tendr									.076	.085	.059		.301
vendr									.025	.021			.100
pondr					.026		.018			.021			.100
porn							.018	.024	.008				
vern		.290	.049	.023	.053	.184	.284	.501	.364	.042	.059		
tern	8.646	.145		.023		.018	.160	.122	.119	.106	.273	.566	.200

-ndr-は 1350 年代に出現

L: レオン王国（半島西）：千語率片側集中分析

L	1100	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650
pondr										.203	.442
tendr							.044			.406	.442
porn										.203	
tern					.088	.037	.087			.203	
vern		.528	.131			.037	.174	.265	.289		.442

音声変化ではなく、語形の地理的移動・変異


-ndr-：アラゴン（東）→カスティーリャ（中央）→レオン（西）



## 7.3. 二重子音文字

CODEA «Corpus de Documentos Españoles Anteriores a 1700» (Pedro Sánchez Prieto Borja, GITHE: (Grupo de Investigación de Textos para la Historia del Español, Universidad de Alcalá). Contiene 1502 textos provenientes de distintas regiones de España de los siglos XI al XVII.

CORHEN: «Corpus Histórico del Español Norteño» (María Jesús Torrens Álvarez, CSIC). Contiene 253 textos del norte de Castilla la Vieja de los siglos X al XIII.


  
 (1) <ff>      (2) <f>      (3) <ss>      (4) <s>

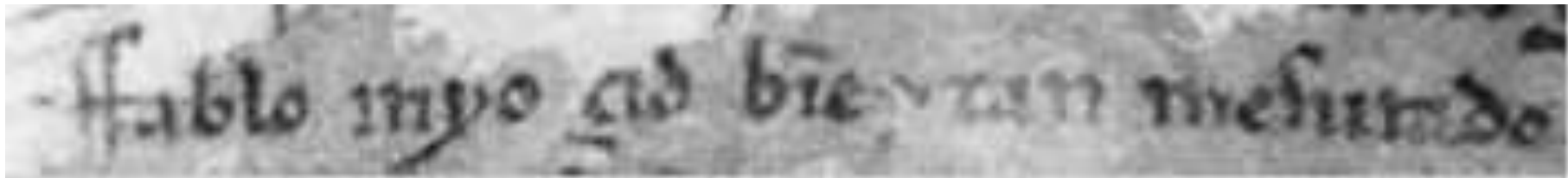


Fig. (a): LEMI, Cid, letra gótica libraria, ms. s. XIV  
 {7} *ffablo myo çid bie<n> e tan mesurado*

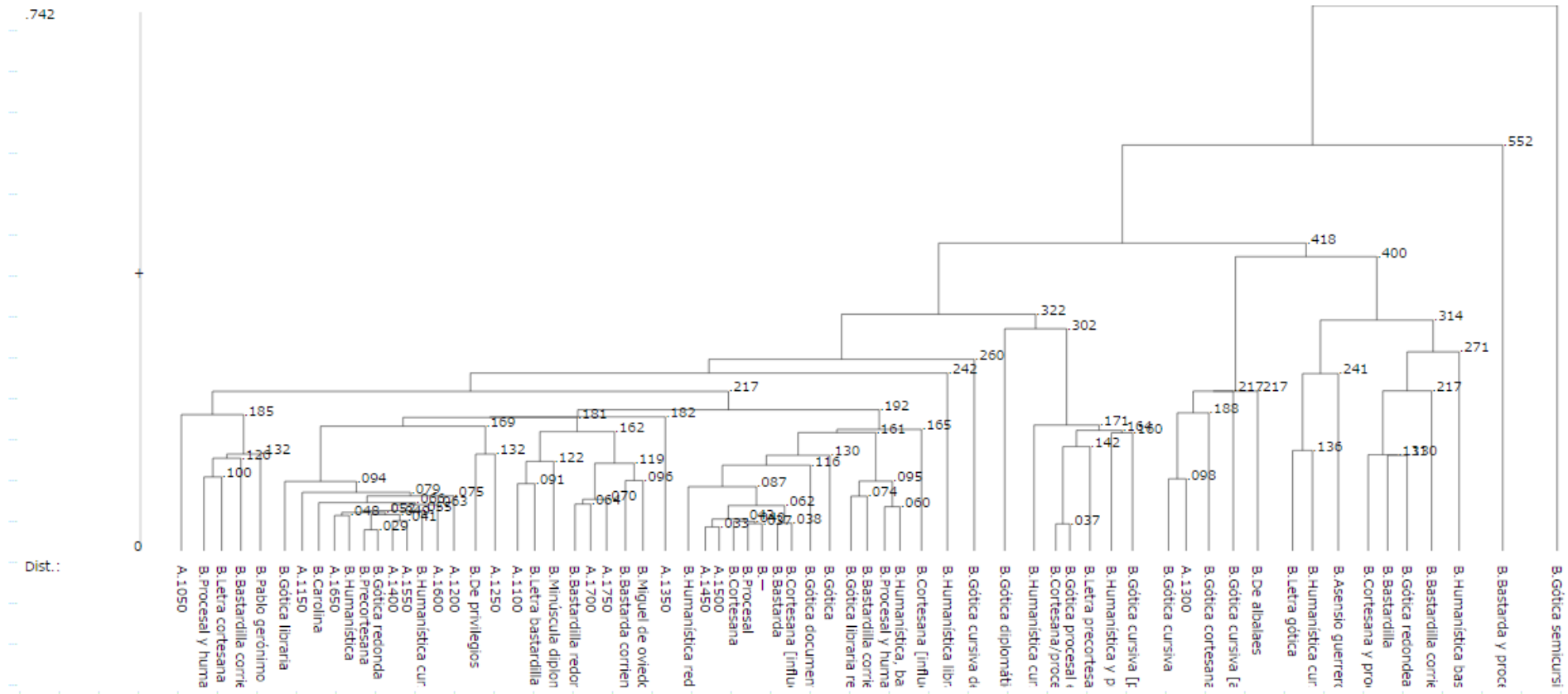
LEMI («Letras Españolas en Manuscritos e Impresos»)

CODEA / CORHEN: Castilla (Ueda, H. 2016)

C	1200	1225	1250	1275	1300	1325	1350	1375	1400	1425	1450	1475	1500	1525	1550	1575	1600
ffue / ffui			.050	.420	.580	1.470	1.320	.370	.090	.080			.020				
ffirmar								.040									
ffray			.230	.060	.090								.020				
fue / fui	.060	1.550	.550	.630	.470	.070	.220	1.240	.820	1.480	1.260	1.630	2.080	2.100	1.830	1.110	.740
firmar	.060	.040	.390	1.700	1.800		.110	.210	.250	.990	.650	.740	.690	.550	.750	.680	.660
fray			.010		.020				.160		.150	.700	1.150		.070	.240	

C	1300	1325	1350	1375
ffue / ffui	.580	1.470	1.320	.370
ffirmar				.040
ffray	.090			
fue / fui	.470	.070	.220	1.240
firmar	1.800		.110	.210
fray	.020			

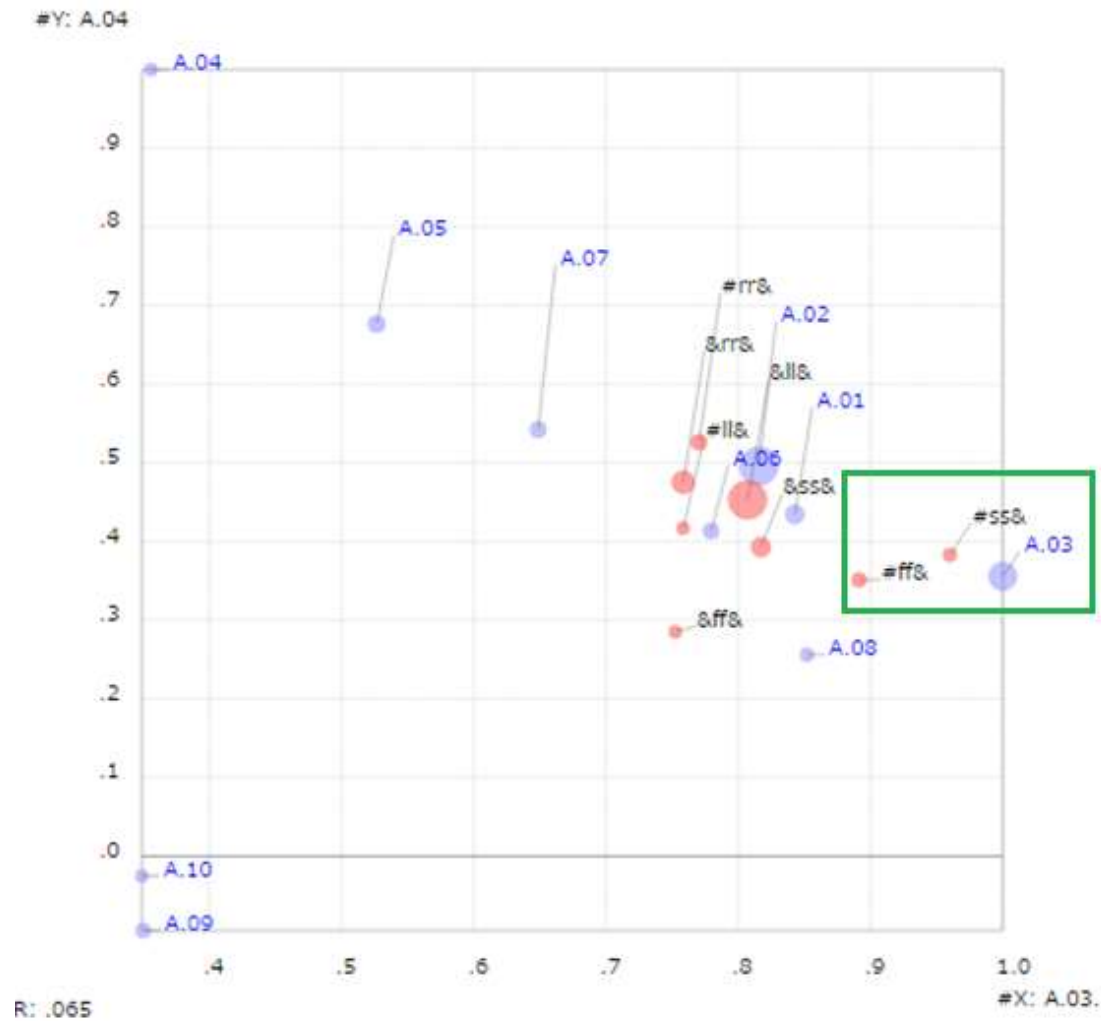
# クラスター



Group	Attribute
A.01	A.1050; B.Bastardilla corriente mercantil; B.Letra cortesana; B.Procesal y humanística; B.Pablo gerónimo bargón; B.Humanística libraria;
A.02	A.1100; B.Letra bastardilla redonda castellana; A.1150; B.Gótica libraria; B.Minúscula diplomática; A.1200; B.Carolina; A.1400; A.1550; A.1450; A.1500; B.Cortesana; B.Procesal; B.—; B.Gótica; B.Humanística redonda; A.1600; A.1650; B.Humanística; B.Gótica redonda; B.Precortesana; B.Humanística cursiva; B.Bastarda; B.Cortesana [influencia humanística]; B.Cortesana [influencia procesal]; A.1700; B.Bastardilla redonda castellana; A.1750; B.Humanística cursiva corriente; B.Bastarda corriente; B.Miguel de oviedo; B.Letra gótica;
A.03	A.1250; B.De privilegios; A.1350; B.Gótica documental; B.Humanística cursiva [redonda]; A.1300; B.Gótica cursiva; B.Gótica cursiva [albalaes]; B.De albalaes; B.Gótica cortesana;
A.04	B.Asensio guerrero;

A.05	B.Bastarda y procesal; B.Gótica cursiva [precortesana]; B.Humanística y procesal;
A.06	B.Bastardilla; B.Gótica redondeada; B.Cortesana y procesal; B.Bastardilla corriente castellana; B.Gótica libraria redonda; B.Humanística, bastarda; B.Procesal y humanística redonda; B.Bastardilla corriente;
A.07	B.Cortesana/procesal; B.Gótica procesal encadenada; B.Letra precortesana; B.Gótica diplomática;
A.08	B.Gótica cursiva de albaes;
A.09	B.Gótica semicursiva;
A.10	B.Humanística bastarda española;

# 選択軸分析 : [A:03] (Interest) - [A:04] Minimal correlation



[A.04] と #ss / #ff: [A:08]

文字種の集中分析：(Ueda, H. 2016)

Letra	1075	1100	1125	1150	1175	1200	1225	1250	1275	1300	1325	1350	1375	1400	1425	1450	1475	1500	1525	1550	1575	1600	1625	1650
(1) Carolina	1	2	1	7	2	3	3	1																
(2) G. libraria				1	5	14	48	45	13	8	3	3		1										
(3) De privilegios							1	4	5															
(4) Gótica				1		10	12	8	1		1			1	1		1							
(5) G. c. albalaes									22	1														
(6) De albalaes								2	14	4	1	1												
(7) G. cursiva								14	32	46	29	25	7	2										
(8) G. redonda					1	5	5	4	3						1	2	3	4	1					
(9) G. cursiva p. (10) Precortesana												5	6	20	3			3						
(11) Cortesaa													6	4	12	30	9	20	8	1	1			
(12) H. redonda														1			1	1	1	11	3	3		
(13) Procesal																1			2		2		1	
(14) H. cursiva																			1	3	2	4	8	4



## 参考文献：

- ALARCOS LLORACH, EMILIO. (1979): «Bases para un comentario diacrónico (fonético y fonológico)», en *Comentarios lingüísticos de textos*. Universidad de Valladolid, pp. 7-18.
- BASSOLS DE CLIMENT, Mariano. (1976): *Fonética latina*. Madrid, CSIC.
- BLAKE, ROBERT. J. (1988a): «Aproximaciones nuevas al fenómeno de [f] > [h] > [Ø]. *Actas del I Congreso Internacional de Historia de la Lengua Española*, I, Madrid Arco Libros, 71-88.
- \_\_\_\_\_. (1988b): «*Ffaro, Faro or Haro?*: F doubling as a source of linguistic information for the early middle ages», *Romance Philology*, 41, pp, 267-289.
- CATALÁN, DIEGO. (1967): «La pronunciación [ihante] por /iffante/ en la Rioja del siglo XI. Anotaciones a una observación dialectológica de un historiador árabe», *Romance Philology*, 21, 410-485, recogido en D. Catalán (1989): *El español. Orígenes de su diversidad*. Madrid, Paraninfo, 267-293.

- CHOMSKY, NOAM. / HALLE, MORRIS. (1968): *The sound pattern of English*. New York, Harper and Row.
- DEROLEZ, ALBERT. (2003): *The paleography of Gothic manuscript books. From the twelfth to the early sixteenth century*. Cambridge, Cambridge University Press.
- HANSSEN, FEDERICO. (1913): *Gramática histórica de la lengua castellana*. Paris, Ediciones Hispano-americanos.
- JAKOBSON, R, FANT, C. G. M. / HALLE, MORRIS. (1951): *Preliminaries to speech analysis. The distinctive features and their correlates*. The M. I. T. Press.
- LAPESA, RAFAEL. (1981): *Historia de la lengua española*. Madrid, Gredos.
- LAUSBERG, HEINRICH. (1976): *Lingüística románica. I. Fonética*. Madrid, Gredos.
- LÁZARO CARRETER, FERNANDO. (1971): *Diccionario de términos filológicos*. Madrid, Gredos.
- MARÍN MARTÍNEZ, TOMÁS. (1996): *Paleografía y diplomática*. vol. 1. Madrid, Universidad

Nacional de Educación a Distancia.

MARTÍNEZ CELDRÁN, EUGENIO. (1994): *Fonética*. Barcelona, Teide.

\_\_\_\_\_ / FERNÁNDEZ PLANAS, ANA MARÍA (2007): *Manual de fonética española. Articulaciones y sonidos del español*. Barcelona, Ariel.

MENÉNDEZ PIDAL, RAMÓN. (1906): «El dialecto leonés», *Revista de Archivos, Bibliotecas y Museos*, Año 10, 128-172, 294-311.

\_\_\_\_\_. (1968): *Manual de gramática histórica española*. 13.<sup>a</sup> ed. Madrid, Espasa-Calpe.

\_\_\_\_\_. (1969): *Cantar de Mío Cid. vol. III. Texto del Cantar y adiciones*. Madrid, Espasa-Calpe.

\_\_\_\_\_. (1980): *Orígenes del español*. Madrid, Espasa-Calpe.

METZELTIN, MICHAEL. (1979): *Altspanisches Elementarbuch. I. Das Altkastilische*. Heidelberg, Carl Winter. Universitätsverlag.

MINISTERIO DE EDUCACIÓN Y CIENCIA. (1961): *Poema de Mio Cid. Edición facsímil del Códice*

- de Per Abat, conservado en la Biblioteca Nacional. Madrid, Ministerio de Educación y Ciencia.*
- MORTERERO Y SIMÓN, CONRADO. (1979): *Apuntes de iniciación a la paleografía española de los siglos XII a XVII*, Madrid, CSIC.
- MUÑOZ Y RIVERO, JESÚS. (1972): *Manual de paleografía diplomática española de los siglos XII al XVII*. Madrid, Lope de Vega.
- NIEDERMANN, MAX. 1997. *Précis de phonétique historique du latin*. Paris, Klincksiek.
- PENNY, RALPH. (2006): *Gramática histórica del español*. Barcelona, Ariel.
- PENSADO RUIZ, CARMEN. (1993): «Sobre el contexto del cambio "F> h" en castellano». *Romance Philology*, 47/2, 147-176.
- POSSNER, REBECCA. (1966): *The Romance languages. A linguistic introduction*. Peter Smith.
- RIESCO TERRERO, ÁNGEL. / RUIZ GARCÍA, ELIZA., DOMÍNGUEZ APARICIO, JESÚS., SÁNCHEZ PRIETO, Ana Belén. (1995): *Aproximación a la cultura escrita. Material de apoyo*.

Madrid, Playor.

RUIZ ASENCIO, J. M. (2008): «Propuesta de elaboración de unas normas de transcripción de textos castellanos medievales», en Beatriz Díez Calleja (ed.): *El primitivo romance hispánico*. Burgos, Instituto Castellano y Leonés de la Lengua, 137-143.

SÁNCHEZ-PRIETO, PEDRO. (ed.): (1995): *Textos para la historia del español. t. II. Archivo Municipal de Guadalajara*. Universidad de Alcalá de Henares.

\_\_\_\_\_. (1998) «Para una historia de la escritura en Castilla», *Actas del IV Congreso Internacional de Historia de la Lengua Española, La Rioja, 1-5 abril de 1997*, editadas por Claudio García Turza, Fabián Bachiller y Javier Mangado, Logroño, vol. I, 289-301.

\_\_\_\_\_. (2004): «La normalización del castellano escrito en el siglo XIII. Los caracteres de la lengua: grafías y fonemas» en R. Cano Aguilar (ed.): *Historia de la lengua española*. Barcelona, Ariel. 423-448.

\_\_\_\_\_. (2012): «Para una historia de la escritura romance en León, Castilla y Aragón: algunas claves interpretativas»”, *Medioevo Romanzo*, XXXVI(2012):, 24-61.

TAMAYO, ALBERTO. (2012): *Historia de la escritura latina e hispánica*. Gijón, Ediciones Trea.

TORREBLANCA, MÁXIMO. (1983-84): «La "F" prerromana y la vasca en su relación con el español antiguo». *Romance Philology*, 37, 273-281.

\_\_\_\_\_. (1991-92): «Sobre los orígenes de la distinción fonológica /f/:/h/ en el castellano medieval». *Romance Philology*, 45/3, 369-409.

TORRENS ÁLVAREZ, MARÍA JESÚS. (2002): *Edición y estudio lingüístico del Fuero de Alcalá (Fuero Viejo)*: Alcalá de Henares, Fundación Colegio del Rey.

\_\_\_\_\_. (2007): *Evolución e historia de la lengua española*, Madrid, Arco/Libros.

UEDA, HIROTO. (2015): «Grafías consonánticas dobles en el español medieval y el origen de la letra ñe», VIII Congreso Internacional de la Asociación Asiática de Hispanistas, Universidad de

Estudios Extranjeros de Shanghái, 964-976.

\_\_\_\_\_. (2016): «La grafía doble <ff> en el castellano medieval: Similitud paleográfica y lingüística con la doble <ss> alta», *Revista de Historia de la Lengua Española*, 10, pp. 105-132.

\_\_\_\_\_ / MORENO-SANDOVAL, Antonio. (2015): «LETRAS and NÚMEROS: two integrated web-based tools for research in Linguistics and Humanities», VII Congreso Internacional de Lingüística del Corpus, Universidad de Valladolid.

WILLIAMS, EDWIN (1962): *From latin to Portuguese*. Philadelphia. Univ. of Pennsylvania Press.

ZAMORA VICENTE, ALONSO. (1967): *Dialectología española*, 2a. ed. Madrid, Gredos.

Fin