Sixteenth International Conference on Methods in Dialectology (METHODS XVI)
National Institute for Japanese Language and Linguistics, Tokyo, 10 August, 2017.

Unilateral Correspondence Analysis

Applied to Spanish linguistic data in time and space

Hiroto Ueda, University of Tokyo

1. Introduction

The correspondence analysis represented by Jean-Paul Benzécri (1973) is used in different scientific disciplines (Clausen 1998, Greenacre 2013). Previously in Japan the Quantification Method Type III of Chikio Hayashi (1956) had been also well known and widely utilized (Akido 1974, Yasuda / Umino 1977, Asano 2000, Kosugi 2007). In dialectology Inoue (1996, 1997) and Inoue and Fukushima (1997) are good examples of its utilization. The aim in common is to find the distribution of the frequencies in the two-dimensional matrix with the highest correlation coefficient, calculating the constitution of the two vectors, horizontal and vertical, which are weighted to the frequencies. As a result the frequencies are presented in a concentrated proximity of the diagonal area, by which we can examine the relation of cases (linguistic forms), the relation of attributes (linguistic and extralinguistic parameters), and the relation of both cases and attributes at the same time.

The analysis of Benzécri as well as the method of Hayashi are characterized by the rearrangement of cases in rows and attributes in columns in order to obtain the best form of diagonalization in the distribution of frequencies. However, researchers sometimes need to rearrange only the order of cases fixing the order of attributes and vice versa. For example, in the historical data the chronological axis of years, decades or centuries should be maintained unchanged in order to study the linguistic forms in accordance with chronological parameter.

In this occasion we propose to carry out the reordering of the rows of the linguistic forms provided that the columns corresponding to, for example, the chronological attributes are not to be moved. In the presentation I will explain the mathematical derivation of the Unilateral Correspondence Analysis in comparison with that of the usual Bilateral Correspondence Analysis and its applications to Spanish dialectal data and historical documents. In addition I propose a method of finding the most concentrated zones within the optimally diagonalized matrix. I believe that the Bilateral and the Unilateral Correspondence Analyses together with the search for the most concentrated areas will allow not only wide range of uses in dialectometric studies but also in other scientific fields in general.

2. Method

2.1. Bilateral correspondence analysis

The method of «Correspondence Analysis» consists in searching for the weight values in both the cases (x_1, x_2, ...) and the variables (y_1, y_2, ...) in the data table D_{np} in form of the concentrated table C_{np} so that the distribution of the frequencies present the highest possible correlation. The values inferred to the two axes are used to see their weights and also to rearrange the table. The result allows us to interpret the frequency distribution in diagonally concentrated form[1].

[1] In this section we follow basically the explanations of Mino (2001: 162-164) and Takahasi (2005: 106-124).

D_{np}	y_1: English	y_2: Physics	y_3: Latin	S_n
x_1: Ana	9	14	18	$s_1 = 41$
x_2: Juan	17	7	11	$s_2 = 35$
x_3: Mary	15	13	14	$s_3 = 42$
x_4: Ken	5	18	8	$s_4 = 31$
T_p	$t_1 = 46$	$t_2 = 52$	$t_3 = 51$	N = 149

C_{np}	y_2: Physics	y_3: Latin	y_1: English	$\mathbf{X_n}$
x_4: Ken	18	8	5	$x_4 = -.473$
x_1: Ana	14	18	9	$x_1 = -.094$
x_3: Mary	13	14	15	$x_3 = .108$
x_2: Juan	7	11	17	$x_2 = .400$
$\mathbf{Y_p}$	$y_2 = -.361$	$y_2 = .028$	$y_2 = .377$	

The objective of the method is to look for the case weight vector $\mathbf{X_n}$: (x_1, x_2, …, x_n) and the variable weight vector $\mathbf{Y_p}$: (y_1, y_2, …, y_p), which render the maximum correlation coefficient.

In order to find the unique solution of the two weight vectors $\mathbf{X_n}$ and $\mathbf{Y_p}$, the first condition is imposed on the two vectors: that the means (Mx, My) of $s_i * x_i$ (i = 1, 2, …, n) and $t_i * y_i$ (i = 1, 2, …, p) are 0:

[1a] Mx

$$= [(9x_1 + 14x_1 + 18x_1)$$
$$+ (17x_2 + 7x_2 + 11x_2)$$
$$+ (15x_3 + 13x_3 + 14x_3)$$
$$+ (5x_4 + 18x_4 + 8x_4)] / 149$$
$$= (41x_1 + 35x_2 + 42x_3 + 31x_4) / 149$$
$$= S_n^T \mathbf{X_n} / N = 0$$

[1b] My

$$= [(9y_1 + 17y_1 + 15y_1 + 5y_1)$$
$$+ (14y_2 + 7y_2 + 13y_2 + 18y_2)$$
$$+ (18y_3 + 11y_3 + 14y_3 + 8y_3)] / 149$$
$$= (46y_1 + 52y_2 + 51y_3) / 149$$
$$= T_p^T \mathbf{Y_p} / N = 0$$

where S_n is a vector of horizontal sums (41, 35, 42, 31), T_p is a vector of vertical sums (46, 52, 51), N is a scalar of Total sum (149).

The second condition is that the variances (Vx, Vy) of $s_i * x_i$ (i = 1, 2, …, n) and $t_i * y_i$ (i=1, 2, …, p) are 1:

$$[2a]\ Vx = [41(x_1 - Mx)^2 + 35(x_2 - Mx)^2 + 42(x_3 - Mx)^2 + 31(x_4 - Mx)^2 / 149$$
$$= [41x_1^2 + 35x_2^2 + 42x_3^2 + 31x_4^2] / 149 \quad \leftarrow [1a]\ Mx = 0$$
$$= \mathbf{X_n}^T S_{nn} \mathbf{X_n} / N = 1$$

$$[2b]\ Vy = [46(y_1 - My)^2 + 52(y_2 - My)^2 + 51(y_3 - My)^2] / 149$$
$$= (46y_1^2 + 52y_2^2 + 51y_3^2) / 149 \quad \leftarrow [1b]\ My = 0$$
$$= \mathbf{Y_p}^T T_{pp} \mathbf{Y_p} / N = 1$$

where

$S_{nn} = dg(S_n)$; $T_{pp} = dg(T_p)$ [dg: diagonal matrix]

S_{nn}	x_1	x_2	x_3	x_4
x_1	41			
x_2		35		
x_3			42	
x_4				31

T_{pp}	y_1	y_2	y_3
y_1	46		
y_2		52	
y_3			51

Considering the matrix D_{np} as a distribution graph, the correlation (R) between the X axis and the Y axis is:

$$[3]\quad R = [9(x_1-Mx)(y_1-My)$$
$$+ 14(x_1-Mx)(y_2-My)$$
$$+ 18(x_1-Mx)(y_3-My)$$
$$+ 17(x_2-Mx)(y_1-My)$$
$$+ \ldots$$
$$+ 8(x_4-Mx)(y_3-My)] / [(Vx*Vy]^{1/2} * 149]$$

$$= (9x_1 y_1 + 14x_1 y_2 + \ldots + 8x_4 y_3) / [(Vx*Vy]^{1/2} * 149]$$
$$\leftarrow [1a, 1b]\ Mx = My = 0$$
$$= (9x_1 y_1 + 14x_1 y_2 + \ldots + 8x_4 y_3) / 149 \quad \leftarrow [2a, 2b]\ Vx = Vy = 1$$
$$= \mathbf{X_n}^T D_{np} \mathbf{Y_p} / N$$

The objective is to find the two weight vectors $\mathbf{X_n}$ and $\mathbf{Y_p}$ when R is maximized. To maximize R, we formulate Q with the two constraints of Variances Vx = 1 and

Vy = 1, with two Lagrange multipliers[2]: Lx and Ly. We perform the differential calculation of Q with respect to $\mathbf{X_n}$ and $\mathbf{Y_p}$, whose results are zero vectors O_n and O_p:

$$Q = R - Lx\,(Vx - 1) - Ly\,(Vy - 1)$$
$$= (\mathbf{X_n}^T D_{np}\, \mathbf{Y_p}) / N - Lx\,[(\mathbf{X_n}^T S_{nn}\, \mathbf{X_n}) / N - 1] - Ly\,[(\mathbf{Y_p}^T T_{pp}\, \mathbf{Y_p}) / N - 1]$$

[4a] $Df(Q, \mathbf{X_n}) = D_{np}\, \mathbf{Y_p} / N - 2\, Lx\, S_{nn}\, \mathbf{X_n} / N = O_n$
←$Df(Q, \mathbf{X_n})$: Differentiate Q with respect to $\mathbf{X_n}$

[4b] $Df(Q, \mathbf{Y_p}) = D_{np}^T\, \mathbf{X_n} / N - 2\, Ly\, T_{pp}\, \mathbf{Y_p} / N = O_p$
←$Df(Q, \mathbf{Yp})$: Differentiate Q with respect to $\mathbf{Yp}$

[5a] $D_{np}\, \mathbf{Y_p} / N = 2\, Lx\, S_{nn}\, \mathbf{X_n} / N$ ←[4a]
$X_n^T D_{np}\, \mathbf{Y_p} / N = 2\, Lx\, X_n^T S_{nn}\, \mathbf{X_n} / N$ ←Multiply both sides by X_n^T
$R = 2\, Lx$ ←[3] $R = X_n^T D_{np}\, \mathbf{Y_p} / N$; [2a] $X_n^T S_{nn}\, \mathbf{X_n} / N = 1$

[5b] $D_{np}^T\, \mathbf{X_n} / N = 2\, Ly\, T_{pp}\, \mathbf{Y_p} / N$ ←[4b]
$\mathbf{X_n}^T D_{np} / N = 2\, Ly\, \mathbf{Y_p}^T T_{pp} / N$ ←Move $\mathbf{Yp}$; T_{pp}: diagonal matrix
$\mathbf{X_n}^T D_{np}\, \mathbf{Y_p} / N = 2\, Ly\, \mathbf{Y_p}^T T_{pp}\, \mathbf{Y_p} / N$ ←Multiply both sides by $\mathbf{Yp}$
$R = 2\, Ly$ ←[3] $R = \mathbf{X_n}^T D_{np}\, \mathbf{Y_p} / N$; [2b] $\mathbf{Y_p}^T T_{pp}\, \mathbf{Y_p} / N = 1$

By [5a] and [5b],

[6] $R = 2\, Lx = 2\, Ly$

[7a] $D_{np}\, \mathbf{Y_p} = 2\, Lx\, S_{nn}\, \mathbf{X_n}$ ←[5a] Multiply both sides by N
$D_{np}\, \mathbf{Y_p} = R\, S_{nn}\, \mathbf{X_n}$ ←[6] 2 Lx = R
$R\, S_{nn}\, \mathbf{X_n} = D_{np}\, \mathbf{Y_p}$ ←Change of sides
$S_{nn}\, \mathbf{X_n} = D_{np}\, \mathbf{Y_p} / R$ ←Move scalar R
$S_{nn}^{-1} S_{nn}\, \mathbf{X_n} = S_{nn}^{-1} D_{np}\, \mathbf{Y_p} / R$ ←Multiply S_{nn}^{-1} in both sides
$\mathbf{X_n} = S_{nn}^{-1} D_{np}\, \mathbf{Y_p} / R$ ←$S_{nn}^{-1} S_{nn} = I_{nn}$ (Identity matrix)

[7b] $D_{np}^T\, \mathbf{X_n} = 2\, Ly\, T_{pp}\, \mathbf{Y_p}$ ←[5b] Multiply both sides by N
$D_{np}^T\, \mathbf{X_n} = R\, T_{pp}\, \mathbf{Y_p}$ ←[6] R= 2 Ly

[2] Then:

$Df(Q, Lx) = (X_n^T S_{nn}\, X_n) / N - 1 = 0$ ← [2a]
$Df(Q, Ly) = [(Y_p^T T_{pp}\, Y_p) / N - 1 = 0$ ← [2b]

Introducing [7a] to $\mathbf{X_n}$ of [7b],

[8] $D_{np}{}^{T} S_{nn}{}^{-1} D_{np} \mathbf{Y_p} / R = R\ T_{pp} \mathbf{Y_p}$

$D_{np}{}^{T} S_{nn}{}^{-1} D_{np} \mathbf{Y_p} = R^2 T_{pp} \mathbf{Y_p}$ ←Move scalar R

$D_{np}{}^{T} S_{nn}{}^{-1} D_{np} (T_{pp}{}^{1/2})^{-1} T_{pp}{}^{1/2} \mathbf{Y_p} = R^2 (T_{pp})^{1/2} (T_{pp})^{1/2} \mathbf{Y_p}$

←$(T_{pp}{}^{1/2})^{-1} T_{pp}{}^{1/2} = I_{pp}$; $(T_{pp})^{1/2} (T_{pp})^{1/2} = T_{pp}$

(The details will be described later.)

Abbreviating $(T_{pp})^{1/2} \mathbf{Y_p}$ by A_p,

[9] $(T_{pp})^{1/2} \mathbf{Y_p} = A_p$

[8] will be:

$D_{np}{}^{T} S_{nn}{}^{-1} D_{np} (T_{pp}{}^{1/2})^{-1} A_p = T_{pp}{}^{1/2} R^2 A_p$

$(T_{pp}{}^{1/2})^{-1} D_{np}{}^{T} S_{nn}{}^{-1} D_{np} (T_{pp}{}^{1/2})^{-1} A_p = (T_{pp}{}^{1/2})^{-1} T_{pp}{}^{1/2} R^2 A_p$

←Multiple both sides by $(T_{pp}{}^{1/2})^{-1}$

$(T_{pp}{}^{1/2})^{-1} D_{np}{}^{T} S_{nn}{}^{-1} D_{np} (T_{pp}{}^{1/2})^{-1} A_p = R^2 A_p$ ←$(T_{pp}{}^{1/2})^{-1} T_{pp}{}^{1/2} = I_{pp}$

Abbreviating by $(T_{pp}{}^{1/2})^{-1} D_{np}{}^{T} S_{nn}{}^{-1} D_{np} (T_{pp}{}^{1/2})^{-1} = W_{pp}$

$W_{pp} A_p = R^2 A_p$

In this way we arrive at the formula of eigen value (R^2) and eigen vector (A_p), which are calculated by the program at the same time. The attribute weight vector $\mathbf{Y_p}$ is obtained by [9]:

$(T_{pp})^{1/2} \mathbf{Y_p} = A_p$ ←[9]

$[T_{pp}{}^{1/2}]^{-1} (T_{pp})^{1/2} \mathbf{Y_p} = [T_{pp}{}^{1/2}]^{-1} A_p$ ←Multiply both sides by $[T_{pp}{}^{1/2}]^{-1}$

$\mathbf{Y_p} = [T_{pp}{}^{1/2}]^{-1} A_p$ ←$X_{pp}{}^{-1} X_{pp} = I_{pp}$ (Identity matrix)

The attribute weight vector $\mathbf{Y_p}$ is reduced quite small because the sum of products with S_n is 0 and for the variance is 1 by [1b] and [2b].

[1b] $My = T_p{}^{T} \mathbf{Y_p} / N = 0$

[2b] $Vy = Y_p{}^{T} T_{pp} \mathbf{Y_p} / N = 1$

According to Takahashi (2005: 127-129), in order to adopt the data dimension, we multiply $\mathbf{Y_p}$ by the square root of the total sum N. It is also advisable to multiply $\mathbf{Y_p}$ by the correlation coefficient (R), to reflect the magnitude of R:

[10] $\mathbf{Y_p}' = [T_{pp}^{1/2}]^{-1} A_p * N^{1/2} * R$

The case weight vector $\mathbf{X_n}$ is obtained by [7a]:

$$\mathbf{X_n} = S_{nn}^{-1} D_{np} \mathbf{Y_p} / R$$

*** Raising the matrix to 1/2 and -1/2**

$A_{pp}^{1/2}$ of the square matrix A_{pp} is defined as X_{pp} having a relation of

$$X_{pp}^2 = X_{pp} X_{pp} = A_{pp},\ X_{pp} = A_{pp}^{1/2}$$

And if A_{pp} has an inverse matrix A_{pp}^{-1}, Y_{pp} in relation:

$$Y_{pp} Y_{pp} = A_{pp}^{-1}$$

is A_{pp} raised to -1/2, $App^{-1/2}$:

$$Y_{pp}^2 = Y_{pp} Y_{pp} = A_{pp}^{-1},\ Y_{pp} = A_{pp}^{-1/2}$$

*** Inverse of the diagonal matrix**

When T_{pp} is a diagonal matrix, the elements of its inverse matrix T_{pp}^{-1} are inverse values of T_{pp}:

$$T_{pp}^{1/2} T_{pp}^{1/2} = T_{pp}$$

T_{pp}	1	2	3
1	A		
2		B	
3			C

T_{pp}^{-1}	1	2	3
1	1/A		
2		1/B	
3			1/C

When T_{pp} is a diagonal matrix, the elements of its matrix raised to 1/2 in $T_{pp}^{1/2}$ are square roots of T_{pp}:

$$(T_{pp}^{1/2})^{-1} T_{pp}^{1/2} = T_{pp}$$

T_{pp}	1	2	3
1	A		
2		B	
3			C

$T_{pp}^{1/2}$	1	2	3
1	$\sqrt{A}$		
2		$\sqrt{B}$	
3			$\sqrt{C}$

Therefore, $(T_{pp}^{1/2})^{-1}$ is:

$(T_{pp}^{1/2})^{-1}$	1	2	3
1	$1/\sqrt{A}$		
2		$1/\sqrt{B}$	
3			$1/\sqrt{C}$

*** Correspondence between cases and variables**

The lower left table (D_{np}) is the input data and the lower right table is the weight vector X_n, corresponding to the cases:

D_{np}	y_1: English	y_2: Physics	y_3: Latin
x_1: Ana	9	14	18
x_2: Juan	17	7	11
x_3: Mary	15	13	14
x_4: Ken	5	18	8

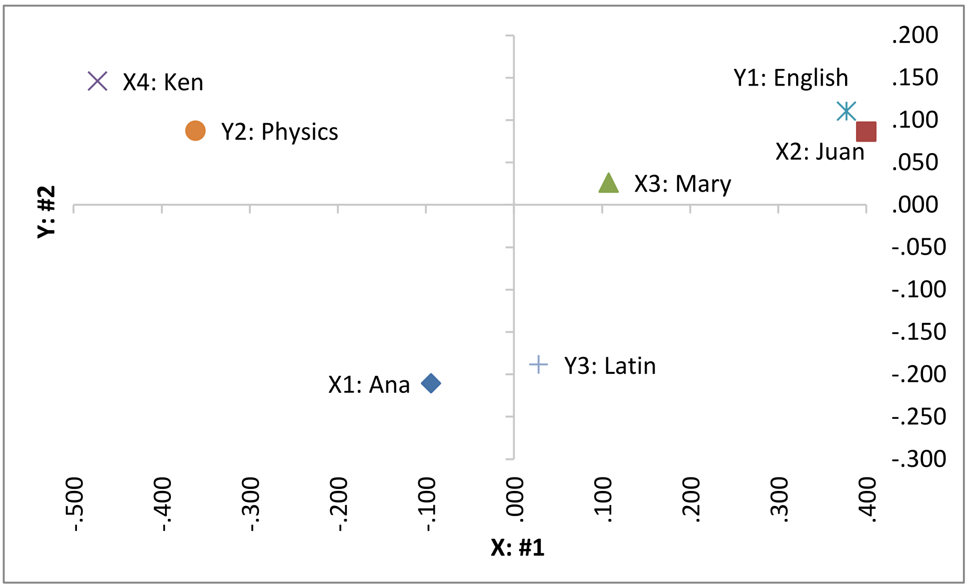
CA.d. (X_n)	#1	#2
x_1: Ana	-.094	-.211
x_2: Juan	.400	.086
x_3: Mary	.108	.026
x_4: Ken	-.473	.146

The lower left table shows the correlation and the right the vector of weight Y_p, corresponding to the variables (English, Physics, Latin):

Corresp.	1	2
Correl.	.300	.136

CA.v. (Yp)	#1	#2
y_1: English	.377	.110
y_2: Physics	-.361	.087
y_3: Latin	.028	-.189

The following graph shows the distribution of cases and variables located in the two vectors # 1 and # 2:



It is observed that there is a strong relationship between Ken and Physics, Juan and English, Ana and Latin. Mary is neutral with a certain proximity to English.

* Program (Excel VBA)

```
Sub BCA 'Bilateral Correspondence Analysis (Ueda 2002)
   Dim Xn, Yp, Snn, Tpp, App, Rp, Epp: ReDim Xn(n, 1), Yp(p, 1)
   Snn = d(uMt(n), smH(Dnp)) 'Row sum reverse diagonal matrix
   Tpp = d(uMt(p), e(smV(Dnp), 0.5)) 'Column sum root reverse diagonal matrix
   App = x(x(x(x(Tpp, t(Dnp)), Snn), Dnp), Tpp) 'Eigen value equation
   Rp = e(eigenV(App), 0.5) '√ Eigen value vector→Correlation coefficient
   Epp = eigenM(App) 'Eigen vector matrix
   Yp = extC(m(m(x(Tpp, Epp), Sqr(smA(Dnp))), Rp), 2) 'Yp=Tpp Ap*√N*Rp
   Xn = extC(d(x(x(Snn, Dnp), Yp), Rp), 2) 'Xn=Snn Dnp / Rp
 End Sub
```

In this program we use functions **d**: division, **UMt**: Unit matrix, **smH**: horizontal sum, **smV**: vertical sum, **e**: exponent, **x**: matrix product, **t**: transpose, **eigenV**: eigen value vector, **eigenM**; eigen vector matrix, **extC**: extract column.

2.2. Unilateral correspondence analysis

We have seen that by Bilateral Correspondence Analysis the unknown vectors of cases $\mathbf{X_n}$ as well as the vector of variables $\mathbf{Y_p}$ are obtained. In this place we propose to perform a Unilateral Correspondence Analysis where we determine *a priori* one of the two vectors and look for the other unknown vector. To the pre-established vector we provide standardized score of the consecutive numbers (1, 2, ..., n or p), for example if p = 3, they are (1-2)/(.816) = -1.225, (2-2)/(.816) = 0, (3-2)/(.816) = 1.225. By setting this vector as an external criterion, we look for the other unknown vector, which we also assume that the Mean of the weighted values is 0 and its Variance, 1. Starting from the data table D_{np}, our objective is to maximize the correlation coefficient in the concentrated table U_{np}:

D_{np}	y_1: English	y_2: Latin	y_3: Physics	S_n
x_1: Ana	9	14	18	$s_1 = 41$
x_2: Juan	17	7	11	$s_2 = 35$
x_3: Mary	15	13	14	$s_3 = 42$
x_4: Ken	5	18	8	$s_4 = 31$
T_p	$t_1 = 46$	$t_2 = 52$	$t_3 = 51$	N =149

U_{np}	y_1: English	y_2: Physics	y_3: Latin	$\mathbf{X_n}$
x_2: Juan	17	7	11	$\mathbf{x_2 = -.210}$
x_3: Mary	15	13	14	$\mathbf{x_3 = -.029}$
x_4: Ken	5	18	8	$\mathbf{x_4 = .119}$
x_1: Ana	9	14	18	$\mathbf{x_1 = .269}$
Y_p	$y_1 = -1.225$	$y_2 = .000$	$y_3 = 1.225$	

Let us first see the case in which we fix the vector Y_p of the variables and look for the unknown vector of case weight $\mathbf{X_n}$.

We impose the condition that the mean Mx of the product of $s_i * x_i$ (i=1, 2, …, n) is 0 and its variance Vx = 1:

[11] $\quad Mx = 0$ ←See [1]

[12] $\quad Vx = \mathbf{X_n}^T S_{nn} \mathbf{X_n} / N = 1$, where $S_{nn} = dg(S_n)$ [dg: diagonal matrix], see [2]

The correlation R of U_{np} is:

[13] $\quad R = \mathbf{X_n}^T D_{np} Y_p / N$ ←See [3]

In order to maximize R, we formulate Q with Lagrange multiplier L with the constraint of [12] $Vx = \mathbf{X_n}^T S_{nn} \mathbf{X_n} / N = 1$:

$$Q = R - L(Vx - 1)$$
$$Q = (\mathbf{X_n}^T D_{np} Y_p) / N - L (Vx- 1)$$
$$= (\mathbf{X_n}^T D_{np} Y_p) / N - L [(\mathbf{X_n}^T S_{nn} \mathbf{X_n}) / N - 1]$$

The derivative of Q with respect to $\mathbf{X_n}$ is O_n (zero vector):

[14] $Df(Q, \mathbf{X_n}) = D_{np} Y_p / N - 2 L S_{nn} \mathbf{X_n} / N = O_n$ (zero vector)

[15] $D_{np} Y_p / N = 2 L S_{nn} \mathbf{X_n} / N$ ←[14]
$\mathbf{X_n}^T D_{np} Y_p / N = 2 L \mathbf{X_n}^T S_{nn} \mathbf{X_n} / N$ ←Multiply both sides by $\mathbf{X_n}^T$
$R = 2 L \mathbf{X_n}^T S_{nn} \mathbf{X_n} / N$ ←[13] $R = \mathbf{X_n}^T D_{np} Y_p / N$;
$R = 2 L$ ←[12] $\mathbf{X_n}^T S_{nn} \mathbf{X_n} / N = Vx = 1$

[16] $D_{np} Y_p = R S_{nn} \mathbf{X_n}$ ←[15] $D_{np} Y_p / N = 2 L S_{nn} \mathbf{X_n} / N$; $R = 2 L$
$R S_{nn} \mathbf{X_n} = D_{np} Y_p$ ←Swap both sides
$S_{nn} \mathbf{X_n} = D_{np} Y_p / R$ ←Move scalar R
$S_{nn}^{-1} S_{nn} \mathbf{X_n} = S_{nn}^{-1} D_{np} Y_p / R$ ←Multiply both sides by S_{nn}^{-1}
$\mathbf{X_n} = S_{nn}^{-1} D_{np} Y_p / R$ ←$S_{nn}^{-1} S_{nn} = I_{nn}$ (Identity matrix)

In the same way as in the Bilateral Correspondence Analysis [10], we multiply $\mathbf{X_n}$ by R (correlation coefficient):

[17] $\mathbf{X_n}' = \mathbf{X_n} * R = S_{nn}^{-1} D_{np} Y_p / R * R = S_{nn}^{-1} D_{np} Y_p$

On the other hand, in order to find the vector of variables $\mathbf{Y_p}$, by fixing the vector of cases (X_n), we deviate the processes from [12] in:

[22] $Vy = \mathbf{Y_p}^T T_{pp} \mathbf{Y_p} / N = 1$

[23] $R = X_n^T D_{np} \mathbf{Y_p} / N$
$Q = (X_n^T D_{np} \mathbf{Y_p}) / N - L [Vy - 1]$
$= (X_n^T D_{np} \mathbf{Y_p}) / N - L [(\mathbf{Y_p}^T T_{pp} \mathbf{Y_p}) / N - 1]$

The derivative of Q with respect to $\mathbf{Y_p}$ is O_p (zero vector):

[24] $Df(Q, \mathbf{Y_p}) = D_{np}^T X_n / N - 2 L T_{pp} \mathbf{Y_p} / N = O_p$ (zero vector)

[25] $D_{np}^T X_n / N = 2 L T_{pp} \mathbf{Y_p} / N$ ←[24]

$X_n^T D_{np} / N = 2 L \mathbf{Y_p}^T T_{pp} / N$ ←$A^T B = B^T A$; T_{pp}: Diagonal matrix

$X_n^T D_{np} \mathbf{Y_p} / N = 2 L \mathbf{Y_p}^T T_{pp} \mathbf{Y_p} / N$ ←Multiply both sides by Y_p

$R = 2 L$ ←[3] $R = X_n^T D_{np} \mathbf{Y_p} / N$; [2b] $\mathbf{Y_p}^T T_{pp} \mathbf{Y_p} / N = 1$

[26] $D_{np}^T X_n = R T_{pp} \mathbf{Y_p}$ ←[25] $D_{np}^T X_n / N = 2 L T_{pp} \mathbf{Y_p} / N$; $R= 2 L$

$R T_{pp} \mathbf{Y_p} = D_{np}^T X_n$ ←Swap both sides

$T_{pp} \mathbf{Y_p} = D_{np}^T X_n / R$ ←Move scalar R

$T_{pp}^{-1} T_{pp} \mathbf{Y_p} = T_{pp}^{-1} D_{np}^T X_n / R$ ←Multiply both sides by T_{pp}^{-1}

$\mathbf{Y_p} = T_{pp}^{-1} D_{np}^T X_n / R$ ←$T_{pp}^{-1} T_{pp} = I_{pp}$ (Identity matrix)

The same as the previous case [17], we multiply $\mathbf{Y_p}$ by R:

[27] $\mathbf{Y_p}' = \mathbf{Y_p} * R = T_{pp}^{-1} D_{np}^T X_n / R * R = T_{pp}^{-1} D_{np}^T X_n$

* Program (Excel VBA)

```
Sub UCA(Dnp, sel)    'Unilateral Correspondence Analysis (Ueda 2002)
   Dim Xn, Yp, Snn, Tpp: ReDim Xn(n, 1), Yp(p, 1)
   Snn = d(uMt(n), smH(Dnp)) 'Row sum reverse diagonal matrix
   Tpp = d(uMt(p), smV(Dnp)) 'Column sum root reverse diagonal matrix
   If sel = 1 Then 'Row concentration
      Yp = stdV(seqC(Yp, 1)) 'Standarized sequential number
      Xn = x(x(Snn, Dnp), Yp) 'Weight calculation
   End If
   If sel = 2 Then 'Column concentration
      Xn = stdV(seqC(Xn, 1)) 'Standarized sequential number
      Yp = x(x(Tpp, t(Dnp)), Xn) 'Weight calculation
   End If
End Sub
```

In this program we use functions **d**: division, **UMt**: Unit matrix, **smH**: horizontal sum, **smV**: vertical sum, **stdV**: standardize, **seqC**: consecutive number, **x**: matrix product, **t**: transpose.

3. Application

3.1. Spanish preposition «de» in «sospechar de que»

In Takagaki et al. (2004-2014), we conducted surveys in the nine cities of Spain on different Spanish syntactic issues. One of them is a prominent use of the preposition «de» (English 'of') behind the verb 'sospechar' (English 'suspect'). The context is: *Sospecho de que me mintió*. (English: 'I suspect he lied to me.'). In the following I present tables of Absolute Frequency, Vertical Relative Frequency and a concentrated table presented by the Unilateral Correspondence Analysis (Columns: changed / Rows: fixed), where we can appreciate the admission gradation in Lp (the most generous), passing through Pa, Ma, Al, Se, Te, Ov, Hu and finally reaches Ba (the most rigid).

Absolute Frequency

AF	1.Ov	2.Pa	3.Al	4.Ma	5.Ba	6.Se	7.Hu	8.Te	9.Lp	Total
1: I say	0	2	1	2	0	1	2	0	4	12
2: I hear	20	16	7	15	14	17	9	15	19	132
3: No	7	3	3	4	6	6	7	5	1	42
Total	27	21	11	21	20	24	18	20	24	186

Vertical Relative Frequency

VRF	1.Ov	2.Pa	3.Al	4.Ma	5.Ba	6.Se	7.Hu	8.Te	9.Lp
1: I say		.10	.09	.10		.04	.11		.17
2: I hear	.74	.76	.64	.71	.70	.71	.50	.75	.79
3: No	.26	.14	.27	.19	.30	.25	.39	.25	.04

Unilateral Correspondence Analysis (rows: fixed / columns: changed)

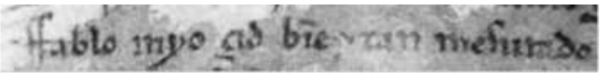
UCA: Column	9.Lp	2.Pa	4.Ma	3.Al	6.Se	8.Te	1.Ov	7.Hu	5.Ba
1: I say	.17	.10	.10	.09	.04			.11	
2: I hear	.79	.76	.71	.64	.71	.75	.74	.50	.70
3: No	.04	.14	.19	.27	.25	.25	.26	.39	.30

Abbreviation: 1.Ov:Oviedo, 2.Pa:Pamplona, 3.Al:Alcalá, 4.Ma:Madrid, 5.Ba:Barcelona, 6.Se:Sevilla, 7.Hu: Huelva, 8.Te: Tenerife, 9.Lp: Las Palmas de Gran Canaria.

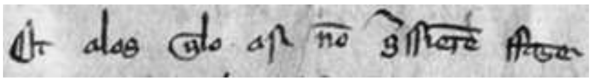
* Ueda (2017)

3.2. Relative chronology of Spanish letters

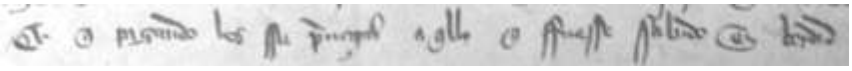
The following photocopies show different types of letters used in literary works and notarial documents:



(a) Cid, 1207, Letra gótica libraria {7} *ffablo myo çid bien e tan meſurado*



(b) CODEA:0287, Madrid, 1340, Letra de albalaes
{14} *Et a los que lo aſi non quiſſieren ffazer*



(c) CODEA:3931, Madrid 1386, Letra gótica cortesama
{31} *E que pagando los ſſu prinçipal aquello que ffueſſe ſabido en verdad*

* Data origin:
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.
http://www.corpuscodea.es/
LEMI «Letras Españolas en Manuscritos e Impresos», H. Ueda.
http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/index.html

Other types of letters are added up to 14. The following table shows the chronology of letter types presented in alphabetical order:

Letra	Carolina	Cort	De al.	De priv	G. c.	G. c. al	G. c. p.	G. lib	G. r.	Gótica	H. c	H. r.	Prec	Proc.
1075	1													
1100	2													
1125	1													
1150	7							1		1				
1175	2							5	1					
1200	3							14	5	10				
1225	3			1				48	5	12				
1250	1		2	4	14			45	4	8				
1275			14	5	32	22		13	3	1				
1300			4		46	1		8						
1325			1		29			3		1				
1350			1		25			3					5	
1375		6			7		3						6	
1400		4			2			1		1		1	20	
1425		12							1	1			3	
1450		30							2					1
1475		9							3	1		1		
1500		20							4			1	3	
1525		8							1		1	1		2
1550		1									3	11		
1575		1									2	3		2
1600											4	3		
1625											8			1
1650											4			
1675											7			

Fixing the chronology (years) on the vertical axis, we performed the Unilateral Correspondence Analysis to obtain the diagonalized distribution. The result of the new order of letter is as follows:

Letra	(1) Carolina	(2) G. lib	(3) De priv	(4) Gótica	(5) G. c. al	(6) De al.	(7) G. c.	(8) G. r.	(9) G. c. p.	(10) Prec	(11) Cort	(12) H. r.	(13) Proc.	(14) H. c
1075	1													
1100	2													
1125	1													
1150	7	1		1										
1175	2	5						1						
1200	3	14		10				5						
1225	3	48	1	12				5						
1250	1	45	4	8		**2**	**14**	4						
1275		13	5	1	**22**	**14**	**32**	3						
1300		8			**1**	**4**	**46**							
1325		3		1		**1**	**29**							
1350		3				**1**	**25**			5				
1375							**7**		3	6	6			
1400		1		1			**2**			20	4	1		
1425				1				1		3	12			
1450								2			30		1	
1475				1				3			9	1		
1500								4		3	20	1		
1525								1			8	1	2	1
1550											1	11		3
1575											1	3	2	2
1600												3		4
1625													1	8
1650														4
1675														7

Letters: (1) Carolina, (2) Gothic library, (3) Privileges, (4) Gothic, (5) Gothic cursive (albalaes), (6) Albalaes, (7) Gothic cursive, (9) Gothic cursive (precourtesan), (10) Precourtesan, (11) Courtesan, (12) Humanistic round, (13) Procedural, (14) Humanistic cursive

Let us look at the specific cases of the initial <ff>, <f>, <h> and Ø in a frequent verb, *fazer* 'to do, to make', a noun, *fijo*, *fija* ('son', 'daughter'), and a preposition of Arab origin, *fasta* ('until') with its multiple graphic variants in notarial documents:

Year	ffijo	ffazer	ffasta	fijo	fazer	fasta	hacer	hasta	hijo
1200				.280		.060			
1225				2.300	.990	.560			
1250	**.970**	**.380**	**.050**	2.070	2.550	.690			
1275	**.950**	**1.440**	**.660**	1.630	1.810	.470			.010
1300	**1.930**	**1.910**	**.640**	1.510	.950	.530			
1325	**1.770**	**2.720**	**.590**	.150					.070
1350	**.880**	**1.100**	**.440**	.220	1.430	.110			
1375	**.170**	**.290**	**.370**	1.240	2.030	1.700			
1400	.030	.030		1.220	3.080	1.260			
1425				.410	2.640	.490			
1450				.610	2.140	1.110			
1475				.580	1.590	1.080	.430	.040	.080
1500				.270	1.060	.350	.970	.310	.690
1525					.130	.130	2.020	1.470	.380
1550					.040	.070	1.180	.720	.250
1575						.050	1.260	.720	.870
1600						.070	1.470	1.550	.520
1625				.090			.570	.090	
1650							.120	.540	.120
1675							.880	.350	.880

Row	Xn
1200	-1.648
1225	-1.474
1250	-1.301
1275	-1.127
1300	-.954
1325	-.780
1350	-.607
1375	-.434
1400	-.260
1425	-.087
1450	.087
1475	.260
1500	.434
1525	.607
1550	.780
1575	.954
1600	1.127
1625	1.301
1650	1.474
1675	1.648

Column	ffijo	ffazer	ffasta	fijo	fazer	fasta	hacer	hasta	hijo
Yp	-.921	-.872	-.839	-.837	-.457	-.371	.888	.954	.962

***Fazer*, *fijo*, *fasta* in Castile (per thousand words)**

We are impressed by the concentration of the double graph in *ffijo*, *ffazer*, *ffasta* in years from 1250 to 1400, which coincide with the time of (5) Gothic cursive (albalaes), (6) Albalaes, (7) Gothic cursive, which are characterized by their cursiveness of writing.

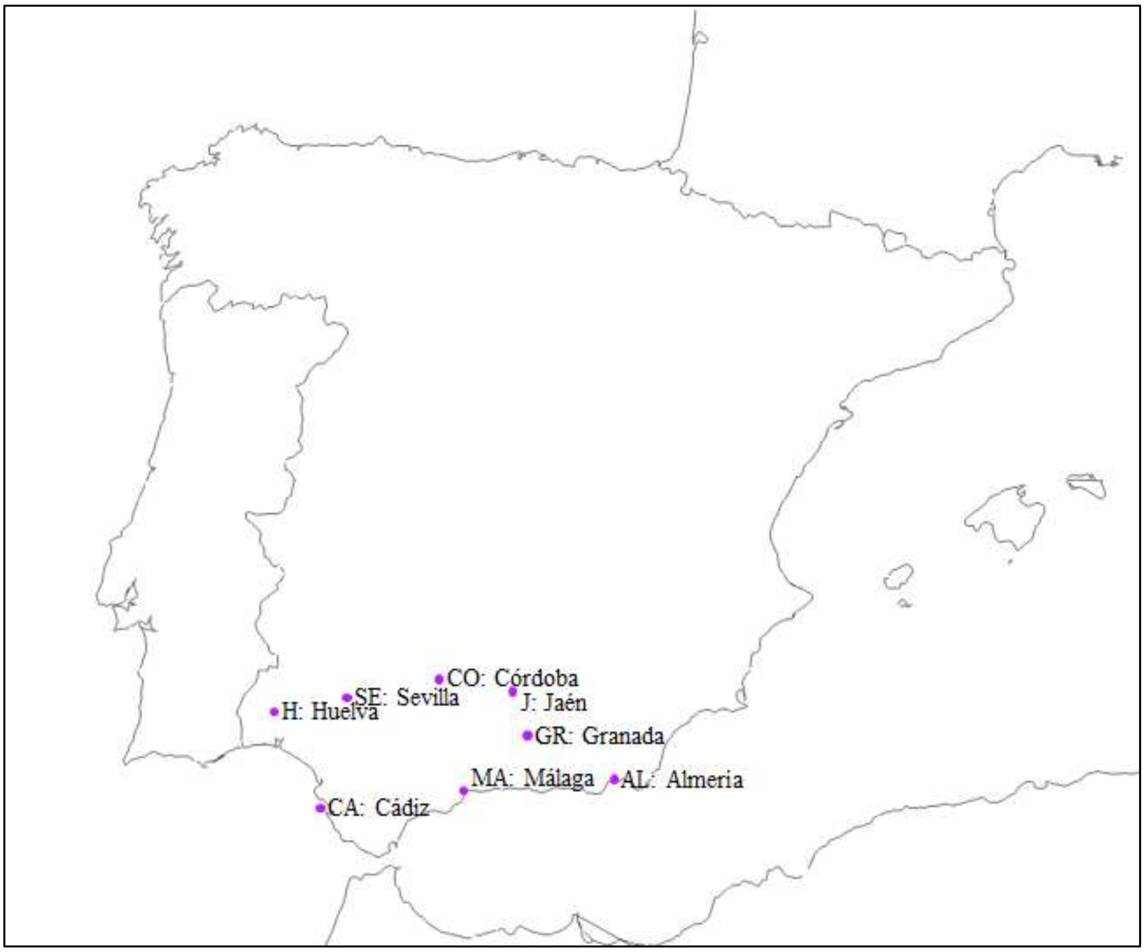
* Ueda (2016)

3.3. Andalusian phonetic features

The following table shows the relative frequencies of open and closed vowel cases in different provinces of Andalusia:

R / N * 100	H	SE	CA	MA	CO	J	GR	AL
1533B:miel:el>e+	17	10	9	15	20	29	46	30
1533C:miel:el>e:	11	6	4	16	12	11	16	3
1615A:caracol:-ól>ó+(:)	2	3	3	5	15	14	19	11
1615B:caracol:-ól>ó(:)	18	27	15	16	3	1	6	2
1616A:árbol:-ol>o+	2	1			6	8	6	6
1616B:árbol:-ol>o	23	30	17	26	18	11	23	11
1618A:sol:-ól>ó+(:)	7	9	3	13	13	12	19	11
1618B:sol:-ól>ó(:)	15	21	15	13	1	1	6	1
1623A:beber:-ér>é+l	2			1	10	11	19	20
1623B:beber:-ér>é+	4	7	3	6	13	17	15	8
1623C:beber:-ér>é	19	24	15	19	2		4	
1626A:tos:-ós>ó+h	6	2	2	4	7	13	17	9
1626C:tos:o++				2	7	10	18	12
1626C:tos:-ós>ó+	7	7	5	13	18	17	27	19
1626D:tos:-ós>ó	10	22	11	9	2	1	2	
1627A:nuez:-éθ>é+h	5	2	1		2	3	8	3
1627B:nuez:-éθ>é+	7	13	5	17	20	25	39	26
1627C:nuez:e++				5	14	18	26	18
1627C:nuez:-éθ>é	12	16	12	9	3	1	1	
1629A:voz:-óθ>óh	5	3		1	1	1	5	3
1629B:voz:-óθ>ó+	3	5	3	12	22	30	44	30
1629C:voz:-óθ>ó	18	23	14	13	2	1	2	1
1689A:niños:-os>-o+	1	2		4	22	31	44	30
1689B:niños:-os>oh[os)	4	1			2	3	8	8
1690A:pared:-éd>é+	6	8		10	17	19	24	11
1693A:redes:redes>rede	3	2	1	1	4	12	8	18
1693B:redes:redes>re+	14	6	4	12	3	6	16	6
1693C:redes:redes>reh	1		2		1	4	7	
1694A:clavel:-él>-él	3	2	1	3	6	5	11	15
1694B:clavel:-él>é+,		6	3	15	20	24	40	29
1694C:clavel:-él>ér						1	5	1
1695A:claveles:e-es>-e-e+		2		4	2	2	4	3
1695B:claveles:e-es>-e+-e+		1		7	18	24	33	21
1695C:claveles:-e-es>-e-e:		1		3	1	2	1	1
1695D:claveles:e-es>-e-eh	3	1			5	4	9	5

We perform the Unilateral Correspondence Analysis, with a fixed external criterion: provinces ordered from West to East. Our goal is to find the vector of phonetic features by which we rearrange the rows as presented in the following table:



Cor.R	H	SE	CA	MA	CO	J	GR	AL
1629C:voz:-óθ>ó	18	23	14	13	2	1	2	1
1627C:nuez:-éθ>é	12	16	12	9	3	1	1	
1626D:tos:-ós>ó	10	22	11	9	2	1	2	
1623C:beber:-ér>é	19	24	15	19	2		4	
1618B:sol:-ól>ó(:)	15	21	15	13	1	1	6	1
1615B:caracol:-ól>ó(:)	18	27	15	16	3	1	6	2
1616B:árbol:-ol>o	23	30	17	26	18	11	23	11
1693B:redes:redes>re+	14	6	4	12	3	6	16	6
1629A:voz:-óθ>óh	5	3		1	1	1	5	3
1533C:miel:el>e:	11	6	4	16	12	11	16	3
1627A:nuez:-éθ>é+h	5	2	1		2	3	8	3
1618A:sol:-ól>ó+(:)	7	9	3	13	13	12	19	11
1695C:claveles:-e-es>-e-e:		1		3	1	2	1	1
1623B:beber:-ér>é+	4	7	3	6	13	17	15	8
1690A:pared:-éd>é+	6	8		10	17	19	24	11
1533B:miel:el>e+	17	10	9	15	20	29	46	30
1626C:tos:-ós>ó+	7	7	5	13	18	17	27	19
1695A:claveles:e-es>-e-e+		2		4	2	2	4	3
1627B:nuez:-éθ>é+	7	13	5	17	20	25	39	26
1626A:tos:-ós>ó+h	6	2	2	4	7	13	17	9
1693C:redes:redes>reh	1		2		1	4	7	
1615A:caracol:-ól>ó+(:)	2	3	3	5	15	14	19	11
1695D:claveles:e-es>-e-eh	3	1			5	4	9	5
1689B:niños:-os>oh[os)	4	1			2	3	8	8
1616A:árbol:-ol>o+	2	1			6	8	6	6
1694A:clavel:-él>-él	3	2	1	3	6	5	11	15
1629B:voz:-óθ>ó+	3	5	3	12	22	30	44	30
1694B:clavel:-él>é+,		6	3	15	20	24	40	29
1693A:redes:redes>rede	3	2	1	1	4	12	8	18
1695B:claveles:e-es>-e+-e+		1		7	18	24	33	21
1689A:niños:-os>-o+	1	2		4	22	31	44	30
1627C:nuez:e++				5	14	18	26	18
1623A:beber:-ér>é+l	2			1	10	11	19	20
1626C:tos:o++				2	7	10	18	12
1694C:clavel:-él>ér						1	5	1

The high frequencies in the upper left section show that in the provinces of western Andalusia (H: Huelva, SE: Seville, CA: Cádiz, MA: Malaga), cases of no vowel opening are detected, while the lower right section shows high frequencies corresponding to cases of the open vowel in the eastern provinces: CO: Córdoba, J: Jaen, GR: Granada. AL: Almeria. This is a general trend and, of course, numerous cases are found in the remaining sections. However, they are relatively few compared to the mentioned sections.

* Data origin: Manuel Alvar y Antonio Llorente (1973) / Ueda (1993)

3.4. Spanish words of 'farmer' in Latin America.

The following figure shows the result of the Bilateral Corresponding Analysis of the words of the concept of 'farmer' in Latin America (Cahuzac 1980):

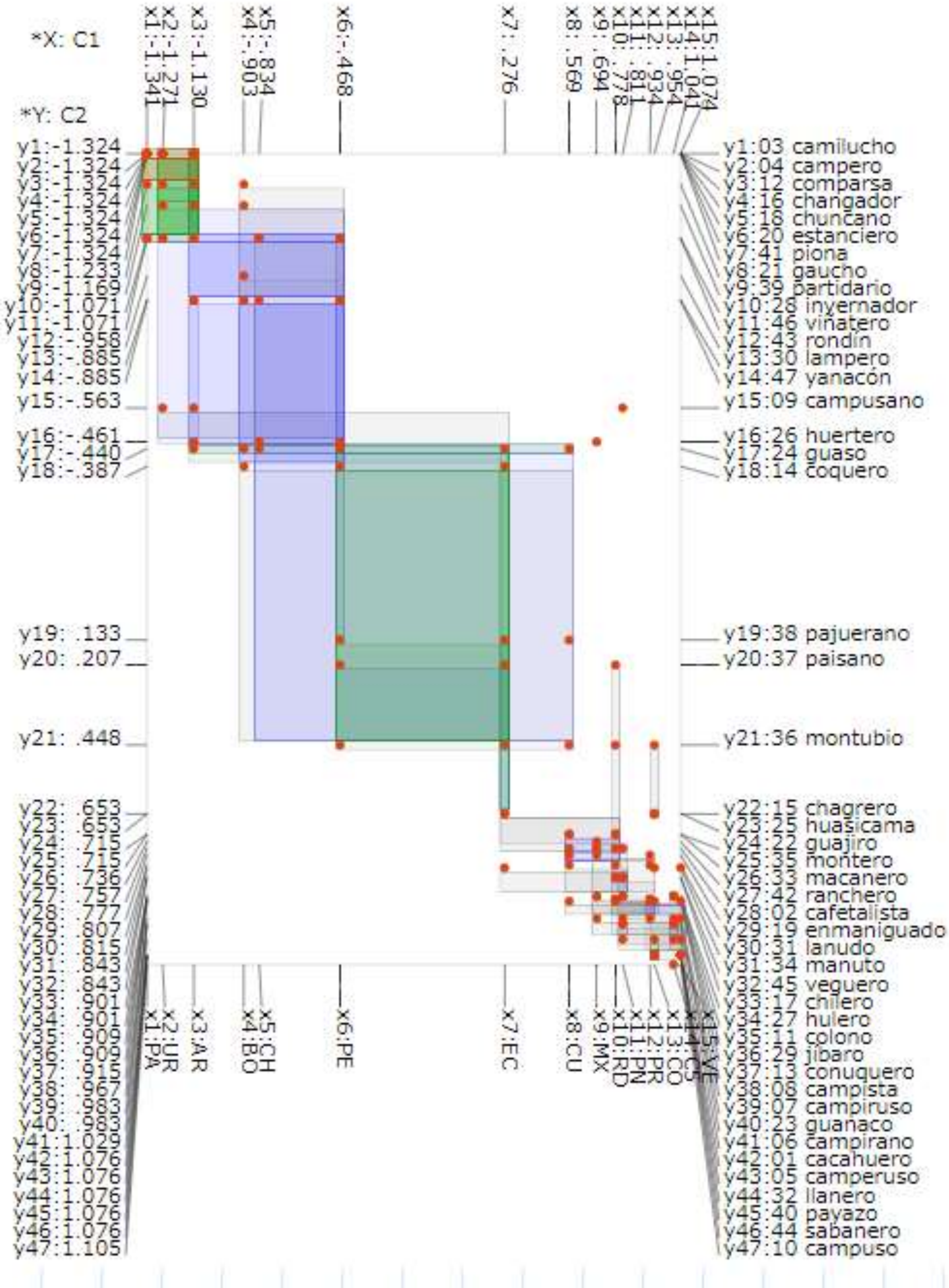
F.A.	PA	UR	AR	BO	CH	PE	EC	CU	MX	RD	PN	PR	CO	C5	VE	Val.
03 camilucho	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1.324
04 campero	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1.324
12 comparsa	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1.324
16 changador	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1.324
18 chuncano	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1.324
20 estanciero	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1.324
41 piona	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1.324
21 gaucho	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	-1.233
39 partidario	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	-1.169
28 invernador	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	-1.071
46 viñatero	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	-1.071
43 rondín	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	-.958
30 lampero	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	-.885
47 yanacón	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	-.885
09 campusano	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	-.563
26 huertero	0	0	1	0	1	1	0	0	1	0	0	0	0	0	0	-.461
24 guaso	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	-.440
14 coquero	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	-.387
38 pajuerano	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	.133
37 paisano	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	.207
36 montubio	0	0	0	0	0	1	1	1	0	1	0	0	1	0	0	.448
15 chagrero	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	.653
25 huasicama	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	.653
22 guajiro	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	.715
35 montero	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	.715
33 macanero	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	.736
42 ranchero	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	.757
02 cafetalista	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	.777
19 enmaniguado	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	.807
31 lanudo	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	.815
34 manuto	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	.843
45 veguero	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	.843
17 chilero	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	.901
27 hulero	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	.901
11 colono	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	.909
29 jíbaro	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	.909
13 conuquero	0	0	0	0	0	0	0	1	0	1	0	1	1	0	1	.915
08 campista	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	.967
07 campiruso	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	.983
23 guanaco	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	.983
06 campirano	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1.029
01 cacahuero	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1.076
05 camperuso	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1.076
32 llanero	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1.076
40 payazo	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1.076
44 sabanero	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1.076
10 campuso	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1.105
Val.	-1.341	-1.271	-1.130	-.903	-.834	-.468	.276	.569	.694	.778	.811	.934	.954	1.041	1.074	

In this figure we observe that the words 16, 18, ..., 21 are concentrated in PA (Paraguay), UR (Uruguay), AR (Argentina). We call the "concentrated area" to the area where positive reactions are gathered. In the Concentrated Distribution table, we find several Concentrated Areas around the diagonal line. By the program we look for, from a certain point, the successive lower points of the right side to find the point that offers the greatest result of the expectative probability (ProbEx) and order the table in descending way:

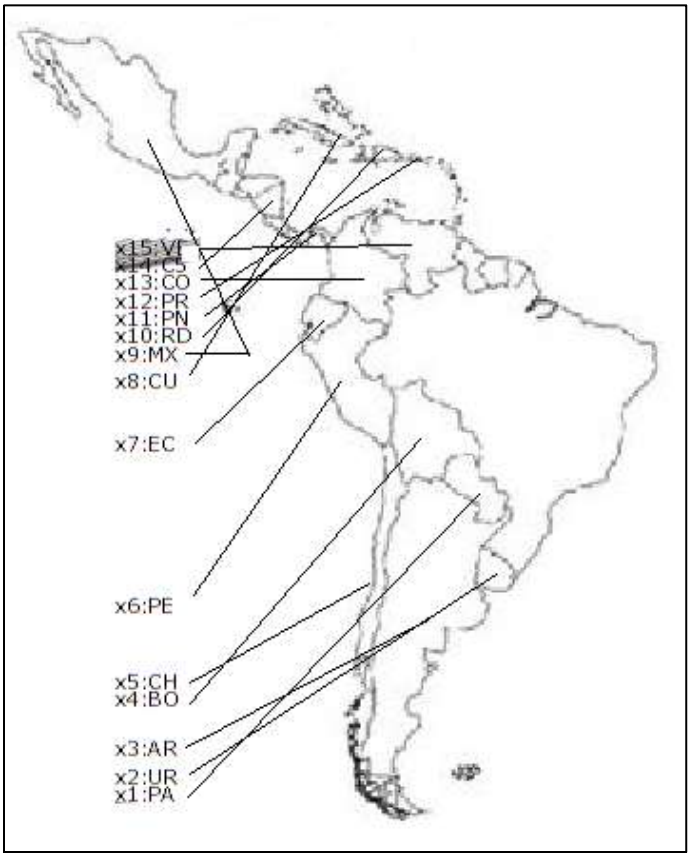
C1	A1	C2	A2	X1	Y1	X2	Y2	Suma	Total	Media	ProEx
03 camilucho	PA	20 estanciero	AR	1	1	3	6	18	18	1.000	.847
04 campero	PA	41 piona	AR	1	2	3	7	18	18	1.000	.847
12 comparsa	PA	21 gaucho	AR	1	3	3	8	18	18	1.000	.847
16 changador	PA	21 gaucho	AR	1	4	3	8	15	15	1.000	.819
03 camilucho	UR	20 estanciero	AR	2	1	3	6	12	12	1.000	.779
04 campero	UR	41 piona	AR	2	2	3	7	12	12	1.000	.779
12 comparsa	UR	21 gaucho	AR	2	3	3	8	12	12	1.000	.779
16 changador	UR	39 partidario	AR	2	4	3	9	12	12	1.000	.779
18 chuncano	PA	21 gaucho	AR	1	5	3	8	12	12	1.000	.779
18 chuncano	UR	28 invernador	AR	2	5	3	10	12	12	1.000	.779
20 estanciero	UR	46 viñatero	AR	2	6	3	11	12	12	1.000	.779
20 estanciero	PA	46 viñatero	AR	1	6	3	11	17	18	.944	.762
41 piona	UR	46 viñatero	AR	2	7	3	11	10	10	1.000	.741
24 guaso	PE	36 montubio	EC	6	17	7	21	10	10	1.000	.741
41 piona	PA	46 viñatero	AR	1	7	3	11	14	15	.933	.721
21 gaucho	UR	46 viñatero	AR	2	8	3	11	8	8	1.000	.688

For example, the case of 03 *camilucho* is placed in the coordinate (1, 1) in the concentrated distribution table and, therefore, the coordinate (1, 1) is the starting point. The program reaches the coordinate (3, 6) and the coordinate Zone enclosed between (1, 1) and (3, 6) gives the sum of reactions 18 and the total case count: 3 * 6 = 18. Now the mean is 18/18 = 1, which is the maximum value. The expectative probability (ProEx) of *camilucho* is .847.

The following chart shows the first 170 concentrated areas. Each area represents the relationships between cases and attributes.



* Cahuzac (1980) / Ueda (2012)



By means of the Bilateral Correspondence Analysis, we notice that the countries of the North (MX: Mexico, C5, 5 Central American countries and Caribbean countries, PR: Puerto Rico, RD: Dominican Republic, CU: Cuba), Andean countries (EC: Ecuador, PE: Peru, BO: Bolivia) and La Plata (AR: Argentina, UR: Uruguay, PA: Paraguay). The members of the northern group and those of La Plata are closely linked, while the members of the Andean group maintain a certain distance among them.

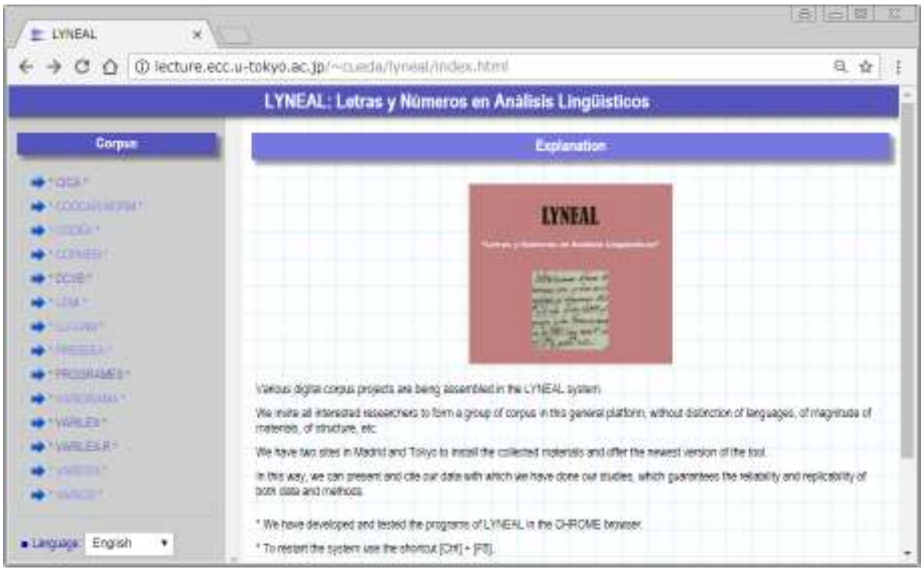
4. Conclusion

Both the method of Correspondence Analysis (Benzécri 1973) and the Quantification Theory Type III (Hayashi 1956) are useful in investigating the reactions obseved in two-dimensional space of cases and attributes. The merit of both methods is to facilitate a vision of distribution in a diagonally concentrated manner. In this way it is mandatory to analyze the cases and attributes at the same time changing their order of presentation according to the magnitude of weight.

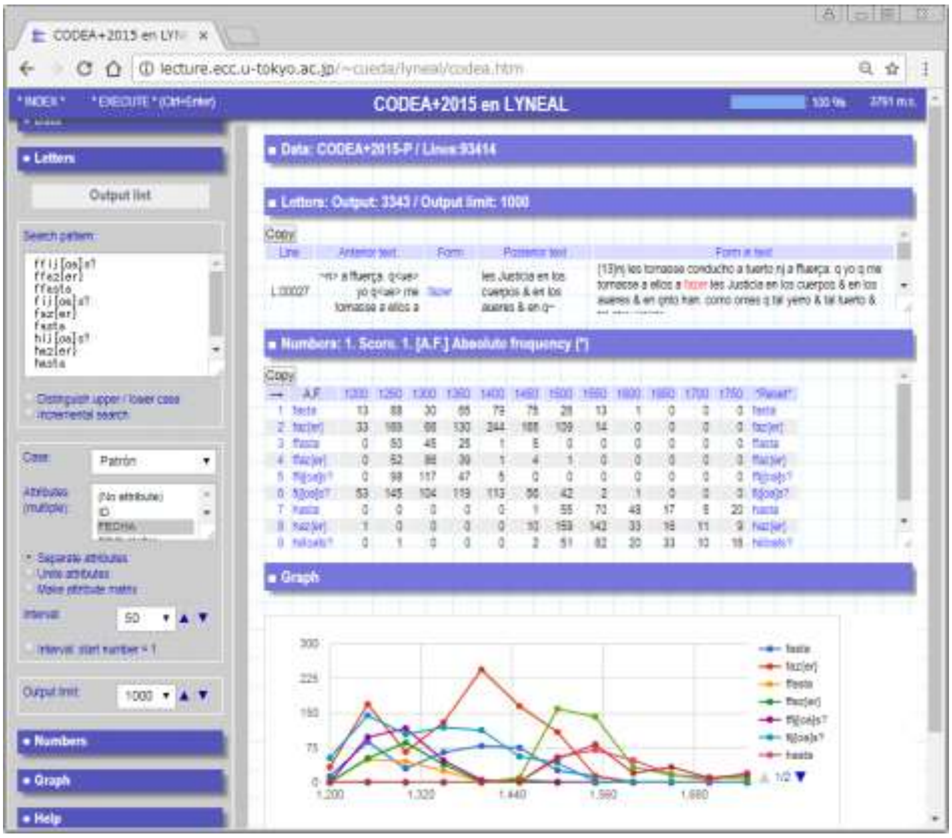
On the other hand, researchers are sometimes forced to analyze cases with pre-established attributes. For example, linguistic phenomena throughout the chronology of years, the order of time parameter should not be changed so as to observe the phenomena along the chronological axis. In the survey data the responses have a logical order (for example *never*, *sometimes*, *usually*, *always*). Even in the dialectological data it is interesting to perform the observation with ordered axes, from north to south, from west to east, from center to periphery, etc. For this purpose I had developed my own Concentration Analysis by Distance, which permits both unilateral and bilateral analyses (Ueda 1993, 2006).

By the method proposed here, Unilateral Correspondence Analysis, the researcher is capable of studying the cases in accordance with the preordained attributes, in the same way as my method of Unilateral Concentration Analysis by Distance, both of which. are installed in the following site:

LYNEAL (Letras y Números en Análisis Lingüísticos):



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



http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/codea.htm

Reference:

Akido, Hiroshi. 飽戸弘(1974)「社会心理学」林知己夫・池内一・森田優三・寺田和夫（編）『計量的研究－我が国人文･社会科学研究の最近の動向』東京：南窓社, pp.69-91.

Asano, Hirohiko 朝野熙彦(2000)『入門多変量解析の実際』東京：講談社サイエンティフィック

Alvar, Manuel / Llorente, Antonio. (1973) *Atlas lingüístico y etnográfico de Andalucía*, 6 tomos, Granada, Universidad de Granada-CSIC, 1961-1973

Benzécri, Jean-Paul. (1973) *L'analyse des données, T2, l'analyse des correspondances*, Dunod, Paris.

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

Clausen, Sten-Erik. (1998) *Applied Correspondence Analysis. An Introduction*. London. Sage Publications. 藤本一男訳(2015)『対応分析入門：理論から応用まで』東京：オーム社

Greenacre, Michael. (2013) *Theory and Applications of Correspondence Analysis* http://www.carme-n.org/?sec=books5 [2017/7/29]

Hayashi, Chikio. (1956) "Theory and example of quantification (II)", *Proceedings of the Institute of Statistical Mathematics*. 4 (2): 19-30 (in Japanese).

Inoue, Fumio. (1996) "Computational dialectology (1)", *Area and Culture Studies* (Tokyo University of Foreign Studies) 52, pp. 67-102.
http://dictionary.sanseido-publ.co.jp/affil/person/inoue_fumiódoc/E16.pdf

_____ (1997) "S-shaped curves of language standardization", in Alan R. Thomas (ed.) *Issues and Methods in Dialectology*. University of Wales Bangor, pp. 79-93..
http://dictionary.sanseido-publ.co.jp/affil/persoñinoue_fumiódoc/E23.pdf

_____ / Fukushima, Chitsuko. (1997) "A quantitative approach to English dialect distribution. Analyses of CLAE morphological data", in Viereck (ed.) *The Computer Developed Linguistic Atlas of England*, pp. 57-65.
http://dictionary.sanseido-publ.co.jp/affil/person/inoue_fumiódoc/E20.pdf

Kosugi, Koji. 小杉考司(2007)『社会調査士のための多変量解析法』京都：北大路書房

Mino, Tairai. 三野大來(2001)『統計解析のための線形代数』共立出版

Okumura, Haruhiko. 奥村晴彦(1986)『パソコンによるデータ解析入門．数理

とプログラミング実習』技術評論社

Takagaki, Toshihiro / Fukushima, Noritaka / Miyamoto, Masami / Ruiz Tinoco, Antonio / Ueda, Hiroto. (2004-2014) *VARIGRAMA, Variación Gramatical del Español*
http://lecture.ecc.u-tokyo.ac.jp/~cueda/varigrama/index.html

Takahashi, Shin. 高橋信(2005)『Excel で学ぶコレスポンデンス分析』オーム社

Ueda, Hiroto. (1993) "División dialectal de Andalucía. Análisis computacional", *Actas del Tercer Congreso de Hispanistas de Asia,* Asociación Asiática de Hispanistas, Tokio, pp. 407-419.

_____ (2002) *Program package NUMEROS*.
http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/4-numeros/numeros.xlsm

_____ (2006) "Patternized configuration of geolinguistic data: Application to Spanish lexical variation", *Topics in Dialectal Variation,* Markku Filppula, Juhani Klemola, Marjatta Palander, Esa Penttila (eds.) University of Joensuu, Faculty of Humanities, pp. 163-187

_____ (2012) *Análisis de datos cuantitativos para estudios lingüísticos,*
http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/4-numeros/doc/numeros-es.pdf

_____ (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.

_____ (2017) "Queísmo y dequeísmo observados en los datos de VARIGRAMA en el sistema LYNEAL. Aproximación a la realidad de variación sintáctica española a través de análisis multivariados pluridimensionales", *Spanish Dialect Syntax,* Universidad Complutense de Madrid, 21 de abril de 2017.

Yasuda, Saburou / Umino, Michirou 安田三郎・海野道郎(1977)『社会統計学』東京：丸善

[FIN]