THE DEGREE OF BINDING RESULTING FROM REACTION POINTS EXPRESSED IN A DIATOPIC TABLE.

AN APPLICATION TO A CATALAN VERB MORPHOLOGY DATABASE

Hiroto Ueda, University of Tokyo

Maria Pilar Perea, University of Barcelona

2012

0. Introduction

Antoni M. Alcover: 500,000 verb forms corresponding to the complete conjugation of 75 verbs in 149 localities in the Catalan-speaking territory.

“La flexió verbal en el dialectes catalans” (1929 – 1933).

1. Techniques of Integration

2. Assessing the degree of integration

3. Geolinguistic interpretation

4. Conclusion

1. Techniques of Integration

Form	1	2	3	4	5	6	7	8	9	10
cantar-1-cant										
cantar-1-canta										
cantar-1-cante										
cantar-1-canti	V	V	V	V	V	V	V	V	V	V
cantar-1-cantic										
cantar-1-canto										
cantar-1-cantoc										
cantar-1-cantot										
cantar-2-cantas										
cantar-2-cantes	V	V	V	V	V	V	V	V	V	V
cantar-3-canta	V	V	V	V	V	V	V	V	V	V
cantar-4-cantam				V			V			V
cantar-4-cantem	V	V	V		V	V		V	V	V
cantar-5-cantau				V			V			V
cantar-5-canteu	V	V	V		V	V		V	V	V
cantar-6-canten	V	V	V	V	V	V	V	V	V	V

Table 1.

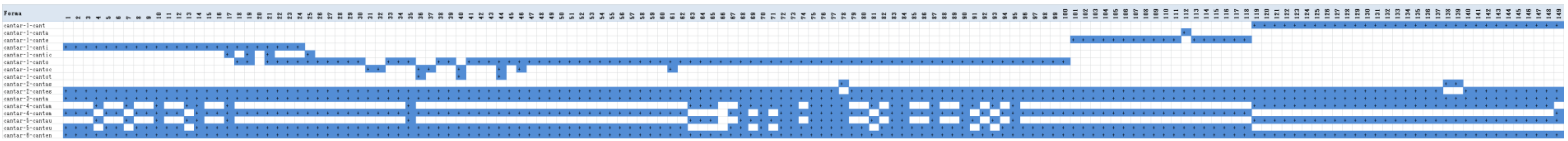


Figure 1. Initial data matrix

1.1 Average distance with respect to the zero point (ADZP)

Minkowski distance:

$$ADZP = [(\Sigma_i x_i h_i^N) / \Sigma_i x_i]^{1/N}$$

where x_i is 1 if there is reaction in form of “v”; or zero (0) if there is no reaction; h_i is the sequential number (1, 2, ..., p) counting from the point of origin (0).

D	L-1	L-2	L-3	L-4	Dist
d-1	v	v			1.581
d-2			v		3.000
d-3		v			2.000
d-4			v	v	3.536
d-5	v	v	v		2.160
Dist	3.606	3.416	3.873	4.000	

Table 2.

$d\text{-}1: [(1^2 + 2^2) / 2]^{1/2} = 1.581$

$d\text{-}2: [(3^2) / 1]^{1/2} = 3.000$

$d\text{-}3: [(2^2) / 1]^{1/2} = 2.000$

$d\text{-}4: [(3^2 + 4^2) / 2]^{1/2} = 3.535$

$d\text{-}5: [(1^2 + 2^2 + 3^2) / 3]^{1/2} = 2.160$

$L\text{-}1: [(1^2 + 5^2) / 2] = 3.606$

$L\text{-}2: [(1^2 + 3^2 + 5^2) / 3] = 3.416$

$L\text{-}3: [(2^2 + 4^2 + 5^2) / 3] = 3.873$

$L\text{-}4: [(4^2) / 1] = 4.000$

D	L-1	L-2	L-3	L-4
d-1	v	v		
d-2			v	
d-3		v		
d-4			v	v
d-5	v	v	v	

→

D	L-1	L-2	L-3	L-4
d-1	v	v		
d-3		v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Table 3.

Lv	L-1	L-2	L-3	L-4
d-1	v	v		
d-2			v	
d-3		v		
d-4			v	v
d-5	v	v	v	

→

LL	L-2	L-1	L-3	L-4
d-3	v			
d-1	v	v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Table 4.

$$\text{d-1: } [(1^2 + 2^2) / 2]^{1/2} = 1.581 \quad (\ldots 1)$$

$$\text{d-2: } [(3^2) / 1]^{1/2} = 3.000 \quad (\ldots 4)$$

$$\text{d-3: } [(2^2) / 1]^{1/2} = 2.000 \quad (\ldots 2)$$

$$\text{d-4: } [(3^2 + 4^2) / 2]^{1/2} = 3.535 \quad (...5)$$

$$\text{d-5: } [(1^2 + 2^2 + 3^2) / 3]^{1/2} = 2.160 \quad (...3)$$

Lv	L-1	L-2	L-3	L-4	Value
d-1	v	v			1.581
d-3		v			2.000
d-5	v	v	v		2.160
d-2			v		3.000
d-4			v	v	3.536

Table 5.

$$\text{L-1: } [(1^2 + 3^2) / 2]^{1/2} = 2.236 \quad (...2)$$

$$\text{L-2: } [(1^2 + 2^2 + 3^2) / 3]^{1/2} = 2.160 \quad (...1)$$

$$\text{L-3: } [(3^2 + 4^2 + 5^2) / 3]^{1/2} = 4.082 \quad (...3)$$

$$\text{L-4: } [(5^2) / 1]^{1/2} = 5.000 \quad (...4)$$

Lv	L-2	L-1	L-3	L-4	Value
d-1	v	v			1.581
d-3	v				1.000
d-5	v	v	v		2.160
d-2			v		3.000
d-4			v	v	3.536
Value	2.160	2.236	4.082	5.000	

Table 6.

d-1: $[(1^2 + 2^2) / 2]^{1/2}$ = 1.581 (...2)

d-3: $[(1^2) / 1]^{1/2}$ = 1.000 (...1)

d-5: $[(1^2 + 2^2 + 3^2) / 3]^{1/2}$ = 2.160 (...3)

d-2: $[(3^2) / 1]^{1/2}$ = 3.000 (...4)

d-4: $[(3^2 + 4^2) / 2]^{1/2}$ = 3.535 (...5)

Lv	L-2	L-1	L-3	L-4	Value
d-3	v				1.000
d-1	v	v			1.581
d-5	v	v	v		2.160
d-2			v		3.000
d-4			v	v	3.536
Value	2.160	2.550	4.082	5.000	

Table 7.

$$\text{L-2: } [(1^2 + 2^2 + 3^2) / 3]^{1/2} = 2.160 \quad (...1)$$

$$\text{L-1: } [(2^2 + 3^2) / 2]^{1/2} = 2.550 \quad (...2)$$

$$\text{L-3: } [(3^2 + 4^2 + 5^2) / 3]^{1/2} = 4.082 \quad (...3)$$

$$\text{L-4: } [(5^2) / 1]^{1/2} = 5 \quad (...4)$$

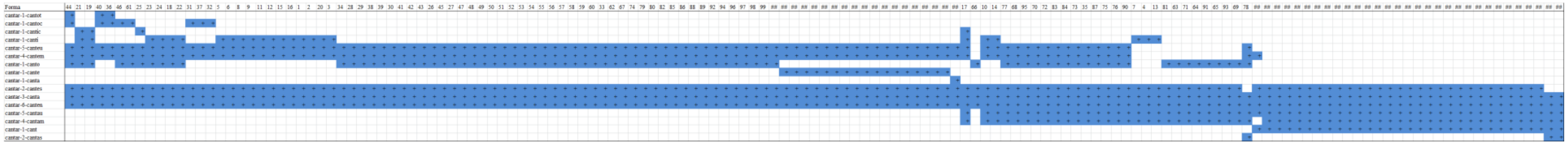


Figure 2. Integration by the distance regarding the point of origin

1.2. Other multivariate methods

1.2.1 Correlation coefficient matrix

P1	v-1	v-2	v-3	v-4
v-1	1.000	0.667	-0.167	-0.408
v-2	0.667	1.000	-0.667	-0.612
v-3	-0.167	-0.667	1.000	0.408
v-4	-0.408	-0.612	0.408	1.000

P1	v-2	v-1	v-3	v-4
v-2	1.000	0.667	-0.667	-0.612
v-1	0.667	1.000	-0.167	-0.408
v-3	-0.667	-0.167	1.000	0.408
v-4	-0.612	-0.408	0.408	1.000

Table 8

P1	d-1	d-2	d-3	d-4	d-5
d-1	1.000	-.577	.577	-1.000	.577
d-2	-.577	1.000	-.333	.577	.333
d-3	.577	-.333	1.000	-.577	.333
d-4	-1.000	.577	-.577	1.000	-.577
d-5	.577	.333	.333	-.577	1.000

P1	d-4	d-3	d-1	d-2	d-5
d-4	1.000	-.577	-1.000	.577	-.577
d-3	-.577	1.000	.577	-.333	.333
d-1	-1.000	.577	1.000	-.577	.577
d-2	.577	-.333	-.577	1.000	.333
d-5	-.577	.333	.577	.333	1.000

Table 9.

P1	L-2	L-1	L-3	L-4
d-4			v	v
d-3	v			
d-1	v	v		
d-2			v	
d-5	v	v	v	

Table 10.

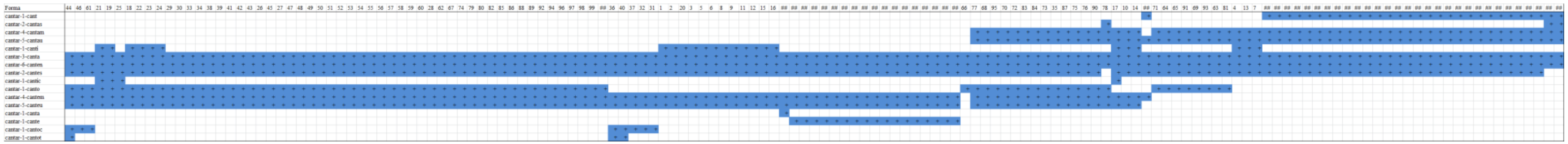


Figure 3. Integration by Correlation Coefficient

1.2.2 Principal Component Analysis

P1	L-3	L-2	L-1	L-4
d-2	v			
d-5	v	v	v	
d-3		v		
d-1		v	v	
d-4	v			v

Table 11.

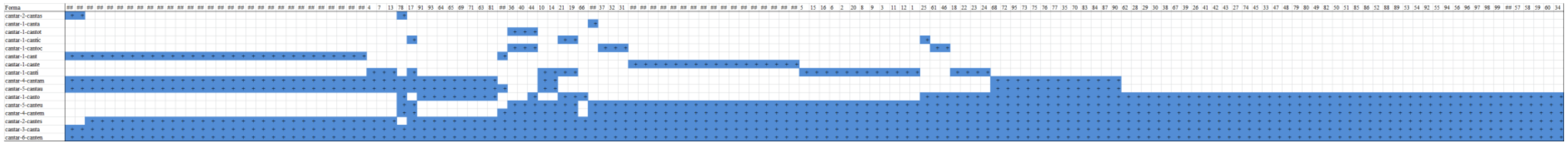


Figure 4. Integration by Principal Component Analysis

1.2.3 Factor Analysis

P1	L-3	L-4	L-1	L-2
d-2	v			
d-4	v	v		
d-1			v	v
d-5	v		v	v
d-3				v

Table 12.

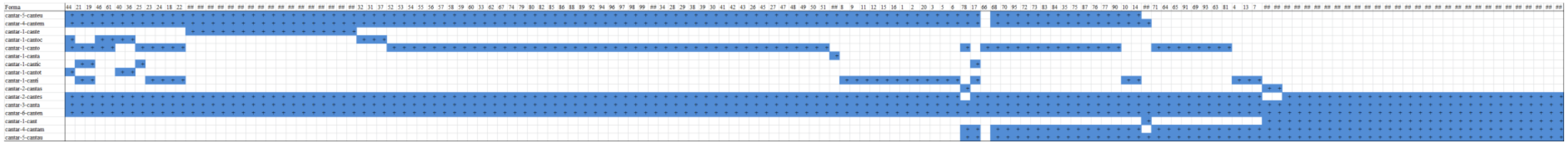


Figure 5. Integration by Factor Analysis

1.2.4 Quantification Type-III

P1	L-2	L-1	L-3	L-4
d-3	v			
d-1	v	v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Table 13.

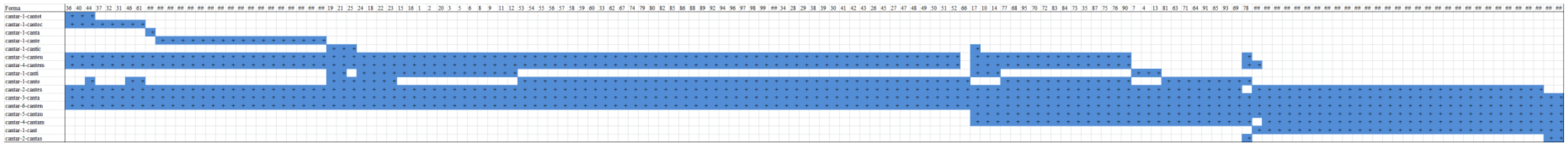


Figure 6. Quantification type-III

1.2.5 Cluster Analysis

P1	L-1	L-2	L-3	L-4
d-1	v	v		
d-3		v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Table 14.

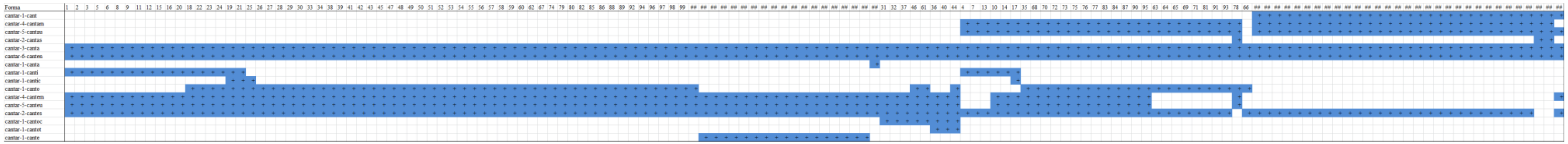


Figure 7. Integration by Cluster Analysis

2. Assessing the degree of integration

2.1. Sequential Average Distance (DMS)

$$DMS = \Sigma_i \Sigma_j \Sigma_a \Sigma_b [(i - a)^2 + (j - b)^2]^{1/2} |x_{ij}\ x_{ab}|^{1/2} / n$$

where i and j are sequences of a point of the compared pair; a and b are the same as another point on the compared pair, n is the total number of compared pairs

2.2. Referential Average Distance (RAD)

$$RAD = \Sigma_i \Sigma_j \Sigma_a \Sigma_b [(\mathbf{v}_i - \mathbf{h}_a)^2 + (\mathbf{v}_j - \mathbf{h}_b)^2]^{1/2} |x_{ij}\ x_{ab}|^{1/2} / n$$

where **v** y **h** are vectors of vertical and horizontal references.

2.3. Sequential Correlation Coefficient (SCC)

Lv	L-1	L-2	L-3	L-4
d-1	v	v		
d-2			v	
d-3		v		
d-4			v	v
d-5	v	v	v	

→

Lv	L-2	L-1	L-3	L-4
d-3	v			
d-1	v	v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Tables 15a and 15b.

Data：(X, Y) = (1, 1) (2, 1) (2, 2) (3, 1) (3, 2) (3, 3) (4, 3) (5, 3) (5, 4)

SCC = 0.82

2.4. Referential Correlation Coefficient (RCC)

Lv	L-2	L-1	L-3	L-4	Value
d-3	v				-1.42
d-1	v	v			-0.71
d-5	v	v	v		-0.01
d-2			v		0.76
d-4			v	v	1.38
Value	-1.10	-0.82	0.58	1.34	

Table 16.

Data：(X, Y) = (-1,10, -1.42) (-1.10, -0.71) … (1.34, 1.38)

RCC = 0.84

2.5. Mean Contact Index (MCI)

P1	L-1	L-2	L-3	L-4
d-1	v	v		
d-2			v	
d-3		v		
d-4			v	v
d-5	v	v	v	

→

P1	L-2	L-1	L-3	L-4
d-3	v			
d-1	v	v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Tables 17a and 17b.

P2	L-1	L-2	L-3	L-4
d-1	1	1	2	3
d-2	2	4	3	4
d-3	1	3	2	3
d-4	3	3	2	4
d-5	2	3	2	4

Table 18.

MCI: we divide the sum of links by the number of reactions.

2.6. Normal Contact Index (NCI)

```
+ - + - +
```

```
+ - +
|
+
```

«U(3)= 2» **«U_{max}(3)=2»**

When N = 4, Umax(4) = 4 in the square distribution in the right-hand figure:

```
+ - + - + - +
```

```
+ - +
|   |
+ - +
```

«U(4)=3» **«U_{max}(4)=4»**

```
+ - + - +
|   |
+ - +
```

```
+ - + - +
|   |   |
+ - + - +
```

```
+ - + - + - +
|   |   |
+ - + - +
```

```
+ - + - + - +
|   |   |   |
+ - + - + - +
```

«U_{max}(5)=5» **«U_{max}(6)=7»** **«U_{max}(7)=8»** **«U_{max}(8)=10»**

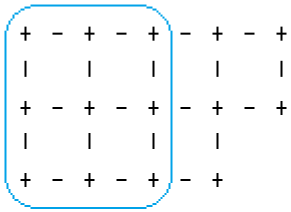
```
+ - + - +
|   |   |
+ - + - +
|   |   |
+ - + - +
```

«U_{max}(9)=12»

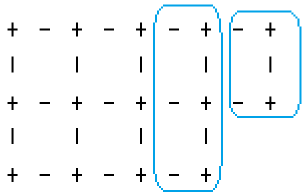
```
+ - + - + - + - +
|   |   |   |   |
+ - + - + - + - +
|   |   |   |
+ - + - + - +
```

«U_{max}(14) =20»

R = Int(Sqr(N))



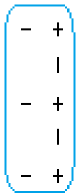
$$U_{max}(R) = 2 * R * (R - 1)$$



$$Int((N - R^2) / R)$$

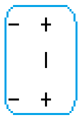
$$Int((N - R^2) / R) = Int((14 - 3^2) / 3) = 1$$

$$R + R - 1 = 2 * R - 1$$



«N=3, U = 5»

$$(N \text{ Mod } R) + (N \text{ Mod } R) - 1 = 2 * (N \text{ Mod } R) - 1 = 2 + 2 - 1 = 3$$



«N=2, U = 3»

```
R = Int(Sqr(N))
U = 2 * R * (R - 1) 'Square part
U = U + Int((N - R ^ 2) / R) * (2 * R - 1) 'Rectangular part
If N Mod R > 0 Then U = U + 2 * (N Mod R) - 1 'Remainder part
```

2.7. Comparison of multivariate methods

Evaluation	Original	ADZP	Correlation M.	P.C.A.	Factor A.	Quant.-III	Cluster
Seq. dist.	7.318	11.657	3.821	8.616	11.473	11.615	4.605
Ref. dist.	3.523	5.897	.506	1.712	.291	.018	.212
Seq. correl.	-.053	.495	-.467	.231	.435	.521	-.375
Ref. correl.	-.123	.587	-.602	.305	.442	.625	-.446
Mean contact	1.309	1.643	1.575	1.660	1.531	1.667	1.540
Norm. contact	.677	.850	.814	.858	.791	.862	.796

Table 19