

de mí, por todas las cosas,  
de los dos por todo el *pueblo*.  
(CRA, 57, v. 8)

## Apéndice 1.º:

### METODOLOGIA DEL ANALISIS FACTORIAL DE CORRESPONDENCIAS

La formulación del método que hemos utilizado para el desarrollo del programa operativo corresponde a la realizada por Rafael Romero Villafranca<sup>48</sup>, basada en los estudios de Benzécri<sup>49</sup> y Lebard<sup>50</sup>, en los que se define el Análisis factorial de correspondencias. Es, por tanto, en los trabajos de estos autores, donde se encontrará una justificación amplia y detallada de la base matemática que sustenta la metodología, cuestión que desborda el marco de aplicación de este trabajo, que no es una investigación sobre métodos, sino la utilización de uno que consideramos pertinente para llegar a conclusiones sobre la estructuración de los lenguajes de un poeta.

Es necesario de todas formas fijar lo que son los principios generales del Análisis factorial de correspondencias y su formulación, aplicada concretamente al material en estudio, definido en sus elementos, mediante las páginas anteriores, como una *matriz* multivariante en la que se relacionan términos / poemas (obras) y que se define a partir de ahora con la fórmula:

$$X_{ij}$$

siendo X el elemento genérico indicador de la matriz, j el indicador del término; i el de poema (obras).

La matriz, desde la primera aproximación, nos presenta la existencia de sensibles diferencias en los perfiles de los términos en ciertas obras, así como las semejanzas en otras. Con el Análisis factorial de correspondencias se pretende cuantificar de forma objetiva los fenómenos de diferencia y semejanza anteriores. Para lograrlo, se deben determinar los ejes fundamentales a lo largo de los cuales se manifiestan las diferencias / semejanzas entre poemas y, entre términos, hallando las coordenadas de aquéllos y éstos sobre los mismos. De esta forma es posible obtener una representación en la que las obras con frecuencias de términos parecidas se sitúen próximas entre sí y lejos de aquéllas que las tienen muy diferentes, y en la que, de forma análoga, se agrupen o se separen los términos en función de la mayor o menor semejanza de los perfiles de repartición en los poemas / obras.

El Análisis de correspondencias es una variante del Análisis de componentes principales

---

<sup>48</sup>Estas páginas son, con mínimas variantes, las propuestas por Rafael Romero Villafranca como metodología del análisis factorial de correspondencias. Cf. Romero Villafranca y Zunica Ramajo, *Geografía electoral española. Una aplicación del análisis factorial de correspondencias a las elecciones del 10 de Marzo de 1979*, «Revista española de investigaciones sociológicas», n.º 9, enero-marzo de 1980.

<sup>49</sup>Benzécri y otros, *L'Analyse des Données, II, L'Analyse des correspondances*, Ed. Dunod, París, 1976.

<sup>50</sup>Lebard, Morineau y Tabard, *Techniques de la description statistique*, Ed. Dunod, París, 1977.

de matrices, cuyas características hacen aconsejable la utilización de un sistema peculiar de ponderación de la importancia de las distintas filas y columnas.

La formulación concreta de sus principios generales, aplicados al material en estudio, es:

$$X_{ij} = \text{número total de presencias de término } j \text{ en el poema } i.$$

$$N = \sum_{ij} X_{ij} = \text{total de presencias de todos los términos en todos los poemas.}$$

$$P_{ij} = \frac{X_{ij}}{N} = \text{Proporción de presencias del término } j \text{ en el poema } i \text{ respecto al total.}$$

$$P_{i.} = \sum_j P_{ij} = \text{Proporción de presencias en el poema } i \text{ (todos sus términos) respecto al total.}$$

$$P_{.j} = \sum_i P_{ij} = \text{Proporción de presencias del término } j \text{ (en todos los poemas) respecto al total.}$$

$$P_{j/i} = \frac{P_{ij}}{P_{i.}} = \text{Proporción de presencias del término } j \text{ en el poema } i \text{ respecto al total de presencias en } i.$$

$$P_{i/j} = \frac{P_{ij}}{P_{.j}} = \text{Proporción de presencias en el poema } i \text{ del término } j \text{ respecto al total de presencias de } j.$$

Para cuantificar la diferencia entre la estructura de presencias en los poemas  $i$  e  $i'$ , en el Análisis de correspondencias, no se utiliza la distancia euclidiana ordinaria entre las  $P_{j/i}$ , sino una distancia especial, la distancia GI - dos, definida como sigue:

$$d(i, i') = \sqrt{\frac{1}{P_{.j}} (P_{j/i} - P_{j/i'})^2}$$

Como puede apreciarse, cada proporción  $(P_{j/i} - P_{j/i'})^2$  se pondera de forma inversamente proporcional a  $P_{.j}$ . Ello significa que una misma diferencia absoluta en porcentaje de presencias tiene un peso menor si se refiere a un término con alta frecuencia (por ejemplo, *luz*: 202 presencias) que a uno con baja frecuencia (por ejemplo, *limón*: 35 presencias), lo cual resulta razonable, pues cualquiera consideraría más relevante una diferencia entre el 10% y el 20% del término *limón* que entre el 60% y el 70% del término *luz*.

El Análisis de correspondencias permite obtener una representación de las obras en un espacio de dimensión dada,  $K$ , que satisface la propiedad de minimizar la suma de las

discrepancias entre los cuadrados de las distintas GI -dos primitivas y los de las existentes entre los puntos representados; en esta suma las discrepancias correspondientes a cada par de obras se ponderan proporcionalmente al producto  $P_i$ .  $P_i$  de las importancias relativas de las mismas, de forma que, cuanto mayor es el peso de una obra en el total, más fielmente se intentan reproducir sus distancias a las restantes.

El espacio seleccionado viene definido por  $K$  ejes, que participan de forma progresivamente decreciente en la explicación de las diferencias existentes entre las obras. Así, el primer factor, o eje, es siempre el más importante, seguido por el segundo, el tercero, etc. Generalmente no se fija de antemano el número de factores a obtener, es decir, la dimensión  $K$  del espacio de representación, sino que se van obteniendo factores de forma sucesiva hasta que la parte de la variabilidad total que resta por explicar es poco importante.

La coordenada  $f_{ik}$  de la fila (en este caso, obra)  $i$  en el factor  $K$  viene dada por la expresión:

$$f_{ik} = P_{ji} \frac{U_{kj}}{P_j}$$

donde las  $U_{kj}$  son las componentes del vector  $U_k$ , que es el vector propio asociado al  $K$ -grupo de obras, valor propio de una determinada matriz simétrica  $S$ , que se calcula a partir de la matriz inicial. El cociente, entre dicho valor propio y la traza de  $S$ , indica la parte de variabilidad, total, o inercia, que es explicada por el factor correspondiente.

Todo el análisis puede realizarse de forma completamente simétrica para obtener una representación de las columnas (en nuestro caso, términos). Es habitual en el Análisis de correspondencias superponer ambas representaciones, la de las filas y la de las columnas, en gráficos comunes. La interpretación en estos gráficos de las posiciones de los elementos de un conjunto (obras) frente a las del otro (términos) se basa en las relaciones siguientes entre las coordenadas  $f_{ik}$  y  $C_{jk}$  de filas y columnas sobre un eje  $K$  cualquiera:

$$f_{ik} = \frac{1}{k} \sum_j P_{j/i} C_{jk}$$

$$C_{jk} = \frac{1}{k} \sum_i P_{i/j} f_{ik}$$

Dichas relaciones indican que la coordenada de la obra  $i$  no es más que el baricentro de las coordenadas de los diferentes términos, ponderados por la importancia relativa de cada uno en la obra, multiplicado por el coeficiente de dilatación  $\frac{1}{\lambda_k}$ . De forma simétrica, la coordenada del término  $j$  no es más que el baricentro de las coordenadas de las distintas obras, ponderadas por la importancia relativa de cada una en las presencias de cada término, multiplicado por el mismo coeficiente de dilatación.

La interpretación de los factores se facilita con el cálculo de unos coeficientes denominados contribuciones absolutas y relativas. La contribución absoluta de un elemento,