

JK-Meta-Biplot y STATIS Dual como herramientas de análisis de tablas textuales múltiples

Daniel Caballero-Juliá¹, M^a Purificación Galindo Villardón², Marie-Carmen García³

daniel.caballero-julia@univ-tlse3.fr, pgalindo@usal.es, marie-carmen.garcia@univ-tlse3.fr

¹ CRESCO Université Toulouse III Paul Sabatier, 118 Route de Narbonne, 31062, Toulouse, Francia

² Universidad de Salamanca, Calle Alfonso X El Sabio s/n. 37007 Salamanca, España

³ CRESCO Université Toulouse III Paul Sabatier, 118 Route de Narbonne, 31062, Toulouse, Francia

DOI: 10.17013/risti.25.18-33

Resumen: El artículo, en primer lugar, expone brevemente los modos de análisis estadístico de datos textuales, así como de análisis cualitativo comúnmente utilizado en ciencias sociales. En un segundo momento, se propone un nuevo tratamiento de los datos con el fin de poder utilizar ambos modos de análisis de forma combinada a la hora de trabajar con múltiples textos de manera simultánea. En el artículo se detalla un procedimiento en el que se utilizan el STATIS Dual y el JK-Meta-Biplot, técnicas multivariantes que representan gráficamente una nube de puntos multidimensional en un espacio de dimensión reducida, para representar la estructura consenso de distintas participaciones a un cuestionario cualitativo. Dicho cuestionario recoge las trayectorias biográficas de profesionales españoles de la Expresión Corporal.

Palabras-clave: Análisis Estadístico de Datos Textuales; JK-Meta-Biplot; STATIS-Dual; Expresión Corporal; cuestionario

JK-Meta-Biplot and STATIS Dual as multiple textual tables tools

Abstract: First, the article briefly shows the manner of statistical analysis of textual data and qualitative analysis that are usually used in social sciences. In a second place, we propose a new treatment of data in order to be able to use both modes of analysis in combination when working with multiple texts simultaneously. The article details the procedure which uses the STATIS Dual and JK-Meta-Biplot analysis, two multivariate techniques graphing a multidimensional cloud of points in low dimension, in order to represent the consensus structure of different entries to a qualitative questionnaire. This questionnaire collects biographical trajectories of Spanish professionals of Corporal Expression.

Keywords: *Statistical Analysis of Text; JK-Meta-Biplot; STATIS Dual; Corporal Expression; Questionnaire*

1. Introducción

El desarrollo del Análisis Estadístico de Datos Textuales (AEDT) fue posible, a partir de los años sesenta, gracias a la aparición de herramientas informáticas que permitían la aplicación de nuevos algoritmos complejos (Beaudouin, 2016).

En un primer momento abordaremos la Expresión Corporal (EC), disciplina relativamente reciente y muy variada ligada a las actividades artísticas (danza, teatro, mímica...), a partir de un corpus textual procedente de un cuestionario de respuestas abiertas. Acto seguido, explicaremos cómo analizarlo de forma automática, tras realizar una breve exposición de los métodos estadísticos multivariantes empleados tradicionalmente en AEDT, incorporando la utilización del JK-Meta Biplot (Vicente et al., 2001) y el STATIS Dual (Escoufier, 1985; L'Hermier des Plantes, 1976) teniendo en cuenta el trabajo de codificación cualitativa realizado en una fase previa de análisis cualitativo clásico.

El objetivo principal de este artículo reside entonces en mostrar las posibilidades de estas herramientas a la hora de analizar distintos documentos/discursos de manera conjunta desde la perspectiva del AEDT. Así los objetivos específicos serán: primero, aprovechar la codificación cualitativa del objeto; segundo, aprovechar la capacidad de reducción de la complejidad de las técnicas multivariantes; tercero, obtener una representación gráfica de la estructura consenso de los distintos discursos y cuarto, mostrar las ventajas e inconvenientes de la aplicación de cada técnica multivariante a la hora de obtener dicha estructura.

2. Objeto y método de investigación.

2.1. El análisis cualitativo de las trayectorias biográficas de los profesionales en Expresión Corporal en España

La EC es, en España, una disciplina relativamente reciente (a partir de los 60) desarrollada por múltiples personalidades provenientes de espacios como la danza o el teatro (Ferrari, 2012; Sánchez, 2009) y cuyos dominios de aplicación son diversos (artístico, pedagógico o terapéutico). Nuestra investigación aborda el análisis de las trayectorias de profesionalización en EC en territorio español trabajando con una población compuesta por profesionales españoles de EC.

Con el fin de analizar las trayectorias de dichos profesionales sobre cuatro grandes dimensiones (formación, vida profesional, ocio y vida diaria, con especial insistencia sobre el papel que juega la EC en cada una de ellas), se ha obtenido un material compuesto por el texto procedente de las respuestas a preguntas abiertas a un cuestionario administrado en Google Forms® el cual emula las preguntas realizadas en el transcurso de entrevistas semi-estructuradas. El corpus textual ha sido analizado en diferentes etapas asociando los métodos de análisis cualitativos de las ciencias sociales con el AEDT.

La primera etapa de nuestro análisis reposa sobre la comprensión de los fenómenos y de las representaciones sociales. Está fundamentada en el análisis cualitativo propuesto por Murillo et Mena (2006) y Escobar & Román (2011). Y se trata de extraer las unidades de significación aportadas por los propios actores y analizar su manera de estructurarse en el discurso.

Para tal extracción, hemos procedido a la codificación del texto, procedimiento habitual del análisis cualitativo. Más concretamente utilizamos el sistema de codificación

propuesto por Murillo y Mena (2006) y por Escobar y Román (2011), que diferencia entre códigos de “sentido” (unidades de significación) y de “referencia” (objeto/s mencionados por el individuo). Atribuyendo tanto códigos emergentes como códigos previstos *a priori* a las frases del texto obtenemos 32 códigos de “sentido” y 22 códigos de “referencia”.

2.2. Del análisis cualitativo al análisis estadístico textual: un método de análisis textual combinado

La participación de la escuela francesa de análisis de datos, donde una de las figuras más representativas es Benzècri, ha sido de una importancia crucial para el desarrollo del AEDT (Beaudouin, 2016). Benzècri (1973) postuló los principios del Análisis Factorial de Correspondencias AFC que permitía el análisis de Tablas de Contingencia y sobre todo, ponía el acento en la visualización de las proximidades entre diferentes variables ubicándolas sobre dos o más ejes factoriales. Haciendo así posible la interpretación por parte del investigador, gracias a la proyección de una nube de datos de un espacio multidimensional dentro de un espacio de dos dimensiones, “incluso cuando éste no conoce las sutilezas del método” (Beaudouin, 2016, p. 20).

Siguiendo esta tradición, Lebart, Salem et Bécue (1988, 1994; 2000) han elaborado una “caja de herramientas” de los métodos estadísticos más importantes para en el análisis de datos textuales, incluyéndose entre ellos el AFC (simple y múltiple) y el Análisis de Clasificación Jerárquica Ascendente (CJA). Éste último método ha sido utilizado y criticado, por ejemplo, por Dalud-Vincent (2011) para el análisis de datos procedentes de entrevistas semi-estructuradas utilizadas en estudios sociológicos con el apoyo del software ALCESTE¹. Martin, Adèle et Reutenauer (2016) más recientemente, han utilizado otros programas tales como Sonal o TXM (Heiden, Magué, & Pincemin, 2010) para el análisis de este tipo de entrevistas.

En línea con los trabajos de Lebart, Osuna (2006) propone como alternativa al AFC la utilización de los métodos Biplots (Gabriel, 1971) que permitía, igualmente, la representación simultánea de los datos fila y columna de una matriz $X_{n \times p}$ proyectando la nube de puntos de un espacio multidimensional en un espacio de dos dimensiones.

2.3. Análisis Biplot de múltiples tablas léxicas

La segunda fase de nuestro proceso de análisis implica la aplicación del AEDT descrito por Caballero, Vicente y Galindo² (2014b) donde se utiliza la codificación cualitativa obtenida en la primera fase para crear la matriz de datos (38 en nuestro caso). En cada una de ellas se encuentra todo el contenido textual recogido a través de las preguntas del cuestionario utilizado. Esto implica que cada tabla (que llamaremos Tabla Léxica) es capaz de representar de forma independiente todo el discurso de cada cuestionario así como el análisis cualitativo que previamente se ha realizado sobre el mismo.

1 Resulta interesante acercarse a los trabajos de Reinert (1991, 1992), alumno de Benzècri, que retoman el AFC y CJA en el programa ALCESTE.

2 Este método de análisis textual propone la conversión del texto procedente de grupos de discusión en tablas de frecuencia corregidas y después analizadas con ayuda del HJ-Biplot (Galindo, 1986).

La metodología que aquí se propone, aprovecha dicho análisis transformando en variables o columnas de la matriz, los códigos obtenidos en la primera fase. Esto nos permitirá, como veremos, conocer su estructura. Al mismo tiempo, se reconstruye el discurso a partir de la recomposición de las frases más características de cada código cualitativo. Existen diversas propuestas para seleccionar las unidades más significativas (por ejemplo, palabras) así como maneras de representarlas. De entre ellas, cabe destacar la propuesta de Contreras, Arias Masa, Luengo y Casas García (2015) que utiliza la Teoría de los Conceptos Nucleares en asociación con las Redes Asociativas Pathfinder. Nuestra propuesta parte, sin embargo, del interés por obtener el discurso característico de cada código cualitativo. Es por ello que nuestras tablas léxicas disponen en las filas de la matriz cada una de las palabras que han sido utilizadas para responder al cuestionario.

El resultado es una tabla léxica con palabras en las filas que han sido agrupadas dentro de códigos cualitativos, los cuales figuran como variables en las columnas de la matriz. Dado que dicha operación se realiza para cada una de las intervenciones (cuestionarios en nuestro caso, pero aplicables a entrevistas, grupos de discusión, etc.), esta operación tiene, en suma, como resultado la acumulación de múltiples tablas (ver Figura 2).

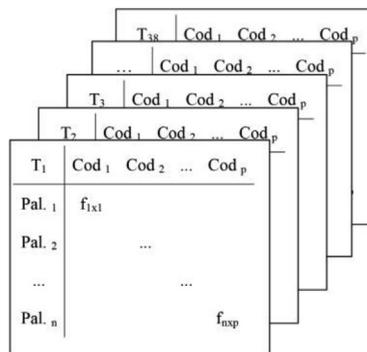


Figura 1 – Conjunto de tablas léxicas

Para reducir el volumen total de palabras y facilitar así el análisis, una vez obtenidas todas las tablas de frecuencias aplicamos el protocolo³ propuesto por Lebart, Salem et Bécue (2000) de *normalización*, que previene posibles errores de interpretación debidos a los diferentes modos de escritura; *segmentación*, que delimita las unidades léxicas que van a ser contabilizadas (palabras, frases...) y *lematización*, que estandariza y simplifica las expresiones contenidas en las unidades léxicas (verbos al infinitivo, nombres y adjetivos al singular⁴, etc.).

Acto seguido, aplicamos a cada tabla Tk el Valor de Caracterización descrito por Caballero (2011, pp. 33-35; Caballero et al., 2014b). Se trata pues de un factor de corrección que

3 Para más detalles sobre el protocolo consultar (Caballero, 2011; Caballero et al., 2014b; Lebart et al., 2000; Osuna, 2006)

4 El protocolo propuesto por Lebart et al. Así como por Osuna cambia además, los nombres y los adjetivos al masculino. Nosotros aquí conservamos ambas formas (ej: 1 chico + 1 chica = 2 chico/a).

además de reducir el volumen de datos a analizar, permite la redistribución de los pesos con el criterio de exclusividad en lugar del de “altas frecuencias”. Responde a la fórmula:

$$f'_{np} = \frac{f_{np}}{\sqrt{\max_i \sqrt{\max_j}}}$$

Donde f'_{np} representa el nuevo valor de la frecuencia, f_{np} el valor original $n \times p$ y \max_i \max_j contienen el máximo de línea y el máximo de columna respectivamente.

El material queda así listo para su análisis estadístico. Para ello, proponemos una comparación entre una técnica “clásica” para el análisis de tablas de tres índices, el STATIS Dual, y el JK-Meta Biplot, que sigue la línea de análisis abierta por los estudios del Biplot.

El análisis con el método STATIS fue introducido por L’Hermier des Plantes (1976) y Escoufier (1980, 1985) ⁵ para el tratamiento conjunto de varias tablas de datos. El método reposa sobre el principio de una configuración inicial de la matriz $n \times p$ con p variables que están medidas para n individuos en q ocasiones. Para el análisis STATIS Dual, cuya explicación matemática ha sido resumida en la Figura 2, se requieren tres niveles de análisis (Lebart, Morineau, & Piron, 1995) : en primer lugar, se captura la estructura factorial de cada tabla de datos M_k a partir de la covariancia/matriz de correlación entre las variables:

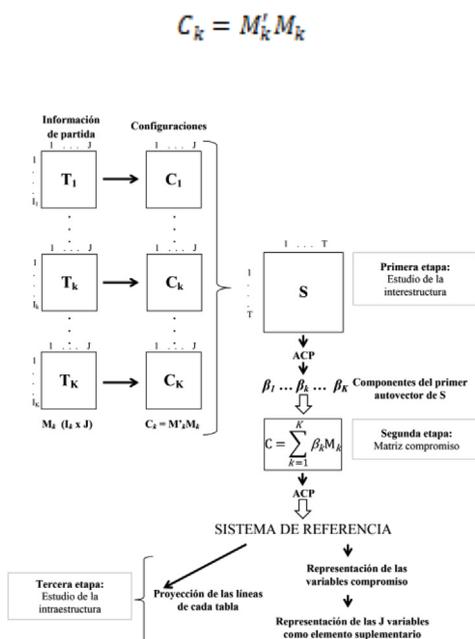


Figura 2 – Esquema de análisis STATIS Dual.

5 Ver también Pagès et al. (1976).

La comparación entre las configuraciones C_k requiere de un producto interno. En este caso el producto interno de Hilber-Schmidt:

$$\langle C_k | C_k \rangle_{\text{Hilber-Schmidt}} = \text{tr}(C_k C_k')$$

A partir de esta matriz de correlaciones se calculan los valores y vectores propios de la matriz con el Análisis de Componentes Principales (ACP) con el fin de reducir las dimensiones y producir una imagen euclídea donde cada matriz está representada por un punto dentro de este nuevo espacio.

Dentro del espacio estructurado por la distancia Hilber-Schmidt donde cada matriz está representada por un punto y su conexión con el origen del sistema de coordenadas permite la estimación gráfica del coeficiente vectorial entre pares de matrices.

Por otro lado, el análisis Biplot⁶ es una técnica de análisis estadístico propuesta por Gabriel (1971) la cual permite la representación gráfica en baja dimensión de cualquier matriz X ($n \times p$) con p variables y n individuos (Gabriel, 1971; Galindo, 1986; Galindo & Cuadras, 1986; Martín, Galindo, & Vicente, 2002). Esta técnica ha sido especialmente desarrollada en el seno del Departamento de Estadística de la Universidad de Salamanca. Hoy en día representa un amplio abanico de posibilidades de análisis multivariante. El Biplot parte siempre de la descomposición en valores singulares:

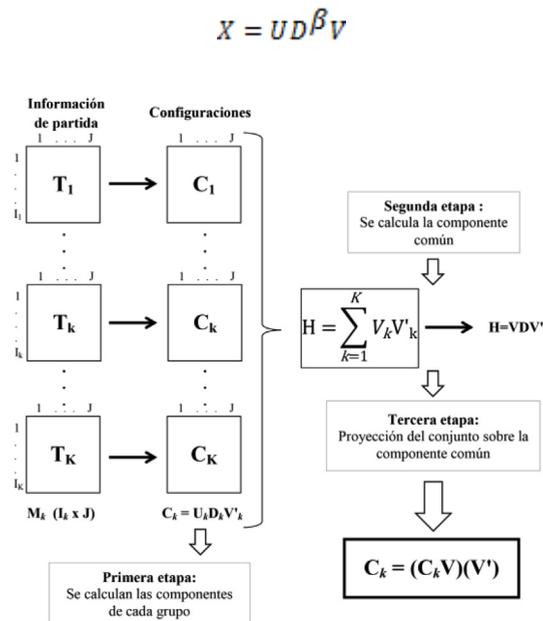


Figura 3 – Esquema de análisis JK-Meta-Biplot.

6 Agradecemos ampliamente a José Luis Vicente Villardón sus aportes y clarificaciones sobre los cálculos Biplot y Meta Biplot.

Donde U representa la matriz de vectores propios de la matriz XX' y D^β representa la matriz de « valores propios » de esta misma matriz en orden decreciente. V representa la matriz de vectores propios de $X'X$. A partir de esta descomposición, la técnica calcula los marcadores g_1, g_2, \dots, g_n para las filas (individuos) y los marcadores h_1, h_2, \dots, h_p para las columnas (variables) de la matriz para cada casilla ij de la matriz original. Se permite así una representación gráfica bi(tri)dimensional⁷. Esto es posible gracias a la factorización de la matriz preservando bien la métrica de las filas (GH-Biplot) o bien la de columnas (JK-Biplot).

Tras la integración del conjunto de datos en una sola matriz M_k , el JK-Meta-Biplot considera cada matriz T_k (discurso procedente de cada cuestionario en nuestro ejemplo) como una ocasión o grupo. Así, la técnica elimina las diferencias entre las medias de los grupos y procede a calcular las componentes de cada matriz individual:

$$C_k = U_k D_k V_k'$$

Acto seguido, se calcula la componente común al conjunto de cuestionarios (grupos) en una segunda etapa (ver Figura 3) y se proyecta el conjunto sobre esta nueva componente común.

Los resultados gráficos que se obtienen tras realizar los cálculos JK-Meta-Biplot y STATIS Dual se fundamentan en el JK-Biplot y en las correlaciones entre las variables respectivamente. En otras palabras, tendremos a las variables mejor representadas (es decir, nuestro códigos cualitativos)⁸. Todos estos cálculos son realizados de forma automática gracias al software « MultBiplot <http://biplot.usal.es> » desarrollado por Vicente (2014) en el seno del Departamento de Estadística de la « Universidad de Salamanca ».

3. Resultados

El trabajo de análisis cualitativo que podemos hacer a partir de este gráfico es bastante amplio y depende en gran medida de los objetivos planteados en el inicio de investigación. Aquí mostraremos simplemente la manera de “leer” rápidamente los resultados obtenidos a partir del ejemplo.

En el análisis del STATIS Dual, la proyección euclídea de los objetos es observable en la Figura 4. En ella, cada vector representa un cuestionario y su posición relativa debe ser interpretada como la (di)similaridad entre los diferentes discursos contenidos en cada uno de ellos. Gracias al análisis de la interestructura somos capaces de evaluar la pertinencia del análisis multidimensional de las tablas múltiples y al mismo tiempo nos permite guardar las distancias respecto de los discursos y reagruparlos de una manera más objetiva.

En la Figura 5 observamos la estructura de los códigos que configuran el “discurso compromiso” de los cuestionarios analizados.

⁷ Consultar Gabriel (1971) y Galindo (1985, 1986) para más detalles.

⁸ El STATIS clásico y el GH-Meta-Biplot operan sobre las filas de la matriz y no sobre las variables.

Más en detalle vemos que en el 2º cuadrante se encuentran, con una fuerte correlación, los vectores *RTrabajo* y *DProf* que retoman las expresiones sobre el trabajo/puesto de empleo y las trayectorias profesionales de los individuos respectivamente. Los vectores *RActExp* y *Herramienta* se inclinan hacia el primer cuadrante guardando una cierta similitud con los dos vectores que acabamos de mencionar.

Además, la interpretación del gráfico exige centrarse en el ángulo que forman *RTrabajo-REducForm/REducnoform* et *DProf-DFormación*. Un ángulo muy cercano a los 90º que indica al investigador la independencia estadística entre dichas variables. Así, desde un punto de vista cualitativo este fenómeno puede ser interpretado como la presencia de dos discursos completamente diferentes y sin relación aparente entre ellos. Dicho de otro modo, las palabras empleadas dentro de estos códigos son claramente diferentes.

Este fenómeno se repite entre los vectores *RActExp* y *RYo* el cual forma un pequeño subgrupo con una fuerte correlación *RYo*, *Lenguaje* et *RCuerpo*. Tal situación es coherente con la lectura de los datos ya que los profesionales participantes hablan de la EC como un medio de comunicación hacia el otro. Se trata ésta de una referencia que encontramos entre los vectores más pequeños, símbolo de la expresión del yo interior.

Los ejes del gráfico pueden, además, ser caracterizados por la presencia/ausencia de uno o más vectores (variables-código) en su proyección al eje, es decir, fijándonos en los cosenos más elevados de los ángulos que forman el eje y los vectores que representan a las variables de la matriz. Así, podemos observar que los vectores representados en el primer y segundo cuadrante caracterizan más fuertemente al segundo eje. En el eje de ordenadas van todos en la misma dirección. No obstante, los vectores del tercer cuadrante se contraponen a aquellos que se encuentran en el cuarto cuadrante. Su correlación es negativa salvo para los códigos *Lenguaje-RCuerpo* y *DFormación* cuyo ángulo se aproxima a los 90º. Cuando el ángulo se aproxima a los 180º no hablamos de independencia sino de asociación negativa o de discursos diferentes autoexcluyentes.

Así, en el plano 1-2 tal situación es fácilmente detectable entre el grupo de vectores *RTrabajo/DProfy Lenguaje/RCuerpo* y *RYo* y *REducnoform/REducForm*. Si ampliamos sobre los vectores más pequeños veremos este tipo de ángulo también entre *DFormación* (3º cuadrante) y *DnoProf*, datos no profesionales; *RVida*, referencia a la vida personal del individuo; y *ModeDeVida*, que retoma las representaciones que tiene el individuo de la EC como modo de vida. No obstante se ha de ser prudente con los vectores que tienen una menor calidad de representación y verificar sus ángulos con otros vectores (teniendo en cuenta la posibilidad de una representación tridimensional o superior)

Para este ejemplo, hemos pedido al programa MultBiplot el cálculo de un tercer eje para poder verificar lo que acabamos de decir. Los resultados pueden observarse en la Figura 6 la cual desvela, tras rotar la proyección de la nube de puntos, una asociación casi nula entre las variables *REducForm/REducnoform/DFormación* y *ModeDeVida/RVida* y las otras variables que no estaban representadas en el plano 1-2. Esta nueva representación nos muestra una longitud mayor de las variables *Cambio/Descubrimiento* (1º cuadrante) y *RActExp/RActLud/DNoProf* (4º cuadrante). Se aprecia una fuerte asociación entre las referencias a las Actividades expresivas, las referencias a Actividades lúdicas y los Datos no profesionales. Dicho de otro modo, el plano 1-3 expone sobre el eje 3 dos discursos bien diferenciados y fuertemente asociados.

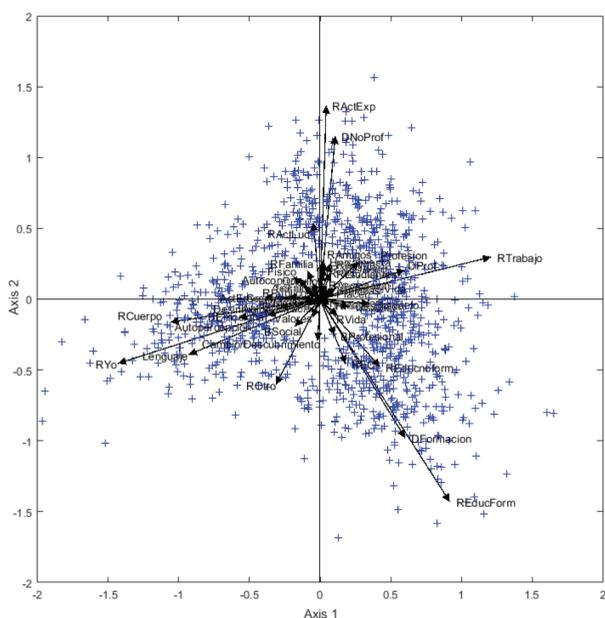


Figura 7 – Análisis JK-Meta-Biplot. Plano 1-2.

Podemos observar como los vectores de los códigos *RActExp*, *DNoProf* y, con una varianza más baja, *RActLud* tienen una asociación muy fuerte con el eje Y en sus valores positivos. Así pues, podemos interpretar los valores positivos del eje 2 como la dimensión “ocio” del discurso de los profesionales de la EC. Por otro lado, los valores negativos reciben la influencia de los vectores *ROtro* del 3er cuadrante y *REducForm*, *DFormación*, *REducnoform*, *REC* muy fuertemente asociados entre ellos, visibles sobre el 4º cuadrante. Por añadidura, estas variables (y en especial las últimas) tienen también una influencia importante sobre el segundo eje formando ángulos cercanos a 45°. En este caso hablamos de variables “de plano”, es decir, aquellas que caracterizan al cuadrante del plano. Sobre el segundo eje, podemos proyectar las variables *RYc*, *RCuerpo* y *Lengaje* entre un buen número de otros códigos con una longitud mucho menor.

La interpretación sociológica es similar a la ya expuesta por el método del STATIS Dual. La formación de los grupos de variables mejor representados se mantiene en el sentido en que ambos métodos consiguen aportar conclusiones similares. No obstante, el método JK-Meta-Biplot organiza las variables en función de su proyección en todas las dimensiones de la matriz (todos los factores provenientes del cálculo multidimensional) mientras que el método STATIS Dual utiliza únicamente la primera componente. Esta particularidad tiene como consecuencia una distribución ligeramente diferente pero mucho mejor adaptada a los datos.

Aunque, de manera general, los vectores han sido distribuidos de forma simétrica respecto a los resultados obtenidos por el método STATIS Dual, el JK-Meta-Biplot distribuye las

los códigos la posible atribución a uno u otro plano. Así, rápidamente nos percatamos que *DNoProf* (56,57% de calidad de representación en el 2º eje y 13,30% sobre el 3º), *RActExp* (50,24% sobre el 2º y 8,79%) y *RActLud* (31,19% sobre el 2º y 21,25% sobre el 3º) están mucho mejor representados en el plano 1-2 que en el 1-3. No obstante, *RVida*, *ModeDeVida* et *Cambio/descubrimiento* están bien representadas en el 3º eje con 51,78%, 49,98% y 26,30% respectivamente. Estos resultados nos permiten afirmar que los discursos que caracterizan al tercer eje reposan sobre el “cambio vital” que supone la EC para estos profesionales.

En suma, los factores (ejes) pueden ser caracterizados por las variables que forman entonces un ángulo cuyo coseno se aproxima a 1. Las palabras pueden ser utilizadas para la reconstrucción del discurso presente dentro de cada variable a partir de la proyección perpendicular de cada palabra sobre el vector de la variable que deseamos interpretar.

4. Conclusiones

A la vista de estos resultados, podemos extraer algunas conclusiones importantes según los objetivos que inicialmente se habían planteado:

Objetivo 1: Aprovechar la codificación cualitativa del objeto.

A partir de la lectura crítica, somos capaces de identificar las unidades de significación utilizadas por los actores sociales y generar así una lista de códigos que nos permita analizar objetivamente el texto teniendo en cuenta la subjetividad de los agentes.

El modo de construcción de las tablas léxicas requiere de la aplicación conjunta de la codificación por sentido y por referencia. Esto permite una reconstrucción completa del análisis cualitativo de los textos de una sola pasada; lo que, a su vez, permite no tener que analizar cada tipo de código por separado. Como se ha visto, la frecuencia relativizada de los códigos gracias al Valor de Caracterización nos permite ir, en nuestras interpretaciones, más allá de la simple caracterización por altas frecuencias (Caballero, 2011; Caballero, Vicente, & Galindo, 2014a; Caballero et al., 2014b).

Objetivo 2 y 3: Aprovechar la capacidad de reducción de la complejidad de las técnicas multivariantes y obtener una representación gráfica de la estructura consenso de los distintos discursos.

Los resultados muestran que podemos reconstruir la estructura del discurso a partir de los resultados STATIS Dual y JK-Meta-Biplot. Además, ambos métodos estadísticos pueden ser adaptados a este tipo de datos y de investigación. La gráfica aportada por estos métodos puede ser muy útil para la exposición de los resultados del análisis cualitativo. Ésta permite representar la estructura general del discurso de un colectivo o de un conjunto de texto. No obstante, los resultados obtenidos no son iguales para ambas técnicas.

Objetivo 4: Mostrar las ventajas e inconvenientes de la aplicación de cada técnica multivariante a la hora de obtener dicha estructura.

El método JK-Meta-Biplot aporta una mejor explicación de los datos proyectando la nube de puntos compromiso que es capaz de explicar un porcentaje de la varianza más

grande y esto teniendo en cuenta todas las componentes del cálculo multivariante. No obstante, los resultados aportados por los métodos JK-Meta-Biplot y STATIS Dual son satisfactorios en el sentido en que ambas facilitan la comprensión de la estructura interna del (de los) discurso(s) compartido (s) por un cierto número de individuos/textos. Desde un punto de vista matemático, las cifras de absorción de la inercia no parecen muy elevadas (especialmente si se las compara con matrices de datos puramente cuantitativas). La selección de un número más modesto de variables aumentaría dicha absorción, lo que facilitaría a su vez la interpretación de las variables. Sin embargo, como ya ha sido dicho (Caballero, 2011; Caballero et al., 2014b), la aceptación de variables con una baja calidad de representación no supone un problema desde el punto de vista cualitativo.

Referencias

- Beaudouin, V. (2016). Retour aux origines de la statistique textuelle : Benzécri et l'école française d'analyse des données. En *13ème Journées internationales d'Analyse statistique des Données Textuelles* (Vol. 3, pp. 17-27).
- Benzécri, J. P. (1973). *L'Analyse des Données: L'Analyse des correspondences*. Paris: Dunod.
- Caballero, D. (2011). *El HJ-Biplot como Herramienta en el Análisis de Grupos de Discusión*. Salamanca: Repositorio Institucional Gredos de la Universidad de Salamanca.
- Caballero, D., Vicente, M. P., & Galindo, M. P. (2014a). Análisis de Grupos de Discusión Basado en HJ-Biplot. En A. P. Costa, L. Paulo Reis, F. Neri de Souza, & R. Luengo (Eds.), *III Congreso Ibero Americano en Investigación Cualitativa Vol. 3: Artículos de Ciencias Sociales*. Badajoz: Ludomedia.
- Caballero, D., Vicente, M. P., & Galindo, M. P. (2014b). Grupos de Discusión y HJ-Biplot: Una Nueva Forma de Análisis Textual. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E2), 19–36. <https://doi.org/10.17013/risti.e2.19-35>
- Contreras, J. Á., Arias Masa, J., Luengo, R., & Casas García, L. M. (2015). Índices de Nuclearidad (Completo y Reducido), como aportación a la Teoría de Conceptos Nucleares. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E4), 16–34. <https://doi.org/10.17013/risti.e4.16-34>
- Dalud-Vincent, M. (2011). Alceste comme outil de traitement d'entretiens semi-directifs : essai et critiques pour un usage en sociologie. *Langage et société*, 135(1), 9. <https://doi.org/10.3917/l35.0009>
- Escobar, R. M., & Román, H. (2011). Presentation of self in cyberspace: an analysis of self-definitions in blogs and social networks. *Revista de Psicología Social*. 26(2), 207–222.
- Escoufier, Y. (1980). L'analyse conjointes de plusieurs matrices de données. En Jolivet E. (Ed.), *Biométrie et Temps* (pp. 59-76).
- Escoufier, Y. (1985). Objectifs et procédures de l'analyse conjointe de plusieurs tableaux. *Statist et Anal des Données*, 10(1), 1–10.

- Ferrari, H. (2012). Las cuatro Calidades Elementales de Movimiento. Método de Expresión Corporal de Marta Schinca. En G. Sánchez, & J. Coterón (Eds.), *Expresión Corporal, Investigación en la práctica* (pp. 13-30). Madrid: AFYEC.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), 453–467.
- Galindo, M. P. (1986). An alternative for simultaneous representation: HJ-Biplot. *Questiio: Quaderns d'Estadística, Sistemes, Informatica i Investigació Operativa*. Barcelona: Universitat Politècnica de Catalunya.
- Galindo, M. P., & Cuadras, C. M. (1986). Una extensión del método Biplot y su relación con otras técnicas. *Publicaciones de Bioestadística y Biomatemática Universida*(17).
- Heiden, S., Magué, J.-P., & Pincemin, B. (2010). TXM: Une plateforme logicielle open-source pour la textométrie – conception et développement. En *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010* (pp. 1021-1032).
- L'Hermier des Plantes, H. (1976). *STATIS : Structuration de tableaux à trois indices de statistique*. Montpellier.
- Lebart, L., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Lebart, L., & Salem, A. (1988). *Analyse statistique des données textuelles*. Paris: Dunod.
- Lebart, L., & Salem, A. (1994). *Statistique Textuelle*. Paris: Dunod.
- Lebart, L., Salem, A., & Bécue, M. (2000). *Analisis Estadístico de Textos*. Lleida: Milenio.
- Martin, A., Adelé, S., & Reutenauer, C. (2016). Stratégies du voyageur : analyse croisée d'entretiens semi-directifs. En *13ème Journées internationales d'Analyse statistique des Données Textuelles*.
- Martín, J., Galindo, M. P., & Vicente, J. L. (2002). Comparison and integration of subspaces from a biplot perspective. *Journal of Statistical Planning and Inference*, 102(2), 411–423. [https://doi.org/10.1016/S0378-3758\(01\)00101-X](https://doi.org/10.1016/S0378-3758(01)00101-X)
- Murillo, S., & Mena, L. (2006). *Detectives y camaleones, el grupo de discusión : una propuesta de investigación cualitativa*. Madrid: Talasa.
- Osuna, Z. (2006). *Contribuciones al Análisis de Datos Textuales*. Salamanca: Universidad de Salamanca.
- Pagès, J. P., Escoufier, Y., & Cazes, P. (1976). Opérateurs et analyse de tableaux à plus de deux dimensions. *Cahiers du Bureau universitaire de recherche opérationnelle Série Recherche*, 25, 61–89.
- Reinert, M. (1991). Proposition d'une méthodologie d'analyse des données séquentielles. *Bulletin de la Société Française pour l'Etude du Comportement Animal*, 1, 53-60.
- Reinert, M. (1992). Système Alceste: Une méthodologie d'analyse des données textuelles. En M. Bécue, L. Lebart, & Rajadell, N. (Eds.), *JADT 1990* (pp. 144-161). Barcelone: Université Polytechnique de Catalogne.

- Sánchez, G. (2009). La Expresión Corporal-Danza en Patricia Stokoe. En *Expresión Corporal y Educación*. Wanceulen: Editorial Deportiva.
- Vicente, J. L. (2014). MULTBILOT: A package for Multivariate Analysis using Biplots. Salamanca: Departamento de Estadística. Universidad de Salamanca.
- Vicente, J. L., Galindo, M. P., Avila, C., Fernandez, M. J., Martín, J., & Bacala, N. (2001). JK-META-BILOT: una alternativa al método STATIS para el estudio espacio temporal de ecosistemas. En *Conferencia Internacional de Estadística en Estudios Medioambientales* (p. 200). Cádiz: Universidad de Cádiz.