

UNIVERSITAT POLITÈCNICA DE VALENCIA

TRABAJO FINAL DE MÁSTER



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Caracterización del espacio perceptual olfativo con
técnicas estadísticas multivariantes.

Autor: Alejandro Arenal Monfort

Dirigido por: Dr. Manuel Zarzo Castelló

Año de presentación: 2016

Máster Universitario en Ingeniería de Análisis de Datos, Mejora de Procesos
y Toma de Decisiones.

Departamento de Estadística e Investigación Operativa Aplicadas y Calidad

Índice

1	Introducción	3
2	Objetivos	6
2.1	Objetivo general.....	6
2.2	Objetivos específicos	6
3	Metodología	7
3.1	Estructura y descripción de los datos.....	7
3.2	Software empleado	8
3.3	Técnicas estadísticas	8
3.4	Estudio de las diferentes dimensiones olfativas.....	11
3.5	Clasificación de los descriptores olfativos.....	12
3.6	Estudio las dos dimensiones olfativas más importantes en perfumería.....	14
3.7	Estudio del carácter hedónico de los descriptores	14
3.8	Estudio de descriptores olfativos ambiguos.....	15
4	Resultados	17
4.1	Estudio de las diferentes dimensiones olfativas.....	17
4.2	Clasificación de los descriptores.....	25
4.2.1	Análisis de conglomerados por agrupamiento difuso.....	25
4.2.2	Análisis de partición iterativa con el método K-medias.....	29
4.3	Estudio las dos dimensiones olfativas más importantes en perfumería.....	35
4.3.1	Pretratamiento A: normalización del rango de los descriptores y centrado 37	
4.3.2	Pretratamiento B: centrado y escalado a varianza unitaria.....	45
4.3.3	Discusión de los mapas obtenidos y comparación con estudios precedentes 53	
4.4	Estudio del carácter hedónico de los descriptores	58
4.5	Estudio de descriptores olfativos ambiguos.....	62
4.5.1	Diferencias entre los descriptores “fragante” y “aromático”	62
4.5.2	Diferencias entre los descriptores “perfumería” y “colonia”	69
5	Conclusiones	75
6	Referencias	78
	Anexos	81

1 Introducción

Los olores son difíciles de describir, medir y cuantificar, y por ello la investigación de la percepción olfativa y su aplicación en perfumería es todavía bastante empírica. Se han propuesto en la bibliografía diversos sistemas para clasificar los olores, pero ninguno de ellos tiene todavía amplia aceptación. Así pues, queda mucho trabajo por hacer hasta llegar a un consenso en la forma de describir y cuantificar los olores. Se pueden emplear distintos procedimientos para proporcionar una descripción detallada de la percepción olfativa, lo que se conoce como “perfil olfativo”, los cuales pueden estar basados en un conjunto de términos asignados para describir el olor (métodos semánticos) o bien basados en descripciones numéricas (perfiles olfativos cuantitativos).

Perfiles olfativos semánticos

Los métodos semánticos de descripción olfativa son los procedimientos más comunes, ya que permiten la generación rápida de datos (Harper *et al.*, 1968). Estos consisten en asignar las palabras que vienen a la memoria al oler una sustancia, describiendo cómo el olor percibido se parece a otros olores comunes que hemos oído y “recordado” con anterioridad. Estas palabras se llaman descriptores olfativos o notas olfativas, y los más útiles en la química de fragancias son aquellos que la mayoría de gente es capaz de entender, normalmente asociados a la fuente de aquel olor. Debido a que el empleo de descriptores olfativos requiere asignar las mismas palabras de la misma manera, se requiere cierto entrenamiento y algo de experiencia. Pero los resultados están influidos por la subjetividad personal, por lo que se recomienda emplear un panel de catadores para evitar cualquier sesgo.

Pocas bases de datos de perfiles olfativos semánticos han sido publicadas (Arctander, 1969; Burdock, 2004; Sigma-Aldrich, 2005). Un estudio ha compilado 1396 sustancias puras del manual de Arctander (1969), el manual de Fenaroli (Burdock, 2004) y otras fuentes, obteniendo una base de datos con 135 descriptores olfativos. El análisis de los datos dio lugar a un modelo descriptivo del espacio perceptual (Jaubert *et al.*, 1986; 1987). Otro estudio seleccionó 2467 compuestos del manual de Arctander. Aplicando diferentes métodos estadísticos multivariante, 60 descriptores fueron reagrupados en 27 clústeres, cada uno formado entre 2 a 4 notas, y 14 permanecieron como notas aisladas (Chastrette *et al.*, 1988). Un estudio similar analizó 1573 compuestos de esta base de datos, y mediante el uso del análisis clúster se clasificaron 126 descriptores en 19 clústeres (Abe *et al.*, 1990). Las descripciones olfativas de la base de Arctander básicamente fueron obtenidas por sólo una persona (S. Arctander), ocasionando un discutible grado de subjetividad personal, y algunos de los descriptores eran bastante subjetivos. Tal vez por esta razón algunos autores (Pintore *et al.*, 2006) han encontrado diferencias significativas entre la base de datos de Arctander y otra comercial de perfiles olfativos semánticos (PMP, 2001).

Los métodos semánticos han sido considerados bastante 'ruidosos' debido a las diferencias entre individuos en la interpretación de los descriptores. Para generar cierto consenso, Firmenich S.A. obtuvo una descripción de 628 compuestos puros con un comité de 7 perfumistas mediante la utilización de una lista restringida de 32 descriptores, que fueron clasificados en 10 clústeres según los resultados de cuatro métodos estadísticos multivariantes (Chastrette *et al.*, 1991). El catálogo de fragancias y aromas de la empresa Sigma-Aldrich Fine Chemicals (SAFC) (Sigma-Aldrich, 2005) contiene un inventario de descripciones olfativas semánticas para más de 1700

compuestos químicos. Un estudio seleccionó 278 descriptores para 851 compuestos de este catálogo, y aproximadamente fueron identificadas 32 dimensiones usando escalado multidimensional (Madany-Mamlouk *et al.*, 2003; Madany-Mamlouk y Martinetz, 2004). Otro estudio estadístico del catálogo SAFC clasificó 82 descriptores en 15 clases usando análisis de componentes principales (PCA) (Zarzo y Stanton, 2006).

Perfiles olfativos cuantitativos

Un procedimiento diferente para caracterizar olores consiste en evaluar la semejanza del olor por comparación directa a una serie de referencias olfativas (Schutz, 1964; Wright y Michels, 1964; Yoshida, 1975). Con este método, un panel de 6 perfumistas evaluó 309 compuestos, cuantificando sus semejanzas respecto a 30 materiales de referencia. El uso de PCA y análisis factorial permitió la clasificación de los compuestos en 14 clústeres (Boelens y Haring, 1981). Otro estudio de la misma base de datos usando PCA y análisis clúster clasificó los compuestos en 27 grupos (Ennis *et al.*, 1982). Las relaciones entre los 30 descriptores olfativos fueron analizadas y discutidas con más profundidad en otro estudio (Zarzo y Stanton, 2009). El análisis multivariante de este tipo de descripciones proporciona información fiable para estudiar las dimensiones subyacentes en el espacio perceptual olfativo. Pero es restringido a un número reducido de referencias, ya que de otra manera se necesitaría mucho tiempo y sería poco práctico.

Para estudiar las relaciones entre un alto número de descriptores, un método alternativo consiste en oler una sustancia y valorar en una escala numérica qué descriptores son los más convenientes para describir el olor, a partir de una lista predeterminada (Harper *et al.*, 1968). El empleo de una lista restringida trata de evitar cierta subjetividad que surge cuando se emplean descripciones semánticas abiertas, como es el caso de los manuales de Arctander y Fenaroli. En un estudio que emplea este método, un conjunto de observadores evaluaron 15 muestras olfativas empleando 17 descriptores (Moskowitz y Gerbers, 1974). Otro trabajo ha obtenido perfiles olfativos basados en 10 descriptores, con valores de 0 a 4 (Brud, 1986). Este número reducido de notas puede parecer demasiado restringido para una descripción apropiada del olor, y algunos estudios han usado un conjunto de 146 descriptores. Tan alto número requiere más tiempo que en el caso de obtener perfiles semánticos, y puede parecer poco exacto ya que ningún material de referencia fue provisto para cada descriptor. Pero los resultados experimentales han mostrado que, en general, el promedio de los perfiles olfativos tienden a estabilizarse mediante la utilización de un número elevado de catadores (Dravnieks, 1982).

Con esta lista de 146 descriptores, un panel de aproximadamente 120 individuos evaluó 160 muestras, obteniendo una base de datos de perfiles olfativos conocida como el Atlas de Dravnieks (1985). Aunque los métodos para describir olores basados en comparación con un conjunto de referencias parezcan *a priori* más exactos, un experimento llevado a cabo con 49 catadores reveló que este método tenía una reproductibilidad muy buena (Dravnieks *et al.*, 1978). Ya que el número de compuestos en este Atlas no es tan grande como en otras compilaciones similares basados en métodos semánticos, un estudio publicado evaluó muestras adicionales mediante la utilización de los mismos descriptores con un panel de aproximadamente 20 individuos (Jeltema y Southwick, 1986), resultando un total de 415 muestras. Una prueba de reproductibilidad con 35 muestras mostró una correlación buena entre el Atlas de Dravnieks y los resultados de este reducido panel. Un análisis factorial de esta base de datos permitió la clasificación de los descriptores en 17 clústeres, y 10 descriptores fueron considerados como

intermedios entre dos clases. La mayor parte de los clústeres propuestos agruparon descriptores similares, pero algunos de ellos incluyeron olores bastante diferentes. Éste es el caso del clúster ‘animal – fétido’, formado por olores como “rancio”, “leche rancia”, “animal”, “orina”, “sangre”, “queso” o “aceitoso”. Este estudio era uno de los primeros análisis estadísticos de una amplia base de datos de perfiles olfativos, y por tanto los resultados no pudieron ser comparados con otros estudios similares.

Posteriormente, el Atlas de Dravnieks fue utilizado en otro estudio para reducir el número de dimensiones olfativas a través de análisis multivariante (PCA), y de este modo relacionar la dimensión hedónica obtenida con la estructura física de las moléculas (Khan *et al.*, 2007). Este Atlas también fue utilizado para estudiar la estructura del espacio perceptual olfativo humano, reduciendo el número de dimensiones a través de un procedimiento de remuestreo (*jackknife*), obteniendo que más de la mitad de la varianza del espacio perceptual olfativo puede ser representado a través de una curva con dos dimensiones (Koulakov *et al.*, 2011).

Dicho Atlas también ha sido analizado a través de la técnica Factorización Matricial no Negativa, obteniendo 10 dimensiones en el espacio perceptual olfativo (Castro *et al.*, 2013). Un estudio reciente ha empleado este Atlas así como la base de datos obtenida por Boelens y Haring (1981) y las descripciones de Crocker y Herderson (1927), de forma que se ha obtenido un mapa en tres dimensiones del espacio perceptual olfativo (Zarzo, 2015).

Hasta la fecha, todavía no se ha publicado una clasificación completa de los 146 descriptores de Dravnieks en base a un PCA sobre esta base de datos. Tampoco ningún estudio ha propuesto un mapa sensorial en dos o tres dimensiones de las familias olfativas más características.

2 Objetivos

2.1 Objetivo general

Estudiar las dimensiones latentes del espacio perceptual olfativo humano a partir del análisis multivariante del Atlas de Dravnieks (1985), analizando en profundidad las dimensiones más importantes que tienen que ver con la tecnología de fragancias.

2.2 Objetivos específicos

- Obtener una clasificación de los 146 descriptores, y discutir los resultados en función de otros estudios similares realizados con la misma base de datos o de otros estudios que han empleado los mismos 146 descriptores.
- Estudiar si las dos dimensiones latentes del espacio perceptual de fragancias están de alguna forma presentes también en esta base de datos. Una de estas dimensiones describe los olores como frescos frente a cálidos, y la otra discrimina los olores femeninos frente a los masculinos. Esta segunda dimensión ha sido encontrada en el estudio del Atlas realizado por Castro *et al.* (2013), por lo que tiene interés estudiar si también puede encontrarse con PCA.
- Obtener un mapa sensorial en dos dimensiones con los descriptores más habituales en perfumería, que sea coherente con otros mapas sensoriales publicados.
- Estudiar la dimensión hedónica de los olores en base a las dimensiones latentes del espacio perceptual olfativo.
- Estudiar qué diferencias hay entre los descriptores subjetivos “aromático” y “fragante”, así como entre los descriptores “colonia” y “perfumería”.

3 Metodología

3.1 Estructura y descripción de los datos

El Atlas de Dravnieks comprende 160 muestras aromáticas (138 productos químicos puros y algunas mezclas), seleccionadas para cubrir un amplio espectro olfativo tanto de olores agradables como desagradables. Las muestras fueron disueltas en dipropilenglicol de modo que aquellos productos químicos con un olor más intenso fueron evaluados a una concentración inferior. Los datos fueron recogidos a partir de 120-140 individuos de los 12 laboratorios participantes. Los catadores fueron reclutados de diferentes organizaciones, y no todos ellos estaban familiarizados con la descripción y la clasificación de olores.

El panel fue provisto de una lista con 146 descriptores olfativos comúnmente usados, seleccionados a partir de una extensa colección con 830 términos recogidos por el comité de evaluación sensorial de la Sociedad Americana de Evaluación y Materiales (ASTM) después de revisar diversas fuentes bibliográficas e industriales (Dravnieks *et al.*, 1978; Dravnieks, 1982). Cada panelista olió la muestra y seleccionó para cada descriptor el número que mejor caracterizó el “grado de presencia” de aquella nota olfativa en la muestra: 0 (ausente), 1 (ligero), 2, 3 (moderado), 4 o 5 (extremo).

Para cada descriptor y cada muestra, se calcularon dos valores: la puntuación media (AS, *average score*) y el porcentaje de los panelistas que emplearon el término (PU, *percentage of usage*). Ambos parámetros están correlacionados ya que valores altos de puntuación media se corresponden con aquellos términos seleccionados por la mayor parte de catadores. No obstante, ambas variables no proporcionan la misma información. En caso de que dos descriptores fueran puntuados con el mismo AS pero con un PU muy diferente, se consideró razonable asignar un valor más alto ‘corregido’ al descriptor con más frecuencia escogido. Por esta razón, en vez de AS, el Atlas de Dravnieks muestra un parámetro denominado Porcentaje de Aplicabilidad (PA) para cada muestra y cada descriptor, calculado como la media geométrica entre PU y el porcentaje de puntuación máxima posible (la puntuación media expresada en una escala de 0-100, esto es, $20 \cdot AS$) según la ecuación 1 (Dravnieks *et al.*, 1978):

$$PA = \sqrt{20 \cdot AS \cdot PU} \quad [\text{ec. 1}]$$

Se dispone de una hoja Excel con la matriz PA, que consta de 160 observaciones (muestras olfativas, en filas) por 146 variables (descriptores olfativos, en columnas). En este estudio sólo se han analizado los valores PA, ya que se considera que son lo que aportan información más fiable. Así pues, no se analizan los valores de AS ni los PU.

También se ha utilizado otro estudio publicado para discutir la dimensión hedónica de la percepción olfativa (Dravnieks *et al.*, 1984), es decir, el carácter agradable o desagradable de un olor. En dicho estudio, se obtuvo la valoración hedónica, denominada “tono hedónico”, a partir de un experimento llevado a cabo por el comité de evaluación sensorial de la ASTM. El tono hedónico de cada descriptor fue valorado en una escala de 1 a 9, siendo el valor 1 el más agradable y 9 el más desagradable. Estos

valores fueron recopilados por la compañía Warner-Lambert a partir de 429 individuos. Para cada uno de ellos, los descriptores de la lista fueron mostrados en un orden diferente controlado por ordenador. Al final, se calcularon los valores medios para cada descriptor. Para facilitar su uso, los datos en la escala 1-9 fueron convertidos a forma bipolar (positivos los agradables y negativos los desagradables), restando a 5 el valor medio, haciendo que los olores desagradables tuvieran un tono hedónico negativo y los agradables, un valor positivo.

3.2 Software empleado

El software empleado en el presente estudio ha sido el siguiente:

- ProSensus MultiVariate: programa de análisis multivariante diseñado fundamentalmente para análisis de componentes principales y regresión PLS.
- Statgraphics: programa que combina procedimientos estadísticos básicos con gráficos interactivos, e incluye funciones estadísticas avanzadas.
- R: entorno y lenguaje de programación con un enfoque al análisis estadístico. Se trabaja con R a través del entorno de desarrollo integrado (IDE) RStudio.
- Hojas de cálculo Excel 2010 para el pre-procesamiento de los datos.

3.3 Técnicas estadísticas

Las técnicas estadísticas utilizadas para el análisis del Atlas de Dravnieks han sido:

PCA (*Principal Components Analysis*):

Este método estadístico condensa la información contenida en k variables primitivas en un número reducido de nuevas variables latentes, incorrelacionadas entre sí. Estas nuevas variables son definidas como combinaciones lineales de las variables primitivas. El método fue propuesto por primera vez por Pearson (1901) y Hotelling (1933), aunque su uso se extendió con el desarrollo de los ordenadores y el cálculo secuencial de las componentes a partir del algoritmo NIPALS (Esbensen y Geladi, 1987).

Para cada una de las componentes que se calculan, pueden proyectarse las observaciones sobre estas direcciones de variabilidad, las cuales cuentan con un valor de la proyección que se denomina “*score*”. La dirección sobre la que se proyecta la observación es definida por unos pesos (*loadings*). Así pues, los resultados de un PCA se discuten en términos de puntuaciones o “*scores*” y a partir de los pesos o “*loadings*” (Jolliffe, 2002).

La idea genérica de “condensar la información” proyectando sobre un nuevo subespacio puede concretarse mediante diferentes criterios, los cuales conducen a la definición de las componentes como el mismo ente matemático, que maximiza la inercia o dispersión de la nube proyectada. Los criterios son:

- Subespacio que mejor se “ajusta” a la nube: Según este criterio, inspirado en la idea de que la nube de puntos se modifique lo menos posible al proyectarse, las componentes habría que seleccionarlas como ejes ortogonales de forma que minimizasen las sumas de las distancias de los puntos de la nube primitiva a la nube proyectada. Se demuestra que esto implica seleccionar las componentes de forma que sea máxima la varianza de los puntos al proyectarse sobre los ejes.
- Subespacio que minimiza la “deformación” de la nube proyectada: Con el fin de poner de manifiesto las relaciones existentes entre las observaciones, se trataría con este criterio de conseguir que las distancias entre los individuos en la nube proyectada sean lo más parecidas posibles a las existentes en la nube primitiva.
- Predicción óptima de las k variables primitivas a partir de aquellas variables latentes (combinaciones lineales de las variables originales) que recogen la información relevante en la matriz de datos.

PLS (*Partial Least Squares Regression*):

Método estadístico utilizado para encontrar las relaciones fundamentales entre dos matrices (X e Y), basado en variables latentes, para modelar la estructura de covarianza en estos dos espacios. De forma intuitiva se puede decir que un modelo de PLS trata de encontrar la variable latente del espacio de X que explica la dirección de la máxima varianza multidimensional en el espacio Y .

Interpretación geométrica: Para cada matriz X e Y se construye un espacio de k y m dimensiones, respectivamente. La longitud de cada eje coordenado se escala generalmente a varianza unitaria. Cada observación se representa como un punto en el espacio X y otro en el espacio Y , por lo que las matrices de datos X e Y representan dos nubes de puntos conectadas entre estos dos espacios. Se calcula la media de las variables, y se restan a la matriz de datos, lo que equivale a mover el origen del sistema de coordenadas al centro de la nube de datos. La primera componente PLS es una recta en el espacio X y una recta en el espacio Y , de manera que las rectas aproximen bien los datos y las proyecciones estén convenientemente correlacionadas. Las rectas pasan por los puntos medios de las nubes. Las coordenadas de las proyecciones de los individuos en los dos subespacios (t_1 en X y u_1 en Y), en caso de haber relación lineal, están relacionadas a través de la relación interna: $u_{i1} = b_1 t_{i1} + h_i$ siendo b_1 es el coeficiente de regresión PLS, y h_i es el residuo.

La segunda componente PLS también se representa por dos rectas, una en cada espacio de variables, que pasan por la media de la nube de puntos. En el espacio X , esta segunda componente es ortogonal a la primera, sin embargo en el espacio Y esto no se debe cumplir necesariamente. El segundo par de vectores “*score*”, t_2 y u_2 , también están correlacionados linealmente.

Regresión lineal múltiple:

El análisis de regresión lineal es un proceso estadístico clásico para predecir una variable dependiente en función de varias variables explicativas (o predictoras).

Se considera que los valores de la variable dependiente (Y) son generados por una combinación lineal de los valores de una o más variables explicativas (x_1, x_2, \dots, x_k) y un término aleatorio (u):

$$Y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k + u$$

Los coeficientes son elegidos de forma que la suma de los cuadrados entre los valores observados y los pronosticados sea mínima, es decir, que se minimiza la varianza residual.

PCR (*Principal Component Regression*):

Método que combina PCA y regresión lineal. Primero se realiza un PCA sobre las variables explicativas y se obtienen las variables latentes. A continuación se realiza una regresión lineal múltiple a partir de las nuevas variables latentes como variables explicativas y la variable dependiente como respuesta.

Análisis de conglomerados (*cluster analysis*):

Engloba una amplia variedad de técnicas utilizadas para clasificar un conjunto heterogéneo de elementos (individuos o variables) en una serie de grupos que reflejen las relaciones existentes entre los mismos. Estas técnicas generan grupos formados por elementos similares entre sí. En este trabajo se han utilizado dos técnicas:

- Agrupamiento difuso (*fuzzy clustering*): Este algoritmo se caracteriza por el hecho de que los individuos pueden pertenecer a más de un grupo. Esta característica es de interés en este trabajo ya que el espacio perceptual es continuo, y lógicamente es posible clasificar algunos olores como intermedios de dos categorías. Esa asociación a cada grupo está medida por un nivel de pertenencia. Este nivel de pertenencia indicará el grado de relación entre el individuo y un grupo en particular. Se ha utilizado el paquete e1071 de R, diseñado por Meyer *et al.* (2014), y el algoritmo fuzzy C-Means, que es el más usado para esta técnica (Pal *et al.*, 1996).
- Partición iterativa con el método K-medias: es un método de agrupamiento que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo más cercano a la media. Se minimiza la suma de los cuadrados dentro de cada grupo. El algoritmo utilizado es el de Hartigan-Wong, ya que suele dar mejores resultados (Würtz *et al.*, 2009). El paquete R utilizado es stats, de *R Core Team and contributors worldwide* (2016).

El método de agrupamiento difuso se ha escogido porque hay descriptores olfativos que pueden encajar en diferentes clústeres, debido a que se trata de un espacio perceptual continuo. Este método indica la probabilidad de pertenencia de cada descriptor a cada clúster. Por lo que en aquellos descriptores que tengan valores elevados de probabilidad de pertenencia a diferentes clústeres, se considera que el descriptor es intermedio de esos clústeres, lo cual tiene bastante sentido en este caso.

La partición iterativa con el método K-medias se ha utilizado para poder comparar los resultados respecto al método anterior y otros estudios publicados. A diferencia de los

métodos jerárquicos, en este caso es necesario especificar *a priori* los grupos a formar, y de este modo se puede elegir el número de clústeres que se desea obtener.

Antes de abordar el análisis multivariante de una matriz, es necesario decidir cuál es el pretratamiento más adecuado de los datos. Para el presente trabajo se han utilizado diferentes técnicas de pretratamiento de los datos:

- Centrado a media 0: A cada valor de la matriz se le resta la media de la columna, de modo que en la matriz centrada, todas las columnas tienen media cero.
- Escalado a varianza unitaria: Cada valor de la matriz es dividido por la desviación típica de la columna. De este modo, la varianza resultante de todos los descriptores es 1.
- Normalización del rango por columnas (descriptores): Cada valor de la matriz es dividido por el rango de la columna, es decir el valor máximo menos el mínimo (Moreda-Piñeiro *et al.*, 2001).
- Normalización del rango por filas (muestras): Cada valor de la matriz es dividido por el rango de la fila, es decir, el valor máximo menos el mínimo (Moreda-Piñeiro *et al.*, 2001).
- Transformación medio rango y valor central de los descriptores: Esta transformación convierte los valores mínimos en -1, los valores máximos en +1, manteniendo la media de los valores a 0 (Moreda-Piñeiro *et al.*, 2001):

$$X_{ij}^{Transformada} = \frac{X_{ij} - cv_j}{hr_j}$$

X_{ij} = valor de la muestra i para el descriptor j .

cv_j = valor medio del descriptor j (máximo más mínimo del descriptor dividido entre 2).

hr_j = Máximo del descriptor j menos mínimo del descriptor j dividido entre 2.

- Homogeneización de los perfiles olfativos, de forma que la suma de cada fila vale uno: se ha dividido cada valor entre la suma de los valores disponibles para esa muestra. De esta forma, todas las filas suman lo mismo. Con ello se pretende mostrar la contribución que tiene ese descriptor en el olor de cada muestra.
- Transformaciones de datos: en algunos casos de descriptores con distribuciones claramente asimétricas, se han probado diversas transformaciones (raíz cuadrada, logaritmo, etc.)

3.4 Estudio de las diferentes dimensiones olfativas

Para estudiar el grado de dimensionalidad del espacio perceptual olfativo humano se ha utilizado la metodología PCA, pues el Atlas de Dravnieks pretende ser una recopilación suficientemente exhaustiva de todos los olores existentes, y el número de descriptores es muy amplio.

Como la traducción de algunos descriptores es un tanto incierta, se ha decidido poner en las figuras y las tablas los descriptores originales en inglés, pero en el texto se refiere a

ellos en castellano. Por ejemplo, “cool” y “fresh” se traducen como fresco o frío, que en castellano más o menos son casi lo mismo aplicado a un olor. Sin embargo, en inglés tienen matices bastante distintos: “cool” se aplica a los olores percibidos como frescos debido a un efecto trigeminal (como el olor a menta), mientras que “fresh” se refiere a olores frescos no asociados a este efecto trigeminal. Otro ejemplo es “powdery” y “dusty”, que se traducen por pulverulento o polvoriento. Sin embargo, el primero alude a un olor seco y cálido, mientras que el segundo se suele aplicar a olores más terrosos.

Se han probado los diferentes pretratamientos descritos en el punto anterior. Para elegir cuál de ellos es el más adecuado, se ha realizado una comparación de los gráficos de *loadings*, eligiendo aquel que mejor refleja las distintas dimensiones del espacio olfativo, atendiendo a estudios previos publicados en la bibliografía.

Se han eliminado 5 descriptores porque se emplean muy poco y por tanto su información no es relevante: “hojas de laurel”, “cerveza”, “cuerda”, “huevos frescos” y “acuoso”. A partir de la matriz de correlación entre todos los descriptores, se han identificado un par de anomalías. En concreto, “manzana” y “sopa” tienen los mismos valores (coeficiente de correlación igual a uno), lo cual es altamente improbable pues estos olores son claramente distintos. Tras verificar la correlación de estos con el resto de variables, se ha decidido eliminar el descriptor “sopa”. De modo análogo, “queso” y “comino” también tienen los mismos valores, y se ha eliminado uno de ellos (“queso”).

Para estudiar el número de componentes principales (dimensiones latentes) del espacio perceptual olfativo que aportan información relevante, se han considerado varios criterios: cantidad de varianza explicada por cada componente (R^2), el estadístico Q^2 obtenido por validación cruzada (que debe ser positivo), y el valor propio, el cual debe ser superior a uno.

3.5 Clasificación de los descriptores olfativos

La clasificación de los descriptores del Atlas de Dravnieks se ha realizado mediante análisis clúster utilizando dos técnicas: el agrupamiento difuso y la partición iterativa con el método K-medias. El primer método se ha escogido porque hay descriptores que pueden encajar en diferentes clústeres, y este método indica la probabilidad de pertenencia de cada descriptor a cada clúster. Por lo que en aquellos descriptores que tengan valores elevados de probabilidad de pertenencia a diferentes clústeres, se considera que el descriptor es intermedio de ellos. La partición iterativa con el método K-medias se ha utilizado porque, a diferencia de las técnicas jerárquicas, es necesario especificar *a priori* los grupos a formar, y de este modo se puede elegir el número de clústeres a obtener. Como se comenta más abajo, el método de agrupamiento difuso ha clasificado los descriptores en 8 grupos, lo cual es un valor bajo de agrupaciones en comparación con otros estudios similares publicados. De hecho, algunos trabajos sugieren que el número de dimensiones de la percepción olfativa puede ser hasta de 30. Por ello, se ha decidido realizar otra clasificación con un número más elevado de clústeres a través de la partición iterativa K-medias. De este modo, se consiguen así dos

clasificaciones bastante diferentes, una con pocas categorías y otra con muchas más categorías, pero la comparación de ambas tiene mucho interés. De este modo se pretende en cierta forma reproducir los múltiples estudios de la bibliografía que han abordado esta cuestión, pues algunos de ellos han propuesto un número de clasificaciones más bien bajo, y otros por el contrario más elevado. En resumen, el objetivo ha sido obtener una clasificación general, con pocos grupos, y otra más específica, con elevado número de grupos.

Para realizar el análisis clúster se ha usado la matriz completa definida en el punto “estructura y descripción de los datos”, pero hay dos parejas de descriptores que tienen correlación 1 como se ha mencionado, lo cual parece ser una errata, ya que los descriptores no guardan relación entre ellos. Por ello, se ha usado uno de los descriptores de la pareja (pues tienen los mismos valores), y en función del clúster al que pertenecen, se elige como descriptor correcto aquel que mejor encaja con el resultado obtenido. Así pues, la matriz empleada tiene 144 descriptores y 160 muestras olfativas. Previamente al análisis, hay que realizar un pretratamiento de los datos. Se han probado dos tratamientos distintos para estudiar con cuál de ellos se obtienen mejores resultados en los análisis clúster:

- El primer pretratamiento consiste en aplicar la normalización del rango de los descriptores y centrado (debido a que en el estudio de las diferentes dimensiones olfativas se observó que este pretratamiento es el que mejor refleja las distintas dimensiones del espacio olfativo), y a continuación transponer la matriz obtenida. Esta operación de transposición tiene por objetivo que los descriptores pasen a ser los individuos, y las muestras las variables, debido a que el software empleado necesita para realizar la clasificación que los elementos a clasificar aparezcan en filas.
- El segundo pretratamiento está inspirado en el PCA *clustering*. En primer lugar se analiza con PCA la matriz tras aplicarle el pretratamiento de normalización del rango de los descriptores y centrado, debido a que en el estudio de las diferentes dimensiones olfativas se observó que este pretratamiento es el que mejor refleja las distintas dimensiones del espacio olfativo. A continuación se selecciona el número de componentes principales más conveniente a utilizar. Se obtienen los *loadings* de estas componentes, y luego se transpone esta matriz. De este modo, resulta una matriz cuyas filas son los descriptores y cuyas columnas corresponden a las distintas componentes principales que aportan información relevante.

El análisis de conglomerados por agrupamiento difuso se ha realizado a partir de la matriz de distancias escalada, la cual contiene las distancias entre las filas (descriptores olfativos) de la matriz. Los tipos de distancia que se han probado para obtener la matriz escalada de distancias son los que ofrece el software (euclídea, máximo, Manhattan, Canberra y Minkowski). Se ha seleccionado el tipo de distancia que mejores clasificaciones de los descriptores ofrece tras estudiar varias de las clasificaciones que realiza cada una de las medidas de la distancia.

3.6 Estudio las dos dimensiones olfativas más importantes en perfumería

Estudios previos publicados en la bibliografía han caracterizado convenientemente las dos dimensiones olfativas más importantes en perfumería. Una de ellas discrimina los olores frescos frente a los cálidos, y otra dimensión ortogonal separa los masculinos frente a los femeninos. A partir de ambas dimensiones es posible obtener mapas olfativos de los descriptores, para lo cual una opción es aplicar la metodología PCA a una base de datos olfativos. En estos mapas, a efectos prácticos de facilitar la comparación entre ellos, suele representarse en el eje horizontal la dimensión fresco-cálido, quedando los olores cálidos en la parte izquierda y los olores frescos en la parte derecha. En el eje vertical se suele representar la otra dimensión, apareciendo los olores femeninos en la parte superior y los olores masculinos en la parte inferior (Zarzo y Stanton, 2009). Así pues, en el presente estudio se pretenden obtener mapas que reproduzcan ambas dimensiones latentes de la percepción olfativa de fragancias.

Se aplican dos pretratamientos distintos a los datos:

- Centrado y normalización del rango de los descriptores, ya que en el estudio de las diferentes dimensiones olfativas se observó que este pretratamiento es el que mejor refleja las distintas dimensiones del espacio olfativo.
- Centrado y escalado a varianza unitaria, debido a que este pretratamiento es el más empleado habitualmente.

3.7 Estudio del carácter hedónico de los descriptores

Los olores pueden evaluarse en una escala de agradable frente a desagradable, lo que suele denominarse carácter hedónico. Así pues, es posible obtener evaluaciones de este carácter para un conjunto de muestras olfativas. También se puede asignar valores hedónicos a los descriptores, pues por ejemplo los olores florales o afrutados suelen ser agradables, mientras que otros como “moho” o “pútrido” son desagradables. Dravnieks *et al.* (1984) obtuvieron el carácter hedónico de los 146 descriptores a partir de un panel de catadores. Tiene interés estudiar la relación entre esta evaluación sensorial con las variables latentes de la matriz del Atlas.

Para ello, la metodología empleada en esta parte del estudio es similar a la técnica PCR. En primer lugar se extraen las variables latentes relevantes, a continuación se transpone esta matriz (filas: descriptores; columnas: componentes), y se incorpora como variable dependiente el carácter hedónico según Dravnieks *et al.* (1984). A partir de esta estructura de datos (figura 1) pueden aplicarse técnicas de regresión. Así pues, en vez de utilizar en la regresión los valores de los *scores* obtenidos de las observaciones al realizar PCA, se utilizan los *loadings* para cada uno de los descriptores. En la regresión, los descriptores pasan a ser los individuos, y las variables explicativas son los *loadings* correspondientes a las distintas componentes principales. La variable dependiente será la puntuación otorgada para cada descriptor por los expertos en la escala hedónica (Dravnieks *et al.*, 1984).

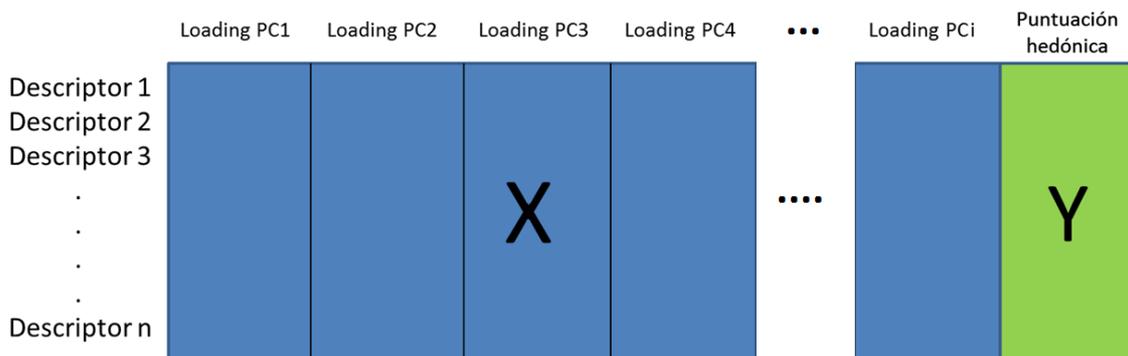


Figura 1: Estructura de datos para predecir el tono hedónico con regresión lineal. Las columnas de la matriz contienen los *loadings* obtenidos tras aplicar PCA al Atlas de Dravnieks.

En resumen, los pasos a seguir son:

- Se realiza el PCA con el pretratamiento escogido.
- Se obtienen los *loadings* de cada descriptor para las componentes principales relevantes.
- Se realiza una regresión lineal múltiple paso a paso (*stepwise*), comparando la opción *backward* y *forward*, considerando los descriptores como individuos, y la puntuación hedónica del descriptor como variable dependiente, tal como se muestra en el esquema de la figura 1.
- En el modelo de regresión se estudian los posibles efectos cuadráticos y la normalidad de los residuos.
- Se analizan los resultados obtenidos y se propone el modelo de regresión definitivo.

Este estudio se ha llevado a cabo mediante el pretratamiento previo a PCA de normalización del rango de los descriptores y centrado de la matriz, ya que en el estudio de las diferentes dimensiones olfativas se observó que este pretratamiento es el que mejor refleja las distintas dimensiones del espacio olfativo. Se han eliminado 7 descriptores, cinco de ellos porque se emplean muy poco de modo que su información no es relevante, y dos de ellos porque sus valores son erróneos tal como se ha justificado más arriba.

3.8 Estudio de descriptores olfativos ambiguos

El análisis de la información proporcionada por las parejas “fragante” - “aromático” y “perfumería” - “colonia” se realiza mediante regresión PLS. Para cada pareja, se parte de la matriz tras eliminar los 7 descriptores mencionados en el párrafo anterior. A continuación se realiza una regresión PLS considerando la pareja de descriptores a estudiar como variables dependientes, y el resto como variables explicativas. En el modelo obtenido se identifican aquellos descriptores de mayor VIP (influencia de la variable X en la proyección) en las diferentes componentes. Para simplificar el modelo, se repite el PLS descartando aquellos descriptores de menor VIP, y se visualiza el

gráfico de *loadings* obtenido para las dos primeras componentes PLS para estudiar cómo las variables se combinan para formar la relación cuantitativa entre X e Y.

A continuación se ajusta un PLS para cada descriptor de la pareja por separado, en función del resto de descriptores como variables X. Se obtiene un modelo para cada elemento de la pareja, y se comparan los resultados, estudiando los descriptores de mayor VIP en la primera componente. El modelo se simplifica descartando las variables de menor VIP. Finalmente, se comparan los coeficientes PLS obtenidos para cada miembro de la pareja.

4 Resultados

4.1 Estudio de las diferentes dimensiones olfativas

Se han realizado diversos modelos PCA para evaluar las diferencias entre los distintos preprocesamientos mencionados en el punto 3. Para elegir cuál de ellos es el más adecuado, se estudian los gráficos de *loadings* para las primeras componentes principales, analizando qué gráficos reflejan de manera más adecuada las diferentes dimensiones del espacio perceptual olfativo humano, comparando los resultados con los estudios de Castro *et al.* (2013) y de Jeltema y Southwick (1986). Tras las diversas pruebas realizadas, el pretratamiento que ofrece resultados más coherentes con otros estudios es el de normalización del rango de los descriptores y posterior centrado. No se adjuntan en la presente memoria estos gráficos debido al elevado número de figuras que se han obtenido, ya que se han probado muchos pretratamientos, y para cada uno de ellos sería necesario adjuntar varios gráficos.

A continuación se discute el número de dimensiones relevantes del espacio perceptual olfativo. En la figura 2 se aprecia que a partir de 18 componentes aproximadamente, la varianza explicada (R^2) aumenta muy lentamente (menos del 1%), mientras que el parámetro Q^2 es ya casi horizontal. Por ello, parece adecuado seleccionar unas 18 componentes principales. En la tabla 2 se muestran los valores propios para las 18 componentes principales, los cuales son superiores a 1.

La misma matriz de datos se ha analizado también con PCA empleando el programa SIMCA-P 10.0, aunque con el pretratamiento de autoescalado (datos centrados y escalados a varianza unitaria). En este caso, el estadístico Q^2 es positivo hasta la 9ª componente, a partir de la cual pasa a ser negativo (Tabla 3). Con 15 componentes principales, los valores propios obtenidos son similares a los obtenidos en la Tabla 2. Hay que tener en cuenta que el Atlas de Dravnieks pretende abarcar un amplio abanico de olores, pero para algunos de ellos la cantidad de muestras es muy reducida. Por este motivo, el algoritmo de validación cruzada que emplea SIMCA-P para calcular Q^2 da lugar a valores negativos. Así pues, si nos quedamos sólo con 9 componentes con Q^2 positivo según la Tabla 3, este valor es quizás insuficiente. No es posible determinar con exactitud el número de dimensiones relevantes de la percepción olfativa a partir del Atlas de Dravnieks, pero los resultados sugieren que el número máximo de dimensiones podría ser quizás de unas 15 a 18 dimensiones.

Se concluye que el número de componentes principales que aportan información relevante del espacio perceptual olfativo, a partir de los resultados obtenidos, es aproximadamente 18, lo cual es coherente con otros estudios citados en la bibliografía, que las sitúan entre 10 y 30 (ver Tabla 1 de Zarzo, 2008).

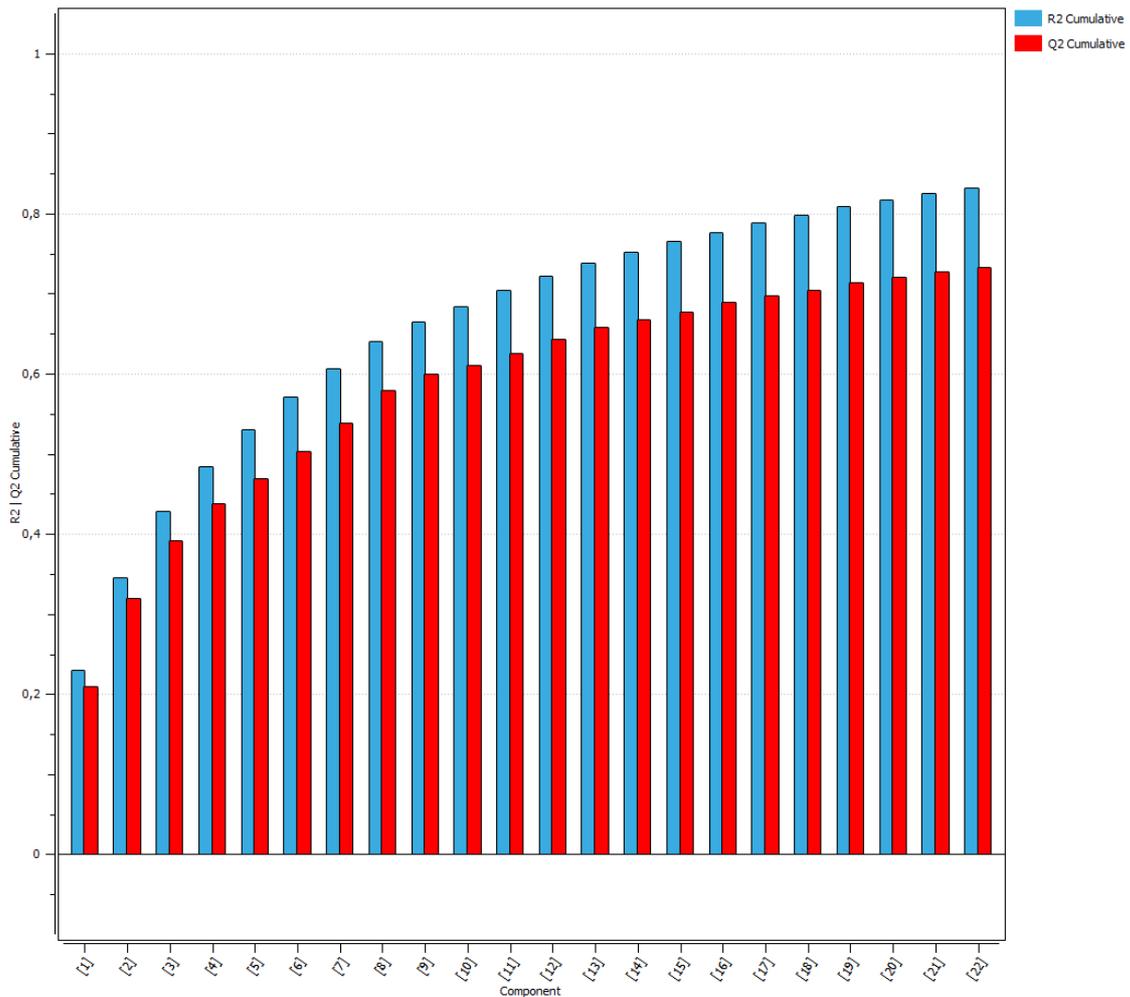


Figura 2: Varianza explicada (R^2) acumulada y varianza explicada por validación cruzada (Q^2) acumulada para cada componente principal. El pretratamiento empleado es el de normalización del rango de los descriptores y centrado.

Tabla 1: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 18 componentes principales (R^2 total = 0.797; Q^2 total = 0.704). Pretratamiento empleado: normalización del rango de los descriptores y centrado.

Componente	R^2	Q^2	Componente	R^2	Q^2
1	0.228	0.209	10	0.020	0.011
2	0.116	0.109	11	0.019	0.015
3	0.083	0.072	12	0.018	0.017
4	0.056	0.046	13	0.016	0.015
5	0.045	0.032	14	0.014	0.009
6	0.041	0.033	15	0.013	0.010
7	0.036	0.035	16	0.012	0.0115
8	0.033	0.041	17	0.011	0.008
9	0.024	0.020	18	0.010	0.007

Tabla 2: Valores propios para las 18 componentes principales obtenidas con el pretratamiento de normalización del rango de los descriptores y centrado.

Número componente	Valor propio	Número componente	Valor propio	Número componente	Valor propio
1	24,391	7	4,919	13	2,459
2	13,855	8	4,385	14	2,394
3	11,181	9	3,708	15	2,269
4	8,402	10	3,555	16	1,868
5	6,128	11	3,296	17	1,768
6	5,690	12	2,703	18	1,697

Tabla 3: Características de las 15 componentes principales obtenidas para el Atlas de Dravnieks con pretratamiento autoescalado: varianza explicada (R^2) y acumulada ($R^2(\text{cum})$), valores propios, y varianza explicada por validación cruzada (Q^2).

Componente	R2	R2(cum)	Valor propio	Q^2
1	0.131	0.131	18.2	0.108
2	0.085	0.216	11.8	0.062
3	0.064	0.280	9.0	0.028
4	0.050	0.330	7.0	0.012
5	0.041	0.371	5.7	0.018
6	0.037	0.408	5.1	-0.009
7	0.035	0.443	4.8	0.010
8	0.033	0.476	4.6	0.009
9	0.029	0.505	4.1	0.003
10	0.026	0.531	3.6	-0.001
11	0.024	0.556	3.3	-0.003
12	0.022	0.577	3.0	-0.014
13	0.020	0.597	2.8	-0.024
14	0.019	0.616	2.7	-0.021
15	0.017	0.633	2.4	-0.020

A continuación se discuten las componentes principales obtenidas que aportan la información más relevante. Las dos dimensiones latentes fundamentales del espacio perceptual de fragancias quedan reflejadas en el Atlas de Dravnieks analizado. Una de estas dimensiones discrimina los olores femeninos frente a los masculinos. Los olores florales, dulces y afrutados son típicamente los más femeninos, pues estas notas son las más habituales en fragancias destinadas a mujeres. Por el contrario, las notas más masculinas son las amaderadas, herbáceas, etc. La siguiente figura muestra los pesos correspondientes a la tercera frente a la primera componente. Los descriptores más femeninos (“dulce”, “floral”, “afrutado”) aparecen a la izquierda, mientras que los más masculinos (“amaderado”, “nuez”, “terroso”, “hierba machacada”, “ahumado”) aparecen en la zona superior derecha.

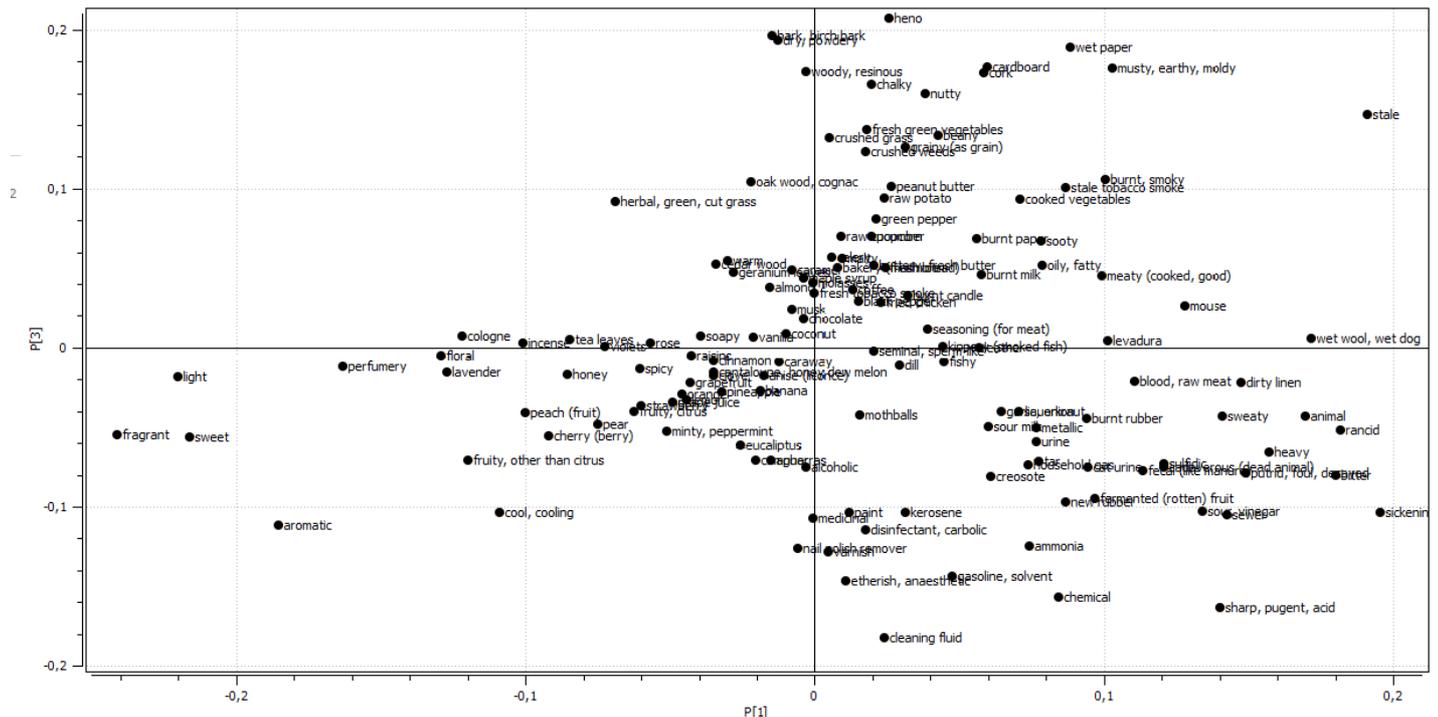


Figura 3: PCA con 139 variables: *loadings* correspondiente a las componentes 1 y 3.

Se observa en la figura 3 que la primera componente tiene valores negativos elevados para descriptores de fragancias femeninas, como “fragancia”, “aromático”, “dulce”, “perfumería” y “floral”. Estos descriptores coinciden con la dimensión 1 obtenida en el estudio de Castro *et al.* (2013).

La tercera componente (eje vertical de la figura 3) toma valores elevados para descriptores de fragancias masculinas, como “madera”, “terroso” o “hierba”. Estos descriptores coinciden con la dimensión 2 de Castro *et al.* (2013).

Lo esperado era que estos dos grupos de descriptores contrapuestos (masculino frente a femenino) hubieran aparecido en una sola componente, pues son olores opuestos perceptualmente. Por el contrario, se ha obtenido en este caso (figura 3) que los olores desagradables son opuestos a los femeninos (eje horizontal) y también a los masculinos (eje vertical), lo cual es lógico, ya que en un perfume se buscan olores que sean agradables. Además, se ha observado que estos olores desagradables coinciden con la cuarta dimensión obtenida por Castro *et al.* (2013). Los descriptores de olores químicos son los más opuestos a los descriptores olfativos masculinos, lo cual resulta difícil de interpretar. No obstante, este grupo de descriptores químicos coincide con la quinta dimensión hallada por Castro *et al.* (2013).

Tras verificar distintas combinaciones de componentes, se ha encontrado que básicamente la cuarta componente discrimina si los olores son cálidos o frescos, tal como se refleja en la figura 4:

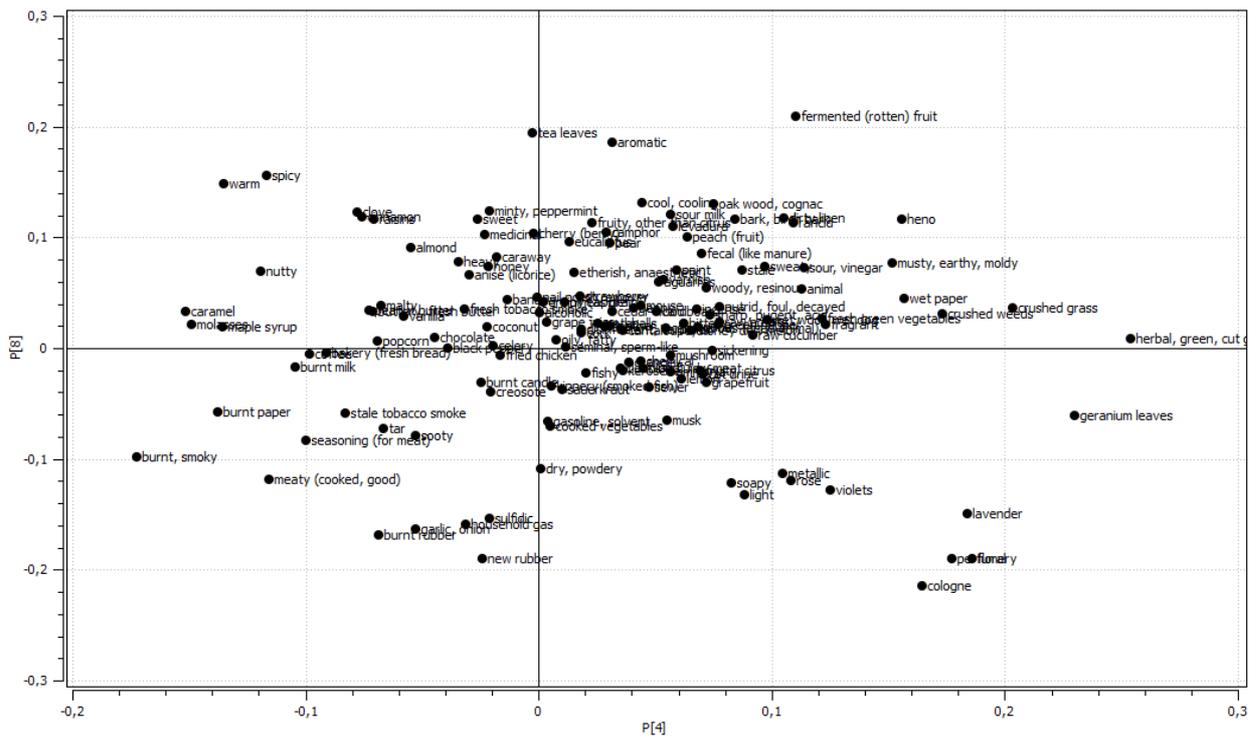


Figura 4: PCA con 139 variables: *loadings* correspondiente a las componentes 4 y 8.

En el gráfico de la figura 4 se observa que en la parte izquierda se hallan olores cálidos (“warm”), tal como “quemado-humo”, “caramelo”, “melaza” o “clavo”, mientras en la parte derecha se encuentran olores frescos como “verde”, “herbáceo” y “colonia”.

Los olores afrutados son típicamente femeninos. Sin embargo, el olor de las frutas cítricas es más fresco que las del resto. Este hecho, bien conocido en perfumería, queda de manifiesto en el siguiente gráfico que muestra los *loadings* de la séptima y la décima componentes (figura 5).

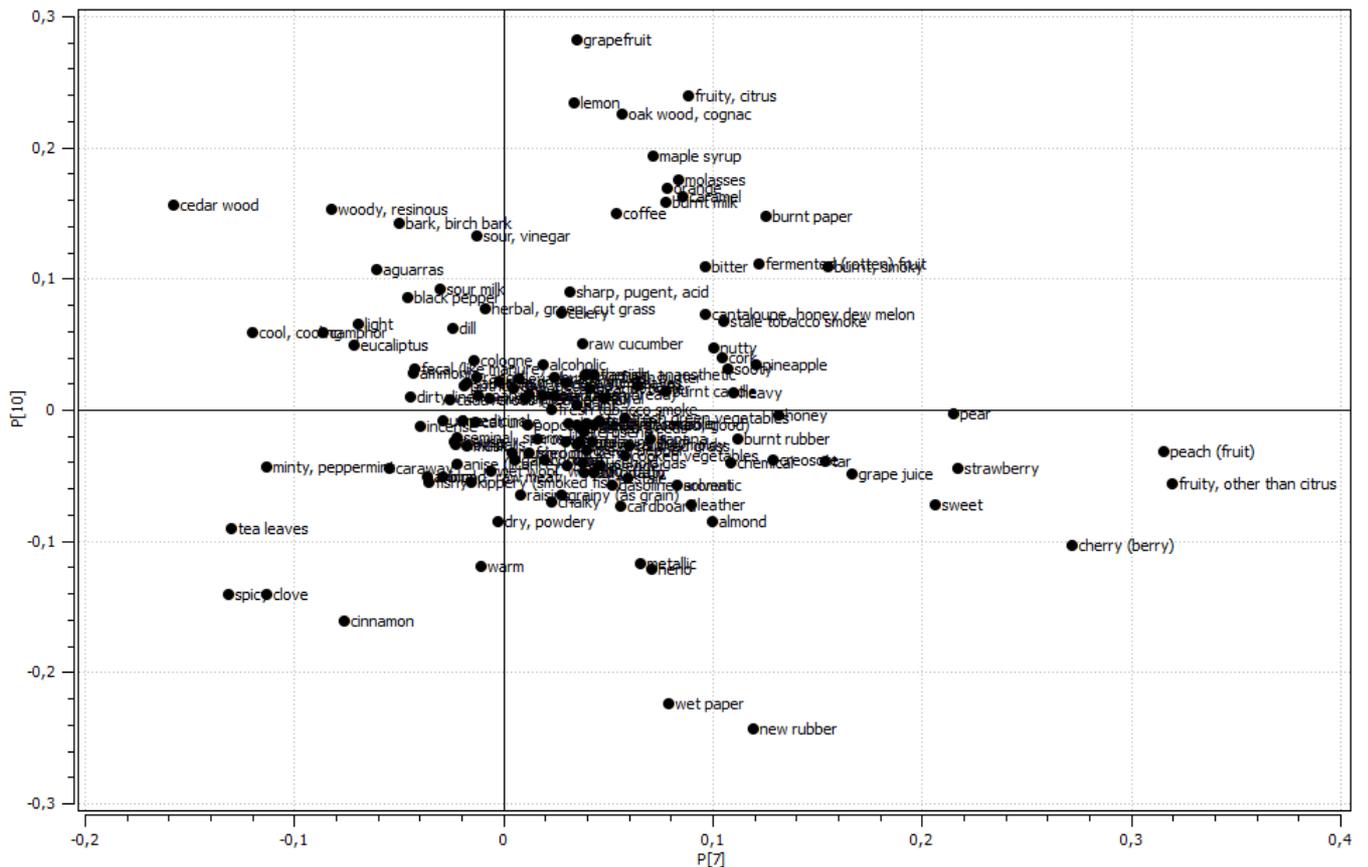


Figura 5: PCA con 139 variables: *loadings* correspondiente a las componentes 7 y 10.

Se puede apreciar que la séptima componente está básicamente determinada por los olores afrutados no cítricos, ya que encontramos descriptores como “fruta no cítrica”, “melocotón”, “cereza”, “fresa” y “pera”. Esta dimensión coincide con la tercera del estudio de Castro *et al.* (2013). También coincide con el factor 3 (frutas no cítricas) en el estudio de Jeltema y Southwick (1986). Es interesante observar que estos descriptores son perceptualmente opuestos a olores especiados (“clavo”, “canela”) y a los canforáceos (“cool” y “menta”).

La décima componente está básicamente determinada por descriptores asociados a frutas cítricas, ya que los descriptores con mayor peso son: “pomelo”, “fruta cítrica” y “limón”. Esta dimensión coincide con la décima de Castro *et al.* (2013) y el factor 10 (denominado “cítricos”) obtenido por Jeltema y Southwick (1986). Se puede apreciar que los olores especiados (“clavo” y “canela”) son opuestos a los cítricos, ya que se hallan en una posición opuesta en la figura 5.

PCA tras reducir el número de descriptores:

Tras realizar un PCA con las 139 variables y seleccionar el mejor pretratamiento, se ha considerado la posibilidad de reducir el número de variables para obtener gráficos más fáciles de interpretar. La idea es eliminar variables redundantes. Por ejemplo, “crushed grass” (hierba machacada) y “crushed weeds” (malas hierbas machacadas) aluden

básicamente al mismo olor y ambos descriptores aparecen muy juntos en los gráficos mostrados. Así pues, tiene sentido dejar solamente uno de ellos. Con este criterio, tras identificar las parejas de descriptores de mayor correlación y observar su posición en las figuras, se han eliminado algunos descriptores “redundantes”. En total se han eliminado 48 descriptores, quedando 91 variables en el PCA. Este procedimiento ha facilitado el estudio de relaciones entre descriptores de familias que no se apreciaban en el PCA anterior.

A continuación se muestran las familias olfativas de las cuales se ha eliminado algún descriptor, indicándose el descriptor considerado como más representativo que se ha mantenido en el modelo:

- Disolvente: “líquido limpiador”.
- Afrutado no cítrico: “afrutado no cítrico”.
- Afrutado cítrico: “cítrico”
- Balsámico-miel: “miel”
- Amaderado: “madera-resinoso” y “madera de cedro”.
- Herbáceo: “herbáceo, verde, césped cortado” (se trata de un solo descriptor).
- Vegetal: “vegetales verdes frescos”.
- Quemado: “quemado, ahumado”.
- Animal: “animal”.
- Pescado: “pescado”.

A continuación se muestran los nuevos resultados obtenidos con PCA (91 variables). La componente cuatro discrimina los olores florales frente a los especiados, como se muestra en el siguiente gráfico (figura 6):

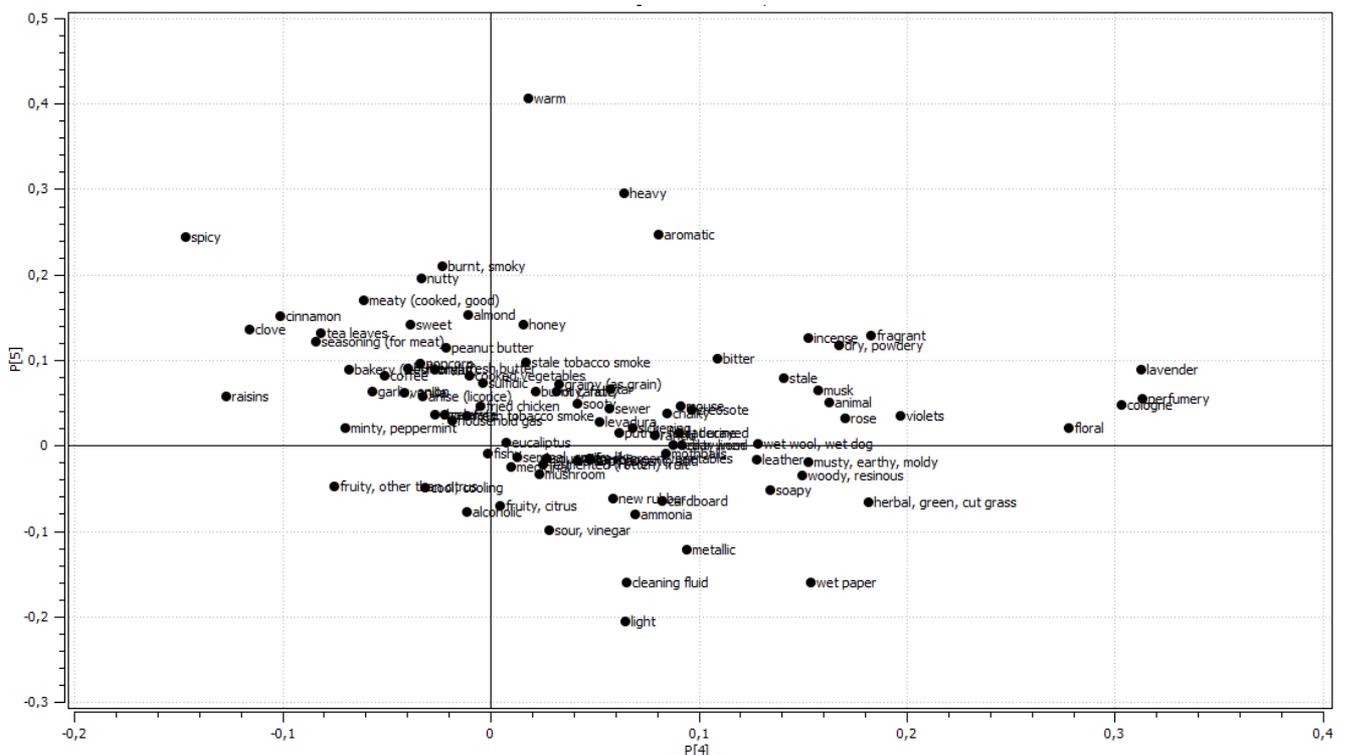


Figura 6: PCA con 91 variables: *loadings* correspondientes a las componentes 4 y 5.

En la parte izquierda del gráfico se encuentran los descriptores especiados: “especia”, “clavo” y “canela”. Estos descriptores coinciden con el factor 13 (“especiado”) del estudio de Jeltema y Southwick (1986). En la parte derecha encontramos olores habituales en cosméticos (“floral”, “lavanda”, “violetas”, “perfume” y “rosa”), los cuales coinciden con el factor 7 (denominado “floral”) en el estudio de Jeltema y Southwick (1986).

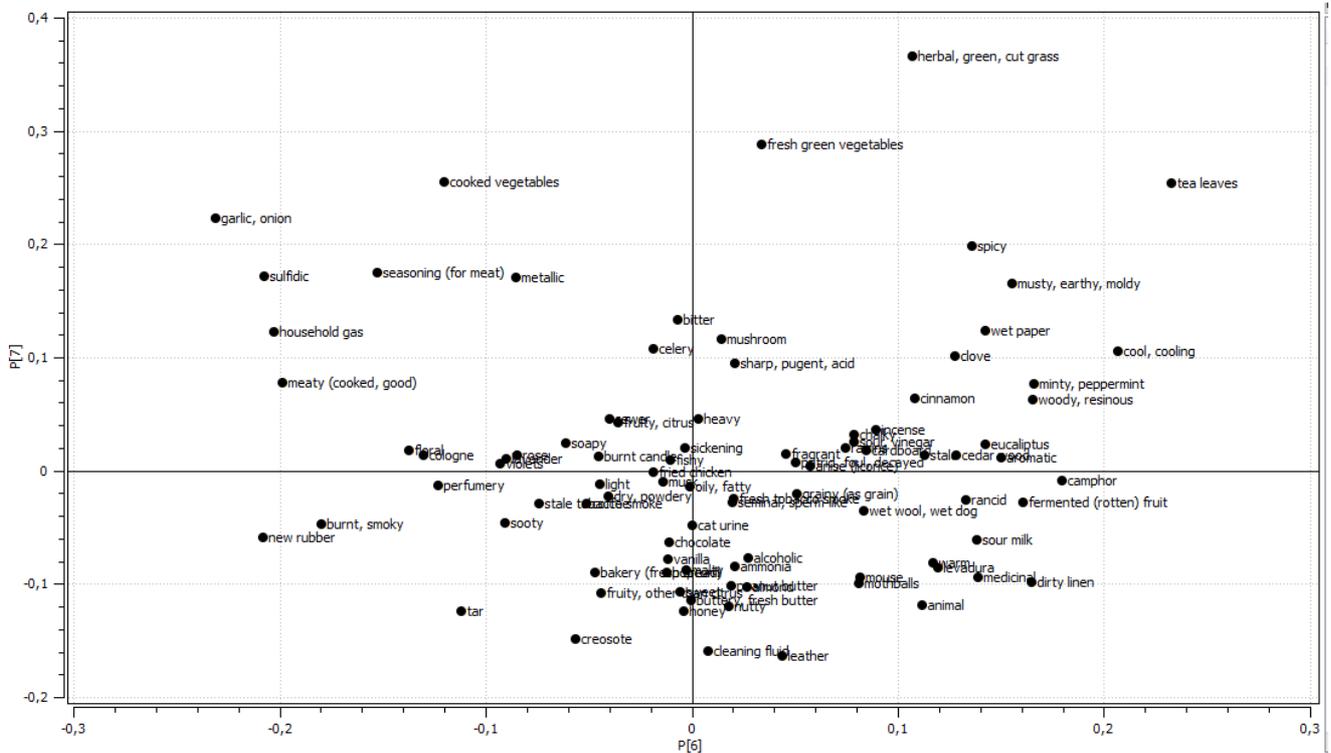


Figura 7: PCA con 91 variables: *loadings* correspondientes a las componentes 6 y 7.

En el gráfico anterior se muestran los pesos (*loadings*) correspondientes a las componentes 6 y 7 (figura 7). En la parte superior izquierda aparecen los olores azufrados (“sulfuroso”, “ajo”, “cebolla”, “gas doméstico”), los cuales coinciden con el factor 11 de Jeltema y Southwick (1986). En la parte derecha aparecen los olores canforáceos, como los descriptores “menta”, “cool”, “alcanfor”, “eucaliptus”, lo cual encaja con la componente 6 de Castro *et al.* (2013) y el factor 8 denominado “cool-menta” según Jeltema y Southwick (1986).

4.2 Clasificación de los descriptores

4.2.1 Análisis de conglomerados por agrupamiento difuso

Se ha realizado el análisis de conglomerados difuso con las dos matrices obtenidas de los dos pretratamientos aplicados a la matriz original (descritos en el punto 3.5). El pretratamiento inspirado en la metodología PCA-clúster es el que mejor resultados ha proporcionado, tal como se detalla a continuación.

Tras realizar el PCA con la matriz de 144 descriptores por 160 muestras olfativas, la tabla 4 muestra la bondad de ajuste (R^2) y la varianza explicada por validación cruzada (Q^2) correspondiente a las 10 componentes principales:

Tabla 4: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 10 componentes principales. Pretratamiento: normalización del rango de los descriptores y centrado. R^2 total = 0.670; Q^2 total = 0.592.

Componente	R^2	Q^2	Componente	R^2	Q^2
1	0.223	0.204	6	0.042	0.040
2	0.112	0.104	7	0.036	0.030
3	0.081	0.070	8	0.032	0.039
4	0.057	0.048	9	0.024	0.018
5	0.044	0.029	10	0.019	0.010

Se ha realizado el análisis clúster de agrupamiento difuso con un distinto número de componentes principales. Se ha obtenido que con 4 componentes, que explican el 47,3 % de la varianza total, se consiguen los mejores resultados, al comparar los grupos de descriptores obtenidos con los estudios de Castro *et al.* (2013) y de Jeltema y Southwick (1986). Emplear solamente cuatro componentes parece insuficiente *a priori*, pues otras más aportan también información relevante y parecería lógico incluirlas también en el análisis clúster, pero curiosamente las agrupaciones de descriptores obtenidas eran bastante diferentes a las obtenidas por los estudios de Castro *et al.* (2013) y de Jeltema y Southwick (1986). También ocurre al realizar el análisis con la matriz original.

A partir de los *loadings* de las cuatro componentes principales, transponiendo los datos se obtiene la matriz con los descriptores en filas y los *loadings* en columnas. Esta matriz se ha analizado con la técnica de agrupamiento difuso considerando los diferentes métodos que ofrece el software para medir la distancia entre filas (descriptores en este caso) a la hora de calcular la matriz de distancias. Para cada tipo de distancia, se ha comprobado el número de agrupaciones obtenidas para las particiones que dan lugar a valores más bajos del índice Xie-Beni (Xie y Beni, 1991), con el requisito de que el número de clústeres sea mayor a 5. Tras analizar los grupos de descriptores que se obtienen con cada uno de los métodos para medir la distancia en la matriz de distancias, el que ofrece agrupaciones de manera más similar a los estudios publicados que se han mencionado es el método de distancia máxima. Los distintos métodos para calcular la distancia de la matriz de distancias también se han probado para un distinto número de componentes principales.

A continuación se muestran los resultados obtenidos para la matriz de distancias con el método óptimo encontrado (distancia máxima).

En la tabla 5 se observa, para cada posible partición (de 2 a 20 clústeres), los valores obtenidos de los índices Xie-Beni (xb) (Xie y Beni, 1991), Fukuyama-Sugeno (fs) (Fukuyama y Sugeno, 1989), y *partition entropy* (pe). La agrupación es mejor cuanto más bajo sea su valor. Por el contrario, mayores valores del coeficiente de partición (pc) indican una mejor partición.

Tabla 5: Valores de los índices Xie-Beni (xb), Fukuyama-Sugeno (fs), *partition entropy* (pe) y coeficiente de partición (pc) para un distinto número de particiones.

Número clusters	2	3	4	5	6	7	8
xb	0,00293	0,00135	0,00197	0,00121	0,00092	0,00212	0,00141
fs	-0,28774	-0,58061	-0,60603	-0,75102	-0,81242	-0,78758	-0,76522
pc	0,70434	0,64024	0,58425	0,58087	0,57248	0,55259	0,53792
pe	0,45163	0,642	0,791	0,851	0,910	0,980	1,046
Número clusters	9	10	11	12	13	14	15
xb	0,00160	0,00209	0,00175	0,00183	0,00167	0,00266	0,00162
fs	-0,73917	-0,72481	-0,70833	-7525,91500	-0,81119	-0,72832	-0,77496
pc	0,52462	0,50115	0,49455	0,51186	0,51356	0,48313	0,49380
pe	1,104	1,176	1,213	1,186	1,205	1,301	1,288
Número clusters	16	17	18	19	20		
xb	0,00167	0,00266	0,00250	0,00205	0,00207		
fs	-0,77504	-0,70566	-0,65609	-0,72153	-0,71617		
pc	0,49489	0,47019	0,46890	0,48649	0,48773		
pe	1,304	1,381	1,415	1,366	1,384		

Se observa que con dos clústeres se obtienen los mejores valores en los indicadores fs, pc y pe. El valor de xb proporciona los mejores resultados con 6 clústeres. Debido a que nos interesa tener un número relativamente elevado de clústeres para clasificar convenientemente los descriptores, se descarta la opción con 2 clústeres, ya que son demasiado pocos. La tabla indica que con un número de agrupaciones de 5, 6, 8 y 9 clústeres, el indicador xb toma un valor bastante bajo, lo cual sugiere centrar la atención en alguna de estas opciones. Tras analizar detenidamente la coherencia de los descriptores clasificados en los clústeres obtenidos con 5, 6, 8 o 9 grupos, se ha obtenido que los mejores resultados corresponden a la opción con 8 clústeres.

En el anexo 1 se muestra la probabilidad de pertenencia de cada descriptor a los distintos clústeres, señalando aquel clúster cuya probabilidad de pertenencia es mayor. Estas probabilidades se han tenido en cuenta para determinar si algún descriptor está asociado a más de un grupo, en caso de que la mayor probabilidad de pertenencia a un clúster sea cercana a la de pertenencia a otro u otros clústeres. Como criterio orientativo para establecer si un descriptor pertenece a dos o más clústeres, se ha considerado que la probabilidad de pertenencia de un clúster secundario es al menos la mitad que la probabilidad de pertenencia del clúster principal.

En la tabla 6 se representan las agrupaciones obtenidas. Los descriptores se han ordenado de mayor a menor probabilidad de pertenencia al clúster. Se indica con color amarillo aquellos descriptores clasificados únicamente en ese clúster. En color naranja, aquellos descriptores que están en más de un clúster y ese clúster es el que mayor probabilidad de pertenencia tiene. Finalmente, en color azul se indican aquellos descriptores que están en más de un clúster y ese clúster no es el de mayor probabilidad de pertenencia.

El primer clúster contiene descriptores asociados a olores fétidos. Coincide con la dimensión 4 obtenida por Castro *et al.* (2013) y con la mayoría de descriptores incluidos en el factor 1 según Jeltema y Southwick, llamado “animal-fétido”.

El segundo clúster contiene descriptores agradables de perfumes femeninos, y coincide con la dimensión 1 de Castro *et al.* (2013), excepto porque no aparecen los descriptores “rosa” ni “violetas”. También se parece al factor 7 denominado “floral” obtenido por Jeltema y Southwick (1986), el cual incluye también los descriptores “violeta” y “rosa”.

El tercer clúster es de interpretación un poco dudosa, pues la mayoría de los descriptores están asociados a comida no dulce ni frutal, pero hay algunos que no guardan relación con comida, estando relacionados en su mayoría con olores animales desagradables (orina, orina de gato y leche agria). La razón es que el análisis se ha llevado a cabo sólo con 4 componentes, por lo que cierta parte de la información no se ha tenido en cuenta. No obstante, si en lugar de 4 componentes se consideran más, lo cual sí que se ha comprobado, resulta que el número de clústeres es mayor pero su interpretación es algo más confusa. Estos resultados no se muestran debido a que ocupan un tamaño considerable y no aportan información relevante. Es decir, aumentando el número de componentes no se llega necesariamente a una clasificación más clara de los descriptores, debido a que estas componentes adicionales hacen que las componentes principales más importantes pierdan relevancia al clasificar.

El cuarto clúster agrupa descriptores de perfumes masculinos, y está asociado a la segunda componente de Castro *et al.* (2013). En la dimensión de estos autores aparecen descriptores asociados a la perfumería como “fragante”, “ligero” y “aromático”, pero en este clúster no se incluyen. Esto se explica porque la dimensión de Castro *et al.* (2013) no es un análisis clúster, por lo que los descriptores no pertenecen a un solo grupo. Sin embargo, en el presente análisis clúster realizado, estos descriptores asociados a la perfumería femenina encajan muy bien en el segundo clúster, por lo que solo se ubican en un solo grupo, a pesar de ser descriptores también empleados en perfumes masculinos pero con mucha menor relevancia que en los femeninos.

El quinto clúster hace referencia a descriptores de olor cálido, destacando en el grupo los olores asociados a quemado y olores balsámicos tipo caramelo. Contiene los factores 6 (“brown”) y 9 (“quemado”) obtenidos por Jeltema y Southwick (1986).

El sexto clúster contiene descriptores de olores químicos. Encaja con la dimensión 5 de Castro *et al.* (2013) y el factor 2 denominado “disolvente” según Jeltema y Southwick (1986), salvo por los descriptores asociados a olores canforáceos (“alcanfor”, “cool”, “eucaliptus”), pero estos descriptores son los que menor probabilidad de pertenencia tienen del grupo y aparecen también en otros grupos.

Tabla 6: Clasificación de los descriptores mediante análisis de agrupamiento difuso

Cluster 1	Cluster 2	Cluster 3	Cluster 4
sweaty	perfumery	oily, fatty	bark, birch bark
stale	sweet	fishy	fresh green vegetables
putrid, foul, decayed	lavender	nutty	musty, earthy, moldy
sour, vinegar	floral	kippery (smoked fish)	wet paper
fecal (like manure)	cologne	eggy (fresh eggs)	dry, powdery
blood, raw meat	light	soupy	crushed weeds
cadaverous (dead animal)	fragrant	peanut butter	hay
animal	aromatic	urine	cardboard
wet wool, wet dog	herbal, green, cut grass	cat urine	chalky
yeasty	laurel leaves	fried chicken	crushed grass
dirty linen	cool, cooling	sauerkraut	cork
mouse	geranium leaves	stale tobacco smoke	beany
rancid		sooty	raw potato
sewer		seminal, sperm-like	geranium leaves
sickening		cooked vegetables	rope
fermented (rotten) fruit		garlic, onion	woody, resinous
rope		sour milk	raw potato
sour milk		buttery, fresh butter	herbal, green, cut grass
		popcorn	grainy (as grain)
		celery	laurel leaves
		malty	
		sulfidic	
		grainy (as grain)	
		mushroom	
		fermented (rotten) fruit	
		black pepper	
		burnt candle	

Cluster 5	Cluster 6	Cluster 7	Cluster 8
burnt rubber	medicinal	anise (licorice)	pear
burnt paper	disinfectant, carbolic	cinnamon	cherry (berry)
coffee	gasoline, solvent	banana	grapefruit
burnt milk	varnish	clove	lemon
maple syrup	paint	raisins	honey
molasses	nail polish remover	caraway	soapy
bakery (fresh bread)	kerosene	fresh tobacco smoke	orange
caramel	creosote	coconut	fruity, citrus
household gas	etherish, anaesthetic	almond	cantaloupe, honey dew melon
burnt, smoky	chemical	metallic	rose
seasoning (for meat)	cleaning fluid	spicy	strawberry
warm	turpentine (pine oil)	vanilla	oak wood, cognac
heavy	tar	mothballs	fruity, other than citrus
bitter	new rubber	chocolate	musk
sharp, pungent, acid	camphor	black pepper	tea leaves
meaty (cooked, good)	cool, cooling	minty, peppermint	violets
ammonia	eucaliptus	burnt candle	woody, resinous
sulfidic	leather	beery	peach (fruit)
malty	alcoholic	leather	grape juice
popcorn		eucaliptus	cedarwood
buttery, fresh butter		alcoholic	raw cucumber
garlic, onion		dill	dill
		pineapple	green pepper
		cool, cooling	pineapple
		celery	incense
		camphor	mushroom
		ammonia	beery
			cool, cooling

El séptimo clúster es el más difícil de interpretar, ya que aparecen olores asociados a especias (“clavo”, “canela”), a olores canforáceos (“bolas antipolillas de alcanfor”, “menta”, “eucaliptus”, “cool”, “alcanfor”), a balsámicos dulces (“vainilla”, “chocolate”, “anís”), y otros descriptores. El clúster se corresponde en cierta medida con los factores 8 (cool-menta), 13 (especiado) y 14 (coco-almendra) de Jeltema y Southwick (1986).

El octavo clúster contiene descriptores de frutas cítricas y no cítricas, fundamentalmente. Se corresponde con los factores 3 (frutas no cítricas) y 10 (cítricas) de Jeltema y Southwick (1986), siendo los descriptores de estos factores los que mayor probabilidad tienen de pertenencia al clúster.

4.2.2 Análisis de partición iterativa con el método K-medias

Se ha realizado la partición iterativa (clasificación) de los descriptores con el método K-medias para las dos matrices obtenidas de los dos pretratamientos de la matriz original (descritos en el punto 3.5). Tras comparar las agrupaciones obtenidas con los estudios de Castro *et al.* (2013) y de Jeltema y Southwick (1986), se ha comprobado que el pretratamiento de normalización del rango de los descriptores y centrado (transponiendo la matriz posteriormente) es el que ofrece agrupaciones más similares a estos estudios.

Se ha ejecutado esta partición iterativa con el método K-medias empleando el programa R, indicando un número de clústeres entre 2 y 50. La figura 8 muestra la suma de cuadrados total dentro de cada clúster en función del número de clústeres:

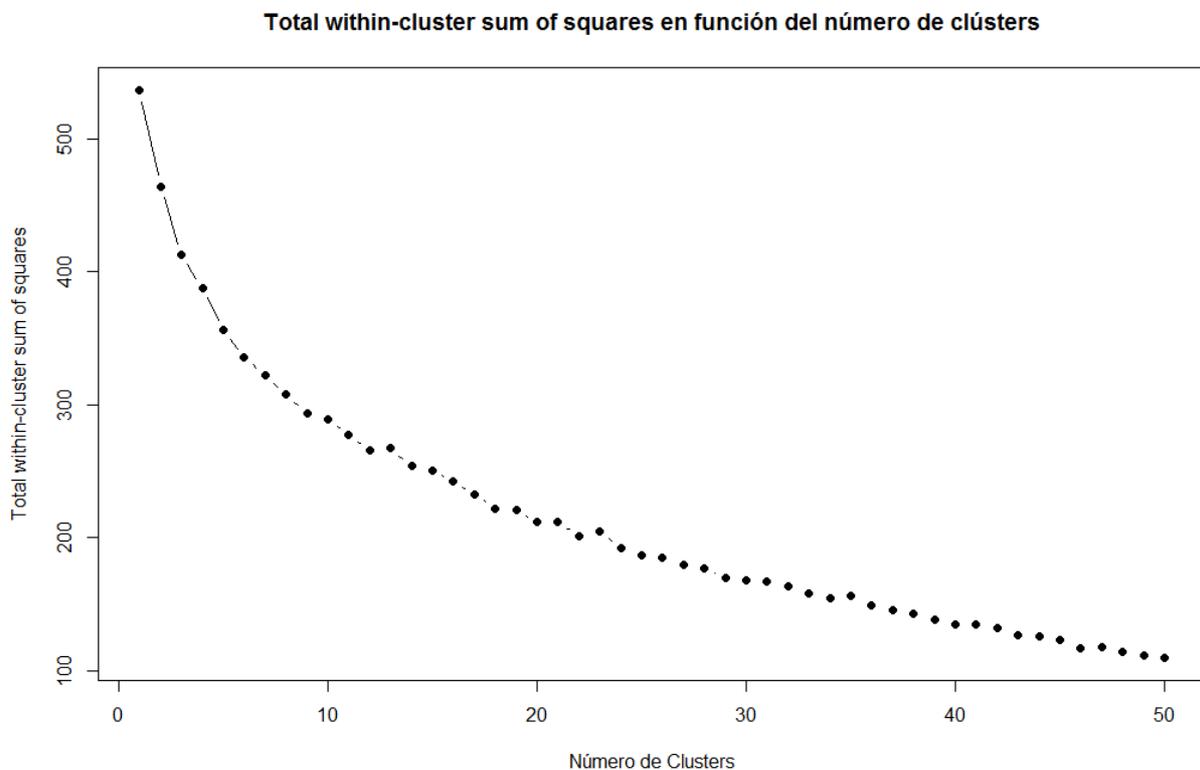


Figura 8: suma de cuadrados total dentro de cada clúster en función del número de clústeres obtenido con el análisis de partición iterativa (método K-medias).

Para elegir el número óptimo de clústeres, una regla práctica habitualmente empleada consiste en utilizar aquel en el que se produce un “codo” en la gráfica, aunque en este caso no se produce ninguno con claridad. Para elegir el número de particiones, se procede a analizar las agrupaciones obtenidas para un distinto número de clústeres, comparándolos con los estudios de Castro *et al.* (2013) y de Jeltema y Southwick (1986). Tras analizar las diferentes clasificaciones de los descriptores, se obtiene que la mejor partición es la obtenida con 37 clústeres.

El primer método aplicado que daba lugar a 8 clústeres pretendía obtener grupos más genéricos. En este caso el objetivo es algo distinto: obtener de alguna forma el máximo número de agrupaciones de descriptores que se pudiera obtener. Este valor podría interpretarse de alguna forma como el máximo número de clústeres olfativos que se puede dar. Esto justifica la elección del método de partición iterativa con el método K-medias, ya que, a diferencia de los métodos jerárquicos, en esta técnica es necesario especificar *a priori* el número de clústeres a formar.

En relación al número de clústeres en los que se pueden clasificar los distintos descriptores olfativos, este tema fue revisado por Zarzo (2008), comparando los diferentes números de grupos obtenidos en diferentes estudios. En la tabla 7 se muestra un resumen.

Tabla 7: Número de categorías olfativas descritas por diversos estudios, a partir del análisis estadístico de descripciones olfativas semánticas (ver Tabla 1 de Zarzo, 2008).

Número muestras	Número descriptores	Fuente de las descripciones olfativas	Resultado del análisis de los datos	Referencia
2467	60	Arctander (1969)	27 clústeres	Chastrette <i>et al.</i> (1988)
1573	126	Arctander (1969)	19 clústeres	Abe <i>et al.</i> (1990)
628	32	Un equipo de 7 perfumistas	10 clústeres	Chastrette <i>et al.</i> (1991)
1396	135	Recopilado de la literatura	Modelo descriptivo de la percepción olfativa	Jaubert <i>et al.</i> (1995)
851	278	Sigma-Aldrich (1996)	32 dimensiones	Madany-Mamlouk y Martinetz (2004)
881	82	Sigma-Aldrich (2003)	17 clases de descriptores	Zarzo y Stanton (2006)

El valor mínimo de esta tabla es 10, el cual es similar a 8 clústeres obtenidos en el presente trabajo. El valor máximo es 32, lo cual de nuevo es similar a los 37 clústeres obtenidos en el análisis de partición iterativa con el método K-medias. Por tanto, tiene mucho interés la comparación de los resultados, es decir, ver en qué medida las agrupaciones obtenidas con un número bajo de clústeres se corresponden con las agrupaciones resultantes con un número de clústeres alto.

En la tabla 8 se muestran los distintos clústeres formados, y el nombre asignado a cada uno de ellos para facilitar la interpretación de los resultados.

Tabla 8: Clasificación de los descriptores obtenida mediante análisis de partición iterativa con el método K-medias

Clúster	Descriptores	Clúster	Descriptores	Clúster	Descriptores
Químico medicinal	etherish, anaesthetic disinfectant, carbolic medicinal	Cartón	cardboard wet paper cork	A cerveza	beery
Químico disolvente	cleaning fluid gasoline, solvent chemical kerosene paint varnish	Floral	floral rose violets lavender	Balsámico caramelo	maple syrup caramel molasses celery coffee
Químico empireumático	leather tar creosote new rubber	Amaderado	oak wood, cognac woody, resinous cedarwood bark, birch bark	Balsámico vainilla	honey almond chocolate vanilla malty warm
Químico alcohol	nail polish remover alcoholic	Frutal	fruity, other than citrus pear peach (fruit)	Comida cocinada	soupy seasoning (for meat) black pepper dill sauerkraut meaty (cooked, good) cooked vegetables
Repulsivo	dirty linen rancid sweaty animal sewer putrid, foul, decayed fecal (like manure) cadaverous (dead animal) sickening	Frutal dulce	grape juice strawberry cherry (berry)	Frutal y otros	pineapple cantaloupe banana coconut anise (licorice) mushroom fresh tobacco smoke seminal, sperm-like
Animal	wet wool, wet dog stale mouse	Cítricos	fruity, citrus lemon grapefruit orange	Cuerda	rope
Irritante	bitter sharp, pugent, acid heavy	Herbáceo	laurel leaves geranium leaves herbal, green, cut grass	Pescado carne crudos	eggy (fresh eggs) fishy kippery (smoked fish) blood, raw meat
Agrio	yeasty sour milk fermented (rotten) fruit sour, vinegar	Hojas verdes	green pepper fresh green vegetables crushed weeds crushed grass raw cucumber raw potato beany	Almizcle	musk soapy incense dry, powdery
Quemado	burnt, smoky stale tobacco smoke burnt paper burnt milk sooty burnt candle	Amoniaco	ammonia urine cat urine	Panadería o tostado	grainy (as grain) bakery (fresh bread) oily, fatty buttery, fresh butter popcorn fried chicken
Fragancia	fragrant aromatic sweet	Sulfuroso	burnt rubber household gas sulfidic garlic, onion	Metal	metallic
Perfume	cologne perfumery	Canforáceo Químico	mothballs camphor eucaliptus turpentine (pine oil)	Especiado	spicy clove cinnamon tea leaves raisins
Ligero	light	Canforáceo herbáceo	caraway minty, peppermint cool, cooling		
Terroso	hay musty, earthy, moldy	Frutos secos	nutty peanut butter chalky		

La dimensión 5 de Castro *et al.* (2013) hace referencia a los olores químicos, al igual que el factor 2 llamado “disolvente” de Jeltema y Southwick, y el clúster 6 obtenido en el análisis clúster difuso. Esta agrupación también aparece en la tabla 8, pero se halla dividida en cuatro clústeres, que son “químico-medicinal”, “químico-disolvente”, “químico-empireumático” y “químico-alcohol”. Los cuatro grupos formados tienen sentido, ya que los descriptores de cada grupo aluden a olores similares. El término

empírico alude a todos los olores de sustancias sometidas a fuego, tal como olor a tostado, ahumado, quemado, etc. El alquitrán (*tar*) lógicamente tiene un olor a humo. La creosota es un derivado del alquitrán y también tiene un olor similar. El olor a cuero también se describe frecuentemente con notas ahumadas. El descriptor “aguarrás” (aceite de trementina) es el único que no aparece clasificado en ninguno de los cuatro grupos de olores químicos en la tabla anterior, pero sí se clasifica como químico en los análisis de Jeltema y Southwick (1986), de Castro *et al.* (2013) y con el agrupamiento difuso (Tabla 6).

También aparecen en esta clasificación (Tabla 8) los olores desagradables, que se reflejan en la dimensión 4 de Castro *et al.* (2013), en el factor 1 (“animal, sucio”) de Jeltema y Southwick (1986), y en el clúster 1 del análisis agrupamiento difuso (Tabla 6). Estos olores desagradables aparecen divididos en cinco clústeres, que son: “repulsivo”, “animal”, “irritante”, “agrio” y “amoniaco”. El clúster “agrio” parece correcto excepto por el descriptor “levadura”, que no debería aparecer clasificada como un olor desagradable, ya que en la escala hedónica obtuvo una puntuación neutra. El clúster “irritante” parece correcto, ya que de hecho engloba descriptores aplicados a olores con un carácter irritante. Los clústeres “repulsivo” y “animal” deberían haber aparecido en uno solo, ya que algunos términos se refieren a olores relacionados. El clúster “amoniaco” contiene descriptores asociados a la orina y el amoniaco.

En el clúster de olores a quemado aparecen descriptores similares, y equivale al factor 9 denominado “quemado” de Jeltema y Southwick (1986), salvo que en la clasificación de estos autores aparecen también “goma quemada” y “creosota”, pero no se incluye “humo de tabaco” ni “añejo”. Parece adecuado que en el clúster obtenido apareciera “goma quemada”, pero debido a que este olor también tiene una nota sulfurosa, no aparece. El descriptor “creosota” está clasificado correctamente como un olor químico, ya que se trata de un derivado del alquitrán. También es clasificado de manera adecuada el descriptor “humo de tabaco añejo” como un olor a quemado.

Los descriptores de perfumería aparecen en la Tabla 8 en tres clústeres diferentes, que son “fragancia”, “perfume” y “ligero”. Esta clasificación tiene sentido, ya que los descriptores que aluden a olores más parecidos aparecen en el mismo grupo (colonia-perfume y fragante-aromático), y el clúster denominado “fragancia” es el que está más asociado a los olores dulces, lo que justifica la presencia del descriptor “dulce”.

El factor 4 obtenido por Jeltema y Southwick (1986), denominado “verde”, se encuentra repartido en la Tabla 8 en 3 clústeres, que son “terroso”, “hojas verdes” y “herbáceo”. La separación entre los descriptores de los clústeres “verde” y “herbáceo” no parece bien definida del todo, y de hecho se trata de olores relacionados. En general, los olores descritos como “hojas verdes”, son bastante frescos y algo ásperos o penetrantes. El más típico es el olor a césped recién cortado. Por el contrario, se suele describir como herbáceo los olores que recuerdan también las plantas pero son más agradables. Son los olores de las infusiones (manzanilla, tomillo, etc.) o especias como el orégano o el romero.

La separación del descriptor “enmohecido-terroso” respecto a los olores “hojas verdes” parece adecuada, ya que aunque ambos son olores típicos de la naturaleza y de ambientes rústicos, perceptualmente son claramente distintos. Lo que resulta llamativo

es la asociación encontrada entre “heno” y “enmohecido-terroso”, pues estos olores son bastante distintos.

El clúster “cartón” no se ha encontrado en otras clasificaciones, pero parece correcto, ya que tanto el corcho, el papel mojado y el cartón tienen olores similares.

El clúster “floral” coincide con el factor 7 denominado “floral” según Jeltema y Southwick (1986). Estos autores incluyen en dicho clúster los descriptores asociados a la perfumería, lo cual no sucede en este caso y parece tener sentido.

El clúster “amaderado” contiene descriptores asociados a diferentes tipos de madera, y concuerda con el factor 12 llamado “madera” del estudio de Jeltema y Southwick (1986), salvo por el descriptor “corcho”.

La clasificación de los descriptores asociados a olores frutales no es muy satisfactoria. Estos descriptores se han clasificado en tres clústeres: “afrutado”, “afrutado-dulce” y “afrutado y otros”. Los clústeres “afrutado” y “afrutado-dulce” contienen sólo descriptores asociados a frutas, y al parecer la diferencia entre ellos es que uno alude a olores afrutados más dulces que el otro, aunque tendría sentido haber aparecido en un mismo clúster. El clúster “afrutado y otros” es el que parece más erróneo de todos los clústeres en la Tabla 8. De los 8 descriptores que contiene, 5 se refieren a olores afrutados, y tres descriptores no tienen relación entre ellos ni con olores frutales. El motivo es incierto.

El clúster “cítrico” contiene todos los descriptores de olores cítricos, y coincide de manera exacta con el factor 10 denominado “cítrico” según Jeltema y Southwick (1986).

El clúster “sulfuroso”, que también puede llamarse “azufrado”, coincide con el factor 11 llamado “sulfuroso” de Jeltema y Southwick (1986), salvo que en el clúster obtenido en el presente estudio aparece también el descriptor “goma quemada”. Esto es debido a que dicho descriptor tiene tanto notas sulfurosas como ahumadas, y en la clasificación obtenida ha quedado más cerca del clúster “sulfuroso”.

Los olores canforáceos se subdividen en dos clústeres, “canforáceo-químico” y “canforáceo-herbáceo”, y son similares al factor 8 denominado “*cool-minty*” (mentolado-refrescante) según Jeltema y Southwick (1986). La distinción entre estos dos clústeres parece correcta, ya que los olores a comino y menta son herbáceos pero presentan notas canforáceas que producen una sensación refrescante por activar el nervio trigémino. En el factor 8 de Jeltema y Southwick (1986) no aparece el descriptor “aguarrás”, posiblemente porque este descriptor tiene notas químicas, canforáceas y a madera, de modo que en la clasificación de dichos autores ha prevalecido la nota química. Por el contrario, en el presente estudio ha prevalecido la nota canforácea, aunque se aprecia la nota química ya que se ha clasificado en el clúster “canforáceo-químico”, y no como “canforáceo-herbáceo”.

El clúster “frutos secos” contiene los descriptores “nuez” y “mantequilla de cacahuete”, que son olores parecidos, y el término “tiza”, que no encaja en el clúster. La razón puede ser que la tiza no presenta un olor bien definido, siendo casi inodora. En la

clasificación de Jeltema y Southwick (1986), aparecen más descriptores para el factor 5 “nuez”, y no se incluye el descriptor “tiza”.

Los olores balsámicos aparecen en dos clústeres: “balsámico-caramelo” y “balsámico-vainilla”, aunque la diferencia perceptual entre ellos no está bien definida. Coinciden la mayoría de descriptores con el factor 5 de Jeltema y Southwick (1986), curiosamente denominado “marrón”. Este término quizás alude a la caramelización del azúcar, por lo que podría interpretarse como “olor caramelizado”, que tiene sentido. En el factor 5 de Jeltema y Southwick (1986) no aparece “apio” (*celery*), esto se explica porque tiene notas herbáceas. El descriptor “almendra” lógicamente encaja más en el clúster “frutos secos”, pero su olor es más dulce que otros frutos secos. De hecho, el olor a turrón, que encaja perfectamente en este clúster, es almendrado.

El clúster “comida cocinada” contiene descriptores asociados a olores producidos al cocinar. Esta agrupación de descriptores parece coherente, pero curiosamente no aparece un grupo similar en las clasificaciones de Jeltema y Southwick (1986) ni en la de Castro *et al.* (2013).

El clúster “pescado/carne crudos” contiene dos descriptores asociados a pescado, los cuales aparecen en el factor 16 denominado “pescado” según Jeltema y Southwick (1986). Curiosamente, junto a estos descriptores de pescado aparecen otros dos que recuerdan el olor a carne cruda: “sangre/carne cruda” y “huevos frescos”. Aunque el olor a pescado es muy característico pues se asocia a determinados compuestos químicos (aminas), la asociación con olores a carne tiene cierto sentido. No obstante, estos descriptores tienen una baja incidencia en el Atlas de Dravnieks, de modo que la clasificación de estos descriptores es un tanto incierta.

El clúster “almizcle” contiene descriptores con cierta similitud. De hecho, estos olores se suelen describir como “pulverulentos”. La presencia de “jabón” en este clúster puede resultar llamativa, pero no hay que olvidar que la mayoría de jabones domésticos llevan almizcle en la composición de sus esencias. Este clúster no aparece en la agrupación de Jeltema y Southwick (1986) ni en la de Castro *et al.* (2013).

El clúster “panadería/tostado” contiene descriptores que recuerdan a los olores de pan tostado. La presencia de “pollo frito” dentro de este clúster puede resultar llamativo, pero de hecho este producto huele a tostado. No hay un grupo similar en la clasificación de Jeltema y Southwick (1986).

El clúster “especiado” coincide con el factor 13 con el mismo nombre de Jeltema y Southwick (1986), pero estos autores no incluyen en dicho factor los descriptores “hojas de té” ni “pasas”. El olor a hojas de té encaja con este clúster, pues su olor es especiado. El olor a pasas debería haber aparecido junto con los descriptores afrutados dulces. De hecho, Jeltema y Southwick (1986) lo clasifican dentro del grupo de olores afrutados.

Se han obtenido tres clústeres con un solo descriptor, que son “metal”, “cuerda” y “cerveza”. El olor a metal es bastante desagradable pero distinto a otros olores repulsivos, por lo cual tiene sentido que aparezca como un olor independiente. Por el contrario, no está claro que “cuerda” y “cerveza” sean olores tan distintos del resto. El motivo podría ser que ambos descriptores han sido muy poco empleados por el panel de

catadores, y por tanto sus puntuaciones son bastante inciertas a efectos de una correcta clasificación de estos descriptores.

A la vista de los resultados del análisis, en relación a las dos parejas de descriptores con correlación uno, “manzana” – “sopa”, y “queso” – “comino”, se ha obtenido que la primera pareja aparece clasificada en el clúster de comida cocinada. Así pues, el descriptor “sopa” parece ser el correcto, mientras que “manzana” debe ser descartado. De modo similar, respecto a “queso - comino”, aparecen en el clúster canforáceo-herbáceo, lo que sugiere que “comino” es el correcto. Este resultado justifica eliminar “manzana” y “queso” del resto del estudio.

4.3 Estudio las dos dimensiones olfativas más importantes en perfumería

En este apartado se muestran los mapas olfativos obtenidos para cada uno de los pretratamientos y se discuten los resultados obtenidos.

La figura 4 muestra que la cuarta componente principal básicamente discrimina los olores frescos frente a los cálidos. De hecho, diversos estudios sugieren que los olores a hojas verdes son muy frescos, y éstos aparecen con peso elevado en la cuarta componente. Al lado opuesto aparecen los olores especiados, balsámicos, frutos secos... que son cálidos.

Para facilitar la comparación con otros estudios, hay que tener presente que los olores frescos se suelen representar al lado derecho y los cálidos al izquierdo (Zarzo y Stanton, 2009) ya que de este modo se puede interpretar esta dimensión como “frescor”.

En cambio, la dimensión latente que clasifica los olores femeninos frente a los masculinos no queda de manifiesto en el análisis PCA realizado. La componente masculina no ha resultado ser totalmente opuesta a la femenina, pues la figura 3 muestra que los olores femeninos quedan de manifiesto en la primera componente, mientras que los masculinos, en la tercera. Ambas componentes son ortogonales, de modo que en el Atlas de Dravnieks, los catadores han considerado los olores femeninos como distintos de los masculinos, pero no como opuestos. Este resultado causa problemas a la hora de obtener mapas perceptuales comparables con otros estudios, ya que habitualmente aparece una dimensión que discrimina los olores masculinos y femeninos. Así pues, se ha planteado aplicar alguna metodología estadística para conseguir que los descriptores masculinos queden contrapuestos a los femeninos. Con este objetivo, se han planteado dos métodos diferentes, los cuales se exponen a continuación.

Método 1: Combinación de las dimensiones masculina y femenina

1. Se realiza un PCA con todos los descriptores, excepto aquellos asociados a productos químicos, olores desagradables y a comida cocinada, ya que se trata de olores no utilizados en perfumería. Se ha comprobado que de esta forma se

- obtienen mejores resultados con PCA a la hora de caracterizar las dimensiones de perfumería.
2. Se identifica qué componentes son las que están asociadas a las dimensiones de olores femeninos, olores masculinos y cálidos frente a frescos. Se guardan los valores de los *loadings* para estas tres dimensiones, para cada descriptor.
 3. En la componente asociada a los olores femeninos, se comprueba si los *loadings* de los descriptores de olores femeninos son positivos o negativos. En caso de ser positivos, no se modifican; en caso de ser negativos, se cambian de signo. Posteriormente, se eliminan los descriptores que tienen valores negativos.
 4. En la componente asociada a los olores masculinos, se estudia si los *loadings* de los descriptores de olores masculinos son positivos o negativos. En caso de ser negativos, no se modifican; en caso de ser positivos, se cambian de signo. Posteriormente, se eliminan los descriptores que tienen valores positivos.
 5. Se comprueba que ningún descriptor importante ha sido eliminado en el punto 3 ni en el punto 4, ya que si esto ocurre, el descriptor no aparecerá en el mapa. En el caso que esto ocurra, se le otorga un valor nulo de *loading*, ya que se considera que se trata de un descriptor neutro, es decir, ni femenino ni masculino.
 6. Se combina en una sola lista los descriptores y sus *loadings* obtenidos en los puntos 3, 4 y 5. El objetivo es asignar un único *loading* a cada descriptor, con valores negativos para los masculinos y positivos para los femeninos. En caso de que en esta lista el descriptor aparezca una sola vez, la asignación está clara. En caso de que un descriptor aparezca dos veces, con un valor positivo (femenino) y con uno negativo (masculino), se elimina aquel *loading* que tenga un menor valor absoluto de los dos. Por ejemplo, “madera de cedro”, que se utiliza principalmente en fragancias masculinas, y en menor medida en las femeninas, aparece en la lista con los valores 0.015 y -0.101. Nos quedamos con el *loading* negativo puesto que es mayor en valor absoluto, lo cual tiene sentido pues corresponde a un olor masculino.
 7. Se obtiene un gráfico de dispersión, representando en el eje horizontal los valores en la dimensión cálido / fresco de los descriptores, y en el eje vertical, los *loadings* obtenidos según el punto anterior.
 8. Se colorea cada descriptor del gráfico de dispersión del punto 7 en función de la frecuencia de uso del descriptor en la base de datos de Dravnieks. Este mapa se compara con otros similares de la bibliografía (Zarzo y Stanton, 2009). Aquellos descriptores cuya posición no es coherente con la bibliografía y que han sido poco empleados por el panel de catadores se pueden eliminar, debido a que está justificado por el poco uso que se hace del descriptor.
 9. Se representa el mapa de fragancias sin los descriptores considerados como erróneos según el punto anterior.

Método 2: Suma de *loadings* de la componente femenina y masculina

En este caso, la metodología empleada para construir un mapa sensorial de descriptores es similar al método anterior, pero con ciertas variaciones respecto a la estimación del carácter masculino frente al femenino:

1. Se realiza un PCA con todos los descriptores, excepto aquellos asociados a productos químicos, olores desagradables y a comida cocinada, ya que se trata de descriptores asociados a olores que no se utilizan en perfumería. Se ha verificado que con esta selección se obtienen mejores resultados para el PCA en las dimensiones de la perfumería.
2. Se identifican las componentes asociadas a los olores femeninos, masculinos y cálidos frente a frescos. Se guardan los valores de los *loadings* de estas tres dimensiones para cada descriptor.
3. Se comprueba en la componente asociada a los olores femeninos si los *loadings* de los descriptores asociados a olores femeninos son positivos o negativos. En caso de ser positivos, no se modifican; en caso de ser negativos, se cambian de signo.
4. Se comprueba en la componente asociada a los olores masculinos si los *loadings* de los descriptores asociados a olores masculinos son positivos o negativos. En caso de ser negativos, no se modifican; en caso de ser positivos, se cambian de signo.
5. Se suma para cada descriptor los *loadings* obtenidos en el punto 3 y 4. De este modo, se obtienen valores positivos para olores femeninos y valores negativos para los masculinos.
6. Se comprueba en la componente que discrimina los olores cálidos frente a frescos si los descriptores frescos quedan a la derecha (*loadings* positivos) y los cálidos a la izquierda. En caso de no ser así, se cambian de signo para esta componente.
7. Se obtiene un gráfico de dispersión, representando los *loadings* obtenidos en el punto 5 en el eje vertical, y los del punto 6 en el eje horizontal. De este modo se obtiene un mapa sensorial de fragancias para todos los descriptores.
8. Se colorean los descriptores de este mapa en función de la frecuencia de uso en la base de datos de Dravnieks. El mapa se compara con otros similares obtenidos en bibliografía (Zarzo y Stanton, 2009) para identificar qué descriptores poco utilizados están erróneamente posicionados en el mapa de fragancias obtenido, los cuales se pueden eliminar. Esto está justificado por el poco uso que se hace del descriptor.
9. Se representa el mapa de fragancias sin los descriptores considerados erróneos según el punto anterior.

4.3.1 Pretratamiento A: normalización del rango de los descriptores y centrado

a) Método1: Combinación de las dimensiones masculina y femenina

En el gráfico de la figura 9 se muestra el mapa olfativo obtenido para los distintos descriptores. Se han eliminado algunos, según el procedimiento descrito para este método. Adicionalmente, algunos otros descriptores también han sido descartados: los que tienen notas canforáceas (“bolas antipolillas de alcanfor”, “eucaliptus”, “menta”, “cool” y “alcanfor”). El motivo es que estos descriptores aparecían en la zona de los

olores cálidos, lo cual es erróneo, pues estas notas se perciben como refrescantes por activar el nervio trigémino. Está justificado eliminar estos descriptores, ya que tienen que ver con una dimensión olfativa ortogonal. De hecho, los olores canforáceos activan el nervio trigémino y producen una sensación refrescante. Sin embargo, los olores frescos no canforáceos no producen esta sensación.

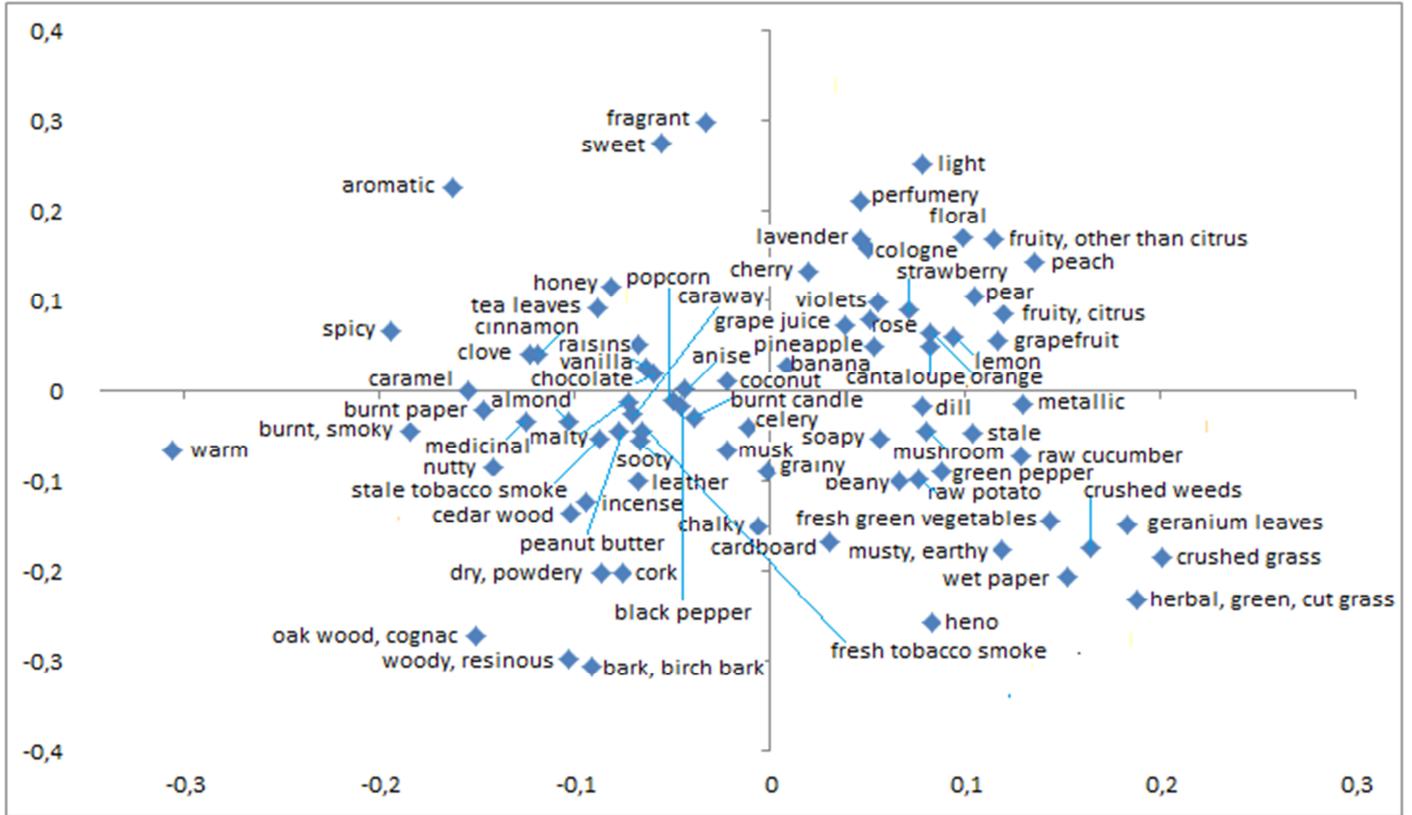


Figura 9: mapa olfativo obtenido con pretratamiento de normalización del rango de los descriptores y centrado (valores del eje vertical calculados por el método 1).

Posteriormente se ha coloreado cada descriptor en el mapa en función de la frecuencia de uso que tienen, a partir de la media de la puntuación del descriptor para las muestras olfativas incluidas en el Atlas de Dravnieks, lo cual se muestra en la figura 10.

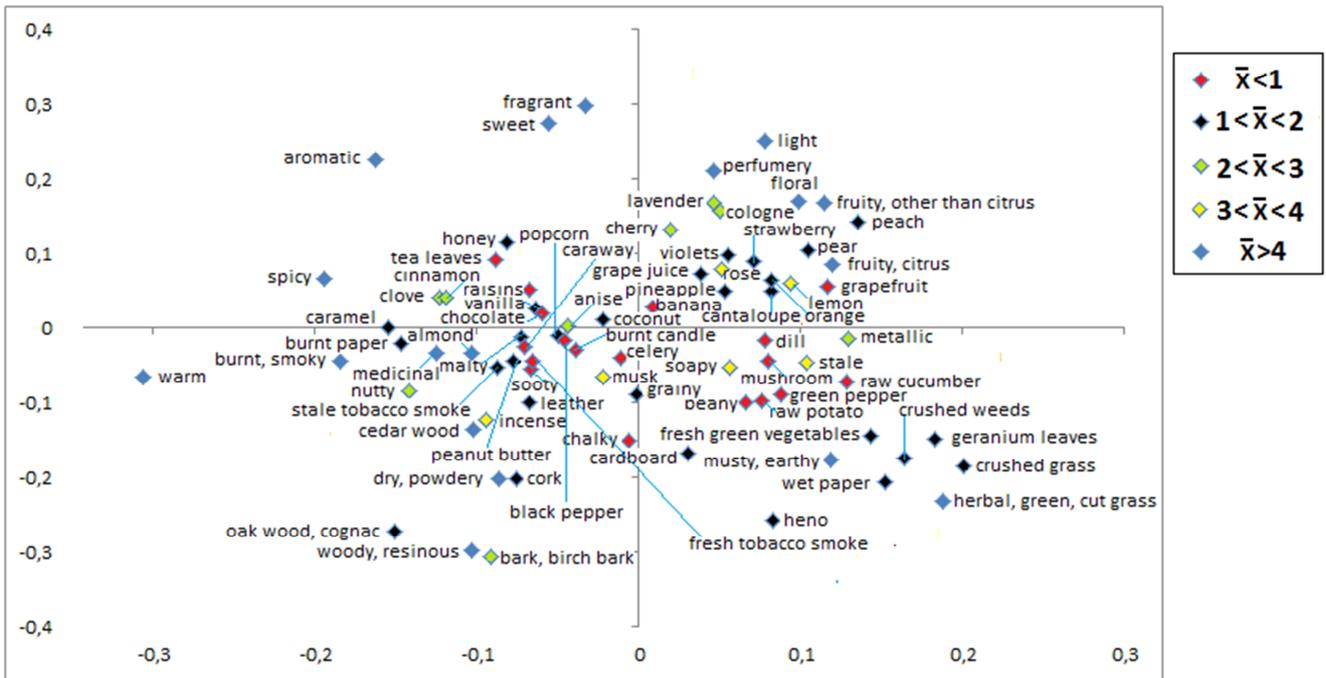


Figura 10: mapa olfativo indicando la frecuencia de uso de los descriptores, con pretratamiento normalización del rango de los descriptores y centrado (valores del eje vertical calculados por el método 1).

En el gráfico anterior (figura 10) se han identificado aquellos descriptores poco utilizados que tienen una posición incorrecta según otros estudios de bibliografía, y que pueden ser eliminarlos. Está justificada su eliminación por el poco uso del descriptor. A continuación, en la figura 11, se señalan mediante un cuadrado en rojo aquellos que han sido seleccionados como incorrectos:

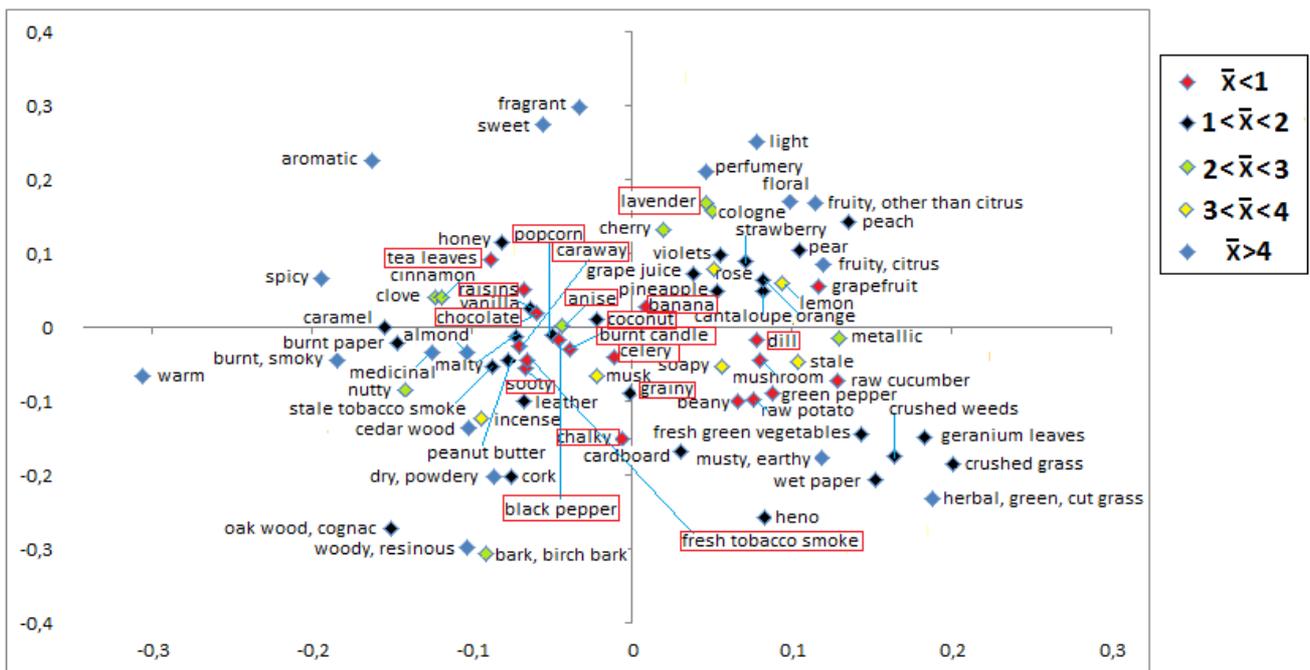


Figura 11: mapa olfativo indicando los descriptores a eliminar, con pretratamiento normalización del rango de los descriptores y centrado.

Tras eliminar los descriptores erróneos de la figura 11, se obtiene el siguiente mapa (figura 12):

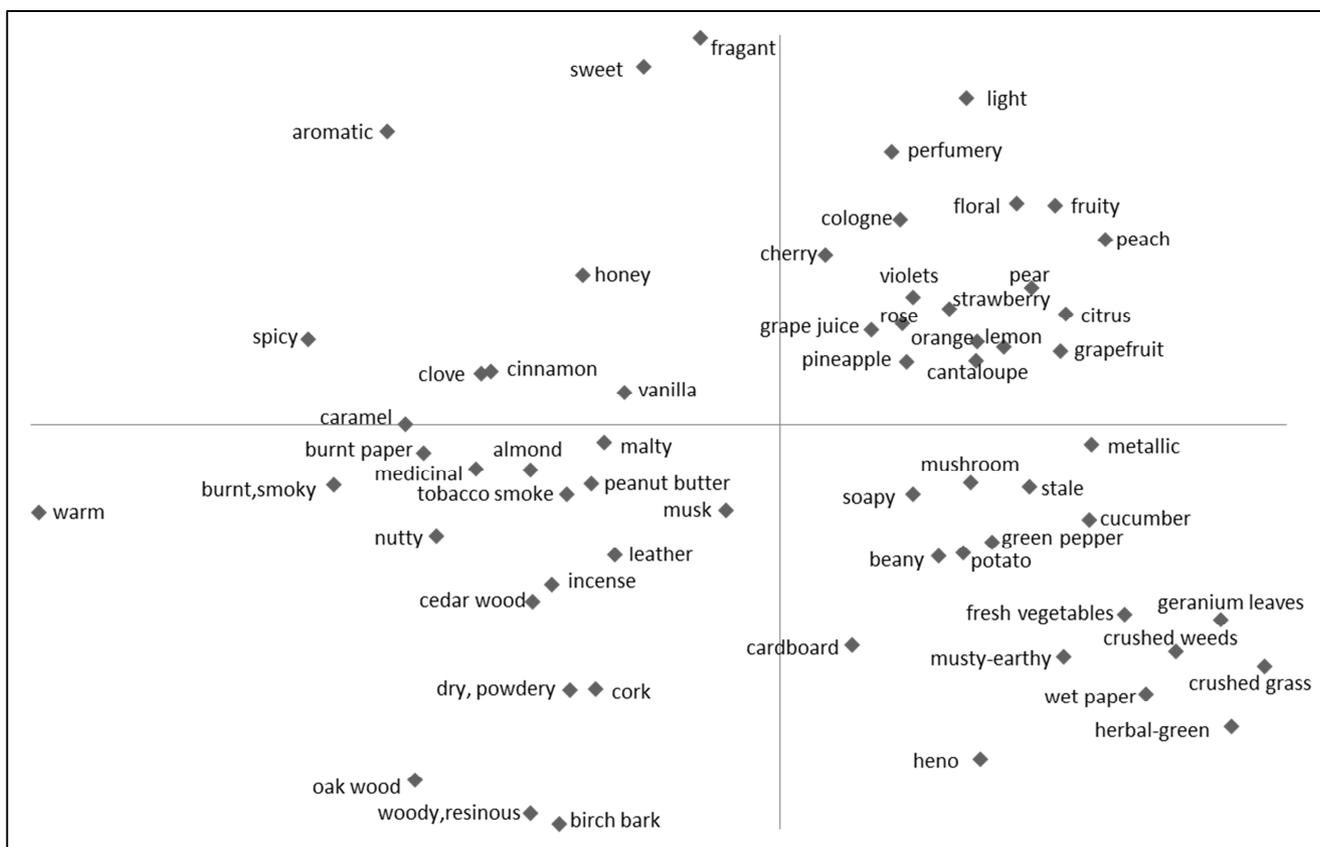


Figura 12: mapa olfativo definitivo con pretratamiento normalización del rango de los descriptores y centrado (valores del eje vertical calculados por el método 1).

El mapa obtenido en la figura 12 es consistente con otros mapas sensoriales (Zarzo y Stanton, 2009, ver figura 25; Zarzo, 2013, ver figura 26). Tras comparar la posición de los descriptores que comparten en común, se ha observado que la mayoría ocupan posiciones similares.

b) Método 2: Suma de *loadings* de la componente femenina y masculina

En el siguiente gráfico (figura 13) se muestra el mapa olfativo para todos los descriptores seleccionados. Como se ha comentado para el método anterior, se han eliminado los descriptores con notas canforáceas (“bolas antipolillas de alcanfor”, “eucalipto”, “menta”, “cool” y “alcanfor”), ya que aparecían como cálidos, lo cual es erróneo pues producen un olor refrescante. Está justificado eliminarlos, ya que la percepción trigémina refrescante asociada a los olores canforáceos, es ortogonal al resto de olores percibidos como frescos.

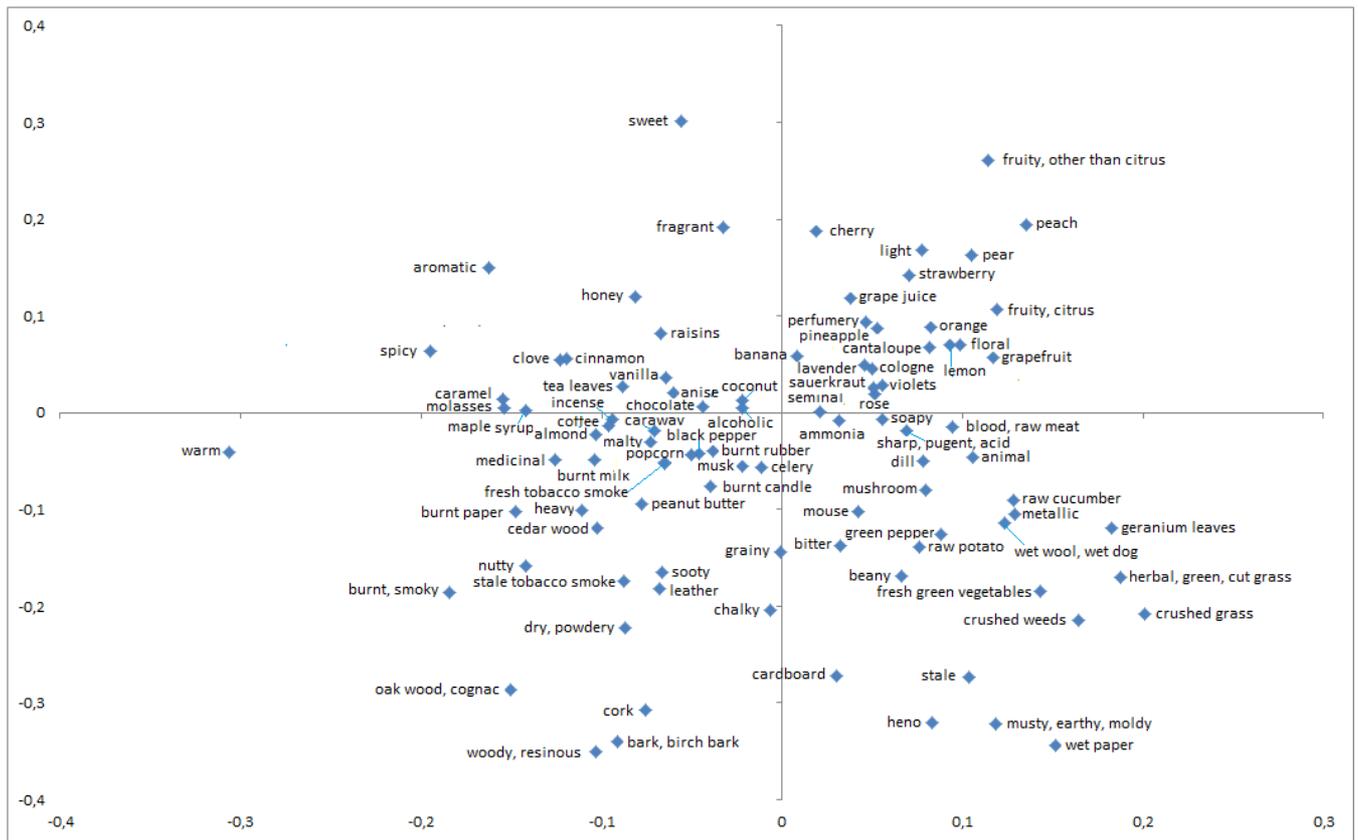


Figura 13: mapa olfativo con pretratamiento normalización del rango de los descriptores y centrado (valores del eje vertical calculados por el método 2).

Posteriormente se ha coloreado cada descriptor en el mapa anterior en función de la frecuencia de uso que tienen, a partir de la media de la puntuación del descriptor para las muestras del estudio, obteniéndose el gráfico de la figura 14.

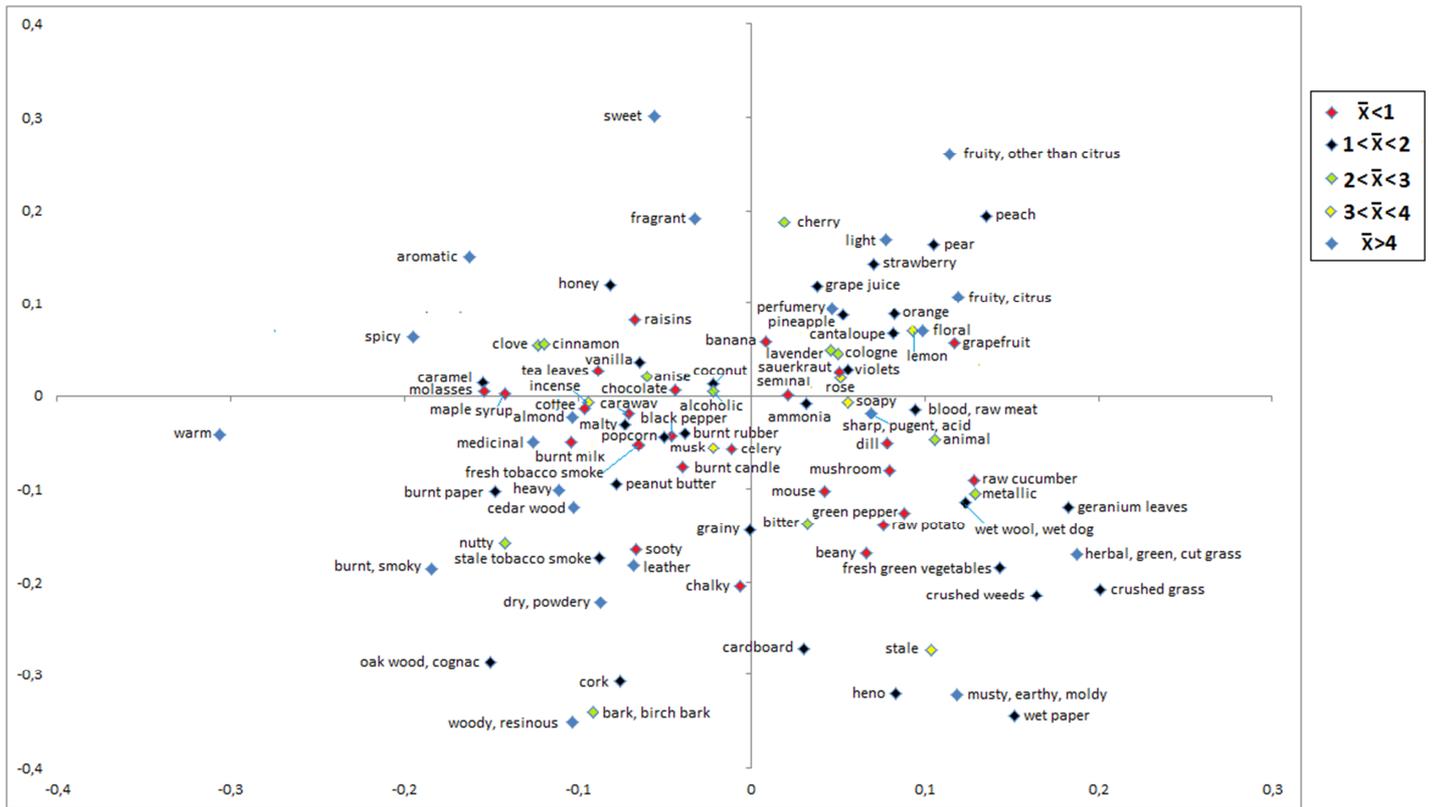


Figura 14: mapa olfativo indicando la frecuencia de uso de los descriptores, con pretratamiento normalización del rango de los descriptores y centrado (valores del eje vertical calculados por el método 2).

En el gráfico anterior (figura 14) se han identificado aquellos descriptores poco utilizados cuya posición no es coherente con otros estudios y que pueden ser eliminados. Está justificada su eliminación por el poco uso del descriptor. En la figura 15 se remarcan en rojo aquellos que han sido seleccionados como incorrectos.

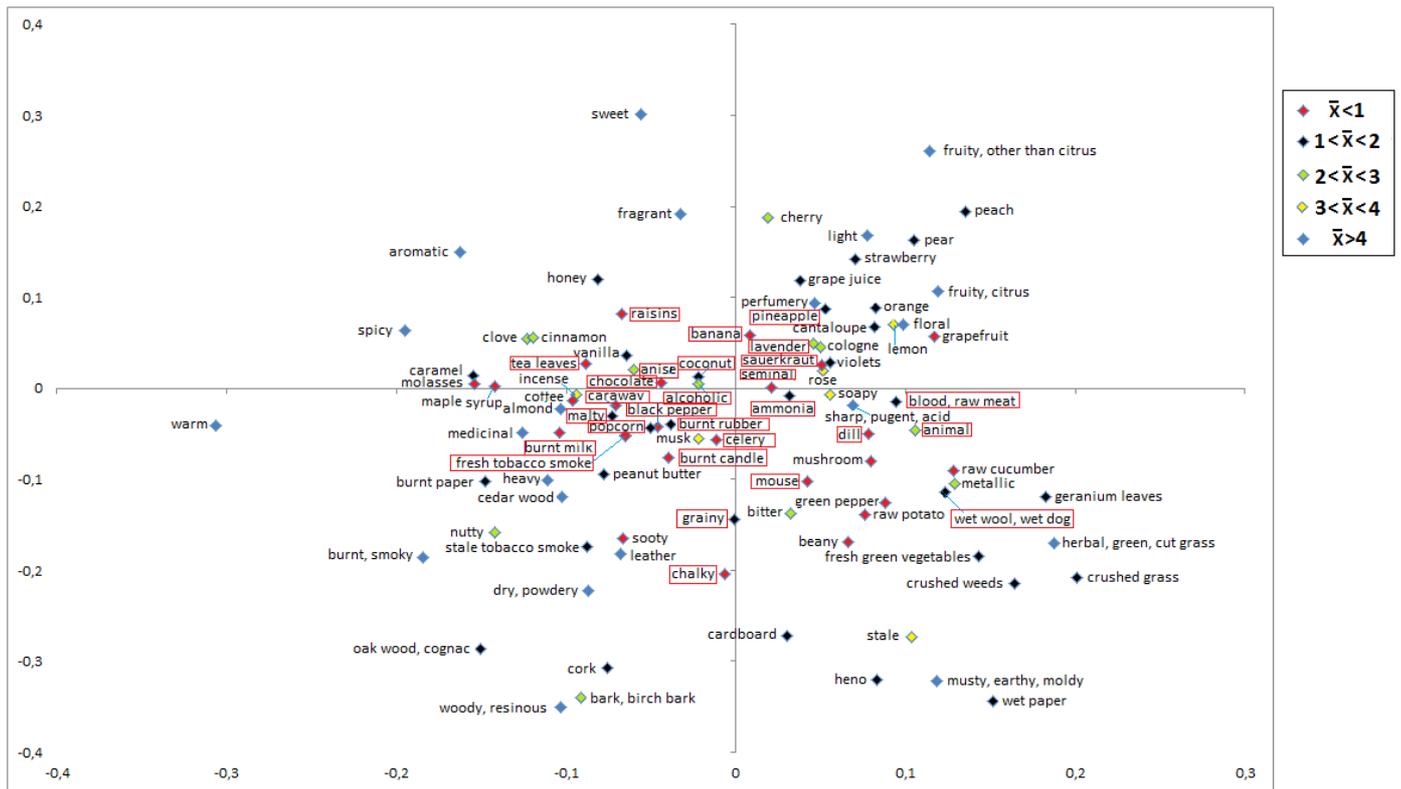


Figura 15: mapa olfativo indicando los descriptores a eliminar, con pretratamiento normalización del rango de los descriptores y centrado (valores del eje vertical calculados por el método 2).

Tras eliminar los descriptores erróneos de la figura 15, se obtiene el siguiente mapa (figura 16):

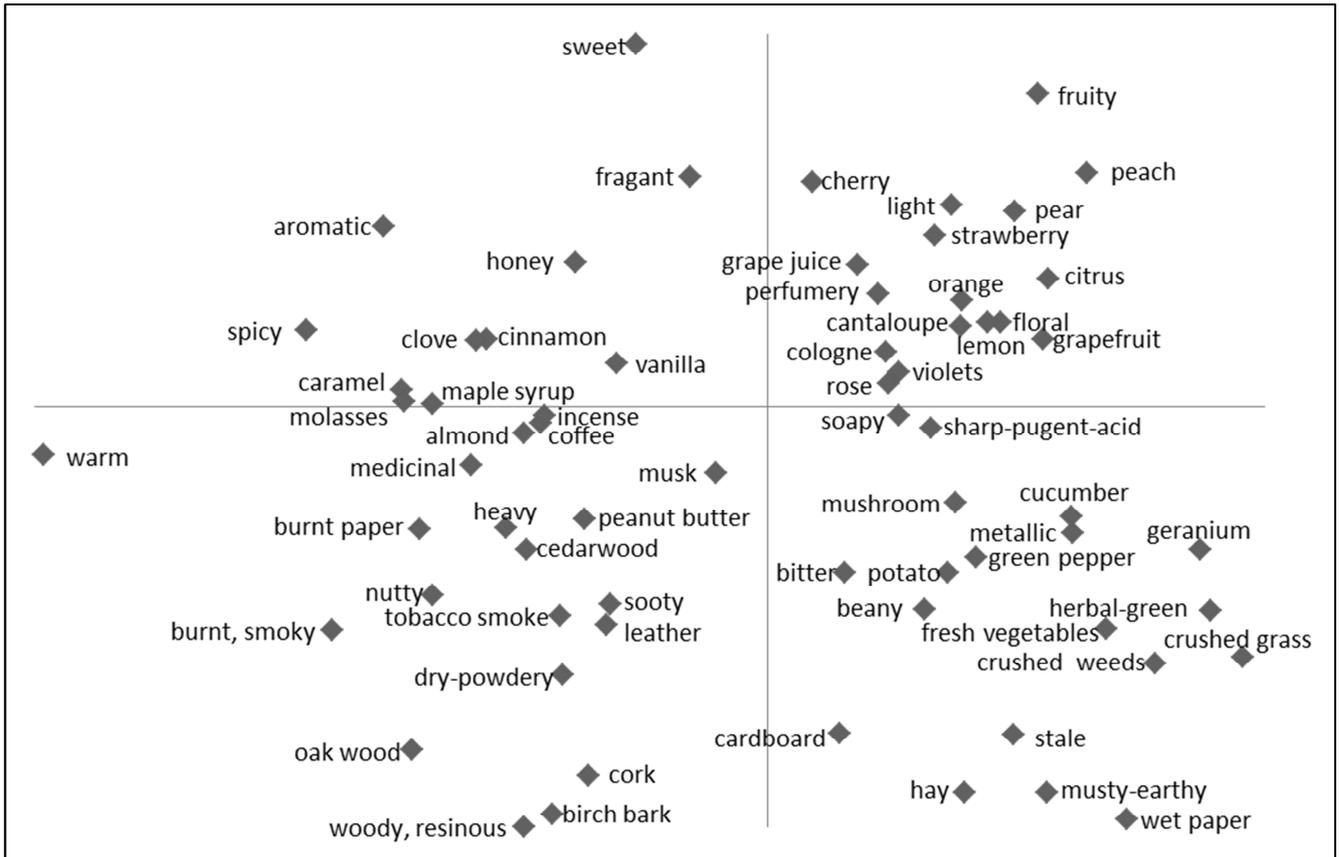


Figura 16: mapa olfativo definitivo con pretratamiento normalización del rango de los descriptores y centrado, valores del eje vertical obtenidos por el método 2.

El mapa obtenido en la figura 16 es consistente con otros mapas sensoriales (Zarzo y Stanton, 2009, ver figura 25; Zarzo, 2013, ver figura 26). De hecho, tras comparar los descriptores que comparten, se ha observado que la mayoría ocupan posiciones similares.

c) Comparación de resultados entre método 1 y método 2

En este apartado se discute qué método de los dos propuestos es el que mejor resultados ofrece. Para ello, se compara la posición de los descriptores en los mapas definitivos obtenidos respecto al eje vertical, que discrimina los olores en la escala femenino/masculino, ya que las coordenadas respecto al eje horizontal (escala fresco/cálido) son las mismas en ambos mapas.

Se observa que los descriptores asociados a notas dulces, pulverulentas, especias, madera, frutos secos, medicinales, afrutadas, cítricas y verdes, se sitúan en una situación similar en ambos mapas.

El método 1 clasifica mejor los descriptores asociados a olores femeninos, ya que aquellos relativos a tonos florales y de perfumería (“fragante”, “ligero” y “perfumería”) son posicionados como más femeninos que en el método 2, puesto que ocupan una posición superior en el eje vertical.

El método 2 clasifica mejor los descriptores asociados a olores masculinos. De hecho, “mohoso-terroso”, “papel mojado”, “añejo”, y los asociados a notas ahumadas, están posicionados como más masculinos que con el método 1, lo cual concuerda mejor con clasificaciones realizadas en otros estudios similares. El descriptor “metálico” es clasificado como masculino en el método 2, mientras que el método 1 lo clasifica como neutro, lo cual parece menos correcto. Este método también clasifica mejor los olores frutales, aparecen como más femeninos que con el otro método. Finalmente también clasifica de manera correcta el descriptor “incienso”, colocándolo como neutro, mientras que el método 2 lo clasifica como bastante masculino, lo que no parece correcto.

En resumen, al parecer el método 2 posiciona más descriptores de forma correcta que el método 1, por lo que se selecciona este método como el más adecuado y se elige este mapa para ser comparado con el otro pretratamiento realizado.

4.3.2 Pretratamiento B: centrado y escalado a varianza unitaria

a) Método 1: Combinación de las dimensiones masculina y femenina

En el gráfico de la figura 17 se muestra el mapa olfativo obtenido a partir de todos los descriptores seleccionados:

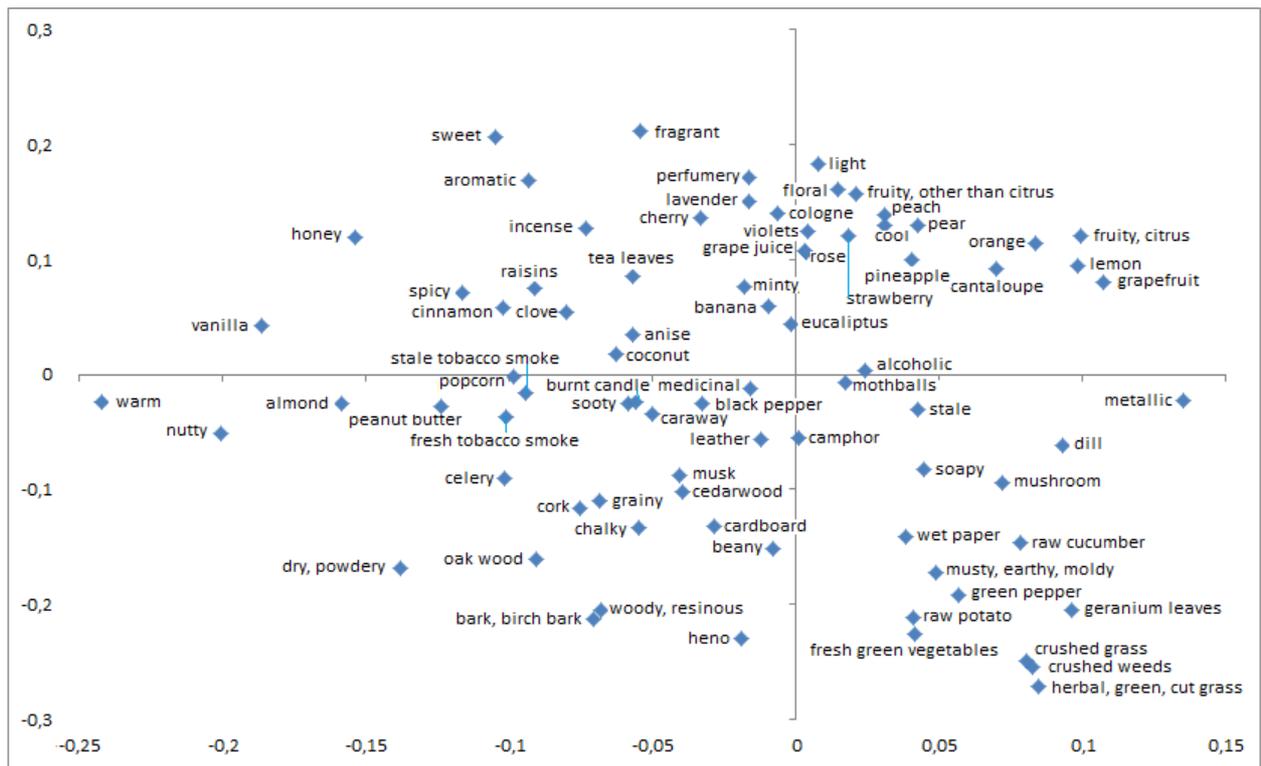


Figura 17: mapa olfativo con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical calculados por el método 1.

Posteriormente se ha coloreado cada descriptor en el mapa en función de la frecuencia de uso que tienen, a partir de la media de la puntuación del descriptor para las muestras del estudio, tal como se indica en la figura 18.

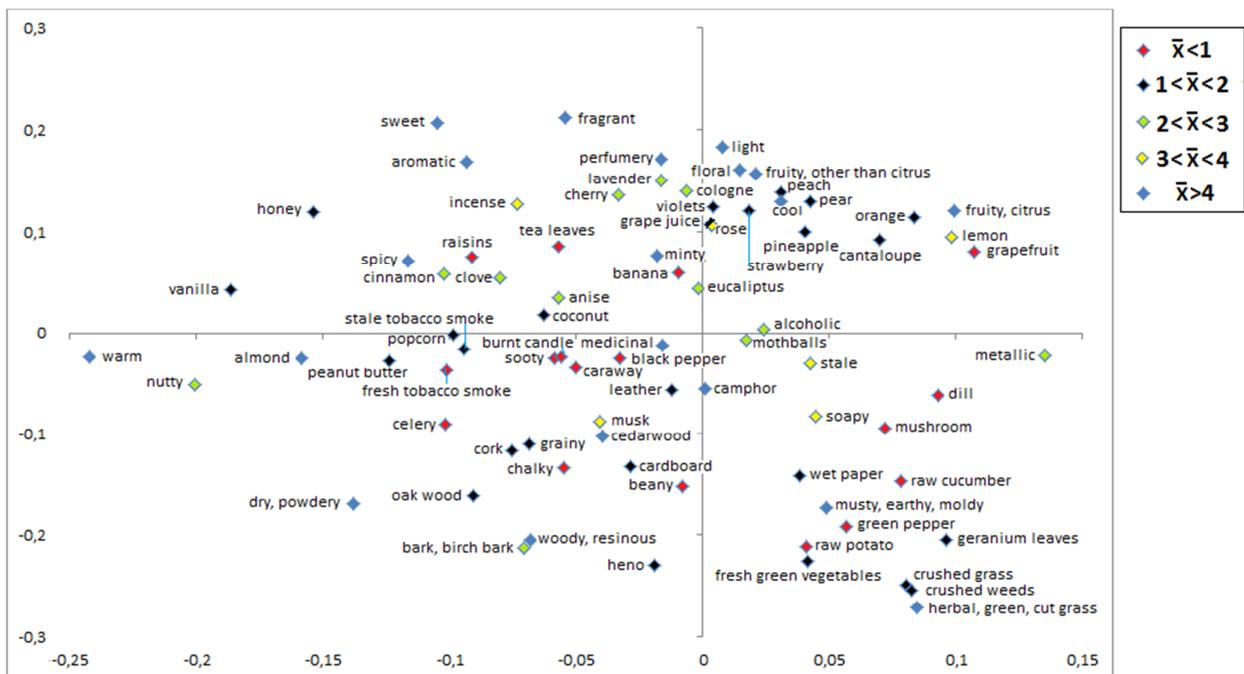


Figura 18: mapa olfativo indicando la frecuencia de uso de los descriptores, con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical obtenidos por el método 1.

En el gráfico de la figura 18 se han identificado aquellos descriptores poco utilizados que tienen una posición incorrecta según otros estudios similares y que, por tanto, pueden ser eliminados. Se trata de términos poco empleados cuya información es poco fiable. En el gráfico de la figura 19 se señalan remarcado aquellos considerados como incorrectos.

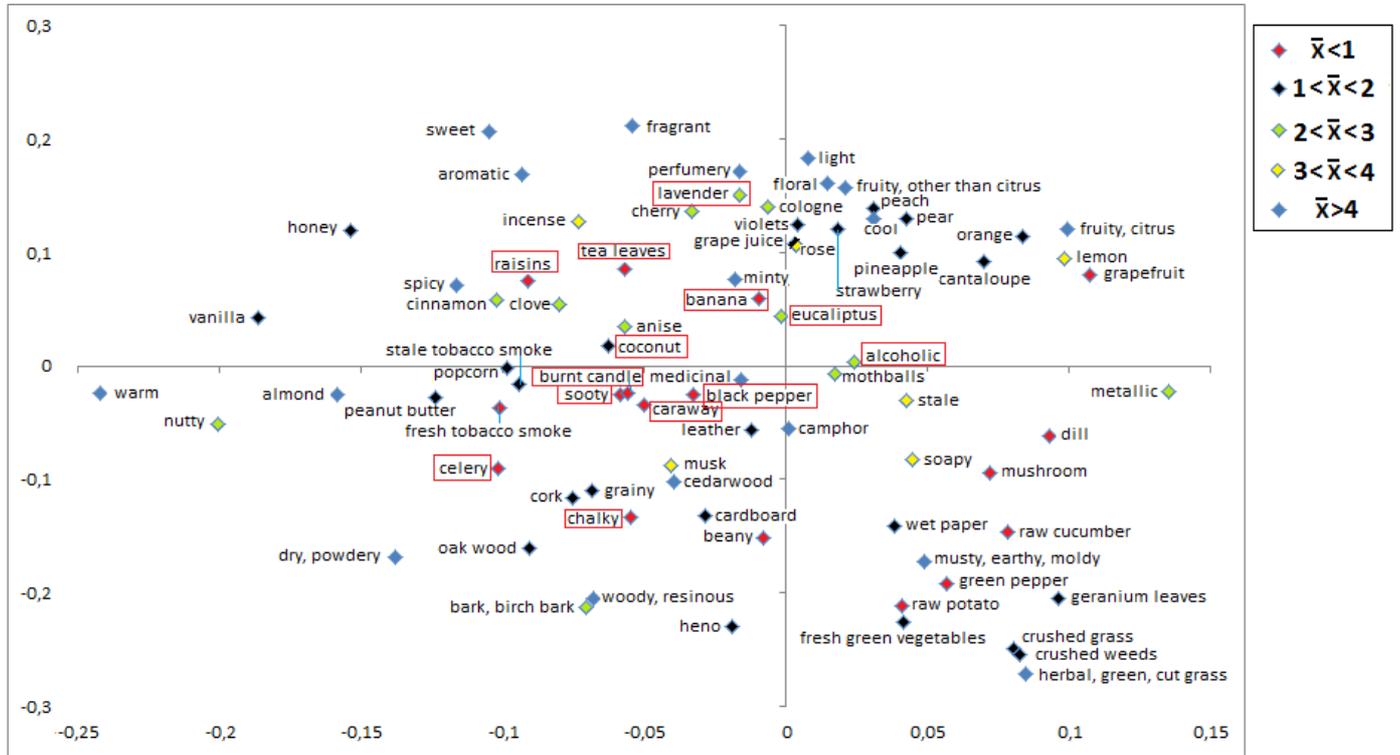


Figura 19: mapa indicando los descriptores a eliminar, con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical calculados por el método 1.

Tras eliminar los descriptores erróneos de la figura 19, se obtiene el siguiente mapa (figura 20):

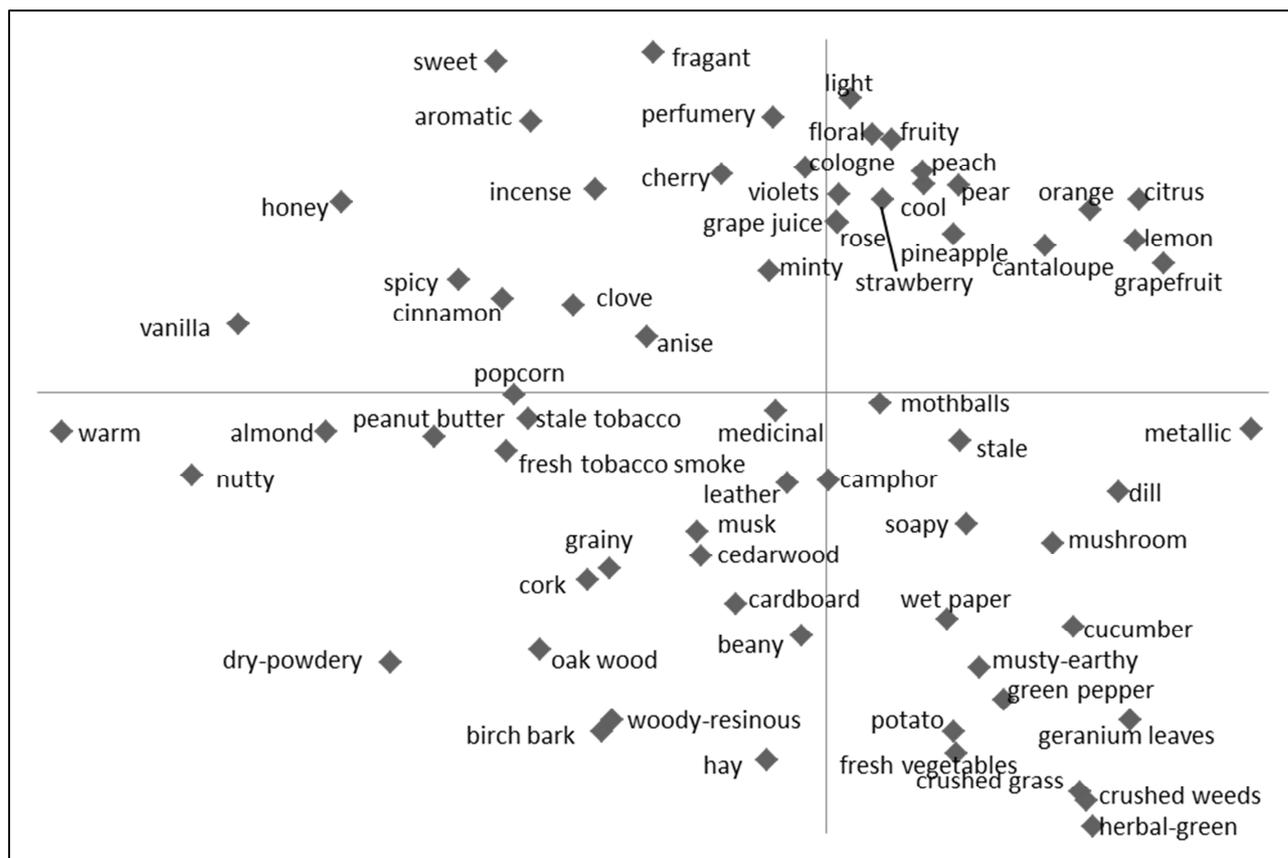


Figura 20: mapa olfativo definitivo con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical calculados por el método 1.

El mapa obtenido en la figura 20 es consistente con otros mapas sensoriales (Zarzo y Stanton, 2009, ver figura 25; Zarzo, 2013, ver figura 26). Así pues, tras comparar los descriptores en común entre dichos mapas, se ha observado que la mayoría ocupan posiciones similares.

b) Método 2: Suma de *loadings* de la componente femenina y masculina

En el gráfico de la figura 21 se muestra el mapa sensorial olfativo para todos los descriptores seleccionados.

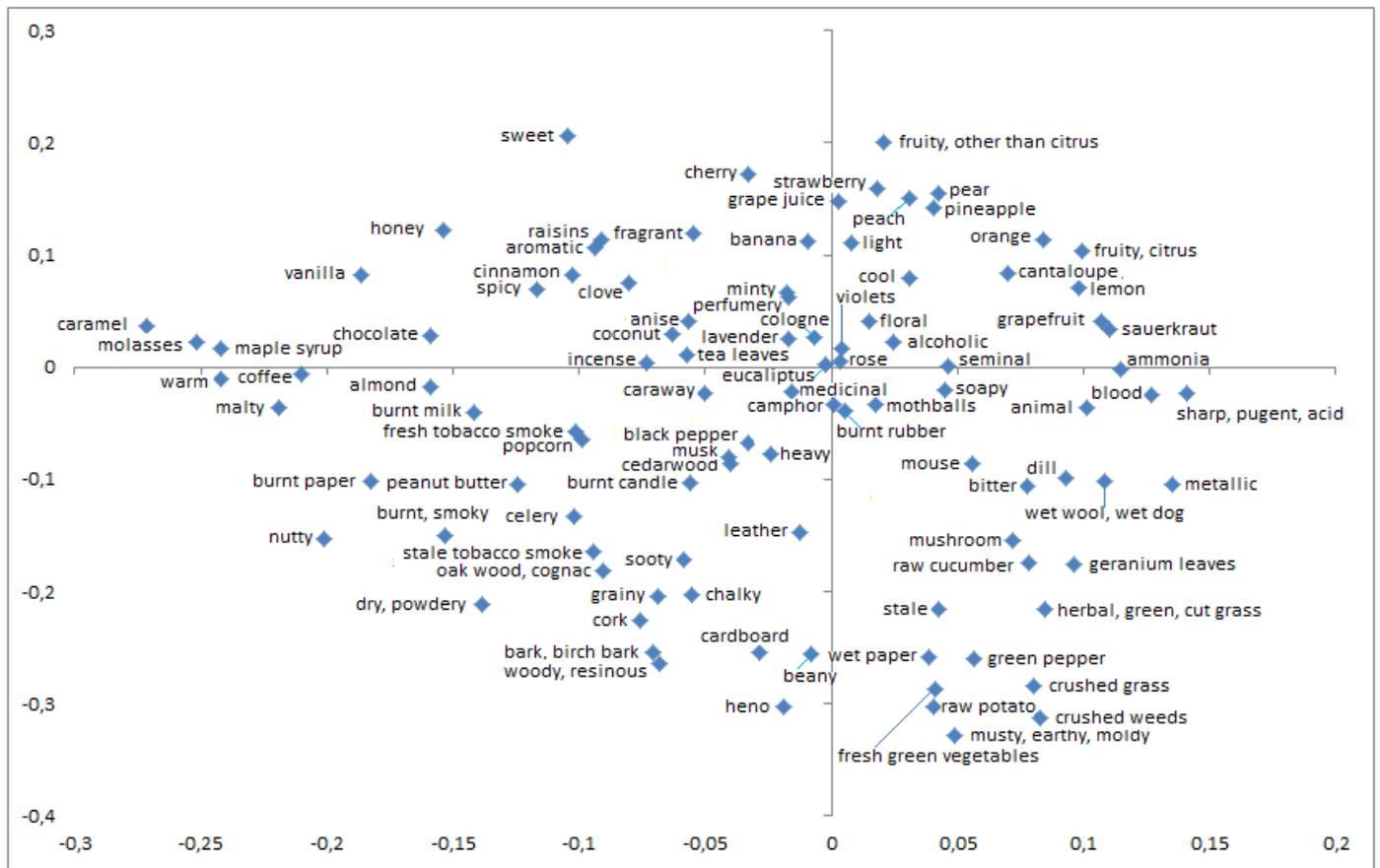


Figura 21: mapa olfativo con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical calculados por el método 2.

Posteriormente se ha coloreado cada descriptor en el mapa en función de la frecuencia de uso que tienen, a partir de la media de la puntuación del descriptor para las muestras olfativas del Atlas de Dravnieks, tal como se muestra en la figura 22.

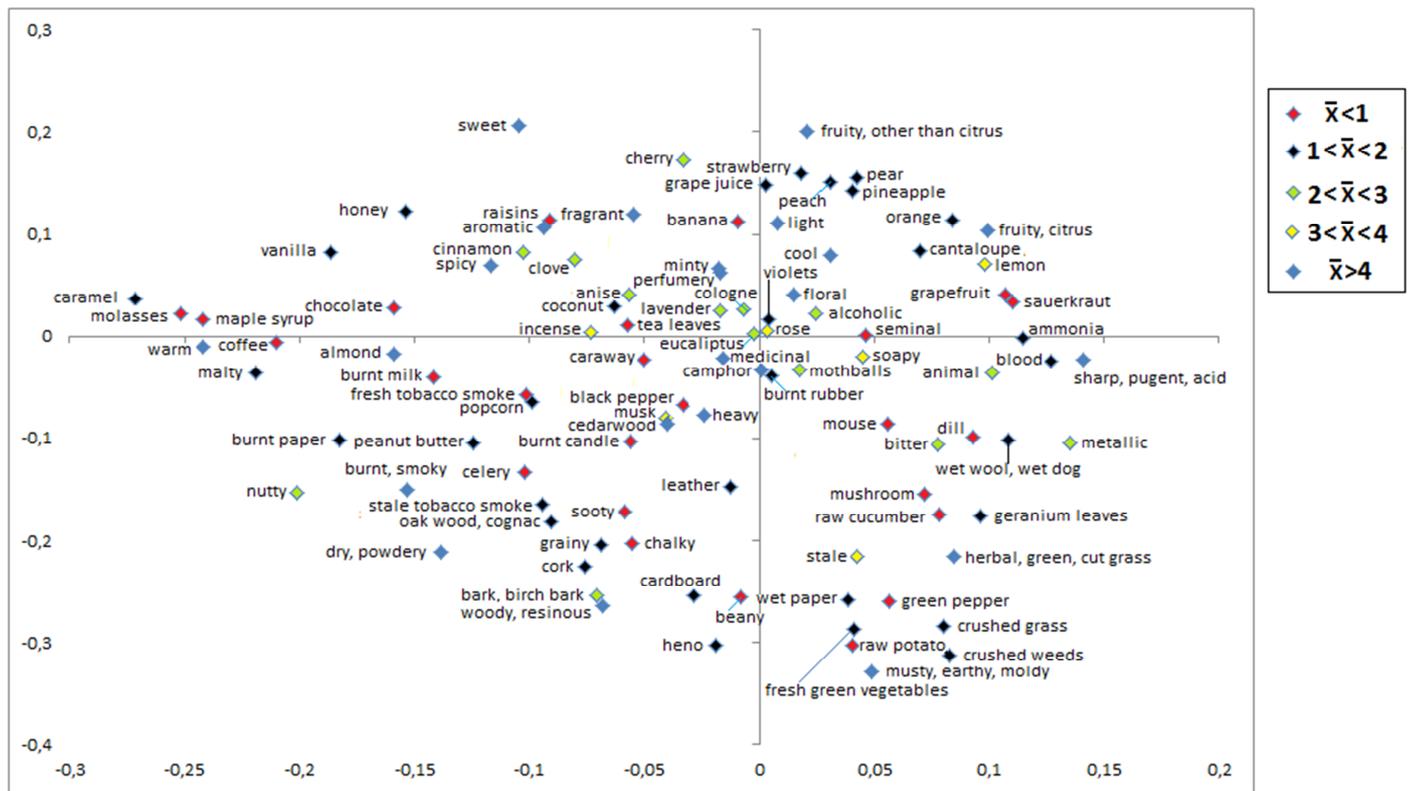


Figura 22: mapa olfativo de la figura 21 indicando la frecuencia de uso de los descriptores, con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical calculados por el método 2.

En el gráfico de la figura 22 se han identificado aquellos descriptores poco utilizados cuya posición resulta incorrecta en comparación con otros estudios similares y que, por tanto, pueden ser descartados. En la figura 23 se señalan mediante un recuadro aquellos que han sido considerados como incorrectos.

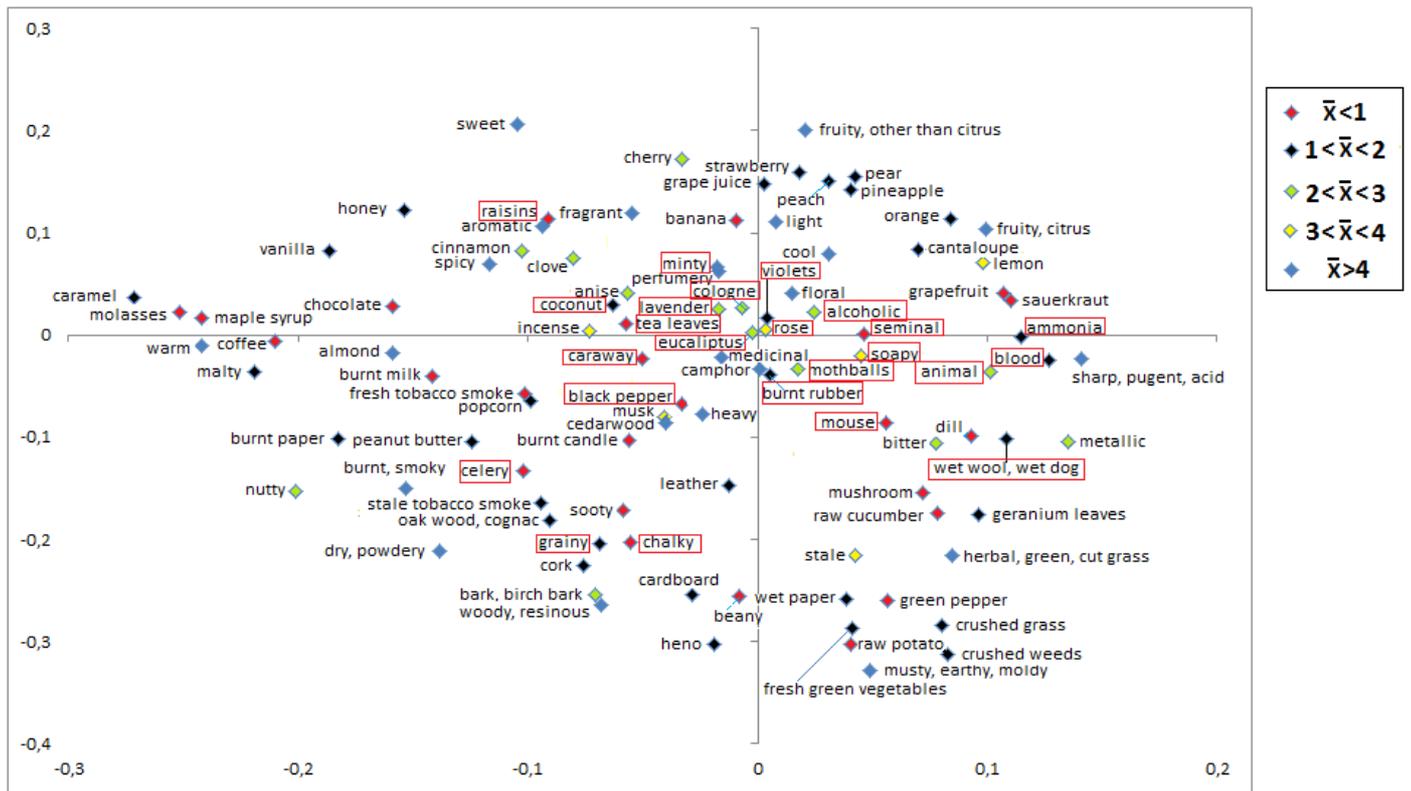


Figura 23: mapa olfativo de la figura 22 indicando los descriptores a eliminar, con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical calculados por el método 2.

Tras eliminar los descriptores erróneos de la figura 23, se obtiene el siguiente mapa (figura 24):

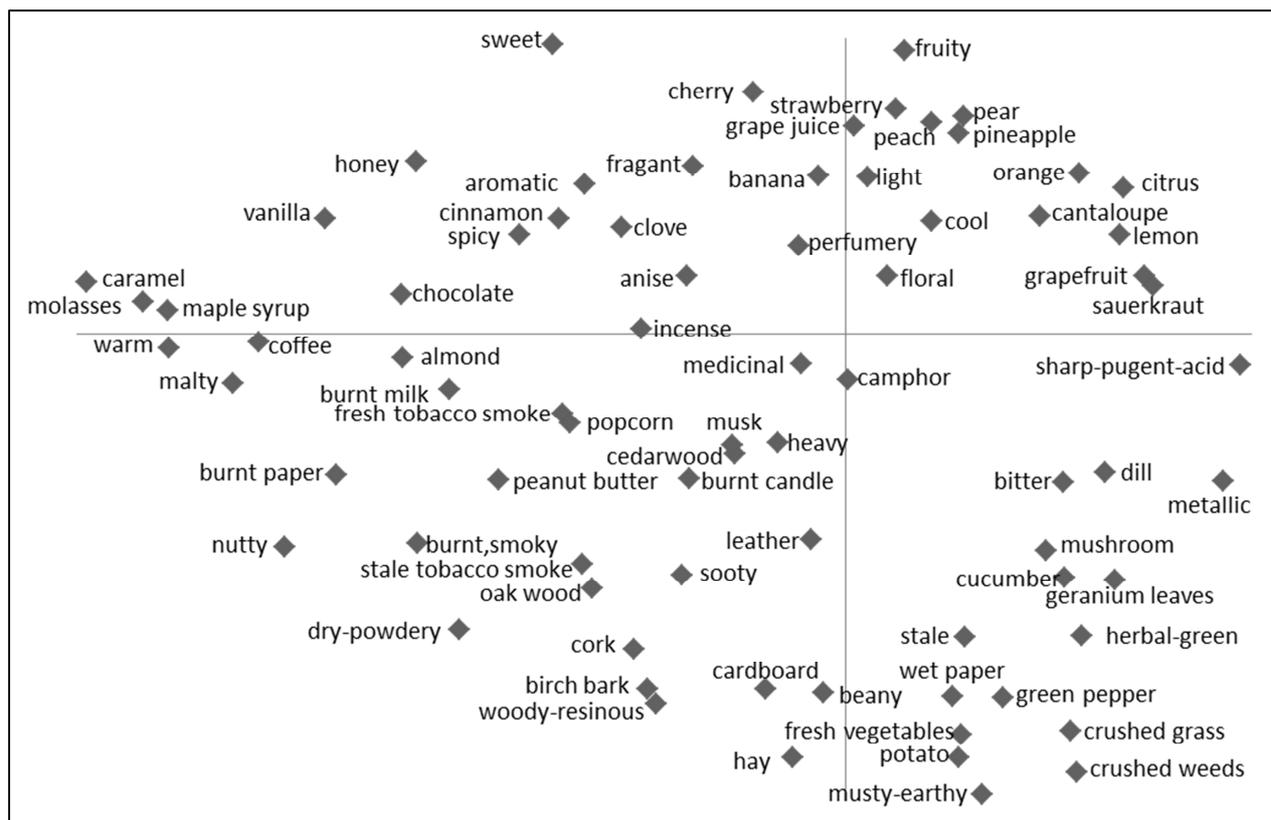


Figura 24: mapa olfativo definitivo con pretratamiento centrado y escalado a varianza unitaria. Valores del eje vertical calculados por el método 2.

El mapa obtenido en la figura 24 es consistente con otros mapas sensoriales (Zarzo y Stanton, 2009; Zarzo, 2013). De hecho, tras comparar los descriptores en común entre estos mapas, se ha observado que la mayoría ocupan posiciones bastante similares.

c) Comparación de resultados con el método 1 y método 2

En este apartado se discute cuál es el método de los dos propuestos que mejores resultados ofrece. Para ello, se compara la posición de los descriptores en los mapas definitivos obtenidos, tanto respecto al eje vertical que clasifica los olores en la escala femenino/masculino, como al eje horizontal (fresco/cálido). En general, los descriptores aparecen en la misma posición en ambos mapas.

A grandes rasgos, se observa que los descriptores relativos a notas pulverulentas, especiadas, amaderadas, medicinales, afrutadas, y de hojas verdes, se sitúan en una situación similar en ambos mapas.

En comparación con otros estudios de la bibliografía, el método 1 clasifica mejor los descriptores asociados a olores florales y de perfumería. Así pues, “fragante”, “ligero”, “cool” y “perfumería” aparecen posicionados como más femeninos que en el método 2, puesto que ocupan una posición superior en el eje y, lo cual es más coherente con el estudio de Zarzo y Stanton (2009).

El método 2 clasifica mejor los descriptores asociados a olores masculinos. Así pues, el descriptor “mohoso-terroso” y los asociados a notas ahumadas, frutos secos y vegetales, están posicionados como más masculinos que con el método 1, lo cual concuerda mejor con otros mapas sensoriales (Zarzo y Stanton, 2009, ver figura 25; Zarzo, 2013, ver figura 26). Los descriptores “metálico” y “añejo” son clasificados como masculinos según el método 2, mientras que el método 1 los clasifica como neutros, lo cual parece menos correcto. Este método también clasifica mejor el descriptor “vainilla”, pues aparece como más femenino que el otro método, estando más cerca de “dulce”. Finalmente, también clasifica de manera más correcta el descriptor “aromático” y los olores cítricos, situándolos como menos femeninos que en el otro método, lo cual es coherente con otros estudios precedentes realizados.

En conclusión, al parecer el método 2 clasifica más descriptores de forma correcta que el método 1, es decir, de forma más consistente en comparación con otros estudios similares publicados en la bibliografía. Por este motivo, se selecciona este método como el más adecuado y se elige este mapa sensorial para compararlo con el otro pretratamiento realizado.

4.3.3 Discusión de los mapas obtenidos y comparación con estudios precedentes

En este apartado se comparan los mapas seleccionados para cada uno de los pretratamientos realizados, estudiando cuál de los dos posiciona los descriptores de manera más adecuada. Posteriormente se compara este mapa sensorial de descriptores con otros similares obtenidos en estudios precedentes de la bibliografía.

Tras comparar los mapas de cada pretratamiento, se ha observado que los descriptores “incienso”, “cálido” y “almizcle”, así como los asociados a notas ahumadas, frutos secos, madera y cítricos, se sitúan en una situación similar en ambos mapas.

A continuación se enumeran los descriptores que han sido posicionados de manera más adecuada para cada uno de los pretratamientos:

Normalización del rango de los descriptores y centrado:

- Notas especiadas: estos descriptores están más cerca de los masculinos que en el otro pretratamiento, lo cual es más correcto.
- Notas a verde y herbáceas: están mejor localizados en la dimensión masculino-femenino, debido a que no aparecen como tan femeninos como en el otro pretratamiento.
- Descriptor “metálico”: se encuentra posicionado como menos fresco en este pretratamiento.
- Descriptor “ligero”: está posicionado como más fresco que en el otro pretratamiento, lo cual es correcto.

Centrado y escalado a varianza unitaria:

- Notas dulces: están mejor posicionadas como cálidas, pues en el otro pretratamiento aparecen demasiado cercanas al centro.

- Descriptor “mohoso-terroso”: está posicionado más cerca del centro en la escala de cálido-fresco, que es su posición correcta.
- Descriptor “medicinal”: está mejor situado en la dimensión cálido-fresco, pues en el otro pretratamiento aparece como demasiado fresco.
- Descriptor “seco-pulverulento”: se halla localizado como más cálido, lo cual es más adecuado.
- Notas vegetales: están posicionadas más hacia el centro en la dimensión fresco / cálido y como olores más masculinos.
- Descriptor “añejo”: está localizado como menos fresco.
- Descriptor “ácido-irritante”: está clasificado como más fresco.
- Notas frutales: los descriptores están posicionados como más neutros en la escala cálido / fresco.
- Notas florales: están más cerca del neutro en la dimensión cálido / fresco.

En conclusión, el pretratamiento “centrado y escalado a varianza unitaria” clasifica más descriptores correctamente que el pretratamiento “normalización del rango de los descriptores y centrado”. Así pues, se selecciona el mapa obtenido por este pretratamiento como el propuesto por el estudio para describir el espacio perceptual olfativo con arreglo a estas dos dimensiones.

El mapa obtenido se ha comparado con dos mapas olfativos sensoriales obtenidos en estudios precedentes:

- El obtenido por Zarzo y Stanton (2009), el cual se muestra en la figura 25. Este mapa fue obtenido al aplicar sobre la base de datos de Boelens y Haring (1981) un análisis PCA. Los descriptores se indican mediante triángulos negros. En este mapa también se muestra con círculos vacíos la clasificación de algunos descriptores propuesta por P. Jellinek en 1951, la cual se elaboró tras una dilatada carrera como perfumista.
- El obtenido por Zarzo (2013), el cual se muestra en la figura 26. Este mapa fue obtenido al aplicar PCA sobre la base de datos de Boelens y Haring (1981), y los descriptores son señalizados mediante triángulos blancos. En esta representación también se muestra con diamantes negros la clasificación de algunos descriptores propuesta por Thiboud (1991) en su mapa de fragancias, obtenido por medio del análisis estadístico de descripciones olfativas de fragancias a partir de consumidores procedentes de diversos países.

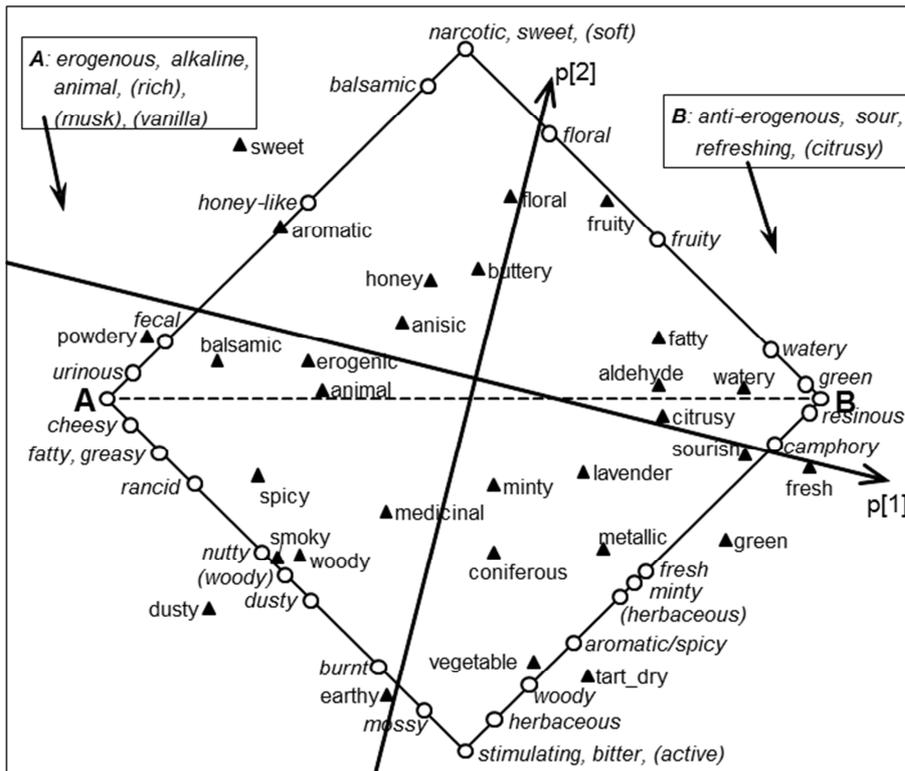


Figura 25: mapa obtenido por Zarzo y Stanton (2009) al aplicar PCA sobre la base de datos de Boelens y Haring (1981) (triángulos negros). También se superpone el mapa propuesto por P. Jellinek (círculos blancos).

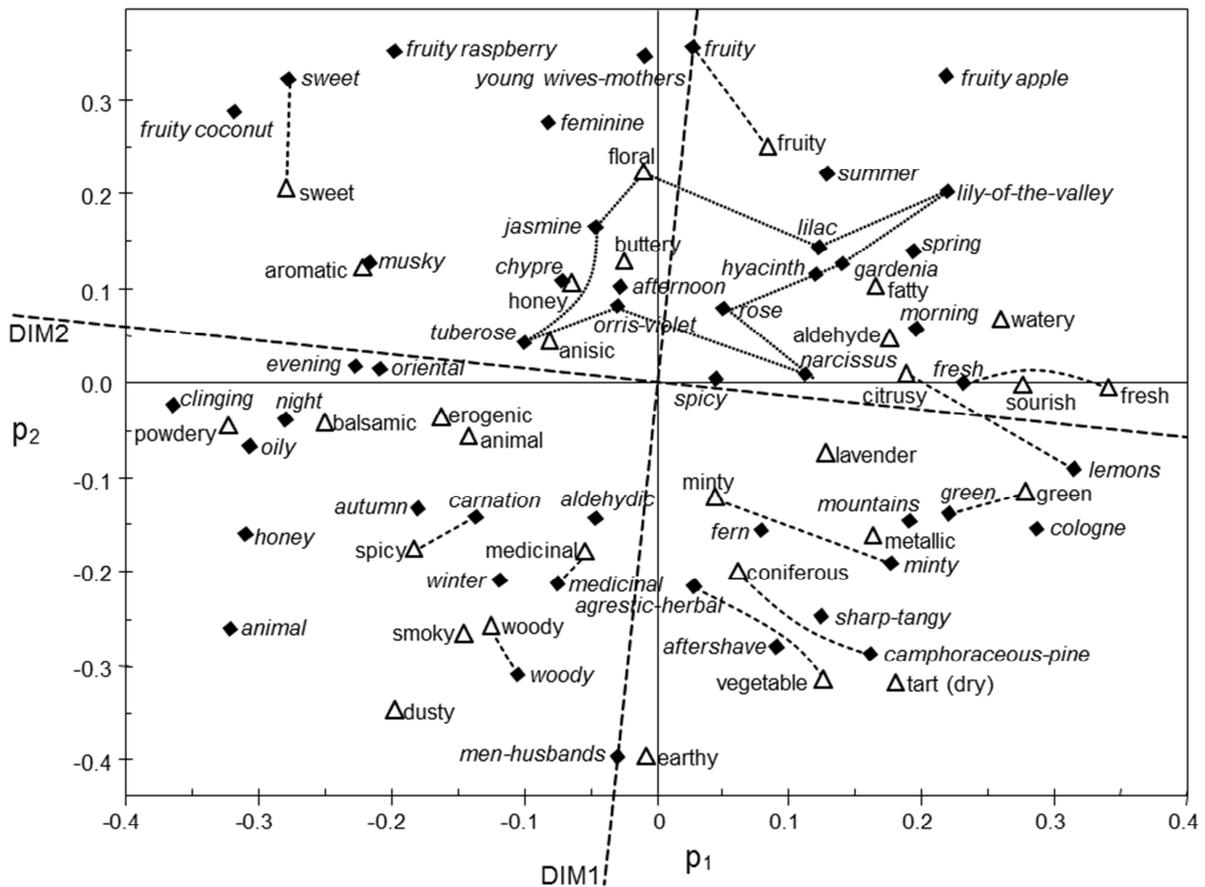


Figura 26: mapa obtenido por Zarzo (2013) al aplicar PCA sobre la base de datos de Boelens y Haring (1981) (triángulos blancos). Se ha superpuesto el mapa de fragancias desarrollado por Thiboud (1991) (diamantes negros).

A nivel general, el mapa propuesto no cumple la simetría en las dimensiones cálido-fresco y masculino-femenino de manera tan satisfactoria como los mapas con los que se compara. Se aprecia que el espacio ocupado por los descriptores cálidos es el doble que el espacio ocupado por los descriptores frescos. El espacio ocupado por los descriptores masculinos es casi el doble que el espacio ocupado por los descriptores femeninos. En cambio, en los mapas de la figura 25 y 26, el espacio ocupado por los descriptores masculinos es similar al de los femeninos, y ocurre también con el espacio ocupado por los cálidos y el que ocupa los descriptores frescos.

En los descriptores con notas dulces, se observa que “dulce” está situado de manera satisfactoria en cuanto a su nivel de femenino, pero en los otros mapas está posicionado como más cálido que en el mapa propuesto en este estudio, aunque esta distancia es pequeña. Los otros descriptores caracterizados fuertemente por notas dulces, como “caramelo”, “vainilla”, “chocolate” y “miel” deberían estar posicionados más cerca de “dulce”, pero aparecen como menos femeninos, y curiosamente estos descriptores sí que aparecen como muy cálidos. El único descriptor que coincide con nuestro estudio con notas dulces, aparte de “dulce”, es “miel”, y encaja con la posición propuesta por P. Jellinek (figura 25).

El descriptor “aromático” de la figura 24 encaja perfectamente con la posición que ocupa en los estudios con los que se ha comparado.

Los descriptores especiados (“clavo”, “canela” y “especia”) no están situados de manera consistente con los otros mapas. Se observa que “especia” (el único coincidente especiado) en los estudios precedentes está clasificado como masculino, mientras que en el mapa propuesto en este estudio aparece como femenino. No obstante, sí que ocupa una posición similar como cálido.

El descriptor “nuez” queda localizado de modo muy similar a la propuesta por P. Jellinek (figura 25).

Los descriptores con notas amaderadas están situados muy próximos a la posición que ocupan en el gráfico 26 según las clasificaciones obtenidas por Zarzo y Stanton (2009) así como la de Thiboud (1991).

Los descriptores con notas ahumadas encajan con los mapas de los estudios que se compara, ya que ocupan una posición similar al descriptor “ahumado” de estos estudios, así como a “quemado, ahumado”.

El descriptor “medicinal” debería estar situado como mucho más masculino, pero en el presente estudio aparece muy cerca del neutro.

El descriptor “seco-pulverulento” en el mapa propuesto aparece como bastante masculino, mientras que en el mapa de la figura 26, el descriptor “pulverulento” es clasificado como masculino casi neutro, y su nivel de calidez es mucho más notable.

El descriptor “mohoso-terroso” es muy cercano al descriptor “terroso” de la figura 26, aunque este último no aparece como fresco, y en la clasificación obtenida en el estudio sí.

Los descriptores asociados a vegetales encajan perfectamente con la posición que ocupan en los mapas de las figuras 25 y 26. Los descriptores asociados a notas herbáceas del estudio encajan con las clasificaciones que se realizan de éstos en los estudios precedentes.

El descriptor “floral” no ocupa exactamente la misma posición que en los estudios precedentes, ya que en el mapa propuesto aparece como poco femenino. No obstante, se encuentra dentro del recinto floral remarcado en la figura 26, en el cual se unen los descriptores asociados a notas florales, lo que hace pensar que es una posición adecuada.

Los descriptores asociados a notas cítricas aparecen en el mapa del estudio como ligeramente femeninos, lo cual diverge con la posición que ocupan en los mapas de los estudios precedentes, ya que se les clasifica como neutros. El nivel de frescor que obtienen es similar.

Los descriptores con notas frutales encajan con los estudios precedentes, en especial en la clasificación que se realiza en la figura 26.

4.4 Estudio del carácter hedónico de los descriptores

Partiendo del análisis PCA realizado en el estudio de las dimensiones olfativas, en el que resultaban 18 componentes relevantes, se guardan los *loadings* de estas componentes y se transpone la matriz resultante. Para cada descriptor está disponible su carácter hedónico a partir del estudio de Dravnieks *et al.* (1984). Esto permite realizar una regresión lineal múltiple paso a paso (*stepwise*), tanto *backward* como *forward*, con las 18 componentes como variables de entrada. Se obtiene el mismo modelo tanto con la opción *backward* como *forward*.

Estudiando los residuos de este modelo, se ha encontrado un marcado efecto cuadrático asociado a la primera componente. Para estudiar otros posibles efectos cuadráticos, se ha realizado de nuevo la regresión *stepwise*, añadiendo el efecto cuadrático de los *loadings* asociados a las diferentes componentes.

Selección *backward*:

Se ha obtenido el modelo indicado en la Tabla 9, con un coeficiente de determinación $R^2 = 0,757$.

Tabla 9: Resultados del modelo de regresión lineal *stepwise*, con selección *backward*

<i>Parámetro</i>	<i>Coefficiente Estimado</i>	<i>Error</i>	<i>T Estadístico</i>	<i>P-Valor</i>
Constante	0,559	0,138	4,049	0,0001
P[1]	16,660	1,008	16,529	0,0000
P[11]	3,327	1,016	3,273	0,0014
P[15]	-4,620	0,989	-4,673	0,0000
P[2]	-6,026	1,183	-5,095	0,0000
P[3]	4,620	1,033	4,471	0,0000
P[6]	3,467	1,051	3,300	0,0012
P[1] ²	-42,827	8,896	-4,814	0,0000
P[9] ²	36,180	10,695	3,383	0,0009

A continuación se indica la interpretación de las distintas variables del modelo:

- P[1], pesos en la formación de la primera componente: está básicamente determinado por los olores más típicos de perfumes femeninos, que se contraponen a descriptores de olores desagradables.
- P[2]: olores químicos. Dado que el coeficiente de regresión es negativo indica que estos olores tienden a percibirse como desagradables, lo cual tiene sentido.
- P[3]: olores más típicos de perfumes masculinos. El coeficiente de regresión positivo sugiere que estos olores tienden a percibirse como agradables, como es lógico.
- P[6]: olores herbáceos y vegetales, con coeficiente de regresión positivo.
- P[9]: esta variable sólo aparece en el modelo como cuadrática, pero el efecto lineal no es significativo, lo cual dificulta la interpretación.
- P[11]: olores cítricos y a nuez. El coeficiente positivo tiene sentido.

- P[15]: la componente está básicamente determinada por descriptores balsámicos los cuales aparecen con *loadings* negativos. El coeficiente de regresión negativo tiene sentido porque estos olores se perciben como agradables.

Selección *forward*:

Se obtiene el modelo de la Tabla 10, con un coeficiente de determinación $R^2 = 0,759$ (prácticamente el mismo que con la opción *backward*).

Tabla 10: Resultados del modelo de regresión lineal *stepwise*, con selección *forward*

<i>Parámetro</i>	<i>Coefficiente Estimado</i>	<i>Error</i>	<i>T Estadístico</i>	<i>P-Valor</i>
Constante	0,652	0,137	4,747	0,0000
P[1]	15,123	1,045	14,469	0,0000
P[12]	-2,836	1,023	-2,771	0,0064
P[15]	-3,817	0,997	-3,830	0,0002
P[3]	6,521	1,102	5,919	0,0000
P[6]	2,397	1,009	2,377	0,0189
P[7]	2,886	1,081	2,669	0,0086
P[1] ²	-27,121	8,080	-3,356	0,0010
P[11] ²	28,256	7,617	3,710	0,0003
P[3] ²	-42,397	9,761	-4,344	0,0000

La interpretación de cada variable en este modelo es la siguiente:

- P[1]: olores más típicos de perfumes femeninos, que se contraponen a descriptores de olores desagradables.
- P[3]: olores más típicos de perfumes masculinos, con coeficiente de regresión positivo lo cual es coherente.
- P[6]: olores herbáceos y vegetales.
- P[7]: descriptores afrutados (no cítricos), con coeficiente de regresión positivo, lo cual tiene sentido.
- P[11]: discrimina olores cítricos frente a balsámicos caramelo, pero se trata de una variable cuadrática cuyo efecto lineal no es significativo, lo cual dificulta la interpretación.
- P[12]: los descriptores amaderados son los de mayor influencia en esta componente con *loadings* negativos, de modo que el coeficiente de regresión negativo tiene sentido ya que estos olores son agradables.
- P[15]: esta componente está básicamente determinada por olores balsámicos que aparecen con *loadings* negativos. El coeficiente negativo de esta componente parece lógico, pues estos olores son agradables.

Diversos estudios de la bibliografía han sugerido que la primera componente principal (PC1) del Atlas de Dravnieks puede interpretarse como la dimensión hedónica (Khan *et al.*, 2007; Koulakov *et al.*, 2011). El hecho de que la variable P[1] sea la más importante de estos modelos de regresión es parcialmente consistente con esta teoría. No obstante,

han aparecido otras variables latentes cuyo efecto en estos modelos es claramente significativo. Esto parece indicar que la interpretación de PC1 como la dimensión hedónica es discutible. De hecho, en este trabajo se interpreta de una forma distinta, como un factor que contrapone los olores más típicos de fragancias femeninas frente a olores desagradables.

Los resultados indican que es necesario emplear la información de unas 6 componentes para predecir convenientemente el tono hedónico de los descriptores, lo que evidencia que interpretar PC1 como la dimensión hedónica es un criterio discutible.

Múltiples estudios han encontrado que cuando se evalúa un conjunto amplio de olores, tanto agradables como desagradables, generalmente la primera componente refleja el carácter hedónico. Es decir, cuando se percibe un olor, lo más inmediato que puede decirse es si huele bien o mal. Por este motivo, se había considerado hasta ahora que PC1 del Atlas de Dravnieks reflejaba de hecho el tono hedónico. Sin embargo, los resultados obtenidos ponen en duda esta cuestión.

Elección del modelo:

El coeficiente de determinación es prácticamente el mismo tanto con la opción *forward* como *backward*. Las variables que entran en el modelo son muy similares: las variables latentes de *loadings* asociadas a las componentes 1ª, 3ª, 6ª, 11ª y 15ª. En el caso *backward* entran en el modelo las componentes 2ª y 9ª, mientras que en el otro caso aparecen la 7ª y 12ª.

Debido a que las variables del modelo parecen más adecuadas, y a que el R^2 ajustado es superior, se selecciona el modelo obtenido por selección *forward*.

A continuación se valida el modelo obtenido por selección *forward*:

Hipótesis de normalidad

Se estudia la normalidad de los residuos representando los mismos sobre un Papel Probabilístico Normal. Se observa que la mayoría de los residuos se ajustan a la recta normal. Los coeficientes de asimetría y curtosis normalizados valen 0,42 y -1,27 respectivamente, por lo que se encuentran dentro del intervalo de -2 a 2 de modo que la hipótesis de normalidad de los residuos es admisible.

Hipótesis de homocedasticidad de los residuos

Se estudia si existe homocedasticidad en los residuos mediante los siguientes métodos gráficos. En el gráfico de valores observados frente a predichos se observa una distribución de los puntos homogénea respecto a la recta. En la figura 28 se observa que la variabilidad de los residuos no depende de los valores de $p[1]$, lo que sugiere que se cumple la hipótesis de homocedasticidad.

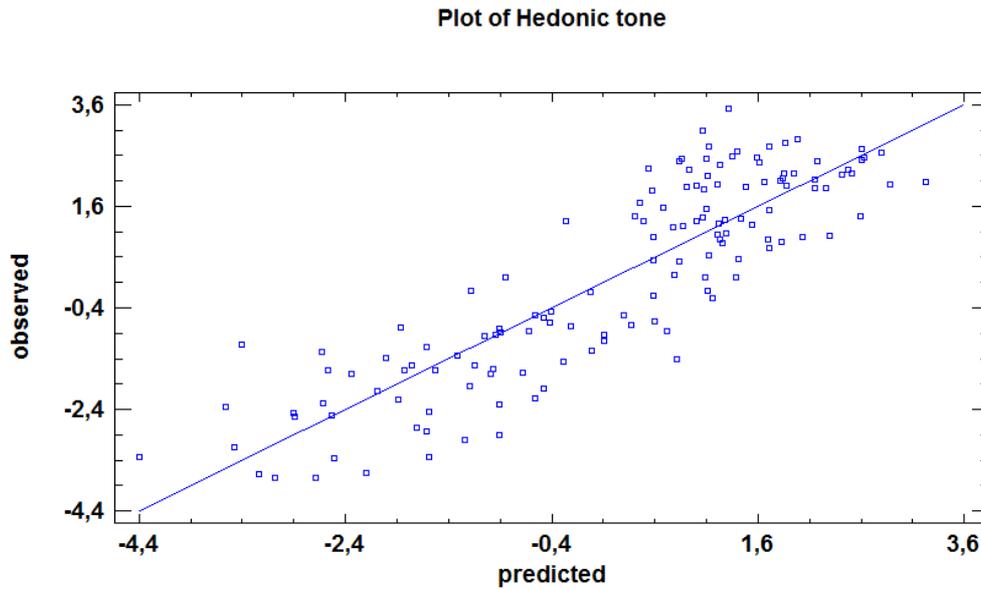


Figura 27: Gráfico de los valores observados del tono hedónico frente a los predichos.

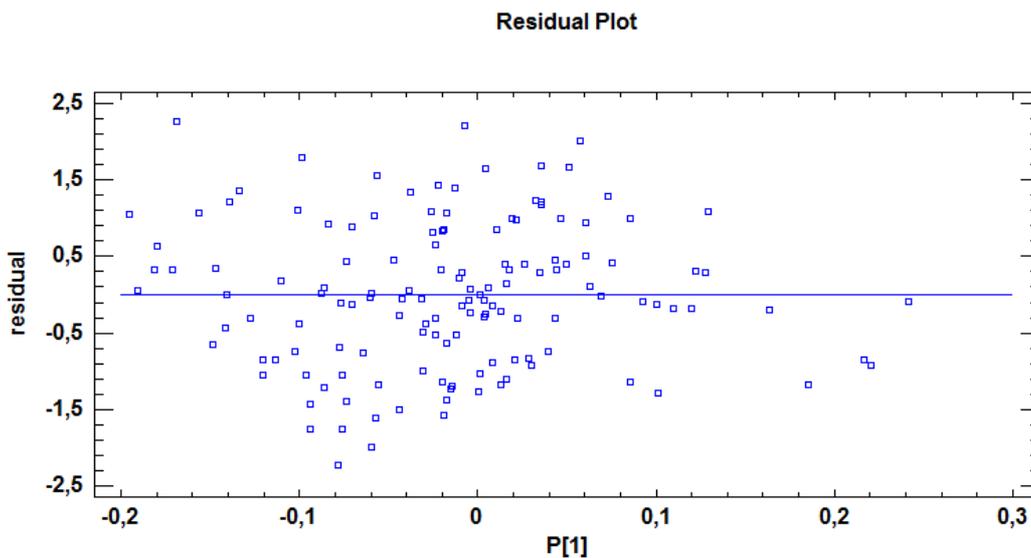


Figura 28: Gráficos de los residuos frente a los pesos de la 1ª componente principal.

Estas comprobaciones se han llevado a cabo para todas las variables incluidas en el modelo final de regresión múltiple, pero sólo se muestran frente a la primera componente principal (figura 28) debido a que no aportan información relevante para el estudio. Se ha elegido mostrar la figura para la primera componente porque es la más importante.

La ecuación resultante del modelo de regresión lineal (Tabla 10) es la siguiente:

$$\text{Tono hedónico} = 0,65 + 15,12 \cdot P[1] - 2,84 \cdot P[12] - 3,82 \cdot P[15] + 6,52 \cdot P[3] + 2,40 \cdot P[6] + 2,89 \cdot P[7] - 27,12 \cdot P[1]^2 - 42,40 \cdot P[3]^2 + 28,26 \cdot P[11]^2$$

A partir de los valores del estadístico t indicados en la Tabla 10 se deduce que la dimensión 1, que discrimina olores típicos de perfumes femeninos frente a olores desagradables, es la que más capacidad tiene para predecir el tono hedónico de los descriptores. La dimensión 3, que hace referencia a los olores característicos de perfumes masculinos, es la segunda que más contribuye a la bondad de ajuste. El resto de las dimensiones del modelo tiene valores similares del estadístico t , por lo que no es posible establecer un rango de importancia en cuanto a la capacidad predictiva del modelo.

4.5 Estudio de descriptores olfativos ambiguos

4.5.1 Diferencias entre los descriptores “fragante” y “aromático”

Los términos “fragante” y “aromático” son ambiguos en perfumería. De hecho, según el diccionario de la Real Academia de la Lengua, el primero significa olor suave y delicioso. Un olor aromático se define como agradable, según este diccionario. Por tanto, ambos términos aluden en castellano a olores agradables, de modo que no está clara la diferencia entre ellos.

En el campo de la perfumería esta ambigüedad también existe. Por ejemplo, la vainilla es considerada a veces como referencia para olores aromáticos. Sin embargo, otros autores asocian “aromático” a las notas típicas de las fragancias masculinas.

En conclusión, la interpretación de ambos términos es totalmente ambigua y no está claro qué concepto tenían de ellos los catadores del panel en el estudio de Dravnieks (1985). Para profundizar en esta cuestión, se ha realizado un PLS con todos los descriptores, considerando como variables dependientes “aromático” y “fragante”. Los parámetros del modelo obtenido se muestran en la Tabla 11.

Tabla 11: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 4 componentes del modelo PLS que considera como variable dependiente los descriptores “fragante” y “aromático”. (R^2 total = 0.921; Q^2 total = 0.878)

Dimensión	R^2	Q^2
1	0.669	0.878
2	0.180	0.644
3	0.049	0.164
4	0.024	0.020

Dada la elevada cantidad de variables, resulta conveniente realizar una cierta selección de aquellas más importantes en este modelo, a efectos de facilitar la interpretación de los resultados. A continuación, se realiza un nuevo PLS, solamente con aquellos descriptores que han tenido un valor de VIP superior a 0.8 considerando el modelo con

4 componentes. El estadístico VIP cuantifica la importancia de las variables en la capacidad predictiva del modelo de regresión. Este criterio de eliminar las variables con un valor inferior a 0.8 corresponde a los descriptores poco relevantes para el modelo. En total se han incorporado 74 variables de entrada en el nuevo modelo. A continuación se muestran los parámetros del modelo obtenido en la tabla 12.

Tabla 12: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 4 componentes del modelo PLS que considera como variable dependiente los descriptores “fragante” y “aromático”, con 74 variables de entrada, seleccionadas por el VIP obtenido del modelo de la tabla 11. (R^2 total = 0.917; Q^2 total = 0.887)

Dimensión	R^2	Q^2
1	0.673	0.654
2	0.183	0.178
3	0.036	0.040
4	0.025	0.016

El siguiente gráfico muestra la superposición de los pesos del modelo PLS correspondiente a las dos primeras componentes. Las variables predictivas se muestran en rojo. La bondad de ajuste con dos componentes es $R^2 = 85\%$.

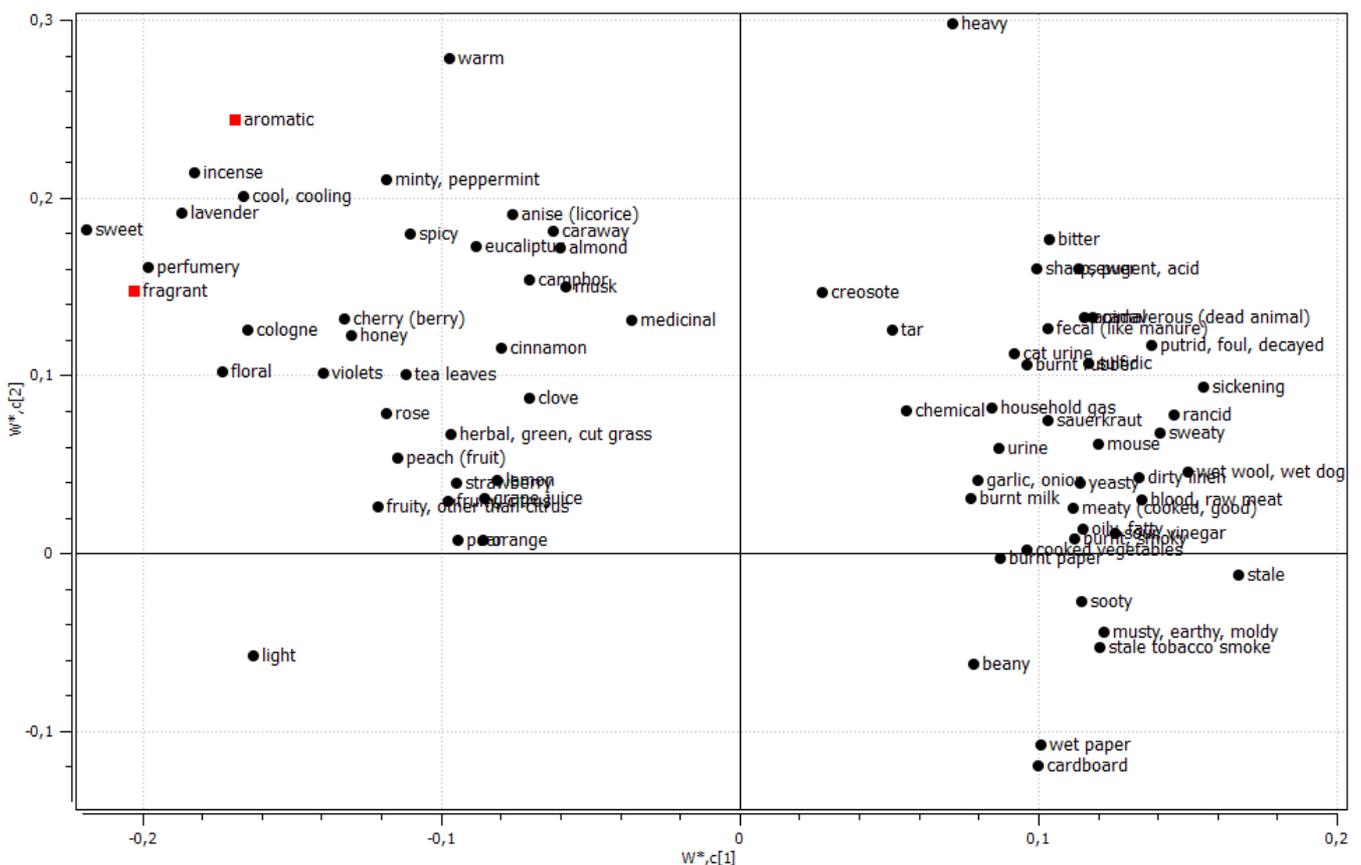


Figura 29: Gráfico de *loadings* obtenido para las dos primeras componentes del modelo PLS con $Y =$ “fragante” y “aromático”.

La posición de los descriptores “fragante” y “aromático” indica que son similares, ya que ambos descriptores se encuentran próximos y en el mismo cuadrante. Se observa

que “fragante” obtiene una mayor puntuación negativa en la primera dimensión que el descriptor “aromático”. Esta primera componente del modelo está asociada a olores típicos de perfumes femeninos, ya que los descriptores con mayor peso negativo en el eje horizontal (“dulce”, “perfumería”, “floral” y “ligero”) coinciden con la primera dimensión de Castro *et al.* (2013), asociada a los olores de perfumes femeninos. Por ello, se puede decir que “fragante” fue preferentemente aplicado por el panel de catadores de Dravnieks (1985) para describir notas típicas de perfumes femeninos. Esto tiene sentido, pues “fragante” se asocia a fragancia, y las fragancias femeninas son más abundantes en el mercado que las masculinas.

Respecto a la segunda dimensión del modelo PLS, “aromático” obtiene mayor puntuación que “fragante”. El descriptor de mayor peso en esta componente es “pesado”, y en la posición opuesta se encuentra “ligero” (*light*). Ambos términos son semánticamente opuestos, lo cual resulta consistente con su posición opuesta en la figura 29. Por otra parte, los descriptores “menta”, “cool” y “lavanda” aluden a olores canforáceos y quedan cerca de “aromático”. Las notas canforáceas son más típicas de las fragancias masculinas. Por tanto, los resultados sugieren que el panel de catadores tuvo cierta tendencia a describir las notas masculinas como “aromático”, más que como “fragante”. Este criterio es el más aceptado en perfumería, y de hecho en la clasificación de Edwards (2010), la familia de perfumes denominada “aromático” o “fougère” es la que comprende la mayor cantidad de perfumes masculinos.

Para profundizar en el estudio de las diferencias entre estos descriptores, se ha ajustado un PLS para cada descriptor de la pareja por separado, y se han comparado los coeficientes PLS obtenidos para la primera componente de cada modelo.

Modelo PLS para predecir el descriptor “fragante”:

Se ha realizado un PLS con todos los descriptores, excepto “aromático”, considerando como variable dependiente “fragante”. La Tabla 13 muestra los parámetros del modelo obtenido.

Tabla 13: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 3 componentes del modelo PLS con $Y = \text{“fragante”}$ (R^2 total = 0.954; Q^2 total = 0.926)

Dimensión	R^2	Q^2
1	0.796	0.775
2	0.136	0.124
3	0.023	0.027

Dada la elevada cantidad de variables, resulta conveniente realizar una cierta selección de aquellas más importantes en este modelo, a efectos de facilitar la interpretación de los resultados. A continuación, se realiza un nuevo PLS, solamente con aquellos descriptores que han tenido un valor de VIP superior a 1 considerando el modelo con una componente. El estadístico VIP cuantifica la importancia de las variables en la capacidad predictiva del modelo de regresión. Este criterio de selección de variables resulta adecuado para facilitar la interpretación de los resultados. Las variables incluidas en el modelo, que son las de mayor VIP, son aquellos descriptores cuya información es

relevante para la predicción. En total se han incorporado 52 variables de entrada en el nuevo modelo. A continuación se muestran los parámetros del modelo obtenido en la Tabla 14.

Tabla 14: Varianza explicada (R^2) y parámetro Q^2 del PLS obtenido con $Y = \text{“fragante”}$, con las 52 variables de mayor capacidad explicativa (R^2 total = 0,945; Q^2 total = 0,928)

Dimensión	R^2	Q^2
1	0.786	0.776
2	0.120	0.120
3	0.039	0.032

Posteriormente se han calculado los coeficientes de regresión del modelo y su intervalo de confianza al 95%, que se muestran en el gráfico siguiente (figura 30).

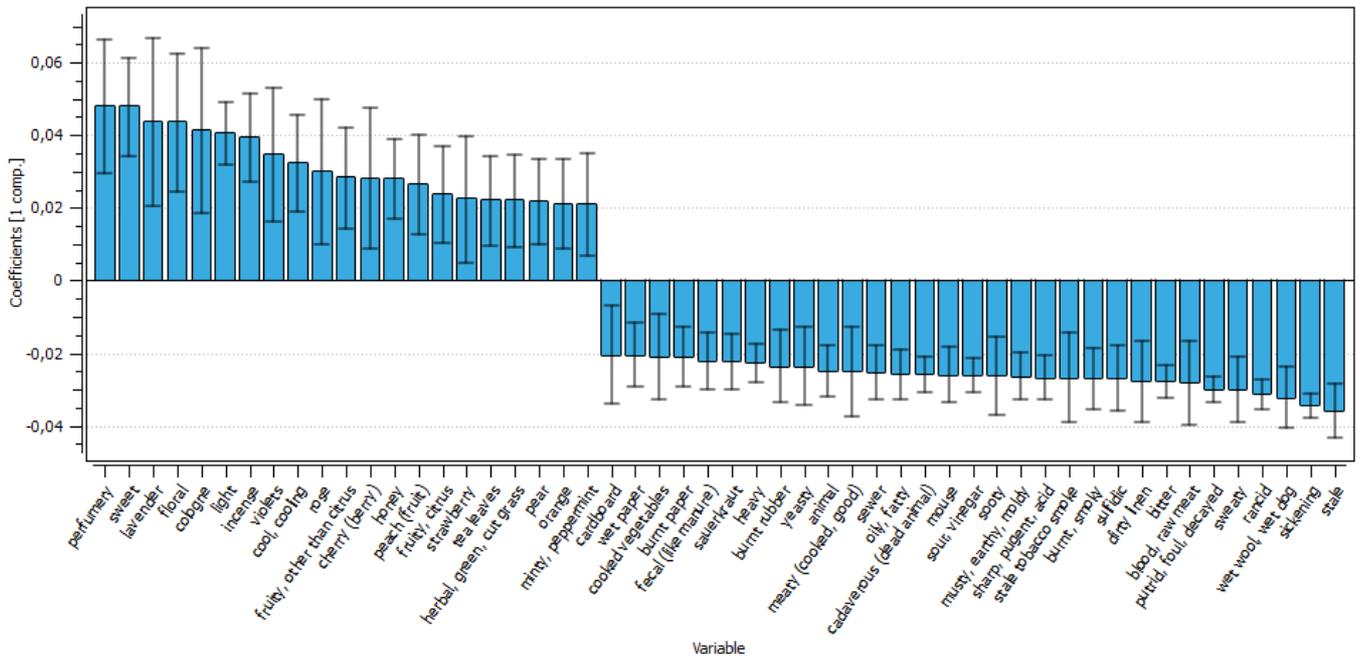


Figura 30: Coeficientes de regresión PLS de la primera componente para la variable dependiente “fragante”, del modelo con las 52 variables de mayor capacidad predictiva.

Modelo PLS para predecir el descriptor “aromático”:

Se ha ajustado un PLS con todos los descriptores, excepto “fragante”, considerando como variable dependiente “aromático”. Los parámetros del modelo se indican en la Tabla 15.

Tabla 15: Varianza explicada (R^2) y estadístico Q^2 para las 3 componentes del modelo PLS con Y = “aromático” (R^2 total = 0,900; Q^2 total = 0,839)

Dimensión	R^2	Q^2
1	0.608	0.573
2	0.244	0.221
3	0.048	0.046

Dada la elevada cantidad de variables, resulta conveniente realizar una cierta selección de aquellas más importantes en este modelo, a efectos de facilitar la interpretación de los resultados. A continuación, se realiza un nuevo PLS empleando solamente aquellos descriptores con un VIP superior a 1 considerando el modelo con una componente. El objetivo es repetir el modelo empleando únicamente aquellos descriptores cuya información es relevante para la predicción. En total se han incorporado 54 variables explicativas al nuevo modelo. En la Tabla 16 se muestran los parámetros del modelo obtenido.

Tabla 16: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) del modelo PLS ajustado para Y = “aromático”, con las 54 variables de mayor capacidad predictiva (R^2 total = 0.848; Q^2 total = 0.813)

Dimensión	R^2	Q^2
1	0.622	0.599
2	0.188	0.179
3	0.038	0.035

Posteriormente se han calculado los coeficientes de regresión del modelo y su intervalo de confianza al 95%, que se muestran en el gráfico de la figura 31.

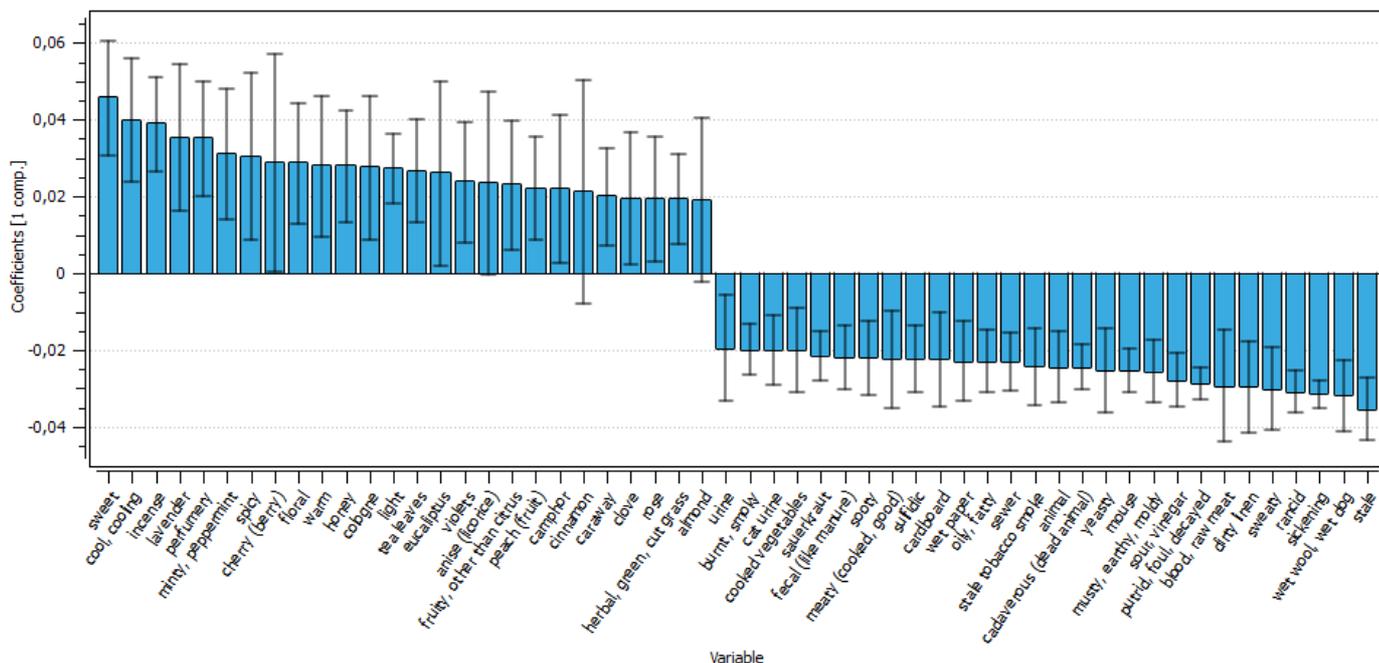


Figura 31: Coeficientes de regresión PLS de la primera componente para la variable dependiente “aromático”, del modelo de la tabla 16.

Diferencias entre los modelos PLS para “fragante” y “aromático”:

En este punto se comparan los coeficientes de regresión PLS obtenidos para cada uno de los modelos, el PLS con variable respuesta “fragante” (figura 30) y el PLS con variable respuesta “aromático” (figura 31).

Tras analizar detenidamente el orden de las variables que aparecen ordenadas con los mayores valores de VIP, la conclusión que se deduce es que “aromático” se aplicó preferentemente a los olores canforáceos (“cool”, “menta”, “alcanfor” y “eucaliptus”) así como también a olores especiados (“especia”, “clavo” y “canela”). Curiosamente, ambos tienen en común un carácter trigeminal, por activación del nervio trigémino. Así pues, “cool” aparece en segunda posición en la predicción de “aromático” pero en posición 9ª respecto a “fragante”. “Menta” aparece en 6ª posición para “aromático” pero en posición 21ª para “fragante”. “Eucaliptus” y “alcanfor” aparecen en la lista de mejores predictores para “aromático” pero no para “fragante”. Esto es consistente con el análisis realizado del modelo PLS conjunto para “fragante” y “aromático” (figura 29). De modo análogo, “especia” y “clavo” figuran entre los mejores descriptores para “aromático”, y han salido como significativos, ya que el intervalo de confianza del coeficiente se encuentra por encima de 0. En cambio, en el modelo “fragante” no han aparecido descriptores asociados a olores especiados. Por tanto, los resultados sugieren que el panel de catadores tuvo cierta tendencia a describir las notas especiadas como “aromático”, más que como “fragante”. Las notas especiadas y canforáceas son más típicas de las fragancias masculinas. Por tanto, estos resultados sugieren que el panel de catadores tuvo cierta tendencia a describir las notas masculinas como “aromático”, más que como “fragante”.

Por el contrario, “fragante” se aplicó preferentemente para los olores florales. De hecho, “floral” aparece en posición 4ª para la predicción de “fragante” pero en 9ª posición para “aromático”. En el caso de “violetas”, 8ª posición para “fragante” pero 16ª para “aromático”. El descriptor “rosa” ocupa la 10ª posición para “fragante” y la 24ª para “aromático”. También “fragante” se aplicó preferentemente para los descriptores asociados a perfumería femenina. En el caso de “fragante”, los descriptores asociados a la perfumería femenina “perfumería”, “floral”, “colonia” y “ligero” ocupan las posiciones 1ª, 4ª, 5ª y 10ª, respectivamente, mientras que en el modelo para “aromático” ocupan las posiciones 5ª, 9ª, 12ª y 13ª, respectivamente. Esto es consistente con el PLS realizado para la predicción conjunta de “fragante” y “aromático” (figura 29).

En general, estos resultados son coherentes con la interpretación de olores fragantes y aromáticos que se emplea actualmente en perfumería: “aromático” suele aplicarse preferentemente a las notas más frecuentes de perfumes masculinos, y “fragante” para los femeninos, fundamentalmente caracterizados por las notas florales. De hecho, las notas canforáceas son básicamente las responsables de catalogar una fragancia como “aromática”, según el criterio de Edwards (2010).

En el modelo “aromático” el descriptor “cálido” tiene un coeficiente de regresión significativo. Esto no ocurre para “fragante”, por lo que sugiere que el panel de catadores tuvo cierta tendencia a describir las notas cálidas como “aromático”, más que como “fragante”.

Lo que más llama la atención en este estudio es que “dulce” ocupa la segunda posición en la predicción de “fragante” pero la primera posición en “aromático”. Este resultado resulta extraño, pues los olores dulces son típicamente femeninos, de modo que la asociación entre olores aromáticos y dulces no era previsible. La interpretación de esta asociación, que resulta llamativa, pudiera tener su explicación en el concepto de “aromático” empleado antiguamente, hasta la década de los años 1980. En química se denominan aromáticos a los compuestos derivados del benceno, pues la mayoría de ellos presentan olores intensos. Al parecer, bastantes compuestos derivados de la química aromática tenían olores dulces, y por este motivo se asoció “aromático” con “dulce”. De hecho, en el estudio de Boelens y Haring (1981), la vainillina fue el compuesto tomado como referencia para “aromático”, el cual tiene un olor marcadamente dulce.

Curiosamente, el Atlas de Dravnieks se publicó en 1985, y es posible que algunos de los catadores todavía consideraran la asociación entre “dulce” con “aromático” en el sentido de “compuesto derivado de la química aromática”. Sin embargo, al parecer en esa época esta asociación ya comenzaba a entrar en desuso, por lo que da la sensación de que otros catadores asociaron “aromático” a los olores típicos de fragancias masculinas, que es la interpretación actual.

La asociación entre “fragante” y “aromático” con olores femeninos y masculinos, respectivamente, se remonta al famoso botánico Lineo (Linnaeus, 1756). En esa época no era tan habitual el uso de fragancias entre hombres, de modo que describir un olor como “fragante” era casi sinónimo de “fragancias femenina”. Así pues, Lineo describía como aromático aquellos olores de plantas o flores con un carácter agradable pero que

no recordaban las fragancias femeninas. Curiosamente, la misma distinción entre olores fragantes y aromáticos, propuesta en 1756, es la que se emplea actualmente.

4.5.2 Diferencias entre los descriptores “perfumería” y “colonia”

Según el diccionario de la Real Academia de la Lengua, el término “colonia” se refiere al “agua de Colonia”, que es un perfume fresco de carácter cítrico. Estrictamente hablando, una colonia alude a un tipo de perfumes, por lo que cabe esperar diferencias entre los descriptores “perfumería” y “colonia” en el Atlas de Dravnieks.

Para profundizar en esta cuestión, se ha realizado un PLS considerando “perfumería” y “colonia” como variables Y, y el resto de descriptores como variables X. Los parámetros del modelo se muestran en la Tabla 17.

Tabla 17: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para el modelo PLS con Y = “perfumería” y “colonia” (R^2 total = 0.883; Q^2 total = 0.828)

Dimensión	R^2	Q^2
1	0.531	0.492
2	0.287	0.285
3	0.039	0.039
4	0.025	0.012

Dada la elevada cantidad de variables, resulta conveniente realizar una cierta selección de aquellas más importantes en este modelo, a efectos de facilitar la interpretación de los resultados. Por este motivo, se ha ajustado un nuevo PLS con aquellos descriptores que han tenido un VIP superior a 0.8 considerando el modelo con 4 componentes. Este criterio de eliminar las variables con un VIP inferior a 0.8 corresponde a los descriptores poco relevantes para el modelo. En total se han incorporado 66 variables explicativas en el nuevo modelo. A continuación se muestran los parámetros del modelo obtenido (Tabla 18).

Tabla 18: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 4 componentes del modelo PLS con Y = “perfumería” y “colonia”, obtenido con las 66 variables explicativas de mayor capacidad predictiva (R^2 total = 0.893; Q^2 total = 0.853)

Dimensión	R^2	Q^2
1	0.537	0.503
2	0.292	0.303
3	0.039	0.029
4	0.024	0.018

El gráfico de la figura 32 muestra los pesos de las variables en la formación de las dos componentes PLS. Las variables dependientes aparecen en rojo.

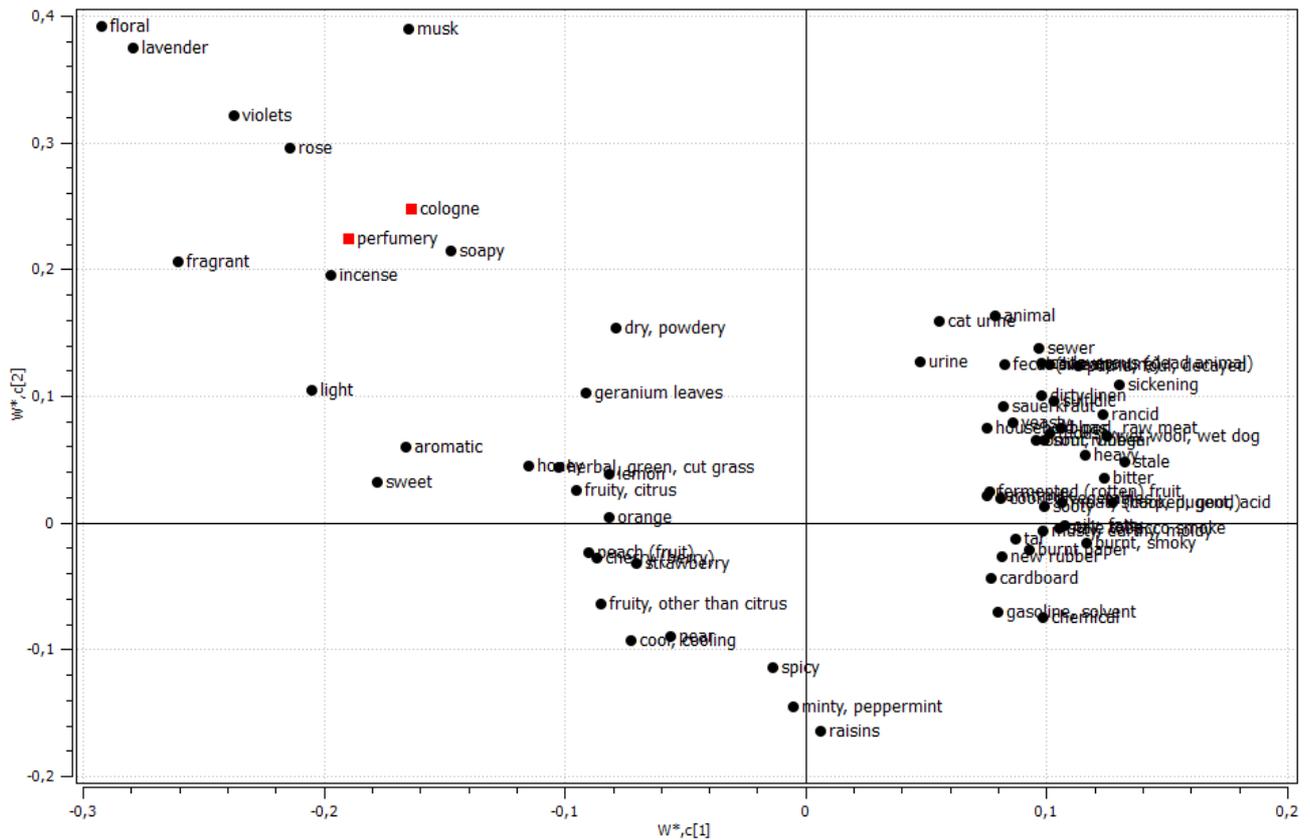


Figura 32: Gráfico de *loadings* obtenido para las dos primeras componentes del modelo PLS “perfumería” y “colonia” correspondiente a la Tabla 18.

Lo más relevante del gráfico es que “perfumería” y “colonia” aparecen muy próximos, lo que sugiere que el panel de catadores interpretó ambos términos casi como sinónimos. Este resultado tiene sentido, pues de hecho es habitual que se empleen los términos perfume o colonia indistintamente, aunque estrictamente las colonias son perfumes más frescos y generalmente cítricos.

Se observa que la primera componente hace referencia básicamente a la dimensión hedónica, encontrándose en la parte izquierda del gráfico los olores más agradables, y en la parte derecha aquellos descriptores asociados a olores desagradables. Respecto a la primera componente, la variable respuesta “perfumería” alcanza valores un poco superiores (en valor absoluto) en comparación a “colonia”, lo que podría indicar que “perfumería” tiende a asociarse a olores más agradables que en el caso de “colonia”. La segunda dimensión hace referencia a olores florales, ya que “floral” es el descriptor de mayor peso, y los descriptores “rosa” y “violetas” también presentan puntuaciones elevadas. En la segunda dimensión, la variable respuesta “colonia” alcanza un peso ligeramente superior al descriptor “perfumería”. Esto sugiere que quizás “colonia” tiende a asociarse a olores más florales que en el caso de “perfumería”, en que el olor estará asociado con menos frecuencia a notas florales, siendo más amplio su abanico de fragancias, como por ejemplo a notas dulces o frutales.

Para profundizar en el estudio de las diferencias entre “perfumería” y “colonia”, se ha ajustado un PLS para cada descriptor de la pareja por separado, y se han comparado los coeficientes PLS obtenidos para la primera componente de cada modelo.

Modelo PLS para predecir el descriptor “perfumería”:

Se ha ajustado un PLS con todos los descriptores, excepto “colonia”, considerando como variable dependiente “perfumería”. Los parámetros del modelo se muestran en la Tabla 18.

Tabla 18: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 4 componentes del modelo PLS con $Y = \text{“perfumería”}$ (R^2 total = 0,913; Q^2 total = 0,874)

Dimensión	R^2	Q^2
1	0.594	0.563
2	0.265	0.263
3	0.035	0.035
4	0.019	0.012

Dada la elevada cantidad de variables, se han seleccionado las más importantes de este modelo, a efectos de interpretar los resultados. A continuación se ha realizado un nuevo PLS, solamente con aquellos descriptores con un VIP superior a 1 considerando el modelo con una componente, que son las variables de mayor capacidad predictiva. En total se han incorporado 42 variables explicativas. Los parámetros del modelo resultante se muestran en la Tabla 19.

Tabla 19: Varianza explicada (R^2) y estadístico Q^2 para las 4 componentes del modelo PLS con $Y = \text{“perfumería”}$, empleando las 42 variables de mayor VIP (R^2 total = 0.916; Q^2 total = 0.890)

Dimensión	R^2	Q^2
1	0.649	0.626
2	0.228	0.237
3	0.025	0.015
4	0.014	0.012

Posteriormente se han obtenido los coeficientes de regresión del modelo y su intervalo de confianza al 95%, que se muestran en el gráfico de la figura 33.

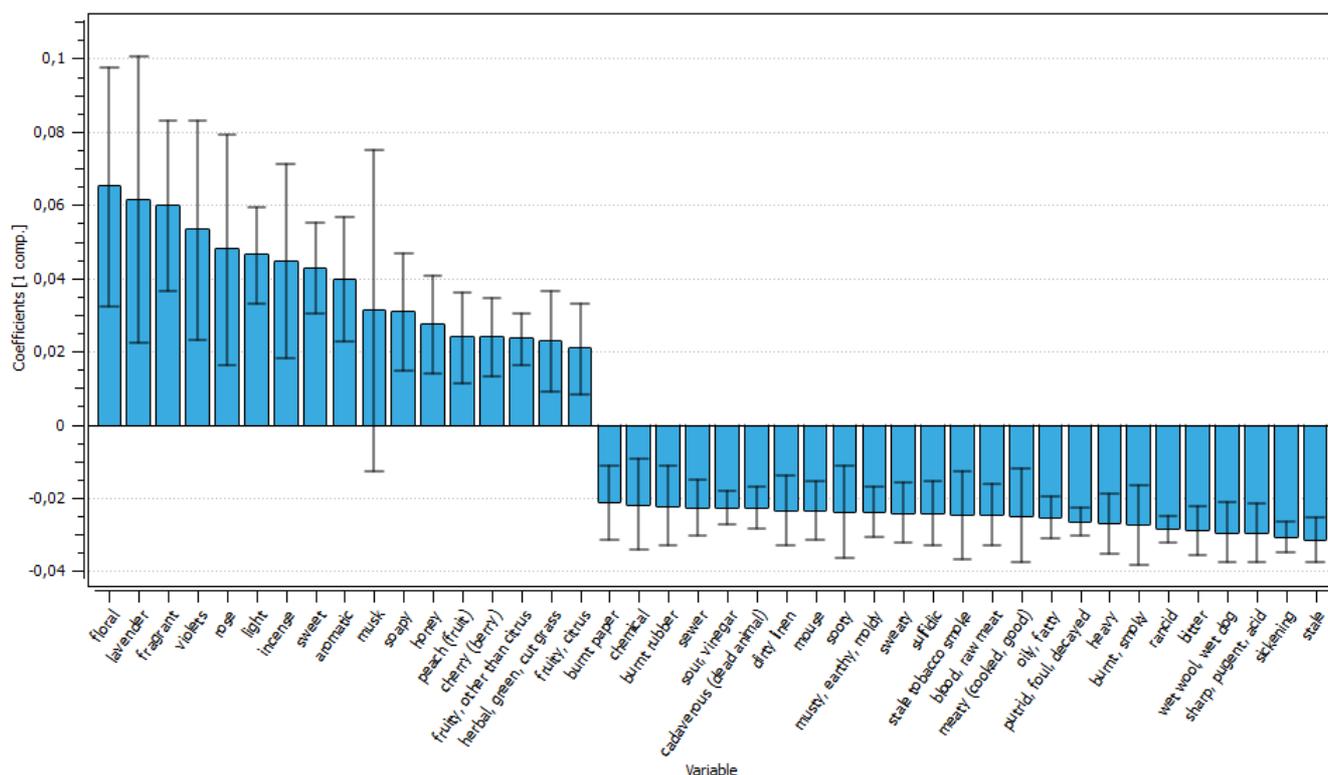


Figura 33: Coeficientes de regresión PLS de la primera componente para la variable dependiente “perfumería”, del modelo de la tabla 19.

Modelo PLS para predecir el descriptor “colonia”:

Se ha realizado un PLS con todos los descriptores, excepto “perfumería”, considerando como variable dependiente “colonia”. Los parámetros del modelo se muestran en la Tabla 20.

Tabla 20: Varianza explicada (R^2) y varianza explicada por validación cruzada (Q^2) para las 4 componentes del modelo PLS con $Y = \text{“colonia”}$ (R^2 total = 0,872; Q^2 total = 0,788)

Dimensión	R^2	Q^2
1	0.481	0.430
2	0.311	0.302
3	0.046	0.045
4	0.032	0.012

Para facilitar la interpretación del modelo, se ha realizado una selección de las variables más importantes. A continuación, se realiza un nuevo PLS con aquellos descriptores cuyo VIP es superior a 1 considerando el modelo con una componente. Estos descriptores son los de mayor capacidad predictiva. En total se han incorporado 41 variables explicativas en el nuevo modelo. Los parámetros del modelo resultante se muestran en la Tabla 21.

Tabla 21: Varianza explicada (R^2) y estadístico Q^2 del modelo PLS con $Y = \text{“colonia”}$, a partir de las 41 variables de mayor capacidad predictiva (R^2 total = 0,866; Q^2 total = 0,814)

Dimensión	R^2	Q^2
1	0.546	0.503
2	0.250	0.264
3	0.051	0.032
4	0.019	0.015

Posteriormente se han calculado los coeficientes de regresión del modelo y su intervalo de confianza al 95%, que se muestran en el gráfico de la figura 34.

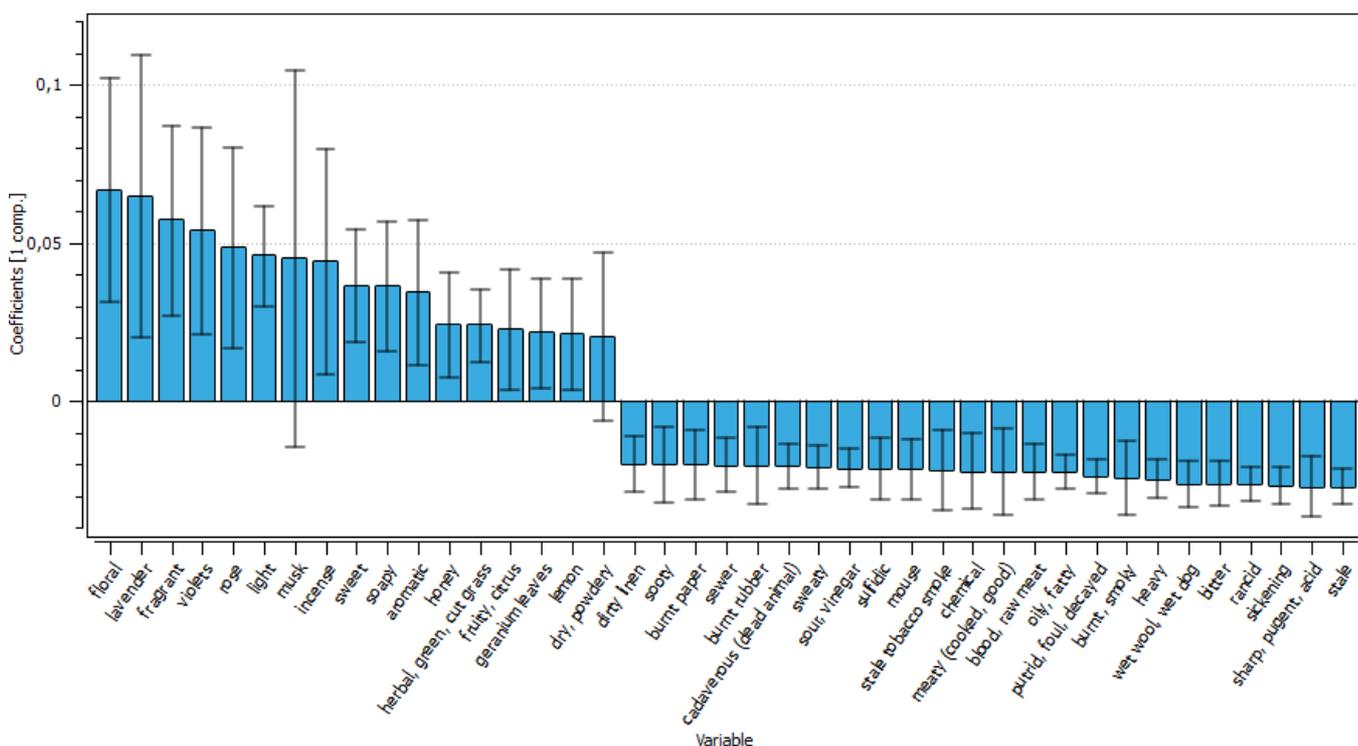


Figura 34: Coeficientes de regresión PLS de la primera componente para la variable dependiente “colonia”, del modelo correspondiente a la Tabla 21.

Diferencias entre los modelos PLS para “perfumería” y “colonia”:

En este punto se comparan los coeficientes de regresión PLS obtenidos para cada uno de los modelos (figuras 33 y 34). Se puede observar que los descriptores de mayor coeficiente y el orden de los descriptores de mayor a menor son muy similares. Comparando el orden de las variables de mayor coeficiente de regresión, se deduce que los resultados son muy similares. De hecho, las 6 primeras variables son las mismas.

La principal diferencia radica en las notas afrutadas. En el modelo “perfumería”, han aparecido tres descriptores con notas frutales como significativos (el intervalo de confianza del coeficiente se encuentra por encima de 0), que son: “frutal”, “cereza” y

“melocotón”, mientras que en el modelo “colonia” no ha aparecido ningún descriptor con notas afrutadas. Esto sugiere que el panel de catadores tuvo cierta tendencia a describir las notas frutales como “perfumería”, más que como “colonia”.

Se observan diferencias entre los modelos para los descriptores con notas cítricas. Al ordenar los coeficientes de regresión de mayor a menor, se observa que en el modelo “colonia” el descriptor “cítrico” ocupa la posición 13^a mientras que para “perfumería”, la 17^a. Además, en el modelo “colonia” se ha obtenido como relevante el descriptor “limón”, que tienen notas cítricas, pero no en el modelo de “perfumería”. Esto sugiere que el panel de catadores tuvo cierta tendencia a describir las notas cítricas como “colonia”, más que como “perfumería”. Esto es coherente con la definición del diccionario de la Real Academia de la Lengua al término “colonia”, al que se refiere como “agua de Colonia”, que es un perfume fresco de carácter cítrico.

5 Conclusiones

En este apartado se presentan las conclusiones más relevantes obtenidas en los diferentes capítulos que abarca el presente trabajo:

Estudio de las dimensiones del espacio perceptual olfativo

El número de componentes principales que aportan información relevante del espacio perceptual olfativo, a partir de los resultados obtenidos, es aproximadamente 18, lo cual es coherente con otros estudios citados en la bibliografía, que las sitúan entre 10 y 30.

Se han comparado los resultados con la clasificación de los mismos descriptores olfativos según el estudio de Castro *et al.* (2013), mediante la técnica de Factorización Matricial No-negativa. En general, el análisis PCA con un adecuado pretratamiento da lugar a dimensiones olfativas comparables con las que proponen Castro *et al.* (2013). No obstante, existen ciertas diferencias. En el estudio realizado por Castro *et al.* (2013) la interpretación de variables latentes es más clara en algunos casos, en comparación con los resultados obtenidos con PCA.

Los gráficos de pesos obtenidos con PCA permiten discutir las similitudes y diferencias entre descriptores. Así pues, la primera componente principal está básicamente determinada por olores típicos de fragancias femeninas, los cuales se contraponen a olores desagradables. Otra dimensión ortogonal refleja los olores masculinos, los cuales aparecen opuestos a los descriptores de sustancias químicas, que se refieren a olores “artificiales”, no presentes en la naturaleza. Los descriptores relativos a frutas cítricas aparecen agrupados, opuestos a olores especiados y canforáceos. Cabe resaltar que los olores cítricos aparecen independientes de los afrutados, lo cual es bien conocido en perfumería. Otro factor está asociado a olores florales, que aparecen opuestos a los especiados. Los descriptores con notas canforáceas también aparecen relacionados, lo cual tiene sentido dado su carácter trigeminal. Finalmente, los olores sulfurosos también aparecen como ortogonales debido a que se asocian a la presencia de compuestos azufrados.

Clasificación de los descriptores olfativos

Se han propuesto dos clasificaciones de los descriptores del Atlas de Dravnieks (1985), con diferente número de agrupaciones. La primera consta solamente con 8 clústeres, que es un valor bajo en comparación con otros estudios similares publicados, y que tiene como objetivo proporcionar una clasificación general de los descriptores. También se ha propuesto una clasificación con un 37 clústeres, que es un valor elevado en comparación con otros estudios similares, pero pretende identificar los descriptores más similares entre sí. En general, las clasificaciones obtenidas son coherentes con las realizadas previamente en los estudios de Castro *et al.* (2013) y de Jeltema y Southwick (1986).

Estudio las dos dimensiones olfativas más importantes en perfumería

En el estudio del Atlas de Dravnieks (1985) por medio de PCA ha quedado de manifiesto que las dos dimensiones fundamentales del espacio perceptual de fragancias también se encuentran subyacentes. Una de estas dimensiones discrimina los olores como frescos frente a cálidos y la otra dimensión ortogonal clasifica los olores como femeninos o masculinos.

Se ha propuesto un mapa sensorial de los descriptores en función de las dos dimensiones latentes del espacio perceptual de fragancias. Este mapa es consistente con otros publicados previamente (Zarzo y Stanton, 2009; Zarzo, 2013). Una de las dificultades encontradas es que los olores masculinos suelen aparecer opuestos a los femeninos en múltiples estudios, pero en el Atlas de Dravnieks aparecen como ortogonales. Para resolver este inconveniente a efectos de construir los mapas sensoriales, se han propuesto dos métodos que consiguen resultados satisfactorios a la hora de reproducir una única componente que contrapone los olores masculinos frente a los femeninos.

Estudio del carácter hedónico de los descriptores

Diversos estudios han sugerido que la primera componente principal (PC1) del Atlas de Dravnieks puede interpretarse como la dimensión hedónica (Khan *et al.*, 2007; Koulakov *et al.*, 2011). El hecho de que esta componente sea la más importante en el modelo de regresión obtenido es parcialmente consistente con esta teoría. Sin embargo, han aparecido otras variables latentes cuyo efecto en el modelo es claramente significativo. Esto parece indicar que la interpretación de PC1 como la dimensión hedónica es discutible. En este trabajo se interpreta PC1 de una forma distinta, como un factor que contrapone los olores más típicos de fragancias femeninas frente a olores desagradables.

Múltiples estudios han encontrado que cuando se evalúa un conjunto amplio de olores, tanto agradables como desagradables, generalmente la primera componente refleja el carácter hedónico. Por este motivo, se había considerado hasta ahora que PC1 del Atlas de Dravnieks reflejaba de hecho el tono hedónico. Sin embargo, los resultados obtenidos ponen en duda esta cuestión.

Estudio de descriptores olfativos ambiguos

El estudio realizado comparando los descriptores “fragante” y “aromático”, sugiere que el descriptor “fragante” fue preferentemente aplicado por el panel de catadores de Dravnieks (1985) para describir notas típicas de perfumes femeninos. También sugiere que dicho panel tuvo cierta tendencia a describir las notas masculinas como “aromático”, más que como “fragante”, como ha quedado de manifiesto por la mayor asociación de las notas canforáceas y especiadas en el descriptor “aromático”. Este criterio es el más aceptado en perfumería, y de hecho en la clasificación de Edwards (2010), la familia de perfumes denominada “aromático” o “fougère” es la que comprende la mayor cantidad de perfumes masculinos.

El estudio realizado que compara los descriptores “perfumería” y “colonia”, sugiere que ambos descriptores fueron interpretados por el panel de catadores de Dravnieks (1985) prácticamente como sinónimos. No obstante, se ha detectado que el panel de catadores tuvo cierta tendencia a describir las notas afrutadas como “perfumería”, más que como “colonia”, mientras que las notas cítricas tuvieron cierta tendencia a ser descritas como “colonia”, más que como “perfumería”. Este resultado tiene sentido ya que el término colonia alude al Agua de Colonia, que originalmente era una marca de perfumes de carácter cítrico.

6 Referencias

Abe, H.; Kanaya, S.; Komukai, T.; Takahashi, Y.; Sasaki, S. (1990) Systemization of semantic descriptions of odors. *Anal. Chim. Acta*, 239, 73-85.

Arctander, S. (1969) *Perfume and Flavor Chemicals (Aroma Chemicals)*, Vol. I and II. S. Arctander publisher, Montclair, NJ.

Boelens, H.; Haring, H.G. (1981) *Molecular Structure and Olfactive Quality*. Internal Report, Naarden International, Bussum, The Netherlands.

Brud, W.S. (1986) Words versus odours: how perfumers communicate. *Perf. Flav.*, 11 (4), 27-44.

Burdock, G.A. (2004) *Fenaroli's Handbook of Flavor Ingredients*, fifth edition. CRC Press, Boca Raton, FL.

Castro J.B.; Ramanathan A.; Chennubhotla C.S. (2013) Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. *PLoS One*. 8(9):e73289.

Chastrette, M.; Elmouaffek, A.; Sauvegrain, P. (1988) A multidimensional statistical study of similarities between 74 notes used in perfumery. *Chem. Senses*, 13, 295-305.

Chastrette, M.; de Sain Laumer, J.Y.; Sauvegrain, P. (1991) Analysis of a system of description of odors by means of four different multivariate statistical methods. *Chem. Senses*, 16, 81-93.

Crocker E.C.; Henderson L.F. (1927) Analysis and classification of odors: an effort to develop a workable method. *Am. Perf. Essent. Oil Rev.*, 22, 325- 356.

Würtz, D.; Chalabi, Y.; Chen, W.; Ellis, A. (2009) *Portfolio Optimization with R/Rmetrics*, p. 131. <https://books.google.es/books?id=keDLpq86DU4C>, Rmetrics: Zurich.

Dravnieks A. (1982) Odor quality: semantically generated multidimensional profiles are stable. *Science*, 218, 799-801.

Dravnieks, A.; Bock, F.C.; Powers, J.J.; Tibbetts, M.; Ford, M. (1978) Comparison of odors directly and through profiling. *Chem. Senses Flav.*, 3, 191-225.

Dravnieks, A. (1985) *Atlas of Odor Character Profiles, Data Series DS 61*. American Society for Testing and Materials, Philadelphia, PA.

Dravnieks, A.; Masurat, T.; Lamm, R.A. (1984) Hedonics of odors and odor descriptors. *J. Air Pollut. Control Assoc.*, 34, 752-755.

Edwards M. (2010) *Fragrances of the World: Parfums du Monde 2010*. 26th ed. Sydney (Australia): *Fragrances of the World*.

- Ennis, D.M.; Boelens, H.; Haring, H.; Bowman, P. (1982) Multivariate analysis in sensory evaluation. *Food Technology*, 36(11), 83-90.
- Esbensen, K.; Geladi, P. (1987) Principal Component Analysis. *Chemometric and Intelligent Laboratory Systems*, 1, 41-56.
- Fukuyama, Y.; Sugeno, M. (1989) A new method of choosing the number of clusters for the fuzzy c-means method. In *Proc. 5th Fuzzy Syst. Symp* (Vol. 247, pp. 247-250).
- Harper, R.; Bate Smith, E.C.; Land, D.G. (1968) *Odour Description and Odour Classification: a Multidisciplinary Examination*. Elsevier, New York.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6 & 7), 417–441 & 498–520.
- Jaubert, J.N.; Gordon, G.; Dore, J.C. (1986) Classification of odours and their sensorial perception. *Quintessenza*, 5, 27-42.
- Jaubert, J.N.; Gordon, G.; Dore, J.C. (1987) Une organisation du champ des odeurs. II. Modèle descriptif de l'organisation de l'espace odorant. *Parfum, Cosmétiques, Aromes*, 78, 71-82.
- Jeltema, M.A.; Southwick, E.W. (1986) Evaluation and applications of odor profiling. *J Sens Stud.*, 1(2), 123-136.
- Jolliffe, I.T. (2002). *Principal Component Analysis*, 2nd ed. Springer-Verlag: New York.
- Khan, R.M.; Luk, C.H.; Flinker, A.; Aggarwal, A.; Lapid, H. (2007) Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. *J. Neurosci*, 27, 10015–10023.
- Koulakov, A.A.; Kolterman, B.E.; Enikolopov, A.G.; Rinberg, D. (2011) In search of the structure of human olfactory space. *Front. Syst. Neurosci.*, 5, 65.
- Linnaeus, C. (1756) *Odores medicamentorum*. In: *Amoenitates Academicae*, Vol. 3. Stockholm (Sweden): Lars Salvius. pp. 183–201.
- Madany-Mamlouk, A.; Chee-Ruiter, C.; Hofmann, U.G.; Bower, J.M. (2003) Quantifying olfactory perception: mapping olfactory perception space by using multidimensional scaling and self-organizing maps. *Neurocomputing*, 52-54, 591-597.
- Madany-Mamlouk, A.; Martinetz, T. (2004) On the dimensions of the olfactory perception space. *Neurocomputing*, 58-60, 1019-1025.
- Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. (2014) Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.6-4. <http://CRAN.R-project.org/package=e1071>

- Moreda-Piñeiro, A.; Marcos, A.; Fisher, A.; Hill, S. (2001) Evaluation of the effect of data pre-treatment procedures on classical pattern recognition and principal components analysis: a case study for the geographical classification of tea. *Journal of Environmental Monitoring*, 3, 352-360.
- Moskowitz, H.R.; Gerbers, C.L. (1974) Dimensional salience of odors. *Ann. N.Y. Acad. Sci.*, 237, 1-16.
- Pal, N. R.; Bezdek, J. C.; Hathaway, R. J. (1996) Sequential competitive learning and the fuzzy c-means clustering algorithms. *Neural Networks*, 9(5), 787-796.
- Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11), 559-572.
- Pintore, M.; Wechman, C.; Sicard, G.; Chastrette, M.; Amaury, N.; Chretien, J.R. (2006) Comparing the information content of two large olfactory databases. *J. Chem. Inf. Model*, 46, 32-38.
- PMP (2001) Database of Perfumery Materials and Performance. BACIS, The Netherlands.
- R Core Team and Contributors Worldwide (2016) R Documentation, Version 3.2.3
- Schutz, H.G. (1964) A matching-standard method for characterizing odor qualities. *Ann. N.Y. Acad. Sci.*, 116, 517-526.
- Sigma-Aldrich (2005) Flavors and Fragrances 2005-2006 Catalog. Sigma-Aldrich Fine Chemicals Company, Milwaukee, WI.
- Wright, R.H.; Michels, K.M. (1964) Evaluation of far infrared relation to odors by a standard similarity method. *Ann. N.Y. Acad. Sci.*, 116, 535-551.
- Xie, X. L.; Beni, G. (1991) A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 841-847.
- Yoshida, M. (1975) Psychometric classification of odors. *Chem. Senses Flav.*, 1, 443-464.
- Zarzo, M.; Stanton, D. T. (2006) Identification of latent variables in a semantic odor profile database using principal component analysis. *Chemical Senses*, 31, 713-724.
- Zarzo, M. (2008) Psychologic dimensions in the perception of everyday odors: pleasantness and edibility. *Journal of Sensory Studies*, 23(3), 354-376.
- Zarzo, M.; Stanton, D.T. (2009) Understanding the underlying dimensions in perfumers' odor perception space as a basis for developing meaningful odor maps. *Atten. Percept. Psychophys.*, 71(2), 225-247.
- Zarzo, M. (2015) A sensory 3D map of the odor description space derived from a comparison of numeric odor profile databases. *Chemical Senses*, 40 (5), 305-313.

Anexos

Anexo 1: probabilidad de pertenencia de cada descriptor a cada uno de los clústeres, señalado aquel clúster cuya probabilidad de pertenencia es mayor.

Descriptor	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6	Clúster 7	Clúster 8
alcoholic	0,0360	0,0324	0,0870	0,0292	0,1856	0,2541	0,2851	0,0906
almond	0,0159	0,0053	0,0930	0,0092	0,0913	0,0106	0,7455	0,0291
ammonia	0,0398	0,0249	0,0960	0,0262	0,3216	0,1654	0,2565	0,0696
animal	0,8246	0,0065	0,0657	0,0215	0,0307	0,0076	0,0252	0,0181
anise (licorice)	0,0010	0,0005	0,0042	0,0007	0,0040	0,0010	0,9855	0,0031
aromatic	0,0289	0,5352	0,0463	0,0547	0,0443	0,0977	0,0746	0,1184
bakery (fresh bread)	0,0233	0,0048	0,1036	0,0085	0,7240	0,0112	0,1058	0,0188
banana	0,0042	0,0019	0,0222	0,0033	0,0141	0,0032	0,9371	0,0141
bark, birch bark	0,0030	0,0027	0,0037	0,9772	0,0018	0,0010	0,0031	0,0075
beany	0,0958	0,0296	0,1435	0,4235	0,0452	0,0181	0,0853	0,1591
beery	0,0317	0,0194	0,1297	0,0397	0,0552	0,0202	0,3873	0,3169
bitter	0,0828	0,0123	0,1383	0,0214	0,5822	0,0304	0,0973	0,0352
black pepper	0,0323	0,0084	0,3178	0,0180	0,0907	0,0130	0,4659	0,0540
blood, raw meat	0,8854	0,0027	0,0605	0,0095	0,0155	0,0031	0,0141	0,0092
burnt candle	0,0323	0,0081	0,3183	0,0171	0,1011	0,0131	0,4602	0,0499
burnt milk	0,0167	0,0029	0,0564	0,0052	0,8562	0,0071	0,0450	0,0105
burnt paper	0,0091	0,0023	0,0229	0,0035	0,9203	0,0072	0,0276	0,0071
burnt rubber	0,0035	0,0007	0,0101	0,0012	0,9701	0,0020	0,0100	0,0024
burnt, smoky	0,0442	0,0098	0,0846	0,0151	0,7062	0,0300	0,0828	0,0273
buttery, fresh butter	0,0515	0,0075	0,4626	0,0159	0,2553	0,0138	0,1602	0,0332
cadaverous (dead animal)	0,8275	0,0037	0,0924	0,0120	0,0260	0,0046	0,0211	0,0126
camphor	0,0291	0,0430	0,0637	0,0293	0,1002	0,4576	0,1809	0,0961
cantaloupe, honey dew melon	0,0198	0,0167	0,0616	0,0378	0,0269	0,0125	0,1168	0,7079
caramel	0,0296	0,0097	0,0669	0,0131	0,7160	0,0382	0,0988	0,0276
caraway	0,0085	0,0056	0,0329	0,0077	0,0301	0,0105	0,8648	0,0397
cardboard	0,0815	0,0249	0,0881	0,6127	0,0327	0,0141	0,0536	0,0925
cat urine	0,1113	0,0049	0,7513	0,0156	0,0435	0,0064	0,0453	0,0217
cedarwood	0,0244	0,0224	0,0821	0,0347	0,0473	0,0230	0,3224	0,4435
celery	0,0436	0,0122	0,3738	0,0308	0,0727	0,0152	0,3514	0,1003
chalky	0,0747	0,0255	0,0964	0,5808	0,0338	0,0144	0,0602	0,1142
chemical	0,0157	0,0248	0,0259	0,0149	0,0407	0,8018	0,0447	0,0315
cherry (berry)	0,0011	0,0016	0,0029	0,0027	0,0016	0,0009	0,0058	0,9834
chocolate	0,0303	0,0078	0,2474	0,0152	0,1371	0,0140	0,5051	0,0432
cinnamon	0,0015	0,0007	0,0075	0,0011	0,0068	0,0013	0,9773	0,0040
cleaning fluid	0,0315	0,0536	0,0492	0,0307	0,0727	0,6234	0,0788	0,0603
clove	0,0046	0,0023	0,0198	0,0033	0,0242	0,0051	0,9278	0,0128
coconut	0,0170	0,0058	0,1160	0,0113	0,0528	0,0093	0,7467	0,0410
coffee	0,0117	0,0028	0,0446	0,0047	0,8567	0,0074	0,0614	0,0106
cologne	0,0127	0,7798	0,0204	0,0424	0,0152	0,0181	0,0292	0,0822
cooked vegetables	0,1430	0,0150	0,5054	0,0687	0,0605	0,0145	0,1071	0,0858
cool, cooling	0,0370	0,1580	0,0719	0,0548	0,0786	0,2480	0,1593	0,1924
cork	0,1027	0,0288	0,1368	0,4540	0,0440	0,0176	0,0786	0,1376
creosote	0,0082	0,0108	0,0160	0,0074	0,0285	0,8742	0,0354	0,0194
crushed grass	0,0461	0,1176	0,0590	0,5103	0,0347	0,0258	0,0592	0,1472
crushed weeds	0,0336	0,0656	0,0446	0,6515	0,0244	0,0168	0,0436	0,1197
dill	0,0336	0,0208	0,1311	0,0464	0,0521	0,0197	0,2975	0,3989
dirty linen	0,7388	0,0108	0,0913	0,0420	0,0393	0,0111	0,0367	0,0300
disinfectant, carbolic	0,0015	0,0024	0,0026	0,0014	0,0042	0,9796	0,0050	0,0034
dry, powdery	0,0492	0,0193	0,0584	0,7278	0,0224	0,0101	0,0384	0,0743
eggy (fresh eggs)	0,0416	0,0028	0,8773	0,0092	0,0223	0,0035	0,0293	0,0139
etherish, anaesthetic	0,0128	0,0212	0,0214	0,0124	0,0332	0,8344	0,0377	0,0269
eucaliptus	0,0341	0,0410	0,0874	0,0342	0,1266	0,1672	0,3694	0,1400
fecal (like manure)	0,8968	0,0025	0,0535	0,0091	0,0138	0,0028	0,0129	0,0086
fermented (rotten) fruit	0,4320	0,0178	0,2344	0,1010	0,0572	0,0159	0,0735	0,0683
fishy	0,0179	0,0021	0,9123	0,0064	0,0167	0,0027	0,0299	0,0120
floral	0,0111	0,8202	0,0172	0,0379	0,0128	0,0152	0,0237	0,0620
fragrant	0,0337	0,5633	0,0479	0,0684	0,0436	0,0749	0,0657	0,1025
fresh green vegetables	0,0066	0,0069	0,0085	0,9452	0,0041	0,0024	0,0073	0,0189
fresh tobacco smoke	0,0155	0,0051	0,1007	0,0092	0,0699	0,0093	0,7602	0,0301
fried chicken	0,0321	0,0058	0,6984	0,0145	0,0581	0,0081	0,1480	0,0351
fruity, citrus	0,0070	0,0169	0,0157	0,0216	0,0092	0,0068	0,0277	0,8951
fruity, other than citrus	0,0250	0,0617	0,0480	0,1309	0,0261	0,0184	0,0651	0,6248
garlic, onion	0,0641	0,0075	0,4886	0,0164	0,2533	0,0137	0,1250	0,0314
gasoline, solvent	0,0062	0,0078	0,0110	0,0054	0,0198	0,9168	0,0208	0,0123
geranium leaves	0,0453	0,2629	0,0627	0,2934	0,0401	0,0349	0,0709	0,1899

Descriptor	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6	Clúster 7	Clúster 8
grainy (as grain)	0,1323	0,0250	0,3108	0,1568	0,0641	0,0198	0,1291	0,1621
grape juice	0,0329	0,0210	0,1233	0,0480	0,0490	0,0191	0,2613	0,4454
grapefruit	0,0020	0,0040	0,0048	0,0057	0,0027	0,0019	0,0089	0,9700
green pepper	0,0584	0,0344	0,1211	0,2799	0,0426	0,0191	0,1025	0,3419
hay	0,0534	0,0558	0,0565	0,6445	0,0310	0,0196	0,0475	0,0917
heavy	0,0806	0,0115	0,1395	0,0203	0,5916	0,0279	0,0949	0,0336
herbal, green, cut grass	0,0487	0,3128	0,0642	0,2535	0,0431	0,0394	0,0719	0,1665
honey	0,0046	0,0067	0,0122	0,0099	0,0069	0,0043	0,0276	0,9278
household gas	0,0262	0,0049	0,1156	0,0089	0,7153	0,0112	0,0987	0,0191
incense	0,0313	0,2618	0,0609	0,0687	0,0506	0,0728	0,1205	0,3336
kerosene	0,0075	0,0103	0,0147	0,0069	0,0257	0,8838	0,0327	0,0183
kippery (smoked fish)	0,0149	0,0020	0,9020	0,0054	0,0222	0,0029	0,0393	0,0113
laurel leaves	0,0562	0,3114	0,0702	0,2363	0,0491	0,0462	0,0766	0,1539
lavender	0,0053	0,9157	0,0082	0,0166	0,0063	0,0078	0,0115	0,0286
leather	0,0336	0,0394	0,0877	0,0335	0,1266	0,1537	0,3865	0,1390
lemon	0,0020	0,0033	0,0050	0,0047	0,0028	0,0019	0,0104	0,9699
light	0,0255	0,6445	0,0378	0,0525	0,0345	0,0624	0,0544	0,0885
malty	0,0422	0,0082	0,3281	0,0161	0,2968	0,0160	0,2547	0,0379
maple syrup	0,0164	0,0051	0,0426	0,0072	0,8266	0,0183	0,0678	0,0160
meaty (cooked, good)	0,0935	0,0099	0,2219	0,0197	0,4996	0,0204	0,1020	0,0330
medicinal	0,0001	0,0001	0,0001	0,0000	0,0002	0,9993	0,0002	0,0001
metallic	0,0171	0,0131	0,0578	0,0155	0,0685	0,0302	0,7241	0,0737
minty, peppermint	0,0297	0,0355	0,0840	0,0330	0,0943	0,0801	0,4635	0,1799
molasses	0,0267	0,0089	0,0628	0,0121	0,7312	0,0349	0,0973	0,0259
mothballs	0,0277	0,0202	0,0838	0,0221	0,1545	0,0725	0,5378	0,0815
mouse	0,6741	0,0061	0,1853	0,0190	0,0484	0,0079	0,0380	0,0213
mushroom	0,0616	0,0244	0,2324	0,0900	0,0604	0,0202	0,2009	0,3101
musk	0,0270	0,0200	0,0931	0,0438	0,0398	0,0170	0,2002	0,5590
musty, earthy, moldy	0,0312	0,0165	0,0322	0,8311	0,0147	0,0076	0,0232	0,0436
nail polish remover	0,0068	0,0085	0,0129	0,0060	0,0238	0,9001	0,0271	0,0148
new rubber	0,0341	0,0288	0,0690	0,0252	0,1681	0,4635	0,1494	0,0619
nutty	0,0274	0,0019	0,9071	0,0054	0,0234	0,0027	0,0233	0,0088
oak wood, cognac	0,0196	0,0595	0,0447	0,0436	0,0318	0,0297	0,1015	0,6695
oily, fatty	0,0083	0,0007	0,9709	0,0019	0,0063	0,0009	0,0079	0,0032
orange	0,0059	0,0069	0,0165	0,0129	0,0083	0,0046	0,0336	0,9111
paint	0,0061	0,0117	0,0110	0,0064	0,0165	0,9112	0,0216	0,0155
peach (fruit)	0,0333	0,0750	0,0594	0,2438	0,0313	0,0214	0,0722	0,4637
peanut butter	0,0264	0,0033	0,8243	0,0079	0,0542	0,0051	0,0629	0,0160
pear	0,0008	0,0012	0,0019	0,0023	0,0010	0,0006	0,0034	0,9888
perfumery	0,0009	0,9863	0,0014	0,0025	0,0011	0,0016	0,0020	0,0042
pineapple	0,0356	0,0202	0,1472	0,0458	0,0561	0,0200	0,3351	0,3400
popcorn	0,0465	0,0079	0,3810	0,0161	0,2947	0,0151	0,2030	0,0357
putrid, foul, decayed	0,9268	0,0023	0,0307	0,0075	0,0129	0,0028	0,0102	0,0069
raisins	0,0073	0,0030	0,0412	0,0054	0,0236	0,0049	0,8924	0,0222
rancid	0,6676	0,0145	0,1105	0,0464	0,0590	0,0164	0,0486	0,0370
raw cucumber	0,0449	0,0358	0,0934	0,2397	0,0366	0,0180	0,0919	0,4397
raw potato	0,0673	0,0334	0,1313	0,3322	0,0442	0,0190	0,0997	0,2728
rope	0,2317	0,0284	0,1946	0,2690	0,0590	0,0206	0,0871	0,1095
rose	0,0198	0,0656	0,0394	0,0778	0,0234	0,0185	0,0610	0,6946
sauerkraut	0,1486	0,0058	0,6942	0,0189	0,0490	0,0074	0,0508	0,0253
seasoning (for meat)	0,0358	0,0061	0,1689	0,0114	0,6241	0,0132	0,1164	0,0239
seminal, sperm-like	0,0411	0,0094	0,5094	0,0237	0,0709	0,0124	0,2668	0,0663
sewer	0,5597	0,0095	0,2087	0,0258	0,0944	0,0138	0,0579	0,0303
sharp, pungent, acid	0,0474	0,0197	0,0961	0,0238	0,5119	0,0997	0,1515	0,0498
sickening	0,4820	0,0182	0,1714	0,0441	0,1323	0,0270	0,0782	0,0468
soapy	0,0050	0,0086	0,0128	0,0112	0,0076	0,0052	0,0287	0,9209
sooty	0,0449	0,0067	0,5700	0,0146	0,1664	0,0116	0,1545	0,0313
soupy	0,0420	0,0027	0,8697	0,0075	0,0318	0,0038	0,0305	0,0119
sour milk	0,2397	0,0131	0,4712	0,0583	0,0591	0,0132	0,0835	0,0619
sour, vinegar	0,9070	0,0030	0,0399	0,0119	0,0130	0,0032	0,0125	0,0095
spicy	0,0180	0,0123	0,0613	0,0147	0,0901	0,0333	0,7101	0,0603
stale	0,9764	0,0007	0,0102	0,0027	0,0036	0,0008	0,0032	0,0023
stale tobacco smoke	0,0839	0,0068	0,6175	0,0163	0,1493	0,0113	0,0873	0,0277
strawberry	0,0244	0,0191	0,0719	0,0542	0,0287	0,0131	0,1058	0,6829
sulfidic	0,1898	0,0121	0,3184	0,0270	0,2867	0,0218	0,1040	0,0402

Descriptor	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6	Clúster 7	Clúster 8
sweaty	0,9992	0,0000	0,0004	0,0001	0,0001	0,0000	0,0001	0,0001
sweet	0,0034	0,9476	0,0053	0,0084	0,0045	0,0069	0,0079	0,0160
tar	0,0286	0,0269	0,0584	0,0224	0,1306	0,5433	0,1326	0,0571
tea leaves	0,0274	0,1201	0,0586	0,0597	0,0454	0,0519	0,1287	0,5082
turpentine (pine oil)	0,0236	0,0398	0,0490	0,0246	0,0744	0,5862	0,1256	0,0767
urine	0,0962	0,0053	0,7624	0,0185	0,0377	0,0064	0,0481	0,0255
vanilla	0,0187	0,0059	0,1144	0,0105	0,1065	0,0118	0,6997	0,0325
varnish	0,0064	0,0114	0,0111	0,0064	0,0171	0,9119	0,0208	0,0148
violets	0,0282	0,1603	0,0522	0,1127	0,0333	0,0299	0,0793	0,5041
warm	0,0233	0,0080	0,0817	0,0116	0,6026	0,0245	0,2180	0,0302
wet paper	0,0366	0,0213	0,0372	0,7993	0,0177	0,0094	0,0275	0,0511
wet wool, wet dog	0,8063	0,0058	0,0807	0,0171	0,0384	0,0076	0,0273	0,0170
woody, resinous	0,0331	0,0522	0,0618	0,2430	0,0299	0,0184	0,0717	0,4898
yeasty	0,7542	0,0077	0,1070	0,0344	0,0305	0,0077	0,0324	0,0261

Anexo 2: Puntuación media obtenida para cada descriptor de los 139 seleccionados para las 160 muestras:

Descriptor	Media puntuación
fruity, citrus	4,20
lemon	3,52
grapefruit	0,68
orange	2,00
fruity, other than citrus	7,67
pineapple	1,47
grape juice	1,16
strawberry	1,73
pear	1,09
cantaloupe, honey dew melon	1,40
peach (fruit)	1,04
banana	0,95
floral	8,47
rose	3,26
violets	1,65
lavender	2,05
cologne	2,97
musk	3,13
perfumery	10,22
fragrant	23,88
aromatic	20,73
honey	1,38
cherry (berry)	2,67
almond	4,69
nail polish remover	1,88
nutty	2,65
spicy	7,61
clove	2,36
cinnamon	2,80
tea leaves	0,85
seasoning (for meat)	0,96
black pepper	0,84
green pepper	0,60
dill	0,89
caraway	0,46
oak wood, cognac	1,67
woody, resinous	13,19
cedarwood	5,41
mothballs	2,71
minty, peppermint	5,26
camphor	4,79
eucalyptus	2,82
chocolate	0,65
vanilla	1,77
sweet	19,10
maple syrup	0,99
caramel	1,08

Descriptor	Media puntuación
malty	1,28
raisins	0,70
molasses	0,57
coconut	1,25
anise (licorice)	2,57
alcoholic	2,58
etherish, anaesthetic	6,38
cleaning fluid	2,92
gasoline, solvent	3,59
turpentine (pine oil)	4,71
geranium leaves	1,23
celery	0,81
fresh green vegetables	1,32
crushed weeds	1,76
crushed grass	1,74
herbal, green, cut grass	8,43
raw cucumber	0,59
heno	1,06
grainy (as grain)	1,01
yeasty	1,39
bakery (fresh bread)	0,27
sour milk	0,94
fermented (rotten) fruit	1,28
soapy	3,15
leather	1,74
cardboard	1,68
wet paper	1,11
wet wool, wet dog	1,53
dirty linen	1,43
stale	3,20
musty, earthy, moldy	8,74
raw potato	0,70
mouse	0,60
mushroom	0,97
peanut butter	1,36
beany	0,37
bark, birch bark	2,81
cork	1,28
burnt, smoky	6,57
fresh tobacco smoke	0,33
incense	3,39
coffee	0,41
stale tobacco smoke	1,11
burnt paper	1,37
burnt milk	0,24
burnt rubber	1,19

Descriptor	Media puntuación
tar	2,15
creosote	2,21
disinfectant, carbolic	5,29
medicinal	7,42
chemical	9,46
bitter	2,09
sharp, pungent, acid	10,50
sour, vinegar	6,07
sauerkraut	0,47
ammonia	1,66
urine	1,44
cat urine	0,96
fishy	2,23
kippery (smoked fish)	0,72
seminal, sperm-like	0,40
new rubber	1,18
sooty	0,87
burnt candle	1,00
kerosene	1,32
oily, fatty	7,27
buttery, fresh butter	1,04
paint	4,38
varnish	2,55
popcorn	1,26
fried chicken	0,20
meaty (cooked, good)	1,21
cooked vegetables	0,88
rancid	5,68
sweaty	6,83
household gas	1,36
sulfidic	2,66
garlic, onion	3,22
metallic	2,51
blood, raw meat	1,29
animal	2,07
sewer	1,95
putrid, foul, decayed	5,47
fecal (like manure)	2,35
cadaverous (dead animal)	1,59
sickening	11,73
dry, powdery	4,50
chalky	0,59
light	14,34
heavy	15,12
cool, cooling	8,82
warm	8,76