Food Quality and Preference xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

A Factorial Approach for Sorting Task data (*FAST*)

Marine Cadoret, Sébastien Lê *, Jérôme Pagès

Applied Mathematics Department, Agrocampus Rennes, 65 rue de Saint-Brieuc, 35042 Rennes Cedex, France

ARTICLE INFO

Article history:
Received 10 May 2008
Received in revised form 19 January 2009
Accepted 20 February 2009
Available online xxxx

Keywords:
Sorting task
Qualified categorization
Multiple correspondence analysis
Multiple factor analysis
Ellipses
DISTATIS
Perfumes

ABSTRACT

The aim of this paper is to present a new approach to analyze categorization data called *FAST* that stands for Factorial Approach for Sorting Task data. This approach, based on multiple correspondence analysis (MCA), provides an optimal representation of the products, an optimal representation of the consumers, which are to be interpreted jointly. It provides also elements of validation based on confidence ellipses. In the case of "labelled" categorization, where a verbalization task is asked to describe the groups of products, it provides an optimal representation of the words which is directly linked to both representations of the products and of the consumers. The *FAST* approach will be illustrated by an example where 98 consumers were asked to group 12 luxury perfumes.

1. Introduction

Sorting task or categorization is a cognitive process that has been used for a long time (e.g. Healy & Miller, 1970) to collect data, in particular in sensory analysis where one seeks to describe products according to their sensory properties. This task consists in asking assessors to group products in function of their sensory resemblances. Following this task, a verbalization task can also be asked to describe the groups (or categories) and to supplement the original data: in this case we will speak about "labelled" categorization. Categorization is becoming more and more popular since it does not need to be performed by trained assessors. A presentation of the method and a complete bibliography can be found in Abdi, Valentin, Chollet, and Chrea (2007).

Sorting task data can be analyzed with numerous methods. The current practice consists in using multidimensional scaling (MDS) (Torgerson, 1958). For example, during the last two years, all the papers published in Food Quality and Preference about free sorting data used MDS (or DISTATIS for the paper of Lelièvre et al.): Ballester, Patris, Symoneaux, & Valentin, 2008; Blancher et al., 2007; Cartier et al., 2006; Faye et al., 2006; Lelièvre, Chollet, Abdi, & Valentin, 2008; Parr, Green, White, & Sherlock, 2007. In 2007, Abdi et al. propose a new method called DISTATIS which combines MDS and STATIS (Lavit, 1976; Robert & Escoufier, 1976). In addition to a representation of the products, this method provides a representation of the consumers. It also allows a representation of the words associated with the groups provided by the consumers in the case of "labelled" categorization, as well as a representation of the variability around the products (Abdi & Valentin, 2007).

The aim of this paper is to present a new approach to analyze categorization data called *FAST* that stands for Factorial Approach for Sorting Task data. This approach is based on multiple correspondence analysis (MCA) (Greenacre, 1984; Lebart, 1975), also known as Homogeneity analysis (Tenenhaus & Young, 1985). The idea of using MCA to analyze categorization data is not new (a comparison between MDS and MCA can be found in Van der Kloot and Van Herk (1991): on the suggested examples, the two methods provide similar results) but doesn't seem to be used in practice (see references above in Food Quality and Preference). The main interest of our approach is that it complements MCA with different results and representations often needed in sensory analysis. To the representation of the products and the categories provided by MCA, the FAST approach adds a representation of the consumers directly linked to the two previous ones. It provides also elements of validation by means of confidence ellipses. In the case of "labelled" categorization, the representation of the words coincides with the one of the categories.

The *FAST* approach will be illustrated by an example where 98 consumers were asked to group twelve luxury perfumes and to describe their respective groups. In this article, this approach will also be compared to DISTATIS, which is at the same time the most recent method and the closest one to ours in terms of objectives. This comparison will be done both on the perfumes data and on the data used in the DISTATIS paper.

* Corresponding author. Tel.: +33 2 23 48 58 81; fax: +33 2 23 48 58 71.
E-mail addresses: sebastien.le@agrocampus-ouest.fr, Sebastien.Le@agrocampus-rennes.fr (S. Lê).

0950-3293/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2009.02.010

2. Data

Ninety eight consumers carried out a “labelled” categorization on twelve luxury perfumes: *Angel*, *Aromatics Elixir*, *Chanel n°5*, *Cinéma*, *Coco Mademoiselle*, *L'instant*, *Lolita Lempicka*, *Pleasures*, *Pure Poison*, *Shalimar*, *J'adore* (eau de parfum), *J'adore* (eau de toilette). The consumers were mostly women (69.4%) and rather young (mean age = 25 years; range: 18–58).

Consumers were placed in individual booths, each perfume was sprayed on a small piece of cotton wool placed in a pill box, and all twelve pill boxes were presented at each consumer. The pill boxes were aligned and ordered according to Williams Latin squares. Consumers were asked to evaluate the products in this order and they were allowed to return to a sample; they were asked to make at least two groups of perfumes and at most eleven groups. Following this task, they were also asked to carry out a verbalization task to describe the groups.

Data can be gathered in a table with I rows and J columns, in which each row i corresponds to a perfume ($I = 12$), each column j corresponds to a consumer ($J = 98$) and the cell (i, j) corresponds to the number of the group to which perfume i belongs to for consumer j. Each consumer j can be assimilated to a qualitative variable with K_j categories, where K_j denotes the number of groups used by consumer j; in the case of a “labelled” categorization, the label of each category is the sequence of words associated with each one of the K_j groups.

3. The fast approach

3.1. Preliminary study

Data can be first analyzed with bar plots that describe the number of groups per consumer, or the number of perfumes per group. These bar plots indicate if the consumers made many groups or not, and if the groups comprise many perfumes or a few.

A synthetic summary of the associations between perfumes is provided by the co-occurrences matrix, of dimension $I \times I$, between perfumes: the cell (i, j) corresponds to the number of consumers that have put perfumes i and j in the same group. This matrix makes it possible to see the most frequent associations between the perfumes. Let us notice in passing that this matrix is the sum of the individual co-occurrences matrices in which the cell (i, j) is equal to1 if perfumes i and j are in the same group and 0 if not.

In the case of the “labelled categorization”, an assessment on the use of the words is provided by several numerical and graphical indicators as the total number of words used, the number of words per group (respectively per perfume), the number of consumers who did not use any word and the number of “unlabelled” groups.

A first characterization of the perfumes is obtained by listing for each perfume the words that are both frequently used to characterize one perfume and rarely used to characterize the others. These two pieces of information are integrated in a test which compares the number of times that a word was associated to a perfume to the number of times that it was used for all the perfumes (cf. Lebart, Piron, & Morineau, 2006, pp. 292–294). For each perfume, the words are sorted according to the p-values associated with this test.

3.2. Representation of the perfumes and the words

By construction, the data table can be typically analyzed by multiple correspondence analysis (MCA) which is dedicated to the analysis of qualitative variables. In the analysis, these data are taken into account via the so called disjunctive table which comprises here I lines and $K = \sum K_j$ columns: the consumer j is represented by a set of K_j dummy variables, each variable corresponding to a group: a dummy variable associated with a group is equal to 1 if the perfume belongs to the group and 0 if not (cf. Fig. 1).

MCA analyzes the set of perfumes from the disjunctive table. This set of perfumes lies in a K-dimensional space. In this space, the distance between two perfumes i and l is thus defined:

$$d^2(i, l) = \frac{1}{j} \sum_k \frac{I}{I_k} (x_{ik} - x_{lk})^2$$

where x_{ik} is equal to 1 if perfume i belongs to group k and 0 if not; I_k denotes the number of perfumes in the group k.

For this distance, (1) two perfumes i and l are superimposed if they were put together by all the consumers and (2) two perfumes i and l are all the more close as they were placed in the same group by a great many of consumers. Conversely, two products are all the more distant as they were placed in two different groups by a great many of consumers.

More precisely a group k contributes to this distance in a way inversely proportional to its size (denoted I_k): the assignment to a group of small size moves away a perfume from all the others. This is illustrated by the example Fig. 2. In this example, assessor 1 contributes three times more to the distance between products A and B than assessor 2. This property specific to MCA seems to be adapted to categorization data, in the following way: assessor 1 considers A and B both as very particular and not having the same characteristic, whereas consumer 2 considers only A and B as being different (and not very particular).

In the borderline case when a product is always alone, the reordered disjunctive table reveals a four-block structure on the data

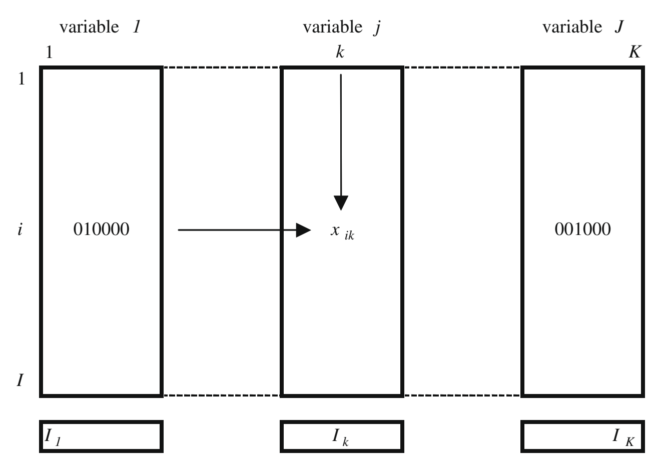

Fig. 1. Disjunctive table, x_{ik} is equal to 1 if perfume i belongs to group k; I_k is the number of perfumes in the group k.

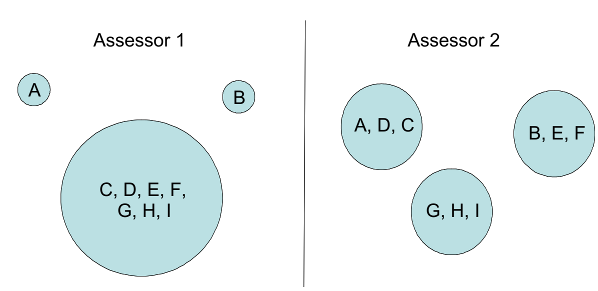

Fig. 2. About nine perfumes categorized by two assessors that have made three groups each (a disc represents a group). The two assessors distinguished between products A and B, but in a more remarkable way for assessor 1.

	A				Ā			
A	1	1	…	1	0	0	0	0
B	0	0	…	0	1	0	…	1
…	…			…	…			…
I	0	0	…	0	0	1	…	0

Fig. 3. Disjunctive table of a whole of categorizations for which product A was insulated by all the consumers. The first columns of the table correspond to the groups to which product A belongs, the last columns correspond to the groups to which product A does not belong.

(cf. Fig. 3) just as a block matrix: the top right-hand block and the lower left-hand block of the table contain only the value 0.

In this particular case, the first dimension of MCA opposes this product to the others (which is a result the user can legitimately expect). This first dimension is associated with an eigenvalue of 1.

MCA analyzes also the scatterplot of the categories, a category corresponding to a group. In the case of the "labelled" categorization, the scatterplot of the categories corresponds to the scatterplot of the labels used to characterize the groups; by misuse of language, we will call it the scatterplot of the words. In this scatterplot, two labels are all the more close as they were frequently used to characterize the same perfumes.

One of the main features of MCA is that these two representations can be superimposed: a perfume is at the barycentre of the labels that characterize it and a label is at the barycentre of the perfumes with which it was associated.

The basic representation of the words is based on all the words. Another solution would have been to use the barycentre of a same word used by different consumers, but it can be a risky solution: for example the word "spicy" can refer to different meanings: pepper, cinnamon, etc.

With a great number of words, this representation can be overloaded, or even unreadable. In this case, the user will make successive selections: the words that are the most distant from the origin, the words that have been used by a particular consumer, etc.

3.3. Stability of the representation of the products

As it is often the case in exploratory multivariate analysis, one may wish to represent the variability around the individuals by means of confidence ellipses when it is possible. In our example, such confidence ellipses aim at answering the question: "what would have happened to the relative positioning of the products if we had used different consumers (*i.e.* different from those used initially)?". We decided to use partial bootstrap (Lebart, Piron, & Morineau, 2006), "which is less heavy computationally and which yields highly satisfactory results" (Greenacre, 1984, p. 215). The application of this technique in the sensory area was introduced by Lê, Husson, & Pagès, 2004 (see also Husson, Lê, & Pagès, 2005). More precisely, to obtain the confidence ellipses around the products, we use (1) the barycentric properties of MCA for which an individual (*i.e.* a product) is in the centre of gravity of the categories (in our case the words) which characterize it and (2) the characteristic of our data set according to which a variable can be assimilated to a consumer on the other hand. The idea consists in using resampling techniques on our initial set of consumers using random draws with replacement and in projecting the categories of the resampled consumers (considered as a virtual panel, composed of "as many consumers" as the original one) as supplementary elements (in the same way as categories of active variables). The categories (*i.e.* words used by the consumers to characterize their groups) associated with our virtual panel once projected, we calculate the new coordinates of the products according to the barycentric properties mentioned before. This process is then reiterated several times in order to get confidence ellipses which include 95% of the representations of the "virtual" products thus obtained.

3.4. Representation of the consumers

One may wish to have a representation of the consumers in which each consumer is represented by a point and in which two consumers are all the more close as they carried out similar categorizations. To get that representation of the consumers, we use as a starting point multiple factor analysis (MFA) in the sense of Escofier and Pagès (1998) and the equivalence between MCA and MFA performed on the same variables by considering each variable as a group (and as many groups as there are variables) (Pagès, 2002). The representation of the groups of variables in MFA becomes in our particular case a representation of the consumers through their respective individual co-occurrences matrix of dimension $I \times I$. The consumers constitute a space of dimension $I \times I$. In this representation, the scalar product between two consumers is equal to the Phi square between two qualitative variables: so, this scalar product expresses the resemblance between the partitions associated to the two qualitative variables. MFA provides an optimal low-dimensional representation of the consumers related to the representation of the perfumes on the one hand and of the words on the other hand (Escofier & Pagès, 1998, p. 168). This representation has the important following property: the coordinate of a consumer j on the dimension of rank s is equal to the squared correlation ratio between, on the one hand the qualitative variable associated to consumer j, and on the other hand F_s the vector of the coordinates of the individuals (perfumes) on the dimension of rank s. This squared correlation ratio measures the link between a qualitative variable and a quantitative variable: its range goes from 0 to 1 (Greenacre, 1984, p. 106).

3.5. Comparison with DISTATIS

DISATIS first analyzes the individual co-occurrences matrices and then provides a representation of the consumers. This representation is optimal in itself, but the dimensions on which the consumers are represented are not interpretable in term of individuals (*i.e.* perfumes). The first dimension sorts the consumers according to their degree of resemblance with respect to all the consumers; it can be regarded as a linear combination of the individual co-occurrences matrices in which each matrix is all the more important as it is similar to the others.

Then DISTATIS diagonalizes this linear combination of individual co-occurrences matrices to provide a representation of the perfumes. This representation of the perfumes which does not give the same importance to the consumers constitutes a second difference with the *FAST* approach. It draws aside the consumers who differ too much from the others.

In DISTATIS, the representation of the words results from the representation of the perfumes by using barycentric properties and is not optimal in itself.

The two methods will be confronted on the same data.

4. Results

4.1. Preliminary study

On the whole, consumers constituted between four and five groups (cf. Fig. 4), generally composed of one to three perfumes (cf. Fig. 5).

The co-occurrences matrix (cf. Table 1) does not present any cell of size null: there is no consensus between the consumers (whatever the couple of products, at least six consumers put them in the

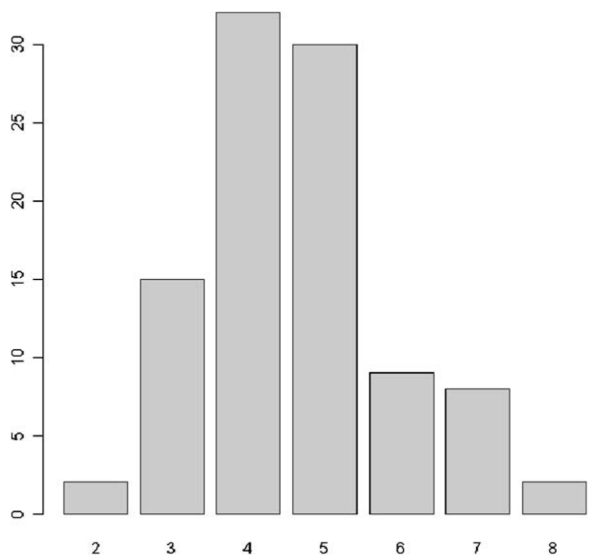

Fig. 4. Number of groups defined by all 98 consumers.

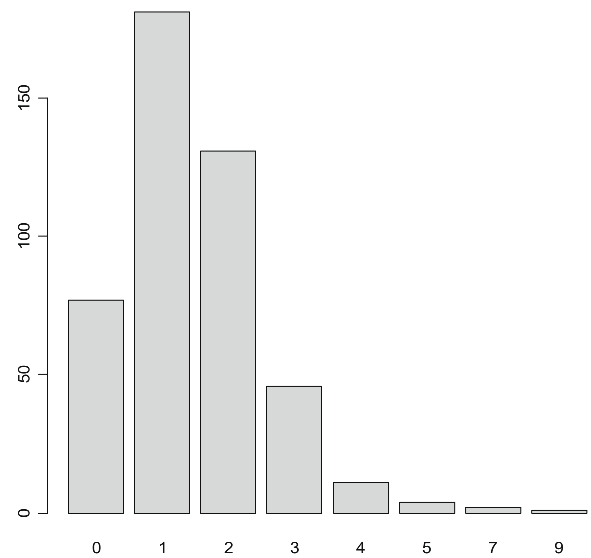

Fig. 6. Number of words per group.

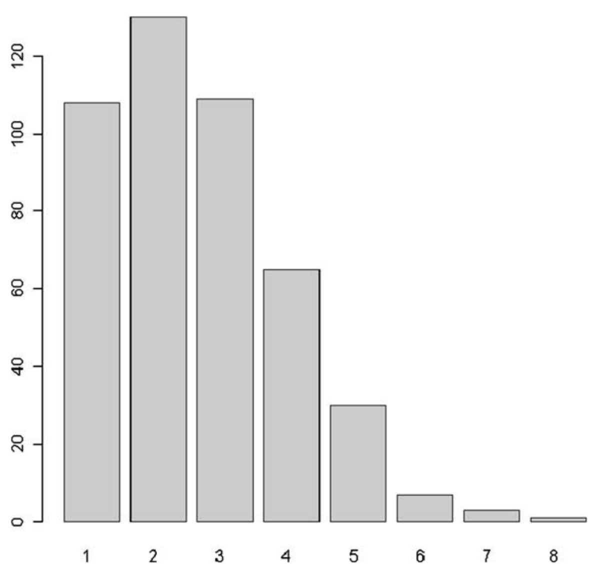

Fig. 5. Number of perfumes per group.

same group). Some perfumes were more often put alone and thus perceived as particular perfumes: *Shalimar* and *Chanel n°5*. *J'adore* (*ET*) and *J'adore* (*EP*) were placed 56 times in the same group, which represents the most important frequency. To be better understandable, this matrix was reordered, according to the coordinates of the products on the first dimension of MCA, in order to reveal sub-groups of perfumes.

On the whole, 224 words were used to characterize the perfumes. In majority, a group was characterized by one to two words (cf. Fig. 6). Only five assessors did not use any word, and 74 groups were not characterized. On the whole, consumers were able to qualify the groups of perfumes they provided.

Each perfume can be directly characterized by the words which are associated to it in a "privileged way" (cf. Table 2). For example, the words *spicy*, *vanilla*, *strong* and *sweet* are much more often used for *Angel* than for the rest of the perfumes; in the same way the words *strong*, *oriental*, *aggressive*, *drug*, *masculine*, *old* are much more often used for *Shalimar*.

4.2. Representation of the perfumes and the words

First, one can analyze the eigenvalues table (cf. Table 3). The eigenvalues decrease regularly: there is still a slight discontinuity between the second and the third eigenvalue. The two first dimensions reconstitute 30% of the total inertia of the data.

The first dimension opposes the perfumes *Shalimar*, *Aromatics Elixir* and *Chanel n°5* to the others (cf. Fig. 7). The second dimension opposes *Angel*, *Lolita Lempicka* and to a lesser extent *Cinéma* to the other perfumes. The relative distances between some perfumes are

Table 1
Co-occurrences among the perfumes.

	Shalimar	*Aromatics Elixir*	*Chanel n°5*	*Angel*	*Lolita Lempicka*	*L'instant*	*Pure Poison*	*Cinéma*	*Coco Mademoiselle*	*Pleasures*	*J'adore (EP)*	*J'adore (ET)*	Alone
Shalimar	98	42	30	21	9	13	11	10	9	6	6	7	24
Aromatics Elixir	42	98	51	27	6	13	12	8	12	11	12	7	6
Chanel n°5	30	51	98	15	8	10	21	9	11	14	12	14	17
Angel	21	27	15	98	36	14	10	18	10	11	11	12	13
Lolita Lempicka	9	6	8	36	98	22	18	42	21	18	18	18	7
L'instant	13	13	10	14	22	98	25	26	20	23	28	22	9
Pure Poison	11	12	21	10	18	25	98	28	33	30	29	28	7
Cinéma	10	8	9	18	42	26	28	98	30	22	23	24	5
Coco Mademoiselle	9	12	11	10	21	20	33	30	98	28	28	38	8
Pleasures	6	11	14	11	18	23	30	22	28	98	38	48	8
J'adore (EP)	6	12	12	11	18	28	29	23	28	38	98	56	2
J'adore (ET)	7	7	14	12	18	22	28	24	38	48	56	98	2

Please cite this article in press as: Cadoret, M., et al. A Factorial Approach for Sorting Task data (*FAST*). *Food Quality and Preference* (2009), doi:10.1016/j.foodqual.2009.02.010

Table 2
Description of the products *Angel* and *Shalimar* sorted by descending order of significantivity.

Angel	Intern (%)	Global (%)	*p*-Value
Vanilla	2.58	0.55	0.007
Spicy	3.23	0.94	0.011
Sweet	9.03	5.18	0.025
Strong	10.32	6.45	0.036
Shalimar			
Strong	12.42	6.45	0.003
Aggressive	3.92	1.05	0.004
Mint	1.31	0.11	0.007
Oriental	1.31	0.17	0.020
Old	2.61	0.77	0.025
Drug	1.31	0.22	0.038
Peppery	1.96	0.55	0.045
Masculine	1.96	0.55	0.045

Table 3
Eigenvalues table of multiple correspondence analysis.

	Eigenvalue	Percentage of variance (%)
Dim 1	0.64480997	17.80
Dim 2	0.49427323	13.64
Dim 3	0.39745161	10.97
Dim 4	0.35182556	9.71
Dim 5	0.32561992	8.99
Dim 6	0.2851883	7.87
Dim 7	0.28374358	7.83
Dim 8	0.25248057	6.97
Dim 9	0.23489308	6.48
Dim 10	0.20866116	5.76
Dim 11	0.14350202	3.96

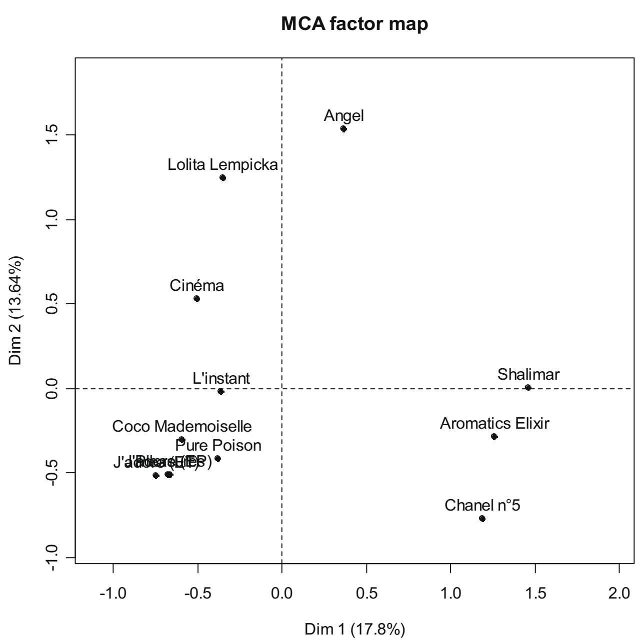

Fig. 7. Representation of the perfumes in the plane defined by dimensions 1 and 2 of MCA.

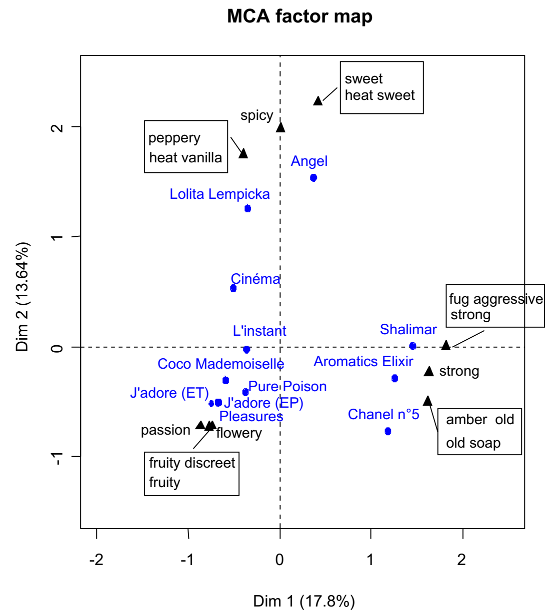

Fig. 8. Representation of the words in the plane defined by dimensions 1 and 2 of MCA.

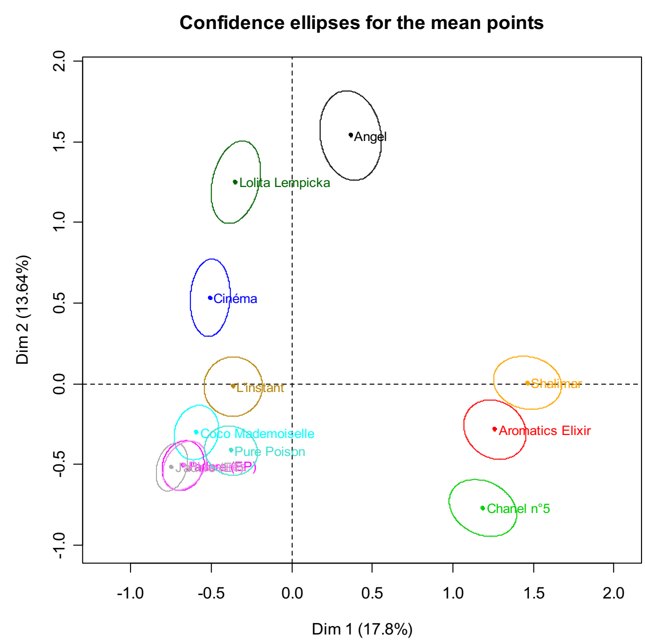

Fig. 9. Representation of the perfumes and their respective confidence ellipses in the plane defined by dimensions 1 and 2 of MCA.

to be connected with the number of times that these perfumes were put in a group alone: it is the case for example for *Shalimar*, *Chanel n°5* and *Angel* (cf. last column of Table 1). The proximities are to be connected to the frequency perfumes were put in the same group: it is the case for *Aromatics Elixir* associated 42 times with *Shalimar* and 51 times with *Chanel n°5*; and for *Lolita Lempicka* associated 36 times with *Angel* (cf. Table 1). Also let us note the proximity between the two *J'adore* that were put together 56 times.

The representation of the perfumes is supplemented by superimposing the representation of the words: by construction, a perfume is at the barycentre of the words with which it was associated (cf. Fig. 8). By selecting the words that are the most distant from the origin, the first dimension opposes the perfumes associated with the words *strong*, *old*, with the perfumes described as rather *flowery* and *fruity*. The second dimension opposes the perfumes associated with the words *heat*, *sweet*, *spicy* to the other perfumes. The characterization of the groups following the sorting task appears invaluable for the interpretation of the sensory dimensions and is integrated in a natural way in the analysis.

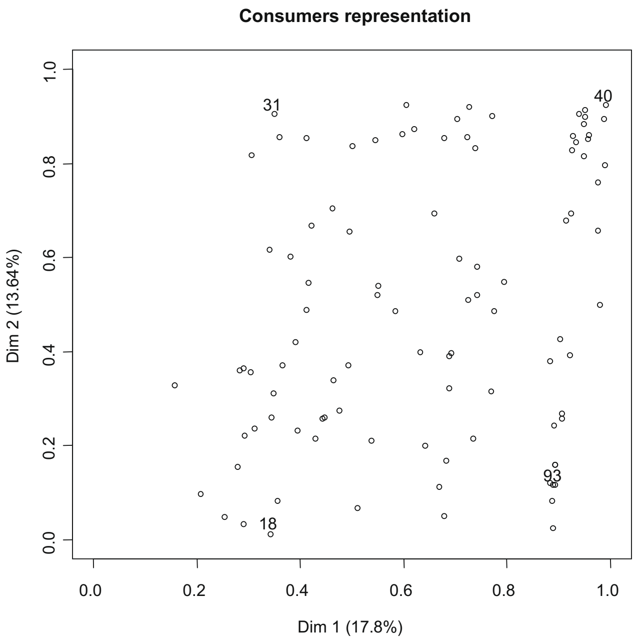

Fig. 10. Representation of the consumers in the plane defined by dimensions 1 and 2 of MFA.

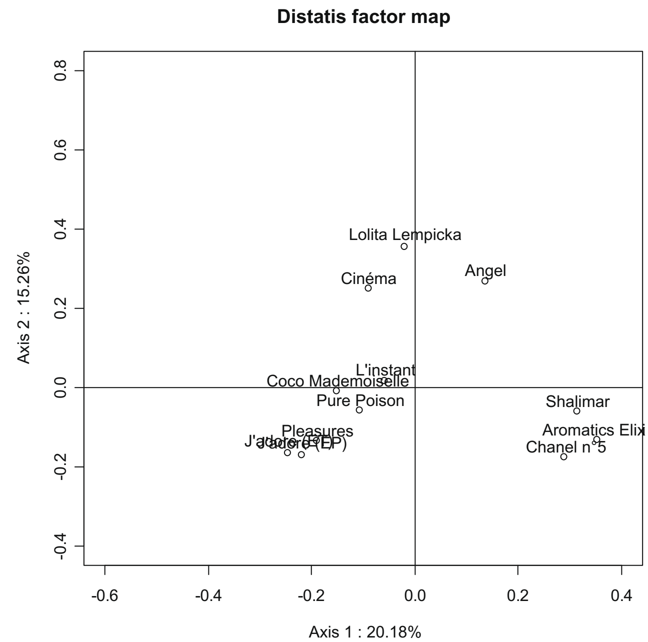

Fig. 11. Representation of the perfumes in the plane defined by dimensions 1 and 2 of DISTATIS.

Table 4
Categorizations of assessors 18, 31, 40 and 93.

	Assessor 18	Assessor 31	Assessor 40	Assessor 93
Shalimar	2	1	4	4
Aromatics Elixir	2	2	5	3
Chanel n°5	3	4	5	3
Coco Mademoiselle	3	1	2	2
J'adore (EP)	1	1	1	1
J'adore (ET)	3	1	1	2
L'instant	2	1	2	1
Pleasures	3	1	1	1
Pure Poison	1	2	2	2
Angel	3	5	6	1
Cinéma	3	3	3	2
Lolita Lempicka	1	3	3	2

4.3. Stability of the representation of the products

The size of these ellipses is relatively small given the dispersion of the perfumes (cf. Fig. 9), it results a weak overlapping between the ellipses even for the close perfumes. Although they are close, the perfumes *L'instant*, *Coco Mademoiselle* and *Pure Poison* are well differentiated by the consumers. In spite of a lack of consensus (cf. Section 4.1), there is a structure emerging from the associations between perfumes: the "significant" differences between the perfumes are well highlighted by the *FAST* approach. Let us notice the overlapping of both *J'adore* that is directly linked to the co-occurrences matrix.

4.4. Representation of the consumers

The representation of the consumers highlights various types of categorization (cf. Fig. 10). On the first dimension, consumers 93 and 40, who both have high coordinates, are opposed to consumers 18 and 31. The coordinate of a consumer on a dimension being equal to the correlation ratio between its partitioning variable and the dimension, it results that consumers 40 and 93 clearly individualized the perfumes *Shalimar*, *Aromatics Elixir* and *Chanel n°5* unlike consumers 18 and 31 (cf. Table 4). According to the second dimension, consumers 31 and 40, who have strong coordinates, are opposed to consumers 18 and 93. Indeed, consumers 31 and 40 individualized well *Angel* and with less degrees *Lolita Lempicka* and *Cinéma*, which is not the case for consumers 18 and 93 (cf. Table 4).

Table 5
Correlation coefficients between the first two dimensions provided by MCA and DISTATIS.

	$F_{1\text{MCA}}$	$F_{2\text{MCA}}$
$F_{1\text{DISTATIS}}$	0.9805	−0.139
$F_{2\text{DISTATIS}}$	−0.156	0.9376

4.5. Comparison with DISTATIS

The comparison with DISTATIS is done first on the data which have just been analyzed, then on the beers data used in the presentation of DISTATIS (Abdi et al., 2007).

4.5.1. Data perfumes

The first two dimensions resulting from DISTATIS (cf. Fig. 11) are relatively close to those of MCA: in both cases, the first dimension highlights *Aromatics Elixir*, *Shalimar* and *Chanel n°5*; on the second dimension, *Lolita Lempicka*, *Angel* and *Cinéma* are opposed to the other perfumes. This proximity between results can be quantified using the correlation coefficients calculated between the dimensions resulting from the two methods (cf. Table 5). In detail, *Shalimar* and *Angel* are represented in a more extreme way by MCA than they are by DISTATIS. MCA particularly draws aside the perfumes frequently isolated from the others; one finds here the property mentioned Section 3.2 (a perfume systematically alone generates a dimension by itself).

In DISTATIS the representation of the consumers is interpreted the following way: consumers that have the highest coordinates on dimension 1 are those which have provided the categorizations closest to the whole of categorizations: in a direct way, one can see on Table 4 that consumer 40 put together the perfumes most frequently associated (*Chanel n°5–Aromatics Elixir*, *Lolita Lempicka-Cinéma*); conversely, those having the weakest coordinates are atypical: consumer 18 for example put together *Chanel n°5* and *Cinéma* which were put together only by nine consumers out of ninety eight (cf. Table 1). According to Fig. 12, dimension 1 is in phase with the representation of the consumers resulting from MCA which opposes consumer 40 (strongly related to the first two axes of MCA) to consumer 18 (far from related to the first two axes of MCA). Consumers 93 and 31 are connected only partially with the average configuration, but one does not know according to which characteristics of this average configuration. The second dimension of DISTATIS is not interpretable in terms of categorizations; whereas the dimension of MCA indicates that consumer 93 isolated well *Aromatics Elixir*, *Chanel n°5* and *Shalimar*, and not *Angel*, *Lolita Lempicka* and *Cinéma*.

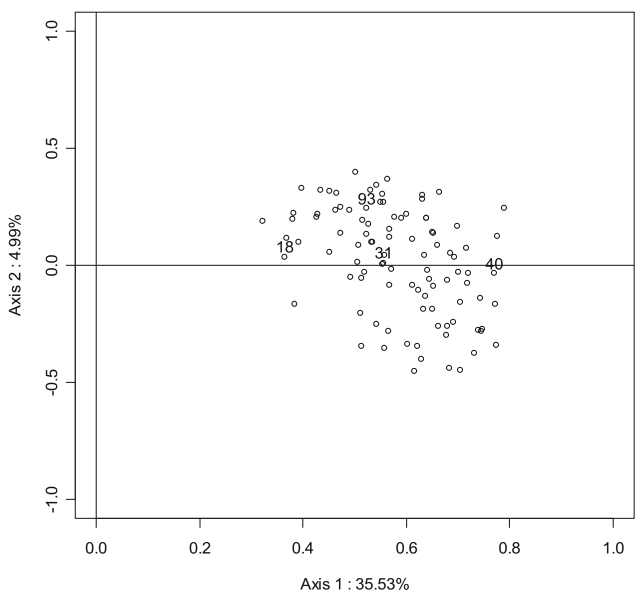

Fig. 12. Representation of the consumers in the plane defined by dimensions 1 and 2 of DISTATIS.

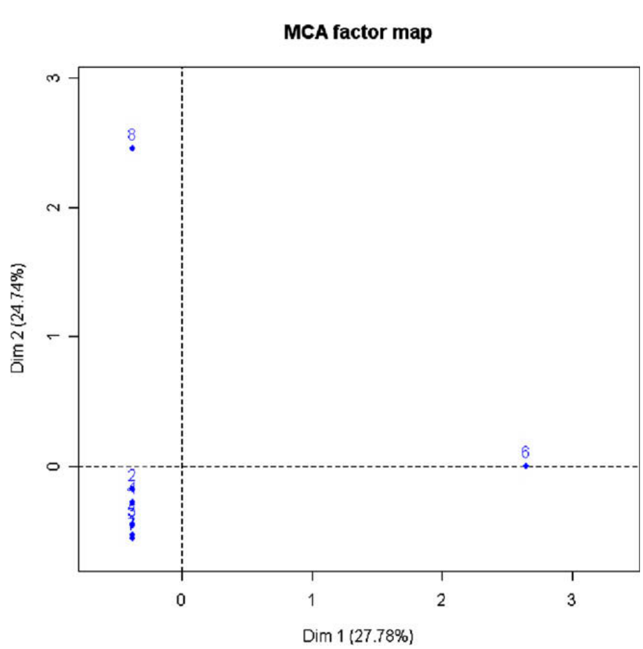

Fig. 13. Representation of beers in the plane defined by dimensions 1 and 2 of MCA.

Table 6
Co-occurrences among the beers, products 2 and 5 have been put together five times; product 6 has been isolated ten times out of ten.

	8	2	4	3	5	7	1	6	Alone
8	10	1	1	0	0	0	0	0	8
2	1	10	2	3	5	0	1	0	1
4	1	2	10	2	5	5	5	0	0
3	0	3	2	10	2	0	1	0	5
5	0	5	5	2	10	4	6	0	0
7	0	0	5	0	4	10	8	0	1
1	0	1	5	1	6	8	10	0	0
6	0	0	0	0	0	0	0	10	10

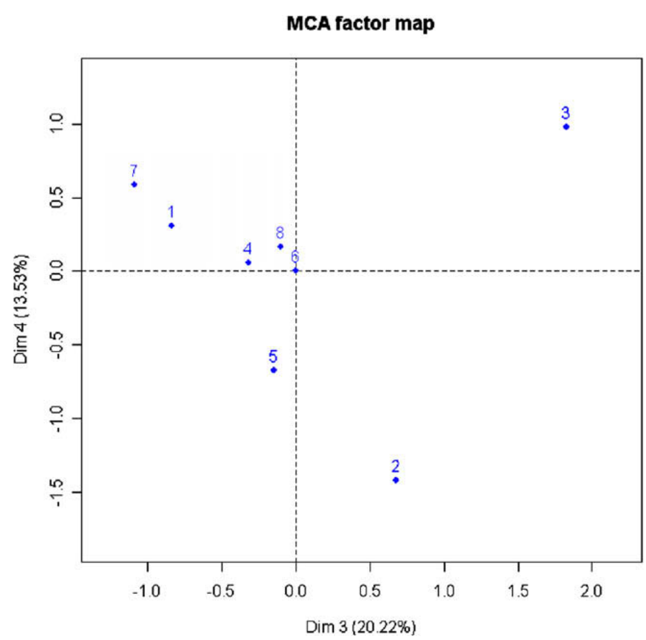

Fig. 14. Representation of beers in the plane defined by dimensions 3 and 4 of MCA.

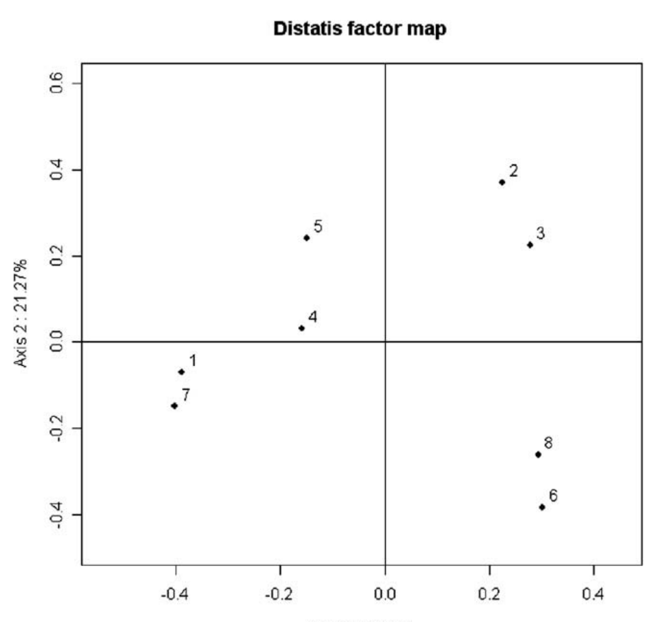

Fig. 15. Representation of beers in the plane defined by dimensions 1 and 2 of DISTATIS.

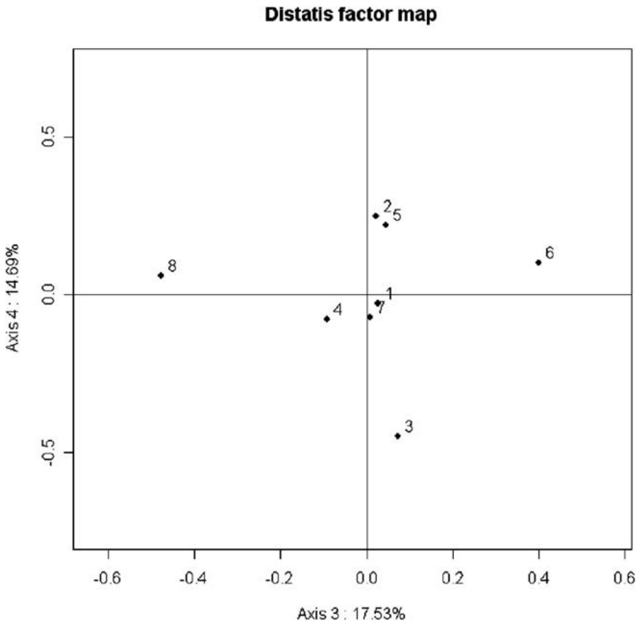

Fig. 16. Representation of beers in the plane defined by dimensions 3 and 4 of DISTATIS.

4.5.2. Data beers

In this data file (Abdi et al., 2007), ten consumers carried out a sorting task on eight beers.

The representation of the beers provided by MCA highlights on the first dimension (associate with an eigenvalue of 1) beer 6, perfectly represented, opposed to all the others (cf. Fig. 13). Indeed, this beer was systematically isolated from the others by all the consumers (cf. Table 6). One finds the borderline case mentioned previously. The second dimension highlights an almost similar case for beer 8: this one was insulated from the others by eight consumers out of ten. Nothing can be said concerning the proximity between the rest of the beers given their bad quality of representation; to do so, let's refer to dimensions 3 and 4.

For dimensions 3 and 4 (cf. Fig. 14), all the beers are well represented, except beers 6 and 8 (whose proximity cannot thus be interpreted), and their proximities express their co-occurrences: beers 1 and 7 which are close were put eight times together out of ten, whereas beers 2 and 3, dispatched all over the plane, were gathered only three times.

The two first dimensions of DISTATIS (cf. Fig. 15) highlight the particular case of beers 6 and 8, which is to be put in relation to the fact that they were never put together with the other beers, but which is in contradiction with the fact that they were never put together. On the other hand, the plane highlights beers 1 and 7 which were actually put together by eight consumers out of ten.

The plane associated with dimensions 3 and 4 of DISTATIS highlights another dimension on which beers 6 and 8 are well represented (cf. Fig. 16): beers 6 and 8 are far away from the origin along dimension 3, which does not correspond exactly to the data since beer 6 is not opposed particularly to beer 8 but to the whole of beers.

Remark. Even if it is not the purpose of this article, we have also analyzed these data by metric MDS (using the *cmdscale* function implemented in R) and we have obtained an RV coefficient of 0.987 between the first two dimensions provided by metric MDS and DISTATIS and an RV coefficient of 0.455 between the first two dimensions provided by MCA and metric MDS.

5. Conclusion

The *FAST* approach is a simple and easy to interpret method since it derives from MCA for which there is a simple relation between the representation of the perfumes and the words. This simplicity allows the natural construction of confidence ellipses starting from the properties of MCA and resampling methods. Equivalence here between MCA and MFA also allows an optimal representation of the consumers related to the representation of the perfumes. The presence of words is invaluable concerning the interpretation: the example of the perfumes shows that it is possible to ask consumers to characterize the groups they constituted with words.

This method is implemented in the SensoMineR package *via* the *FAST* function.

References

Abdi, H., & Valentin, D. (2007). Some new and easy ways to describe, compare, and evaluate products and assessors. *Proceedings of SPISE, 2007*, 5–18.

Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference, 18*, 1–16.

Ballester, J., Patris, B., Symoneaux, R., & Valentin, D. (2008). Conceptual vs. perceptual wine spaces: Does expertise matter? *Food Quality and Preference, 19*, 267–276.

Blancher, G., Chollet, S., Kesteloot, R., Nguyen Hoang, D., Cuvelier, G., & Sieffermann, J.-M. (2007). French and Vietnamese: How do they describe texture characteristics of the same food? A case study with jellies. *Food Quality and Preference, 18*, 560–575.

Cartier, R., Rytz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., et al. (2006). Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference, 17*, 562–571.

Escofier, B., & Pagès, J. (1998). *Analyse factorielle simple et multiple*. Dunod.

Faye, P., Brémaud, D., Teillet, E., Courcoux, P., Giboreau, A., & Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference, 17*, 604–614.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

Healy, A., & Miller, G. A. (1970). The verb as the main determinant of the sentence meaning. *Psychonomic Science, 20*, 372.

Husson, F., Lê, S., & Pagès, J. (2005). Confidence ellipse for the sensory profiles obtained by principal component analysis. *Food Quality and Preference, 16*, 245–250.

Lavit, C. (1976). *Analyse conjointe de tableaux quantitatifs*. Paris: Masson.

Lê, S., Husson, F. & Pagès, J. (2004) Confidence ellipses in HMFA applied to sensory profiles of chocolates. In *The 7th sensometrics meeting*. Davis, USA.

Lebart, L. (1975). L'orientation du dépouillement de certaines enquêtes par l'analyse des correspondances multiples. *Consommation* (Vol. 2, pp. 73–96). Dunod.

Lebart, L., Piron, M., & Morineau, A. (2006). *Statistique exploratoire multidimensionnelle*. Dunod.

Lelièvre, M., Chollet, S., Abdi, H., & Valentin, D. (2008). What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, 19, 697–703.

Pagès, J. (2002). Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes. *Revue de Statistique Appliquée, 50*(4), 5–37.

Parr, W. V., Green, J. A., White, K. G., & Sherlock, R. R. (2007). The distinctive flavour of New Zealand Sauvignon blanc: Sensory characterisation by wine professionals. *Food Quality and Preference, 18*, 849–861.

Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV coefficient. *Applied Statistics, 25*, 257–267.

Tenenhaus, M., & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika, 50*, 91–119.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Van der Kloot, W. A., & Van Herk, H. (1991). Multidimensional scaling of sorting data: A comparison of three procedures. *Multivariate Behavioral Research, 26*(4), 563–581.