Food Quality and Preference 32 (2014) 2–15

Contents lists available at ScienceDirect

Food Quality and Preference

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

Tracking verbal-based methods beyond conventional descriptive analysis in food science bibliography. A statistical approach

Mónica Bécue-Bertaut *

Department of Statistics and Operational Research, Universitat Politècnica de Catalunya, North Campus, C5, Jordi Girona 1-3, 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 14 October 2012
Received in revised form 22 August 2013
Accepted 22 August 2013
Available online 31 August 2013

Keywords:
Verbalisation tasks
Textual statistics
Correspondence analysis
Multiple factor analysis for contingency tables
Characteristic words
Bibliographic study

ABSTRACT

The usage of verbal-based methods beyond conventional descriptive analysis is increasing in sensory analysis, either as full-methods or as a complement to holistic methods. They contribute to a better understanding of the consumers' likes and willingness, important factors to the food industry. A prime objective of this paper is to give a global vision of the scientific publications in food science related to this topic from their abstracts. Textual statistics, combining multidimensional methods such as correspondence analysis, multiple factor analysis for contingency tables and characteristic words, are proving to be useful for extracting information from the corpus of abstracts. These abstracts have evolved over time towards a higher concern for research about methodology, which has become more complex and requires sophisticated statistical methods. Sensory methods, such as free choice profile, flash profile, repertory grid, sorting task, napping, word association and CATA, have emerged or have been revitalised. New statistical methods, such as multiple factor analysis, have been introduced to analyse data issued from verbalisation tasks. However, correspondence analysis, a reference method for dealing with texts and, more generally, frequency tables, is used with too much restraint.

1. Introduction

Since the pioneer work by Cairncross and Sjostrom (1950), conventional descriptive analysis (CDA) and its variants have been widely used in the food and beverage industry. CDA aims at providing an objective measure of the sensory properties of a set of products. A tasting sheet, as a list of descriptor words, is established, usually associated with references, covering all of the essential sensory aspects. Trained panellists score each descriptor for each product to establish the sensory profile of the products individually. Furthermore, these sensory scores, globally taken into account through principal component analysis (PCA), offer a global configuration of the products, showing how closely the products are related based on the similarity of their scoring on each descriptor. CDA is classified in verbal-based methods as the quality of its results relies on the panellist's ability to match perceptions and descriptor words (Murray, Delahunty, & Baxter, 2001; Strigler et al., 2009; Valentin, Chollet, Lelièvre, & Abdi, 2012). This method has proven to be essential and remains the basis of sensory descriptions.

However, the need to innovate and to place new or updated products on the market has led to new considerations beyond the sensory characterisation of the products and to looking for less time-consuming methods (Strigler et al., 2009; Valentin et al., 2012; Varela & Ares, 2012; Worch, Lê, & Punter, 2010). The study of the interactions between sensory attributes and consumers' acceptance, likes and dislikes or even emotions has led to give a voice to the consumer (Van Kleef, Van Trijp, & Luning, 2005). To this end, the verbal-based approach, proving to be an asset, has diversified. Techniques have been imported from market research and psychology (Simeone & Marotta, 2010). The techniques of collecting and analysing the verbal data have evolved and have shaped new methods, gathered here under the label "verbal-based" because they rely on either pre-established or freely formulated verbal descriptions from the panellists, usually consumers. The free comments used to enrich similarity-based methods are also included. This point of view differs from the one proposed by Valentin et al. (2012) who globally encompass the similarity-based methods and their verbal supplements as a whole in the similarity-based family. Here, both tasks, similarity-based and verbal supplement, are separately assigned to their own family.

The main characteristics of the methods of interest in this work are described hereafter.

Lexicon development in a session prior to the CDA, leading the panellists to finally agree on a common set of descriptors, is a current practice used to describe new products (Barcenas, Pérez Elortondo, Salmerón, & Albisu, 1999; Civille, Lapsley, Huang, Yada, & Seltsam, 2010; Kinski et al., 2006; Lawless, Hottenstein, & Ellingsworth, 2012). However, the panellists may have difficulties

* Tel.: +34 93 401 70 31; fax: +34 93 401 58 55.
E-mail address: monica.becue@upc.edu

0950-3293/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2013.08.010

embracing a common definition of descriptors (Quarmby & Ratkowsky, 1988); thus, the idea of letting them completely free to select their own descriptors but avoiding hedonic words is the basis of *Free choice profiling* (FCP, Williams & Langron, 1984), *Flash profile* (FP; Dairou & Sieffermann, 2002; Delarue & Sieffermann, 2004; Sieffermann, 2000) and *Repertory Grid* (RG, Veinand, Godefroy, Adam, & Delarue, 2011). In FCP, the panellists rate the intensity of each attribute for each product individually, whereas in FP, they rank the products according to each attribute successively, thereby favouring the choice of discriminative descriptor words. In RG, the products are randomly grouped in triads, and the panellists have to describe, using free verbal descriptions, how two products of each triad differ from the third, another means to generate discriminative words. Then, the panellists are required to score the intensity of their own descriptors for each product. The data issued from these three methods, which are alternatives to CDA, are analysed through Generalised Procrustes analysis, which yields a global configuration of the products. Discussed points include the huge variability of language when individual words are used and disagreements about the meaning of certain terms, even after training (Quarmby & Ratkowsky, 1988; Strigler et al., 2009), which is a common problem in all verbal-based methods.

More recently, free verbal descriptions expressed by the panellists have been introduced to complement non-verbal methods. In the methods called *answers to open-ended questions* or *free comments*, the consumers are required to complete their liking scores by writing down free remarks with a view towards a better understanding of their preferences. Variants are observed. Ten Kleij and Musters (2003) gave the consumers an option whether to answer and, if they do, to either explain why they gave particular liking scores or to express whatever crossed their mind. Ares, Giménez, Barreiro, and Gámbaro (2010) forced the consumers to give a remark but limited the remark to 4 words. Symoneaux, Galmarini, and Mehinagic (2011) gave consumers the non-mandatory option of separately stating through free comments what they liked and what they disliked about each sample. The verbal tasks called *Labelling* (Blancher et al., 2007; Bécue-Bertaut & Lê, 2011; Cadoret, Lê, & Pagès, 2009; Faye et al., 2004) and *ultra flash profiling* (UFP; Perrin & Pagès, 2009; Perrin et al., 2008) have been used to enrich similarity-based methods such as sorting task (labelled sorting) or napping/projective mapping (napping + UFP). In a labelled sorting task, the panellists are asked to form groups of products depending on the perceived similarities and then to *label* each group with some words. In napping, each panellist places the samples on a two-dimensional space depending on their similarities and then describes each sample with words. Through these verbal tasks, the consumers provide information about the characteristics of the products to support the similarities and dissimilarities that they perceive, in addition to providing descriptions. Sensory but also hedonic words are usually provided. The latter can be considered in the analysis to link sensory and hedonic aspects and to underline the characteristics that are relevant in the consumers' view. A drawback of these free descriptions is the wide variability of vocabulary, from which information can be arduous to extract.

To avoid this variability, *Check-all-that-apply* (CATA; Ares, Deliza, Barreiro, Giménez, & Gámbaro, 2010; Dooley, Lee, & Meullenet, 2010; Lancaster & Foley, 2007; Lee, Findlay, & Meullenet, 2013; Puyares, Ares, & Carray, 2010), recently introduced in sensory analysis, turns to pre-established lists of words or sentences that are not limited to sensory attributes. Therefore, this method maintains the benefits of free comments to explore descriptions by the consumers while also collecting information on preferences and even emotions. CATA requires the consumers to choose, within a list, all of the words or sentences that they consider appropriate to describe a product. This type of questioning has been used in consumer studies to determine which sensory attributes consumers perceive in a food product. The possibility of letting the panellists use their own words is also considered, which turns CATA into a variant of *Free Choice Profiling*, but relying on citation counts and not scores (Dooley et al., 2010).

Word association, a simple technique recently imported from psychology into food science, constitutes a tool to grasp the meaning of specific words, to explore food choices, to elicit the attributes that are drivers of liking or disliking and to understand the consumers' perceptions of new and undefined concepts (Ares, Giménez, & Gámbaro, 2008; Guerrero et al., 2010; Roininen, Arvola, & Lähteenmäki, 2006).

Reviews and comparative works highlight the advantages and disadvantages of the different methods (Ares et al., 2010; Dooley et al., 2010; Moussaoui & Varela, 2010; Tournier, Martin, Guichard, Issanchou, & Sulmont-Rossé, 2007; Valentin et al., 2012; Varela & Ares, 2012; Veinand et al., 2011).

The relevance of verbalisation tasks is reflected in the increasing number of publications devoted to them compared to the total number of articles published in food science journals. Whereas the average number of articles published per year in these journals has doubled from 1990–1994 to 2008–2012, the average number of those devoted to verbal methods has increased by a factor of 12. Several works have considered the contributions of these methods as essential. The relationships between consumers' acceptance and vocabulary are valuable to marketing (Carr, Craig-Petsinger, & Hadlich, 2001). In the music domain, where the perception issues are similar to the food industry, verbal description offers a detailed description of the main features used by the panellists in assessing comparative judgments (Stepanek, 2006). Some panellists give very subtle sensory descriptions, whereas others remain at a low descriptive level (Thamke, Dürrschmid, & Rohm, 2009). Verbalisation facilitates the recognition and sharing of a sensory experience (Baccino et al., 2010). Letting the panellists choose their own words is the only way to identify the customary terms used by the consumers (Galmarini, Symoneaux, Chollet, & Zamora, 2013).

The number of verbal-based methods, their increasing use, the growing number of publications devoted to them and their ability to capture the consumers' exact wording, an ever more pressing need, argue for the relevance of verbalisation tasks. This relevance motivates the present study, whose prime aim is to uncover the evolution of the verbal-based methods mentioned above and to detect changes and novelties through a content-oriented bibliographic analysis of the abstracts published in food science journals. To replace the journals in this evolution and to determine the abstracts presenting a vocabulary ahead of their publication date constituted collateral aims. Moreover, as an original methodology, gathering a series of textual statistics methods is proposed and applied for tracking time in the data base of abstracts, another aim is to show the potentiality of this type of bibliographic study.

2. Material and methods

2.1. Base of abstracts

The collection of abstracts was gathered from the *Web of Science* at the end of January 2013 as a response to the query shown in Table 1. The set of words building up the topic was selected to address the verbal-based methods beyond conventional profiling in the widest sense. Therefore, in addition to the methods cited in the introduction, generic terms, such as *vocabulary* and *textual data*, have been included.

Only English-language publications in food science journals were selected. Equivalence between American English and British English was automatically managed by the query system.

Table 1
Request submitted to the *Web of Science*.

```
Topic=(verbalization OR "verbal based " OR  lexicon OR vocabulary OR
"textual data" OR "free text" OR "free comments" OR "open-ended " OR
"ultraflash profile" OR "ultraflash profiling" OR "ultra flash profile"
OR "repertory grid" OR "ultra flash profiling" OR "flash profile" OR
"flash profiling" " OR "free choice profile" OR  "free choice profiling"
OR "check-all-that-apply" OR CATA OR "association word")


Refined by: Subject Areas=(food science technology)

            Publications = (Journals)

            Languages  = (English )

            Lemmatization = on

            Language= English
```

A total of 423 abstracts were downloaded, with the format displayed in Table 2. Thirty-seven of these abstracts were excluded because they were not related to verbal-based methods; they were selected for containing the expression "open-ended coaxial". Each of the 386 conserved abstracts had complete information concerning the title, authors, journal and publication year.

Fig. 1 summarises the distribution of the abstracts by year and journal (60). This figure shows how the topic of interest here, which arose in abstracts in 1990, increases its presence in food science journals over the years, as mentioned in the Section 1. Only five journals have published more than 10 articles concerning this topic, and three of them, *Journal of Sensory Studies*, *Food Quality and Preference* and *Journal of Food Science*, contain 50% of the articles.

2.2. Pre-processing of the abstracts

The set of abstracts was pre-processed as follows:

- The compound words that identify sensory or statistical methods were written so that they were considered as a single word (for example, *descriptive word* or *analysis of variance*).
- The function words, also called stopwords (articles, prepositions, and so on), and some general words were removed from the data to prevent them from playing a major role and inducing unwelcome proximities between abstracts (Lebart, Salem, & Berry, 1998). In this case, the stop-list proposed in R package *tm* (Feinerer, 2013; Feinerer, Hornik, & Meyer D., 2008) was used.
- Minimum thresholds have been imposed on the word frequencies (10) and on the number of abstracts using a word (5), to make comparisons between documents meaningful from a statistical viewpoint.

As a result, 821 of the 6709 different words and 28,259 of the 76,611 occurrences were kept.

2.3. Aims of the statistical analysis of the bibliography

Content-based bibliographic studies attempt to detect the main topics in a research domain, together with their associations or oppositions, and to track their evolution to locate changes, advances and novelties. Here, words account for content or topics because it is established that documents (here, abstracts) using the same words have closely related contents (Benzécri, 1981). Correspondence analysis (CA; Benzécri, 1973, 1981; Lebart et al., 1998; Murtagh, 2005) accounts for similarities among documents (here, abstracts) from words and word co-occurrences, which makes its use a favoured tool for content-oriented bibliometric studies (Bansard, Kerbaol, & Coatrieux, 2006; Bansard et al., 2007; Kerbaol, Bansard, & Coatrieux, 2006; Šilic, Morin, Chauchat, & Dalbelo Bašic, 2012).

Advances in research usually occur through a link of works starting from a first proposal that is later applied and then enriched or discussed in others, leading to an evolution of the topic, a main concern here. This evolution is expressed through vocabulary renewal over time because an innovative proposal brings a new set of words that is later employed in other works, which further develop and discuss the proposal. To track the evolution, a time-oriented viewpoint is adopted, but a word co-occurrence approach is maintained, by resorting to multiple factor analysis for contingency tables (MFACT; Bécue-Bertaut & Pagès, 2004, 2008). Mappings of words and abstracts, as in CA, but taking into account both the chronology and co-occurrences of words are provided, enhancing the vocabulary evolution over time. By taking advantage of the ability of MFACT to compare both the chronology and vocabulary of the abstracts, possible novelties are spotted. MFACT is also able to locate abrupt changes and to identify lexically homogeneous periods in the corpus, further characterised by the vocabulary. Using the possibility of adding illustrative elements, the associations between journals and topics are underlined.

This content-oriented approach strongly differs from the indicator-oriented bibliometric studies, such as those proposed by Alfaraz and Mirta-Calviño (2004) and Vijay and Raghavan (2007), in food science field, which are only concerned with quantitative indicators without any reference to the content of the research.

Table 2
Format of an abstract as downloaded from the *Web of Science*.

```
PT J
AU Ares, Gaston
   Budelli, Eliana
   Bruzzone, Fernanda
   Gimenez, Ana
   Lema, Patricia
TI CONSUMERS' TEXTURE PERCEPTION OF MILK DESSERTS. I - RELATIONSHIP
WITH
   RHEOLOGICAL MEASUREMENTS
SO JOURNAL OF TEXTURE STUDIES
VL 43
…
PD JUN 2012
PY 2012
AB Studying and predicting consumers' texture perception of milk
desserts
   are important for dairy companies during new product development
and the
   design of positioning (. . .) able to predict consumers'
perception of the
   characteristics responsible for the largest texture differences
between
   the desserts. PRACTICAL APPLICATIONS Check-all-that-apply question
   (. . .)   measurements. The advantage of this methodology is that
it only
   requires participants to select words from a list, being easier,
less time
   consuming and more natural to use.
TC 0
Z9 0
SN 0022-4901
UT WOS:000304805700004
ER
```

The methods and types of achieved results used in this work are detailed hereafter.

2.4. Correspondence analysis

CA is an essential tool for organising documental databases from a content viewpoint (Benzécri, 1981; Lebart et al., 1998; Murtagh, 2005). The corpus is encoded into a *lexical* table with documents (here, abstracts) in rows and words in columns. Weights for row-abstracts and column-words are automatically handled by CA from the margins of this lexical table and are proportional to abstract lengths and word counts.

CA embeds the row and column spaces, endowed with chi-square distances, into classical Euclidean spaces and computes the principal axes of both the weighted row and column clouds. On CA displays, the abstracts are mapped closer when they use a similar vocabulary, and the words are mapped closer when they are frequently present in the same abstract or associated with the same words. This result demonstrates the ability of CA to automatically retrieve synonymy relationships without introducing them as external information through ontologies. As a consequence, abstracts whose contents are similar in meaning, although expressed through different words, are brought closer together. Both mappings are related in such a way that links between words and abstracts are uncovered, allowing for the interpretation of the similarities between abstracts in terms of their vocabulary and content (Lebart et al., 1998). Therefore, CA accounts for and visualises the following:

- Similarities among abstracts based on their verbal content.
- Similarities among words based on their distribution among abstracts, taking into account context, that is, associations between words.
- Mutual associations between abstracts and words through a simultaneous representation of rows and columns on a same graphic. On every axis of this graphic, the rows/abstracts are, up to a constant, the weighted average of the columns (words) they use and vice versa.

The interpretation of the CA results consists of looking, axis by axis, for the "metakeys" and "metadocs" that characterise them. For a given axis, "*metakey*+"/"*metakey*−" (Kerbaol et al., 2006) is

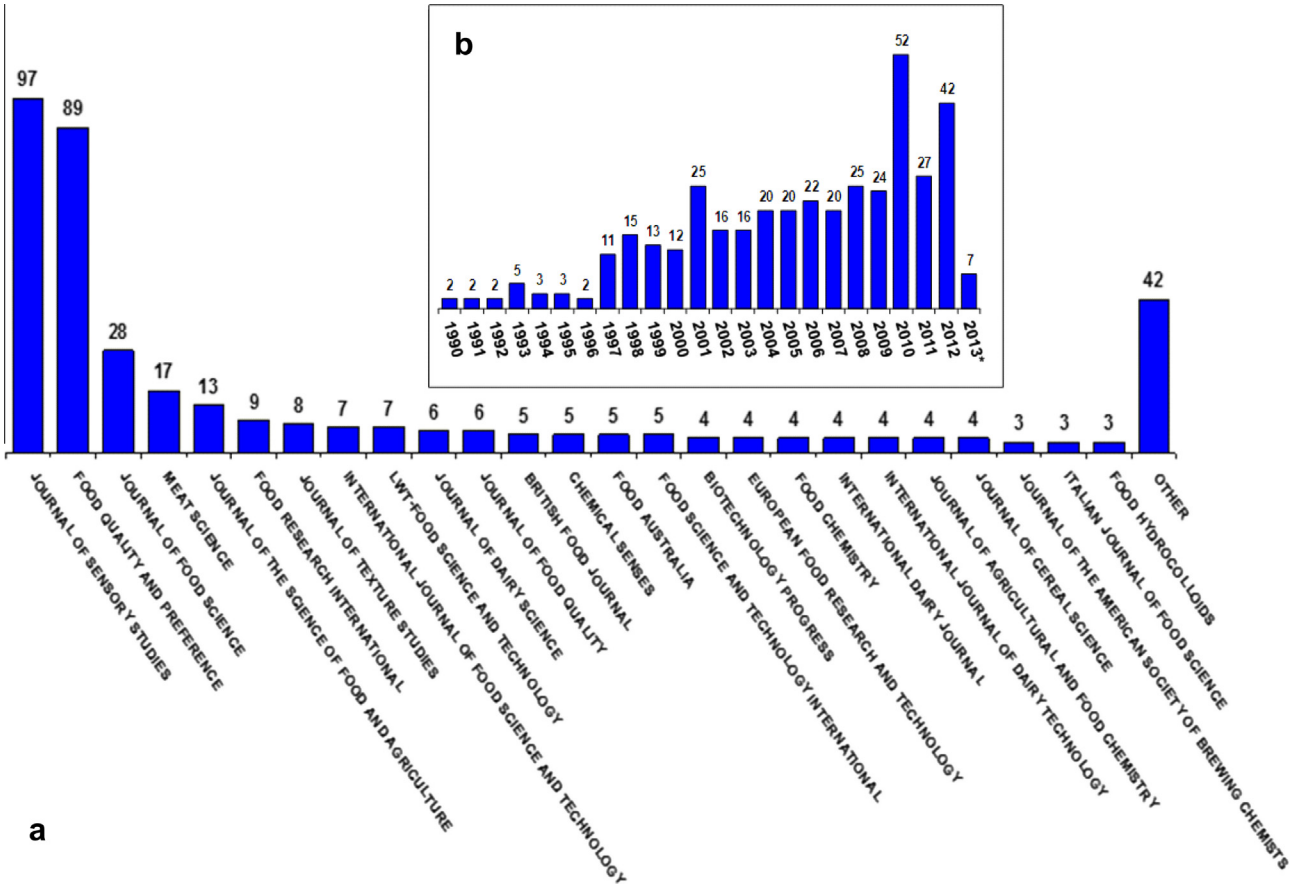

Fig. 1. Distribution of the abstracts a. by journal b. by year; 2013* is incomplete.

the set of the words that most contribute to its inertia and lie on its positive/negative part. Similarly, for a given axis, "*metadoc*+"/"*metadoc*−" is the set of the documents that contribute the most to its inertia and lie on its positive/negative part. Therefore, one or two metakeys/one or two metadocs can characterise each axis depending on the word/abstract configuration, given that all of the highly contributory words/abstracts can lie at only one axis end. The words belonging to a same metakey are frequently used in the same abstracts and, all together, correspond to a given topic. A word can belong to several metakeys, associated in each of them with different other words, that is, with different contexts. Each context corresponds to either a different meaning (for example, *social* acquires a different meaning in *social security* than *in social class*) or a different concern (*green tea* can be cited in an abstract because it is used as a support to present a new sensory method or because its attributes or properties have to be established). The words included in a high number of metakeys, known as high dimensional words, give a good account for the distinct concerns of the corpus.

A metadoc and a metakey characterising the same axis end are associated through CA transition relationships. The abstracts belonging to a metadoc use words belonging to the associated metakey. This allows for the easy location of the abstracts related to the topics revealed by the metakeys.

The CA performed on the abstracts × words table can be enriched by projecting the quantitative variable *publication year*, accounting for the chronology, as illustrative; its correlation with each axis is computed. However, chronology is not the only factor at work on the vocabulary selection, and if not dominant, chronology will not be properly represented on the first axes. Using MFACT eases the tracking-time through vocabulary renewal. In the following, chronology and publication year are used as synonyms.

2.5. Multiple factor analysis for contingency tables

Multiple factor analysis (MFA; Escofier & Pagès, 2008) offers the possibilities of introducing the chronology as active in the analysis and of mapping the abstracts based on their vocabulary, similarly to CA, but also their publication date.

2.5.1. Principles

MFACT (Bécue-Bertaut & Pagès, 2004; Bécue-Bertaut, Álvarez-Esteban, & Pagès, 2008), an extension of MFA, can manage a multiple table juxtaposing several quantitative, categorical and frequency sets of columns. In this work, the data structure is very specific. The abstracts are described by both the vocabulary (frequency set with as many columns as different words) and the chronology (set of only one column whose different values are the publication years, considered as quantitative values). Therefore, the multiple table to be tackled juxtaposes row-wise a lexical table with as many columns as words and a quantitative table reduced to one column. MFACT gives an active role to both sets but balances their influence in the global analysis. If a relationship does exist between chronology and vocabulary, a first axis highly correlated with chronology is provided. Bécue-Bertaut et al. (2008) detail MFACT as applied to a similar data structure.

2.5.2. Types of results

In the case of the particular structure described above, MFACT offers the following:

- Global results on the active rows and columns, as in any principal axes method: eigenvalues, representations of the row-abstracts and column-words in a CA-like way (the abstracts/words are, up to a constant, the weighted average of the words/abstracts that they use/are using and a representation of the chronology through its correlation with the axis, in a PCA-like way. In this global representation, the distance between two abstracts corresponds to the weighted sum of the distances induced by both the vocabulary (as in CA performed on the lexical table) and the standardised chronology (classical Euclidean distance).
- Partial results on the active rows and columns: contribution of each set to each axis inertia; superimposed representations of the abstracts from either the chronology or the vocabulary viewpoint, easing the discovery of those abstracts that are more advanced from vocabulary than from chronology viewpoints (here, called pioneer works because their vocabulary is used more often after than before their publication date).

Illustrative columns can be introduced and displayed on the principal graphs. Year-in-categories and journal, both considered as categorical illustrative variables, are of interest here. Year-in-categories is built from the quantitative variable publication year, with the different years as the categories of this variable. Each category (year or journal) is displayed, as illustrative, at the centroid of the abstracts published during that year or in that journal, either on the global or partial (chronology or vocabulary) representations. In this case, it is worthwhile to display the following:

- The category-years from only the vocabulary viewpoint. Years' trajectory is the trace of the vocabulary evolution and its rhythm, with possible steps forward and backward or decelerations. This trajectory notes gaps in vocabulary renewal and leads to the discovery of abrupt changes and identifies lexically homogeneous periods in the corpus, further characterised by the vocabulary.
- The journals, but limited to those with a minimum of 10 abstracts in the present corpus. In this case, both partial points of view, namely chronology and vocabulary, are of interest. From the first, the journals are placed at the average date of their abstracts, indicating which period concerns them. From the second, the vocabulary viewpoint, they are placed with regard to one another based on their content.

2.5.3. Validation

A test based on permutations assesses the significance of the first global eigenvalue that accounts for the link between chronology and vocabulary. The null hypothesis is that of the exchangeability of the publication year with respect to the abstracts or, equivalently, the non-existence of a chronological dimension in the vocabulary variability. Thus, the rows of the column-year are permuted at random and, for every permutation, MFACT is performed, leading to an empirical distribution of the first eigenvalue under the null hypothesis. A large, number of replications allows for the computation of the *p*-value associated with the observed value of the first eigenvalue (one-tail test).

2.6. Lexical characteristics of the periods

As mentioned above, a chronological partition of the corpus is obtained by cutting the trajectory of the years from the vocabulary viewpoint (vocabulary partial representation) where large gaps between consecutive years point to substantial changes in the vocabulary. The lexical characterisation of the periods facilitates the understanding of the novelties responsible for the changes.

To characterise the periods, it is possible to identify the following (Lebart et al., 1998):

- The *characteristic words* of each period or words whose frequency in a period is significantly greater than what randomness would indicate. This observation provides evidence about the content of the abstracts of the period.
- The *characteristic increments* or words whose frequency significantly increases in a period compared to the previous periods. The increments denote topics that newly appear in the corpus or sharply increase their presence.
- The *chronological characteristic words*, which are the words assigned to the period or groups of consecutive periods that they best characterise. These words qualify the former information by showing that changes are not abrupt.

2.6.1. Characteristic words

With the aim of identifying the highly frequent (versus highly infrequent) words in parts of the corpus, such as those formed by grouping the abstracts on a yearly basis or by journals, the test described hereafter is performed.

The following notation is used:

- $k_{..}$, the grand total, that is, the total number of occurrences in the whole corpus;
- $k_{.j}$, the number of occurrences in part j;
- $k_{i.}$, the total count of word i in the whole corpus;
- k_{ij} the count of word i in part j.

The count k_{ij} of word i in part j is compared to the other counts that are obtained with all the possible samples composed of $k_{.j}$ occurrences randomly extracted from the whole corpus without replacement (which is the null hypothesis).

If the word i is relatively more frequent in part j than in the whole sample, that is, if $k_{ij}/k_{.j} > k_{i.}/k_{..}$ (less frequent in part j than in the whole sample), the p-value of the test is computed through (1) (through (2)):

$$p_{i.j} = \sum_{x=k_{ij}}^{k_{.j}} \frac{\binom{k_{i.}}{x}\binom{k_{..}-k_{i.}}{k_{.j}-x}}{\binom{k_{..}}{k_{.j}}} \tag{1}$$

$$p_{i.j} = \sum_{x=1}^{k_{ij}} \frac{\binom{k_{i.}}{x}\binom{k_{..}-k_{i.}}{k_{.j}-x}}{\binom{k_{..}}{k_{.j}}} \tag{2}$$

2.6.2. Characteristic increments

As a chronological corpus is manipulated, the words whose usage significantly increases in a given period (characteristic increments) are relevant information. The same test as above is performed for every couple (word, period), truncating the corpus at the end of the period under study.

2.6.3. Chronological characteristic words

Certain words better characterise a group of consecutive periods than a single period. Therefore, each word is successively tested as a possible characteristic word of each period (first level), of each group of two consecutive periods (second level), of each group of three consecutive periods (third level) and so on. A *p*-value is associated with each test. At the end of the process, each

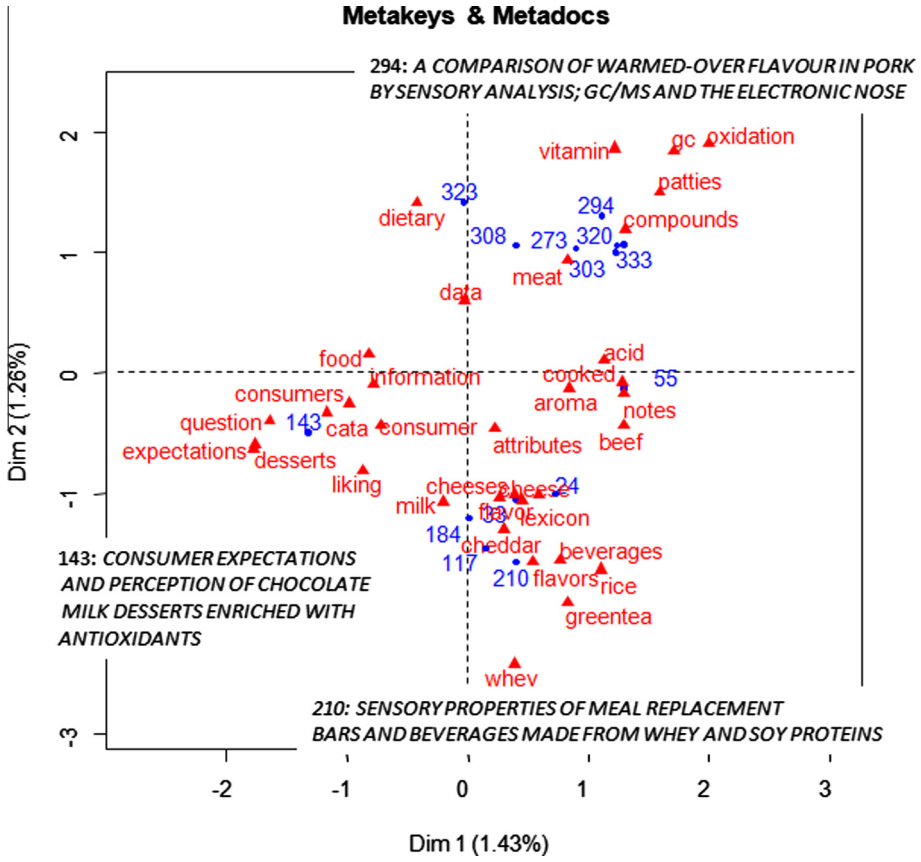

Fig. 2. Representation of the most contributory words and abstracts on the first CA principal plane.

word is assigned as a chronological characteristic word to the period or group of periods that it best characterises, i.e., to the group of periods associated with the lowest *p*-value (under the condition that this *p*-value is under 0.05).

A careful examination of the list of chronological characteristic words allows for an understanding of the flow of the vocabulary.

2.7. Statistical software

Statistical analyses were performed using R version 2.15.1 (R Development Core Team, 2012). The function "MFA" of the package FactoMineR (Husson, Josse, Lê, & Mazet, 2007; Lê, Josse, & Husson, 2008) was used for MFACT (Kostov, Bécue-Bertaut, & Husson, 2013). The R function, developed by Belchin Kostov, for characteristic words and lexical increments is available upon request from the author.

3. Results

3.1. Glossary of frequent words

The glossary provides a first overview of the abstracts through the frequent words. They refer to the following:

- Usual topics in *sensory* analysis: *perceptions* of *flavour*, *texture*, *aroma*, *taste* and *odour* are *evaluated* by *trained panellists*, *assessors* or *consumers*. *Liking*, *preference*, *acceptance*, *feelings*, *emotions* and *willingness* are *elicited*. Among the *attributes*, *descriptors* and *characteristics*, the most cited are *acid*, *sweet* and *sour*.
- Products: *cheese*, *meat* (*beef*), *dairy products*, *fruit* (*apple*), *wine*, *desserts*, *tea* (*green tea*), *bread*, *water*, *sauce*, *soy*, *potato*, *chicken*, *choi* and *whey*.
- Sensory methods: *descriptive analysis* and *profiling techniques* –frequently associated with *lexicon generation* or *development*– *free choice profile*, *flash profile*, *napping*|*projective mapping*, *check_all_that_apply*, *repertory grid* and *word association*.
- Statistical methods: *principal component analysis*, *procrustes* (also known as *procustes*), *analysis of variance* and *multiple factor analysis*.
- Words related to verbalisation: *terms*, *lexicon*, *free* (*text* or *comments*), *language* and *vocabulary*.

The glossary not only draws attention to the presence but also to the absence of words and low frequencies. *Disliking*, despite being an important driver of consumers' choices, is uncommon. *Correspondence analysis* is not as frequent as expected, although it is a reference method for analysing word count-based methods, such as *open-ended questions* but also *check_all_that_apply* and *word associations*.

3.2. CA on the abstracts × words table

3.2.1. Main features of the global analysis on the first plane

CA, performed on the table crossing the 386 abstracts and the 821 retained words, provides the metakeys and metadocs associated with the principal axes.

The first two axes, with eigenvalues equal to 0.30 and 0.26, correspond to the axes to which a high number of words and abstracts contribute (Benzécri, 1973), although they account for a minor part of the global inertia (together, 2.70%). This low percentage, typical when dealing with large spare matrices but highly significant with respect to the hypothesis of independence between words and abstracts, is frequently associated with a satisfactory structure of the data, as discussed by Lebart et al. (1998).

Fig. 2 displays the metakeys and metadocs characterising the first principal plane issued from CA. These metakeys and metadocs gather the words and abstracts whose contribution is over 6 times the mean contribution of their respective set on any of the first two axes. The configuration highlights a tripolar structure, present in both the word and abstract clouds. The first axis contrasts a set of words (or metakeys) related to consumer-oriented sensory methods and their objectives (*consumer*(*s*), *expectation*, *cata*, *liking* and *information*) to two others concerned with (1) profiling food and beverage attributes (*whey*, *cheese*(*s*), *green tea*, *brown rice*, *attributes*) and (2) resulting qualities of heat processing food (*cooked*, *beef*, *acid*, *aroma*, *compounds*, *oxidation*, *meat*, putting the accent on *dietary* properties such as *vitamin*). These two latter metakeys are, in turn, opposed on the second axis. These oppositions are also found in the metadocs, or sets of contributory abstracts. These abstracts are identified by their number on the graph and, in some cases, by their title (Fig. 2). Their consultation places the mentioned topics in their context.

A summary of the information conveyed by these and the following axes (up to 50) is provided by the list of the high dimensional words, ranked from the number of dimensions to which they contribute (information not reproduced here). First, a series of products and their sensory attributes or dietary and health-related properties are mentioned. Then, the *willingness*, *feelings*, *expectations* and *liking* of the *consumers* or close-meaning words are very cited. Finally, sensory methods, such as *napping*/*projective mapping*, *sorting task*, *CATA*, *flash profile*, *free choice profiling* and *word associations* contribute to small number of dimensions, indicating their more specialised role.

3.2.2. Relationship with chronology

The variable *year*, used as an illustrative quantitative column in this analysis, is weakly correlated with the first axes (*corr* (year, axis1) = −0.20; *corr* (year,axis2) = −0.25; *corr* (year,axis3) = 0.17). When 50 axes are considered, the quality of the representation of *year*, as projected on the corresponding subspace, is only 0.40. This analysis, although accounting for the main research topics in this domain, does not explain their evolution over time because the chronology influence is spread over many axes. Many topics are transversal to time, whereas changes are driven by multiple factors. The methods evolve towards the consumer viewpoint, first to capture their perceptions and later their likings, feelings or willingness. The consumers modify their demand towards dietary and healthy aspects, new products are introduced, and new attributes (*umami*, for instance) are considered, in particular because countries such as *Korea* have recently devoted efforts to sensory analysis of their own products. This multiplicity of factors requires introducing chronology for following the evolution of the research, which leads to MFACT.

3.3. MFACT on the multiple table juxtaposing the lexical table and the year column

3.3.1. Main features of the global analysis on the first plane

MFACT is performed on the multiple table juxtaposing the abstracts × words table (with dimensions 386 × 821) and the abstracts × year column (with dimensions 386 × 1), leading to mapping of the abstracts according to both their vocabulary and chronology. The first eigenvalue of MFACT is equal to 1.49. This value is far from randomness, as evidenced by the permutation test (p-value = 0.000 computed from 1000 replications) and is a medium value; in this case, the maximum is equal to 2 (Escofier & Pagès, 2008). Although the abstracts could not be ranked from their vocabulary in chronological order, which was not expected because vocabulary variability is influenced by other factors, a rough chronological order does exist along the first global axis.

MFACT provides a representation of the words on the first global plane that accounts for almost as much variability of the words as CA which, by construction, offers the optimum planar representation of this variability (Table 3). The first axis to which the words contribute 40.6% and the year contributes 59.4% of the inertia is a dispersion direction, present in both sets of columns, words and chronology, that differs from the first principal axis obtained in the separate analysis of the words (that is, in CA). With regard to the words, MFACT conserves a small proportion of the total inertia on the first axes as CA applied to short texts. However, because the first MFACT axis retains a proportion of inertia of the word cloud relatively close to the proportion of inertia explained by the first CA axis, it can be considered as a relevant dispersion direction for this cloud. Regarding chronology, the high correlation of the MFACT first axis with publication year (*corr* = 0.94) suggests that this axis accounts for the variability of the vocabulary related to time, which is a prime concern.

Fig. 3 shows the global representation of the words and abstracts on the first principal plane. In this figure, the abstracts are differentiated by shades of grey depending on three time periods (1990–2004; 2005–2009; 2010–2013). These time divisions are justified in Section 3.3.2. The rules for a conjoint interpretation of both figures, which could be superimposed, are those of CA. The relationship between the abstracts and chronology, not represented here, is summarised by the correlations mentioned above and as interpreted in PCA.

Concerning the representation of the words on the first axis, the information provided by Fig. 3 is completed by Fig. 4. This first axis opposes words related to accurate descriptions of several products from their sensory attributes (on the left) to words related to consumers' perceptions and consumers' willingness (on the right). As observed in Fig. 3b, the abstracts are placed in a rough chronological order on this axis. From the quantitative variable year, the categorical variable "year_in_categories" is built. The category-years (or group of years) are projected as illustrative on the first axis at the centroid of the abstracts published during this period, from the global point of view (Fig. 4). The most relevant words (mainly products and attributes, sensory or statistical methods and their objectives) are also represented on this axis from their coordinates (Fig. 4). This representation allows for the replacement of vocabulary renewal into time flow. As expected, the category-years lie on the first axis in their natural order, although they are separated by different length intervals.

The second axis (Fig. 3), minimally linked to the chronology (*corr* = 0.14), is closely related to the first axis of the separate CA (*corr* = −0.87). Therefore, this axis opposes roughly the same words and abstracts as the first CA axis. Words related to sensory and statistical methodology (negative coordinates) are opposed on this axis to several products and attributes as well as abstracts that favour one or another of these topics (positive coordinates).

3.3.2. MFACT: Partial representation of year_in_categories from a vocabulary viewpoint

The partial representation of *year_in_categories* from only the vocabulary viewpoint (Fig. 5) is of interest because it offers the chronological trajectory of the abstracts grouped on a yearly basis. Far from being regular, this trajectory presents notable gaps

Table 3
Proportion of the inertia of the cloud of words explained by the first two axes of CA and MFACT.

	Axis 1 (%)	Axis 2 (%)	Cumulate (%)
CA	1.43	1.26	2.69
MFACT	0.87	1.36	2.23

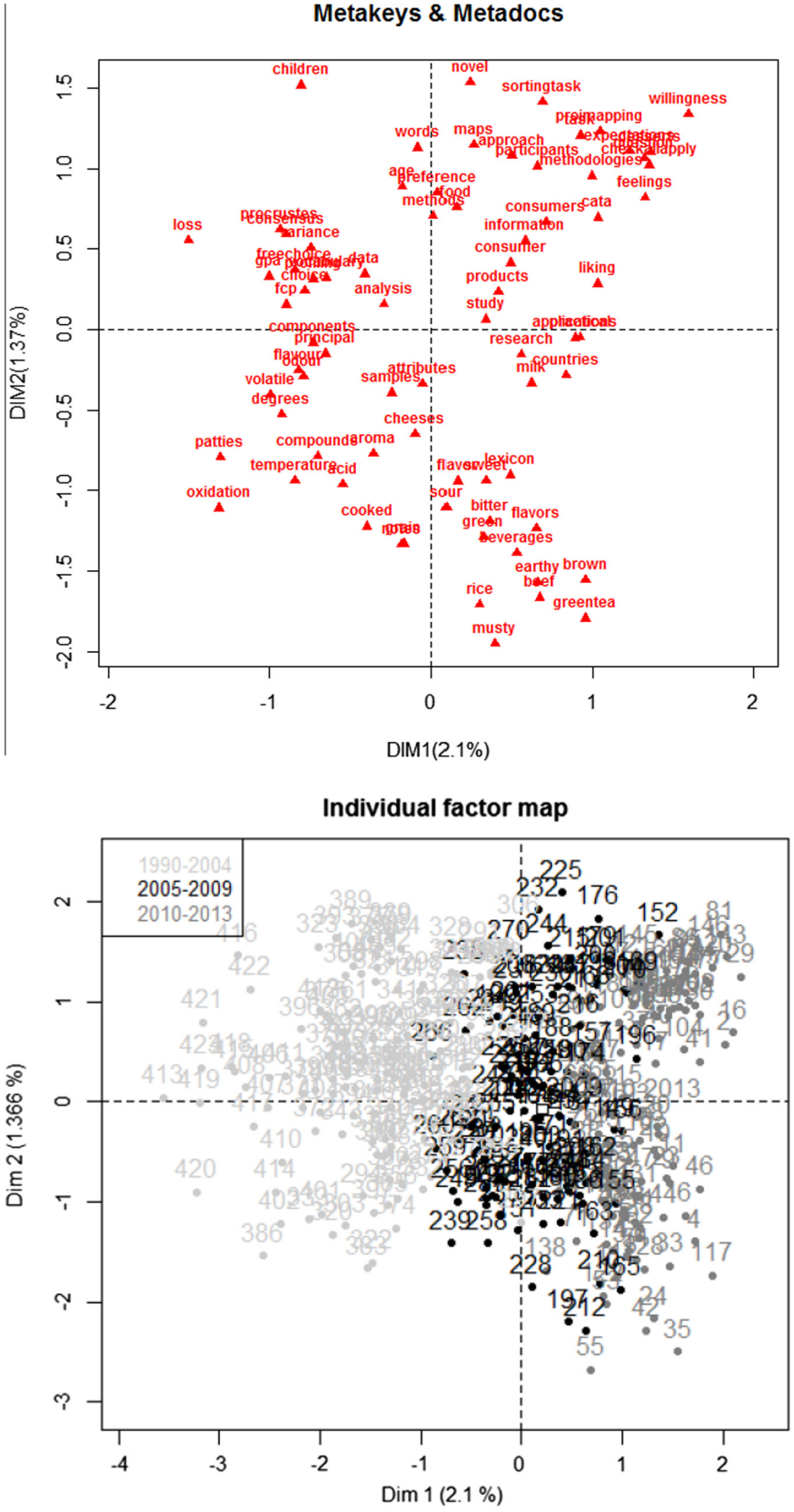

Fig. 3. Global representation of a. an excerpt of the words and b. the abstracts on the first MFACT principal plane.

corresponding to the publication of innovative works and the newly introduced vocabulary in the corresponding year. Behind the observed trajectory, different factors are at work, especially laboratories and researchers that broaden and widen their innovative contributions, which are also applied and discussed by other laboratories. Researchers come from different fields (sensory analysis, statistics, marketing, psychology) with their own concerns but also methods and vocabulary. As the corpus is

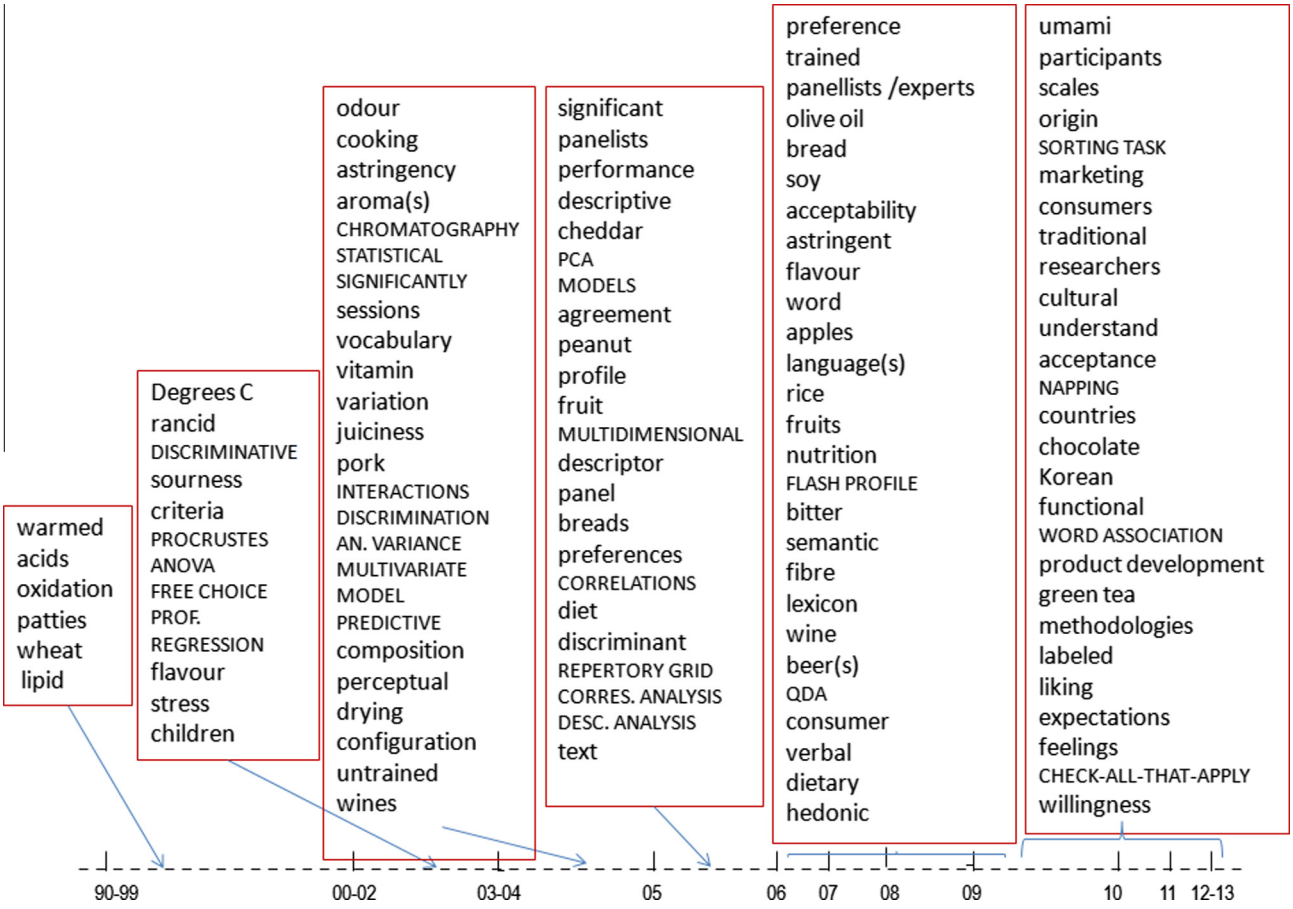

Fig. 4. Global representation of an excerpt of the words (active) and *year_in_categories* (illustrative) from their coordinates on the first MFACT axis.

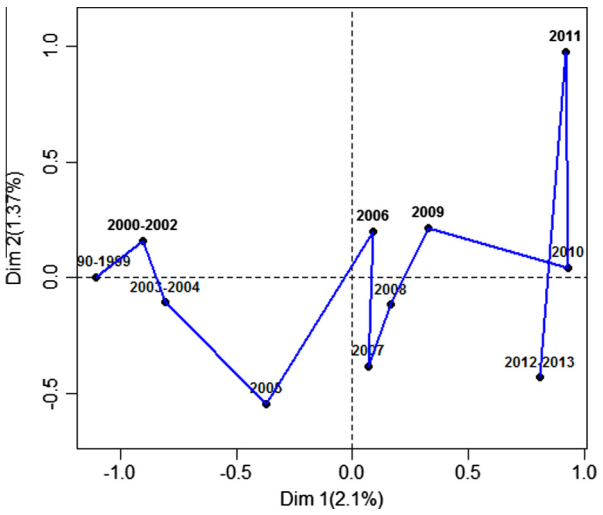

Fig. 5. Partial representation of *years_in_categories* (illustrative) from only the vocabulary point of view on the first MFACT principal plane.

relatively small for this type of study, the local phenomena are not smoothed and should not be considered as relevant.

Nevertheless, general trends can be identified. From 1990 to 2004, little change in vocabulary is observed. A transition occurred in 2005, marked by the introduction of new words that persisted in 2006, as shown in the next section. This vocabulary was used until 2009. Another gap between 2009 and 2010 shows that the latter year is marked by a new vocabulary, which was later used during the "2010–2013" period.

3.4. Segmentation of the corpus into lexically homogeneous periods

The segmentation suggested by Fig. 5 and commented on in the former section is retained, and the periods 1990–2004, 2005–2009 and 2010–2013 are described by their lexical features, enabling the characterisation of the evolution of the vocabulary. The results are consistent with those of the former sections.

Fig. 6 shows how the studies in sensory analysis related to verbal descriptions have diversified from 1990 to 2013.

In the 1990–2004 period, the main issue was about generating a lexicon for the sensory description of products, possibly new products, by trained panels. The descriptors were discussed (*lexicon*, *vocabulary*) and *Free Choice Profile*, introduced by Williams and Langron (1984), was used to overcome the difficulties of forcing panellists to adopt a common definition of descriptors. As a remark, the innovative work by Williams and Langron (1984) has no abstract and therefore does not belong to the database considered here.

Both 2005 and 2006 brought many new words used without noticeable change until 2009, reflecting novelties in objectives and methods. These words show a new interest for consumers' acceptance and preferences, which led to the introduction or revival of holistic methods (*flash profile*, *napping*, *sorting task*, *word association*). New products (*bread*, *beef*) and new aspects of the products (*fibre*, *freshness*), possibly linked to dietary properties were the aim.

1990-2004

Chronological characteristic words

analysis of variance, assessors, descriptive, free choice profiling, regression, terms, training, vocabulary, children

2005-2009

Chronological characteristic words

bitterness, cheese, color, dry, floral, foods, fresh, freshness, languages,novel, panelists, proteins, refrigerated, sausages,sensations, technologies, whey, women

Characteristic increments

acceptance, beef, beverages, bitterness, bread, color, consumer(s), conventional, dietary, fibre, findings, flash profile, flavors, floral, freshness, languages, lexicon, liking, napping, origin, panelists, participants, research(ers), sorting task, technologies, texting, wine, word association

2010-2013

Chronological characteristic words

almond, application(s), bitter, brand, check-all-that-apply, CATA, coffee, consumers, desserts, expectations, feelings, green tea, hedonic, knowledge, liked, methodologies, package, perception, product development, rice, salty, understand(ing), semantic, sweet, umami, vacuum, willingness

Characteristic increments

almond, beef, bitter, brand, check-all-that – apply, coffee, consumer(s), desserts, expectations, feelings, flavors, green tea, hedonic, knowledge, labeled, lexicon, liked/liking, methodologies, napping, origin, product development, professionals, qda, research, rice, salty, scales, semantinc, traditional, umami, understand(ing) vacuum, willingness, wine

1990-2009

Chronological characteristic words

acid(s), appearance, attribute, breads, cheddar, cheese, component, differentiation,fat, fcp, flavour,gene, procrustes/ gpa, juice, multivariate, nose, odour, perceptions, scaling, significantly, prediction, untrained, vitamin, wines

2005-2013

Chronological characteristic words

acceptance, beef, beverages, characterization, chocolate, cluster, consumer, conventional, dietary, elicited, fibre, finding,firmness, flavor(s),functional, industry, labeled, lexicon, liking, napping, origin, participante, professionals, qda; research, researchers, sorting taks, taste, tradicional, wine, word association

Fig. 6. Lexical chronological characteristics and significant lexical increments.

The words introduced in 2010 showed an increasing interest for consumers and not only their *likings* but also their *feelings*/*emotions* and *willingness* that researchers attempt to *understand*. New products were also studied.

Although topics of interest prior to 2005 remain in the following years, there is a more noticeable homogeneity in the vocabulary observed from 2005–2013 than from 1990–2009, as shown by the greater number of chronologically characteristic words relative to 2005–2013 than to 1990–2009. This observation is reinforced by the strong similarity between the lexical increments observed in 2005–2009 and 2010–2013. For instance, *napping* appears from 2006–2009 (5 citations) and its usage increases from 2010–2013 (33), in some cases named as projective mapping. The words *consumer*(*s*) are used 110 times from 1990 to 2004, 149 times from 2005 to 2009 and 306 times from 2010 to 2013. Still, sensory science did not lose sight of its scope and continued exploring more and more diversified products (*desserts*, *beef*, *fillets* and *pomegranate*). In parallel, the methodology becomes more complex and requires more sophisticated statistical methods, frequently imported from other fields.

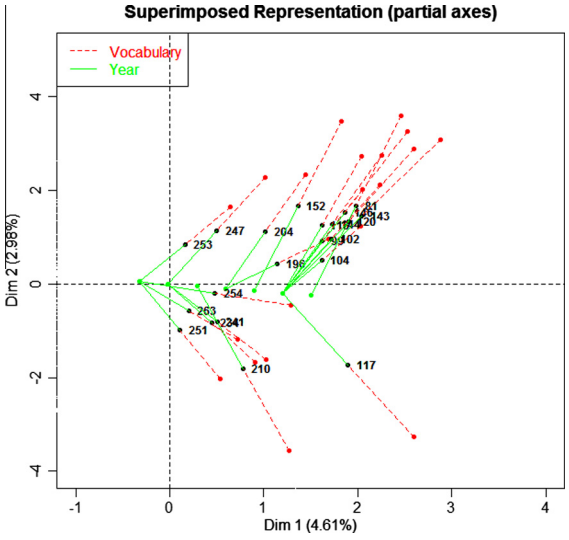

Fig. 7. Representation of the ahead-of-date abstracts on the first MFACT principal plane from 2005 to 2011.

3.5. Abstract s with ahead-of-their-date vocabulary

The superimposed representation of the partial points-abstracts from either the chronology or vocabulary viewpoint allows for spotting the abstracts with a vocabulary ahead of their date. However, the set of words contributes to only 40.5% of the inertia of the first axis, whereas the column-year contributes 59.4%. This unequal contribution of both sets of columns to the first axis inertia makes the partial points "chronology" placed on average farther from the centroid on the first axis than the partial points "vocabulary". This is particularly clear for the oldest dated abstracts due to their low numbers in the same year. These abstracts tend to mechanically present a partial point "vocabulary" ahead of the partial point "chronology". On the contrary, this effect, notably

Table 4
Excerpt of the ahead-of-date abstracts from 2009 to 2011.

Nr.	Title	First author	Journal	Year
81	Consumers' perception of farmed fish and willingness to pay for fish welfare	Solgaard, HS	British Food Journal	2011
143	Consumer expectations and perception of chocolate milk desserts enriched with antioxidants	Ares G.	Journal of Sensory Studies	2010
152	A Factorial Approach for Sorting task data (FAST)	Cadoret M.	Food Quality and Preference	2009

less marked, works the opposite way in the case of the recent abstracts, which allows for interpreting partial points "vocabulary" ahead of partial points "chronology" as corresponding to abstracts using an innovative vocabulary (Fig. 7). These abstracts contain words with a high coordinate on the first axis of MFACT (Fig. 4) and use these words at an early stage. Table 4 gives the title, first author, journal and year of the most ahead-of-its-date abstract in 2009, 2010 and 2011, although all should be consulted. A gap in the number of ahead-of-date abstracts (1 in 2011, 9 in 2010, 2 in 2008, 1 in 2007, 3 in 2006 and 4 in 2005) corresponds to a gap in the trajectory of the years from the viewpoint of vocabulary (Fig. 5). Examining their content facilitates the understanding of the research advances. The ahead-of-date abstracts often correspond to new innovations.

The most recent abstracts cannot be selected from this criterion (no pioneer work has been identified for either 2012 or 2013), and recent innovative articles may have been missed. Nevertheless, it is possible to look for recent abstracts (published either in 2012 or 2103) that favour advanced vocabulary. They are detectable because they present a high partial coordinate on the first axis from the vocabulary viewpoint. For instance, abstract-2 (Thompson & Crocker, 2013) was selected from this criterion. Dealing with investigating the feelings of consumers, it uses many recently introduced words with high coordinates on the first axis, such as *willingness* and *feeling*.

3.6. Profiles of the journals

Fig. 8 offers the global and partial representations of the 5 journals containing at least 10 articles in this corpus (*Journal of Sensory Studies*, *Food Quality and Preferences*, *Journal of Food Science*, *Meat Science* and *Journal of the Science of Food and Agriculture*). Each journal is placed on the axes at the centroid of the abstracts published in the journal. On the first axis, all of the journals have their global and partial coordinates close to each other. This placement denotes, on average, an agreement between the vocabulary and date. On the second axis, all of the journals have partial coordinates from the chronology viewpoint close to the centroid because chronology does not contribute to the inertia of this axis. The partial viewpoint given by the vocabulary is, on the contrary, of interest. The position of each journal from this partial coordinate reveals its focus according to the words opposed on this axis (Section 3.3.1). Thus, the *Journal of Sensory Studies* and the *Journal of the Science of Food and Agriculture* are both product- and methodology-oriented journals, as far as verbalisation tasks are concerned. *Food Quality and Preferences* is methodology oriented, whereas the *Journal of Food Science* is product oriented. Finally, *Meat Science*, which is devoted to products (meat in this case), in accordance with its coordinate on the second axis, is characterised by minimal use of newly introduced words, as indicated by its coordinate on the first axis.

4. Discussion

This study provides several insights regarding the scientific bibliography related to alternative verbal-based methods in food science. The main topics are the different products and their attributes, the effects of processing food, the dietary and health-related qualities of the products and the use, discussion and comparison of the sensory methods. These alternatives have to face objectives in constant evolution, driven by a growing need to better understand

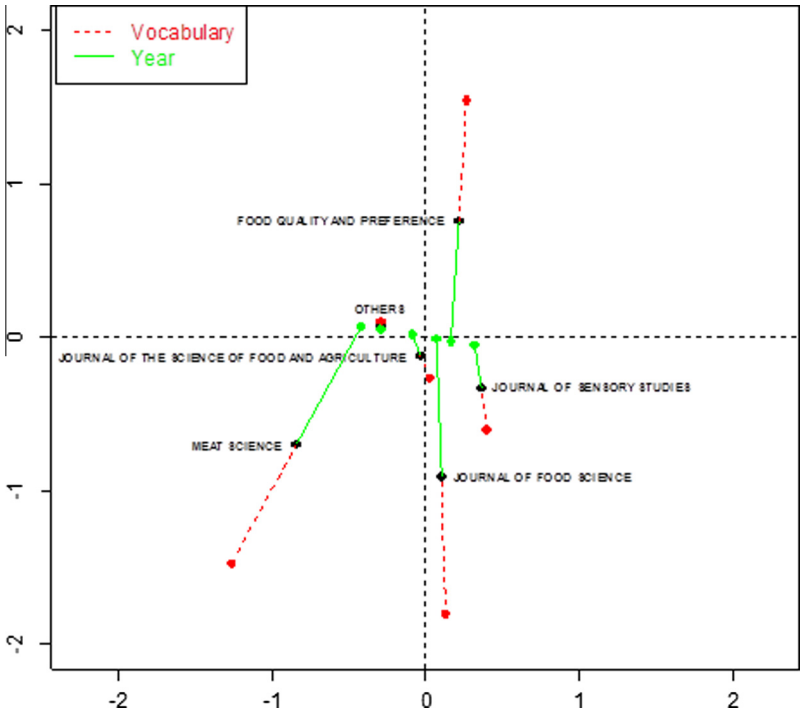

Fig. 8. Superimposed global and partial representations of the journals with more than 10 abstracts on the first MFACT principal plane.

consumers' ever-evolving habits. Their perceptions were first considered, then their likings and, recently, their feelings and willingness as prime concerns. These changing objectives stimulate the search for new sensory methods, possibly imported from the marketing domain. The multiple sources interact, causing their trends to scatter in different directions.

If journals are considered, the study shows that the scientific research concerned with the studied verbal-based methods is concentrated in 5 main journals, which nevertheless differ in their focus towards either more methodological- or more product-oriented methods.

Identifying the abstracts with ahead-of-date vocabulary reveals milestones in research advances. An evident limitation of this methodology is its inability to identify the recent works that will be identified as innovative advances in the near future. Nevertheless, other methods favouring recently introduced words can bring clues to the future value.

Among the statistical methods used to analyse verbalisation data, CA is little used (Section 3), whereas word count-based methods (such as open-ended questions, free text comments, word associations and check-all-that-apply) could take advantage of CA and MFACT properties. Emphasis must be put on the ability of CA to identify words with similar meanings, provided that they are used in similar contexts, and texts with similar content, provided that they use either the same words or synonyms. Therefore, tedious pre-processing of the corpus aimed at unifying synonyms or different forms of a lemma (such as the singular and plural of a noun) is not necessary.

With respect to the statistical approach, relying on an intensive use of correspondence analysis and related methods is indispensable when facing a large set of abstracts. A classical human reading cannot synthesise such a large number of abstracts without making subjective decisions that can lead to the inadvertent omission of essential features. In contrast, the statistical methods systematically operate and render an overall view by counting the words and comparing their distributions in the whole set of abstracts. CA maps the abstracts from the words, closely related to their content, and the words from the viewpoint of the abstracts using them while establishing relationships between both sets of abstracts and words. Therefore, the different end-users can easily find the abstracts of interest by identifying the proper metakeys and corresponding topics, which lead to the associated abstracts.

The application of MFACT to a bibliography study, an innovative contribution of this work, is useful when many sources are at work to stimulate changes in the publications, which is the case here. MFACT forces the computing of the first principal axis to be related to both word dispersion and chronology. This approach allowed for the identification of the common direction of the change trends of the different sources.

The decision to deal only with abstracts related to a targeted question is another issue. The advantage is that it allows for setting a well-delimited reference corpus of abstracts. The drawback is that this corpus reflects a reduced part of the literature. Consequently, the selection step is crucial. With these cautions in mind, the reported results are of interest. This method could be enlarged to the study of the general literature in food science.

However, some limitations must be emphasised to avoid misinterpretations and non-pertinent conclusions. A critical point is that only the most frequent words are taken into account because, from a statistical point of view, words and co-occurrences of words have to be highly frequent to define robust relations among abstracts. Another limitation is that this procedure cannot be fully automated because a complete interpretation of the results requires interaction with specialised users, possibly leading to local studies on a subcorpus of their interest.

The viewpoint of the abstracts provided by this study and its results differs from a classical review, such as those cited in the Section 1. However, this approach can be used as a first step in a systematic review on a well-determined topic. It eases the choice of the articles to be selected, which turns this statistical approach into a strategic tool to follow the trends in food science or other scientific fields.

Acknowledgment

This work has been partially supported by the Catalan Agency for the Administration of University and Research Grants (grant 2009SGR688).

References

Alfaraz, P. H., & Mirta-Calviño, A. (2004). Bibliometric study on food science and technology: Scientific production in Iberian-American countries (1991–2000). *Scientometrics, 61*, 89–102.

Ares, G., Barreiro, C., Deliza, R., Giménez, A., & Gámbaro, A. (2010). Consumer expectations and perception of chocolate milk desserts enriched with antioxidants. *Journal of Sensory Studies, 25*, 243–260.

Ares, G., Deliza, R., Barreiro, C., Giménez, A., & Gámbaro, A. (2010). Comparison of two sensory profiling techniques based on consumer perception. *Food Quality and Preference, 21*, 417–426.

Ares, G., Giménez, A., Barreiro, C., & Gámbaro, A. (2010). Use of an open-ended question to identify drivers of liking of milk desserts. Comparison with preference mapping techniques. *Food Quality and Preference, 21*, 286–294.

Ares, G., Giménez, A., & Gámbaro, A. (2008). Understanding consumers' perception of conventional and functional yogurts using word association and hard laddering. *Food Quality and Preference, 19*, 636–643.

Baccino, T., Cabrol-Bass, D., Candau, J., Meyer, C., Scheer, T., Vuillaume, M., et al. (2010). Sharing an olfactory experience: The impact of oral communication. *Food Quality and Preference, 21*, 43–452.

Bansard, J. Y., Kerbaol, M., & Coatrieux, J. L. (2006). An analysis of IEEE publications in biomedical engineering. *IEEE Engineering in Medicine and Biology Magazine, 25*, 6–12.

Bansard, J. Y., Rebholz-Schuhman, D., Cameron, G., Clark, D., van Mulligen, E., Beltrame, F., et al. (2007). Medical informatics and bioinformatics: A bibliometric study. *IEEE Transactions on Information Technology in Biomedicine, 11*, 237–243.

Barcenas, P., Pérez Elortondo, F. J., Salmerón, J., & Albisu, M. (1999). Development of a preliminary sensory lexicon and standard references of ewes milk cheeses aided by multivariate statistical procedures. *Journal of Sensory Studies, 14*, 161–179.

Bécue-Bertaut, M., Álvarez-Esteban, R., & Pagès, J. (2008). Rating of products through scores and free-text assertions: Comparing and combining both. *Food Quality and Preference, 19*, 122–134.

Bécue-Bertaut, M., & Lê, S. (2011). Simultaneous analysis of categorical and multilingual textual variables structured according to a hierarchy. Application to a cross-cultural study in wine industry. *Journal of Sensory Studies, 26*, 299–310.

Bécue-Bertaut, M., & Pagès, J. (2004). A principal axes method for comparing multiple contingency tables: MFACT. *Computational Statistics & Data Analysis, 45*, 481–503.

Bécue-Bertaut, M., & Pagès, J. (2008). Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Computational Statistics & Data Analysis, 52*, 3255–3268.

Benzécri, J. P. (1973). Analyse des données. *Analyse des correspondances* (Vol. 2). Paris: Dunod.

Benzécri, J. P. (1981). *Pratique de l'analyse des données. Linguistique & lexicologie* (Vol. 3). Paris: Dunod.

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.

Cadoret, M., Lê, S., & Pagès, J. (2009). A factorial approach for sorting task data (FAST). *Food Quality and Preference, 20*, 410–417.

Cairncross, S. E., & Sjostrom, L. B. (1950). "*Flavor* profiles – A new approach to *flavor* problems. *Food Technology, 4*, 308–311.

Carr, B. T., Craig-Petsinger, D., & Hadlich, S. (2001). A case study in relating sensory descriptive data to product concept fit and consumer vocabulary. *Food Quality and Preference, 12*, 407–412.

Civille, G. V., Lapsley, K., Huang, G., Yada, S., & Seltsam, J. (2010). Development of an almond lexicon to assess the sensory properties of almond varieties. *Journal of Sensory Studies, 25*, 146–162.

Dairou, V., & Sieffermann, J.-M. (2002). A comparison of 14 jams characterized by conventional profile and a quick original method, the Flash Profile. *Journal of Food Science, 67*, 826–834.

Delarue, J., & Sieffermann, J. M. (2004). Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference, 15*, 383–392.

Dooley, L., Lee, Y., & Meullenet, J. F. (2010). The application of check-all-that-apply (CATA) consumer profile to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. *Food Quality and Preference, 21*, 394–401.

Escofier, B., & Pagès, J. (2008). *Analyses factorielles simples et multiples* (4th ed.). Paris: Dunod.

Faye, P., Brémaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects; an alternative to descriptive mappings. *Food Quality and Preference, 15*, 781–791.

Feinerer, I. (2013). Introduction to the tm package text mining in R. http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf.

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*, 1–54 [http://www.jstatsoft.org/].

Galmarini, M. V., Symoneaux, R., Chollet, S., & Zamora, M. C. (2013). *Appetite, 62*, 27–36.

Guerrero, L., Claret, A., Verbeke, W., Enderli, G., Zakowska-Biemans, S., Vanhonacker, F., et al. (2010). Perception of traditional food products in six European regions using free word association. *Food Quality and Preference, 21*, 225–233.

Husson, F., Josse, J., Lê, S., & Mazet, J. (2007). FactoMineR: Factor analysis and data mining with R. R package version 1.19. http://www.cran.R-project.org/package=FactoMineR.

Kerbaol, M., Bansard, J. Y., & Coatrieux, J. L. (2006). An analysis of IEEE publications. *IEEE Engineering in Medicine and Biology Magazine, 25*, 6–9.

Kinski, B. F., Drake, M. A., Civille, G. V., Dean, L. L., Hendrix, K. W., & Sanders, T. H. (2006). The development of a lexicon for frozen vegetable soybeans (edamame). *Journal of Sensory Studies, 21*, 644–653.

Kostov, B., Bécue-Bertaut, M., & Husson, F. (2013). Multiple factor analysis for contingency tables in factomineR package. *R Journal, 5*, 29–39.

Lancaster, B., & Foley, M. (2007). Determining statistical significance for choose-all-that-apply question responses. In: *7th Pangborn sensory science symposium*. August 16th 2007, Minneapolis, MN, USA.

Lawless, L. J. R., Hottenstein, A., & Ellingsworth, J. (2012). The McCormick spice wheel: a systematic and visual approach to sensory lexicon development. *Journal of Sensory Studies, 27*, 37–47.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software, 25*.

Lebart, L., Salem, A., & Berry, L. (1998). *Exploring textual data*. Dordrecht: Kluwer.

Lee, Y., Findlay, C., & Meullenet, J. F. (2013). Experimental consideration for the use of check-all-that apply questions to describe the sensory properties of orange juices. *International Journal of Food Science & Technology, 48*, 215–219.

Moussaoui, K. A., & Varela, P. (2010). Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference, 21*, 1088–1099.

Murray, J. M., Delahunty, C. M., & Baxter, I. A. (2001). Descriptive analysis: past, present and future. *Food Research International, 34*, 461–471.

Murtagh, F. (2005). *Correspondence analysis and data coding with Java and R*. Boca Raton: Chapman & Hall.

Perrin, L., & Pagès, J. (2009). Construction of a product space from the ultra-flash profiling method: Application to ten wines from the Loire Valley. *Journal of Sensory Studies, 24*, 372–395.

Perrin, L., Symoneaux, R., Maître, I., Asselin, C., Jourjon, F., & Pagès, J. (2008). Comparison of three sensory methods for use with the Napping® procedure: Case of ten wines from Loire Valley. *Food Quality and Preference, 19*, 1–11.

Puyares, V., Ares, G., & Carray, F. (2010). Searching a specific bottle for Tannat wine using a check-all-that apply question and conjoint analysis. *Food Quality and Preference, 21*, 684–691.

Quarmby, A. R., & Ratkowsky, D. A. (1988). Free-choice flavour and odour profiling of fish spoilage: Does it achieve its objective? *Journal of Science of Food and Agriculture, 44*, 89–98.

R Development Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roininen, K., Arvola, A., & Lähteenmäki, L. (2006). Exploring consumers' perceptions of local food with two different qualitative techniques: Laddering and word association. *Food Quality and Preference, 17*, 20–30.

Sieffermann, J. M. (2000). Le profil Flash: Un outil rapide et innovant d'évaluation sensorielle descriptive. *AGORAL Douzièmes rencontres L'innovation: de l'idée au succè* (pp. 335–340). Paris: Lavoisier.

Šilic, A., Morin, A., Chauchat, J. H., & Dalbelo Bašic, B. (2012). Visualization of temporal text collections based on correspondence analysis. *Expert Systems with Applications, 39*, 12143–12157.

Simeone, M., & Marotta, G. (2010). Towards an integration of sensory research and marketing in new food products development: A theoretical and methodological review. *African Journal of Business Management, 4*, 4207–4216.

Solgaard, H. S., & Yang, Y. (2011). Consumers' perception of farmed fish and willingness to pay for fish welfare. *British Food Journal, 113*, 997–1010.

Stepanek, J. (2006). Musical sound timbre: Verbal description and dimensions. In *Proceedings of the international conference on digital audio effects* (DAFx-06) (pp. 121–126.). Montréal, Quebec, Canada.

Strigler, F., Touraille, C., Sauvageot, F., Barthélémy, J., Issanchou, S., & Pagès, J. (2009). Les épreuves discriminatives et descriptives. In F. Depledt (Ed.), *Evaluation sensorielle. Manuel méthodologique* (pp. 99–197). Paris: Lavoisier.

Symoneaux, R., Galmarini, M. V., & Mehinagic, E. (2011). Comment analysis of consumer's likes and dislikes as an alternative tool to preference mapping. A case study on apples. *Food Quality and Preference, 24*, 59–66.

Ten Kleij, F., & Musters, P. A. D. (2003). Text analysis of open-ended survey responses: A complementary method to preference mapping. *Food Quality and Preference, 14*, 43–52.

Thamke, I., Dürrschmid, K., & Rohm, H. (2009). Sensory description of dark chocolates by consumers. *LWT – Food Science and Technology, 42*, 534–539.

Thompson, D. M. H., & Crocker, C. (2013). A data-driven classification of feelings. *Food Quality and Preference, 27*, 137–152.

Tournier, C., Martin, C., Guichard, E., Issanchou, S., & Sulmont-Rossé, C. (2007). Contribution to the understanding of consumers' creaminess concept: A sensory and a verbal approach. *International Dairy Journal, 17*, 555–564.

Valentin, D., Chollet, S., Lelièvre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology, 47*, 1563–1578.

Van Kleef, E., van Trijp, H., & Luning, P. (2005). Consumer research in the early stages of new product development: a critical review of methods and techniques. *Food Quality and Preference, 16*, 181–201.

Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International, 48*, 893–908.

Veinand, B., Godefroy, C., Adam, C., & Delarue, J. (2011). Highlight of important product characteristics for consumers. Comparison of three sensory descriptive methods performed by consumers. *Food Quality and Preference, 22*, 474–485.

Vijay, K. R., & Raghavan, I. (2007). Journal of food science and technology: A bibliometric study. *Annals of Library and Information Studies, 54*, 207–212.

Williams, A. A., & Langron, S. P. (1984). The use of free choice profiling for the evaluation of commercial ports. *Journal of Science of Food and Agriculture, 35*, 558–568.

Worch, T., Lê, S., & Punter, P. (2010). How reliable are the consumers? Comparison of sensory profiles from consumers and experts. *Food Quality and Preference, 21*, 309–318.