Nudging consumers for relevant data using Free JAR profiling: An application to product development

Alexiane Luc, Sébastien Lê, Mathilde Philippe

Nudging consumers for relevant data using Free JAR profiling: an application to product development

Alexiane LUC - Corresponding Author - alexiane.luc@gmail.com<mailto:alexiane.luc@gmail.com> : (1) (2)

Sebastien Lê : (1)

Mathilde PHILIPPE : (2)

(1) Agrocampus Ouest, Rennes, FRANCE

(2) Strategir, Bordeaux, FRANCE

ABSTRACT

Nudging is nowadays often used in order to influence people's behaviours. However, we are convinced that nudging can also influence people's speeches. In a product development context, consumers are more and more integrated in the process: how integrate them in the most efficient way? Whereas formatted measures can be easily analysed but are far from the consumers, measures that seem to be closer to the consumers are often very difficult to analyse. A trade-off has to be found, in order to get data that are, at the same time, as close as possible to consumers' mind and relevant for the ones who develop new products. In this context, we proposed a new strategy, based on the nudge theory, called "Free JAR Profiling". Its first step invites consumers to classify products into 3 hedonic categories. This classification impact consumers' speeches in a soft way as it places consumers in a hedonic state of mind and allows them to easily structure their thoughts toward product improvement keys. The second step invites consumers to assess the products according to their own vocabulary, but with a constraint on how they give their opinion, based on a JAR scale. This strategy allows to have access to oriented and structured data, which mix the consumer spontaneity strength and the R&D approach of JAR scales. The added-value and the relevance of these nudged data can be checked, and a *Sentiment Mapping* can be performed as a results representation.

Key Words: Product development; Nudge; Free JAR; Data consistency; Sentiment score

1. INTRODUCTION

During the last two decades, the role of the consumer has drastically changed in the process of product development and product innovation.

Initially limited to hedonic assessments, consumers were then asked to provide their own sensory perception of the products. This sensory perception has moved from a very formatted perception, based on a list of descriptors a priori, to a freer perception, as consumers now provide a sensory description based on their own criteria. We have moved from a classic consumer profile to a rather holistic sensory profile, particularly achievable through a sorting step.

In order to get as close as possible to the consumer's thoughts, another strategy was to use the consumer's natural language: the rationale behind is that the less consumers are solicited, the closer the collected data will be to their real thoughts. In the same way, strategies that aim to decode body language, facial language, emotional and eye-tracking measures have also been used.

Nowadays, the question is no longer to integrate the consumer in the product development process, as we have already understood that consumers need to be integrated, but to know how to integrate them in the most efficient possible way. Whereas formatted measures can be easily analysed but can also be far from the consumers, measures that seem to be closer to the consumers are often very difficult to analyse and to interpret. A trade-off has to be found, in order to get data that are at the same time as close as possible to the consumer's mind, relatively easy to handle, and relevant for the ones who develop new products.

In this article, we propose an innovative strategy for providing such data, based on the nudge theory (Thaler et Sunstein 2008): a friendly push to encourage a desired behaviour in a simple, costless and non-coercive way. Nudging is widely used to stimulate healthy behaviours as hand hygiene (Caris et al. 2018), smoking cessation (Sunstein 2015), or reduction of energy consumption (Hunt Allcott 2011; H. Allcott et Mullainathan 2010; Hunt Allcott et Rogers 2014; Costa et Kahn 2013; Momsen et Stoerk 2014). For example, Thaler and Sustein highlight that a nudge can make people reduce their use of energy, in peak period, by 40%. A little ball that glows red when a customer is using lots of energy, but green when energy consumption is modest was implemented. It plays the role of a nudge as it unconsciously encourages people not to want the ball to be red. Without the nudge, the energy consumption is 100 units whereas, thanks to the nudge, the energy consumption is 60 units (Thaler et Sunstein 2008).

Thus, it's demonstrated that nudges have an impact on people's behaviours. But a question appears: can they be used in order to impact people's speeches?

In fact, regarding our issues, we are convinced that a nudge can influence people to the desired speech. We aim to have access to a nudged textual data that present an added-value compared to a non-nudged textual data, by providing relevant, reliable and exploitable results in a product development context. By exploitable and reliable textual data, we mean textual data that highlight product improvement keys.

This nudging may emerge from two strategies' state of minds that are already developed.

One of the developed strategies is free comments data recording, developed in order to let the consumer free during their product assessment, without any constraints and with their own vocabulary. The state of mind of this strategy is to give strength and value to the spontaneity of the consumer. However, the fact is that today, the unconstraint assessment via free comments isn't satisfactory because it provides low quality data, which aren't directly usable in a product development context. In fact, in this context, the main idea is to provide product improvement keys, whereas free comment doesn't allow us to obtain this kind of data. When consumers are asked to answer the question "Why don't you like this product?", we notice that it is often difficult for them to express themselves accurately and in an operational way (Lawless et Heymann 2010; Symoneaux, Galmarini, et Mehinagic 2012). Moreover, the analysis and the interpretation of data coming from this method could be complex and time consuming (Lawrence et al. 2013). Free comments are often use in order to describe products, but to our knowledge, the reflexion of using these descriptions in a product development context has never been done.

Another strategy is the JAR (*Just About Right*) method. JAR scales measure the appropriateness of the level of a specific attribute, and are used to determine the optimum levels of attributes in a product: they are widely used in food industry for product development as it mix hedonic assessment and sensory attributes (Rothman et Parker 2009). In practice, panellists evaluate a list of predefined attributes and have to say whether an attribute is too intense, just about right (JAR) or not intense enough (Richard Popper 2014). This method makes it possible to judge the intensity perceived in a product relative to the desired intensity. The state of mind of this strategy is to provide product improvement keys in a precise and structured way. Nevertheless, this method has some limits and doesn't seem to reflect the reality, due to the predefined list of attributes. This predefined list may be seen as a limitation of the method as it makes these attributes especially salient in consumer's mind, which may represent a bias as it can modify their ideal perception of the product (Ares et al. 2017; Epler, Iv, et Kemp 1998; R. Popper et al. 2004) or drive them to err. Indeed, they would be able to imagine sensory perceptions that do not objectively exist in the product (Lawrence et al. 2013).

In the same way, some major attributes, unperceived by the project leader but significant for the consumers, may be missing. Thus, it severely restricts the consumers because they don't express themselves according to their own sensory attributes and, in a product development context, this information is precisely the one that is relevant. On the other hand, using a predefined list can be useful as it can remind consumers of attributes they might overlook, but it also means taking the risk of emphasize attributes that are, in fine, not important for consumers. In fact, when consumers freely express themselves, they evaluate the product only according to the sensory descriptors they consider important and salient. This makes it possible to highlight the relative importance of the different descriptors in the perception of products.

These two methods are valuable and we aim to combine their state of mind in order to have access to a *nudge* strategy: the spontaneity strength of the free comments and the structure and R&D approach of the JAR. This new method will take into account the respective advantages of those two developed methods while avoiding their respective faults. Let's think about the interaction between the subject and the stimulus: in free comments, the subject is leading while in JAR, this is the stimulus. Our objective is to have access to a textual data balanced between the subject and the stimulus. Therefore, the aim of the present work is to design a new analysis strategy which unconsciously influences the consumers to be free during their product assessment while having an oriented and structured speech: it seems reliable to unlock consumer truth and identify product improvement keys. Beyond the protocol, it is also necessary to test the potential of this new analysis strategy and check the quality of the resulting data.

In this context, a new method called "Free JAR Profiling" was proposed to provide nudged textual data.

2. MATERIALS AND METHODS

2.1. Free JAR Profiling protocol

2.1.1. Protocol

This new developed strategy allows panellists to evaluate a set of products according to their own vocabulary, but based on a JAR scale. The main idea behind is to allow consumers to use their own words, but with a constraint on how they give their opinion. Constraint consumers' speech via a JAR scale highlights faults and qualities of each product according to a consumer point of view, linking sensory descriptors and hedonic assessment.

Free JAR is a two-step-methodology, which is based on the nudge theory:

- The first step is a classification step: after discovering the whole product space, consumers have to classify products into 3 hedonic categories: "I don't like it", "I like it a bit" and "I like it a lot". This graduation may be thinner by including intermediate categories, but a three-point hedonic scale may be acceptable in our case as we only want to place consumers in a hedonic state of mind. Thanks to this classification, we can impact consumers' speeches in a soft way: it allows them not only to be in a hedonic state of mind, but essentially to have a global vision of the product space and to structure their thoughts in this way: they unconsciously highlight product's qualities (if the product was placed in the "I like it a lot" hedonic category) and product's faults (if the product was placed in the "I don't like it" hedonic category). The knowledge of the products' characteristics in relation to each other feed the descriptions and make them richer than free comments.

- Then, once each product is classified, consumers have to describe them using their own descriptors according to a JAR syntax, as explained above: this is the second step of the methodology. Some examples are given to lead consumers to the right expression (**Fig.1**).

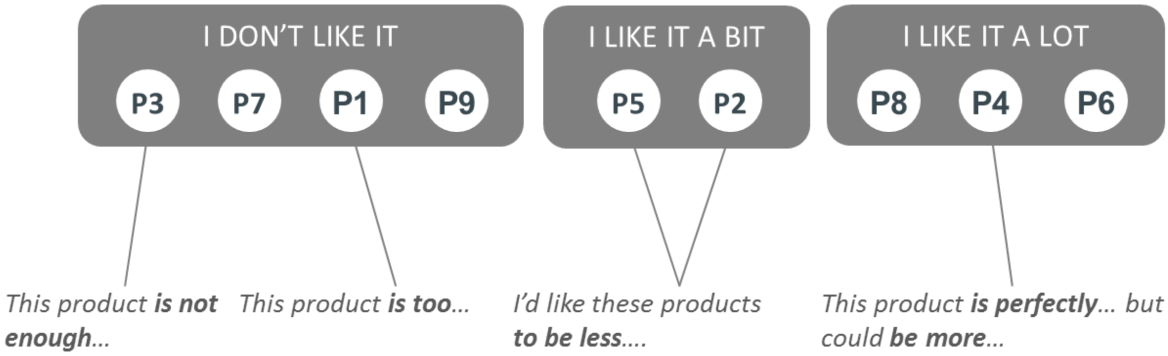

Figure 1: Summary figure of the Free JAR method protocol

In order to validate the protocol, an opinion poll was distributed to consumers who took part of the study to collect their assessment regarding the use of the Free JAR: was it easy to follow? Was it intuitive? Was it enough to describe products?

2.1.2. Delivered datasets

This new methodology conducts to a free and channelled expression, providing us specifically structured datasets. In order to validate our protocol and the resulting data, and to make sure that this new strategy brings an added-value compared to other strategies, we performed two Free JAR tests that allowed us to obtain two different datasets, structured in the same way. The first one, carried out in 2017, was used as a baseline to check the relevance of the implemented nudge, and to carry out a qualitative validation of the data obtained with our new methodology. This first data we recorded deals with perfume data: 15 perfumes were assessed by 93 consumers. The resulting dataset consists of 1395 rows and 4 columns. An extract is presented in **Fig.2**.

The second one was the subject of a data collection realised in 2018, regarding car seat leatherettes: 10 samples of car seat leatherettes were assessed by 60 consumers. We provided a resulting dataset with 600 rows and 4 columns. An extract is presented in **Fig. 3.** It allowed us to check the consistency of the Free JAR data.

Judge	Product	Hedonic category	Free JAR description
1	1	I like it a lot	I like this product because it is floral as I like
1	2	I don't like it	I don't really like this product because it is not sweet enough

Figure 2: Extract of the perfume dataset

Judge	Product	Hedonic category	Free JAR description
1	1	I like it a lot	I like this product because it is very elegant
1	2	I don't like it	I don't like this leatherette because it is too fluffy

Figure 3: Extract of the car seat leatherette dataset

2.2. The added-value of the Free JAR data

2.2.1. The relevance of the Free JAR data compared to free comments

As a reminder, each product is placed into one of the three hedonic categories. Following this ranking, the consumer is invited to describe these products with their own vocabulary, but according to a JAR scale. In order to validate the data quality of the recorded Free JAR data, we use data collected in 2017, regarding perfumes.

Firstly, in practice, we want to make sure that the new implemented method brings a real added value compared to free comments.

To achieve this, the idea is to compare these two types of data recording and to study the provided words used to describe the products:

- On the one hand, 93 people were invited to follow the Free JAR protocol in order to provide Free JAR data
- On the other hand, 142 other people, with a similar profile to people who followed the Free JAR protocol, were invited to follow the free comment protocol. Each consumer evaluated 7 of the 15 perfumes. First, they were asked to evaluate each perfume following a 10-points liking scale. Then, two questions were asked regarding what they liked and disliked about each product: "What do you like about this perfume and why?" and "What do you dislike about this perfume and why?".

A word classification is carried out in order to highlight the different type of words present in both datasets. The person in charge of the analysis realises this classification in regards of the words' meaning and it provides different semantic fields. We separate in priority useful words in a product development context, which means sensory attributes and hedonic assessments, from semantic fields unused in a product development context. For example, when dealing with perfumes, "strong" and "sweet" are usable words whereas "pleasant" doesn't highlight product improvement key, and is thus unusable in a product development context.

The aim of the Free JAR data recording is to highlight product improvement keys, that means words which highlight products' qualities and faults. These kinds of words are usually sensory descriptors linked to hedonic assessments. It is therefore necessary to cluster the different types of words used to describe products in both methods as semantic fields, and then to see to what extent Free JAR data recording makes it possible to influence the consumers' speech towards usable data in an R&D context.

On the other hand, checking the relevance of these data from a qualitative point of view means checking the consistency between a set of words and a hedonic category. Indeed, as each product is first classified in a hedonic way and then described, it is necessary that words used to describe products placed in each hedonic category reflect the reality of the category: words used to describe products placed in the “I like it a lot” hedonic category must highlight qualities whereas words used to describe products placed in the “I don’t like it” hedonic category must highlight faults. Revealing this kind of link will allow to say to what extent Free JAR data is more operational than free comments.

2.2.2. The consistency of the Free JAR data

Checking the consistency of the Free JAR data can be done by converting the textual data into a quantitative data, which would represent the consumer’s speech as accurately as possible. Then, a way to achieve this is to call on sentiment analysis, and especially on *sentiment score*.

The goal of sentiment analysis, also known as opinion mining, is to evaluate the emotional tone related to a sentence (Boullier et Lohard 2012). This tone can be represented by a number that we called “sentiment score”. When the emotional tone induce by a sentence is positive (respectively negative), the sentiment score related to this sentence will be positive (respectively negative).

To calculate the sentiment score, we use a *lexical approach* based on dictionaries and weights assigned to each word and grammatical rule, which appears to be the most natural calculation method. In fact, as it is our first attempt with Free JAR data, we want to be in a controlling position, to be close to the data and close to the meaning we give them. This strategy is in line with the semantic approach we previously used to classify words in 2.2.1.

To achieve this, we firstly need to study the structure of sentences present in the dataset. As a reminder, we use the dataset that deals with car seat leatherettes to assess it. Each sentence is built with one polarized word, which is often the interest sensory descriptor, and with a cluster, which represents the surrounding word group of the polarized word, as a context. An example is displayed below:

*“I REALLY LIKE THIS PRODUCT <u>BECAUSE IT IS VERY</u> **ELEGANT**”*

Different packages allow to calculate the sentiment score on the R software: *hunspell* (Ooms et details 2017), *spelling* (Ooms et Hester 2018) and *sentimentr* (Rinker [2015] 2018) are the main used packages. The three of them are based on weights and dictionaries, but *spelling* isn’t able to propose suggestions of correction and only one of them can store modifications and add other words to its dictionary: the *sentimentr* package. This is the one we chose to use. This package allows to quickly calculate the sentimental polarity, at a sentence level. Each polarized word (also denoted as *pw*) is weighted, firstly by its own weight, and then by the weight of the different types of words present in the surrounding group.

In practice, to calculate the sentiment score, we firstly need to have access to two dictionaries, as Tyler Rinker explains in his package instruction manual (Rinker [2015] 2018):

- a *polarity dictionary* which assigns a positive, neutral or negative nature to each word, tagged with a +1, 0 or -1 respectively.
- a *valency dictionary* which assigns to each type of words present in the surrounding group its nature: negator, amplifiers, de-amplifiers or adversative conjunctions.

We can create our own dictionaries or choose ones that are already filled: we created our own dictionaries for this study, specific to perfumes and car seat leatherettes, which suit as accurately as possible our datasets.

Then, the software checks automatically that each word is present in the polarity dictionary. A group of words, allowing to determine the context of the sentence is thus selected: it corresponds to 4 words before the polarized word and 2 words after. These are default values but the user can choose their own numbers of word to keep before and after the polarized word. Thus, the considered group of word, that we called *cluster*, is define as:

$$C = \{pw - 4, \ldots, pw, \ldots, pw+2\}$$

where *pw* represents the polarized word

The different types of words present in the cluster impact in different ways the polarized word during the calculation:

- *Neutral word*: the equation doesn't take into account this type of word but it affects the number of words present in the cluster.
- *Amplifier*: increases the impact of the polarized word (ex: I **really** like it). It multiplies the polarity of the polarized word by 1 + 0.8. This is a default value, but the user can choose their own amplifier weight. This value must be between 0 and 1, and will multiply the polarized word by 1 + the chosen value.
  It becomes a de-amplifier if the cluster contains an odd number of negators, and aims to decrease the polarity (ex: I **hardly** like it).
- *Negator*: flips the sign of the polarized word (ex: I **don't** like it). Negation is determined by raising -1 to the power of the number of negators present in the cluster. This provides the sign of the polarized word. It also acts on amplifiers / de-amplifiers as explained above.
- *Adversative conjunction*: overrules the previous clause containing the polarized word (ex: I like it **but** it's not worth it). When it's placed before the polarized word, it multiplies the number of words placed before the polarized word in the cluster by 1*0.85 and then adds 1. When it's placed after the polarized word, it multiplies the number of words placed after the polarized word in the cluster by –1 * 0.85 and then adds 1.
  This calculation method is based on the belief that an adversative conjunction increases the weight value of the next clause while lowering the weight value of the previous one (Rinker [2015] 2018).
  The weight 0.85 is a default value, but the user can choose their own weight for adversative conjunctions.

Thus, sentences containing these different types of words will lead to different sentiment scores:

- "I really like this product because it is very elegant" → 0.67
- "I don't like this leatherette because it's too fluffy" → - 0.34
- "I like this product even though it seems a little bit cheap" → 0.29

All these calculations provide a sentiment score for each sentence. Therefore, we have access to a fifth column in our dataset, which allows to check the consistency of Free JAR data. Our resulting dataset is presented in **Fig.4**.

Judge	Product	Hedonic category	Free JAR description	Sentiment score
1	1	I like it a lot	I like this product because it is very elegant	0.67
1	2	I don't like it	I don't like this leatherette because it is too fluffy	-0.34

Figure 4: Resulting dataset after the processing via sentiment analysis

In order to check the consistency of Free JAR data and given our dataset, analyses of variance are carried out. In fact, we can say that a judge is qualified as consistent if the way he perceives the product in Free JAR description is consistent with the way he ranks the product. Thus, we want to check a consistency between a group of words and a hedonic category. As we have already transformed these groups of words from Free JAR method into a numerical one via sentiment analysis, we can use the sentiment score previously calculated to realise this quantitative data checking.

Thus, we can affirm that a consumer is consistent if the sentiment score obtained for a given product matches with the hedonic category in which the product was placed in. An analysis of variance may be used to study the relationship between the sentiment score and the hedonic category. Indeed, analysis of variance is used in order to study the behaviour of a quantitative variable (the sentiment score in our case) explained as a function of one or more categorical variables (the hedonic categories in our case). This relationship study will allow to see to what extent the developed method provides oriented textual data, that are consistent with the hedonic category in which the product has been placed.

2.3. Drivers of liking

The study of the different words used in consumers' speeches in order to describe products allow us to highlight the drivers of liking and disliking related to each product. Indeed, highlighting words significantly more used to describe a product than the others related to the sentiment score associated with this product is a strategic tool regarding the outcome of the study. In fact, it allows to see which particular feature provides a better or a lower sentiment score for a product, compared to other features.

To highlight these drivers of liking or disliking, we carried out a description of frequencies. A word is considered as a driver of liking or disliking if its abundance in the class is considered significantly higher than what we can expect given its presence in the population. We note:

- N_{kj} the number of times the interest word j has been cited to describe the product k among all the words N_k cited to describe the product k
- N_j the number of times the interest word j has been cited to describe all the products among all the words N cited to describe all the products

The abundance of the interest word j is defined by comparing its percentage in the kth class with its percentage among all the products.

We define:

- H0: when the N_k words used to describe the product k are randomly selected without putting them back among the N words used to described all the products, the percentage of the interest word used to describe the product k and the percentage of the interest word used to describe all the products coincide at:

$$\frac{Nkj}{Nk} \approx \frac{Nj}{N}$$

Then, a test corresponding to the hypergeometric distribution is performed and the probability to observe a more extreme value than the one observed is calculated (Lebart, Piron, et Morineau 1995; Francois Husson et al. 2018).

2.4. Data visualisation: the "Sentiment Mapping"

Once the consistency between the sentiment score and the hedonic assessment of the product is checked, it seems reliable to take advantage of this score to obtain an intuitive and ergonomic data visualisation of the results.

To achieve this, we carried out a Multiple Factor Analysis (MFA) on the dataset presented in **Fig. 5**, which comes from the 2018 dataset, dealing with car seat leatherettes. This dataset represents two combined table. The first one represents the hedonic category in which each product was placed in by each judge. The second one is a contingency table between the products and the hedonic categories. This analysis aims to highlight and study the similarities between products from the point of view of all variables (François Husson, Lê, et Pagès 2017). We interpret MFA's results as follows: two products are even closer as they have been placed a large number of times in the same hedonic category at the panel level (Brard et Lê 2018).

Products	**Judge 1**	**Judge 2**	**Judge 3**	**…**	**I like it a lot**	**I like it a bit**	**I don't like it**
P1	I like it a lot	I like it a bit	I like it a lot		37	15	8
P2	I don't like it	I like it a bit	I like it a bit		19	14	27
…							

Figure 5: Dataset used in order to realise the MFA

As in an external preference mapping (François Husson, s. d.), the sentiment score replaces consumers' preferences in order to project on the product space a response surface indicating a proportion of consumers having a good sentiment of the product. This good sentiment represents the mean of the sentiment scores calculated for each product. Then, remarkable drivers of liking or disliking previously highlighted are displayed, allowing a direct interpretation of the map.

This produced map is called "*Sentiment Mapping*".

3. RESULTS

In order to validate the consistency and the reliability of the new developed method, which allows us to highlight structured textual data, oriented towards product improvement keys, we must first validate our protocol from a user's point of view and then check the relevance of the provided data.
On the one hand, it is necessary that the user does not have any problems with the protocol, which has to be easy and intuitive to follow. On the other hand, it is necessary that this developed data recording provides high quality data and makes it possible to meet the objectives previously set.

3.1. Protocol validation from a user's point of view

Thanks to our first data recording trial, dealing with perfume data, we attempted to validate the protocol of our new method. To achieve this, an opinion poll was distributed to consumers who took part of the study. It turns out that 81% of the consumers found the Free JAR useful, 80% managed to use the Free JAR and 77% found the Free JAR enough to describe the products.
Thus, the protocol of our new data recording seems intuitive and easy to follow for users. It seems reasonable to say that it would be adapted as a potential routine development.

3.2. Highlighting the added-value of the Free JAR data

3.2.1. Checking the relevance of the Free JAR data compared to free comment

Firstly, in order to make sure that the new implemented method brings a real added value compared to free comments, we realised a graph as presented in **Fig.6.** The overview of the words present in the dataset allows us to realise a sorting of them according to 7 semantic fields: sensory descriptors, intensity, concept, memory, appreciation, sense and context. We sorted all the words used in both method and it provides the presented graph. It represents the number of times the different semantic fields are used to describe products in both methods.

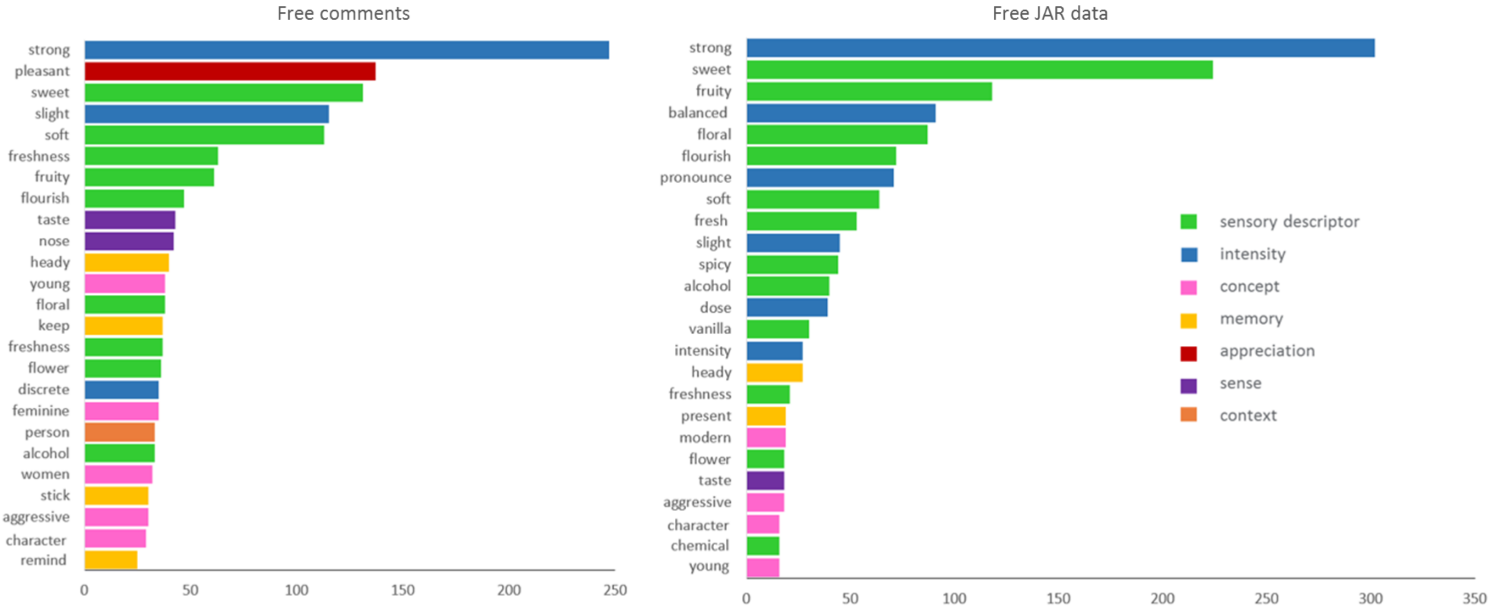

Figure 6: Number of times words relative to the different semantic fields are used to describe products regarding free comments and Free JAR data recording

First of all, we notice that free comments use more diversified type of words than the Free JAR data recording. In fact, words linked to appreciation, context, memory and sense are much more used to describe products in free comments. On the contrary, we note that sensory descriptors or type of words linked to the intensity are much more used to describe products during the Free JAR data recording: the Free JAR strategy uses less words, but they are more usable in a product development context.

Thus, in the light of the kind of textual data we want to obtain, the Free JAR method seems to be able to answer our objectives. In fact, although free comments provide diversified type of words, these textual data are less usable in a product development context, where the goal is to highlight sensory descriptors that allow to define products' qualities and faults. On the contrary, Free JAR data recording, that uses numerous sensory and intensity descriptors, provides more usable data which highlight product improvement keys: consumers are implicitly led to structure their speech towards sensory descriptors. Free JAR data are, in this way, more relevant in a qualitative point of view.

On the other hand, in order to check the relevance of the Free JAR data from a qualitative point of view, we realised the graph presented in **Fig.7**. To answer the objectives previously stated, this graph represents the number of times the terms "Just about right", "Not enough", "Too much" and descriptive words have been used to describe products placed in the 3 different hedonic categories: "I don't like it", "I like it a bit" and "I like it a lot".

We notice in the resulting figure that "Not enough" and "Too much" terms are more used to describe products placed in the "I don't like it" hedonic category than to describe products placed in the "I like it a lot" hedonic category. In the same way, we notice that "Just about right" term is more used to describe products placed in the "I like it a lot" hedonic category than to describe products placed in the "I don't like it" or "I like it a bit" hedonic category. Thus, each hedonic category is characterized by its own words, highlighting faults for the "I don't like it" hedonic category and qualities for the "I like it a lot" hedonic category. We can therefore affirm that the dataset quality is checked from a qualitative point of view.

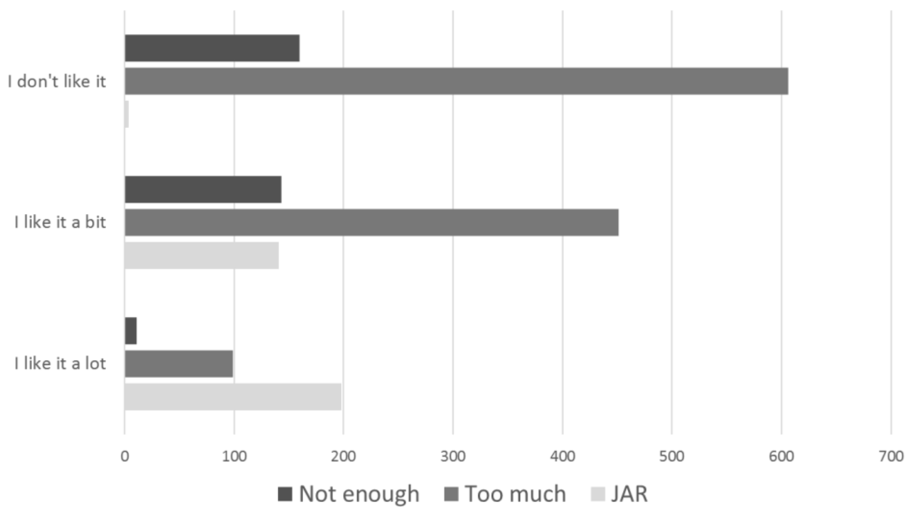

Figure 7: Number of times terms are used to describe products placed in the hedonic categories

In order to confirm this structured distribution, we realised Chi2 tests on contingency tables, which compared the expected distribution versus the real distribution. It confirms our thoughts:

words are significantly structured through the hedonic categories with a p-value lower than 2e-16.

3.2.2. *Checking the consistency of the Free JAR textual data*

In order to check the Free JAR data quality in a quantitative way using the new numerical data previously calculated (the sentiment score), as many analyses of variance are performed as there are judges in the dataset. Given a judge and thanks to the analysis of variance, we have access to the mean of the sentiment score for each hedonic category.
Provided results highlight different types of judges:

- *Perfectly consistent judges in a sentiment score point of view*: the mean of the sentiment score for the "I like it a lot" hedonic category is higher than the one for the "I like it a bit" hedonic category, which is higher than the one for the "I don't like it" hedonic category.
- *Moderately consistent judges in a sentiment score point of view*: the mean of the sentiment score for the "I like it a bit" hedonic category is higher than the one for the "I like it a lot" hedonic category, or lower than the one for the "I don't like it" hedonic category.
- *Inconsistent judges in a sentiment score point of view*: the mean of the sentiment score for the "I like it a lot" hedonic category is lower than the one for the "I don't like it" hedonic category.

As an example, **Table 1** presents the result of the analysis of variance for **a perfectly consistent judge** in a sentiment score point of view.

Ftest	***F value***	***P-value***
Hedonic_category	5.1061	0.0429
Hedonic category	***Coefficient***	***P-value***
I like it a lot	0.80842	0.02675
I like it a bit	- 0.16120	0.55374
I don't like it	- 0.64722	0.02707

Table 1: Result of the analysis of variance for a perfectly consistent judge in a sentiment score point of view

Analyses of variance are realised for each judge present in our dataset that deals with car seat leatherettes. They highlight that 56% of the judges are perfectly consistent while 44% of the judges are moderately consistent. In this dataset, we didn't find any inconsistent judges and we therefore didn't remove any of them from the dataset. This allows us to proceed further with the analysis.

3.3. *Provided graphic representation*

Once the dataset is filled, the sentiment score is used in order to perform the "*Sentiment Mapping*". This representation is presented in **Fig.8**.
Products are represented by letters. Those present in the red area are products that drive a bad opinion, whereas those present in the green area are products that drive a good opinion. Thus, we can notice that products J, C and F generate a better opinion than products G or D. Drivers of liking and disliking projected allow us to say that product G doesn't seem rigid enough while product D seems too fluffy to consumers.

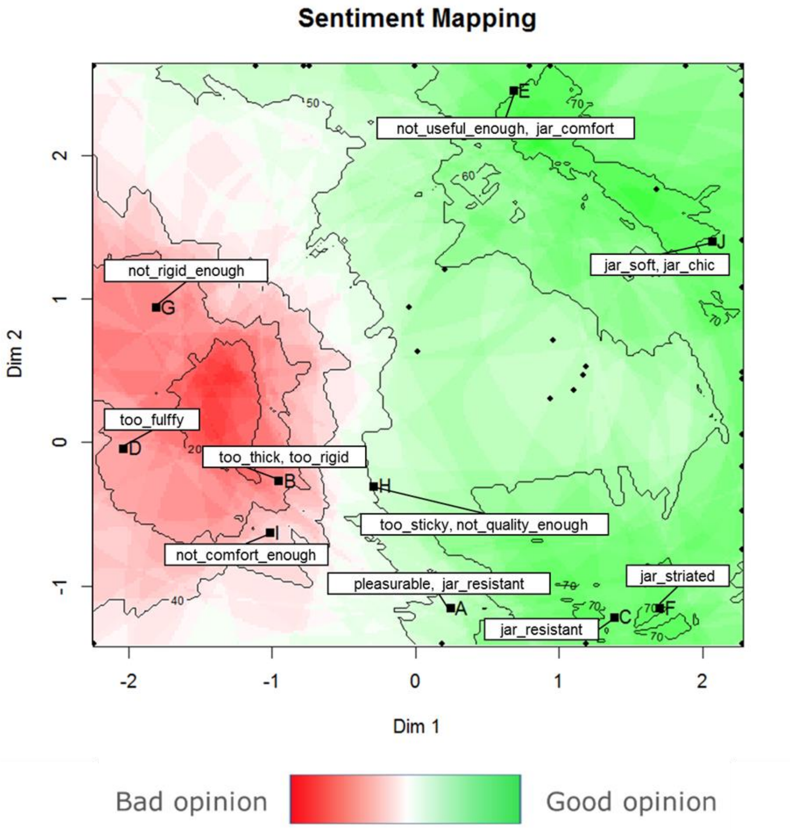

Figure 8: Sentiment Mapping provided thanks to the car seat leatherette dataset

4. CONCLUSION AND DISCUSSION

As a conclusion, we provide today a new way to collect data for market research institutes based on the *nudge* theory, which lead consumers to the desired speech, highlighting product improvement keys, essential in a product development context. This information can be summed up into an ergonomic and intuitive map: the "*Sentiment Mapping*". This new graphic representation puts the consumer at the heart of the process, as expected. In fact, data are only based on consumer speech and the provided map allows an easy access to the faults and qualities that each of the tested products may have. Moreover, it positions products against each other regarding their sentiment score. The developed method lets the consumers freely express themselves, while structuring their thoughts in order to influence their speech toward products' faults and qualities.

However, we can highlight some limits and discussion points regarding the new developed method.

First, regarding the first step of the Free JAR method, a discussion must be done with the client in order to define the labels of the different hedonic categories ("I like it a lot", "I like it a bit", "I don't like it"). Indeed, these labels can influence the representation consumers have regarding the different categories. Moreover, we agree that there may have a potential loss in sensitivity by using a simple three-point hedonic scale and a discussion can be held regarding the different strategies we can set up in order to put consumers in a hedonic state of mind.

Then, the classification of the words used for the qualitative validation of Free JAR data is quite subjective. Indeed, the person in charge of the analysis has to class words according to their perception: it seems unlikely that two people performing this task independently will result the same words' classification. It would be necessary to drive the study further in order to highlight salient and optimal categories, differentiating usable words in a product development context from unusable words in a product development context, and bringing together all types of words present in the dataset. Machine learning strategies could be considered (Pradhan et al. 2005).

Regarding the calculation of the sentiment score, we use today a lexical approach, as it is previously explained. But since the relevance and the interest of the nudge has been demonstrated, we can think about machine learning strategy (Hatzivassiloglou et McKeown, s. d.). This strategy aims to train the software to detect expressions, models, patterns via a first test corpus in order to be able to detect these models in the corpus of interest, or even to detect new ones, close to those it already knows. The main interest of this strategy is that it will allow us to move away from dictionaries, which are time-consuming to set up, specific for each product space and which are not specially filled in an objective way, as they are filled according to our perception.

Then, we wanted to discuss more the concept of “consistency” for one consumer. As we defined the concept of consistency with Thierry Worch in our paper (Worch et al. 2012), we defined the concept of consistency in the Free JAR context by answering two major questions: (1) Are the descriptions in agreement with the hedonic classifications? (2) Have the products placed in the “I like it a lot” category a better sentiment score than the others? More generally, consistency is defined as the fact that a consumer follows the same direction from one task to another. However, we are conscious that people can find drawbacks for a product, even if they like it a lot: that’s why we keep the moderately consistent judges for the analysis.

Thus, the fact that 44% of the consumers are “moderately consistent” for the quantitative validation of Free JAR data encourage us to go further regarding the understanding of these judges. Indeed, it is encouraging but it also means that there is still work to be done. In fact, we don’t know well how to construe them and how optimize the analysis in order to make the calculation of the sentiment score more accurate: here again, machine learning strategies could be considered.

We compared in this manuscript Free JAR data with free comments. However, we didn’t compare Free JAR data with classical JAR data. As we wanted to move away from constraints imposed by the JAR method, it seemed more natural for us to compare Free JAR data with free comments: we aimed to bring added-value to a low-quality data, providing an “oriented” free comment. But the comparison between Free JAR data and JAR data will be the subject of further works.

A discussion should also be hold regarding the size of the product space. Indeed, the Free JAR method aims to analyse large product spaces, thanks to appropriate multivariate analyses. However, when the product space is too small, the method must be reconsidered. In fact, we can still put consumers in a hedonic state of mind and make them express via the Free JAR method, but we are currently considering to what extent comparison between products can be identified as the MFA can’t work with such a small number of products.

Finally, this new method highlights a new kind of data, a quantitative score built from textual data: the sentiment score. It helps to get as close as possible to the consumer by quantifying the feelings a consumer has about a product. Beyond the Free JAR method, the sentiment score allows us to think about other applications, in other contexts and fields.

Acknowledgements

This study is part of a Master's degree project, realised thanks to Agrocampus-Ouest. We want to thank Strategir, that gave us the opportunity to realise this project, that carried out the data recording and allows to go further in future years. A very special thanks to Laure Bonnet, Marie Boutigny, Sonia Guillou and Aubery Lebas for their previous work that helped us a lot. In the end, we particularly wanted to thank our best partners, Mathilde Doyen and Vanessa Morin. Acknowledgment is also expressed to the two reviewers for their insightful comments, as these comments helped us to significantly improve the manuscript.

References

Allcott, H., et S. Mullainathan. 2010. « Behavior and Energy Policy ». *Science* 327 (5970): 1204‑5. https://doi.org/10.1126/science.1180775.

Allcott, Hunt. 2011. « Social Norms and Energy Conservation ». *Journal of Public Economics* 95 (9‑10): 1082‑95. https://doi.org/10.1016/j.jpubeco.2011.03.003.

Allcott, Hunt, et Todd Rogers. 2014. « The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation ». *American Economic Review* 104 (10): 3003‑37. https://doi.org/10.1257/aer.104.10.3003.

Ares, Gastón, Juliana Cunha de Andrade, Lucía Antúnez, Florencia Alcaire, Marianne Swaney-Stueve, Sheri Gordon, et Sara R. Jaeger. 2017. « Hedonic Product Optimisation: CATA Questions as Alternatives to JAR Scales ». *Food Quality and Preference* 55 (janvier): 67‑78. https://doi.org/10.1016/j.foodqual.2016.08.011.

Boullier, Dominique, et Audrey Lohard. 2012. « Chapitre 5. Détecter les tonalités : opinion mining et sentiment analysis ». In *Opinion mining et Sentiment analysis : Méthodes et outils*. Sciences Po | médialab. Marseille: OpenEdition Press. http://books.openedition.org/oep/214.

Brard, Margot, et Sébastien Lê. 2018. « Adaptation of the Q-Methodology for the Characterization of a Complex Concept through a Set of Products: From the Collection of the Data to Their Analysis ». *Food Quality and Preference* 67 (juillet): 77‑86. https://doi.org/10.1016/j.foodqual.2017.06.010.

Caris, M.G., H.A. Labuschagne, M. Dekker, M.H.H. Kramer, M.A. van Agtmael, et C.M.J.E. Vandenbroucke-Grauls. 2018. « Nudging to Improve Hand Hygiene ». *Journal of Hospital Infection* 98 (4): 352‑58. https://doi.org/10.1016/j.jhin.2017.09.023.

Costa, Dora L., et Matthew E. Kahn. 2013. « ENERGY CONSERVATION "NUDGES" AND ENVIRONMENTALIST IDEOLOGY: EVIDENCE FROM A RANDOMIZED RESIDENTIAL ELECTRICITY FIELD EXPERIMENT: Energy Conservation "Nudges" and Environmentalist Ideology ». *Journal of the European Economic Association* 11 (3): 680‑702. https://doi.org/10.1111/jeea.12011.

Epler, Silke, Edgar Chambers Iv, et Kenneth E. Kemp. 1998. « HEDONIC SCALES ARE A BETTER PREDICTOR THAN JUST-ABOUT-RIGHT SCALES OF OPTIMAL SWEETNESS IN LEMONADE ». *Journal of Sensory Studies* 13 (2): 191‑97. https://doi.org/10.1111/j.1745-459X.1998.tb00082.x.

Hatzivassiloglou, Vasileios, et Kathleen R McKeown. s. d. « Predicting the Semantic Orientation of Adjectives », 8.

Husson, François. s. d. « La cartographie des préférences », 25.

Husson, Francois, Julie Josse, Sebastien Le, et Jeremy Mazet. 2018. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining* (version 1.41). https://CRAN.R-project.org/package=FactoMineR.

Husson, François, Sébastien Lê, et Jérôme Pagès. 2017. *Exploratory multivariate analysis by example using R*. Second edition. Boca Raton: CRC Press.

Lawless, Harry T., et Hildegarde Heymann. 2010. *Sensory Evaluation of Food*. Food Science Text Series. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-6488-5.

Lawrence, G., R. Symoneaux, I. Maitre, F. Brossaud, M. Maestrojuan, et E. Mehinagic. 2013. « Using the Free Comments Method for Sensory Characterisation of Cabernet Franc Wines: Comparison with Classical Profiling in a Professional Context ». *Food Quality and Preference* 30 (2): 145‑55. https://doi.org/10.1016/j.foodqual.2013.04.005.

Lebart, Ludovic, Marie Piron, et Alain Morineau. 1995. *Statistique exploratoire multidimensionnelle*. Paris: Dunod.

Momsen, Katharina, et Thomas Stoerk. 2014. « From Intention to Action: Can Nudges Help Consumers to Choose Renewable Energy? » *Energy Policy* 74 (novembre): 376‑82. https://doi.org/10.1016/j.enpol.2014.07.008.

Ooms, Jeroen, et Authors of libhunspell (see AUTHORS file) hunspell author details. 2017. *hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker* (version 2.9). https://CRAN.R-project.org/package=hunspell.

Ooms, Jeroen, et Jim Hester. 2018. *spelling: Tools for Spell Checking in R* (version 1.2). https://CRAN.R-project.org/package=spelling.

Popper, R., W. Rosenstock, M. Schraidt, et B.J. Kroll. 2004. « The Effect of Attribute Questions on Overall Liking Ratings ». *Food Quality and Preference* 15 (7‑8): 853‑58. https://doi.org/10.1016/j.foodqual.2003.12.004.

Popper, Richard. 2014. « Use of Just-About-Right Scales in Consumer Research ». In *Novel Techniques in Sensory Characterization and Consumer Profiling*, édité par Paula Varela et Gastón Ares, 137‑56. CRC Press. https://doi.org/10.1201/b16853-6.

Pradhan, Sameer, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, et Daniel Jurafsky. 2005. « Support Vector Learning for Semantic Argument Classification ». *Machine Learning* 60 (1‑3): 11‑39. https://doi.org/10.1007/s10994-005-0912-2.

Rinker, Tyler. (2015) 2018. *Dictionary based sentiment analysis that considers valence shifters: trinker/sentimentr*. R. https://github.com/trinker/sentimentr.

Rothman, Lori, et Merry Jo Parker, éd. 2009. *Just-About-Right (JAR) Scales: Design, Usage, Benefits, and Risks*. 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959: ASTM International. https://doi.org/10.1520/MNL63-EB.

Sunstein, Cass R. 2015. « Nudging Smokers ». *New England Journal of Medicine* 372 (22): 2150‑51. https://doi.org/10.1056/NEJMe1503200.

Symoneaux, R., M.V. Galmarini, et E. Mehinagic. 2012. « 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 (1): 59‑66. https://doi.org/10.1016/j.foodqual.2011.08.013.

Thaler, Richard H., et Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Worch, Thierry, Sébastien Lê, Pieter Punter, et Jérôme Pagès. 2012. « Assessment of the Consistency of Ideal Profiles According to Non-Ideal Data for IPM ». *Food Quality and Preference* 24 (1): 99‑110. https://doi.org/10.1016/j.foodqual.2011.10.001.

- In a product development context, the nudge theory can be used in order to influence consumers' speeches toward product improvement keys
- JAR and free comments data recording have limits in a product development context
- The Free JAR strategy lets consumers freely express themselves while structuring their speech thanks to a nudge step
- Sentiment analysis can be use in order to check the consistency of the Free JAR data
- The "*Sentiment Mapping*" positions products against each other regarding their sentiment score in an intuitive and ergonomic way