

FORMALIZING DESIGN RULES APPEARING IN EARLY DESIGN : AN APPLICATION OF AGGLOMERATIVE CLUSTERING

Jean-François Omhover, Carole Bouchard, Fabrice Mantelet, Améziane Aoussat

LCPI (Product Design and Innovation Laboratory)
Ecole Nationale Supérieure d'Arts et Métiers

ABSTRACT

The overall goal of this experiment was the identification of expert design rules, taking into account the natural cognitive process of the designers. More specifically, the protocol aimed to define how designers link semantic adjectives to formal attributes and to other semantic adjectives.

Previous studies of the cognitive activity of the designers allowed us to describe particular features of the design activity. The categorization of design information is a major part of this activity, involving a both highly structured and creative manner of organizing design information.

Our hypothesis was that the use of the agglomerative clustering could be a relevant way of formalization of the expertise of the designers because it mimics closely the cognitive processes entailed in the framework of the design activity. Moreover this approach does not require a huge sample of subjects and data in order to extract relevant algorithms. Therefore we experimented agglomerative clustering in order to develop design rules between semantic concepts, semantic adjectives and the low level dimensions of design: colours, textures and shape descriptors. These design rules characterize the expertise of the designers in the early stages of the shoe design process, where the product has yet to be defined.

Keywords: Design rules, agglomerative clustering, generation of solutions

1 INTRODUCTION

During the cognitive process occurring in the early stages of the design, and that leads to the production of solutions, designers are used to manipulate semantic adjective and design elements. *Semantic adjectives* are words that describe the semantic or perceptive environment of the product. *Design elements* are names or labels that physically qualify the product: these are names of matters, colours, textures, shapes, etc. In the context of shoe manufacturing – we focused on this particular field for our study – examples of semantic adjectives can be “sweet”, “warm”, “comfortable”, “playful”, “classical”, etc, and examples of design elements can be “dark”, “black”, “brown”, “silk”, “round”, “metal”, etc. The very expert knowledge of the designers relies on their capability to associate these kinds of terms in order to produce solutions from a description of their concepts (semantic adjectives, values, etc), these are the *design rules*.

Previous studies of the cognitive activity of the designers allowed us to describe particular features of the design activity [1]. The categorization of design information is a major part of this activity, involving a both highly structured and creative manner of organizing design information [2].

Our investigation focuses on the capitalisation and the formalisation of these design rules, in the context of the European project Kensys. The goal of this project was the elaboration of a software system capable of assisting the designers in understanding the user's needs and generating solutions [3][4].

In this context we focused on the extraction of the associations between semantic adjectives and design elements in order to build a recommendation tool for the designers. Many studies are based on the formalisation of design rules according to semantic testing results led with large panels of end users or sometimes of designers. Then the results are managed thanks to statistical tools like Principal Component Analysis and Regression Analysis [5][6] [7].

In this paper we present a new way of building design rules. The related approach lies on the extraction of expert design rules based on a small sample of experts, which is supposed to be sufficient for expressing expert rules, without deploying a huge sample for testing. The underlying idea was to produce formalized rules based on the automatic extraction of groups in the terms expressed by the designers themselves. This formalization has been operated by a specific kind of machine learning algorithms: a clustering algorithm, and more specifically a hierarchical clustering method. Our hypothesis was that the use of the agglomerative clustering could be a relevant way of formalization of the expertise of the designers because it mimics closely the cognitive processes entailed in the framework of the design activity.

In artificial intelligence, the terms “*clustering*” [9, 12] refers to a set of techniques that are used to classify objects. In a population of objects (vectors, points, feature sets, etc.), a clustering algorithm will tend to isolate groups that comply with a simple rule: the objects within one given group should be considered as similar, the objects that belong to two different groups should be considered as not similar. The family of the clustering algorithms has been declined in many different variants depending on three main dimensions: the nature of the objects to be classified, the meaning of the term “*similar*” in the previous rule and the expected nature of the groups. These various algorithms come with their own parameters and requirements. They rely on several formal definitions such as similarity measures (e.g. Euclidian distance) and description spaces (e.g. vector space). In the simplest case, the objects to be classified are points in a vector space and the similarity is quantified as the inverse of the distance between two points.

In our study, the objects to be classified were terms (semantic adjectives and design elements) used in the early stage shoe design process, where the product has yet to be defined. The nature of the groups expected was undetermined but their structure was considered as associations that could be gradually quantified (two terms could be more linked than others). For these reasons and more, we focused on a specific clustering algorithm called hierarchical clustering algorithm. This particular algorithm relies on a simple process of one to one associations that lead to the production of a tree of objects (i.e. terms) where leaves are the semantic adjectives or design elements, and branches are their associations, more or less important depending on the degree of similarity of the terms.

In this paper, we will present the methodology of extraction of design rules in shoe manufacturing by a hierarchical clustering algorithm and the results obtained. We will present the various elements of our work by following the main phases of the study:

- in section 2, we expose the method used to lead the designers to express links between concepts (see section 2.1.), semantic adjectives, design elements, etc. This was based on a semi-open interview scheme. This protocol enabled us to acquire many terms corresponding to each other, this correspondence has been regarded as a design rule between adjectives and design elements (see section 2.2.).
- in section 3, we present the hierarchical agglomerative clustering (H.A.C.) algorithm and discuss about its applicability in the context of design rule extraction. We also present the post-processing procedure used to use the open interviews made in section 2 as inputs for the H.A.C. algorithm.
- in section 4, we present the results obtained by the application of the algorithm, and lead an in-depth discussion about the nature of these results and how these results could be used to generate recommendations for shoe design, and how we could extend this approach to other design domains.

We finally conclude by establishing the perspectives of this work and the future developments of the method.

2 ACQUIRING EXPERT DATA FROM INTERVIEWS

In order to find out design rules, we had to acquire expert data from the designers. This acquisition has taken the form of semi-structured interviews. This kind of protocol let us acquire different kind of terms used in the shoe manufacturing context: semantic adjectives, matters, shape, colour and texture terms. These terms will be used in the following section (section 3) in order to extract clusters or associations which will lead to design rules. In this section, we expose the protocol followed to interview the shoe designers and the nature of the terms acquired.

2.1 The protocol used to acquire expert data from designers

The purpose of the acquisition of expert data was the definition of a “dictionary” of adjectives where the formal features (shape, colour and texture) corresponding to each entry could be established.

The first step of our protocol was the establishment of a list of concepts adapted to the shoe manufacturing sector. First, a large list of words has been derived from a survey of the shoe manufacturing sector. This large list of words (around 300) has then been reduced to a list of 70 concepts. This 70 reference words were then proposed and validated by expert designers in order to specify shoe’s concepts.

The expert shoe designers were submitted to a questionnaire and to semi-structured interviews between the designers and the experiment leader who was in charge of the good progress of the experiment (see Figure 1). The questionnaire was structured into 4 parts respectively related to (1) the identity of the designers, (2) open question on synonyms for the proposed adjectives, (3) open question on which forms (shape, colour, and texture) could be associated to the adjectives and finally (4) the design rules which seem appropriate for the reference semantic adjectives.

Twelve expert designers were involved in the experiment, divided in 4 groups of 3 designers. Each group processed a specific session dealing with about 25 input semantic adjectives. The 3 designers worked independently in order not to be influenced by their mutual results. In the first part of the experiment, the designers in each group had to express words from an initial list of about 25 semantic adjectives (concepts), in an open way (open number of words). In the second part of the experiment they had to express low level attributes like colour, shape and texture descriptors that correspond to the semantic adjectives. These corresponding attributes were asked in order to qualify the initial list of semantic adjectives (concepts), and they were also expresses freely (see figure below). They were asked to describe some concepts with their own terms. For instance, the reference semantic adjective “aggressive” was described in terms of synonyms, design rules (design terms intuitively correlated to the concept), shaped, textures (matters) and colours. In the following study, we did not use the shape because they are drawings that we cannot use as easily as terms (it could be a very interesting extension of the method).

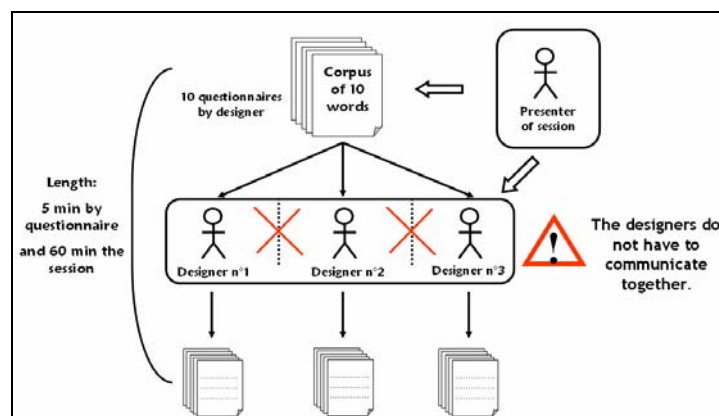


Figure 1 Interview Protocol (60 mins)

2.2 The data acquired and the underlying design rules

After gathering the results data, they were analysed and resumed under the form of a data collection. The resulting database was made of bags of terms. For each concept of the 70 reference words (the column on the left of Figure 2), multiple synonym terms, shape related terms (and drawings), colour terms and texture terms were collected. This constituted the basis of our extraction.

What we call here design rules can be found in this data collection. Indeed, when considering a given bag, the terms belonging to it are related to each other by their semantic relation to the given concept. Also, two bags carry different terms, but sometimes identical terms appear in different bags/concepts. This “correlation” observed between concepts by the co-occurrence of their description terms was the information that we wanted to extract. Indeed, the fact that identical terms do appear under different concepts means they are correlated someway: they may be members of a design rule or semantic rule.

The extraction of these rules was based on a clustering algorithm used to extract groups of identical or correlated terms within the terms collected in the interviews.

Words	Synonyms	Coef	Design rules	Coef	Shapes	Textures	Coef	Colors	Coef
Pleasant	softness	1	round	3		cotton	2	red	1
Pleasant	warm	1	convex	1		leather	1	blue	1
Pleasant	comfortable	1				latex	1	yellow	1
Pleasant	happiness	1				silk	1	pink	1
Pleasant	friendly	1				feather bed	1	brown	1
Pleasant	nice	1						ochre	1
Pleasant	soft	1						beige	1
Pleasant	calm	1						pastel	1
Aggressive	violent	2	overdeveloped	1		leather	1	vivid	3
Aggressive	attractive	1	imposing	1		vinyl	1	black	2
Aggressive	imposing	1	movement	2		mesh	1	red	2
Aggressive	exciting	1	asymetric	2		metal pick	1		
Aggressive	sportive	2	sharp angle	1					
Aggressive	quick	1	complex	1					
Aggressive	nervous	1							
Aggressive	speed	1							
Aggressive	stressed	1							
Aggressive	irascible	1							

Figure 2 Results data: bags of terms (translated) and drawings

3. APPLYING AGGLOMERATIVE CLUSTERING TO EXPERT DATA

In this section we will explain why and how we considered the use of hierarchical agglomerative clustering (H.A.C.) [11] for design rules extraction. First, we will enter an explanation of the principle of H.A.C. and the parameters of this tool (see section 3.1.), then we will explain why and how this method can be useful for design rules extraction (see section 3.2.). Finally, we will expose the transformation made on the results obtained by interviewing the experts in section 2, this transformation was necessary in order to obtain the results shown in the following section (section 4).

3.1. The principle of hierarchical agglomerative clustering

The hierarchical agglomerative clustering [11,12] is one kind of clustering algorithm that relies on the iterative agglomeration of the elements. The principle of the algorithm is shown on Figure 3. In order to cluster the elements A,B,C,D,E in a two dimensional space, we have to go through the following phases :

- **Distance computation:** we need to compute the distance matrix between these elements, i.e. to compute every distance between two elements of the set.
- **Choice of a pair of elements:** we then choose two elements in A,B,C,D,E that are the closest pair of elements of the set, on our figure A and B are the closest and their mutual distance is d_1 .
- **Merging into a group:** These two elements are merged (we draw this merging on Figure 3 by linking the two elements) and now form a new set of elements : {A,B} which are now considered as an element together (a group), and C,D,E which are still single elements.
- **Distance computation (2):** Again, we compute the distance between every pair of elements, but now we have to take into account the fact that A and B form are considered together. Then we have to take that into account for the computation of the distance between on one hand {A,B} and on the other hand C, D or E.
- **Looping:** again, the closest elements are merged, and we loop from the distance computation phase to the merging phase until there is only one single group of elements {A,B,C,D,E} all together.

Given the elements A,B,C,D,E drawn on the right part of Figure 3, the procedure above will lead to the merging of A and B at distance d_1 , then D and E at distance d_2 , then C is merged with the group composed of A and B and the corresponding distance is d_3 , then the two remaining groups are merged all together at a distance of d_4 .

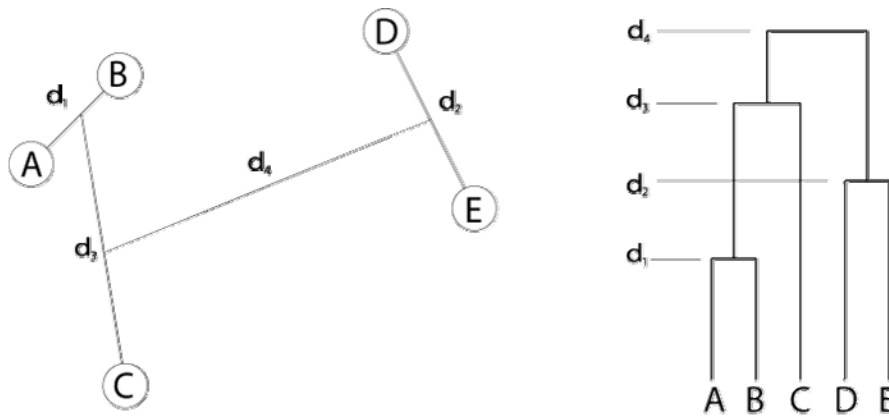


Figure 3. Principle of hierarchical agglomerative clustering in 2D; on the left we show the agglomeration of points, on the right the resulting “dendrogram”

By following the basic principle of the algorithm, we can see how subgroups and groups can be formed using the proximity (or distance) measure between pairs of elements, and how these elements will form subgroups that are larger and larger until there is only one single group. The hierarchical character of the algorithm lies within the remembrance of the distance measures at which the pairs of elements were merged. On the right part of Figure 3, we have plot the *dendrogram* resulting from the merging of the elements drawn on the left. The dendrogram simply represents the elements on the horizontal axis, and the distance on the vertical axis : a tree is drawn which leaves are the elements A,B,C,D,E and which branches are the merging operations, the length of the branches is the distance between the merged elements. We interpret this dendrogram as an observation of the associations made between the elements A,B,C,D,E : the distance quantifies the strength of this association.

In order to set up a H.A.C. algorithm, we need to define three *parameters* :

- **a description space** : these are the dimensions in which we can describe the elements of our population in order to compare them. For instance, on Figure 3, the elements to be compared are considered as points in a 2D space.
- **a distance measure** : how can we quantify the comparison between two elements ? On Figure 3, this comparison is quantified by means of a simple 2D distance measure. (The definition of the description space and the distance measure are closely related to each other).
- **a merging function (linkage)**: the way two elements or two groups are merged, and more specifically the way we will consider the distance between a group of elements and another (e.g. the minimum distance between two pairs of elements belonging to the two groups, or the distance between the average points of the groups, etc.).

Once these three parameters have been defined, the H.A.C. algorithm will output a dendrogram and the results will be exploitable.

3.2. Why using H.A.C. for extracting design rules?

In the machine learning literature [10], many algorithms can be used to extract associations or to establish links between different dimensions or sets of data – it is the goal of our experimentation. We will now explain why we chose to use H.A.C.

The first argument was the intuitive character of H.A.C. that mimics closely the cognitive processes entailed in the framework of the design activity. During the categorization phase which is highly involved in early design activity, the iterative agglomeration of terms converges through the production of groups. This process is there explored automatically.

Different kind of algorithms would have been helpful. Associations Rules [8], for instance, can be useful to draw logical rules from the observation of recurrent pairs in *transactions*. This algorithm could have been used in the context of our experiment, by considering that the interviews of the expert designers are transactions in which we can find facts or symbols (the terms used), and by observing the co-occurrence of the terms in the many transactions collected. But the fact is that the association rules methods rely on the massive co-occurrence of facts, and their exact conjoint observation in many transactions. As the data acquired in section 2 were very sparse, and as the interview led to the collection of many terms that were present roughly in three or four interviews at a time, the exact observation of the co-occurrence of terms would not have been productive enough.

This is the most important reason why we oriented through H.A.C. algorithm: the information gathered in the previous phase of our project (semi-structured interviews) led to the production of many terms in a limited number of documents. In fact, the relations between the terms were at first very weak: it was delicate to find words that did appear in multiple interviews. By automatically building a dictionary of the terms used (see section 3.3.) we obtained a representation of the connectivity of the concepts themselves: we could observe how these concepts were related to each other through their descriptions in design elements. In fact, these relations were observable but not massively recurrent. This is why we chose to orient our study through the use of H.A.C. algorithm because this kind of algorithm doesn't need many data to run and can converge on a small set of objects to classify. On the contrary, statistical techniques would have needed many examples to form a representative result. Here we can use our open set of words that is sparse but very informative in order to build pertinent design rules. We will see in section 4 how the rules extracted tend to validate this hypothesis.

Once parameterized, the H.A.C. algorithm has led us to the association of terms that were co-occurrent, then to the association of the terms that were closely co-occurrent to the groups of these terms, and so on in order to reach all the terms and force the production of associations. The agglomerative character of the H.A.C. then proves to be very useful for the extension of term co-occurrence to their most co-occurrent "neighbour words", and finally to build a structure in the whole set of words used.

3.3. How to apply H.A.C. to the data acquired in semi-structured interviews?

To compute the clustering of terms as described in the principle of the H.A.C. algorithm, it was first necessary to transform the database obtained in section 2. In fact, the data collection was practically impossible to use under its primary form. We had to transform the bags of terms to vectors in order to apply the implementation of the H.A.C. algorithm (using the bags of terms as such would have needed an unnecessarily complex implementation of the algorithm).

This transformation was made by constituting a dictionary of mixed terms (synonyms words, color words, shape words, and texture words). This dictionary covered all the terms used in the whole data collection for describing every concept.

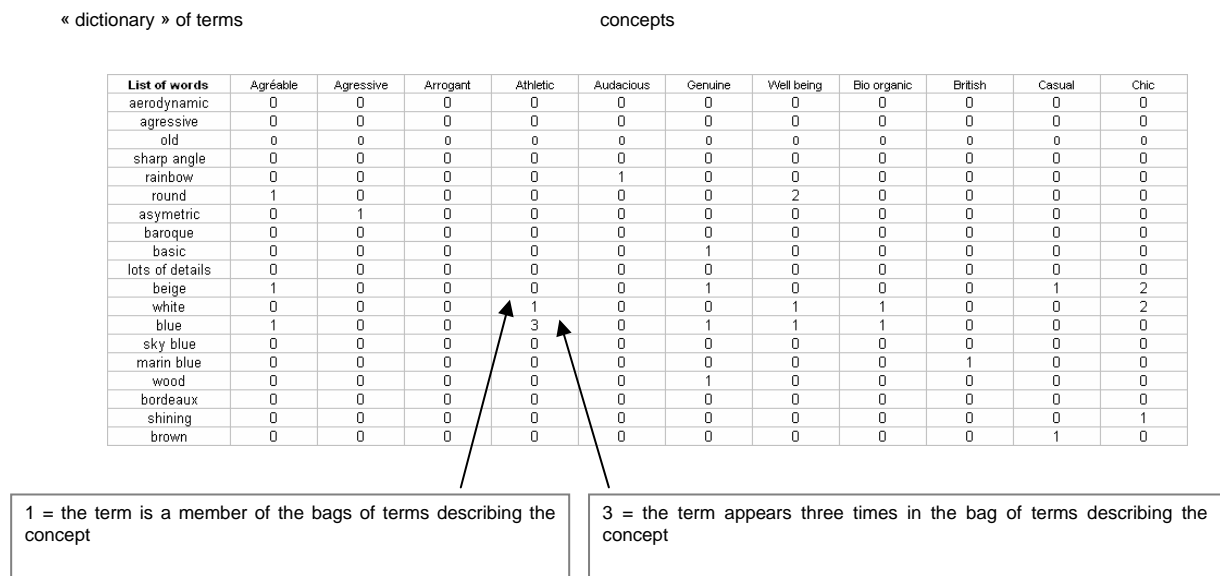


Figure 4 Result data: frequency matrix

The data collection was then transformed as a matrix of term frequency. A part of this huge matrix is shown Figure 4. It consists in putting the dictionary of terms in columns, and the concepts in rows. but To avoid useless noise we only used terms that appear at least two times in different bags/concepts. Each cell of the matrix at concept C (row) and term T (column) is filled up with the frequency of the apparition of the term T in the description of the concept C (can be from 0 to 3 depending on the numbers of the designers who have used T to described C).

For instance on Figure 4, “blanc/white” appears one time in the bag/concept “athletic”, and “bleu/blue” appears three times in this same bag/concept. By counting terms like that, we can establish some “correlations” between bags/concepts AND terms : if two terms appear in the same bags, they must be linked by a design rule, as well as if two concepts carry out some identical terms. By doing this, we just defined the *description space* needed for the H.A.C. to run. The remaining parameters of the algorithm will be defined as follows :

- *distance measure* : we chose to use the intersection measure (sum of the minimum values of the two vectors) between pairs of vectors instead of using the Euclidian distance. This takes into account the common number of terms associated to their frequency of apparition in the documents. In fact, the intersection grows as many terms are found common between semantic adjectives (or design elements), and as these terms tend to appear several times.

- *merging strategy* : we chose to consider a simple linkage where terms where to be linked to each other, and the distance between groups considered as the minimal distance between two of their points. This is mainly because the connectivity of the terms in our interview was weak : it was quite difficult to find common terms in the different interview data. So by merging points one to one and considering a point to point aggregation, the words would soften the limits between groups and let us observe more correspondences between terms.

4. RESULTS DISCUSSED

By applying the H.A.C. algorithm on the frequency matrix obtained in section 3.3., we obtained some relations between terms : either semantic adjectives (concepts) or design elements (shapes, colours and textures). There are in fact two different kind of results.

4.1. Clustering of the semantic adjectives

First, we obtain a tree of semantic adjectives (concepts). This tree – the dendrogram produced by the H.A.C. - let us put together concepts that carry out a similar meaning or value. On Figure 5 we can observe some of the associations between semantic adjectives.

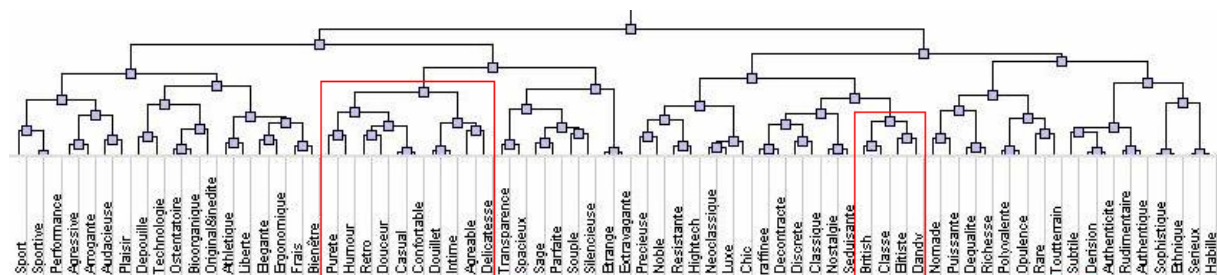


Figure 5 : the clustering of the semantic adjectives (in French)

Most of the associations drawn by the H.A.C. algorithm carry out an intuitive correspondance that tend to validate the results obtained: “sport” associated to “performance”, “casual” associated to “comfortable”, etc. We also observe some interesting groups of concepts like “british” / “classy” / “elitist” / “dandy”, or another group made of “delicate” / “agreeable” / “intimate” / “soft” / “sweet” / “casual” / “comfortable” / “pure” (“humour” and “retro” also are included in this group). Even if these results can be argued (they should be), they indicate a possibility of automatic associations between concepts.

4.2. Revealing design rules

Second, we can obtain a tree gathering the design elements (colors, matters, textures and shapes), but this cannot be meaningfully interpreted without using the frequency matrix itself, once ordered using the dendrograms as coordinates. We have put this representation on the Figure 6 below.

The proximity of terms in this tree-ordered matrix is meaningful because it is supported by the parallel associations made between the concepts and between the design elements. On this figure we can observe “clouds” of associated terms: we have pointed out such “clouds” that are based on semantic adjectives and design elements. These clouds lead to the formulation of design rules, they can only be detected by using the proximity of the concepts and the proximity of the design elements as resulting

from the H.A.C. algorithm. The main revealing results are the following (these results have to be considered specifically in the shoes design field):

- The concept of “sport” was highly linked with fluorescent, aggressive, streamlined, shaped and asymmetric.
- The concept of “originalit”y was associated with unusual colours in the field of shoes design like the fluorescent green or orange, and inedited textures like smooth leathers or rubber.
- The concept of “technology” was mainly linked to other concepts like minimalism (which have a direct impact on shape) and futurism; also tight describing mainly the lines was also correlated to this concept.
- The concept of “precious” was associated with concepts like rare, balanced which implies symmetric forms and baroque, and chrome texture.
- The concept of “tough” was found linked to the concept “hedge”.
- The concept of “powerful” was highly linked to colours like khaki or greys, matters like synthetic clothes or elastomer, and the chrome texture.
- The concept of “rare” was related to colours like pink or purple, also to oversized concepts, and to elastomer matter.
- The concept of “subtle” was above all linked to pastel colours and plastic matters.

These results appear as more or less obvious and are only applicable in the field of shoes design.

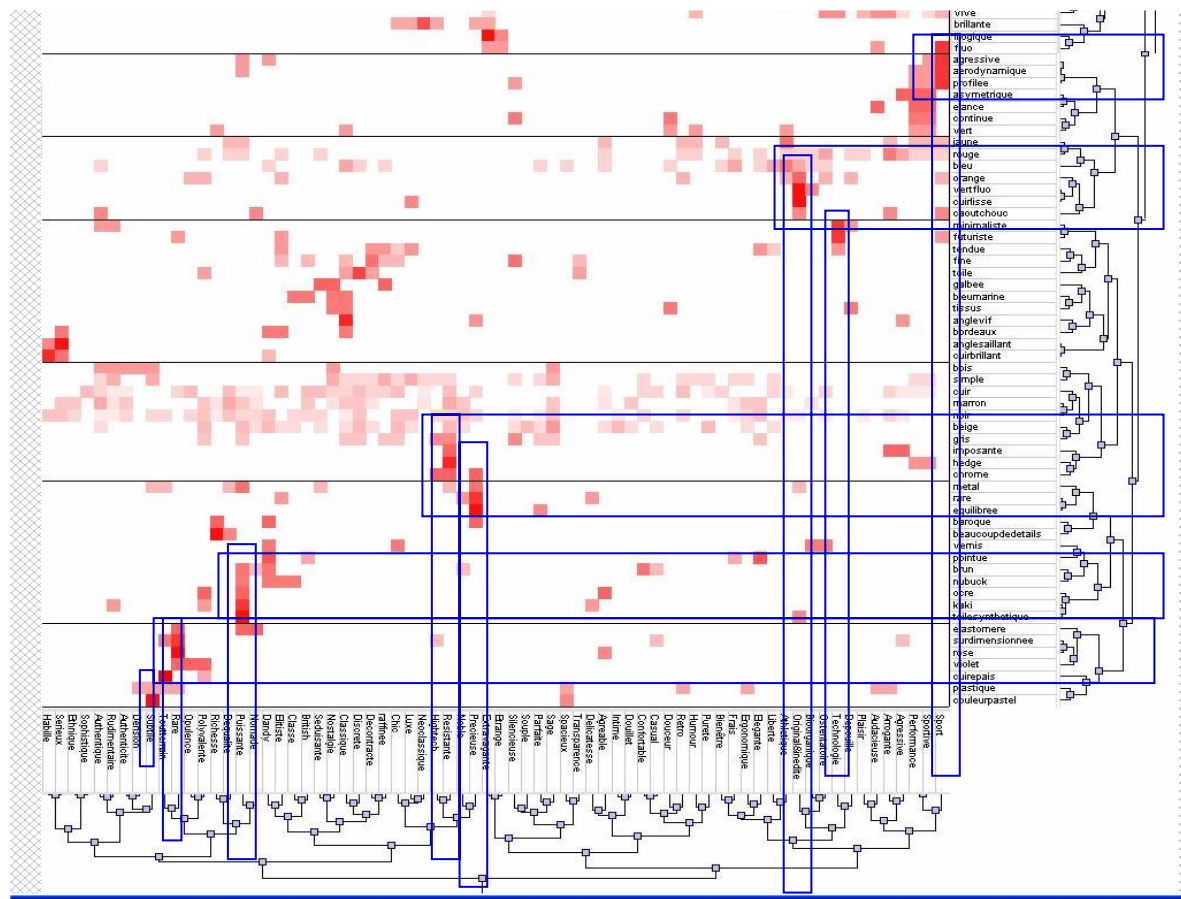


Figure 6. Agglomeration of the design elements

4.3. Extension to other design domains

The experiments led in the context of our project have focused on shoe manufacturing. But the methodology and the tools used for that do not depend on the activity domain of the designers. We could as well extend our approach to other domains.

Other domains more related to engineering activities could be targeted as well. For instance, in the context of human machine interfaces design, similar algorithms are used to classify labels and icons in order to build a hierarchy of modules or functions. Card sorting algorithms similar to the H.A.C. are used to do that. Coupling a classification of functions with a classification of visual attributes related to the interface design could enable us to propose colors or shapes of interface components linked with functional attributes.

Also, the procedure for extracting designer's rules or knowledge could be used to determine some meta-rules in engineering design. For instance we could be asking the designer's to express links between design domains and tools employed for the conception of products. We could also let the designer's express project parameters or tools that would fit in a specific context. This would lead to the formalization of design rules based on the designer's implicit knowledge of how the design tools are linked to each other. It could be employed to build a hierarchy of tools in engineering design.

5. CONCLUSION

Promising results were found, proving that the application of agglomerative clustering is of great interest for formalizing design activity into algorithms, in a flexible manner without losing the expert point of view and requiring a great number of subjects. Indeed this tool coming from the artificial intelligence community can be used without constraining the designers into artificial protocols and biasing potentially so the results of the experiments.

REFERENCES

- [1] BOUCHARD C., AOUSSAT A., Modelization of the car design process, International Journal of Vehicle Design, Vol 31.1, ISSN 0143-3369, 2002, p1-10
- [2] BOUCHARD C., AOUSSAT A., Design process perceived as an information process to enhance the introduction of new tools, International Journal of Vehicle Design, Vol 31.2, ISSN 0143-3369, 2002, P162-175.
- [3] MANTELET Fabrice, BOUCHARD Carole, AOUSSAT Améziane, Analysis of the correlation between words and design elements for the generation of a Kansei Engineering System, International Design Congress IASDR 2005 Taiwan.
- [4] BOUCHARD Carole, MANTELET Fabrice, AOUSSAT Améziane, Semantic structure as input data for Kansei Engineering Systems, International Design Congress IASDR 2005 Taiwan.
- [5] TANOUE, C., ISHIZAKA, K. AND NAGAMACHI M. (1997). Kansei Engineering: A study on perception of vehicle interior image. International Journal of Industrial Ergonomics 19 (1997): 115-128
- [6] YUKIHIRO MATSUBARA, MITSUO NAGAMACHI 1997. Hybrid Kansei Engineering System and design support. International Journal of Industrial Ergonomics 19 (1997) pag. 81-92
- [7] JINDO T., HIRASAGO K. A study of kansei engineering method applying to car interior. Human Factors in Organizational Design and Management – IV.
- [8] AGRAWAL Rakesh, IMIELINSKI Tomasz AND SWAMI Arun N., Mining Association Rules between Sets of Items in Large Databases, Proc of the 1993 ACM International Conference on Management of Data, Washington, 1993, pp 207—216
- [9] RASMUSSEN E., Clustering algorithms, in: W.B. Frakes, R. Baeza-Yates (Eds.), Information Retrieval, Prentice Hall, Englewood Cliffs, NJ, 1992, pp. 419–442.
- [10] MITCHELL, M. T. (1996): Machine Learning. McGraw Hill, New York
- [11] KOJADINOVIC I. Agglomerative hierarchical clustering of continuous variables based on mutual information, Computational Statistics & Data Analysis 46 (2004) 269 – 294
- [12] GORDON, A.D., 1999. Classification, 2nd Edition. Chapman & Hall, London.

Contact: J.F. Omhover
ENSAM - LCPI
151, boulevard de l'Hôpital
75015 Paris
France
+33 (0)1 44 24 63 78
+33 (0)1 44 24 63 59
jean-francois.omhover@paris.ensam.fr
<http://cpi.paris.ensam.fr>