Creating Urban Information for Cartographic Generalization

Annabelle BOFFET
Cogit Laboratory, Institut Géographique National, France
Annabelle.Boffet@ign.fr

1. Requirements for urban area Generalization

The aim of the generalization process is to make a map from cartographic data according to specified scale, objective, and to a given geographical space: that means deleting less important information to emphasize more important information. Nowadays reliable results can be computed interactively, but in the context of automated process, an automated interpretation is still required to define what information to delete or to emphasize. According to numerous authors who have explicitly formulated the requirements for the automation of generalization, e.g. (Brassel *et al.*, 88) (Buttenfield *et al.*, 91) (Mc Master *et al*, 92), the amelioration of preliminary landscape description is a critical prerequisite. Better descriptions are necessary to improve the understanding of spatial organizations and thus to decide on better algorithms for a given space. Many generalization applications developed in GIS systems suffer from lack of such information (i.e. spatial landscape description or contextualization) and provide disappointing results in their would-be “all-automated version.

Whereas interactive generalization makes use of human cognitive and intuitive processes to orchestrate the generalization process, automated generalization has to make do with the available algorithms and the available data description. (Ruas, 98) identified three different conceptual levels of data description that are necessary in the generalization process : micro - meso - macro. Databases are currently described at both the micro level with isolated objects (e.g. building, road) and the macro level of all objects which belong to a specific category not necessarily spatially restricted (i.e. population). These simple levels of description are not sufficient to describe and generalize the complexity of human worlds at small scales, notably so for urban areas that show multi level characters to be kept:

- *At large-scale representations*, urban raw data are often suitable, describing concrete objet ata micro level of description, such as a building, a road, or a community boundary. Automated generalization process may only considering the own characteristics of each object to transform it : urban objects are enlarged, simplified, displaced, squared, or deleted according to their own constraints.

- *At smaller scales*, such micro objects are inadequate, the desired objects, in the urban areas are more abstract, called meso, such as a district, a block, or a town. Those find full meaning in their inner spatial arrangement of buildings and streets. Such meso objects are not described in the current databases and need to be created by derivation of the raw data.

The creation of meso objects consists in grouping elementary features that make up same geographical realities into geographical entities. A meso object is a set of objects that collectively represent a geographic phenomenon significant for the map to be made. The difficulties of deriving meso object from micro objects lie in this notion of geographical phenomenon, which requires knowledge and the human capacity of interpretation in interactive processes. In opposition to human awareness, the computer's “knowledge” in the automated version is a logical succession of computations, of measures on digital representations and of stored resulting data.

Our objective is to propose a spatial analysis of urban areas, providing new urban information of meso level for the whole space of the urban area. First, we need to derive the perceptually natural town limit, which is different from the raw administrative limits. In order to provide an intermediate meso object between the micro buildings and the meso town, we derive blocks, and analyze the characteristics of each block to define districts. In a second turn, the inner organization of each district needs to be studied in order to describe the inner structures that may exist and have to be kept or emphasized such as the alignment of houses along streets in residential districts. Town, block and district are defined as such:

- *The Town* is defined as a closed set of dense building zones of 20 ha minimum, where the maximal distance between two buildings is not more than 100 meters. See other definitions in (INSSE, 96)

- *Urban Blocks* are defined within the town limit by cycling streets, which contain buildings. Also defined by (Ruas, 98).

- *Districts* are visually and functionally characterized blocks. In reality, districts are rather difficult to define as their real limits often superimpose one another, due to their natural fuzziness.

The derivation of such contextual information from raw data in a generalization platform will ameliorate generalization result. We are looking for invariant characteristics of the town, such as districts or inner structures, that have to be kept in the generalization process. We consider that the generalization operation of selection is probably the first beneficiary of such derived information. The selection operation will know which exceptions to keep and to emphasize, and which objects to delete in dense homogeneous districts, in order to maintain the character of the district while keeping its distinction from its neighbors. It allows also homogenizing the generalization of similar districts by using the same sequence of algorithms (which provides a faster and above all more stable generalization).

This short description of the requirements of automated cartographic generalization of urban areas has highlighted the fundamental notion of group and structure description to provide information. We will now focus on the two complementary methods to create such information. Section 2 deals with the classification of predetermined urban blocks to identify urban districts. This method has been implemented on a GIS software, first results are presented. To complete this global description, the section 3 focuses on a theoretical method to analyze the inner organization of each classified block to identify local urban structures. The paper concludes with a global framework where the presented method can be nested.

2. Classification of Urban groups

This section deals with the classification of urban groups. The first sub-section puts forward the key issues of classification, then our urban classification method that has been implemented on the Laser-Scan's object-oriented GIS LAMPS2 is described.

2.1. Key Issues for the Classification

The classification is the creation of homogeneous groups, it provides a continuous structuration of space, each object belongs to a classified group. This analysis method is all the more relevant as it creates information that is an interpretation of the full geographical space. It makes the complexity of the studied space more graspable and more tractable.

Classifying is analyzing a double entry table: *units* and *variables*. It provides groups of units according to variables and eventually dependencies between variables. Key issues of the classification are thus units, variables, and the applied method.

2.1.1. Initial Units

Initial units are individuals to classify. Individuals are located in a coordinate systems and may vary with the resolution of the data.

2.1.2. Variables

Variables are used to classify units. Each unit is informed according to variables, which requires to know how to measure variables on units.

2.1.3. Classification Methods

A great number of reference books on classification methods exist, detailing and comparing their different characteristics, see e.g. (Haggett, 73) (Sanders, 89) (Pumain *et al.*, 97). A major distinction has been made between *goal directed* and *unsupervised* classifications. Both usually build a classification tree with all units to classify at the beginning and groups of classified units at the end, information being progressively discriminated. To each branch correspond one or several discriminant variables, knowing that the number of variables is limited by the rapidly increasing complexity of the tree.

- *Goal-Directed Method* (or supervised classification). Final classes are predetermined according to the user's need. The decision tree itself is decided by the user's model, which justifies the introduction of each variable. The user designs his model according to his needs, purposes, experience and accumulated knowledge and theories of geographical reality. Difficulties lie with choosing the adapted variables, measurements and thresholds and with interpreting intermediate cases. Thresholds depend on the purpose of the study: they can be either of a predetermined type for a general discrimination, or calculated specifically (that well suits the studied case). The chosen thresholds have to be sufficiently sensitive to discriminate significant classes.

- *Unsupervised Classification*, which groups units that most resemble each other according to an unbiased set of variables. The user takes no a priori decision on the shape and characteristics of the classification tree (i.e. choice of thresholds, variables and final classes). In this case, difficulties lie with the interpretation of the created groups.

The three key issues of the classification being described , the following sub-section will detail our urban classification by describing successively the key choices of units, variables and method.

2.2. Urban Classification for a Meso Level of Information

The purpose is to identify homogeneous urban areas to define meaningful districts. The description of our classification method follows the three key issues listed above.

2.2.1. Initial Units

Initial units of *town* and *blocks* that are usually used in several discipline must be available, they belong to the meso level of information. (figure 1) shows the representation of raw data on a test town : Laverune - 2200 inhabitants - located in south-eastern France. The notion of town limit and blocks are missing. Figure 2 shows the units of town and blocks computed by the software in accordance with our method :

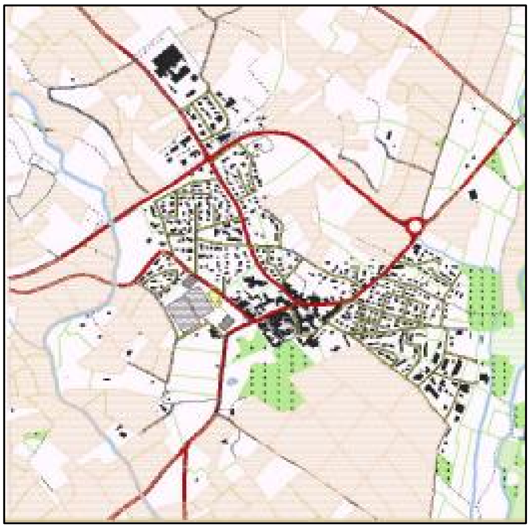

Figure 1: Initial map of Laverune

- To create the initial units of *town* a morphological algorithm has been used (D.Ormsby, pers.comm.) according to the town definition given in the first part of the paper. The town is delimited in accordance with perceptual reading of building density. The result is satisfactory, side effects due to suburbs are well computed : one can see peripheral units that actually do not belong to what we would call the city.

- Then, a structuration of the town in several *blocks*, based on mixed topological and semantic requests is computed. We consider roads, hydrological networks, and town limits as structural objects and use them to make a division of the map space into several parts

Figure 2: Urban situations of Laverune

2.2.2. Classification Variables

Our classification theory is based on the successive assessment of urban blocks, with intra-block to qualify them. Two sets of variables are taken into account:

- *Functional* : Essentially *thematic criteria* are analyzed with *buildings sorts* such as housing, industrial, or sporting buildings.

- *Gestalt* : The emulated perception of urban blocks is analyzed by *geometric criteria* such as size of buildings. To complete this one-by-one analysis, *contextual criteria* such as density and homogeneity are also studied within each block.

For each variable , a majority or a tendency, and particular cases are computed with a view to identify block types. Literature and statistics have been used to define thresholds. Urbanistic thresholds provide a general classification that allows authoritative descriptions of the city. But that may not fit well to the cartographic vision of the city, for which statistical thresholds naturally provide a specific classification.

2.2.3. Classification Method:

We have made the choice of a *goal-directed classification* because we know what kinds of districts are required for generalization, as exposed in the first section. As explained before, the classification method is represented as a hierarchical classification tree with preliminary urban blocks at the beginning, and a priori defined urban districts at the end including residential or collective housing, industrial or sporting zones. Figure 3 presents the whole classification tree we have elaborated.

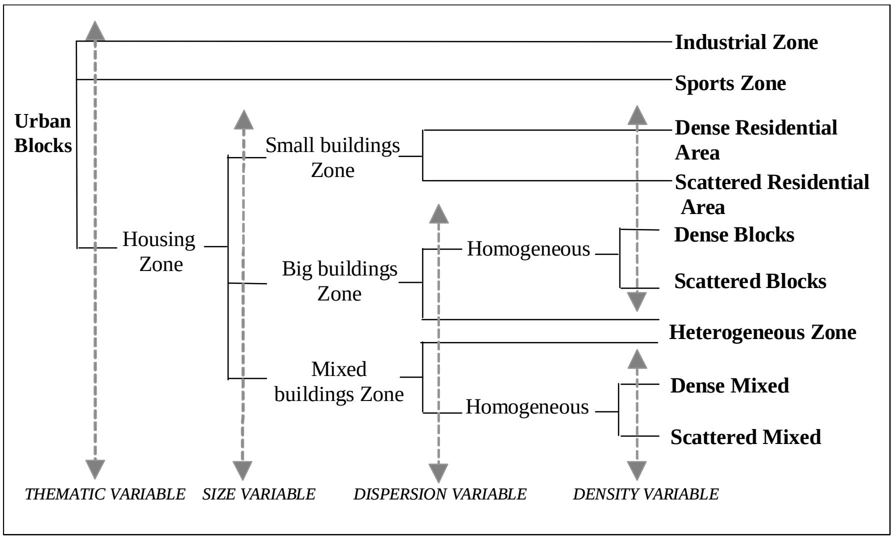

Figure 3: Urban hierarchical classification tree

Our classification tree (figure 3) shows four levels of analysis. The first one is based on *functional* analysis, the three others on *Gestalt* analysis:

- Functional analysis based on *thematic variables* allows the creation of three classes that need to be kept during the generalization process: industrial, sports and housing zones. The majority types are computed on both the number and the surface of building types for each urban block. The industrial and sports zones are already well defined for generalization. Therefore, only "housing zones" will be further classified.

- The first Gestalt analysis is based on *building sizes*, and aims at defining which buildings are big or small within an urban-block. The threshold is hard to define, after tests and thanks to cartographical experience, the limit has been set at 250m^2, a threshold that proves suitable for many towns. Majority of size has then been used to classify blocks: When no majority appears, the class “mixed” is attributed.

The housing blocks are now classified into residential, mixed and collective zones. But the heterogeneity of big and mixed blocks needs to be further analyzed to distinguish blocks where size is homogeneously medium (near $250m^2$) or where there is a mixture between bigger and smaller houses ; variations are bound to be reduced in small housing zones.

- To do this, a fixed numerical threshold for size analysis is not sufficient to identify real ***urban block heterogeneity***, a character that has to be identified, being, just as exceptions, important for the selection operation in the generalization process. To analyze the statistical repartition of the sizes of buildings, which may show a wide range in heterogeneous cases, the standard deviation of building sizes has been used in these areas. Thresholds are entirely statistical, and thus naturally adapted to the town to be classified. We obtain two classes: homogeneous and heterogeneous units. The heterogeneous zones cannot be further studied.

- The last, and most important classification is based *on building density*. Each class is separated into two: dense and scattered, to be treated differently during the generalization process. Different statistical thresholds have been found for each type of housing block.

This analysis provides nine classes of urban districts. The final class names are given with a geographical point of view, which allows recognizing habitual urban districts. We have made the choice of variables and indicators that seemed most relevant according to the generalization requirements on this town after numerous tests on the small town of Laverune. (Coquerel, 99) reports on the implementation off it into *LAMPS2*, a Laser-Scan’s object-oriented GIS. Figure 4 shows the classification result computed for the town of Laverune.

Figure 4a distinguishes the industrial and sports zones from the housing blocks with the first thematic variable of the method. The last three maps detail housing block zones by distinguishing residential blocks (4b), collective blocks (4c) and mixed blocks (4d). The difference between dense and scattered blocks is made in each of these three maps (the finest levels of the classification method).

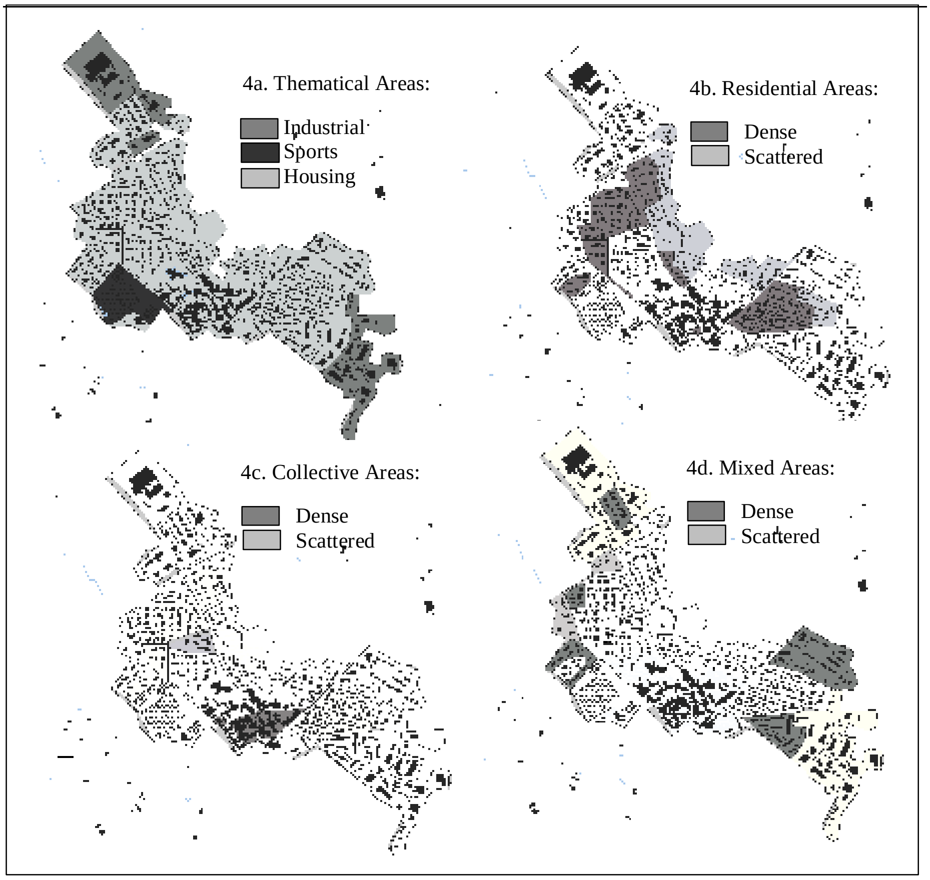

Figure4: Urban Classification of Laverune

To conclude on the presented urban classification, it seems valuable because it provides a meso level of geographical information, intermediate between the local level of buildings and the global level of town. This meso level of information is useful to contextual generalization, helping to maintain the main characteristics of each block during generalization processes, which are building kinds and sizes and block density and homogeneity. What is still missing in such description is the spatial arrangement of the town's elements : There is no information on the possible structures that may appear with the relative positions of the micro objects, such as alignments. A complementary still theoretical method of inner block description, based on the relative position of objects is described below. While the presented urban classification classifies the whole space, the following approach deals with the recognition of local structures that belong to classified urban blocks.

3. recognition of Urban principles

To complete the full description method of the classification for urban areas, this section addresses the identification of *urban structures*, which are made of regular organizations of buildings, and respond to urban construction principles. Such structures are complementary with classification, they describe the inner organization of classified blocks that needs to be emphasized by the generalization process. The first part of this section deals with two existing linear structures identified for generalization requirements of urban areas. Then, our proposition based on urban principles is presented.

3.1. Existing Methods for the Identification of Urban Structures

Works on urban areas carried out at the Cogit laboratory have resulted in cartographic generalization operators. Those, relying on the adapted analysis of geographical space, reveal its utmost importance for generalization purposes of course, but indeed for any kind of GIS applications.

Motivated by the urge for the typification operator, (Hangouët, 98) and (Regnauld, 98) were brought to recognize significant urban distributions for their typification tools. Typifying consists in *carrying into a representation, where the identified distribution is preserved, objects of a same nature that happen to be grouped by some identifiable geographical process* (Hangouët, 98, p. 227). To analyse urban distributions, they have created linear groups of nearby and similar urban objects, but their methods differ.

- (Hangouët, 98, p.191-200) defines the '*inner access'*: both a road arc and the buildings it serves on one of its sides are groups. Proximity criteria are measured by means of a Voronoï diagram. The street is considered determinant as it belongs to the group "inner access".

- (Regnauld, 98, p.56-118) searches for sets of buildings grouped in homogeneous rows. The rows are computed from a Minimum

Spanning Tree, a non-cyclic graph that reveals closest neighboring relationships. The whole graph is divided where the buildings' characteristics suddenly contrast or where roads intersect it. The street is also important but it does not belong to the group.

Whatever the peculiarities of the methods developed by Hangouët and Regnauld, each recognizes linear structures (linear organizations of buildings) to treat it differently. These methods are of great interest for the automation of generalization, being accompanied by generalization tools to treat these linear structures specifically. Both have to be federated in a generic approach that should be able to decide on the appropriate cases where each best applies. Other fundamental structures also need to be recognized in the town.

The next section proposes a method to identify generic urban structures ; the method is based on spatialization principles of which flexibility is the main asset.

3.2. Extending Spatialization Principles

Given the great variety of towns due to their history and territorial characteristics, it seems difficult to give an exhaustive typology of urban structures to recognize. Therefore we renounce using templates to recognize and prefer the other tactic of using construction principles to identify spatial organizations of components. We call those construction principles ***spatialization principles***.

Spatialization of space results from processes of populating, territory appropriation, management, exploitation, network implantation, carried out by a community over a geographical space destined for its usage (Pinchemel, 95, p.64). The main interest of spatialization principles lies in the flexibility and adaptability they provide to the generic structuring approach. Space is no longer divided through fixed templates, but by means of organization processes because a template is the result of the observation of the organization of things, urban objects in our case. Just as identifying templates is snapshot analysis of results, identifying principles is upstream analysis of a process. We are searching for urban construction principles, which belong more generally with spatialization principles.

Urban structures are widely described in urban literature as e.g.(Lynch, 71) (Beaujeu-garnier, 80). A typology of the properties of sub-systems, from which shapes, structures and constructions can be analyzed, is given in (Pinchemel, 95, p.19)

- polarity, centrality, concentricity, periphery,
- linearity, radiality, laterality, angularity, perpendicularity,
- orthogonality, triangularity, equilaterality
- interiority, exteriority,
- symmetry, dissymmetry, anisotropy

This typology fits with the purposes of geographers' analysis, but it is yet too complicated to be applied automatically in a GIS system. We have adapted it to our requirements and defined priorities. Our proposition lies with the analysis of ***Angularity*** between elementary features. The principle of angularity manages the relative positions of objects, it can thus be considered as one of the fundamental keys for local context analysis. We use the deduced principles of ***Alignment, Parallelism, Perpendicularly*** and ***Closure*** for Angularity analysis between street or buildings*:*

- ***Alignment*** is exemplified by residential houses (figure 5).
- ***Parallelism*** is exemplified by houses in their alignment on street (figure 5).
- ***Perpendicularity*** of streets is typical of Roman urban plans (figure 6).
- ***Closure*** is exemplified by the street network of Strasbourg's suburbs (figure 7).

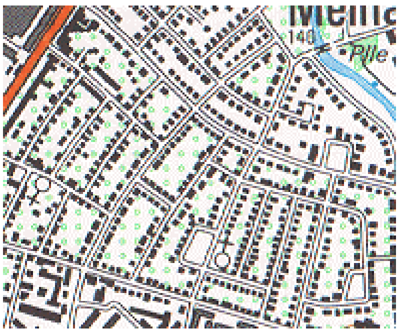

Figure 5: Alignment and parallelism in residential area of Strasbourg

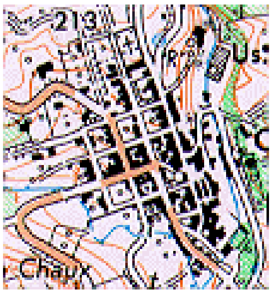

Figure 6: The perpendicular plan of Cenne-Monesties

Figure 7: Closed circulation in a Strasbourg suburb

As the examples show, spatialization principles can be viewed at different scales. Even if we could identify patterns in each example, it is not our objective. We do not want to build a library of urban shapes. Our purpose is to understand how towns are organized, according to what properties. Alignment, parallelism, perpendicularly and closure are our first answer. Typical examples that are completely verified are rare, most current cases present a combination of them. The plurality of principles emphasizes rarely and more likely destroys the evidence of them. The main difficulties lie in determining which principles are effectively present and to asses them.

In the perspective of automated principle detection, existing GIS facilities would provide efficient automatic tools for our analysis of angularity principles. Topology for example, describing relations of connection and inclusion, is helpful for our analysis:

- *Intersection* and *Alignment* can be identified in a " star crossroads " from the connection of aligned arcs of roads (figure 8, the effect is all the more spectacular as the road arcs radiate regularly).

- *Inclusion* and *Closure* can be found in particular towns named by (Pawlowsky, 92) as " circulades"; road arcs closed in a circle containing smaller road arcs which are also closed in a circle and so one. (figure 9)

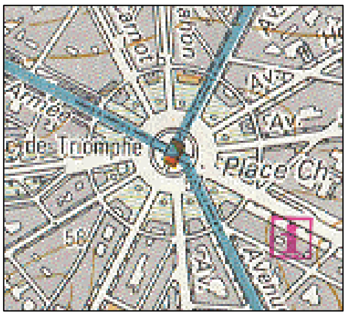

Figure 8: A star crossroads of Paris

Figure 9:Bram'scirculade

The identification criteria for a spatialization principle would be composed of two parts:

- Specific *Methods* that measure the principle on and between elementary features (in the present study, Angularity measures)

- *Thresholds* to decide whether the principle is actually present or not. In intermediate cases, a confidence level can be associated to the principle identified.

Measures and thresholds must be sufficiently and precisely described. The definition must be an efficient compromise between off-hand definition, useless for effective identification, and too rigid a description that would lead to deterministic, exclusive identification.

The theoretical methodology for the analysis of local structures analysis has been presented. No implementation of this method has been done yet. It is our next activity. Nevertheless the next section addresses the organization of both this method and the implemented classification described in the second section of the paper. The creation of a relevant level of urban information requires the combination of both.

4. Complementarity of the two methods

The two methods, classification of urban areas and recognition of urban principles described in paragraphs 2.2 and 3.2 provide complementary urban information. This section addresses the general use of them, how one method is required by the other.

From raw urban data divided in urban blocks the classification provides nine classes of urban districts. Then spatialization principles between buildings can be searched for inside each classified urban block to describe their inner organization. Every principle may not need to be searched in every district. A priority law of searching should be considered. If any principles are found, the confidence level computed for each principle would define its relative importance. The principle that shows the highest level of confidence would be accepted and would increase the information created by the classification.

- The principle of house alignment described above could be searched preferably locally in Housing zones. Figure 10 shows two blocks of Laverune that have been classified as housing zones. Figure 10a shows a block with a grid like organization of buildings (which is the highest level of confidence of the alignment principle on a surface). At the opposite figure 10b shows no principle with a sufficient level of confidence, the block is a spatially unorganized one.

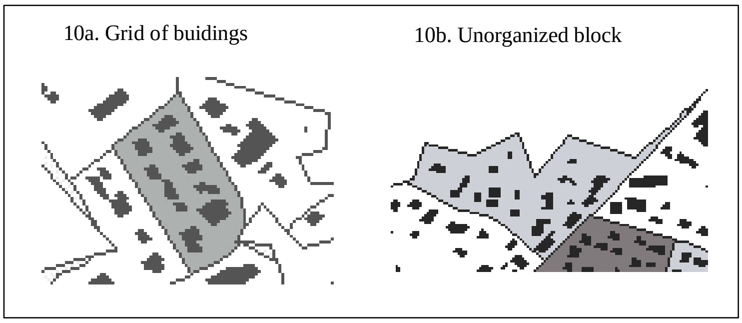

Figure 10: Principles within housing classified blocks of Laverune

This step allows the creation of complementary information locally, at the urban block level. This analysis of principles is all the more interesting as it may detect some spatial organization even in blocks classified as heterogeneous sizes.

- Heterogeneous classes need to be re-classified in a second turn. This requires a new division of heterogeneous urban blocks according to the inner repartitions. The sub-division could be classified with the same method as presented above; the initial units would have a finer resolution. The inner repartition of buildings must show some structure to be efficiently segmented: different homogeneous groups placed side by side have to be identifiable. Figure 11 shows three urban blocks classified as heterogeneous, two of them could be each segmented into two homogeneous units that could be reclassified.

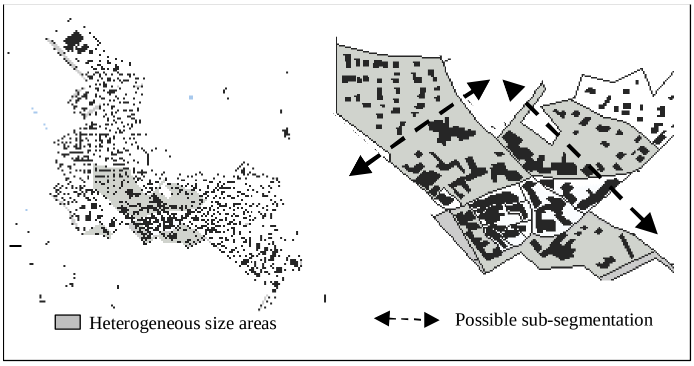

Figure11: Segmentation of heterogeneous blocks of Laverune

Principles could also be searched between several urban blocks or all. It is global analysis : The example of (figure 6) shows the global recurrence of the perpendicularly principles.

5. CONCLUSION

The aim of this paper has been to propose methods to create information that meets the requirements of generalization. We analyze urban areas to create urban information. A short recall of the requirements of generalization has emphasized the need of identifying urban structures. The most important requirement for the urban area is information on meso object : What are their principal types ? Do they contain any particular structure ?

This paper has proposed a framework to create urban information from raw data. It is composed of two complementary kinds of spatial analysis. The first method is a goal-directed classification, that allows to interpret the whole studied urban area. It computes nine classes that are interpreted as districts. Each urban object of the classified area belongs to a district. The classification variables used are both functional and visual. This method of classification has been implemented and provides relevant results to answer to the first question.

To complete the spatially continuous analysis of classification, a complementary method based on spatialization principles has been detailed. This method is local analysis of urban structures that can appear inside classified districts. Angularity principles only have been studied, but they can be associated with topological analysis.

The last section of the paper addresses the complementarity of the two methods. Classification and principles can be used concurrently for the benefit of the derivation of meso urban information. Even if the second method has not been implemented yet, the first results are encouraging and angularity principles are to be implemented and tested in the near future.

Acknowledgments

I would like to thank Caroline Coquerel who worked on the classification during her internship at Cogit laboratory, Anne Ruas and Jean-François Hangouët for their helpful comments. And Daniel Ormsby for sharing his town-delimiting algorithm program.

References

Beaujeu-Garnier J. (1980). *Géographie urbaine*. Armand Colin/Masson, Paris

Brassel KE. & Weibel R. (1988) "A review and conceptual framework of automated map generalization". *Geographical Information Systems,* vol.2, no. 3, pp. 229-244

Buttenfield BP. & MacMaster RB. (1991). *Map generalization: making rules for knowledge representation*. Longman Group, UK

Coquerel C. (1999). *Classification de zones urbaines,* Mémoire de fin d'études ENSAR -IGN

Haggett P. (1973). *L'analyse spatiale en géographie humaine*. Armand Colin, Paris.

Hangouët JF. (1998). *Approche et méthodes pour l'automatisation de la généralisation cartographique ; application en bord de ville'*. Rapport de thèse Université de Marne la Vallée.

INSSE, 1996. Economie et statistique. n° 294 -295.

Lynch K. (1971). *L'image de la cite*. Dunot

McMaster, R.B., Shea, KS., 1992. Generalization in digital cartography. Association of American Geographers, Washington.

Pawlowsky k. (1992). *Ciculades languedociennes de l'an mille.* Presses du languedoc.

Pinchemel P. & G. (1995) *La face de la Terre Eléments de géographie*. Armand Colin

Pumain D. & Saint-Julien T. (1997). *L'analyse spatiale. 1.Localisations dans l'espace.* Armand Colin / Masson, Paris.

Regnauld N. (1998) *Généralisation du bâti : structure spatiale de type graphe et représentation cartographique*. Rapport de thèse Université de Provence, Aix - Marseille I.

Ruas A. (1998). "OO-Constraints modeling to automate urban generalization process". In, *Proceedings 8 th International Symposium on Spatial Data*. 11-15 July, 1998, Vancouver, Canada.p. 225-235

Sanders L. (1989). *L'analyse statistique des données en géographie*. G.I.P.Reclus, Montpellier.