URBAN CLASSIFICATION FOR GENERALIZATION ORCHESTRATION

Annabelle BOFFET, Caroline COQUEREL
Institut Géographique National, France
COGIT Laboratory
annabelle.boffet@ign.fr
caroline.coquerel@free.fr

Technical Commission IV/5

KEY WORDS: Multi-scale analysis, Generalization, Urban objects, Classification, Object-oriented.

ABSTRACT

Nowadays, many generalization algorithms exist but their management is still problematic. If algorithm design needs to be carried on, especially for contextual generalization, 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 a better orchestration of the algorithms. To improve automated analysis, we suggest that a given geographical space should no longer be described by its individual objects, but be analysed by several larger entities created by pertinent structuration processes.
This paper proposes a method of spatial database interpretation to automatically define decisional entities for the generalization process. It focuses on the creation of urban information from urban topographic data. The first section reviews the requirements of automated generalization, which emphasize the lack of useful information for urban generalization. Our method to automatically derive information on urban structure is then described. Implemented on LAMPS2, an object-oriented GIS, the analysis is based on a goal-directed classification. Lastly, the results for the classification of a 8800-inhabitant-strong town are presented. The paper concludes with on-going research improvements.

1 INTRODUCTION

1.1 Generalization requirements

The necessity of generalization comes from the opposition between the constant human limits of vision and the numerous levels of observation of the geographical space (Bertin, 67). The aim of the generalization process is thus to make a map according to specified scale and 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. Many generalization applications developed in GIS systems suffer from lack of information or contextualization and provide disappointing results in their would-be “all-automated version”. Numerous authors have explicitly formulated the requirements of generalization, (Brassel *et al*., 88) (Weibel *et al*., 98). Even if algorithm design needs to be carried on, especially for contextual generalisation, 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, a mechanism called “orchestration” by Mackaness in (Ruas *et al*,97).

Orchestration of the generalization process consists in determining which algorithms to use on which objets and when (Mc Master *et al*, 92). 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, which as isolated objects (e.g. building, road) are still of a simple type. This level of description is not sufficient to describe the complexity of human worlds. (Ruas, 98) proposed three different conceptual levels of description that are necessary in the generalization process, focusing on the second one, which fits with the generalization requirements and must be inferred by analysis:
- Micro-analysis at a very local level such as the analysis of a building (i.e. one)
- Meso-analysis of a set of objects such as a district or a town (i.e. group)
- Macro-analysis of all objects which belong to a specific category not necessarily spatially restricted. (i.e. population)

Following the concept of multi-level analysis, this paper proposes to work at the meso level to define spatial and semantic entities. The creation of meso objects consists in grouping elementary features that make up same geographical realities into geographical entities. A meso object is thus 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. The process of identifying phenomena is a cognitive process, which requires knowledge and the human capacity of interpretation: The knowledge involved in the human process of interpretation is rich and adaptive. It consists of the individual's awareness that entities exist, of his ability to describe and classify them, of his knowing that other entities may exist, also identifiable by means of other known notions. In opposition to human awareness, the computer's "knowledge" is a logical succession of computations, of measures on digital representations and of stored resulting data.

The aim of the paper is to describe a method that builds automatically meso objects that are relevant for the generalization. The following subsection deals with the requirements of generalization on urban areas. Afterwards, our proposed method of urban classification is detailed. The last section deals with experiments and improvements.

1.2 The case of urban areas

The interest of urban areas lies both in their high density in existing objects that needs to be generalized and in their multi level characters for which an abundant literature is already available (Linch, 71), (Barriere *et al*, 80), (Pelletier *et al*, 97).

Urban raw data are often suitable for large-scale representations, describing concrete objets, such as buildings, roads, or community boundaries. At the moment the urban generalization process is interactively computed on raw data: urban objects are enlarged, displaced, squared, or deleted according to the users orders. The rationality of such urban objects is basically legal or technical, because of the single micro level of description.

At smaller scales, such micro objects are inadequate, the desired objects in the urban areas are more abstract. Those are finding full meaning in their relative spatial arrangement. They require varied tools that have to integrate spatial analysis and modeling functions. The description of their spatial organization informs the users on phenomena such as town growth, or town functionality. Some urban organizations have been studied at the Cogit laboratory and algorithms to generalize them, such as homogeneous rows (Regnauld, 98) and inner accesses (Hangouët, 98) tools are efficient to identify their specific entities, and the associated algorithms provide satisfying generalization results. The drawback lies in their spatial restrictions: homogeneous row and inner access are admittedly adapted to certain urban zones but not to the whole urban area of a town.

Our objective is to complete these researches by proposing a spatially continuous analysis of urban areas, providing new urban information for the whole space of the town. First, we need to derive the perceptually natural town limits, different from the raw administrative limits. In order to provide an intermediate meso object between the micro building and the meso town, we derive blocks, and analyze each block to define districts. Town, block and district are defined as:

The town is defined as a close set of dense building zones of 20 ha minimum, where the maximal distance between two buildings is no 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 homogeneous urban blocks. It is rather difficult to define districts as their real limits often superimpose one another, due to their natural fuzziness.

The derivation of such contextual information from raw data in the generalization platform will ameliorate generalization orchestration. We are looking for invariant characteristics of the town, such as districts, 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.

Describing the requirements of urban area generalization has highlighted the fundamental notion of entity recognition to provide information. Analyzing them has listed what entities to find and why. The next section deals with our classification method to create the meso objects blocks, districts and town.

2 CLASSIFICATION OF URBAN ENTITIES

It turns out that classification is an interpretation tool useful for both data description, by providing a new structuration of space and creation of entities: each elementary object is made to belong to a classified group. This section deals with the classification and the construction of urban groups, which relies on a combined analysis of what entities are, how they are perceived, and their functionality in the real world. The first sub-section puts forward the key issues of classification, then our urban classification method is described.

2.1 Key issues for Classification

Initial Units are individuals to classify, which are located in a coordinate systems and may vary with the resolution of the data.

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

Classification Methods: a great number of reference books on classification methods exist, detailing and comparing their different characteristics (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 top and groups of classified units at the bottom, 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 has 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 that are final classes.

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 Proposition of a Urban Classification

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. To create the initial units to classify, we use mixed topological, geometric and semantic queries. Elements like roads, hydrological and railway networks, geographically structuring are used to divide space into several units, which belong to meso-level. (Figure 1.) shows result on a test town: Laverune - 2200 inhabitants - located in south-eastern France. The result is not very satisfactory because of side effects due to suburbs. One can see peripheral units that actually don't belong to what we would call a city.

Therefore an algorithm to detect morphological town limits has been added (D.Ormsby, pers.comm.). The limit is integrated into the previous queries to refine space structuration. (Figure 2.) It is obvious that the town is better delimited, in better accordance with perceptual reading. Moreover this division of space makes downstream spatial analysis more efficient, in preserving the general context between individual objects and the whole volume of the map.

The town of Laverune was the first used to define the method. The classification has also been applied on another French town called Trets.

Figure 1. Segmentation of Laverune
(Boundaries: roads, hydrological and railway networks)

Figure 2. Segmentation of Laverune,
(adding morphological town limit)

2.2.2 Classification Variables. Our classification theory is based on the successive assessment of urban blocks, with intra-block analysis to qualify them. Assessment relies on the study of the characteristics of the buildings in each urban block. Variables of two kinds of are taken into account:

Functional . Essentially *thematic criteria* are analyzed with *buildings* sorts such as housing, industrial, commercial, administrative, religious, or sporting buildings.

Gestalt . The emulated perception of urban blocks is analyzed by *geometric criteria* such as size, height and shape of buildings. To complete this individual 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 Methods. 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 it explained before, the classification method is represented as a hierarchical classification tree with preliminary urban blocks on top, and a priori defined urban districts at the bottom including residential or collective housing, industrial or sporting zones. Figure 3. presents the classification tree we obtained after numerous tests on the small town of Laverune. We have made the choice of variables and indicators that seemed most relevant according to the generalization requirements on this town, and the method has been retested on the city of Trets.

(Coquerel, 99) reports on the implementation off it into *LAMPS2*, a Laser-Scan's object-oriented GIS. It does not take every theoretical criterion quoted above into account. Functional, size and density criteria are used in order to see whether a relevant classification could be done with minimalist criteria.

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 $250m^2$. It has turned out that this threshold may be suitable for many towns.
Majority of size has then been used to classify blocks: When no majority appears, the class “mixed” is attributed (meaning areas where size is homogeneously medium or where there is a mixture between big and small houses).

To recapitulate, 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 (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 can shows 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** (3). Each class is separated into two: dense and scattered, to be treated differently in 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.

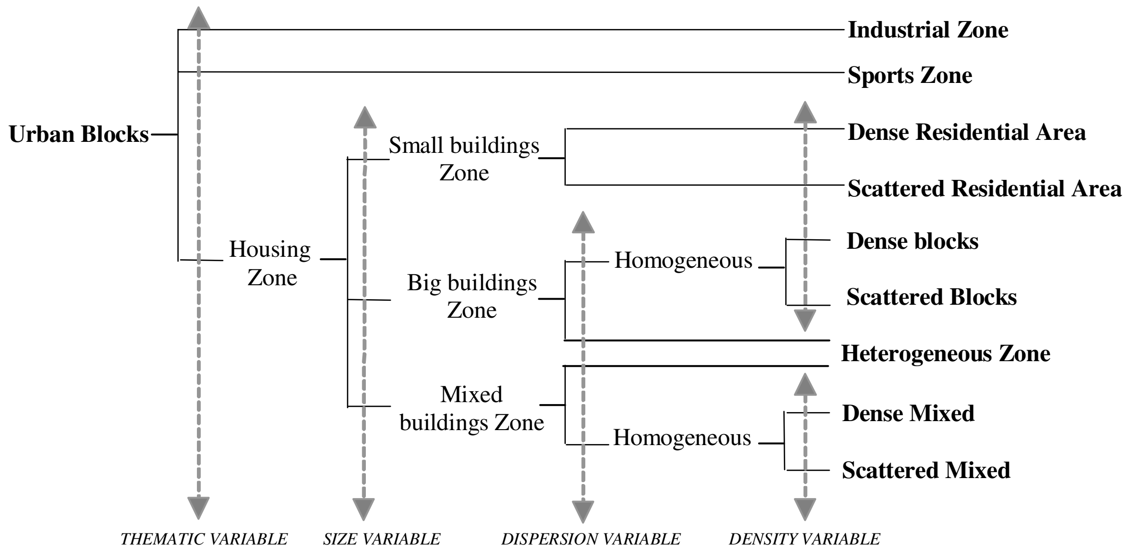

Figure 3. Urban hierarchical classification tree

3 EXPERIMENTS

3.1 Test on TRETS

The classification presented above has been tested on the city of “Trets”, a 880-inhabitant town from the south of France. The result is very promising as the method developed for a small town of 2200 inhabitants, “Laverune”, also produces a meaningful classification for “Trets”. The main results of the classification on Trets are shown in the four maps of Figure 4.

Figure 4a distinguishes the industrial 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).

Criteria such as building heights, shapes, proximities and the complementary themes of commercial, administrative and religious districts, which have been described in the theoretical method, have still to be implemented. Subsequent improvements could also enrich the description of the town; they are described in the next subsection.

Figure 4. Classification of the city of Trets

3.2 Improvements

The experiments on the cities of Laverune and Trets show a discriminant structuration of towns. The classified classes accord with the visual interpretation of the whole urban space. But in its current programmed version, the method needs to be improved. Further precision on heterogeneous zones and further information on downtown and the surrounding organization are still needed for the automated generalization of urban areas.

3.2.1 The heterogeneous blocks could be analysed in a second turn, to find any sub-homogeneous structure without creating minor entities. Some heterogeneous blocks may require new division according to inner repartitions. The hypothetical homogeneous sub-divisions could be classified with the same method as presented above; the initial units would have a finer resolution. The difficulty of this task lies in the subdivision of urban blocks. 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 5.) shows three Laverune urban blocks classified as heterogeneous, each of which could be segmented into two homogeneous units that could be reclassified.

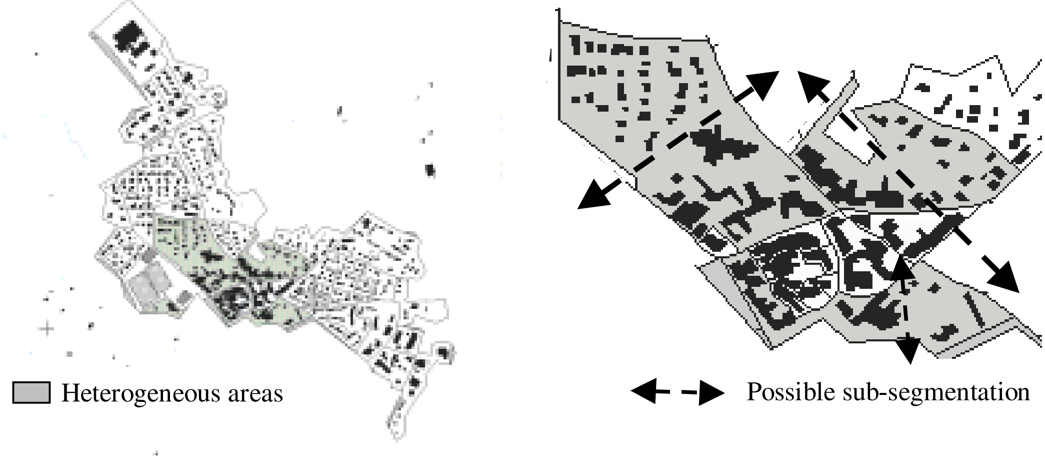

Figure 5. Segmentation of Laverune's heterogeneous areas

3.2.2 Downtown is at the moment computed as the densest big buildings housing zone. The method provides reliable results for Laverune and Trets, but it will certainly not be so for bigger towns. A solution could be to compute several methods adapted to town size. Numerous authors have described "town" and "downtown" in France e.g. (Barriere et al. , 80) or all over the world e.g. (Pelletier *et al.*, 97). The awareness of this notion of downtown is firmly fixed in the collective apprehension of town, but associated to several descriptions depending on both the point of view (i.e downtown inhabitants, activities, or shape) and on size of the associated town. From an urban shape point of view, the French downtown is usually of three types:

- Downtown is associated in small towns to a crossroads or a commercial street.
- The couple of religious edifice and administrative building locates downtown in medium towns. Non built areas are limited and collective flats are higher even if height regularity is preferred to the real number of floors.
- For bigger towns, it seems reliable to define several centres, of historical, administrative or commercial types. In numerous cases such as fortified towns, high density of narrow sinuous streets and circular boulevard are remarkably recurrent in the historical downtown.

From such definitions, the reliable detection of downtowns will be implemented. It will bring new information on town growth and a means to interpret the organisation of the town.

3.2.3 Town organisation will also be analysed by means of town growth models (figure 6.) such as those reviewed in (Beaujeu-Garnier, 80).

- The first model of Park and Burgess explains the concentric circle growth of the town, which is homogeneous in every direction in homogeneous human and natural contexts.
- The second model, of Hoyt, takes communication into account, and defines a favourable zone where activities are preferably installed. It also defines several sectors of activities centred along the communication network.
- Then, the Harris and Ullman model shows the result of the first models combined and the notion of multi-centrality that provides a mosaic organisation. Each of the models highlights the importance of downtown for the analysis of town growth and organisation.

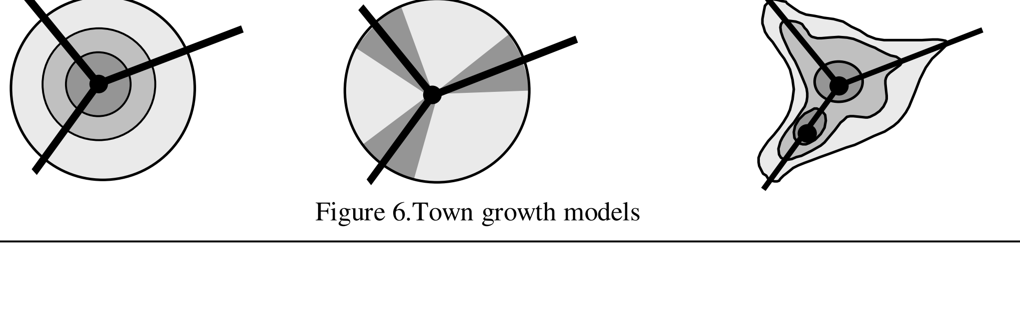

Figure 6.Town growth models

4 CONCLUSIONS

The aim of this paper has been to propose and describe a method to create information that meets the procedural requirements of cartographic generalization on urban areas. We analyze urban areas to create urban information. A short review of the requirements of the generalization process has emphasized the need of identifying urban entities. The paper focuses on the requirements for urban areas that is first, information on town and districts entities and secondly, information on the organization of the districts in the town.
The paper has proposed a framework to create urban information of town and districts from raw data. The method is a goal-directed classification that allows interpreting the whole studied urban area. After delimiting the effective town limit, 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 for a 2200 and a 8800 inhabitant town. It seems valuable and promising because it provides a meso description level of geographical information by considering groups of buildings. This meso level of districts and town is useful to contextual generalisation, helping to maintain the geographical meaning of each urban block and the full expression of the town during generalisation processes. To complete the classification analysis, propositions on heterogeneous districts, town centers and spatial organization of all districts together are to be implemented and tested in the near future for the improvement of the generalization orchestration.

ACKNOWLEDGMENTS

We are grateful to Anne Ruas for her helpful comments and to those who accepted to read the first draft.

REFERENCES

Barriere, P., Cassou-Mounat, M., 1980. Les villes françaises. Masson, Paris.

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

Bertin, J., 1967. Sémiologie graphique. Les ré-impressions, Paris.

Brassel, K.E., Weibel, R., 1988. A review and conceptual framework of automated map generalization. Geographical Information Systems, vol.2, no. 3, pp. 229-244.

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, J.F., 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, K.S., 1992. Generalization in digital cartography. Association of American Geographers, Washington.

Pelletier, J., Delfante, C., 1997. Villes et urbanisme dans le monde. Armand Colin, Paris.

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. Proceedings 8 th International Symposium on Spatial Data. 11-15 July, 1998, Vancouver, Canada.p. 225-235.

Ruas, A., Mackaness, W., 1997. Strategies for urban map generalization. Proceedings International Cartographic Conference, Stockholm, Netherlands, V3 pp. 1387-1394

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

Weibel, R., Dutton, G.H., 1998. Constraint-based automated map generalization, Proceedings 8 th International Spatial Data Handling. 11-15 July, 1998, Vancouver, Canada, 214-224.