SPATIAL ANALYSIS AND AGENT PRINCIPLES TO AUTOMATE GENERALISATION PROCESS

Anne RUAS
Laboratoire COGIT - IGN-France
2 avenue Pasteur
94 165 Saint Mandé FRANCE

ABSTRACT

To generalise digital data, some efficient algorithms are required. Recent researches have enriched the library of algorithms allowing interactive generalisation of GIS. The automation of the process requires that the generalisation system is able to find automatically 'where and how to generalise'. The 'where' depends on the capacity of analysis: the system should be able to identify which objects do not respect the specifications. The 'How' depends on the existence of procedural knowledge which are rules that connect conflicts with the algorithms which are supposed to solve such conflicts. The AGENT prototype has been developed to provide the geographical objects with the capacity to qualify themselves (i.e. to recognise automatically the conflicts) and to apply appropriate algorithms to solve the existing conflicts. Moreover, a mechanism of control has been developed in order to backtrack and try another algorithm whenever the results are not as good as required. This paper explains the principles of self-generalisation at different level of details as well as the importance of spatial analysis for generalisation.

KEY WORDS
Generalisation, Agent, Spatial analysis, Constraints.

1. INTRODUCTION

The process of generalisation aims at producing a new representation of geographical space from a more detailed one. It could be compared to the conception of an abstract from a full paper. The abstract should content less words, but the main ideas should be detected, preserved and enlighten. For geographical data some graphical rules must be satisfied to produce a representation that can be read and understood. The objective is to represent the same geographical space with less details while allowing the readers to build a mental representation reflecting the reality. Research on generalisation aims at designing a system able to build an abstract representation for any type of geographical area and any type of level of details. It requires a good understanding of the process of abstraction.

Early research in the eighties focused on the definition of the operations of data modification such as simplification, aggregation, displacement and object removal (see [1]). Then researchers designed algorithms for each of these operations, starting from the simplest ones to the more complex such as displacement [2]. The result of this research can be seen on some GIS packages of generalisation that allow a user to select an algorithm and apply it to one or several objects selected interactively. But the automation would require an orchestration of these algorithms as suggested by Mackaness [3]. Tests such as [4] clearly showed that 'how' to generalise (i.e. what to apply, where, and when) was not a trivial question as no predefined sequence could produce good results with the existing algorithms. The choice of an algorithm and its parameter values depends on the properties of each object or set objects to generalise.
From that point, three approaches have recently been proposed by the research community:

1. To design new algorithms that would model generalisation as a constraint solving problem [5]. Constraints of proximity, of size are represented at the co-ordinates level. Classical mathematical methods such as least square one are used to displace in an iterative way the position of each point in order to respect constraints,
2. To try a set of combination of operators and to choose the solution that best satisfies a set of constraints. The theoretical number of trials depends on the set of objects and on the possible combination of operators. Simulated annealing technique is used by [6] to reduce the quantity of trials and to choose if not the best at least a good solution,
3. To try to replace the cartographer analytical capacity by spatial analysis tools that would guide step by step the process. This method is presented hereafter.

These three methods are all very interesting. The two first are certainly very valuable and efficient for small scale changes whereas the last one can easily include dramatic changes (such as object removal and collapse) but it requires more spatial analysis tools. A first remark is that all of these methods are based on constraints as it was proposed by K. Beard [7].

This paper presents an approach based on agents and constraints to automate the generalisation process. It begins with a short explanation of the principles of the method before developing the important role of spatial analysis.

2. PRINCIPLES

The model we are proposing is based on constraints, levels of analysis and agents. It has been first proposed by [8] [9] and has been enriched during the European Esprit AGENT project [10]. The model is fully detailed in [11], we will just present hereafter the main principles before focusing on spatial analysis in section 3.

The main global principle of this model is that generalisation is viewed as a state change process. The initial (non-generalised) data are generalised step by step using algorithms that are chosen and triggered dynamically and automatically. The choice depends on two important states : the initial one -which is preserved as the reference- and the current one. The aim of the process is, at each step, to find a good next step. Control mechanisms are part of the process to check dynamically if a 'trial next step' can or not be validated to become the current state. This step by step process follows the human generalisation process.

2.1 THE CONSTRAINTS

Constraints are the representation of 'what should be respected' to obtain a good generalisation. They are formalised by means of functions such as :

size(building) > 350m².

We distinguish two kinds of constraints :

- Constraints of generalisation represent the new rules that should be respected according to the final product specifications. They include graphical constraint -such as building size- and constraints related to information reduction. Usually, constraints of generalisation are not initially satisfied and algorithms are used to improve their state.
- Constraints of maintenance constrain the information to 'look like' the initial state in order to produce a final representation that allows the user to build a mental representation close to the reality. For example a generalised entity should have a shape if not equal at least that 'looks like' its non-generalised shape. These constraints are also used to avoid an homogenisation of the representation. For example, after generalisation, the centre of a city should be different from its suburb (if required in the specifications). Constraint of maintenance are initially satisfied but might be violated after applying the algorithms. They are necessary to avoid or to reduce the possible distortions.

Constraints help to know 1/ 'where to generalise' (e.g. this building should be dilated because it is too small : it does not respect the size constraint on buildings) and 2/ if the proposed solution is acceptable or not (e.g. the solution of building dilation (i.e. size emphasising) should not be validated because the shape of the building changed too much : it does not respect the constraint of shape preservation on building). Of course a large set of constraints (on buildings, on streets, on rivers, etc., related to size, shape, density, etc.) are necessary to perform generalisation and often it is difficult to find one solution that strictly respects all the constraints. The objective is to find a good balance for each object (or set of objects), if possible the best. Whenever a constraint is not satisfied a conflict is created. Some conflicts are more sever than others do. The challenge is to reduce the severity of the conflicts to reach a state where most of the constraints are satisfied.

To facilitate the use of constraints during the process, it has been proposed to associate with each entity (each road, each building) the state of its property towards the function of constraint [8]. To apply an algorithm the system needs to know that this building has a size, which is significantly smaller than the minimum building size. To do so, each entity is described by its initial and current properties (such as size, position, shape, orientation) and each current property (e.g. the current size) is evaluated against the corresponding function of constraint. For example if the current building size of *this_building* is 280 m², and the building size function is 'size(building) > 350m²' then the state of the size of *this_building* is *bad* and requires to be improved by means of a generalisation algorithm such as size_emphasing.

Consequently each entity is described by the state of each of its properties. These states are used to trigger generalisation algorithms and to control that a proposed solution still respects the constraints of maintenance. A building that has been simplified in order to respect its granularity constraint should have a generalised shape that also respects the constraint of shape maintenance (a nearly squared shape should not be transformed into a round or a flat shape). Moreover, as an entity can have a set of non-satisfied constraints, *priority of treatment* are computed to allow the system to choose which constraint should be solved first. These priorities are computed dynamically according to the nature of the constraint and the severity of the constraint violation. As an example, if a set of buildings is too dense and buildings are too close, it is better to reduce the quantity first (by means of object removal) before trying to displace the buildings as there is not enough room for all of them.

2.2 THE AGENTS

As seen previously, each geographical entity is described by a set of properties. The global state of an entity is the synthesis (or the aggregation) of all its property states. In other words, the state of an entity is computed from the conflicts that exist on each of its properties. Whenever an other words, the state of an entity is computed from the conflicts that exist on each of its properties. Whenever an entity has *a bad global state* it should be generalised. The idea is to allow each entity to compute its own global state and to trigger itself its own generalisation. To do so, we used the concept of agent.

An agent is an entity that acts (and interacts) by itself. It means that it can give to itself the order to simplify its border to respect its granularity constraint or to emphasise itself to respect its size constraint. In terms of implementation, it means that each agent has an 'engine' that allows it 1/ to compute its state (the state of each of its property and the synthesis), 2/ to choose an algorithm to improve itself and to trigger it 3/ to control its new state, to validate the solution if it has improved or to backtrack and choose another algorithm if it has not. Phase 1 and 3 depend on the state of the properties towards the function of constraints whereas phase 2 depends on procedural knowledge. The schema of the process of an agent is given in figure 1. In our system, each geographic entity becomes an agent, which means that it has an engine that allows its self-generalisation. The term situation is used to describe a geographical entity whatever its level of detail.

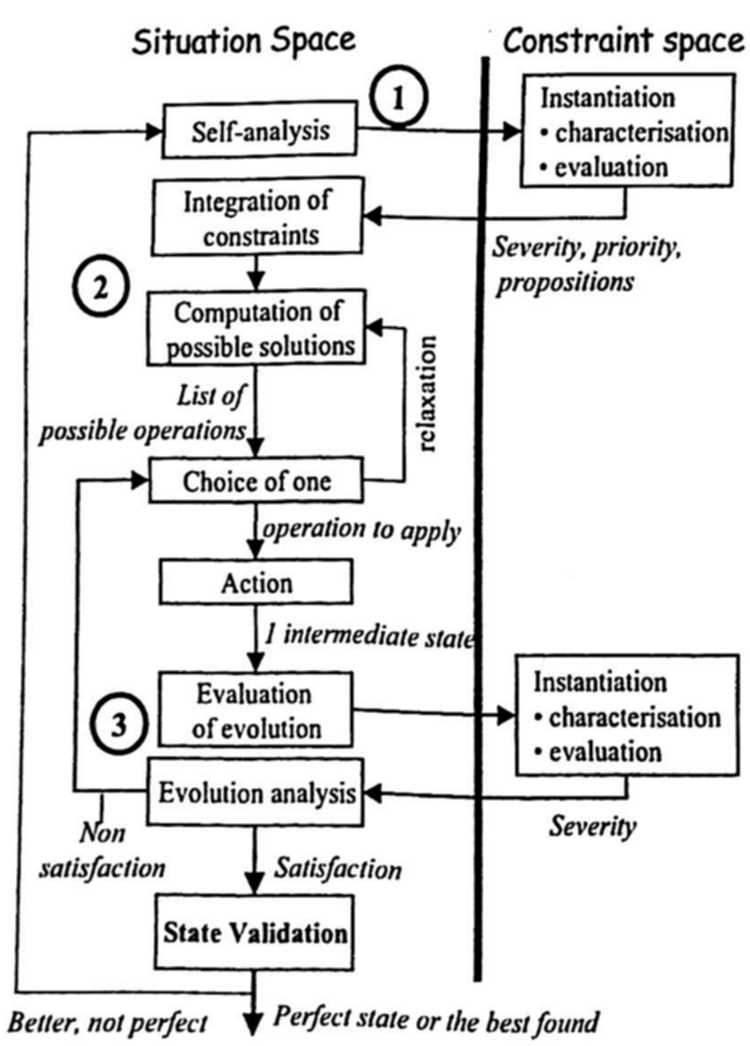

Figure 1: Process of self-generalisation for a single agent [8].

Amongst generalisation operations, some are individual (such as simplification, size emphasising) while others are contextual. Contextual operations are applied to a set of objects. Displacement, object removal, aggregation are contextual operations as the decision of 'where to displace', 'which object can be removed' or 'which objects can be aggregated together' can only be taken with an overview on a set of objects. To do so, we introduced the concept of meso-agents [8]: A meso agent is a group of agents that act collectively as if it was a single entity. A meso agent is an agent (consequently it generalises itself) composed of agents. To distinguish them, we call micro agent an agent that is not composed of other agents. The difficulty is then to understand which kinds of meso agents should be created for generalisation purposes, and to design methods to detect them as they are not initially represented in the database. Different approaches exist : we can create a meso agent each time a conflict can not be solved at the micro level. We can also create meso agents that correspond to geographical meaning such as an urban block, a district, a town or a street network. Generally these two approaches are rather coherent one to the other. Research done at the COGIT laboratory for generalisation purposes led to the definition of a set of meso agents necessary for the generalisation process and is illustrated in figure 2.

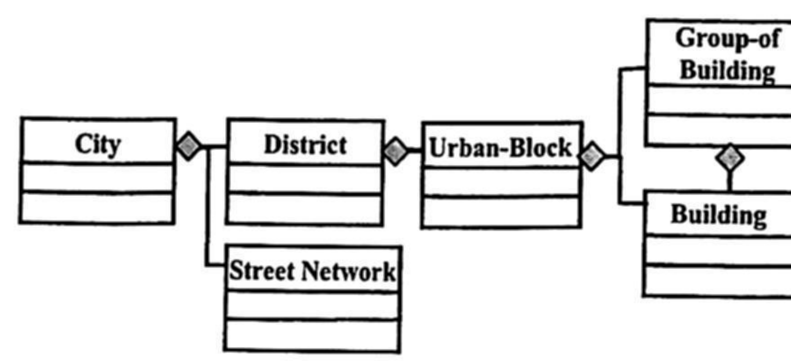

Figure 2: Meso objects type within a town for generalisation [12].

Amongst meso agents, some, such as urban block are used to trigger generalisation algorithms, while others, like districts, are only used to represent some constraints of maintenance. An *urban block agent* triggers and controls building removal and building displacement, a *city agent* triggers and controls street removal, whereas a *district agent* controls the fact that its semantic property is preserved during the process (its diversity of buildings nature).

To choose the appropriate algorithm to reduce a conflict or a set of conflicts, procedural knowledge [7] is described at the constraint level. Each constraint holds this knowledge described by means of rules such as : if state is *very bad* then use such or such algorithm (with such parameter value), if state is *bad*, then use such or such algorithm, otherwise if state is *medium* then use such or such algorithm. Information related to procedural knowledge is given in [11].

This model of generalisation has been used during the European project AGENT. The team developed a GIS package named Agent on top of Laser-Scan Limited Lamps2 GIS. During the project, only methods related to roads and urban areas have been developed. The project proved that this model is appropriate for generalisation. The use of this package in NMAs already allowed to improve the productivity, reducing the quantity of interactive work. Certainly the best result concerns the road network generalisation. More important, the system has proved flexible. It has been used for different databases, at different resolutions. On the other side, the evaluation we made also showed that research should be pushed forward not only to introduce new algorithms but also to design more and better spatial analysis tools.

3. THE ROLE OF SPATIAL ANALYSIS FOR GENERALISATION PURPOSE

The core of the generalisation process is based on our capacity to develop (and to master) measures able to describe geographical information as well as possible. These measures replace the analytical capacity of a cartographer who detects that a situation is not correct and who is able to change it while preserving its main properties. Measures are used to detect conflicts as well as the information and properties that should be preserved during the process: Spatial analysis tools are necessary to trigger and to control the process as suggested by [3]. Functionally, we can decompose the set of measures into two parts :

- The first part is used to recognise and create meso entities such as district, groups of building or ring road
- The second is used to describe the properties of the entities such as measures of distance, size, shape, orientation or density.

Besides the fact that spatial analysis tools are very important for a large set of geographic applications such as risk management, the library of measures is not enough identified as a goal in itself, even if GIS designers are proposing spatial analysis tool packages. At least for generalisation, each developer or researcher develops its own appropriate library from available tools on GIS or from a mathematical library such as CGAL. Designing a measure requires three steps :
1/ the identification of the mental concept coming from our perception,
2/ the transformation of this concept into a mathematical formula that can be implemented,
3/ the interpretation of the value of the measure (the measurement) in the context of generalisation purpose.

The last step is certainly the most complex. The interpretation of a value means that we should be able to give a meaning for each value according to a goal. For example, if a road has a fractal dimension equal to 1.32, what does it mean ? If it is changed to 1.25, is it an important change ? Does it mean that the shape changed dramatically ? Is it acceptable ? If some measures are trivial, others such as measures of shape are much more complex as there is no mathematical definition for shape. Example of this complexity can be found in the research done to describe the sinuosity of roads [13]. Every one can describe the sinuosity of a road in his own way, but how to describe it mathematically ? Often, it is nearly impossible to reach our human subtlety of description even though the measure is mathematically accurate : measures do rarely describe our mental concepts. However it is necessary to progress in the digital description of geographic structures to enrich our very poor data base – where objects are represented only by means of few attributes and their location - with the explicit representation of information necessary to automate a process (whose objective is to simplify the representation while preserving the geographical meaning). If a cartographer is able to visually detect what is important and what should be preserved and how, it is because he has 1/ knowledge on geographical space 2/ a drawing of the specific arrangements of objects that he can connect to its knowledge to reconstitute a mental image of the reality and 3/ experience. For example the generalisation of a river network requires that 1/ we know in theory what is a river network, what are its properties 2/ we see it globally to identify which parts can be removed without losing the global structure of this specific network, and this according to the spatial relationships that exist between the network and other objects. The automation of such a process requires 1/ to reconstitute the river network as a single entity 2/ to analyse its personal and contextual properties and 3/ to develop an algorithm that will reduce the quantity of rivers without losing the global property of a river network nor the main characteristics of *this* river network.

The complexity of such analysis is coming from the fact that most of the meso entities useful for the generalisation process are entities that can not be precisely defined. They are phenomena that can be named (such as a town) and that are useful to geographical reasoning. We are close to the Gestalt theory [14], which looks for *pregnant forms* that emerge from a set of information. Designing methods to detect such phenomena is not easy because no accurate or single definition exist. For example, there are thousands of possible definitions of a town.

To illustrate this difficulty but necessity of analysis let us take the example of building alignments. Building alignments are important on a map because they structure the space. They are the reflect of a certain kind of urban settlement, more specifically of middle class districts. Their maintenance is important during generalisation process because they hold implicitly this geographical (i.e. historical and economical) knowledge. Moreover, generalisation algorithms - such as building displacement, aggregation and removal - tend to distort them. Figure 3 shows an example of AGENT prototype generalisation of an urban block where two alignments are completely lost as they were not detected by the process. At least a 'good generalisation' should have preserved one of both alignments.

© Ruas GEOPRO 2003 – ISBN: 970-36-0103-0

Figure 3: Lost of building alignment with agent prototype.

The detection of such structures would help not only to constrain the displacement of buildings but also to develop a new method of building removal (named typification) adapted to such meso-objects.

Recent research [15] presents a method to detect and characterise this kind of pattern. The alignment in itself is not so difficult to detect by means of projection and clustering techniques (see [15]). The point is to be able to distinguish regularity from irregularity, as, for different reasons, exact regularity never exists. An alignment is perceptually important if and only if it has *some* regularities of arrangement, size, inter-distance, shape and orientation (see figure 4).

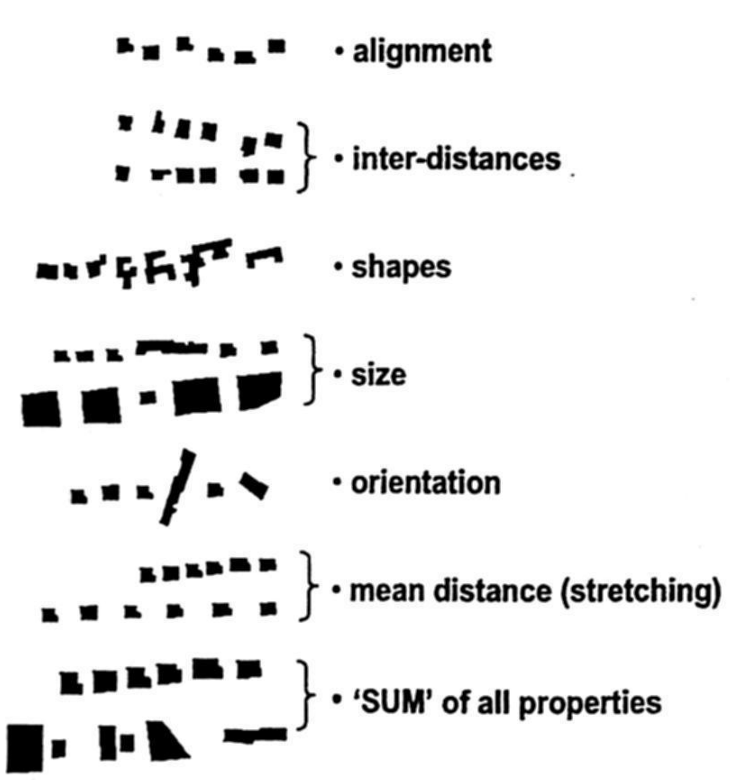

Figure 4: What makes a pattern.

If measures of alignment, distance, building shape, size or orientation, are not too complex, nor their regularity - as the standard deviation can be used to compute regularity-, the difficulty here is to go from digital values to an interpretation necessary to distinguish the main alignments that should be preserved. On going research based on expert knowledge acquisition will soon be published to present our proposed method and results. To sum up, some examples of alignments are proposed to cartographers who mark them. This information is used to find - if possible - the functions between the quantitative values computed by the measure with the qualitative values given by experts. These qualitative values are the result of the knowledge of the cartographers. In such a way their expertise can be introduced in the system. This method allows to compare heterogeneous properties - such as alignment, shape and size - and to build an aggregation function necessary to mark the global quality of each alignment. It then becomes possible to mark them and to select only regular alignments (the smallest mark, the best in figure 5).

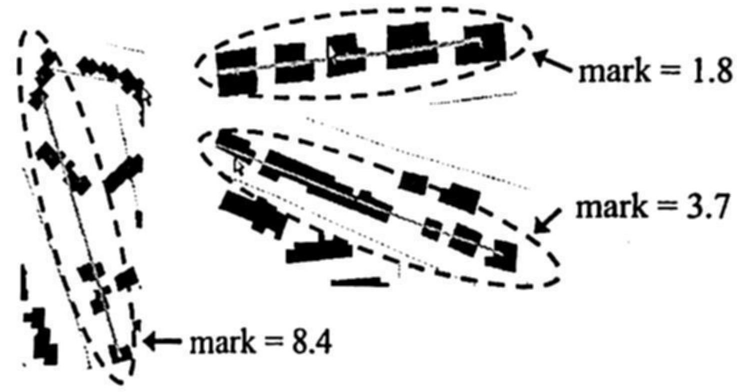

Figure 5: Evaluation of building alignments quality.

The best alignments are introduced as new meso-agents in the generalisation process with appropriate attributes and generalisation methods (Figure 6).

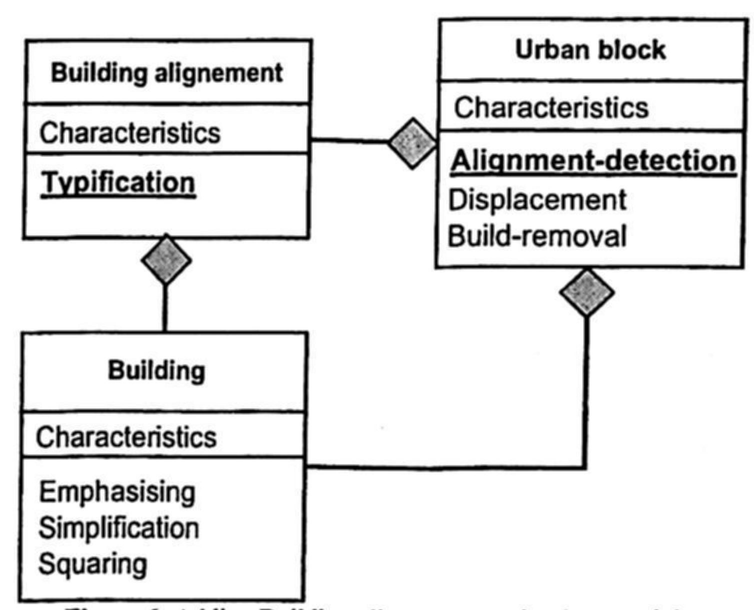

Figure 6: Adding Building alignments to the data model.

As a global result, the fact that we are now able to detect automatically these structures improves the Agent generalisation results : Each *urban block* meso agent has the task to detect and create these new meso agent that would be themselves responsible for their own generalisation : an *alignment* will apply to itself typification to reduce its number of houses while preserving the pattern, and the meso *urban block* will perform displacement while preserving the relative position of buildings (see figure 6).

Another aspect to enlighten is that the progress made in generalisation induced a real change in the data modelling as illustrated in figures 2 and 6. New classes of objects associated with their methods of creation and their characterisation (i.e. the computation of their properties) have been introduced in the initial modelling. This effort of formalisation tends to fill the gap between the abundance of the information of the real world and the weakness of the information contained in classical databases where only single object location and nature are described.

databases where only single object location and nature are described.

The detection and characterisation of building alignments illustrates the use of spatial analysis for generalisation purpose. Of course, lots of progress are still necessary as much more information should be explicitly represented to allow the computer to preserve the main information implicitly contained in the initial database.

4. CONCLUSION

In this paper we have presented the principles of one model of generalisation that has already been implemented by the AGENT team on the Lamps2 package Agent. This model, as opposed to others such as [5] and [6], has the advantage of pushing the explicit representation of the information that governs the process. This new information can be either a better description of the initial data - such as the explicit representation of distance, size, shape, density or orientation - by means of new attributes or the creation of new entities - such as towns, districts, urban blocks or building alignments - that are also characterised by new attributes. This effort towards the explicit representation of information exceeds the simple case of generalisation. For example, it could be used for multiple representation databases as these meso objects are often the real link between existing databases. Moreover it opens the way to new geographical analysis that require different levels of analysis, as the study of the influences of individual entities on more global ones. The phenomena of aggregation - which is the core of generalisation process - is recognised as being a core and complex domain in geography. The introduction of methods of analysis that allow the automated recognition of such entities should certainly improves research in the broader domain of geography.

ACKNOWLEDGEMENT

I want to thank Serguei Levachkine to invite me to the GEOPRO 2002 conference as well as Nicolas Regnauld for his helpful comments.

REFERENCES

[1] K.S. Shea and R. McMaster, Cartographic generalisation in digital environment : When and How to generalize, *Proc. Autocarto 9*, Baltimore USA, 1989 56-67.

[2] M. Bader, Energy minimization methods for feature diaplcement in map generalization *PhD Thesis* University of Zurich, Suisse, 2001.

[3] W. Mackaness, Issues in resolving visual spatial conflicts in automated map design, *Proc. Spatial Data Handing'94*, Edinburgh, UK, 1994, 325-340.

[4] A. Ruas Automatic generalisation project : Learning process form interactive generalisation. *OEEPE Official Publication N°39*, 2001, 11-98.

[5] L. Harrie and T. Sarjakoski, Simultaneous graphic generalization of vector data sets. *GeoInformatica* 6(3), 2002, 233-262.

[6] J.M. Ware, C.B. .Jones and N. Thomas, Automated cartographic map generalization with multiple operators: a simulated annealing approach, to appear in *The International Journal of Geographical Information Systems* 2003.

[7] K. Beard, Constraints on rule formation, in Buttenfield and McMaster (Ed.) *Map Generalization* (Longman, 1991) 121-135.

[8] A. Ruas, OO-constraint modelling to automate urban generalisation process, *Proc. Spatial Data Handling'98*, Vancouver, Canada, 1998, 225-235.

[9] A. Ruas Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. *PhD Thesis* University of Marne La Vallée, ftp://ftp.ign.fr/ign/COGIT/THESES/RUAS, 1999.

[10] M. Barrault, N. Regnauld, C. Duchêne, K. Haire, C. Baeijs, Y. Demazeau, P. Hardy, W. Mackaness, A. Ruas, R. Weibel, Integrating multi-agent, object oriented and algorithmic techniques for improved automated map generalization, *Proc. of 20th International Cartographic Conference*, Beijing, China, 2001, 2110-2116.

[11] A. Ruas, Data modelling for constraints-based generalisation, to appear in *CUES*, 2003,

[12] A. Boffet, Méthode de création d'information multi-niveaux pour la généralisation cartographique de l'urbain *PhD Thesis* Université de Marne la Vallée, ftp://ftp.ign.fr/ign/COGIT/THESES/BOFFET, 2001.

[13] X. Barillot and C. Plazanet, Analyse des formes des routes, in A. Ruas (Ed.) *Généralisation et représentation multiple* (Paris: Lavoisier Hermes Sciences, 2002) 203-224.

[14] M. Wertheimer, Über Gestalttheorie, *Erlangen*, 1925. In the translation by Willis D. Ellis published in his *Source Book of Gestalt Psychology*, New York: Harcourt, Brace and Co, 1938.

[15] S. Christophe and A. Ruas, Detecting building alignments for generalisation purposes, *Proc. Spatial Data Handling*, Ottawa, Canada, 2002, 419-432.