STRATEGIES FOR GENERALIZATION OF BUILDINGS AND BUILT-UP AREAS IN TOPOGRAPHIC GEOSPATIAL DATABASES

Melih BAŞARANER and Mehmet SELÇUK
Yildiz Technical University, Department of Geodetic and Photogrammetric Engineering,
TR-80750 Besiktas-Istanbul, Turkiye
{mbasaran, selcuk}@yildiz.edu.tr

Generalization of geospatial databases has particularly gained importance together with widespread use of GIS. More emphasis is given to convey geospatial semantics and preserve locations of and relations among objects in their generalization. Generalization of buildings and built-up areas is a significant problem in medium scale topographic datasets. A constraint-based approach can be adopted to find more effective solution for this problem. In this paper, first conceptual levels in a geospatial model which form the basis for generalization constraints are recommended, then constraint concept is described and possible constraints for buildings and built-up areas on micro, meso and macro levels are given. Finally some efforts and strategies are mentioned for generalization of these object types in geospatial databases.

1. INTRODUCTION

The automation of the generalization process requires very flexible/adaptive methodology that should make decisions based on geospatial semantics and structure, and not simply the geometry of the object. This requires both detailed spatial analytical techniques and constraint analysis in order to select and prioritise different generalization methods, and to select among resulting candidate solutions [*Ruas*, 1998].

Efficiently modelling of geospatial information is very important for conveying this information consistently to next resolution/scale levels. Geospatial (geographic) information is usually in an implicit form and has to be extracted from complex properties of objects and organizations of groups of objects. Besides, it is difficult to formalize this information.

Formalization of geospatial information encompasses three different issues [*Ruas*, 1998]:

- With respect to application, what are the significant relations and meaning that exist among the objects? i.e. what needs to be characterised (measured)?
- How do we identify and organise such information in a GIS in order to support effective generalization of geospatial databases?
- How do we use the results of characterisation (computed measures)?

Geography relies not only on the location of features but also on their defining characteristics and relationships. When a cartographer generalizes a map, (s)he analyses object shape, orientation, proximity between two objects, alignment, global spatial distribution and also semantic information such as rivers, or urban centre area etc. (S)he also compares between all geospatial phenomena such as towns to ensure homogenous treatment. (S)he never restricts his/her view and analysis to the position of object. If we want to automate the generalization process we need to represent different views of geographic space [*Ruas*, 1998].

More systematic framework can be obtained that may support the efforts for the formalization of geospatial information if its elements and modelling are considered.

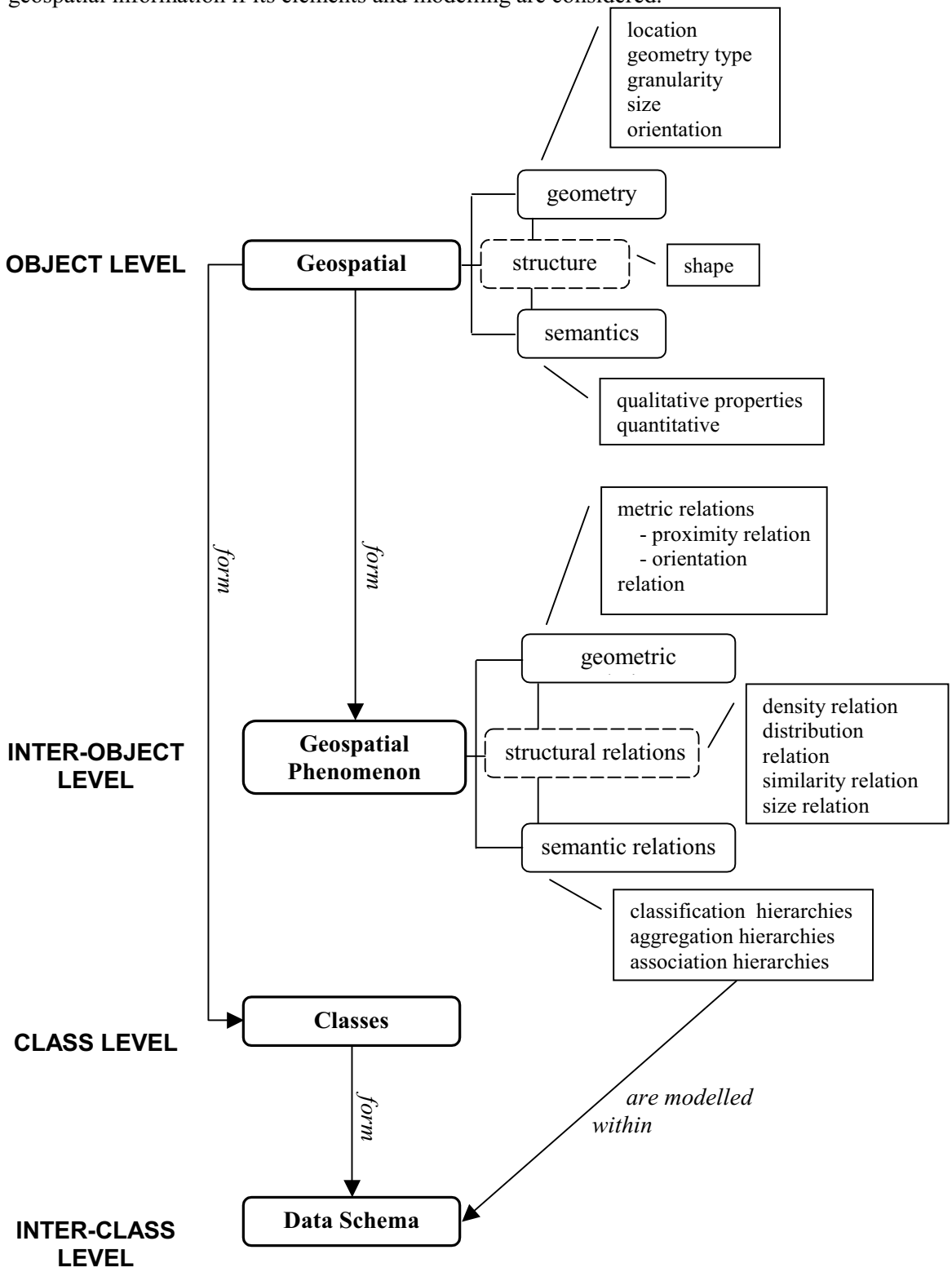


Figure 1: Different conceptual levels in a geospatial model.

Geospatial data generally consists of geometric, semantic and (implicit) structural data (temporal data is out of scope of this paper). Relations among geospatial objects can be geometric relations, semantic relations and structural relations. Structure here means specific organization of geospatial objects individually or as a group and comprises both geometric and semantic properties. Geospatial objects are defined with their geometric, semantic and structural properties and relations among them form various geospatial phenomena. Geospatial objects of common properties and behaviours form classes and classes form data schema.

Different conceptual levels can be formed in a geospatial model by making use of the definitions given above (Figure 1). *Object level* is concerned with individual objects, e.g. a building, a road. *Inter-object* level is concerned with the relations among objects forming a phenomenon, e.g. buildings in a building block or in a built-up area. *Class and data schema level* is concerned with object classes and links among them, e.g. building and road classes and their super and sub classes.

For generalization of buildings and built-up areas, analysis of characteristics such as density, proximity, object sizes are necessary to take appropriate generalization decision. It means that the area composed of buildings and surrounding by streets is a specific view of the geographical space which is called situation (context) [*Ruas*, 1998].

In order to design and develop good algorithms that can actually meet the purpose for which they were conceived it is necessary to carefully and completely define the requirements and criteria that govern the generalization problem at hand. This has lead to the notion of generalization constraints and constraint-based generalization algorithms [*Weibel and Dutton*, 1998].

2. CONSTRAINTS IN GENERALIZATION

Constraints are regarded as design specifications to which solutions should adhere, helping to specify the nature of cartographic or geospatial data products, as well as relations among techniques (algorithms, strategies, etc.) that are necessary to derive them [*Weibel and Dutton*, 1998]. It can be used as triggering and evaluating mechanism for generalization. A constraint describes explicitly either which information does not meet the generalization specifications or which information should be maintained in terms of the content during the generalization process [*Ruas*, 1998].

Different approaches are proposed for constraints in [*Beard*, 1991; *Weibel and Dutton*, 1998; *Ruas*, 1998]. *Ruas* [1998] has proposed micro (local), meso (intermediate) and macro (global) constraints, which are usually interpreted differently depending on the class. Below is given a general list for possible constraints in view of geospatial databases without claiming that it is complete and can fully be implemented:

Micro constraints occur at individual object level. These might be:

- *Geometric constraints* (locational constraint, granularity constraint, size constraint, orientation constraint)
- *Semantic constraints* (qualitative property constraint(s), quantitative property constraint(s))
- *Structural constraints* (shape constraints)

Meso constraints occur within a set of objects which collectively represent a geospatial phenomenon. For example a street network, an urban district, a group of buildings [*Ruas*, 1998]. These might be:

- *Geometric relation constraints* (metric relation constraints (proximity relation constraint, orientation relation constraint), topologic relation constraints)

- *Semantic relation constraints* (classification hierarchy constraints, aggregation hierarchy constraints, association hierarchy constraints)
- *Structural relation constraints* (density relation, distribution relation, similarity relation, size relation)

Macro constraints occur at class and data schema levels. These are the constraints mainly for quantity control and consistency of all objects which belong to one type, e.g. all buildings, all roads. These might be:

- *Class constraints* (quantity constraints, ordering constraints)
- *Data schema constraints* (consistency constraints)

3. CONSTRAINTS FOR BUILDINGS AND BUILTUP AREAS

Micro, meso and macro constraints for buildings and built-up areas in topographic geospatial databases might be as follows [*AGENT Consortium*, 1998; *Ruas*, 1998; *Weibel and Dutton*, 1998]:

3.1 Micro constraints

- **Location**
  Buildings should preserve their initial position.
- **Geometry type**
  Geometry types of buildings can be area or point depending on conceptual model.
- **Granularity***
  Internal outline segments should be of distinguishable dimensions.
- **Size***
  a. Building size should not be smaller than the minimum area threshold.
  b. Building width should be large enough to be distinguishable.
- **Orientation***
  A building should preserve its initial main orientation.
- **Functionality**
  a. Important buildings (according to database purposes) should be maintained.
  b. Isolated buildings can be important buildings as reference objects.
- **Shape***
  a. Buildings should preserve their main orthogonality.
  b. Buildings should maintain their elongation to be recognisable.
  c. If buildings have unusual shapes such as curves, these shapes should be maintained.

3.2 Meso Constraints

- **Proximity**
  a. Smallest distance between two buildings must not be smaller than separation distance.
  b. Relative distance between two buildings should be maintained.
  c. Relative distances between a set of generalized buildings should present the same clustering effect as was initially measured.
  d. Specific patterns and alignments (e.g., along curves) should be maintained.
  e. No specific distribution should be created if it does not exist initially (e.g., if buildings are not organised along a grid, they should not be after generalization).

- **Topology**
  Existing connectivity, adjacent and inclusion relationships must be maintained if buildings are maintained.
- **Orientation***
  Near, neighbouring buildings should try to preserve their relative main orientation.
- **Density**
  a. Groups of buildings should not have too great density.
  b. A group of buildings having too great density can be aggregated/amalgamated or collapsed as a built-up area if it corresponds to the spatial pattern of that portion of a town.
- **Size distribution***
  a. Relative size order of buildings should be preserved whenever their meaning is the same.
  b. Two buildings which have nearly the same sizes can have equal final size after simplification.
- **Shape distribution***
  Within a group of buildings, if some are removed, subset of buildings should preserve buildings which have remarkable shapes (remarkable inside the group); this implies choosing different forms of simplification.
- **Semantics**
  a. Two buildings can be amalgamated together if they do not have contrasting semantic properties.
  b. The distribution of different sub-classes of buildings within an area can change but proportion should try to be maintained. Some indication of semantic exceptions should be maintained (e.g. a commercial area within a neighbourhood of detached houses).

3.3 Macro Constraints

- **Quantity**
  According to database purposes, the area occupied by specific buildings can be maintained.
- **Density distribution**
  a. The ordering of density values among meso objects prior to and after generalization should not greatly change.
  b. Differentiation of densities should be still same after generalization.
- **Shape of districts**
  Globally, districts should have the same kind of shapes (elongation, convexity) as they initially did even if their sizes increase (e.g., resulting from street removal).
- **Size of districts**
  Globally, districts should maintain a difference of sizes according to initial conditions. That is, amalgamation should not alter the relative sizes out of proportion.
- **Semantic distribution**
  If a town has a specific distribution of semantic districts, it should be maintained (town centre, detached houses areas, commercial area, long-building area …).

3.4 *Meso constraints* **(***Street-buildings)*

- **Topology**
  Existing inclusion relationships between a building and a street partition must be maintained (i.e., buildings must not move across streets).
- **Orientation**
  Close street and buildings should try to preserve their relative main orientation.

* only for buildings of area geometry type

* only for buildings of area geometry type

- **Proximity**
  - **a.** Relative distance between a street and a building must be greater than separation distance unless specifically authorised.
  - **b.** Relative distance between a street and a building must be as close as it was initially
- **Functional dependency**
  - **1.** Buildings should be removed/amalgamated if their streets are eliminated.
  - **2.** Generalization of street networks should allow for access to 'important' buildings.

3. GENERALIZATION OF BUILDINGS AND BUILT-UP AREAS IN TOPOGRAPHIC GEOSPATIAL DATABASES

Model (database) generalization creates a dataset at lower resolution from higher one. In the process the amount of data is reduced but the true position of geometry is kept [*Illert*, 2001]. The requirement and some efforts for model generalization of buildings and built-up areas in deriving less detailed topographic geospatial datasets are given in [*Kilpelainen*, 1997; *Anders et al.*, 1999].

Müller [1990] analysed topographic map series and found some facts about buildings and built-up areas, given in Table 1 and Table 2.

Table 1. Size changes for some objects [*Müller*, 1990].

Scale	Streets	Building	Built-up areas
1 : 5 000	no change	no change	no change
1 : 25 000	× 2 - × 4	little change	no change
1 : 50 000	× 4 - × 8	× 1.5 - × 2	× 1.2
1 : 100 000	× 6 - × 16	× 2 - × 4	× 1.5
1 : 200 000	× 32	× 4 - × 8	× 2

Table 2. Changes in number of objects in built-up areas [*Müller*, 1990].

Scale	Dense built-up areas	Scattered built-up areas
1 : 5 000	No change	no change
1 : 25 000	% 60-80 preserved	no change
1 : 50 000	% 30-40 preserved	% 80 preserved
1 : 100 000	% 10 amalgamated in blocks	% 30-50 preserved
1 : 200 000	% 0-3 amalgamated in blocks	% 0-10 preserved

These tables show that at medium scale maps sizes of objects are increased, while their amounts are decreased. However, in model generalization, sizes of objects are not increased while their amounts are decreased. Namely, exaggeration and displacement are not used to be able to preserve the locations of objects.

According to [*Kilpelainen*, 1997], objects have no overlapping conflicts in the model generalization phase, since they have real geospatial (geographic) extent and do not appear in graphic space. Objects do not have to fight for space. The difference between a geospatial database and a cartographic database or a map is schematically shown in Figure 2. A building and a road in a geospatial database at a certain resolution are represented by a point and a line respectively, whereas in cartographic database at corresponding scale by a rectangular point symbol and a line symbol. To render the cartographic representation satisfactory, simplification, smoothing, exaggeration and displacement operations must be performed, which will violate locational accuracy to some extent.

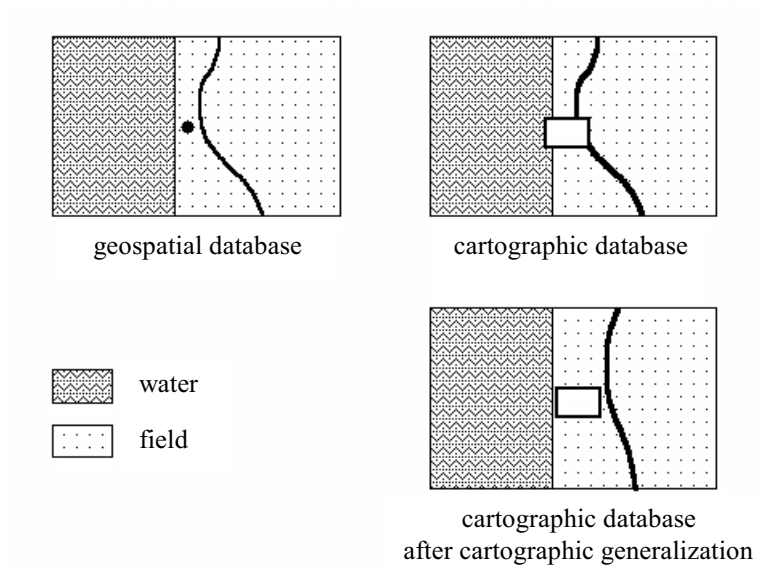


Figure 2: Difference between a geospatial database and a cartographic database [Slightly modified from *Kilpelainen*, 1997].

Geospatial objects have maximal detail at the base level and their representation changes from level to level (Figure 3).

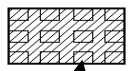
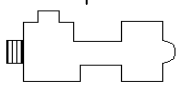
Resolution Levels	**Semantics**	**Abstraction processes**	**Geospatial database representation**	**Geometry**
Level 4	Built-up area	Aggregate / amalgamate the buildings from level 3		Polygon
Level 3	Building	Collapse the building at level 2 to point		Point
Level 2	Building	Simplify the outline of a building at level 1		Simple Polygon
Level 1 (Base Level)	Building	Use the base level representation		Complex Polygon

Figure 3: A building object at different resolution levels in a geospatial database [Slightly modified from *Kilpelainen*, 1997].

4. CONSTRAINT-BASED GENERALIZATION OF BUILDINGS AND BUILTUP AREAS

Constraint-based strategy has particularly gained importance after AGENT project [*AGENT Consortium*, 1998]. A general framework for model (database) generalization based on constraints is given in [*Başaraner*, 2002].

For contextual generalization operations we need objects at different levels: micro objects (individual objects), meso objects (groups of objects) and macro objects (population of objects). They do not perform the same role during the generalisation process, but all are constrained [*Ruas*, 1999]:

- Micro objects generalise themselves when they are autonomous or accomplish orders given by meso objects for contextual operations.
- Meso objects generalise themselves when they perform contextual operations. They also give information to their component objects in order to maintain some group properties.
- Macro objects give information to meso and micro objects to control quantity evolution.

Ruas [1998] gives the attributes enriching the database for more intelligent generalization decisions based on constraints (Table 1).

Table 1. Necessary attributes to govern constraints within the generalization process [*Ruas*, 1998].

Attributes	Uses
Initial and on-going values	Characterisation of constraints and its evolution
Threshold values	Values under or over which a conflict occur
Severity (Non satisfaction)	Evaluation of constraint severity according to thresholds
Priority	Level of priority according to constraint type and severity
Prioritised solving method list	List of advised generalization methods to solve the conflicts
Restrictive methods list	List of non advised generalization methods to maintain a specific information

Whenever a phenomenon studied, a set of constraint objects is created. Each constraint object assesses its severity (non satisfaction) according to 1/its generic thereshold values, 2/ its attributes, by means of evaluation method and proposes a set of prioritised solving methods. Thus the phenomenon or the geospatial object receives a set of proposals and chooses the best according to 1/ constraints priority, 2/ some coherence and procedural rules. Then the phenomenon tries the best generalization method and assesses how the severity of constraints changes. If the constraints are relaxed, then further techniques are applied, if not the system backtracks and the phenomenon tries another solving method. The aim of a phenomenon is to find a local minimum of constraint severity values [*Ruas*, 1998].

A general strategy might be as follows for constraint-based generalization of buildings and built-up areas [*Ruas*, 1998; *Ruas*, 1999]:

Macro objects are first created to control quantity. Some measures like Radical Law might be used to evaluate changes in quantity of objects in classes during generalization. Meso objects are small districts surrounded by streets, and composed of built-up area information, mainly buildings (administrative, industrial, housing etc.). Micro objects are individual objects, i.e. building and surrounding streets constituting a built-up area situation.

Whenever a built-up area-situation is created, it creates some meso constraint objects related to built-up area situation (Figure 4). It means that built-up area situation class owns a built-up area constraint creation method.

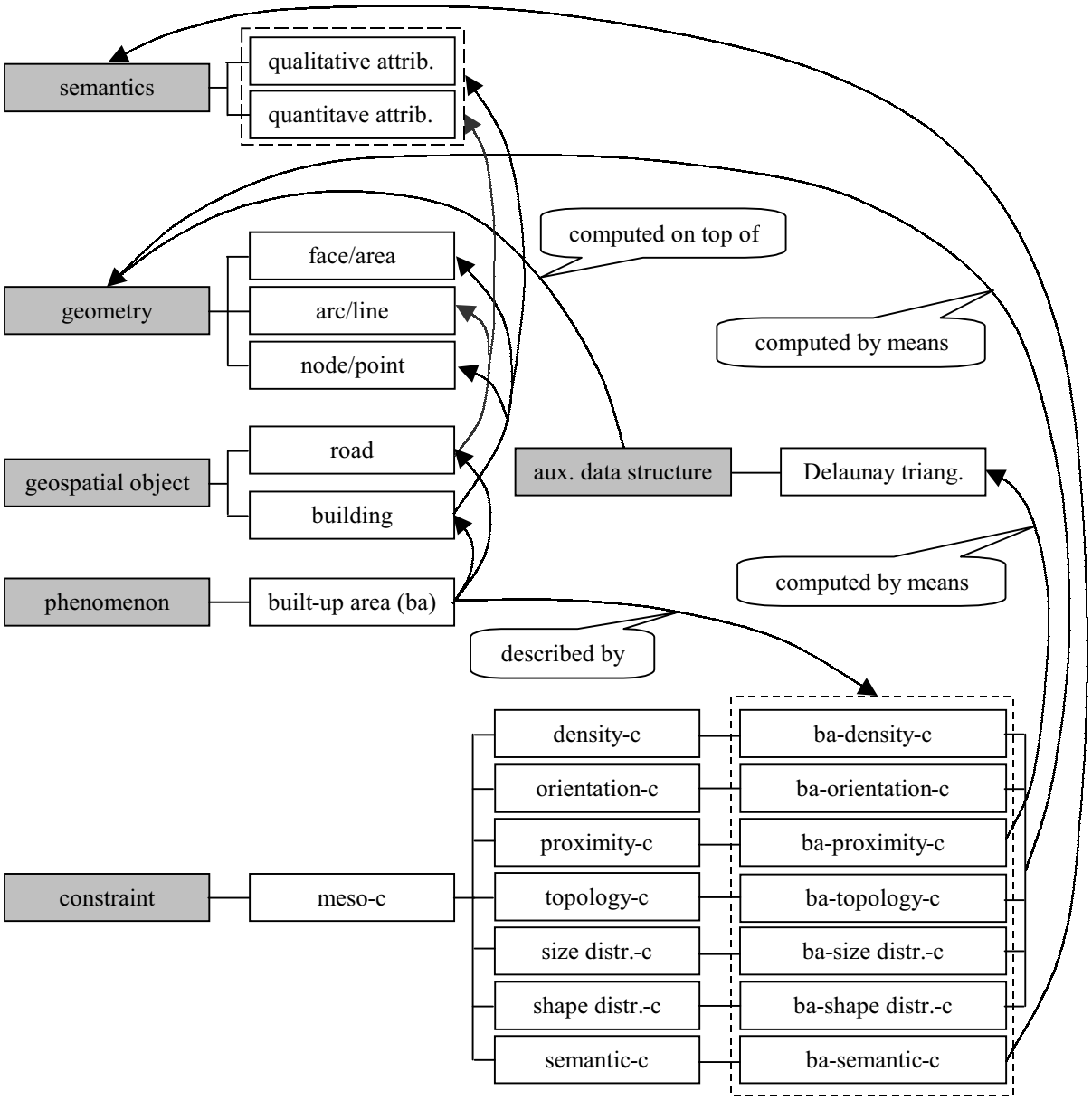


Figure 4: An example of object-oriented constraint modelling applied to built-up area situation [Modified from *Ruas*, 1998].

When a meso situation has reached a state that it considers as being correct, the micro generalization can begin. Then, micro constraint objects can be created. Three cases occur: 1/ The micro objects have powerful interaction capacity to negotiate their order, 2/ The dependency between micro situations is low, so the order of autonomy has no effect on the result, 3/ The micro objects can not communicate easily and their dependency is important. In such a case, the meso situation uses its overview to look for an appropriate order of autonomy: it coordinates the autonomy of its components.

The role of a meso situation is to analyse whether an autonomy order can be found. This order should be a function of 1/ object importance and 2/ object flexibility to support others generalisation. For example an important building surrounded by other buildings should be generalised first. By seeking for an order of autonomy, the meso situation coordinates its components generalisation. After micro

situation generalisations, the meso situation controls that its constraints are still respected and every micro situation found its own solution. If the meso situation is unsatisfied, it can: 1/ carry on its own generalisation, 2/ change the autonomy order, 3/ propose solutions such as constraint relaxation for less important characteristics.

5. CONCLUSIONS

In this paper, a constraint-based strategy is proposed for model (database) generalization of buildings and built-up areas. Constraints can be used in model generalization for meeting the data model specifications. A framework is established for easier understanding and more comprehensive definition of the constraints. Then, a list of constraints and possible strategy are given for buildings and built-up areas considering characteristics of geospatial databases. Topographic geospatial databases involve different strategies in their generalization. Since they are first used for analysis, geospatial information should be preserved at different resolution levels to able to obtain correct analysis results. Regarding these issues and, user and application requirements, next stage will be the development of a conceptual model for these objects at various resolution levels. Then it can be implemented in an object-oriented database environment. After that constraint-based model generalization strategy will be tried to derive lower resolution datasets. After tests, theoritical framework will be clearer. Further studies are needed to find effective solutions for formalization of geospatial information and their use in generalization of geospatial databases.

Acknowledgements

This research has been supported by Yildiz Technical University Scientific Projects Coordination Department. Project Number: 22-05-03-01.

REFERENCES

AGENT Consortium, “Constraint Analysis”, Project Report D A2, ESPRIT/LTR/24 939 (1998).

Anders, K.-H., M.Sester, D.Fritsch, “Analysis of Settlement Structures by Graph-based Clustering”, In: Förstner, W., C.-E. Liedtke, J. Bückner (eds.) SMATI’99, Munich, pp. 41-50 (1999).

Başaraner, M., “Model Generalization in GIS”, In: Proceedings of International Symposium on GIS, September 23-26, Istanbul (2002) (in press).

Beard, M.K., “Constraints on Rule Formation”. In: Buttenfield, B.P. and R.B. McMaster (eds.), Map Generalization: Making Rules for Knowledge Representation, London: Longman, pp.121-135 (1991).

Illert, A., “Multi-Scale GIS: From Generalisation to Data Integration”, In: Proceedings of the 16th ICA International Cartographic Conference, Beijing (2001).

Kilpelainen, T., Multiple Representation and Generalization of Geo-Databases for Topographic Maps. Doctorate Thesis, Publications of the Finnish Geodetic Institute, No: 124, 229 p. (1997).

Müller, J.-C. Rule Based Generalization:Potentials and Improvements”, In: Proceedings of 4th International Symposium on Spatial Data Handling, vol. 1, pp. 317-334 (1990).

Ruas, A., “O-O Constraints Modelling to Automate Urban Generalisation Process”, In: Proceedings of SDH’98, Vancouver, pp. 225-235 (1998).

Ruas, A., “The Role of Meso Level for Urban Generalisation”, Workshop on “Progress in Automated Map Generalization”, ICA’99, Ottawa (1999).

Weibel R. and G. Dutton, “Constraints-based Automated Map Generalization”, In: Proceedings of SDH’98, Vancouver, pp. 214-224 (1998).