Relevant Space Partitioning for Collaborative Generalisation

Guillaume Touya

name.surname@ign.fr

Laboratoire COGIT
Institut Géographique National
73 avenue de Paris 94160 Saint-Mandé – France

1. Introduction

Past and current research in cartographic generalisation led to the development of many automatic processes. Nevertheless, each one is completely efficient only on a sample of the problems raised by cartographic generalisation: a specific landscape (e.g. cities, rural areas, mountains, etc.), a specific theme (e.g. roads, land use, etc.), specific conflicts (road coalescence, etc.) or a mix of the three previous cases (Touya, 2008).

In order to generalise data covering different themes, landscapes and conflicts, it would be interesting to make complementary existing processes collaborate to gather their skills as suggested by (Regnauld, 2007; Duchêne and Gaffuri, 2008). Thus, we propose an approach to divide the initial data into geographic spaces relevant for generalisation by some automatic processes and to generalise each space by the best suited process. We call this generalisation model *Collaborative Generalisation*. Collaborative Generalisation requires a precise definition of these geographic spaces and partitioning methods. This paper deals with the issue of the relevant geographic spaces within a collaborative generalisation framework.

The next part briefly describes the principles of collaborative generalisation. The third part deals with the definition and modelling of a geographic space and its neighbourhood. The fourth part deals with the automatic creation of the spaces and with the issue of the best limits for a given space. Finally, the last part draws some conclusions and proposes further research topics.

2. Principles of Collaborative Generalisation

(McMaster & Shea, 1988; Brassel & Weibel, 1988) specified that automatic generalization processes allowed to answer how, where and when to apply generalisation operators. Many processes now follow this principle and Collaborative Generalisation aims at answering these questions at a higher level: how, where and when to apply generalisation processes. Figure 1 shows in broad outline the principles of Collaborative Generalisation. Automatic generalisation processes are iteratively applied on the appropriate part of the space while side effects are monitored between each application. Thus, a collaborative generalisation can be seen afterwards as a sequence of Space/Process pairs (Touya, 2008). *Collaboration* is used here in the sense of multi-agent system: processes collaborate to reach a common goal, the generalisation of the whole map, and share a common environment as

Figure 1 shows that space is not divided in a partition and so the parts may overlap or cross one another (the road network on which Process 4 is applied crosses all other parts of Figure 1).

(Touya et al, 2010) defines the principles and the mechanisms of collaborative generalisation in five components and three resources: *partitioning, side effects, translator, sequencing* and *registry* components and *geographic spaces, available processes* and *formalised generalisation knowledge* resources. Resources are the inputs of the approach guiding the generalisation. Components are the acting elements of the approach, using resources as inputs and outputs.

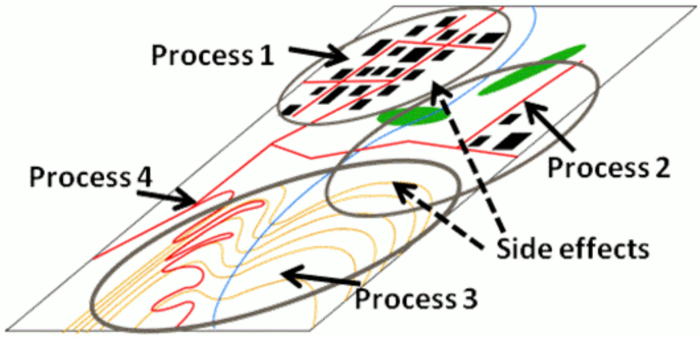

Figure 1. The collaboration principle between generalisation processes. A process 1 is carried out on the town area, a process 2 on the rural area, and then a process 3 on the mountain area and finally a process 4 is carried out on the road network. Side effects are corrected at the neighbourhood (dashed arrows) of application spaces.

The *geographic spaces* are the portions of initial data that are relevant for a generalisation process (e.g. the town relevant for *Process 1*, the road network for *Process 4* in Figure 1). The modelling and definition of the geographic spaces is detailed in section 3 of the paper.

The *available generalisation processes* are the automatic processes that can be executed on the collaborative generalisation platform. For instance, eight automatic processes are currently available on our research platform: an AGENT process (Barrault et al, 2001) customised for urban generalisation, a CartACom process (Duchêne, 2004), a least squares process (Harrie, 2001) or “Beams” process (Bader et al, 2001) among others.

The *formalised generalisation knowledge* gathers the knowledge necessary to parameterise the collaborative generalisation. It can be divided in four main pieces: a generalisation ontology, a formal description of the available processes, a constraint model to formalise map specifications and general sequencing rules. A generalisation ontology gathers the concepts tackled by cartographic generalisation: topographic concepts like buildings and roads, structural concepts like meso objects or geographic relations and procedural concepts like standards operators. The ontology helps to share a common vocabulary between heterogeneous components and resources in collaborative generalisation. The formal description of the available processes allows to describe the capabilities of an automatic generalisation process in order to use it in the collaborative generalisation. Based on ontology concepts, the constraint model helps to express all map requirements in a standardised way. Finally, general sequencing rules help the collaborative generalisation to move on to the big steps of the generalisation as in the *Global Master Plan* of (Ruas & Plazanet, 1996). See more details on each piece of the formalised knowledge in (Touya et al, 2010).

The *partitioning* component helps to divide space in geographic spaces (see §4). Some pieces of the formalised generalisation knowledge guide the partition component to know kind of geographic spaces have to be generated and how.

The *registry* component helps to optimise the association of a space and a process. It chooses the most appropriate process to generalise a given geographic space according to process descriptions and actual conflicts within the space (Touya et al, 2010).

The *translator* component allows translating the formalised constraints into parameters specific to a given process. It consists mainly in a translating function associated with every available process. It allows to launch not interoperable processes like AGENT or the least squares with a single set of constraints (Touya et al, 2010).

The *sequencing* component allows automatically chaining the pairs space/process taking into account conflict seriousness and general sequencing rules. It also manages the online evaluation, in a trial and error approach, dealing thus the emerging spaces (see section 3).

The *side effects* component observes the side effects in the spaces neighbourhood and possibly corrects them. As spaces do not constitute an exact partition and may overlap or cross, conflicts may be caused with outside objects or with previously generalised objects. Section 3.2 deals with the definition a neighbourhood for spaces where such side effects are controlled and corrected.

We propose a collaborative generalisation model, called *CollaGen* that implements the components and resources described above. The rest of the paper deals with the modelling and the role of the geographic spaces in the CollaGen model, particularly in the side effects control.

3. Geographic Spaces Modelling

3.1. Definition of the Geographic Spaces

We define a geographic space as a geographically meaningful extract of the data that can be a relevant input for a given generalisation process. The use of geographic spaces in Collaborative Generalisation is useful for both optimising the use of the existing generalisation processes and partitioning the data to avoid the process of very large datasets, which is a well known limit of automatic generalisation processes (Chaudhry and Mackaness, 2008b).

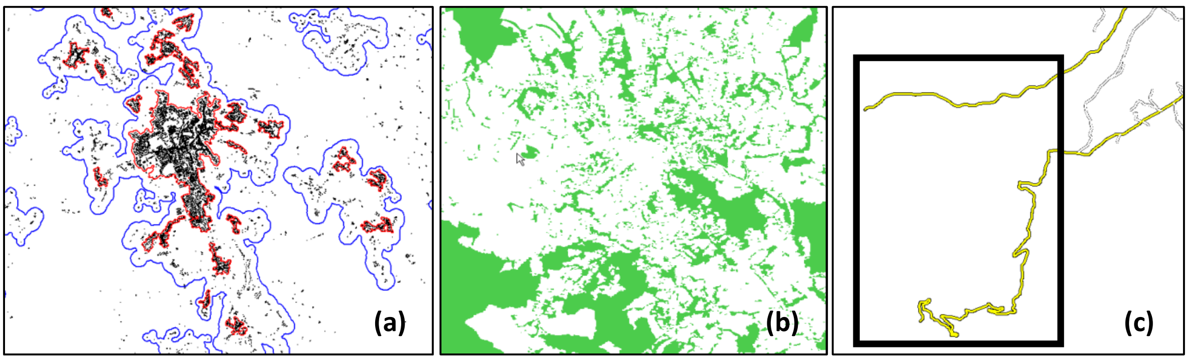

Figure 2. (a) buildings in black, urban areas in red and rurban (Enault, 2004) areas in blue. (b) vegetation thematic space on the same area. (c) mountains roads space (in the black rectangle).

The Geographic Spaces can be divided in three types:

- Metric spaces are delimited by geospatial boundary that include all objects inside the limits (e.g. urban (Figure 2a), rural, mountain or coastal areas).
- Thematic spaces are delimited by geographic theme and include all objects in the initial data that belong to the theme (e.g. road networks, vegetation (Figure 2b) or land use).
- Mixed spaces are both metric and thematic spaces, i.e. metric spaces limited to a (or some) theme(s) (e.g. mountain roads (Figure 2c)).

It can be noticed that with such a definition, the geographic spaces defined in a dataset do not form a mathematical partition as metric spaces can overlap and thematic spaces can cross metric spaces. The Spaces can also be classified according to their role in the collaborative generalisation:

- The *key* spaces are used whatever the available processes are. For instance, urban areas, rural areas and mountain roads are considered as key spaces. The choice of the key spaces is a knowledge input of the collaborative generalisation.
- The *facilitator* spaces are necessary to optimise the use of a given available process. For instance, if a collaborative filtering process (Burghardt & Neun, 2006) is available, the generation of *repetitive spaces* like US residential areas could be required: as such an approach learns online the appropriateness of operations to characterised situations, it would provide optimal results in spaces where many situations share the same characteristics. As another example, the *land use* thematic space is required when a land use aggregation process (Haunert, 2007) is available.
- The *emerging* spaces are sub-spaces where conflicts remain unsolved. During the generalisation of a space by a given process, the conflicts evolution is observed over the process and conflict clusters (close conflicting objects) can emerge as sub-spaces to be generalised by another process than the one processing the whole space. (Duchêne & Touya, 2010) describes the CollaGen implementation of such an emerging approach in relation to comparable emergence in a CartACom process.

The definition of what entity can be considered as a geographic space might be ambiguous. For instance, the flexibility graphs (Lemarié, 2003) can be considered as a facilitator space for the "Elastic Beams" (Bader et al, 2005). The so called flexibility graphs are extracts of the road and river networks that allow to cushion properly the Beams deformations. Flexibility graphs meet the mixed space definition although the graph structure may be understood as a simple auxiliary data structure. To clarify the definition, any extract of the geographic data, limited metrically and/or thematically, that can be a relevant input for a given generalisation process, can be considered as a geographic space.

Figure 3 shows the UML class diagram we have defined to model a generic geographic space. The geographic spaces as Urban areas or Mountain roads extend this model. Geographic Spaces are aggregations of geographic objects and geographic structures with a partonomic relation: the objects and structures are part of the space. In relation to the definition given earlier, geographic spaces are a geographically meaningful extract and thus correspond to a concept in the Generalisation domain ontology (Touya et al, 2010). The next part details the neighbourhood modelling, present in Figure 3.

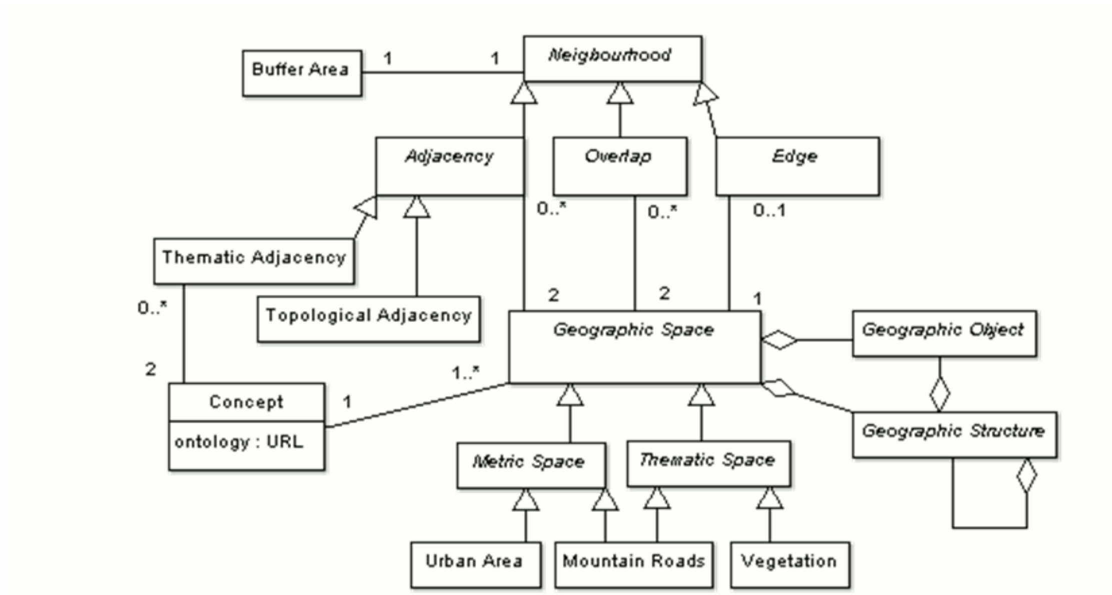

Figure 3. The UM class diagram of a Geographic Space. The three bottom classes are examples of specialised geographic spaces.

3.2.The Neighbourhood of Geographic Spaces

We define the neighbourhood of a geographic space as the part of space where the collaborative generalisation has to look for side effects after the generalisation of this space. Experience shows that side effects can occur in places that belong to several spaces, when some limits are shared by two spaces (occurs frequently as networks are often used to outline the spaces) or outside the spaces when outside objects are not taken into account by the process that generalise the space. Thus, we define the neighbourhood as composed of three parts, the adjacencies, the overlaps and the edge (Figure 4). The adjacency and the overlap correspond to classical topological relationships between polygons (Egenhofer, 1989). The edges are the parts of the outline that do not belong to an adjacency or overlap relationship. The edge is included in the neighbourhood model in order to control potential side effects just the space. In order to define what is just outside or around a neighbourhood component, each of the components is extended by a buffer area (dashed in Figure 4). The buffer area computation depends on the neighbourhood type: for instance, the buffer area of an edge neighbourhood spread only outside the space while the buffer area of an adjacency neighbourhood spreads less but all over the shared limit to find all objects that could be impacted by the movement of the limit.

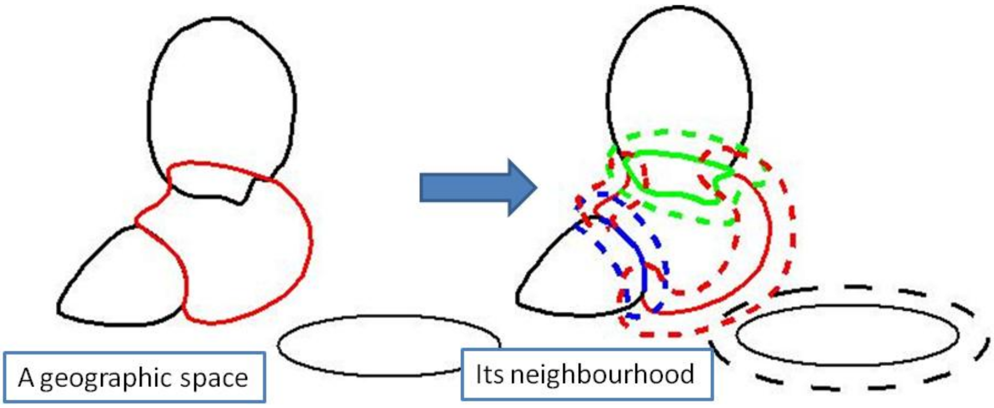

Figure 4. A schematic geographic space in red and its neighbours and the neighbourhood of the space: the edges in red, the adjacencies in blue and the overlaps in green. The buffer areas are represented with dashes.

Figure 5 shows a rural space automatically created with the three kinds of neighbourhood generated: an adjacency neighbourhood with an urban space and a rurban space all around (a space between rural and urban (Enault, 2004)); an overlap neighbourhood with a mountain space and the rurban space and simple edge neighbourhood. Figure 5b also illustrates the need for the use of buffer areas around neighbourhoods: the two rural areas could push buildings outside the area causing conflicts if the areas are to close. The buffer areas allow controlling potential side effects further for better resolution. Thus, we can define a third kind of neighbourhood relationship related to edges, called "close edges", which is shared by close spaces. This relationship can be identified when edge buffer areas are overlapping.

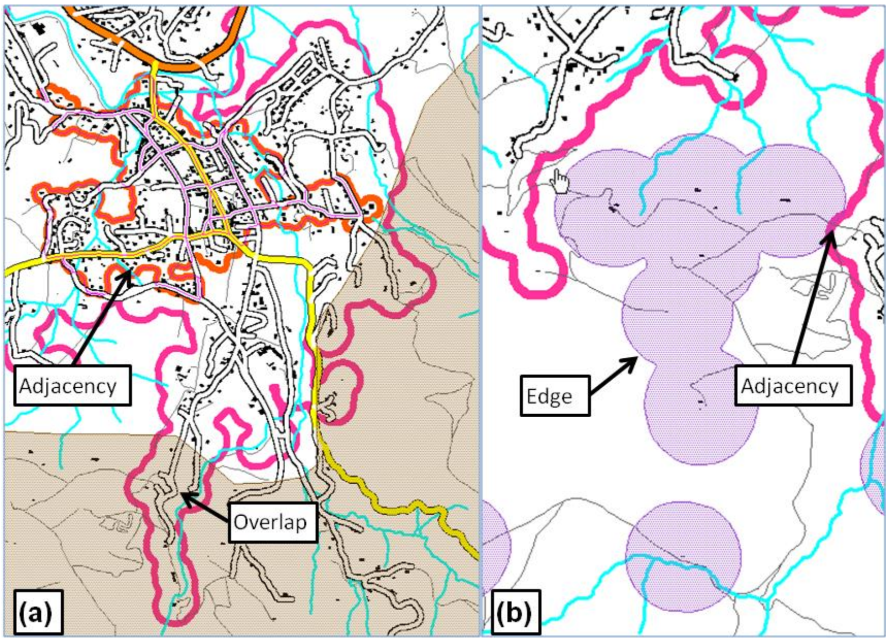

Figure 5. (a) A rural space (outline in purple) extract with the three types of neighbourhood: overlap with the mountain space (brown filled area), adjacency with the rurban area (outline in magenta) and the road network and edges where no space is in the direct neighbourhood. (b) rural spaces (in purple stipple) neighbouring rurban spaces (pink thick outline). The two central rural spaces share a "close edges" kind of neighbourhood.

The class diagram Figure 3 shows that the adjacency neighbourhood can also be specialised in a thematic adjacency. It means that additionally, some spaces can be connected to particular themes of neighbouring spaces. For instance, *forests* are connected to *roads* as forests may edge roads, so a *mountain roads* space is thematically adjacent to the forests and a particular care is taken on the forests edging the mountain roads. This particular thematic connexity and others are described first in the generalisation ontology required in Collaborative Generalisation (Touya et al, 2010) and are then instantiated in the data.

The topology of the thematic spaces is another key issue of space neighbourhood. The difficulty is mainly here to be able to define the outline of spaces that can be point, line, polygon layers (e.g. relief points, road network, vegetation), partitions (land use) or complex layers (electricity networks contain pylons, lines and power plants). Figure 6 shows different existing methods to define the outline of such spaces and the more appropriate one has to be chosen for each of the thematic spaces used. For instance, the union of partitions outlines for a land use space or the extended footprint for a road network seem to be appropriate methods to define specific outlines. Then, edges, adjacencies and overlaps can be identified with other metric or thematic spaces in order to locate possible side effects after the generalisation of the space. Knowing that relief points have adjacencies with roads makes the propagation of road deformations to the adjacent points easier.

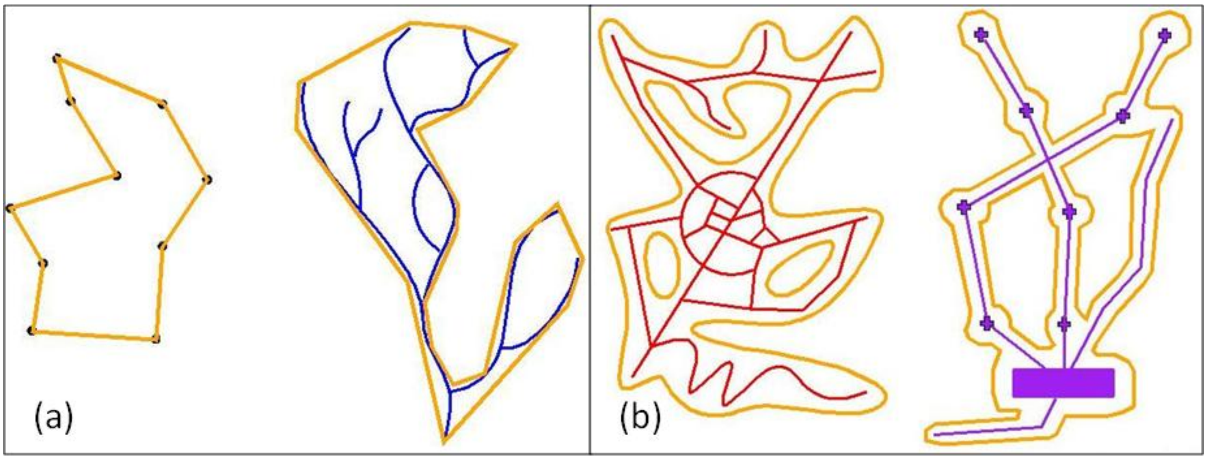

Figure 6. (a) Concave edge method for a point set (Duckham et al, 2008) or a river network (wrapped convex hull). (b) Extended footprint (computed with buffers) method for road and electricity network.

4. Partitioning Space into Geographic Spaces

4.1. Spatial Analysis Algorithms to Create the Spaces

While the thematic spaces can directly be derived from the initial data, the boundaries of the metric and mixed spaces have to be extracted by means of spatial analysis algorithms. For instance, (Boffet, 2000) presents an algorithm that infers urban boundaries from the buildings and their proximities and (Chaudhry & Mackaness, 2008c) presents an algorithm that extracts mountain hierarchical boundaries from contour lines. An algorithm to create the boundaries has to be chosen for each key or facilitator space. We developed some specific algorithms like coastal areas (Figure 6) or rurban areas (Figure 2a) detection in the context of Collaborative Generalisation.

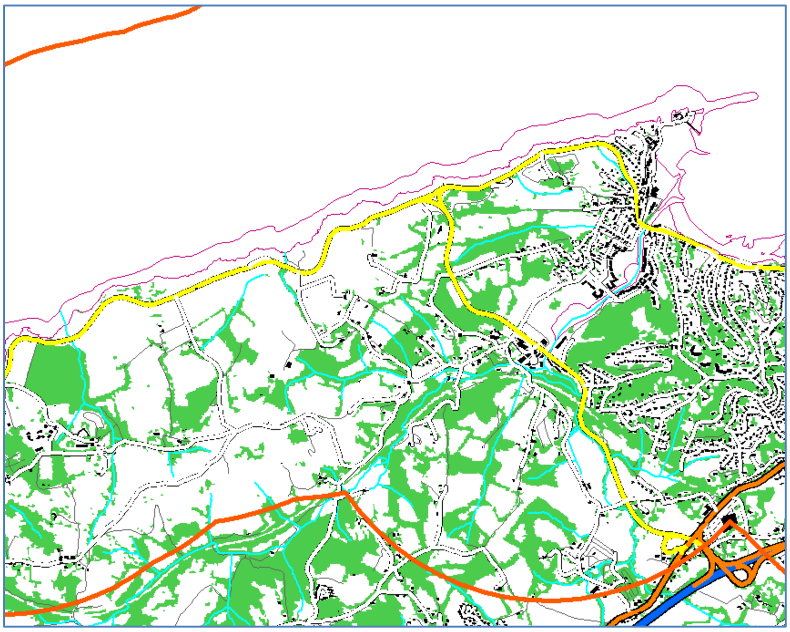

Figure 7. A coastal area (outline in orange) automatically created using coastal lines.

4.2. What Is the Best Outline?

(Chaudhry and Mackaness, 2008a) claim that the metric and mixed geographic spaces considered in this paper, like urban areas, have vague boundaries called *fiat* boundaries (Smith and Varzi, 2000) in opposition to lakes or forests that have *bona-fide* boundaries (i.e. clear physical boundaries caused by spatial discontinuity or qualitative heterogeneity). Although the side effects are controlled and can be corrected, define spaces so as to minimise them would reduce the complexity of the collaborative process. As a consequence, the choice of the boundaries of the space may greatly impact the generalisation of the space and of the objects that can be inside or outside the space depending on the boundaries. Considering the role of the geographic spaces in the collaborative generalisation, the choice of the best boundary is made in two steps: choose the best theoretical boundary (exclude building alignments along the main roads at the limit of urban areas for instance) and then choose the best algorithm to create the boundaries. The best theoretical boundary would be here the one that generally minimises side effects.

Based on the case of urban areas that can be created with several methods (Boffet, 2000; Chaudhry and Mackaness, 2008a; Walter, 2008), a study is being carried out to compare different boundaries for a same space (Figure 8) and assess the sensitivity of generalisation to the choice of the boundary. The idea would be here to guide the partitioning to use the method most appropriate to the available processes for urban areas. For instance, the study showed (Figure 8) that there is more empty space inside the space using (Boffet, 2000) than using (Chaudhry and Mackaness, 2008a). So the second method appears to be more appropriate for the use of an AGENT process that deals well with spaces dense in buildings.

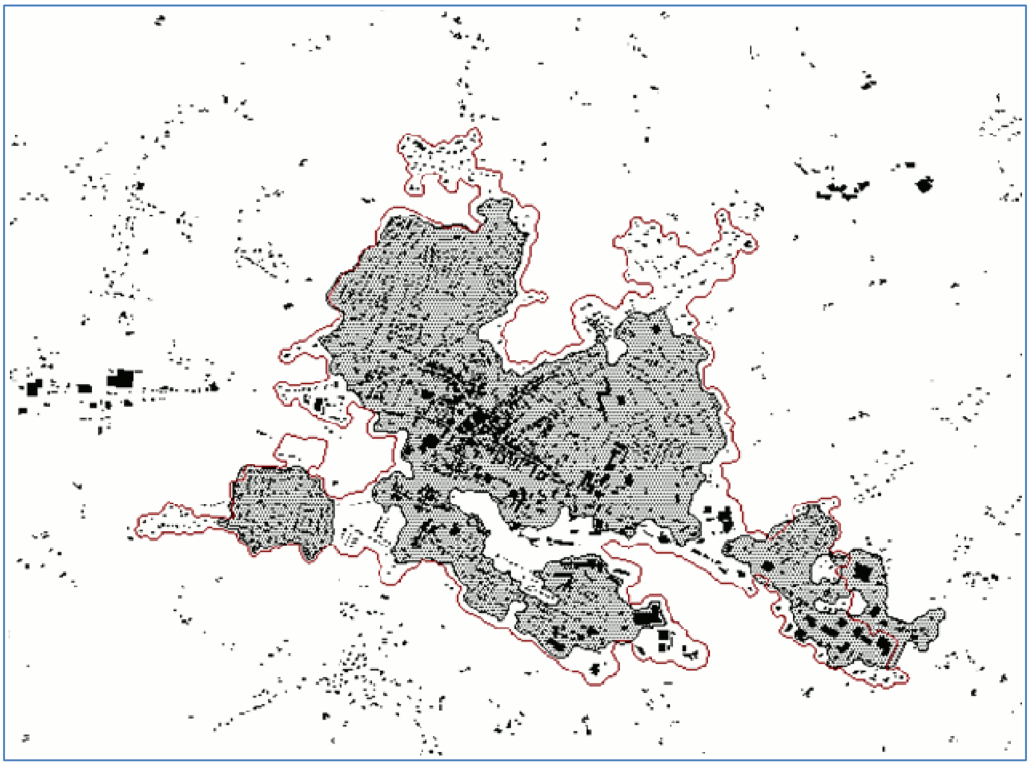

Figure 8. Buildings and urban areas delimited with two methods: building proximity in red (Boffet, 2000) and citiness in grey (Chaudhry and Mackaness, 2008a). The second method allows more buildings just outside the boundary.

5. Conclusion and Further Work

In this paper, we presented the issue of partitioning data into geographic spaces relevant for a particular generalisation process in the context of Collaborative Generalisation. The paper introduces the concept of geographic space for generalisation and its neighbourhood, notably the issue of thematic neighbourhood. The paper also discusses how these spaces can be created, and how to define their boundaries.

To go further, the sensibility tests have to be finalised to really assess the sensibility to the boundaries. Moreover, side effects observation and correction tests have to be carried out to validate the proposed topological and thematic neighbourhood modelling.

References

Bader, M., Barrault, M., Weibel, R., 2005. Building displacement over a ductile truss. International Journal of Geographical Information Science 19 (8), 915-936.

Barrault, M., Regnaud, N., Duchêne, C., Haire, K., Baeijs, C., Hardy, P., Mackaness, W., Ruas, A., Weibel, R., 2001. Integrating multi-agent, object oriented and algorithmic techniques for improved automated map generalization. In: 20th International Cartographic Conference, Beijing, China.

Boffet, A., 2000. Creating urban information for cartographic generalisation. In: International Symposium on Spatial Data Handling (SDH).

Brassel, K. E., Weibel, R., 1988. A review and conceptual framework of automated map generalization. International Journal of Geographical Information Science 2 (3), pp. 229-244.

Burghardt, D., Neun, M., 2006. Automated sequencing of generalisation services based on collaborative filtering. In: M. Raubal, H. J. Miller, A. U. Frank, and M. F. Goodchild (Eds.), Geographic Information Science - 4th International Conference GIScience, IFGI prints, pp. 41-46. Münster, Germany.

Chaudhry, O. Z., Mackaness, W. A., March 2008a. Automatic identification of urban settlement boundaries for multiple representation databases. Computers, Environment and Urban Systems 32 (2), 95-109.

Chaudhry, O. Z., Mackaness, W. A., 2008b. Partitioning techniques to make manageable the generalisation of national spatial datasets. In: ICA Workshop on Generalisation and Multiple Representation. Montpellier, France.

Chaudhry, O. Z., Mackaness, W. A., October 2008c. Creating mountains out of mole hills: Automatic identification of hills and ranges using morphometric analysis. Transactions in GIS 12 (5), 567-589.

Duchêne, C., 2004. The cartacom model: a generalisation model for taking relational constraints into account. In: 7th ICA Workshop on Generalisation and Multiple Representation, Leicester, UK. ICA.

Duchêne, C., Gaffuri, J., 2008. Combining Three Multi-agent Based Generalisation Models: AGENT, CartACom and GAEL. In: Ruas, A., Gold, C. (eds.) Headway in Spatial Data Handling, 13th International Symposium on Spatial Data Handling. LNG&C, pp. 277--296. Springer, Heidelberg (2008)

Duchêne, C., Touya, G., 2010. Emergence de zones de conflits dans deux modèles de généralisation cartographique multi-agents. In : Journées Françaises des Systèmes Multi-Agents, Mahdia, Tunisia.

Duckham, M., Kulik, L., Worboys, M., Galton, A., 2008. Efficient generation of simple polygons for characterizing the shape of a set of points in the plane. Pattern Recognition 41 (10), pp. 3224-3236.

Egenhofer, M., 1989. A Formal Definition of Binary Topological Relationships. In: W. Litwin and H. Schek (eds.), Third International Conference on Foundations of Data Organization and Algorithms (FODO), Paris, France, Lecture Notes in Computer Science, Vol. 367, Springer-Verlag, pp. 457-472, June 1989.

Enault, C., 2004. La dilution : note méthodologique pour l'analyse de l'étalement urbain. *L'Espace Géographique 33* (3), 241-255

Haunert, J.-H., 2007. Efficient area aggregation by combination of different techniques. In: 10th ICA Workshop on Generalisation and Multiple Representation.

Lemarié, C., 2003. Generalisation process for top100: research in generalisation brought to fruition. In: 5th ICA Workshop on progress in automated map generalisation. ICA, Paris, France.

McMaster, R. B., Shea, K. S., 1988. Cartographic generalization in digital environment: A framework for implementation in a gis. In GIS/LIS'88, pp. 240-249.

Regnauld, N., 2007.Evolving from automating existing map production systems to producing maps on demand automatically. In: 10th ICA Workshop on Generalisation and Multiple Representation (2007)

Ruas, A., Plazanet, C., 1996. Strategies for automated generalization. In: 7th International Symposium on Spatial Data Handling, Delft, Netherlands, pp. 319-336.

Smith, B., and A. C. Varzi, 2000. Fiat and Bona Fide Boundaries. Philosophy And Phenomenological Research 60, 401-420.

Touya, G., 2008. First thoughts for the orchestration of generalisation methods on heterogeneous landscapes. In: workshop on generalisation and multiple representation. Montpellier, France.

Touya, G., Duchêne, C., Ruas, A., 2010. Collaborative Generalisation: Formalisation of Cartographic Knowledge to Orchestrate Different Generalisation Processes. In: GIScience’2010. Zurich. Accepted as full paper.

Walter, V., 2008. Automatic interpretation of vector databases with a raster-based algorithm. In: ISPRS Commission II, WG II/4.