Information Aggregation: Automatized Construction of Chorematic Diagrams

Andreas Reimer and Doris Dransch

24th February 2009

The concept of Digital Cities incorporates the need for communicating geographic information. Chorematic diagrams are a special method for communicating geographic information, especially for commucating highly synthesized and aggregated concepts, that used to enjoy great popularity. This report on a dissertation in progress describes an analytical approach towards automatization of chorematic diagram construction.

1 Information Aggregation and the Digital City

There are many definitions, as well as greatly different examples, of the concept of the Digital City. According to KRYSSANOV et al. (2002), a Digital City can be generally defined as a collection of digital products and information resources made of a large distributed database of heterogenous documents of various digital genres.

Most real-world examples only realize a portion of the functionalties many definitions describe: from Amsterdam´s *Digitale Stad*, as example of an internet-based community of real world city residents (VAN DEN BESSELAAR and BECKERS, 2005), to Leica Geosystem´s *Digital City Kunming*, which is a network of GPS-based hardware elements that provide city-wide precise locational data. Even a three dimensional graphical computer model of a city might be called Digital City. So, to a degree, the concept is more of a vision than fully realized reality.

No matter which particular view is taken on the concept, communication of data and information to human end-users is a fundamental task. Although the actual communication tasks and problems are system-specific, one can identify recurring ones. KRYSSANOV et al. (2002), for example, identify information retrieval based on user queries to be a key issue. The Digital City must enable the user to interpret his query results in an uncomplicated and correct manner. Consequently, the use of illustrations, metaphors and analogies for the display of query results is proposed. As the data space of a Digital City is inherently related to a specific geographic area, many information needs will also include a geospatial dimension, best expressed with maps and similiar Geovisualization techniques. Following this argumentation, the Digital City needs uncomplicated

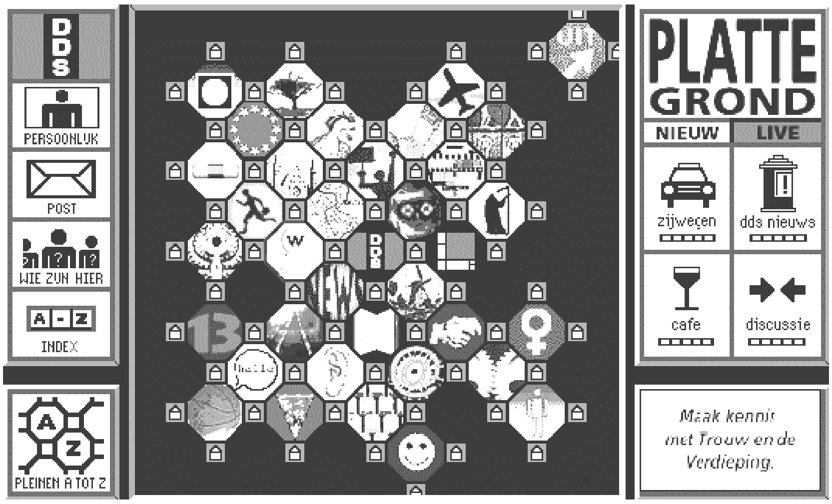

Figure 1: Abstraction and Space in Digital Amsterdam (van den Besselaar and Beckers, 2005)

illustrations of data and information with their relation to geographic space, based on metaphors and analogies. Chorematic diagrams are postulated to exactly match those requirements (Del Fatto et al., 2007). If one wants to facilitate the usage of chorematic diagrams based on user queries, one has to automatize their construction. This work is concerned with this automatization.

2 Related Work: Chorematic Diagrams

From a background of spatial analysis, French geographer and cartographer Roger Brunet developed classes for spatial structures and processes along with a specific way of rendering them graphically (Brunet, 1980, 1987). He called them *chorèmes*, a neologism composed of the Greek word χώρα, meaning space, territory, place and the suffix -ème from theoretical linguistics (Brunet, 1987). The term is to be understood as an analogue to words such as morpheme, phoneme or grapheme, which refer to the smallest linguistic (audible and written repectively) units that carry (semantic) meaning. They were popularized by the work done within the Groupe d´Interêt Public (GIP) RECLUS, which was founded in 1984 under Brunet´s leadership (Ormeling, 1992). Their output included many thematic maps, partly or entirely made from chorèmes. Because it were these chorematic diagrams and maps that garnered attention, the term *chorème* became synonymous with a certain style of graphic depiction of geographic space.

Definition

Chorèmes are a tool for the structural and iconic representation of complex geospatial situations. They consist of terms and graphics that largely abstract from actual objects and precise cartographic symbols. Brunet differentiates between seven classes of basic spatial configurations that take different forms for point, line, area or network-based embodiements of these configurations. Each one is linked to an associative and indexical sign for the respective structure or process (Fig. 1). Chorematic diagrams can either

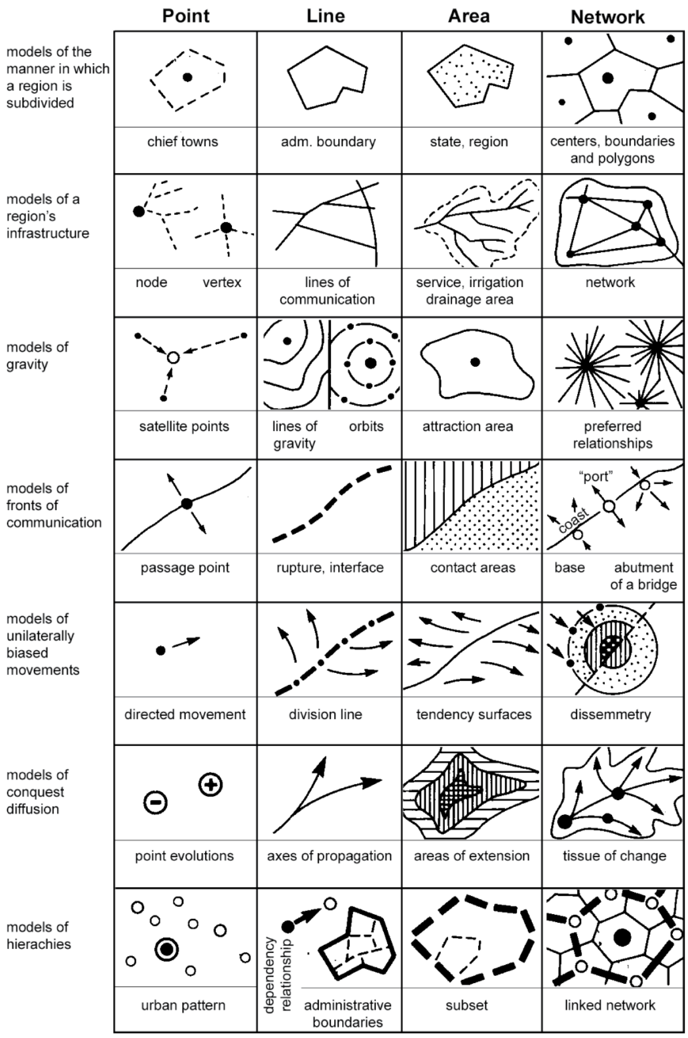

Figure 2: Table of Brunet´s chorèmes with english labels (van Elzakker, 2004)

be maps with additional, highly aggregated layers or map related representations that abstract from the precise topographic and thematic spatial structure. By doing this, the latter enable a high-level comparison of the structures between different regions (abbreviated from Tainz, 2001).

One of the main advantages of chorematic diagrams is that they are in a closer relation to a mental model of the depicted geographic space than a regular thematic map would be. This closer relation goes both ways: a spatial analyst can express his mental model easier with chorematic maps, which in turn is understood easier by a reader. This lead to great successes for Brunet and GIP-RECLUS (Brunet and Dollfus, 1990; Ormeling, 1992), but also highlights the problem of how much trust is actually placed in the analysis that lead to the mental spatial model.

Although *chorèmes* and their proponents lost some of their popularity and nearly all of their institutional backing in the nineties, they did serve as inspiration for several working groups and individuals, mostly European, in more recent times (van Elzakker, 2004; Klippel, 2004; Brunet, 2005; Ligozat, Nowak and Schmitt, 2007; Lardon and Capitane, 2008).

Visual Summary Project

Under the leadership of Robert Laurini from INSI in Lyon, France, an international working group started research on the "automatic generation of chorem maps" in 2006, highlighting the relevance and interest of the subject. Using the existing chorem concept as an inspiration and making a case for the possible benefits, they delineated the problems and presented approaches for tackling them (Laurini, Milleret-Raffort and Lopez, 2006).

Their approach divides the process of automatized construction of chorematic diagrams into an extraction and a visualization subsystem (Fig. 3). From a qualitative review of 50 existing chorematic diagrams the group developed a chorem dictionary. This dictionary consists of the description of available chorems name, type, structure and their visualization (Table 1). The dictionary contains seven chorems, three of which correspond to three of Brunet´s original 28 chorems. Textual annotations are treated as an additional expressive chorematic element. In an iterative observed data mining approach, the user is supposed to distill spatial expressions, divided into phenomena and static situations, which are then mapped to corresponding chorem dictionary entries. In the example given (Migration in Italy 2000), this Spatial Data Mining consisted of a clustering of administrative regions into five macro-regions based on minimal distance, along with the decision to depict the aggregated flow of migrants between these macro-regions. Using ChorML as intermediate format (Rocha Coimbra, 2008), the selected information is passed on to the visualization subsystem. There, the information layers are reduced in geometric complexity via generalization algorithms (a modified Douglas-Peucker algorithm is used in the example) and aggregated. In a further step, the geometries of the base map are "choremized" by reducing the number of vertices in order to reach simpler shapes approaching regular polygons. Arrangement and placement of annotations is automated via a multi-agent approach. The resulting figure is passed on to the chorem editor, where the user can make modifications and change layout parameters and can resolve placement conflicts (Del Fatto et al., 2008).

3 Approach

The aforementioned pioneering work leaves several paths for further research. In the following, a supplemental approach is proposed that can refine as well as expand on

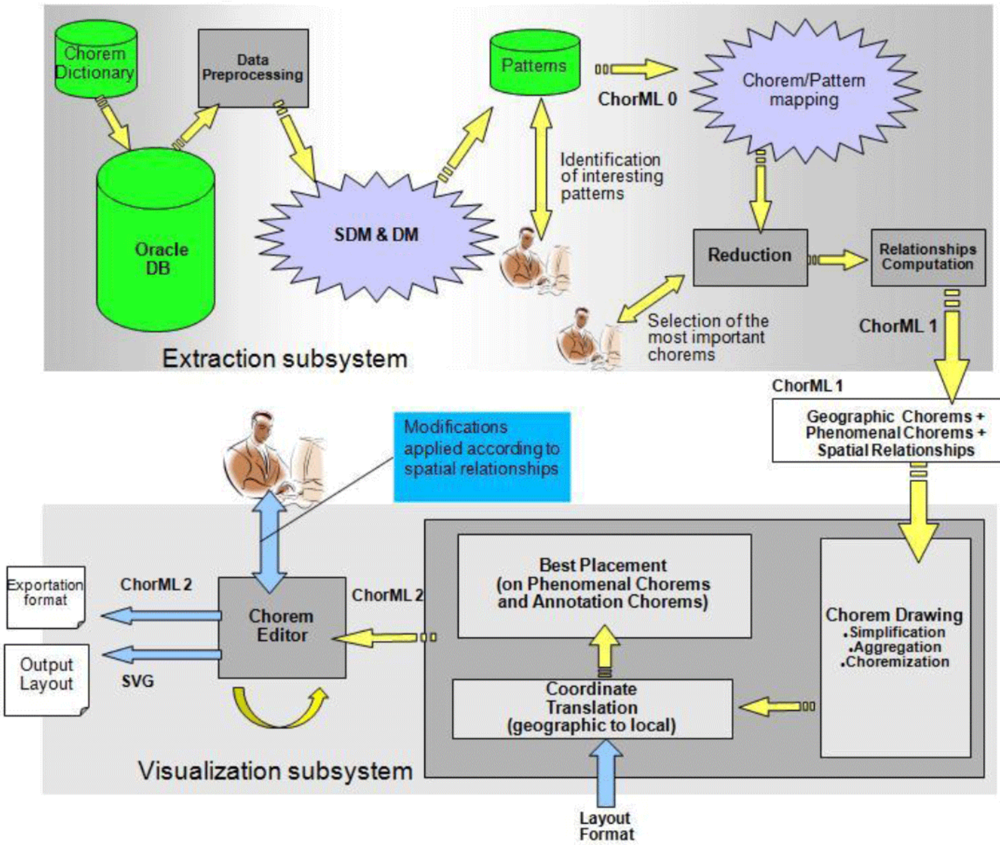

Figure 3: Architecture of *Del Fatto et al.´s* proposed system (Del Fatto et al., 2008)

the current state of automatization of chorematic diagrams. The main focus of this research is the thorough analysis of existing chorematic diagrams. Expert knowledge is valueable for any automatization of tasks that have been done manually. To generate this knowledge for chorematic diagrams and render it into a computable form, is an important part of this approach.

The current level of codified and published knowledge about chorems is rather sparse. The eighties and nineties saw hundreds of chorematic maps, many of which introduced new chorem schemata as well as complex sequences of diagrams. It is unclear if all of these could be expressed with the 7+1 chorems identified in (Del Fatto et al., 2008) (see Table 1). As of now, there is no distinction possible between a chorematic map and a thematic map that only uses the 7+1 visual representations. Currently, heuristic human intervention is needed as a sort of a stop criterion to decide when sufficient "chorem-ness" is reached.

If one had quantitative data on the geometric, graphical and semantic properties

CATEGORY	TYPE
Geographic	Polygon
Geographic	Point
Geographic	Line
Geographic	Network
Phenomenal	Flow
Phenomenal	Tropism
Phenomenal	Diffusion
Annotation	Label

Table 1: The 7+1 types of chorems in the chorem dictionary from DEL FATTO et al. (2008)

of existing and known chorematic diagrams, the need for such intervention would be lowered. Herein such a quantitative approach is outlined.

Geometric Analysis

The first step is the vectorization and annotation of chorematic diagrams that are recognized as such by the community that invented and propagated them, namely GIP-RECLUS and its follow-up organization, projects and former members, as well as adoptors. Once a substantial number of diagrams have been obtained in/transformed into vector format, a quantitative descriptive statistical analysis can be carried out. The first line of analysis will be geometrical, investigating the numerical relationship between the actual topographic situation and the abstracted spatial depictions of the chorematic maps (Fig. 4). This could improve the parametrization of potential conversion and generalization algorithms. Making an informed choice in choosing generalization algorithms themselves is also dependant on a good understanding of the desired end-state (GUILBERT et al., 2008).

Graphical Analysis

The vectorized chorematic diagram data collection will enable other options for analysis. This will be the graphical analysis, which is concerned with the visual variables actually used as well as their respective frequencies and densities. While being essential for parametrizing styling and other drawing algorithms, this also creates the foundation for the semantic analysis and classification. The data gained are annotated and scrutinized following the principles laid down by BERTIN (1973) in his Semiology of Graphics, expanded, discussed and renewed by (SPIESS, 1970; KOCH, 2001; DARU, 2001; BERTIN, 2001).

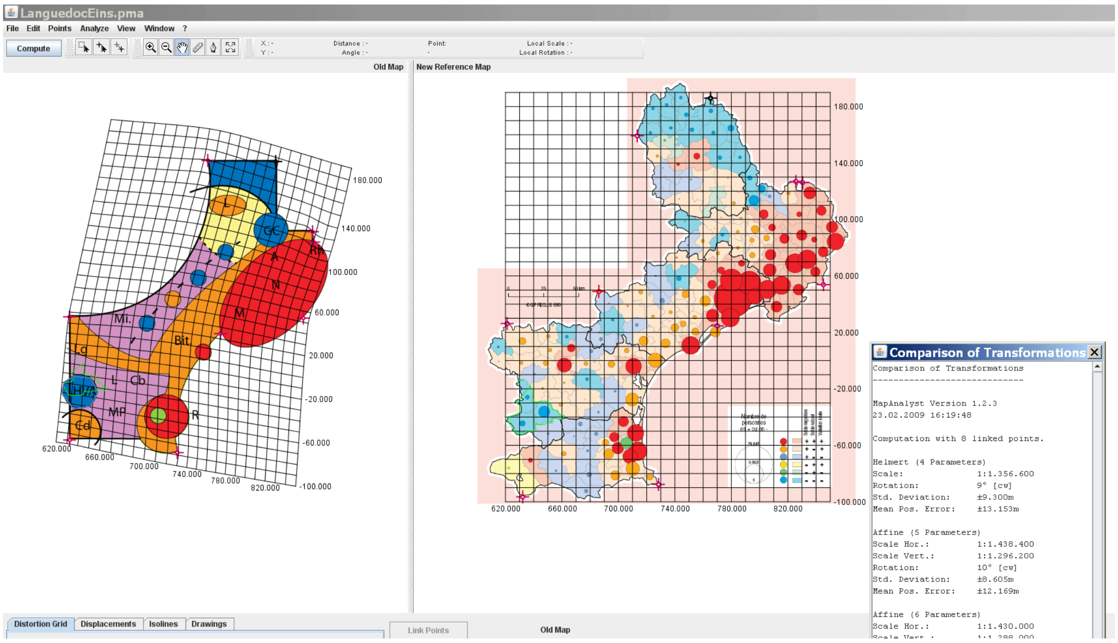

Figure 4: Concept of geometric analysis illustrated with MapAnalyst

Semantic Analysis

This leads into the semantic analysis. Viewing the specific chorematic maps as a whole the respective semantic expressiveness (Mackinlay, 1986) will be investigated. To make the findings comparable and machine readable, a notation for this expressiveness and visualizations in general needs to be used. As Brodlie and Noor (2007) point out, there is no canonical notation for visualizations that is not bound to a specific software or technology. One promising candidate in the context of this research is Humphrey (1999), who introduces a relational visualisation notation (for an exampe see Fig. 5). This notation is based on relational algebra and therefore easily computable and technology independent, while at the same time limited in what it can express (no transitive closures, no arbitrary data structures and no algorithms). Another way of adressing expressiveness and making it comparable could lie in the measurement of information content of signatures and whole diagrams. While current scientific cartography and geovisualization have not followed up on information theory based approaches (Bollmann, 2001), several very interesting measurements have been developed before the 1990s (Ogrissek, 1987, pp. 61f). Implementing the calculation of the information content with, for example, Teterin´s "metric principle", will be feasible within the statistical analysis. It will be interesting to see how such measurements numerically relate to naive measures of graphical complexity, like number of polygons, vertices, diagram elements and so forth.

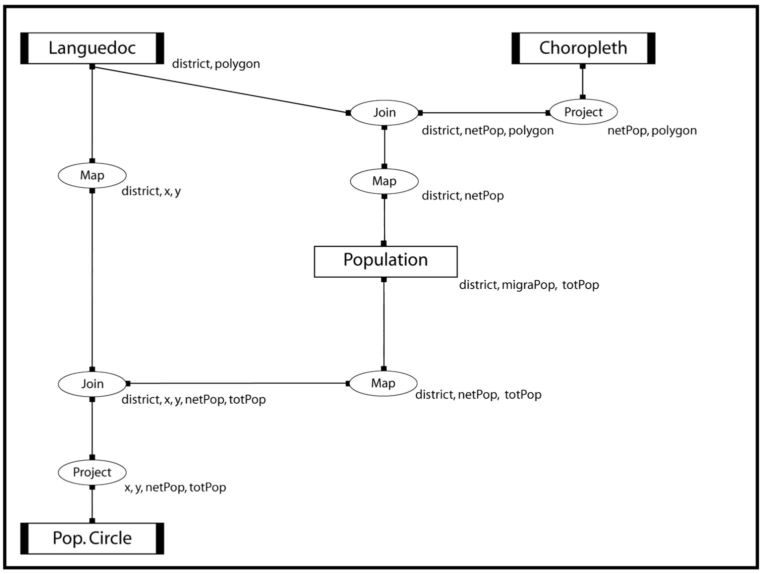

Figure 5: Thematic map of Languedoc-Roussillon (Fig. 4) in the form of a Relational Visualization Notation design graph

Information Aggregation

The original propagators of chorematic diagrams, GIP-RECLUS, already used computer based statistical methods to supplement their traditional spatial analysis techniques for the construction of their spatial models (Ormeling, 1992), a prototypical procedure is outlined in Benoit et al. (1993). Still, such procedures depend on expert knowledge, and are therefore impractical for automatization. A prelimenary literature review seems to indicate that even though great progress is being made in the areas of spatial and geographic data mining, knowledge discovery in databases (KDD) and feature detection, many solutions are also specific to particular problems (Hand, 2002). Therefore, the information aggregation will be adressed in the proof of concept use cases, thereby being a part of the future work. There is hope that a formal semantic description of chorematic diagrams and a classification of the wealth of diagram types will allow pre-selecting applicable methods by defining the expected output.

4 Discussion

While the outlined approach yields quantitative results on particular attributes of chorematic diagrams, they are more than the sum of their parts. The particular depictions of spatial configurations will undergo a qualitative analysis, parallel to the quantitative work. The work done so far indicates that the actual expressive power of chorematic diagrams is not due to clarity via simpllicity alone. The adaptibility of the concept allowed for the dedicated construction of new types of diagrams for many different subjects of interest.

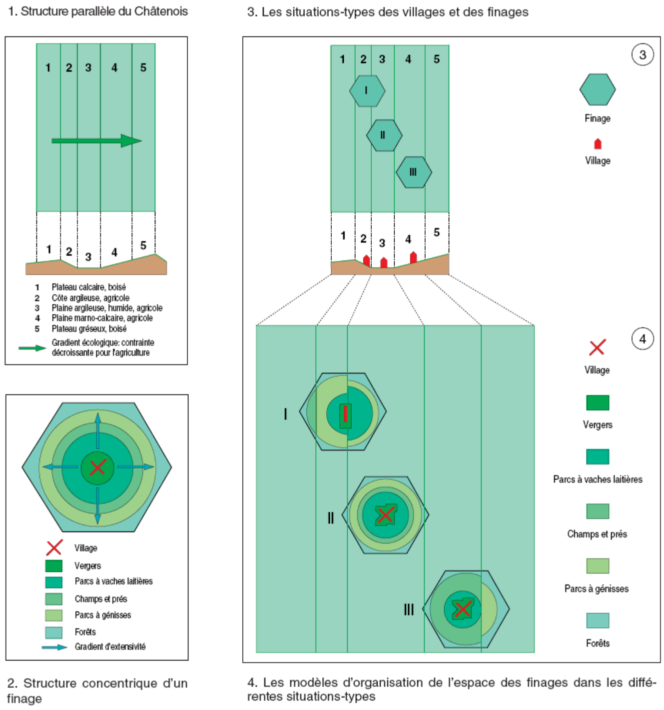

Figure 6: Sequential chorematic diagrams (DEFFONTAINES, 1990)

Whole issues of the journal MappeMonde were dedicated to single domains (e.g. agriculture), and provided a new visual grammar for them (CHEYLAN et al., 1990). Also, chorematic diagrams were seldomly used isolated, but rather as graphical part of an article, being just as important as the text. Through skilled use of the two dimensions of the plane, they often reached what can be called visual argumentation (MERSCH, 2006) (fig. 5). These aspects need to be investigated, too.

Once the analysis has progressed sufficiently, and a formalized description of chorematic diagrams has been created, a proof of concept will be implemented. Possible use cases are disaster management scenarios concerned with earthquake early warning and response for the city of Istanbul, as the represent themselves in the EDIM and SAFER projects. They pose interesting problems at different steps in the whole process: from visualization of sensor networks deployed in the city and their simulation runs to post-desaster decision support visualizations for the coordination of urban disaster relief services.

References

BENOIT, S., HILAL, M., TOUSSAINT, S. (1993): La composition des chorèmes dans la modelisation graphique appliquée à la bourgogne. In: *MappeMonde*, Vol. 4/93, pp. 37–41.

Bertin, J. (1973): *Graphische Semiologie. Diagramme, Netze, Karten.* Gruyter.

Bertin, J. (2001): Matrix theory of graphics. In: *Information Design Journal*, pp. 5–19.

Bollmann, J. (2001): Informationstheorie. In: *Lexikon der Kartographie und Geomatik*, Vol. 1.

Brodlie, K., Noor, N. M. (2007): Visualization Notations, Models and Taxonomies. In: *EG UK Theory and Practice of Computer Graphics (2007).*

Brunet, R. (1980): Le composition des modèles dans l´analyse spatiale. In: *L´Espace Géographique*, Vol. IX, pp. 253–265.

Brunet, R. (1987): *La carte : mode d'emploi.* Fayard-Reclus.

Brunet, R. (2005): *Le développement des territoires : formes, lois, aménagement.* Editions de l'Aube.

Brunet, R., Dollfus, o. (1990): *Mondes nouveaux*, Vol. 1 of *Géographie universelle.* Hachette-Reclus.

Cheylan, J.-P., Deffontaines, J.-P., Lardon, S., Thery, H. (1990): Les chorèmes: un outil pour l´étude de l´activité agricole dans l´espace rural? In: *MappeMonde*, Vol. 4/90, pp. 2–4.

Daru, M. (2001): Jacques Bertin and the graphic essence of data. In: *Information Design Journal*, pp. 20–25.

Deffontaines, J.-P. (1990): Organisation spatiale de l´activité agricole et développement d´une petite région lorraine. In: *MappeMonde*, Vol. 4/90, pp. 12–14.

Del Fatto, V., Laurini, R., Lopez, K., Loreto, R., Milleret-Raffort, F., Sebillo, M., Sol-Martinez, D., Vitiello, G. (2007): Potentialities of Chorems as Visual Summaries of Geographic Databases Contents. In: *Advances in Visual Information Systems*, pp. 537–548.

Del Fatto, V., Laurini, R., Lopez, K., Sebillo, M., Vitiello, G. (2008): A chorem-based approach for visually synthesizing complex phenomena. In: *Information Visualization*, Vol. 7(3-4), pp. 253–264.

Guilbert, E., Saux, E. (2008): Cartographic generalisation of lines based on a B-spline snake model. In: *International Journal of Geographical Information Science*, Vol. 22(8), pp. 847–870.

Hand, D. J. (2002): *Pattern Detection and Discovery: ESF Exploratory Workshop, London, UK, September 16-19, 2002 (Lecture Notes in Computer Science / Lecture Notes in Artific).* Springer, Berlin, 1. ed.

Humphrey, M. C. (1999): A graphical notation for the design of information visualizations. In: *Int. J. Hum.-Comput. Stud.*, Vol. 50(2), pp. 145–192.

Klippel, A. (2004): *Wayfinding choremes: Conceptualizing wayfinding and route direction elements.* Universität Bremen Universitätsbuchhandlung, 1. ed.

Koch, W. G. (2001): Jacques Bertin´s theory of graphics and its development and influence on multimedia cartography. In: *Information Design Journal*, pp. 37–43.

Kryssanov, V., Okabe, M., Kakusho, K., Minoh, M. (2002): Communication of Social Agents and the Digital City — A Semiotic Perspective. pp. 327–336.

Lardon, S., Capitane, M. (2008): Chorèmes et graphes. Production et transformation de représentations spatiales en agronomie. In: *Revue d´anthropologie des connaissances*, Vol. 2(4), pp. 195–217.

Laurini, R., Milleret-Raffort, F., Lopez, K. (2006): A Primer of Geographic Databases Based on Chorems. pp. 1693–1702.

Ligozat, G., Nowak, J., Schmitt, D. (2007): From language to pictorial representations. In: *Proc. of the Language and Technology Conference (L&TC'07), Poznan, Poland, 5-7 September, 2007.*

Mackinlay, J. (1986): Automating the design of graphical presentations of relational information. In: *ACM Trans. Graph.*, Vol. 5(2), pp. 110–141.

Mersch, D. (2006): *Bilder als Diskurse - Bilddiskurse.* Velbrück.

Ogrissek, R. (1987): *Theoretische Kartographie. Eine Einführung.*

Ormeling, F. (1992): Brunet and the revival of French geography and cartography. In: *The Cartographic Journal*, pp. 20–24.

Rocha Coimbra, A. (2008): *ChorML: XML Extension for Modeling Visual Summaries of Geographic Databases Based on Chorems.* Master Thesis, LIRIS, INSA-Lyon France.

Spiess, E. (1970): Eigenschaften von Kombinationen graphischer Variablen. In: *Grundsatzfragen der Kartographie*, pp. 280–293.

Tainz, P. (2001): Chorème. In: *Lexikon der Kartographie und Geomatik*, Vol. 1.

van den Besselaar, P., Beckers, D. (2005): The Life and Death of the Great Amsterdam Digital City. In: *Digital Cities III*, pp. 66–96.

van Elzakker, C. (2004): *The use of maps in the exploration of geographic data.* Dissertation, Universiteit Utrecht / International Institute for Gei-Information Science and Earth Observation.