GRAPHICS AND LANGUAGE AS COMPLEMENTARY FORMAL REPRESENTATIONS FOR GEOSPATIAL DESCRIPTIONS

Hans-Peter Bähr and Marina Müller

Institute of Photogrammetry and Remote Sensing, University of Karlsruhe, Englerstr. 7, 76128 Karlsruhe, Germany – (baehr, mueller)@ipf.uni-karlsruhe.de

ABSTRACT:

In image understanding, the importance of language (e.g. the choice of class names in image analysis) has been widely underestimated up to now. Within this paper, the close interaction between language and graphics is analysed for geospatial applications. It is shown that texts and maps are two complementary representations of knowledge, resulting among other in the use of an equal terminology for verbal and pictorial representations. Two different real scenarios (Brazilian cadastre and message analysing systems for disaster management) are introduced where the direct relationship of language and image analysis becomes not only obvious, but the mutual transformation - of one representation into the other - is the essential task of the application. First results of the presented project already demonstrate how both forms of representation can be transformed to a common symbolic level as a basis for further synergetic analyses.

KURZFASSUNG:

Bei Bildverarbeitungsaufgaben wurde bisher die Bedeutung von Sprache (z.B. bei der Wahl von Klassennamen in der Bildanalyse) weitgehend unterschätzt. In dieser Veröffentlichung wird die intensive Wechselbeziehung zwischen Sprache und Graphik eingehender analysiert. Texte und Karten stellen zwei komplementäre Formen der Wissensrepräsentation dar, wie es u.a. beim Gebrauch der gleichen Terminologie für verbale und bildhafte Repräsentationen deutlich wird. Anhand von zwei Anwendungsszenarios (brasilianisches Kataster und Systeme zur Analyse von Meldungen innerhalb des Katastrophenmanagements) werden nicht nur die direkten Beziehungen zwischen Sprache und Graphik veranschaulicht, sondern es sind auch konkrete Beispiele, bei denen die Transformation von einer Repräsentationsform in die andere die wesentliche Aufgabe darstellt. Erste Ergebnisse des vorgestellten Projekts zeigen, wie Graphik und Sprache auf eine gemeinsame symbolische Ebene transformiert werden können, auf deren Basis integrierte Analysen erst möglich werden.

1. INTRODUCTION

Language has always been considered as a prominent tool for knowledge acquisition, analysis, and storage. Language in spoken or written form is the dominant representation of knowledge in all its very different facets.

However, it will be shown that terms, concepts, language together with meaning and semantics interestingly plays only a minor role in image analysis applications. On the one hand, language defines an attractive field of research of its own e.g. in automatic understanding of spoken language, automatic reading and classification of handwriting, or recognising persons by their individual way of speaking. On the other hand, little has been done up to now to define language – at least names or single words – as a data type within image analysis.
It is exactly this apparent gap which will be focused in the following text.

The situation has been given attention by the Institute of Photogrammetry and Remote Sensing at Karlsruhe University since the early 90ties, but only recently a research project "Abstraktion graphisch und verbal repräsentierter Geoinformation" was established, embedded in a major bundle-type research programme[1] sponsored by the German Science Foundation (DFG).

---

[1] This programme is entitled "Abstraktion von Geoinformation bei der multiskaligen Erfassung, Verwaltung, Analyse und Visualisierung" chaired by Monika Sester/U of Hannover

2. LINGUISTIC ELEMENTS IN GEOSPATIAL ENVIRONMENT

"Geospatial environment" in the context of this paper integrates any georeferenced descriptions. It comprises their analogue and their digital representation: images, maps, graphics, and other kind of 2D data, so to say "clippings" from the "real world".

Navigation marks an application where language has already been introduced (Kray and Baus, 2003). This is, for instance, true for navigation of pedestrians, but especially for navigation of vehicles, which are tracked by GPS. The GPS output is restricted to coordinates, which do not give any meaning to a human unless referred, e.g., to a street (i.e. a name). Piloting a car driver requires spoken language as the optimal tool. Map-based navigation leads to unnecessary and sometimes dangerous distraction of the car driver.

In the course of classifying geospatial data sets, like satellite imagery, physical parameters have very rigorously been modelled, not to speak of the classification algorithms. Classes are completely described by spectral signatures or even by geometrical or topological features. However, much less attention is given to the major parameter: the meaning of that class, represented by its name. This situation had been pointed out by (Bähr and Schwender, 1996). The name of a class including their semantics denotes the final goal of the classification process which is necessarily incompletely described by the number of physical, spectral, and geometrical models.
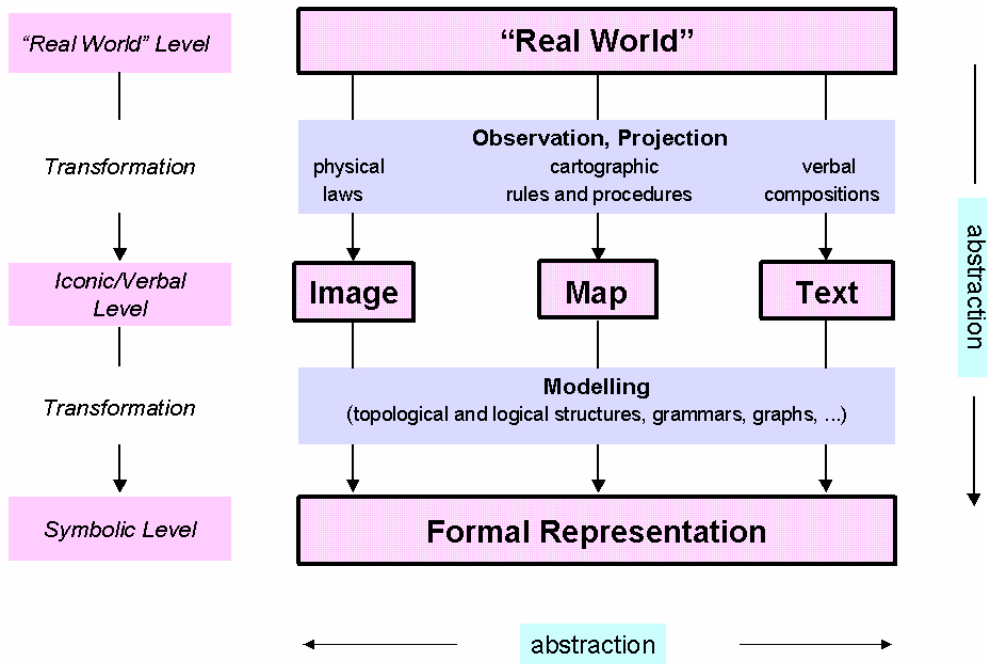
Figure 1. Knowledge transformation in different levels of abstraction.

Nota bene: The name is not at all a self-explaining, an error-free parameter: What is the meaning of a sealed surface, of a forest…..? In class descriptions, there is a contradiction of rigorously modelled physics and mathematics at the one side and of neglected linguistics and semantics at the other.

Finally, most primary information is obviously given by language. This holds true for geospatial descriptions, too. In case of further computer processing or machine vision, spoken or written language has first to be coded and then integrated in the analysis process. Language is an essential element in geospatial analysis. Systems for a more automatic processing and analysis of messages, still at very low level, are under development for a number of environments, e.g. in military applications or emergency/disaster management (Hecking, 2003; Allen et al., 2000). It is the latter field which is addressed in this paper.

The main question is how language can be fused with spatial data, whereas the problem of automatic reading and understanding is left aside.

3. TERMINOLOGY

When performing fusion, or even comparison of language and spatial data, terminological traditions of many disciplines are shared: from linguistics and cartography, cognition sciences and image analysis, from AI and even philosophy (Bähr, 2001; Bähr and Lenk, 2002). Therefore, terms have to be thoroughly defined.

3.1 The three levels: "reality", iconic/verbal, and symbolic

Both language and graphs represent knowledge, but coded by different tools. In order to define knowledge, Makato Nagao´s definition (Nagao, 1990) is taken: "Knowledge = Cognition + Logic". "Logic" implies representation in formal structures, e.g. Minsky´s "frames" (Minsky, 1975): There is no knowledge if

not assigned to organised patterns.

More terms will be discussed by Fig.1 which explains the sequence of reasoning starting from the "real world" level. It should be clear that human perception of "reality" is an indirect process, limited to mere observation of projections. There is no principal difference whether they are projected "on-line" into the human's eye or indirectly transmitted via maps, pictorial, or even by verbal descriptions.

In any case, man's process of perception cannot be separated from his (individual) brain-based inference machine. Hence geospatial analysis starts with secondary information laid down in images, maps, graphs, texts etc.. This second level, accessible to human analysis, shall be called "iconic" or "verbal", relating to imagery/graphs or texts, respectively (Bähr and Lenk, 2002).

Nevertheless, direct comparison or even fusion of knowledge at this low iconic level is not possible by using computers, though the human easily reads and compares maps, texts, and imageries based on his a-priory knowledge.

In order to organise knowledge in a computer-accessible form, a second transformation is required for its formal representation on a third, symbolic level. This step goes far beyond mere digitisation. Explicit modelling is preferred, where knowledge acquisition and knowledge use are strictly separated, like for semantic or Bayesian networks or for blackboard-controlled production systems. The reason for explicit modelling is the more rigorous model, which allows, among others, to determine directly the quality of image analysis. This is less stringent for implicit models, where the computer machine is trained "on the job" by an operator, like multispectral classification or neural networks. For explicit models, quality measure is given by the particular model itself, for implicit models it is kept in the operator's brain.

In Fig. 1 the word "abstraction" indicates scaling in two directions: vertically from the first level ("real world") to the third level (formal representation) as well as horizontally from

images to maps or to texts. The vertical path follows the traditional scheme and is self-explaining. However, it is a new perspective to consider steps from images to maps and/or to texts as abstraction. The justification is that transformations from verbal to pictorial and vice versa can never keep the complete data set from each of the two representations.

Finally, it should be pointed out that semantics is not included in Fig.1. Here, the features extracted from the "real world" are restricted to geometry, topology, names, and words. In computer science the term "semantics" suffers very often from degradation to mere "attribute", although it stands for "meaning", a very challenging concept. Integration of meaning in analysis of pictures and texts absolutely requires context. An isolated pixel, a separated line, or a single word carries no meaning at all until it is put into its context.

3.2 Complementary terminology for verbal and pictorial representation of geospatial descriptions

Since the human brain associates empirically the different representations of knowledge – verbal or pictorial -, he is normally not aware that he uses the same concepts for both domains.

The term "understanding" is a metaphor from language applied in image analysis. In the latter domain, it may replace "complete image description"

"completeness" refers to sufficiency with regard to a particular performance: a manual, describing the operation of an espresso machine, or an image, displaying the complete road net of a particular region.

"neighbourhood" relates to geometric or semantic vicinity. Concepts, verbal descriptions, or styles may be called similar as well as proximity of objects in imagery or adjacent figures in graphs.

"readable" is called a text which is mentally accessible and does not raise any doubts. It is, again, used metaphorically for maps and imagery, where it means easy interpretability due to adequate scale, layout, and design.

"precision" is a quality measure defined quantitatively for geometrical features in maps, graphs, and pictures but also in texts where it stands for a high level of detail in descriptions.

"homogeneity" indicates for both pictorial and verbal representations low variance within a region or a textual passage.

"level of abstraction" is used in a similar way for generalised texts and maps.

"redundancy" means repetition of information or unnecessary volume in verbal or in graphical descriptions. Redundancy is not necessarily negative as it may stabilise a system, e.g. for stochastic observations in least squares adjustment. Nevertheless, in formal or deterministic systems, like coded language or graphs, it is to be avoided.

"context" is an essential term for both the verbal and the pictorial world which has already been mentioned in the previous chapter. Different from the machine, the human puts any signal "automatically" into a context and thus adds semantics (=meaning), regardless of correct or incorrect reasoning.

The complementary nature of graphics and language is not limited to same or similar use of terms. It is true for the processing level, too, and allows parallel analysis and synergetic fusion of both domains. This will be shown in more detail by the applications given in the chapters 4 and 5. As a first example, "transformation" in relation of texts and images will be analysed:

"Transformation" is defined as rule-based changing texts or images from the original (input) level to a processed (output) level. In case of leaving the facts (e.g. of messages) or the objects (e.g. in images) unchanged, this transformation merely gives "a different perspective of view". A text may be formulated in many ways without changing the facts; an object may be imaged in many ways without changing the object itself. The different perspectives of view do not distort neither facts nor objects.

4. EXAMPLE I: BRAZILIAN CADASTRE

4.1 Initial situation

The Brazilian cadastre – unlike most systems used in Europe or America – is not map-based but consists of millions of notarially certified texts, each describing the boundaries of single land parcels. Additional maps are rarely available, and in cases of juridical conflicts, only the texts are committing. This form of administration of land ownership does not meet the necessary requirements to guarantee the legitimacy of ownership or to allow an efficient update and management of ownership information.

4.2 Data

Data base for this project are approximately 70 examples of real texts and accompanying maps. An example map can be seen in Fig. 2.
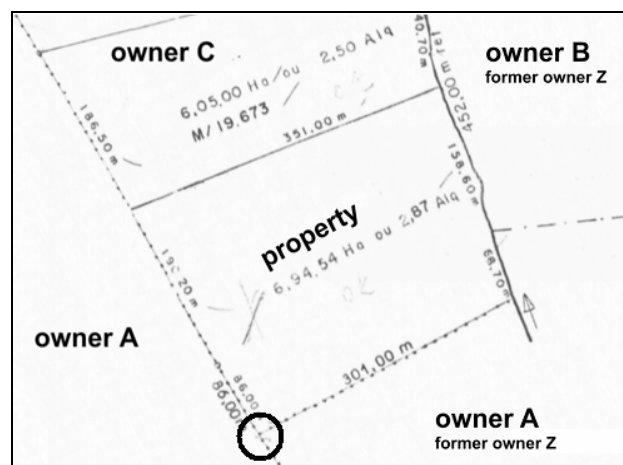

Figure 2. Part of a map of the Brazilian cadastre (real names of the property and owners replaced due to protection of data privacy). The circle indicates the starting point of the textual description (not marked in the original map).

The associated text to the map of Fig. 2 (translated from Portuguese) is presented below:

"It starts at the point of conjunction between a property of

owner A and a different property of owner A (formerly belonging to owner Z). Starting from this point, it follows a straight line for a distance of 301.00 m; at this point it turns left and follows a small creek along its windings for a length of 68.70 m, while it is adjacent to the property A (formerly belonging to Z). From this point it continues along the small creek and accompanies its windings along a distance of 158.60 m, while it is adjacent to the property of B (formerly belonging to Z). At this point, it turns left and follows a straight line for 351.00 m, while it is adjacent to the property of C. At this point, it turns left and follows a straight line for 190.20 m and turns slightly left and follows an already existing fence for 86.00 m up to the starting point and closes the existing boundaries while on this last part it is adjacent to the property of A."

As can be seen in the example, the texts describe the boundary parts of the parcels with precise lengths and names of owners of adjacent properties. Sometimes, the texts can include additional information about buildings, coffee trees, and other objects of value on the land.

4.3 Scientific challenges

The scenario of the cadastre was selected as an apparently "easy" application where the input (the describing text) and the output (the associated map) are a one-dimensional problem (line-following) and known in advance. The main goal of this project is to analyse in how far a transformation from text to map is possible. Related to this aim is the question which levels of abstraction are used in the text and in the map. Are they comparable? Is it possible to transfer the information of the text and the map to a common symbolic level which allows a direct comparison? How can topological and context information be represented?
More possibilities and questions arise after such a common symbolic level has been found: is each representation of a land parcel complete and free of redundancy? Are both representations consistent? Are representations of neighbouring land parcels consistent in their common border area? Is it possible to generate a new text starting from different border points and following the border in the contrary direction than originally described?
A closer inspection of the presented example reveals that the - at first glance - "easy" task of the Brazilian cadastre already comprises a number of difficulties and challenges due to rather compact and fuzzy text passages. Originally, each description referred to a visual inspection of the property, thus passages in the text such as "follows a small creek along its windings for a length of 68.70 m" or "turns slightly left and follows an already existing fence for 86.00 m" should be clear in the field. However, a transformation to a map based solely on the information of the text will have to cope with such incomplete and vague spatial information.

4.4 Transformation

In order to represent both text and map on a common symbolic level, a suitable form of knowledge representation has to be chosen. In this project, semantic networks were selected since they allow an explicit structuring of concepts and their relation within nodes and connecting links. Examples of successful applications of semantic networks in the area of speech and image analysis can be found in Mast et al., 1994; Paulus et al. 2000; Müller et al. 2003 and Kumar et al. 2004. In this work, the applied system shell for semantic networks is ERNEST (Niemann et al., 1990; Quint and Bähr, 1994). Figure 3 shows an excerpt of a semantic network for the Brazilian cadastre.
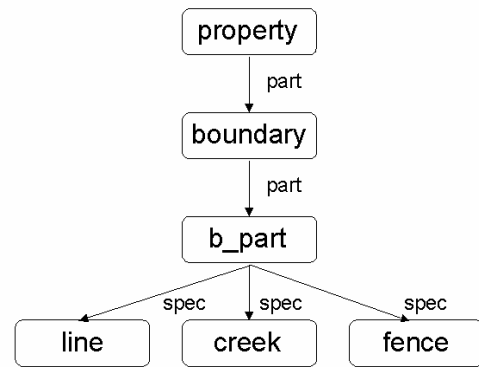

Figure 3. Excerpt of a semantic network for the Brazilian cadastre. Individual concepts (the nodes of the network, e.g. *property* or *b_part*) are connected via part- and specification-links.

This example already presents concepts of the common symbolic level of maps and texts. Both forms of knowledge representation are concerned with a certain "property". Besides some possible additional information about objects on the property, the main part of each text and the essential features of a map are related to the "boundary" of the property. Within the text, as well as in the map, the boundary is divided into segments (boundary parts, "b_parts") that are described in more detail or annotated with additional information. One of the details that are mentioned textually (also as text in the map within the legend to each chosen pattern) is the type (specialisation) of the boundary part. Besides these typical links (part, specialisation) that are used by most approaches for semantic networks, ERNEST allows e.g. to explicitly establish relationships between different levels of abstraction via a "concrete" link. A different level of abstraction occurs e.g. for the map representation between the concepts in Fig. 3 and their concrete realisation at a basic geometrical level (Fig. 4).
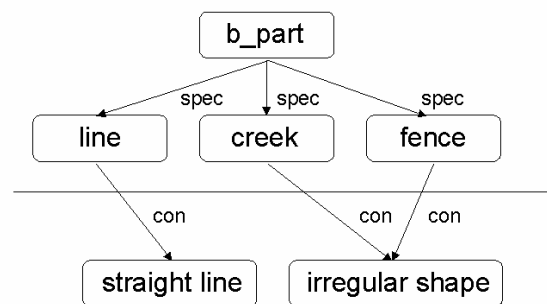

Figure 4. Different levels of abstraction (realised in ERNEST by "concrete" links) within the representation of maps of the Brazilian cadastre: boundary parts and their geometric realisation.

This geometrical level is one example where the difference between texts and maps becomes obvious. The only geometrical information that is always mentioned in the texts, is the length of a boundary part. There are other texts in the given data base of cadastre texts where the precise direction of a line is also given. However, such examples seem to be exceptions. Usually, if some information about the direction exists in the text, it is some relative and often vague description, e.g. "slightly left" (compare section 4.3). Analysing the texts, fences seem to be regarded as rather stable objects in this Brazilian region because

they are used as a reference in the text without further details about their direction. In the example of Fig. 2, fences are quite straight. However, there are other examples of maps where fences show more irregularities in their shape. Creeks do have windings that are explicitly mentioned in the text, but no further information about the direction of these windings is given in the description. Within a digitized map, the irregular shapes of fences and creeks are represented by polylines. If necessary, e.g. in cases of rapid changes in direction along the boundary path, a high number of intermediate points are marked, thus, not only the endpoints of a boundary part, but also its precise course between these endpoints is known and represented at the geometrical level.

4.5 First results

Although the project is an ongoing work and still rather at its beginning, some of the question of section 4.3 will be addressed in this section.

The questions concerning the different used levels of abstraction and a common symbolic level was already briefly examined in the previous section.

Boundary descriptions in cadastre are supposed to be rigorously redundant free. Within the available examples, the information in the descriptive texts is in such compact form that some knowledge even has to be concluded by logical inference or mere presumption. Besides the already mentioned missing information about the direction of fences or the windings of creeks, table 1 reveals more ambiguous attribute values of boundary parts.

No.	type	length in m	adjacent owner	direction of turning
1	line	301. 00	?1	left
2	creek	68.70	A (formerly Z)	continue
3	creek	158.60	B (formerly Z)	left
4	line	351.00	D	left
5	line	190.20	A	slightly left
6	fence	86.00	A	?2

Table 1. Boundary parts with their associated attributes, listed in the sequence of their appearance in the textural description in section 4.2. Some of the information is missing or ambiguous (marked by ?1, ?2).

Right at the first boundary part, the text mentions that the descriptions starts at a point of conjunction between two properties, both belonging to owner A, but one of them formerly belonged to owner Z. The text does not mention, to which of these two properties the first line is adjacent. However, after reading the complete text, it is possible to conclude on the owner of the neighbouring property: The last boundary part (no. 6) is connected to the first part and has as adjacent owner only A, thus it can be deduced that ?1 = A (formerly Z). The direction of turning after the last boundary part - in order to continue with the first boundary part again - is completely missing. Only if it is assumed that the description starts at a turning point and that the bending at the starting point is no exception from the typical walking direction then it can be inferred that ?2 = left. However, there are examples of other properties where the starting point is no turning point at all, therefore ?2 = left is only a hypothesis. An automatic extraction and disambiguation of the knowledge given in the descriptive texts (as it is manually prepared in table 1) is a task within the field of information extraction (Cowie and Lehnert, 1996). This is not part of the presented project, yet.

The maps of the cadastre contain redundancy on purpose since they should provide human interpreters with a quick overview of the situation and precise information. While the length and direction of the boundary parts is already implicitly represented by its drawing and the scale of the map, only the annotated numbers allow a quick reference on the precise distances and directions. However, the annotated numbers are directly taken from the textual description of the properties and are not always consistent with the actually drawn boundary parts. This leads to the question of consistency. The redundancy of the map is useful to match boundary parts of the map and of the associated text which can be checked further on consistency in other attributes (owner of adjacent properties, type). In the given example of section 4.2, there are already some inconsistencies in the type. Boundary parts no. 1 and no. 5 in table 1 are of type "line" while the corresponding boundary parts in the map are of type "fence". Either the description is simply not specific enough for these boundary parts or the map is wrong at these positions.

For the production of a reverse description, not only the sequence of appearance of the boundary parts has to be properly recorded within the chosen knowledge representation. Although the textual descriptions of the cadastre refer to geographic objects (boundary parts) that can be identified in both directions (in contrast to the problem in many navigation tasks where landmarks are only visible in one direction), the information about the turning direction has to be suitably adapted. If they are e.g. given as a certain degree of azimuth, they have to be computed considering ±180 degree. In the example in section 4.2 only relative spatial descriptions such as "left" are used that need to be exchanged by their opposite term. Composed terms, such as "slightly left", can be easily transformed by altering the spatial term ("slightly right").

5. EXAMPLE II: ANALYSING MESSAGES IN DISASTER MANAGEMENT SYSTEMS

Although the scenario of the Brazilian cadastre is already challenging, this second application is of an even higher complexity. Persons in charge in a disaster management centre depend on up-to-date information about events and conditions at different locations. Usually, such information is plotted on a map and regularly updated with the information of incoming messages. In case of a catastrophic event, a large number of such messages in verbal or written form are received and need to be interpreted and evaluated by staff members in the centre. Therefore, the aim in this application is to analyse spatial information in these messages, thus allowing a more efficient update of these maps. Incoming messages in the context of an example scenario "thunderstorm" might be:

- Train derailed at level-crossing A and driven into a magazine of fertilizers at B, fire and a high number of injured persons at B
- Several people in district C complain about shortness of breath due to strong smoke emission at B
- Main road D is blocked between street E and street F because of fallen trees and branches

All these messages contain spatial information, but on different scales (compare Fig. 5). Here, the previously 1D-problem of the cadastre is expanded to all 4 dimensions. To all three possible

spatial dimensions, time has to be taken into account as an additional factor in this scenario. Messages are related to events that occur at a definite point of time (e.g. an accident) or take place during a certain time interval (e.g. the amount of time that a fire-engine needs to get to the location of an accident). Additionally, new messages do not necessarily provide newer or more accurate or complete information. Information of incoming messages can be highly redundant, but this redundancy is desired to fill gaps in the knowledge base. In order to handle incompleteness and fuzziness of the messages, semantic networks will be combined with Bayesian networks in future works.

Figure 5. Messages with information at different spatial resolution: single locations (level crossing A, building B), linear features (streets D, E, F), large areas (district C).

6. CONCLUSIONS

This paper presented the ongoing work in a project about textual and graphic representation. It was shown that both representation schemes are closely connected and two example applications were presented where both representations and their relationship are essential parts of the given tasks. First results of the Brazilian cadastre demonstrate that a common symbolic representation of maps and texts can be established (here as semantic network), enabling a full exploitation of both knowledge resources. Further research in this project will be concerned with a more detailed analysis of the questions raised in section 4.3, especially in the context of analysing systems for messages within disaster management.

REFERENCES

Allen, J., Byron, D., Costello, D., Dzikovska, M., Ferguson, G., and Stent, A., 2000. TRIPS-911 system demonstration. In: Proceedings of the NAACL/ANLP 2000 Workshop on Conversational Systems, May 2000, Seattle, WA, pp. 32-35.

Bähr, H.P., 2001. The power of the links. Photogrammetrie, Fernerkundung, Geoinformation, 1/2001, pp. 23-31.

Bähr, H.P. and Lenk, H., 2002. In: Proceedings of PCV02 – Photogrammetric Computer Vision – ISPRS Commission III-Symposium 2002, Graz, Sept. 9-13, 2002. Available at: http://www.isprs.org/commission3/proceedings/ papers/paper058.pdf (26 April 2004)

Bähr, H.-P. and Schwender, A., 1996. Linguistic confusion in semantic modelling. International Archives of Photogrammetry and Remote Sensing, vol. XXXI, Part B6, pp. 7-12.

Cowie, J. and Lehnert, W.G., 1996. Information extraction. Commun. ACM 39 (1), 80-91.

Hecking, M., 2003. Information extraction from battlefield reports. In: Proceedings of the 8th International Command and Control Research and Technology Symposium, 17-19 June, Washington, DC. Available at:
 http://www.dodccrp.org/events/2003/8th_ICCRTS/pdf/019.pdf (26 April 2004)

Kray, Chr., and Baus, J., 2003. A survey of mobile guides. In: Proceedings of the Workshop 'HCI on mobile guides" at the 5th International Symposium on Human Computer Interaction with Mobile Devices and Services, Sept. 8-11, 2003, Udine, Italy, available at:
http://www.comp.lancs.ac.uk/computing/users/kc/mguides03/ KrayBaus-final.pdf (26 April 2004)

Kumar, A., Sundararajan, S.C., and Lieberman, H., 2004. Common sense investing: bridging the gap between expert and novice. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2004), April 24-29, 2004, Vienna, Austria, ACM 1-58113-703-6/04/0004, available at: http://web.media.mit.edu/~hugo/conceptnet/ (26 April 2004)

Mast, M., Kummert, F., Ehrlich, U., Fink, A.G., Kuhn, T., Niemann, H., and Sagerer, G, 1994. A speech understanding and dialog system with a homogeneous linguistic knowledge base. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(2), 179-193.

Minsky, M., 1975. A framework for representing knowledge. In: The Psychology of Computer Vision, ed. P.H. Winston, McGraw-Hill, New York.

Müller, S., Weis, M., Liedtke, C.-E., and Pahl, M., 2003. Automatic quality surveillance of GIS data with GeoAIDA. In: ISPRS Archives, Vol. XXXIV, Part 3/W8, available at: http://www.tnt.uni-hannover.de/project/kbs/geoaida (26 April 2004)

Nagao, M., 1990. Knowledge and Inference. Academic Press, ISBN 0-12-513662-5.

Niemann, H., Sagerer, F., Schröder, S., and Kummert, F., 1990. ERNEST: a semantic network system for pattern understanding. PAMI, 12(9), pp. 883-905.

Paulus, D., Ahlrichs, U., Heigl, B., Denzler, J., Hornegger, J., Zobel, M., and Niemann, H., 2000. Active knowledge-based scene analysis. Videre: Journal of Computer Vision Research, vol. 1, no. 4, The MIT Press, pp. 6-29.

Quint, F. and Bähr, H.-P., 1994. Feature extraction for map based image interpretation. In: Proceedings of the 3rd Int. Colloquium of LIESMARS, Wuhan, China, pp. 1-8.