LEVELS OF ABSTRACTION IN TEXTUAL REPRESENTATIONS OF GEOINFORMATION – EXAMPLE: CADASTRAL DESCRIPTIONS IN BRAZIL

M. Müller

Institute of Photogrammetry and Remote Sensing, Universität Karlsruhe (TH), Englerstr. 7, D-76128 Karlsruhe, Germany – marina.mueller@ipf.uni-karlsruhe.de

KEY WORDS: abstraction, textual representation of spatial information, cadastre, hierarchical modelling, transformation, uncertainty

ABSTRACT:

Transformations between textual descriptions of spatial information and other types of representation, e.g. maps, require a deep understanding of the actual information given in the texts. The level of detail used in such descriptions depends on various aspects, e.g.: the purpose of the description and thus the necessary information for the application, the author's education and availability of tools for measurements, or the author's assumptions about additional knowledge that a possible reader of the text might have. In this paper, the variability of descriptions is demonstrated by cadastral texts from Brazil. Analysing this data, the original hypothesis had to be rejected that an abstraction hierarchy based on the content of spatial information can be established so that a textual description as a whole can be assigned to one of those abstraction levels. However, by adjusting the idea of the first hypothesis to single spatial attributes and by introducing a hierarchy model of "missing information sources" and uncertainty, it is possible to represent explicitly the desired information content of the texts. The necessary flexibility is achieved by an extendible approach.

ZUSAMMENFASSUNG:

Transformationen zwischen textuellen Beschreibungen räumlicher Informationen und anderen Repräsentationsformen, z.B. Karten, erfordern ein tiefes Verständnis der tatsächlich vorhandenen Information im Text. Der Grad der Detaillierung in diesen Beschreibungen hängt von verschiedenen Faktoren ab, z.B.: dem Zweck der Beschreibung und damit der notwendigen Information für die Anwendung, der Ausbildung des Autors und den vorhandenen Werkzeugen für Messungen, oder auch von Annahmen des Autors, welches zusätzliche Wissen dem Leser des Textes zur Verfügung steht. In diesem Beitrag wird die Variabilität von Beschreibungen anhand von Katastertexten aus Brasilien demonstriert. Aufgrund der Analyse dieser Daten musste die ursprüngliche Hypothese verworfen werden, dass eine Abstraktionshierarchie basierend auf dem Gehalt an räumlicher Information eingeführt werden könne, bei der textuelle Beschreibungen als Ganzes einem einzelnen Abstraktionsniveau zugeordnet werden. Allerdings ist es möglich durch Anpassung der Hypothesenidee an einzelne räumliche Attribute und die Einführung eines hierarchischen Modells mit „fehlenden Informationsquellen" und Unsicherheit den gewünschten Informationsgehalt der Texte explizit darzustellen. Die notwendige Flexibilität wird durch einen erweiterbaren Ansatz erreicht.

1. INTRODUCTION

Textual descriptions of geoinformation are applied in many situations where other types of representation, e.g. an additional map, might be useful, too. Some approaches for a possible transformation from textual to graphical information were already introduced for applications such as visualisation of:

- accident descriptions for evaluations by insurance companies (Egges et al., 2001),
- spatial information in news (Leidner et al., 2003), or
- military reports for tactical reasons (Schade and Frey, 2004).

However, before spatial information of textual descriptions can be transformed into another type of representation, it is essential to understand what information is actually given in the text and if additional knowledge might be necessary for a different type of representation of this data. In this paper, the focus will be basically on the first part, i.e. an analysis of the information content of texts. There are many papers in cognitive literature where the connection between language and space is analysed (see section 3.4 of Mark et al., 1999, for a short survey). However, the reported experiments very often analyze how verbal descriptions are structured and suited for navigation in an unknown environment (see e.g. Streeter et al., 1985, Tversky and Lee, 1999, Fontaine and Denis, 1999). Experiments where people are asked to draw a map based on textual information are rare. An exception is presented in (Tom & Denis, 2003), although in this case the participants drew the map after already walking a route guided by the given textual instructions, therefore the generated maps are not based on the descriptions alone but also on additional environmental information. Usually, the authors of textual descriptions are convinced that the information within their text is sufficient for the task at hand. They omit details that they think are unimportant because these details should be either well-known or the information might be available in the context of the application. This reduction of detail is a form of generalisation, thus abstraction of textual descriptions is the central topic of this paper.

There exist quite a number of models of abstraction for spatial information in various disciplines (cartography, computer science, psychology). Examples are models with clear semantic boundaries between their individual levels of abstraction (sensor layer, geometry layer etc. for image interpretation in

Liedtke et al., 2001; sensory level, control level, causal level etc. for robot navigation based on the spatial semantic hierarchy, see Kuipers, 2000). Other models introduce more artificial definitions for their levels of detail (LoD 0-4 for 3D city models, Ewald and Coors, 2005). Additionally, there is also a close relationship between abstraction and uncertainty (omitting details introduces uncertainty), and a number of publications are also available where uncertainty in the context of geoinformation is analysed. Plewe gives an excellent overview in (Plewe, 2002) of existing literature and introduces a thorough model of types of uncertainty.

For a better understanding and as a good basis for applications such as transformations of textual information into other types of representation, the relationships between information content, uncertainty, and abstraction as well as their consequences on levels of abstraction is analysed in this paper. The study is based on descriptions of the Brazilian ownership cadastre with the assumption that no other sources of information are directly available. Background information concerning this data is given in section 2. In section 3, the different levels of abstraction that can be observed in cadastral texts are analysed. In section 4, the approach for a suitable model of abstraction is given while the paper finishes in section 5 with a conclusion and outlook.

2. DATA

2.1 The Brazilian Ownership Cadastre

Brazil belongs to one of the countries where the ownership cadastre is in principle based on textual descriptions. Additional maps might be available, but only in rare cases. Besides, even if a map exists for a real estate, solely the textual descriptions are considered in judicial decisions. In 2001, a law (no. 10267) was passed in Brazil that became known as “the law of the georeferenced cadastre” (Bähr et al., 2005): For the first time, rules about the form of cadastral texts were introduced, namely that all boundary points of a land parcel must be given in the texts using coordinates of the official Brazilian geodetic system. However, this means that before 2001, no official standards for the recording of cadastral texts existed. The descriptions of the Brazilian cadastre from times before 2001 often resemble “metes and bounds” property descriptions that can still be found e.g. in the US, Canada, or Australia. In such texts, properties are described by reference to the bearings and lengths of the boundary line (metes) together with the names of adjoining properties (bounds) (UN-ECE, 1996). The deeds from the Brazilian cadastre are usually historical documents, sometimes a couple of years, decades, or more than a century old. The persons who recorded the boundary information were often not trained surveyors and/or did not have the appropriate tools for the measurements. There still exist very old, but legal documents that include descriptions such as “from this point onwards three day rides towards the direction of the rising sun” or “the area extends in depth as far as the cows graze” (Bähr et al., 2005). Even if units were used, they were not always as standardised as e.g. “meter” or other SI units. An example is the distance unit “legua”. It was frequently used in Europe and Latin America. Depending on the country or even a local region of a country, one legua corresponds to some distance value of at least 4 km up to more than 6 km. These differences in length can be easily explained by the way how a legua was measured e.g. in Brazil. The procedure is described as follows (Bähr et al., 2005):

At the point of the beginning, the surveyor puts tobacco into his pipe and lights the pipe. Then he mounts his horse and gives it a touch of the spurs. As soon as the tobacco is used up and the pipe begins to get cold, he stops his horse, gets of his horse and marks the reached point. The covered distance equals one legua.

Even if it is assumed that the same rider starts from the same point, uses the same horse, pipe, and amount of tobacco and rides through the same environment, it is extremely unlikely that he will mark the same spot twice.

With the new law, not only descriptions of newly registered properties have to fulfil the new standard but each old description needs to be transferred into the new coordinate format when changes to the property occur and need to be documented. If and how this transfer to a new text (or a suitable map representation) can be done in a more automatic way, is one of the analysis goals of the presented work but will not be further discussed in this paper.

2.2 Project Database and Examples

For the presented project, about 70 examples of cadastral texts with accompanying maps were available, both from rural and urban land parcels of a community in the federal state of São Paulo. A typical example of a description for rural properties is the following text (translated from Portuguese, names substituted by characters due to protection of data privacy; comments of the translator in angular brackets), Figure 1 shows the corresponding map:

Description of the area of 2 ha: “It begins at a common point, at the entrance of the barbwire fence where [the property of] A meets the area of the properties of B; from here it turns in direction 73°55’ NW during which it borders the area 05 by means of a pegged-out borderline for a length of 45.00 m (forty-five meters), from here it continues in the same direction along a pegged-out borderline during which it borders a property of C, for a length of 35.00 m (thirty-five meters); from here it turns left in direction 54°50’ SW during which it borders the area of the property of D for a length of 185.00 m (one hundred and eighty-five meters); from here it follows in direction SE during which it borders the other part of the real estate of E and others for a length of 190.00 m (one hundred and ninety meters), from here it turns left in direction 18°30 NE during which it runs along a fence and borders [the property of] A for a length of 162.00 m (one hundred and sixty-two meters), meeting here the entrance of the fence, the point at which the beginning is located and this description finishes.

Urban lots are usually much smaller than rural land parcels. Additionally, street names are typically used as directional reference in cities. The following description (again translated from Portuguese and characters instead of street names etc.) of an urban lot demonstrates these differences, see Figure 2 for the corresponding map:

Situated in this city of B, SP [São Paulo], second subdistrict, district C, in the residential area called D, a property consisting of lot 02 in block 11, part of the plan of the mentioned residential area, measuring 10.00 meter at the front side along street A, with the same dimension at the back side, where it borders lot no. 23, with a length of 25.00 meters on both sides, from the back side to the front side, with the right side (for someone facing within the property in the direction of *the street) bordering lot no. 03, and with the left side bordering lot no. 01, enclosing an area of 250.00 m².*

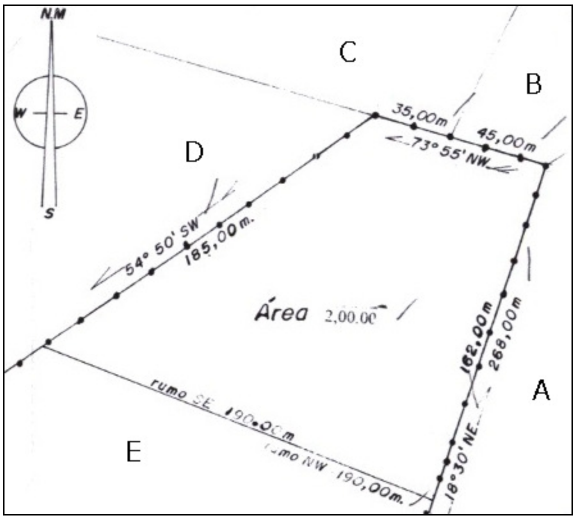

Figure 1. Corresponding map to the example description of a rural land parcel. Names substituted by characters in accordance to the textual description.

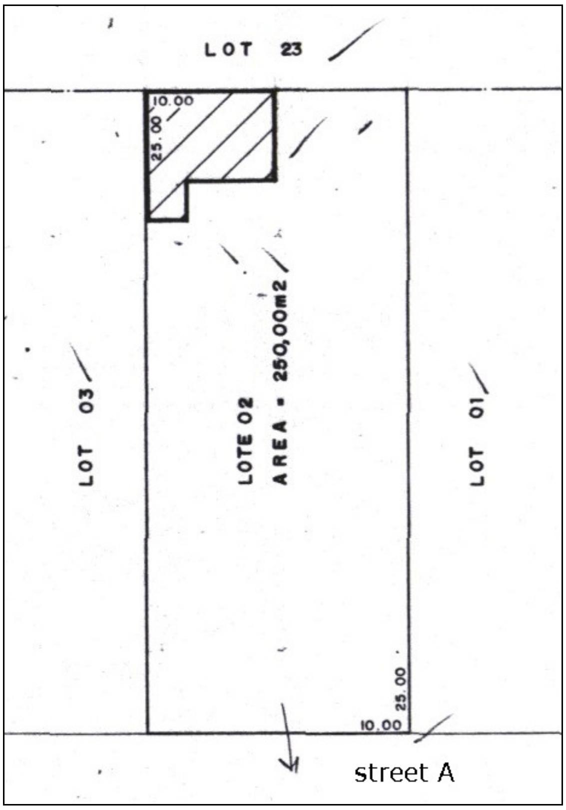

Figure 2. Corresponding map for the example description of an urban lot.

Although no official standard for descriptions existed before 2001, and each author of these cadastral deeds could (and did) set up his own standards for a satisfactorily description, the examples from the community in the federal state São Paulo agree on some typical obligatory parts:

- name of the owner,
- size of the property,
- affiliation to community, city,
- point of beginning (for the rest of the paper abbreviated by *POB*),
- topologic information concerning boundary parts: connectedness of individual boundary parts, names of land parcels and/or their owners for each neighbouring property
- individual attributes of boundary part:
  - length,
  - direction (in cities: names of streets).

Some more optional parts of descriptions might be the type of the boundary (e.g. creek, fence, ditch) or objects on the land (buildings, plantations etc.).

3. OBSERVABLE ABSTRACTION IN CADASTRAL DESCRIPTIONS

With the lack of standards for cadastral texts, each description contains information about the property boundaries at various levels of detail. This can be explained by different reasons, e.g. the educational background of the surveyors, lack of suitable tools for measurements, but also varying assumptions of the authors what kind of information would be available if a reconstruction of a real estate boundary were required. Very often, it seems to be expected that the text is used in a direct field inspection, thus allowing the identification of landmarks or an interview with land owners and neighbours. Due to these reasons, there exist e.g. on the one hand descriptions with lengths in millimetres and directions as azimuth-angles for each boundary part combined with UTM coordinates for the *POB*. On the other hand, there are also other texts with only qualitative spatial terms, such as “turn left”/”turn right” (see e.g. Bähr and Müller, 2004), as directional information and a *POB* given as relative description of a point where the boundaries of several objects meet each other. Since the level of abstraction is obviously connected to the level of detail of each single attribute of the property description, three of the geometric attributes of boundary descriptions (length and direction of boundary parts, *POB*) are analysed in more detail. The following lists summarise examples of expressions for each attribute that were either found in the project data or reported e.g. in (Bähr et al., 2005):

POB

- coordinates, e.g. UTM coordinates
- relative reference e.g. as meeting point between different land parcels of neighbours and the current property

direction

- usually given in degree as azimuth or bearing (see rural example)
- vague directions. The vagueness is either explicit by a textual indicator (“about”, “approximately”) or implicit: in cases where a compass direction (e.g. “south-east”) but no degree value is mentioned, it can be assumed that this is only a vague statement about the direction (see rural example of section 2.2 where the hypothesis of a vague compass direction, i.e. SE is only approximately SE, is confirmed by the map representation in Figure 1)

- qualitative spatial terms such as "turn left", "turn right", "continue" in combination with a straight boundary line. This category also includes boundary parts that are irregular in shape but where it is explicitly mentioned in the text that the direction is given from endpoint to endpoint of this boundary segment
- qualitative spatial terms in combination with irregular/unknown boundary shapes ("turn left and follow the windings of a creek / a fence / a street")
- no direction given at all

length

- given as a quantity of a well-defined standard unit such as "meter"
- a number given but additional uncertainty:
  - about the measured distance itself ("follow the line for approximately …")
  - about the measurement unit ("legua", see section 2.1)
  - about the position to measure the distance, e.g. "… meter along the river": how should the distance be measured in the field? Along the middle axis of the river or along the river bank? Does the position of the middle axis/the river bank change over time (high tide/low tide)?
- vague/ambiguous descriptions without a number ("as far as the cows graze")
- unknown distance

Further levels of abstraction can also be found in non-geometric attributes such as the names of owners or neighbours. At the most detailed level, the names are fully given with all first and last names. On a more abstract level, only the last name is fully given, first names are abbreviated. In case of several owners of a land parcel, another level of abstraction can be identified by the complete omission of the names of the other owners as can be seen in the example of the rural property where only "others" are mentioned. On the most abstract level, all names are completely missing or referred to as "to whom it may belong".

In summary, the most important observations that can be made about abstraction levels in cadastral maps, are the following:

1. <u>Each individual text can rarely be assigned to one single level of abstraction</u>:
   Since each individual attribute of a boundary can be represented on a different level of detail, the texts are usually a mixture of different levels of abstraction. This is demonstrated by the example of the rural land parcel in section 2.2 where the length is given as meter-value while the *POB* is described on an obviously more abstract level as a relative point without any explicit coordinates.
2. <u>The level of abstraction of one single attribute in a description can vary from boundary part to boundary part:</u>
   This is again demonstrated by the rural example where the direction is usually given with degree-values but also for one boundary part only by "SE" which has to be interpreted as "approximately SE" and not as exactly "45°0'0'' SE".

Based on these observations, a very flexible model is necessary to handle all possible levels of detail for spatial attributes and their variety of combinations.

4. MODELLING ABSTRACTION IN CADASTRAL DESCRIPTIONS

4.1 Approach of "Missing Information Sources"

The original hypothesis at the beginning of the project was that each texts is written in a style where the same level of detail is applied throughout the whole description, thus it would be possible to assign a complete text to a certain level of abstraction. With the results of section 3, this hypothesis had to be rejected. However, in order to obtain explicit knowledge about information that is actually given in the texts, a slightly adapted version of the first hypothesis was applied. A straightforward approach was to establish a hierarchy of abstraction for each (in this paper: spatial) attribute. The abstraction hierarchy itself is based on the main criterion "content of spatial information", or more specific: the amount and type of "missing" information. Four levels (already applied as ordering criterion in the lists of section 3) were identified based on the necessary additional sources of information that are required for a reconstruction of the spatial properties of the boundary parts, see Figure 3:

- *none / general knowledge:*
  At the lowest level of abstraction, all necessary information concerning a single boundary attribute is readily available by the text passage concerning this attribute. The author of the description provided all information in numerical values with reference to standard measurement units or systems to enable a direct reconstruction of the geometrical feature. Of course, in order to be able to reconstruct the information some basic general knowledge has to be already assumed on this level, e.g. about coordinate systems, distance units, or directional systems.
- *context dependent knowledge*:
  On the next level of abstraction, the information that is given for a single attribute cannot be directly reconstructed because it includes a reference to other information of this textual description (which is defined here as the *context* of an attribute information). Examples for this dependency are qualitative spatial terms such as "continue" or "turn left". In both cases, the current direction is not directly available but depends on the directional information of the previous boundary part.
- *external sources:*
  Data on this level has references to objects and their spatial attributes that cannot be resolved by the given source of information (the cadastral text = the context) but require external information such as maps or other textual descriptions.
- *all / new acquisition:*
  On this most abstract level, even this reference to a yet unavailable source of information is missing. There is no information given at all concerning the respective spatial attribute. Therefore, a field survey, i.e. a new acquisition

of this attribute information is the only possible solution for a reconstruction of this data.

The two observed levels of abstraction for the *POB* can be clearly assigned to two of these main levels:

- coordinates in standard systems:
  No additional information is necessary, therefore *POB* information at this level of detail belongs to the level of abstraction where the required information sources are: *none / general knowledge.*
- relative reference (e.g. "meeting point"):
  The information given in the text is not sufficient to resolve the missing information but topographic maps or other texts of neighbouring properties with given coordinates might allow the localisation of the point. Therefore, this information is clearly on the level *external sources.*

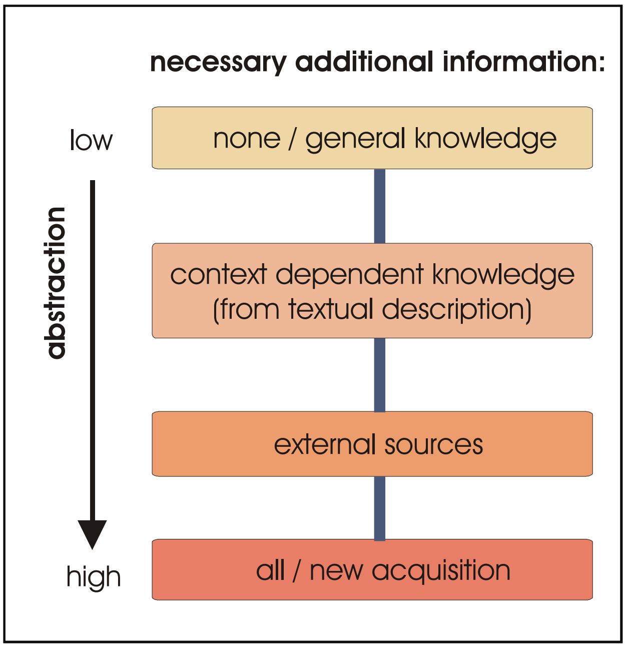

Figure 3. Levels of abstraction based on sources of information that need to be available to reconstruct the spatial geometry and location of a described object.

However, a second factor influences the assignment of descriptions to a level of abstraction which is the uncertainty that is connected with the given geoinformation. Although "missing information sources" themselves are already a cause for the types "imprecise" and "unknown" of uncertainty (Plewe, 2002), the degree of uncertainty on each of the four identified main levels can vary due to other reasons of uncertainty. Thus, the main levels of abstraction resulting from "missing information sources" are further divided by different sublevels of uncertainty as it is displayed by the examples of the attribute *direction*:

- degree-value + standard degree system:
  Although it needs to be considered that there is always uncertainty due to measurement/observation limitations (Plewe, 2002, Frank, 2003), this can be seen as the lowest level of abstraction available for directions. It can be directly geometrically interpreted and no additional information is necessary. The corresponding level is *none / general knowledge.*
- vague direction:
  Some degree value is specified (maybe indirectly by a point of the compass), though the description makes it clear that more uncertainty is connected with the given value than can be expected due to the usual measurement limitations. Nevertheless, the value can be used as a first approximation without other sources of information, thus it can be assigned to the main level *none/ general knowledge*. However, this higher degree of uncertainty (and therefore abstraction) needs to be revealed and considered in a transformation because the correct interpretation of uncertainty that is connected to information in e.g. maps might be an important factor in a decision process (Harrower, 2003).
- qualitative spatial terms + straight boundary:
  As already mentioned, qualitative spatial terms such as "continue" or "turn left" depend on the direction of the previous boundary part. Therefore, the main level is *context dependent knowledge*. However, the interval of possible direction values that are connected to "continue" can be assumed as more restricted than the interval of directions for "turn left" or "turn right" (some few degree values around the previous direction vs. an interval of about 180 possible degrees). Therefore, "continue" is less uncertain, i.e. abstract, than "turn left"/"turn right" and the main level *context dependent knowledge* for directions should be actually at least subdivided into: *"continue" + straight boundary* and *"turn left"/"turn right" + straight boundary*.
- qualitative spatial terms + irregular shape (e.g. "creek"):
  In this case, the knowledge of the previous direction and the given qualitative spatial term only allow assumptions about the direction near the beginning of this boundary part. In order to identify the complex shape and directional changes of the whole boundary part, *external sources* are necessary. Again, the occurrence of "continue" and "turn right"/"turn left" subdivides this level of abstraction, too.
- no direction given at all:
  The lack of explicit directional information requires a field survey including interviews with the land owner and his neighbours as the only reliable source of information. The level of abstraction is obviously the level *all/ new acquisition*.

The observed expressions listed in section 3 for the third attribute *length* can equally be assigned to the main abstraction levels defined by "missing information sources" and they are also subdivided due to further sources of uncertainty. A particularity are the "relative length measures". Clearly, external sources of information are necessary, e.g. to answer questions like: "what influences how far cows graze in the environment of this real estate?" This example shows that *external sources* include not only geometric or map information (topography, vegetation) but also specialised expert knowledge.

NB: Some cases of originally unknown/missing information might be reclassified in this abstraction hierarchy because additional application-dependent knowledge can be exploited. In the case of cadastral texts, such knowledge is e.g. the constraint that the boundary is a closed polygon. If a cadastral texts misses one direction value or one length value of the boundary parts, this value can be calculated based on this constraint and the length and direction information from all other boundary parts (context knowledge). A missing *POB* is another interesting case: if no information is explicitly given for the *POB*, relative information can always be inferred because relative descriptions usually consists of information of the neighbours of the first and last boundary part. The name of neighbours is one of the obligatory parts of a cadastral description and should be available in nearly all texts. Therefore, for the *POB* the case of "unknown information" which usually requires the *new acquisition* of data, coincides here with the level of "relative reference" where *external sources* of information might be sufficient.

4.2 Realisation of the approach

Due to the complexity of possible variations in attribute descriptions, the model is currently realised as a flexible, extendible system: Newly observed abstraction levels can be incorporated in the existing hierarchy similar to the principle of linked lists in computer science (Figure 4).

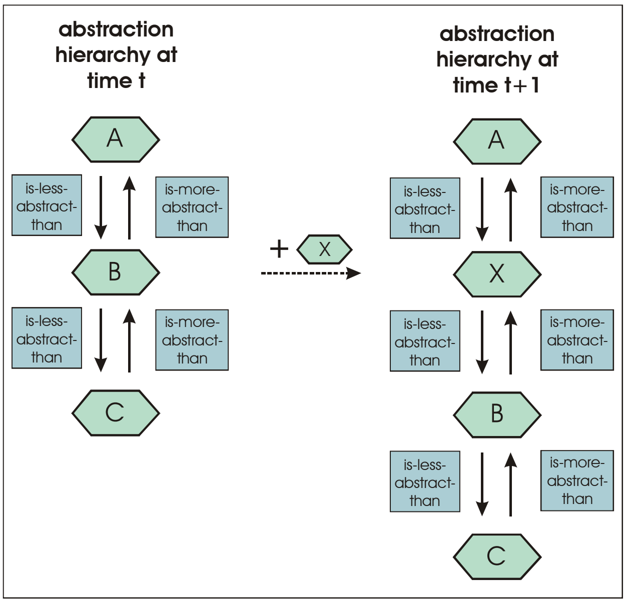

Figure 4. Example of the integration of a new level of abstraction X into the existing hierarchy.

Using such an approach, it is possible to avoid the assignment of a fixed number, i.e. a definite value of "abstractness" for each level of an attribute description. Since the level of abstraction might be equal for two different types of attribute description (e.g. representations of directions in angles of azimuth are equally abstract than directions given as bearing), list elements can have more than one successor or predecessor (Figure 5).

Besides this basically linear structure of abstraction levels represented by the links *is-less-abstract-than/is-more-abstract-than*, a tree structure is established by the main levels that are elements of S = {none/general knowledge, context dependent knowledge, external sources, all/new acquisition} and their sublevels of uncertainty (Figure 6). In order to explicitly model and visualise the concepts and their relationships, an ontology editor (Protégé, 2006) was used in the current project. The presented expandable approach already fulfils the aim that the spatial information that is contained in the textual descriptions is made explicit and that necessary additional information for transformations into other types of representation becomes evident. Ideally, the concepts should be integrated in an existing ontology for spatial information theory where most of the concepts of the current application are already defined and a clear documentation of assumptions about new concepts is encouraged. An ontology of this general form is e.g. suggested as tiers of ontology in (Frank, 2001) where the hierarchy of abstraction levels of the current paper would probably be part of tier 4, the tier that deals with ideas of cognitive agents and particularly with concepts of incomplete knowledge and reliability of sources of information.

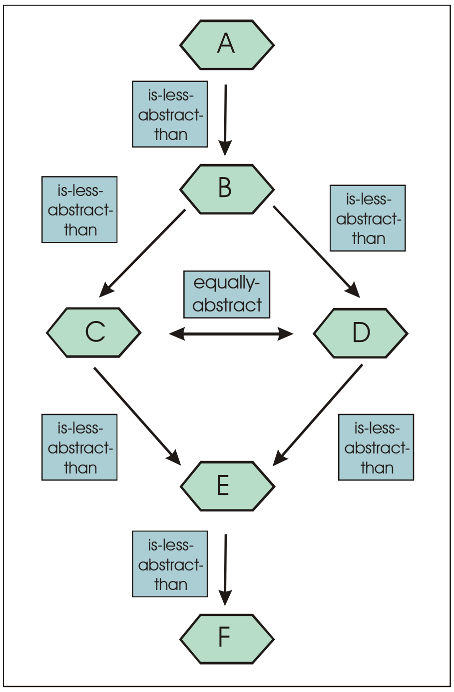

Figure 5. Integration of attribute descriptions with equal level of abstraction. For reasons of readability, the links *is-more-abstract-than* are not displayed here.

4.3 Towards a general quantitative model

If a complete, general quantitative model of abstraction based on "missing information sources" and uncertainty were desired, definite values of abstraction levels ("abstractness") need to be computed by an equation such as

$$L(x) = f(K(x), U(x)) \quad (1)$$

where
x = textual description of the information concerning a spatial attribute
L = level of abstraction of x
K = main level of abstraction of x based on missing sources of information, currently $K(x) \in S$ as defined in section 4.2
U = degree of uncertainty connected to x
f = function that calculates L(x) considering dependencies between $K(x)$ and $U(x)$ where $K(x)$ is the dominant factor

However, such a formulation of "abstractness" raises a number of additional questions, e.g.:

1. Concerning *K(x):*
In the definition of the abstraction hierarchy, it is assumed that the textual description is the only currently available source of information, thus, the levels divide information into: (1) information that is directly available without further computations (2) information that is given in the source of information (the text) but needs at least one step of additional deduction/computation to be available (3) references to other sources of knowledge where the required information can be found (4) no information at all therefore it needs to be newly acquired. However, if more sources of information are assumed to be available that define the *context knowledge*, other factors might be influencing the degree of abstraction, e.g. the reliability of a source of information. Therefore, new main levels (combinations of main levels, see the following item) or sublevels (due to the connection of reliability to uncertainty) might have to be introduced in such situations.

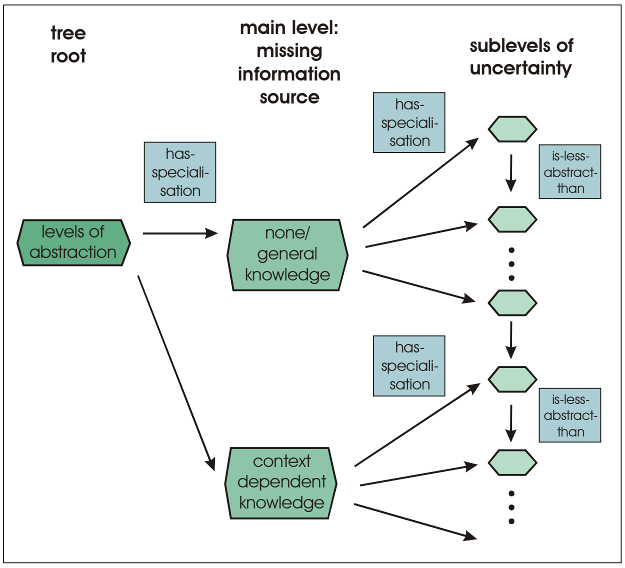

Figure 6. Parallel existence of a tree structure of abstraction levels based on “missing information sources” and uncertainty (arrows to the right) as well as a basically linear structure of abstractness (“down” arrows).

2. Concerning combinations of elements of *S*:
What happens if for one attribute of one boundary part several types of information are available in a text? A good example are descriptions where the degree-value as azimuth or bearing is mentioned. In most cases, they will include turning information (“left”/”right”), as well. The turning information seems redundant in this case because the direction given as azimuth or bearing is sufficient to resolve the required directional information. However, redundant information in cadastral texts can be used for confirmation purposes or as a basis for constraints in computations of unknown information. Therefore, if degree as well as turning information is available, it is certainly less abstract than degree information alone. Yet, more questions remain that need to be solved in the context of a definition of *f(K(x),U(x))*, e.g.: how can combinations of two or more *K(x)* be evaluated, especially if the dominating elements, i.e. less abstract description parts, are influenced by uncertainty? Does redundant information compensate uncertainty? How can they be compared?
3. Concerning *U(x)*:
Although, researchers in GIScience have made great advances in the last 15 years in defining, measuring, modelling, and visualising uncertainty (Harrower, 2003), the problem of assigning suitable quantitative values of uncertainty for each geospatial data is still unsolved. There are e.g. suggestions to use probability for managing errors and fuzzy set theory for vagueness (Fisher, 1999). However, although once the probabilities or fuzzy memberships are given, complete mathematical models to handle them are available, these probabilities or membership values are very often based on an externalisation of expert knowledge. Results from thorough cognitive analyses are rarely available, thus these values are often chosen completely arbitrary as Plewe admits for the case of membership values (Plewe, 2002). Therefore, questions arise such as: how can degrees of uncertainty be sensibly defined? How can uncertainty resulting from different causes be compared? For the examples of this paper, the classification of abstraction levels was based on the degree of freedom that remains for the choice of each spatial parameter e.g. the size of the resulting interval of possible direction values. In the case of “continue” and “turn left”/”turn right”, the differences in the sizes was obvious but is the size of the interval for “continue” smaller or equal than the size of the interval for “turn slightly left”? Although rich literature concerning qualitative spatial predicates already exist (Worboys et al., 2004, Klippel et al. 2005), more research needs to be carried out for a quantitative model.

5. CONCLUSIONS AND FUTURE DIRECTIONS

Due to missing standards concerning the preparation of texts for the Brazilian ownership cadastre, these documents exhibit a high variability in descriptive styles and are therefore an ideal source for an analysis of levels of detail in written spatial information. The most important insights from this analysis are that the texts are not only on different levels of abstraction but also that a whole text cannot be assigned solely to one single level of abstraction. In most cases, it is a mixture of different levels because usually each single spatial attribute (e.g. the length and direction of boundary parts) is described on different levels of abstraction. Hence, a very flexible model is required in order to explicitly represent the given spatial data of a text for transformations that solely rely on this description as a source of information. By introducing an abstraction hierarchy for each attribute based on two different criteria (“missing information sources”, uncertainty), the desired externalisation of the information content is accomplished while the criteria of flexibility is fulfilled by a qualitative, extendible approach. For a model that allows the computation of quantitative values of abstractness, more research needs to be carried out, especially in the field of uncertainty of spatial information and of computational models for cognitive concepts of spatial information. However, the context of this work does not require such a quantitative model and the results of an application of the presented approach for a transformation of textual descriptions into other types of representation (maps) will be presented in future publications.

REFERENCES

Bähr, H.-P., Philips, J., Jacomino, S., Müller, M., 2005. Daseinsvorsorge und Katastersubstanz in Brasilien – von verbaler Grenzbeschreibung zum Koordinatenkataster. zfv - *Zeitschrift für Geodäsie, Geoinformation und Landmanagement,* no.6/2005, pp. 387-393.

Bähr, H.-P., Müller, M., 2004. Graphics and language as complementary formal representations for geospatial descriptions. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science (IAPRSIS),* vol. XXXV, part B, comm. 5, ISSN 1682-1777, CDROM.

Egges, A., Nijholt, A., Nugues, P., 2001. CarSim: Automatic 3D Scene Generation of a Car Accident Description. Technical Report TR-CTIT-01-08, CTIT (Centre for Telematics and Information Technology, University Twente), 12 pp. http://doc.utwente.nl/fid/1253 (accessed Jan. 2006).

Ewald, K., Coors, V., 2005. Appraisal of standards for 3D city models. In: *Proceedings of the 9th international conference on information visualisation IV'05*, London, July 2005.

Fisher, P.F., 1999. Models of uncertainty in spatial data. In: Longley, P.A., Goodchild, M.F., Maguire, D.J., Rind, D.W. (eds.), *Geographical Information Systems: Principles, Techniques, Application & Management*, 2nd edition, John Wiley & Sons, New York, chapter 13, pp. 191-205.

Fontaine, S., Michel, D., 1999. The production of route instructions in underground and urban environments. In: Freksa, C., Mark, D.M. (eds.), *Spatial Information Theory – Cognitive and Computational Foundations of Geographic Information Science (COSIT 1999)*, Springer, Berlin, pp. 83-94.

Frank, A.U., 2001. Tiers of ontology and consistency constraints in geographical information systems. International Journal of Geographical Information Science, 15(7), pp. 667-678.

Frank, A.U., 2003. Ontology for spatio-temporal databases. In: Koubarakis. M. et al. (eds.) Spatio-Temporal Databases : The CHOROCHRONOS Approach, Springer, Lecture Notes in Computer Science, vol. 2520, chapt. 2, pp. 11-81.

Harrower, M., 2003. Representing uncertainty: does it help people make better decisions? Invited white paper of UCGIS Workshop: Geospatial Visualization and Knowledge Discovery Workshop, Landsdowne, VA, USA, Nov 18-20. http://www.ucgis.org/Visualization/whitepapers/Harrower.pdf (accessed Jan. 2006)

Klippel, A., Tappe, H., Kulik, L., Lee, P.U., 2005. Wayfinding chorems – a language for modelling conceptual route knowledge. Journal of Visual Languages and Computing, 16 (2005), pp. 311-329.

Kuipers, B., 2000. The spatial semantic hierarchy. *Artifical Intelligence*, 119, pp. 191-233.

Leidner, J.L., Sinclair, G., Webber, B., 2003. Grounding spatial named entities for information extraction and question answering. In: Proceedings of the HLT_NAACL 2003 Workshop: Analysis of Geographic References, Edmonton, Canada, May/June 2003, pp. 31-38.

Liedtke, C.-E., Bückner, J., Pahl, M., Stahlhut, O., 2001. Knowledge based system for the interpretation of complex scenes. In: Baltsavias, E., Gruen, A., van Gool, L. (eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III),* Balkema, Lisse, pp. 3-12.

Mark, M., Freksa, Chr., Hirtle, St.C., Lloyd, R., Tversky, B., 1999. Cognitive models of geographical space. In: International Journal of Geographical Information Science, 13(8), pp. 747-774.

Plewe, B., 2002. The nature of uncertainty in historical geographic information. *Transactions in GIS*, 6(4), pp. 431-456.

Protégé, 2006. Homepage of the ontology editor Protégé at http://protege.stanford.edu (accessed Jan. 2006).

Schade, U., Frey, M., 2004. Beyond information extraction: the role of ontology in military report processing. In: Buchberger, E. (ed.), *Proceedings of KONVENS 2004* (= Schriftenreihe der österreichischen Gesellschaft für Artificial Intelligence, Band 5), Sept. 2004, Vienna, Austria. http://cogprints.org/3895 (accesssed Jan. 2006)

Streeter, L.A., Vitello, D., Wonsiewicz, S.A., 1985. How to tell people where to go: comparing navigational aids. In: International Journal on Man-Machine Studies (1985) 22, pp. 549-562.

Tom, A., Denis, M., 2003. Referring to landmark or street information in route direction: what difference does it make? In: Kuhn, W., Worboys, M.F., Timpf, S.(eds.) *Spatial Information Theory – Foundations of Geographic Information Science (COSIT 2003)*, Springer Berlin, pp. 362-374.

Tversky, B., Lee, P.U., 1999. Pictorial and verbal tools for conveying routes. In: Freksa, C., Mark, D.M. (eds.), *Spatial Information Theory – Cognitive and Computational Foundations of Geographic Information Science (COSIT 1999)*, Springer, Berlin, pp. 51-64.

UN-ECE (United Nations - Economic Commission for Europe), 1996. Land administration guidelines – with special reference to the countries in transition - Glossary. United Nations Publication, New York and Geneva, ISBN 92-1-116644-6. http://www.unece.org/env/hs/wpla/documentation/laguidelines.html (accessed Jan. 2006)

Worboys, M., Duckham, M., Kulik, L., 2004. Commonsense notions of proximity and direction in environmental space. In: Spatial Cognition and Computation, 4(4), pp. 285-312.

ACKNOWLEDGEMENTS

The author is grateful for the constructive criticism provided by the anonymous reviewers. The work of the presented project is supported by the Deutsche Forschungsgemeinschaft (DFG), project no. BA 686/16.