Dialogue Management For Route Descriptions

Vincent Maya[1]

Abstract. A dialogue system within a geographical information system has to understand and to generate texts in natural language with geographical information. In particular, it applies spatial reasoning in order to identify or to localize the landmarks of the geographical environment. Therefore, the domain knowledge manager (i.e. the module that gathers information) has to access the contents of the geographical database, which represent the information usable by the GIS. Owing to the lack of a standard format, this manager has to deal with various contents and various organizations.

In this paper, we describe the framework of such a dialogue system, the use of a *pivot system* within the domain knowledge manager, the natural language processing component and the solutions that enable to overcome the various difficulties and to differentiate precisely the cases for which the dialogue system has to interrogate the user.

Keywords: Natural Language Processing, Spatial Reasoning, Knowledge Representation

1 Introduction

Different research teams from our laboratory have developed a Geographical Information System (GIS) based on commercial geographical databases describing the geographical environment of the campus of our university and of the city nearby (e.g. the architectural environment, the road network or the railway information). This system allows, among other things, to visualize the environment, to calculate a route between two points in the road network and to identify objects of the environment.

Creating a dialogue system within this GIS requires a domain knowledge manager that enables to manipulate information provided by the geographical database of the GIS. We use a *pivot system* (cf. [6]), which is able to overcome the numerous difficulties implied by the processing of natural language from a database.

In section 2, we present the influence of our framework on the types of text that this system is able to process. In section 3, we describe the organization of the dialogue manager. In section 4, we describe the geographical domain knowledge manager and the pivot system on which it is based. In section 5, we present the funding of the understanding and the generation modules, and their interaction with the dialogue manager.

2 Natural language processing in a GIS

The characteristics of the GIS and of the chosen environment limit the type of the processed texts. These generated texts are either instructions (i.e. descriptions of routes generated by a GIS and not descriptions of a route already followed by the user), or questions. The aim of these instructions is to lead an agent to a final destination or to answer a question of the user about the environment. The aim of the questions is to obtain supplementary information from the user, about his / her position or about his / her preferences.

Symmetrically, the texts provided by the user are requests for geographical information (such as ”*Where is the town hall?*” or ”*Is there a hospital in this town?*”), about a route (such as or ”*How do I go from the train station to the Post office?*”) or answers to the system.

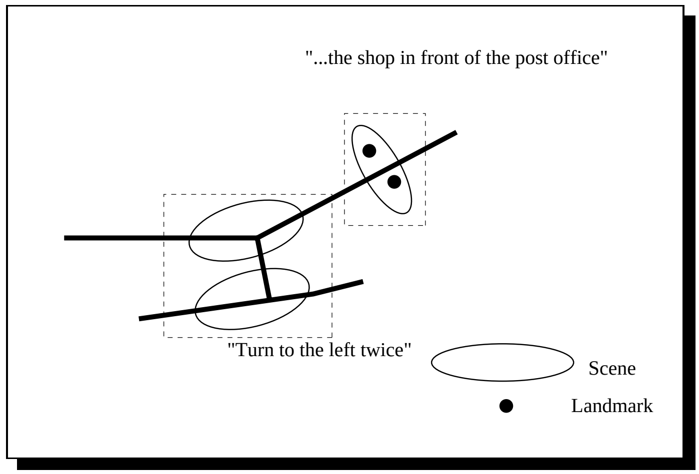

Figure 1. scene in terms of actions and landmarks

Text components usable for understanding may be divided into two groups: landmarks and actions (cf. [3, 8]). In order to avoid ambiguities, the term *landmark*, in this paper, is used only for the references in natural language to elements of the geographical environment.

The actions specify the position of the agent according to the landmarks. These landmarks are explicit (such as in “*cross the river*”), or implicit (such as in “*Turn to the right*”). In an urban environment, this last sentence means that we have to consider the next street section that has a common intersection with the current street and that is oriented to the right.

Among the actions, we are mainly interested in movements such as direction changes, crossing and positioning (e.g. moves according to a direction such as in ”*turn to the right of ...*”, moves toward a landmark or toward a cardinal direction or moves along a landmark such as in ”*pass behind...*”). We note that the actions do not correspond necessarily to verbs but also to prepositions or to nouns (cf. [16, 14]).

In our framework, the representation of the text in terms of land-

[1] Université de Paris-Sud, LIMSI-CNRS, 91403 Orsay France

marks and actions is not sufficient, since the actions have no equivalent in the database and actions and several landmarks may be used in order to identify one place, (such as in "*...the shop in front of the post office*", figure 1). Moreover, landmarks are frequently associated in same sentence and some landmarks are used in order to localize implicitly other landmarks.

Therefore, we define scenes as *places* of an action. In a given environment, the scene is focused on the landmark type that is associated to the agent position. The choice of this type depends on the database content. For databases representing an urban environment with its road network, each position is associated to a street section, and, thus, the scenes are focused on these sections. For a database without a road network, the scenes may be focused on the buildings.

According to this definition, a route, which is an association of landmarks and actions (which themselves correspond to one or several scenes, such as in "*Turn to the left twice*", figure 1), corresponds to a sequence of scenes. Consequently, a route description corresponds to a description of some scenes of this sequence, the choice of these scenes varying according, among other things, to the length of the road, to the user characteristics, to the language, and to the knowledge of the narrator (cf. figure 2).

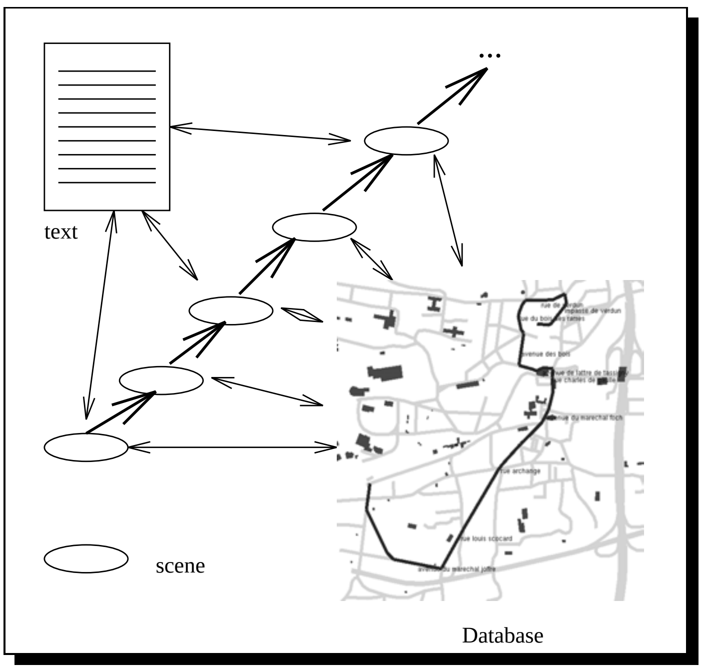

Figure 2. Links between texts, scenes and database contents

3 The dialogue manager

The aim of our dialogue manager is to process scene descriptions or descriptions of a set of scenes in order to describe a route according to the user instructions. The dialogue has to allow to overcome the semantical differences between the information available from the GIS and the one expressed by the user, but also to identify and, if possible, to overcome the differences of knowledge, such as the lack of information in the database content.

Several authors propose models of dialogue (cf. [13, 4]). In our framework, the characteristics of the GIS and of the chosen environment, which limit the interaction between the user and the system limit the required functionalities of the dialogue manager. Therefore, we decompose the dialogue manager into three modules (cf. figure 3) : the understanding module, the generation module and the application manager. The module of understanding is composed of the processing module for natural language interpretation (Interpreter) and the module of access to the domain knowledge (domain knowledge manager). The generation module is composed of a module of natural language generation (Generator), and of the domain knowledge manager. The aim of the application manager is to return information calculated from the application of the GIS. For example, for a user request to have a route description between two points, the application that calculates the route is external to the dialogue manager, and in order to access this information, the dialogue manager uses the application manager.

A dialogue manager handles the dialogue with the user and keeps track of the interaction. In our system, the dialogue manager groups all the other modules. For the author of [7], managing the interaction with the user involves deciding whether a user request is clear enough to access the application or to initiate a sub-dialogue. But, for us, this definition limits the role of the dialogue manager : for a given request type, this manager has to enable the processing of some requests for which the database content is insufficient. In order to do that, it has to indicate the cause of the problem.

3.1 The domain knowledge manager in the understanding module

Firstly, the module of understanding has to analyze the natural language text. The interpreter analyzes syntactically and semantically the user input and delivers a meaning representation. At the level of natural language, we have to take into account the diversity of descriptions that human beings can make. This diversity proceeds from the many possible choices of landmarks and of narrative styles.

From this representation focused on the quoted landmarks and actions, the domain knowledge manager has to identify the scenes described by the text. This identification is made for explicitly quoted landmarks, according to their nature and to their attributes or to other spatial information, but also for the ones that belong to scenes described with actions (i.e. positioning, changes in direction or moves).

The domain knowledge manager has to deal with vague notions, uncertainties, ambiguities, redundancies, and implicit information contained in the texts. Therefore, the reasoning about spatial indications has to use cognitive, linguistic and pragmatic knowledge. We distinguish the knowledge explicitly contained in the databases, and the one that must be inferred from the database contents.

3.2 The domain knowledge module for generation

The module of generation enables us to obtain, from elements of the database, texts with geographical components. For example, from a description of a route in terms of objects of the database, calculated by the GIS, we want to obtain an equivalent description in natural language, easily understandable by a GIS user.

The generator processes texts, according to a narration strategy and according to the database content. The aim is not to use all the available information, but the part of which the user needs.

In order to retrieve the available information about a given scene (i.e. about the object on which the scene is focused), the module of

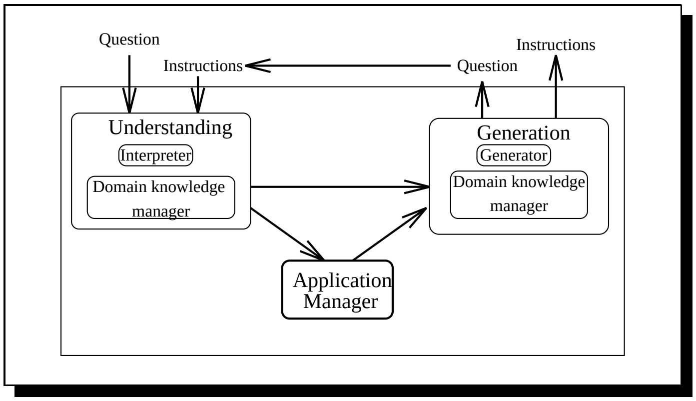

Figure 3. Dialogue manager

generation also uses the domain knowledge manager. We describe the role of this manager in the section 5.2.

4 The geographical domain knowledge manager and pivot system

The domain knowledge manager has to manipulate and to correlate information of, at least, two sources of information: natural language and geographical databases. These databases contain information about some elements of a spatial environment such as an urban environment, a forestry environment or a museum, in terms of rooms, communication between rooms and contained items. They describe their geographical aspects but also their nature and their various attributes.

The difficulties are derived mainly from the opposition between the quantitative nature of the data in the base (such as spatial coordinates) and the qualitative nature of human discourse (ambiguities and undetermined knowledge, such as vague temporal or spatial expressions) and from the various possibilities for a human being to choose and to name landmarks.

Furthermore, there is no standard norm for the databases and, hence, it is not possible to know beforehand either the contents of a database or the distribution of these contents : the database may have no representation of a type of landmark or, if this representation exists, no representation of some elements of this type or no representation of some characteristics of these elements.

Moreover, the organization of the data may lead to a non-pragmatic distribution : elements, which differ according a pragmatic or a cognitive point of view, may be grouped, and conversely, elements of an identical nature may be scattered for technical reasons.

In order to overcome potential semantical differences between information expressed by the user and the information contained by a database, we use a pivot system (described in [14]), based on conceptual graphs (cf. [17, 2, 1]). This ontology-based system introduces a conceptual level that allows to correlate the information of the texts and of the database. Since the database contains only information about landmarks and relations between landmarks, this conceptual level contains three kinds of knowledge: information about the landmarks usable in natural language (in descriptions of routes or of scenes), information related to the organization and the contents of the database, and finally information about the correspondence between objects and landmarks.

At this level, the information about landmarks is defined by the domain ontology of the environment considered (cf. [11, 12, 9] for a definition of ontologies, and [14] for the design of the domain ontology). This domain ontology describes, in terms of concepts, all the elements of an environment type (landmarks in two dimensions or in three dimensions, such as respectively, for an urban environment, roads and buildings) and all information about them (i.e. relations between landmarks and attributes of landmarks).

These conceptual relations, such as relations that express, for example, proximity, adjacency, topological relations or distance relations, do not correspond necessarily either to natural language relations or to database relations. They are links between information in NL (expressed by prepositions, but also, by propositions or by verbs) and information of the database (not necessarily expressed by database relations but also retrievable from coordinates or from other attributes). The choice of the conceptual relations varies according to the environment. Among these conceptual relations, there are all those that can be inferred from landmark coordinates, and those that are frequently used in route descriptions (cf. [3, 10]).

The concepts that represent landmarks correspond to classes of *strict synonyms*: ”*Mairie*” and ”*Hôtel de ville*”, translations of ”*Town hall*”, have the same meaning and, consequently, correspond to the same concept. The hierarchy between these concepts, expressed by a type lattice, is defined from the conventional meaning of the landmarks (e.g. a bridge or a tunnel are particular types of street sections).

Information about the organization of the database and information expressed by the domain ontology enable us to link the types of object to natural language terms, and thus to store meta knowledge. Since we assure that we have the complete set of the elements of the environment, all the object types are linked to a concept that represents a landmark. Conversely, if a landmark concept is not linked to one or several object types, we can deduce that the domain knowledge does not contain information about this landmark. Thus we can define the existence or the non-existence of the explicit information related to a type of landmark in the database: a landmark has a representation in the database if and only if its representation defined from the domain ontology is linked to one or several concepts that represent the content of the database. That does not mean that a corresponding object exists, only that it may exist.

The existence or the non-existence of links from concepts that represent landmarks defines also a level of precision on the contents of the database. For a database containing a representation of all the existent landmarks and defining their nature, the level of precision corresponds to the level of precision of these concepts (e.g. a town-hall is a particular administrative building, which is a particular building, which is a particular architectural element, which is a particular landmark). But if the database does not distinguish a town hall from other buildings, ”*Town-hall*” is, by extension, conceptually equivalent to ”*Building*”. For generation, this enables the system to obtain automatically the most adequate terms for describing an object, and for understanding, to identify the classes that may contain a representation of a landmark, without taking into account the level of precision of the database.

Likewise, this system enables us to express inferable information. By linking concepts that represent two landmarks, we can express equivalence between landmarks from a reasoning viewpoint. For example, in a database that has no representation of the hydrographical network, the concept *River* may be associated to the concpet *bridge*, the presence of a bridge indicating the presence of a river. By linking a landmark concept to an attribute concept, we can express equivalence between a landmark and other landmarks that have this attribute. For example, in an urban environment, for a one-way street, we can infer a one-way sign at one end of the street section.

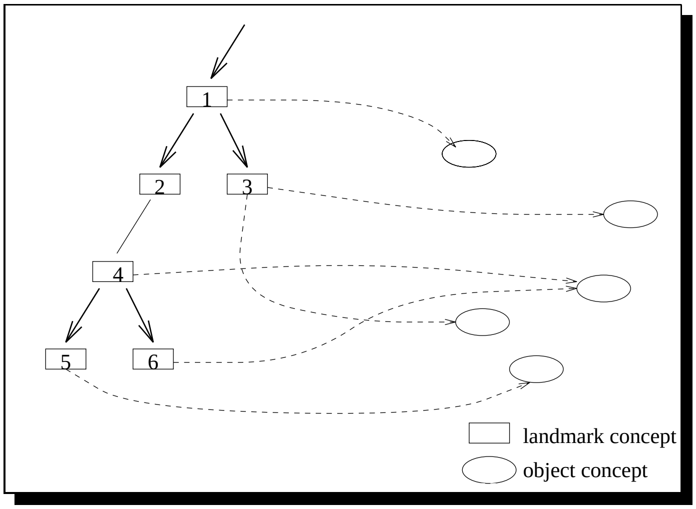

Figure 4. Association between landmarks and objects

By classifying the data according to a cognitive point of view and by linking them to landmarks, we no longer have to take into account the problems of the non-cognitive scattering or the "*abusive*" regrouping of information (such as, in figure 4, landmark concept #3, which is linked to several object classes, or concepts #4 and #6, which are linked to the same class). In the case of an object class that represents different landmarks (e.g. a type which groups the urban structures, such as buildings, squares, road stations or industrial areas) the concept that represents this type is linked to several concepts that represent landmarks. In the case of landmarks of the same nature represented by several classes (e.g. car parks divided between two types, according to their area), the concept that represents these landmarks is linked to several concepts that represent classes of the database.

The database information is expressed in terms of attributes and relations defined in the domain ontology. In order to retrieve the information related to these relations and these attributes, the pivot system returns mapping relations (cf. [9]).

5 Management of the modules

5.1 Understanding

Our aim is, from a route description, from a position description, or from a user request (equivalent to a set of scenes), to find the objects of the database related to the scenes. This means we have to interpret correctly the successive instructions that constitute the description or the question, in spite of ambiguities, imprecision, and implicit information. For example, from the question "*How do I go from the train station to the Post office?*", the understanding module has to identify the scenes corresponding to the train station and to the post office. In a city that contains more than one post office, it has to retrieve, according to the context (i.e. the position and the orientation of the agent and the local environment), the nearest one. From the scenes, the module retrieves the landmarks on which they are focused, or the ones that the application needs (thanks to the application manager).

For *yes-no* questions, the system has to determine whether the scene exists or not in terms of database objects.

Firstly, the system identifies objects of the database, if they exist, from the terms that describe the landmark. From the list of objects obtained thanks to the mapping relations, we have to identify the one that corresponds to the landmark. Information about its attributes, if they are referenced in the pivot system, enables us to filter the list of objects that may correspond to the landmark. Furthermore, relations of the domain ontology that are identified in the text enable the system to filter the list, using spatial reasoning about the associated mapping relations. We have to consider the context: the system takes into account geographical and spatial considerations, pragmatic and cognitive knowledge and linguistic aspects.

After this filtering, if the information contained in the sentence is sufficient to identify the scene focused on the quoted landmark, the list contains a single object. This includes the cases where the database contains a single object of a given type (e.g. with a database which has a single object which corresponds to a bridge, all mentions to a bridge are identifiable as referring to this object). This includes also the cases where the landmark is identified indirectly by another landmark. For example, for processing the sentence "*pass the library that is in front of the square.*", it is not necessary to identify both the library and the square, a single one being sufficient to identify the scene. We use for this case the term of redundancy (according the database contents).

At this stage, the identification of a scene may fail, that is the object list that may represent the landmark on which the scene is focused is empty or contains more than one item. The causes of failure are :

- missing information in the text : the information about a scene is under-specified and the list contains more than one object (i.e. many objects that can represent the landmark), either because of implicit information that a human being can easily understand but the domain knowledge manager cannot (e.g. "*pass the bridge*", which needs to take into consideration the current position of the mobile agent and information on his perceptual area) or because of missing information according to the database point of view (e.g. "*behind the red house...*" with a database that has no information about the colors of the buildings);
- missing information in the database : the list does not contain any object, the database does not contain, for example, the representation of type of landmark (e.g. "*At the Campus entrance*", with a database that has no representation of the campus outskirts);
- impossibility to access an identified landmark : route error during generation or understanding.

The role of the dialogue manager is to identify the failures and their causes and to generate questions in order to continue the processing.

5.2 Generation

The aim of the generator is to choose, according to narration strategies, the scenes that it describes among all those that composed, for example, the route description, and to choose the way to describe these scenes, again according to narration strategies. These narration strategies take into account the contents of the database and the user preferences.

The available information is stored in an intermediate landmark graph. This representation corresponds to a filtering of the scenes that are focused on the database objects that compose the route (i.e. the available information about the objects that belong to the scene). To this end, we obtain on the one hand the potential attributes of the object on which the scene is focused (i.e. a street section for a database which represents an urban environment and its road network), and on the other hand, the objects that, by construction, are associated to it. This information is enriched with knowledge that can be inferred from the contents of the database (e.g. spatial relations computed from spatial coordinates), and with relations between scenes, i.e. numerical and directional information that indicates the way to go to the following scene and that can be exploited in description such as "*Turn twice to the right*" or "*Turn to the right after the second street light*".

Following Denis, in [3], we consider two kinds of “interesting” scenes : the “semantically” interesting ones and the ones that correspond to changes in the environment (direction changes or street changes). The semantic interest is defined with regard to the pivot system. In fact, the pivot system enables us to use automatically the database level of precision in order to describe any object in terms of natural language. In a broader sense, scenes with semantic interest are the ones that are linked to landmarks with semantic interest.

For question generation, the strategy varies according the problem type (cf. [15]).

- insufficient information : ask for supplementary information about the landmark (i.e. about conceptual attributes, retrievable from the database contents or about conceptual relations with others landmarks).
- missing information : ask for supplementary information about a new landmark or about appropriate actions, according the current position of the agent (i.e. the last identified position).
- incoherent information : verification of the information provided by the user. The system expresses the scene that it has to retrieve, and asks for confirmation. If this is not enough to identify the error, the process stops.

6 Conclusion and perspectives

In this paper, we presented a system for natural language processing that enables to deal simultaneously with geographical information and with geographical databases. We have introduced, in particular, the notion of scene that associates, according to natural language viewpoint and to a geographical database viewpoint, the interpretation of landmarks and actions, which compose route descriptions.

The main advantages of our approach is that the domain knowledge manager, as well as the dialogue manager, are reusable for various database types and for various environment types.

At the practical level, the understanding and the generation modules are have been implemented (for texts in French) as well as the domain knowledge manager and the pivot system.

In this framework, we have considered only spatial information. But a dialogue system based on a GIS could also use temporal information, such as bus time-tables or information from spatio-temporal databases (cf. [5]).

REFERENCES

[1] C. Boksenbaum, B. Carbonneill, O. Haemmerlé, and T. Libourel, ‘Conceptual graphs for relational databases’, in *Conceptual Graphs for Knowledge Representations*, ed., Mineau et al., LNAI n 699, pp. 142–161, Quebec, Canada, (Aout 93). ICCS’93, Springer Verlag.

[2] M. Chein and M.-L. Mugnier, ‘Conceptual graphs: Fundamental notions’, *Revue d’intelligence artificielle*, **6**(4), 365–406, (1992).

[3] M. Denis, ‘The description of routes : A cognitive approach to the production of spatial discourse’, in *Cahiers de Psychologie Cognitive / Current Psychology of Cognition*, pp. 409–458, (1997).

[4] J. Allen D.J. Litman, ‘A paln recognition model for subdialogues in conversations’, *Cognitive Sciences*, **11**, 163–200, (1987).

[5] C. du Mouza and P. Rigaux, ‘Bases de données spatio-temporelles’, in *Le Temps, l’Espace et l’Évolutif en Sciences du Traitemenet de l’Information*, eds., H. Prade, R. Jeansoulin, and C Garblay, volume 2 of *Ecole thématique "Document & Évolution" du GROUPEMENT DE RECHERCHE "Information, Interaction, Intelligence"*, CÉPADUÈS-ÉDITIONS, Toulouse, (Septembre 2000).

[6] G. Edwards, G. Ligozat, A. Gryl, L. Fraczak, B. Moulin, and C. M. Gold, ‘A voronoï-based pivot representation of spatial concepts ans its application to route descriptions expressed in natural language’, in 7^{th} *International Symposium on Spatial Handling*, Delf, The Netherlands, (1996).

[7] A. Flycht-Eriksson, ‘A domain knowledge manager for dialogue system’, in *ECAI 2000*, (2000).

[8] L. Fraczak, *Description d’itinéraires : de la référence au texte*, Phd thesis, Université Paris XI, Septembre 1998.

[9] J. H. Gennari, S. W. Tu, T. E. RothenFluh, and M. A. Musen, ‘Mapping domains to methods in support of reuse’, *International Journal of Human and Computer Studies*, **41**, 399–424, (1994).

[10] A. Gryl, *Analyse et modélisation des processus discursifs mis en oeuvre dans la description d’itinéraires*, Phd thesis, Université Paris VI, France, 1995.

[11] N. Guarino, ‘Understanding, building and using ontologies : A commentary to *using explicit ontologies in KBS development*, by Van Heijst, Schreiber and Wielinga’, *International Journal of Human and Computer Studies*, **46**, 293–310, (1997).

[12] N. Guarino, ‘Formal ontology and informations systems’, in *Formal Ontology in Informations Systems*, ed., N. Guarino, IOS Press, Amsterdam, (1998).

[13] D.J. Litman, *Plan Recognition and Discourses Analysis : An Integated Approach for Understanding dialogues*, Phd thesis, University of Rochester, 1985.

[14] V. Maya, *Accès aux bases de données géographiques en langage naturel : Applications aux descriptions d’itinéraires*, Phd thesis, Université Paris XI, December 2001.

[15] P. Nerzic, *Erreurs et échecs dans le dialogue oral homme-machine. Détéection et réparation*, phd thesis, Université Rennes 1, January 1993.

[16] F. Rastier, M. Cavazza, and A. Abeillé, *Sémantique pour l’Analyse: De la Linguistique à l’Informatique*, Sciences cognitives, MASSON, 1994.

[17] J.F. Sowa, *Conceptual Structures - Information Processing in Mind and Machine*, Addison-Wesley, 1984.