Voice knowledge acquisition system for the management of cultural heritage

Stefan du Château, Danielle Boulanger and Eunika Mercier-Laurent

MODEME, IAE Research Center Lyon University 6, av Albert Thomas F-69008 Lyon

Abstract: This document presents our work on a definition and experimentation of a voice interface for cultural heritage inventory. This hybrid system includes signal processing, natural language techniques and knowledge modeling for future retrieval. We discuss the first results and give some points on future work.

1   Introduction

The inventory of the cultural heritage includes several tasks such as the study, analysis, description of the masterpieces still existing, preserved as vestiges, destroyed or disappeared but known through documents (Verdier and al ., 1999).

All categories of masterpieces are concerned, such as religious, civil, military, in a perimeter as large as are the human activities.

The work of the researchers of the inventory consists partially in collecting the on field information available in the cities, villages and specific places. It can take the form of text files, pictures, drawings, video, or plans. Researchers can also conduct a study about a given place or topic before collecting.

The study documents and the collected information are registered in a data base which could be general or personal.

When back to the office the researcher can improve the gathered information on a given object or add some elements from archives to update the content of the data base. Collecting of information into paper files or recording them on a laptop is demanding and time consuming. The amount of the completed and corrected information is still very large and heterogeneous.

Each masterpiece has its own history, past and present, it can be moved from one historical context to the other or it can be modified. In brief, it can change its spatiotemporal context. This kind of information and related knowledge is impossible to represent just in a classic data base.

Our objective is to design and experiment a new collecting support system to help the cultural heritage inventory researchers to perform their work better and in a more efficient way. It is also to help indexing and retrieving information and

knowledge on a given masterpiece and its context. This paper describes our work on such a hybrid system using a voice interface (signal processing), natural language processing and knowledge modeling for the information gathering, management and retrieval. Sect. 2 presents relative work, Sect. 3 describes our work and in Sect. 4 we discuss directions for future work.

2   State of the art knowledge modeling for cultural heritage

2.1 The inventory of the cultural heritage

Most projects in the field of the inventory of the cultural heritage are based on data bases systems. The structures of these data bases, which correspond with well defined specifications on a given application, are not easily extensible. The relative lack of flexibility of these systems makes them incompatible with the notion of knowledge based systems, which are flexible.

The existing systems contain a lot of incompatible data recorded in various data bases using several languages. In such situations an intelligent system with the ability to manage this huge amount of data effectively will be very useful.

Among the different European projects we can quote MICHAEL[1] the purpose of which is to valorize Europe's cultural heritage. This project provides a multilingual interface to encourage the interoperability of different national heritage data bases.

Other projects such as HEDD[2], a project of the English Heritage committee brings together 22 museums, 3 libraries and deposits of archives and uses ontologies to model common knowledge distributed in heterogeneous data. .

Other projects such as those of the national Gallery of Finland[3], the University of Queensland in Australia[4] and SCULPTOR[5], use the ontology CIDOC-Conceptual Reference Model[6] (CRM) (Doerr, 2006) as a tool for knowledge modeling. They unite big galleries and European cultural institutions.

---

[1] http://www.michael-culture.eu/project.html
[2] http://www.fish-forum.info
[3] http://www.fng.fi/fng/rootnew/en/vtm/etusivu.htm
[4] http://www.metadata.net/harmony/MW2002_paper.pdf
[5] http://www.sculpteurweb.org/html/approach.htm
[6] http://cidoc.ics.forth.gr/index.html

2.2 Automatic knowledge acquisition and speech apprehension

Since the beginning of artificial intelligence many researchers have been working on the topics such as automatic knowledge acquisition and speech apprehension, mainly using signal processing techniques. The first voice interface was probably this of a workstation called Buroviseur, built in INRIA in 1981 (Kayak, 1982 Mercier-Laurent, 1980). The voice interface was also used for knowledge acquisition for expert systems (Balaram, 1988) or for human-machine dialog in machine learning systems (Michalski, 1985). This technology is now mature and can be integrated in applications using a large vocabulary with more than 60 000 words (Haton, 2006)

The quality of voice acquisition systems depends of many parameters such as external acoustic environment (noisy or silent) and the quality of the equipment employed.

The main specialists of the field state that these performances provide 90 % of a correct recognition (Veronis 2000).

This performance can seem insufficient in a system with full automatic transcription; however it is acceptable in the half automatic system, where the results are validated by an expert, especially when it is a question of not validating the whole of the re-transcribed text, but only a part corresponding to the predefined information.

The outlines of the extraction of information systems were defined during several *Message Understanding Conference* (MUC) conferences, which took place between 1987 and 1998. It can be said that between the first conference in 1987 and the last one in 1998, the initial ambitions - the understanding of a text by computers - were revised to finally become systems of information extraction.

The goal of the information extraction is to produce a structured representation of unstructured texts by searching for given patterns in the texts which are relevant to an application (Ibekwe-SanJuan, 2007). Basically, we are looking for the partners of a transaction, the names of the bodies, the transactions (sale, purchase) etc.

The text mining information extraction systems are based mainly on two technologies: the one uses automatic learning, the other one uses natural language processing (NLP).

The techniques of machine learning provide the possibility to automatically extract dictionaries and specialized grammar, as well as annotations. They allow reducing the time needed to construct linguistic resources. Their main disadvantage is that they need an important text corpus for each application domain. Information extraction techniques based on NLP use morphosyntactic analyses of text documents. This technique splits a text in sentences and terms. The tagging is based on external resources and grammar, defined by the user for a given field.

While the voice recognition and text mining are not new, the association of both is, based on our knowledge, not really deployed. Our work links these two domains and applies them to knowledge modeling.

There are only a few publications on knowledge modeling in the field of cultural heritage. The main known contribution is the domain ontology CIDOC-CRM, based on object knowledge representation which is flexible and convertible into various formats such as **RDF**, **XML**, **DAML+OIL**, **OWL**. The CRM covers all information required for the scientific documentation in the field of cultural heritage.

In terms of concepts and relations between concepts and the construction of ontologies, text mining has been described in numerous works. Among them we quote Charlet (Charlet, 2002) and Bourigault (Bourigault, 2003), who work on the construction of ontologies from texts in the medical domain.

3   Our work

Our voice acquisition system is presented in Fig 1.
It follows four steps:
  1. Voice acquisition of a given masterpiece description
  2. Automatic transcription of the voice file into the text file by Dragon[7],
  3. Extraction of concepts and relations between concepts
  4. Validation of the extracted concepts found in the previous stage by expert.

The validated descriptors are registered in a data base and will be used to update the existing ontologies. The acquired voice information is distributed in fields of the data base such as: DENOMINATION, CATEGORY, MATERIAL, DESCRIPTION, and INSCRIPTION without constraining the speaker to say the name of the descriptive field. These fields constitute the descriptive system defined by the heritage inventory department (Verdier and al ., 1999). Some of these fields are compulsory, the others optional. The contents of certain fields are defined by a lexicon, the contents of the other fields remaining free.
Usually the acquisition of the data is made using a keyboard and needs to strictly respect a data acquisition model. In the case of a voice acquisition, there is no structure to guide the acquisition. The involved people are specialists in the given field, thus we can expect a coherent and well structured text.
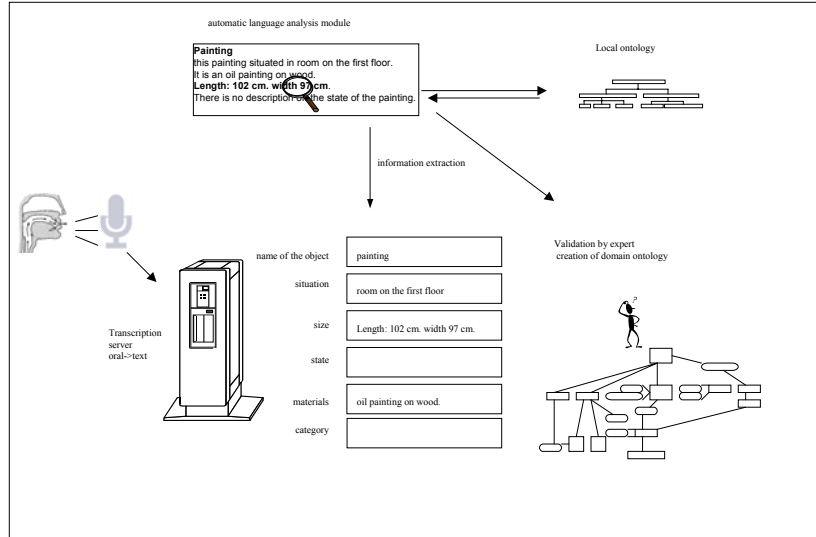
---

[7] http://www.nuance.fr/naturallyspeaking/

Fig. 1 Voice acquisition system helping assistant in knowledge acquisition**.**

3.1 *Robust syntactic analysis*

Despite the good performances of the re-transcription software, some syntactical and semantic errors can occur in the re-transcribed files. The origin of its errors can be directly connected to the way the speaker dictates the text (waiting time, hesitation, back on sentences or words). The transcription process itself may also cause errors.

We started the acquisition without any text archives what made the applying of machine learning methods impossible. We have chosen the robust incremental syntactic analyzer (Hagège and al., 2003). Such an analyzer always insures good results even with a badly structured or erroneous input text.

Incremental means that the rules of disambiguation, category, construction of constituents and the extraction of syntactical dependencies are applied one after the other. The specific and reliable rules are first to filter the rare or exceptional configurations, while the more general rules are at the end of grammar. (Hagège and al., 2003).

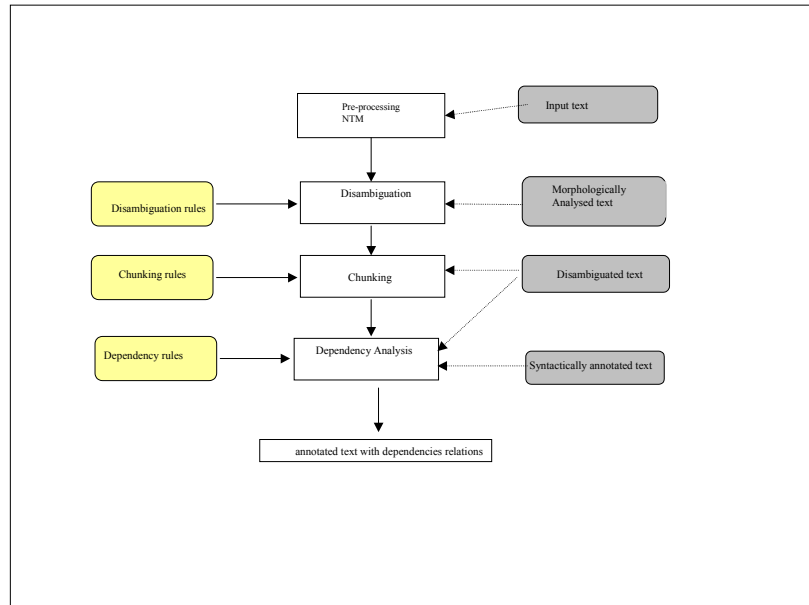For our experiments we use the XIP[8] analyzer created by XRCE[9], whose architecture is presented in Fig 2.

Fig. 2 XIP Architecture

Legend: NTM[10]; Disambiguation[11]; Segmentation[12]; Analysis of the dependences[13].

---

[8] XIP (Xerox Incremental Parser) by AïtMokhtar, Chanod et Roux .

[9] Xerox Research Center Europe

[10] Normalization(Standardization), Tokenisation (Division of a text in words), Morpholgy

[11] Disambiguation: Disambiguation of the categories of the words according to their context of appearance

[12] Segmentation: cut the serial linguistic units (unities) of constituents nucleus(kernel)

[13] Analysis of the dependency: identify syntactical links between the words

3.2 From the data entry form to the extraction patterns[14]

As we mentioned before, the information to find is defined by the descriptive system of the inventory (Verdier and al 1999). It indicates the type of information to be looked for, but also controls, in certain cases, the vocabulary to be used. The terms have to correspond with the entry of a lexicon.

The descriptive system of the inventory will partially guide the conception of the extractions patterns and local grammar.

The collected information on the field can be split in two categories:
- Physical aspects: material of manufacturing, structure, place.
- All the information relative to the historical, social, ethnographical context.

It is the type of information that can be known only by experts of a given domain. Our system of extraction of information has to be able to take it into account.

Two scenarios are possible:
1. The result corresponds exactly to a defined entry of a lexicon. In this case the local grammar must be defined to insure that the analysis and the result of extraction is a word or a constituent, which corresponds exactly to an entry of this lexicon.
2. The result is an incomplete description of a given place, for example:

« *... le retable comprend 4 tableaux : Baptême du Christ, Christ au Jardin des oliviers, la Cène et la Résurrection... »*.

The constituent « *Baptême du Christ »* will be tracked down in the text without problem because it exists in the lexicon, then thanks to an analysis of dependence; it can be associated with the representation. The constituent *« Christ au Jardin des oliviers »* will not be recognized as representation because it does not exist in this lexicon. The system has to be able to recognize this entry as a constituent, and to suggest it as a possible entry. A local syntactic analysis must be triggered by one of the words of the constituent because they belong to the lexicon, or because the sentence contains a word or a constituent which is associated with the idea of the representation: *the representation, are represented*.

In our example, the constituent « *Jardin des oliviers »* and the word « *Christ »* exist separately in the lexicon representation, which is the condition to propose the constituent *Christ in the Garden of olive trees* as a possible descriptor of the representation. According to the principle of relations « sort of » the representation of the « *Christ au Jardin des oliviers » is* a specific case of a representation of Christ.

The identification of the words or the constituents is not the only difficulty, which we have to face. The language of the cultural heritage is extremely rich and

---

[14] Extraction pattern: set(group) rules allowing to identify the expected, relevant information

words can have multiple meanings, which means that the system has to be able to deal with   ambiguities. A word or a constituent can be used in various contexts as well as to describe the representation of a masterpiece or a masterpiece itself. In the example *a picture representing a chalice* the name could be the name of the person represented on the chalice or the artist's name. It frequently happens that the described belongs to a group. The description of this type of objects can hint at the contained or containing elements. We are thus in a situation where several names of a masterpiece are quoted. How can we know which one is the object of the study?

The resolution of ambiguities requires an analysis and the understanding of the local context. Some ambiguities can be decided by using a morphosyntactic analysis of the following or previous words or by searching for linguistic indications according to the given topic.

3.3   *The initial position*

The study of the organization of descriptors in a text can be of considerable help, notably for the resolution of certain types of ambiguities. The study of the initial position, which leans on the cognitive consideration (Enkvist, 1976), (Ho-Dac, 2007), states that the beginning of a sentence has a great importance, as we place important information in an initial position of sentences.

In this perspective, the extraction of the information from the text:

Musée de la société archéologique de Montpellier.
Panneau de Saint Guilhem et Sainte Apolline (87 x 136) en cours de restauration par Anne Baxter.
C'est une peinture à l'huile de très grande qualité, panneau sur bois représentant deux figures à mi corps sur fond de paysage, saint Guilhem et sainte Apolline, peintures enchâssées sous des architectures à décor polylobés; Saint Guilhem est représenté en abbé bénédictin (alors qu'à sa mort en 812 il n'était que simple moine); sainte Apolline tient l'instrument de son martyre, une longue tenaille.

will prefer the descriptor *Panneau* over the descriptor *Peinture*, to indicate the naming of the studied object.

3.4 *Semiautomatic generation of ontology*

The collected knowledge on a masterpiece is partial; it is valid only for a lapse of time and cannot be limited to a fix frame defined for a given application.

The knowledge is flexible, the masterpieces of the cultural heritage have a past, a present and maybe a future "life", and they can change in time. As we mentioned before the extraction of information in our case has to correspond to a precise specification.

We have to face two requirements: to fill a data base defined by the descriptive system of the inventory and allow the flexibility of a knowledge management system. For the first the information found by extraction can be adjusted, and validated by an expert if it is necessary. We think that it is also a convenient moment to satisfy the second point; the validated information composed of descriptors and their relation, which describes the material and immaterial aspects of masterpiece, will feed the ontology of a domain in a vaster and more flexible way.

How to define the ontology regarding the problem of modeling, opening, and knowledge sharing?

There is a vast variety of definitions of ontology, and that of Gruber (Gruber, 93) seems to correspond the best in case: *« ontology is an explicit and formal specification of a conceptualization being the object of a consensus »*.

In other words the ontology of a domain is a set of concepts and relations between these concepts defined by means of a formal language by involved actors and for a particular domain.

According to Charlet (Charlet, 2002), in an ontology we represent and classify concepts and their characteristics (properties); we also represent relations between these concepts.

In our case, we have to describe of what material the object of cultural heritage is made, by whom, when, why, what transformations were done, what is its state of preservation as well as the masterpieces movements. We can say that a certain number of concepts is outlined: time, place, actor (person) and state of preservation.

Intuitively, we guess that some of these concepts are connected to each other, as for example the state of preservation and time, transformations and time, movements and place, transformation and person.

The CIDOC-CRM ontology, already quoted in Sect. 2, presents the necessary formalism allowing reporting relations, which an object can have in time and space.

The heart of CRM is constituted by the temporal entity expressing the dependence between time and the various events in the life of the historical object.

If we consider an example of a sculpture described by the inventory system, information such as author, naming, materials (…) are easily expressed. Because this system is not able to model the various movements of a given object, this information is described using free text and mixed with other type of information in the historic field.

The same information can be easily expressed by the CRM ontology, presented in Fig 3.
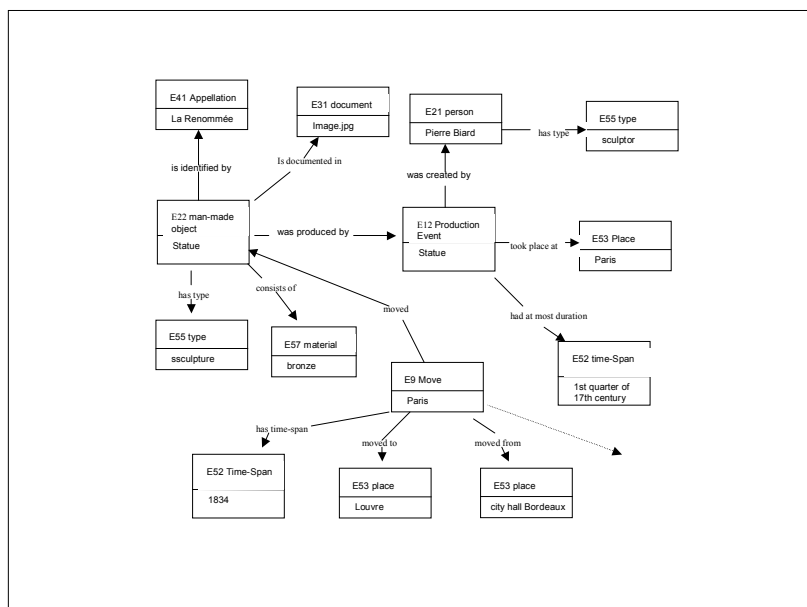
Fig. 3 Example of a sculpture modeling in CIDOC-CRM.

The evolution from the model defined by the inventory descriptive system to the CIDOC-CRM ontology is possible by the search for the correspondences between the fields of the descriptive system, in which the content be considered as the instance of one of the classes of the CRM ontology.

For the cases, in which this correspondence could not be found because the information does not exist in the descriptive system, it will be necessary to extract it from the re-transcribed text, under the condition that the speaker registered it. Otherwise it will be necessary to enter it during the validation of the information extracted automatically by the system.

4   Conclusion and perspectives

This paper presents our work on a voice assistant for knowledge acquisition in the domain of cultural heritage. The originality of our system is the link between three distinctive research domains such as signal processing, ontology and natural language processing. We experimented on field voice knowledge acquisition, "translation" of voice into a text file, the work on text files in order to extract the relative concepts and relation between them in semiautomatic way. The voice interface provides a considerable help and efficiency for an expert working in the

field. The knowledge modeling with ontology adds the flexibility to the classic inventory systems and allows future knowledge retrieval.

We expect to continue this work by incorporating a control of the voice acquisition, in the form of a dialogue human-machine. So the "knowledge collector" would have a real-time feedback on the understanding by the machine of what he dictates. We believe that the implementation of a transcription system and the extraction of information will be shortly possible on mobile devices.

References

Balaram M (1998), PC Version of a Knowledge-Based Expert System with Voice Interface. IEA/AIE (Vol. 2) 1988: 1168-1173

Bourigault D, Aussenac-Gilles N (2003), Construction d'ontologies à partir de textes, TALN 2003, Batz-sur-Mer.

Boufaden N, Bengio Y, Lapalme G (2004), Approche statistique pour le repérage de mots informatifs dans les textes oraux, TALN 2004, Fés, 2004

Boufaden N (2004), Extraction d'information à partir de transcriptions de conversations téléphoniques spécialisées, Thèse de doctorat, Université de Montréal, 2004.

Burns G , Cheng WC (2006), Tools for knowledge acquisition within the NeuroScholar system and their application to anatomical tract-tracing data, Journal of Biomedical Discovery and Collaboration, http://www.j-biomed-discovery.com/content/1/1/10

Charlet J., Zacklad M., Kassel G. & Bourigault D. (eds) (2000), Ingénierie des connaissances, Evolutions récentes et nouveaux défis, Editions Eyrolles et France Télécom-CENT, Paris 2000.

Charlet J. (2002), L'ingénierie des connaissances : résultats, développements et perspectives pour la gestion des connaissances médicales. Mémoire d'habilitation à diriger des recherches, Université Pierre et Marie Curie.

Cole R. A., Hirschman L., et al. (1992), Workshop on spoken language understanding. Technical Report CSE 92-014, Oregon Graduate Institute of Science & Technology, P.O.Box 91000, Portland, OR 97291-1000 USA, September 1992.

Condamines A. et Rebeyrolles J (2000), Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthode. In Charlet J,

Doerr M, Crofts N, Gill T, Stead S, Stiff M (editors) (2006), Definition of the CIDOC Conceptual Reference Model, October 2006.

Enkvist N.E, (1976), Notes on valency, semantic scope, and thematic perspective as parameters of adverbial placement in English". In: Enkvist, Nils E./Kohonen, Viljo (eds.) (1976): Reports on Text Linguistics: Approaches to Word Order.

GRUBER T. R. (1993). A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5, 199.220

Hagège C, Roux C (2003), Entre syntaxe et sémantique : Normalisation de la sortie de l'analyse syntaxique en vue de l'amélioration de l'extraction d'information à partir de textes, TALN 2003, Batz-sur-Mer, 11–14 juin 2003

Haton J-P, Cerisara C, Fohr D, Laprie Y, Smaïli K (2006), Reconnaissance automatique de la parole, Du signal à son interpétation, Donod.

Hernandez N (2005), Ontologies de domaine pour la modélisation de contexte en recherche d'information, Thèse de doctorat, Université Paul Sabatier, Toulouse,2005.

Ho-Dac L (2007), La position Initiale dans l'organisation du discours : une exploration en corpus. Thèse de doctorat, Université Toulouse le Mirail.

Kaufmann K., Michalski R.S. (1986) EMERALD: An Integrated System of Machine Learning and Discovery Programs to Support AI Education and Experimental Research, Center for Artificial Intelligence George Mason University.

Le Priol F (2000), Extraction et capitalisation automatiques de connaissances à partir de documents textuels.Thèse de doctorat, Université Paris-Sorbonne, 2000.

Mercier-Laurent E,(1980), Réalisation de communications dans un processeur de consultation de données textuelles, Thèse Docteur-Ingénieur, INRIA 1980.

Mercier-Laurent E (2007) Role de l'ordinateur dans le processus global de l'innovation à partir de connaissances, HDR, University Jean-Moulin, Lyon

Mykowiecka A, Kupsc A, Marciniak M (2005), Rule-Based Medical Content Extraction and Classification, Intelligent Information Systems 2005

Pavia N G (2003), Un modèle d'analyseur syntaxique robuste fondé sur la modularité et la lexicalisation de ses grammaires, Thèse de doctorat, Université Paris XI, 2003.

Poibeau T (2003), Extraction automatique d'information, Du texte brut au web sémantique, Hérmes, Paris.

Roche Ch (2003), La Construction d'Ontologies : Quel Constat ?,EGC 2003, Lyon 22-23-24 janvier 2003.

Séguéla P (2001), Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques. Thèse de doctorat; Université Toulouse III, 2001.

Troncy R (2004), Formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies : application à la description de documents audiovisuels , Thèse de doctorat, Université Joseph Fourier, 2004.

Verdier H. (1999),- Système descriptif des objets mobiliers. Paris, 1999.- Editions du Patrimoine.

Veronis J (2000), « Annotation automatique de corpus : panorama et état de la technique », Ingénierie des langues, Hermes,2000.

Yangarber R (2001), Scenario Customization for Information Extraction. Thèse de doctorat, New York University, 2001.