

# The CHROME Manifesto: integrating multimodal data into Cultural Heritage Resources

**Francesco Cutugno**  
Università degli Studi  
di Napoli “Federico II”  
cutugno@unina.it

**Felice Dell’Orletta**  
Istituto di Linguistica Computazio-  
nale del CNR - Pisa  
ItaliaNLP Lab - [www.italianlp.it](http://www.italianlp.it)  
felice.dellorletta  
@ilc.cnr.it

**Isabella Poggi**  
Università di Roma3  
isabella.poggi@uniroma3.it

**Renata Savy**  
Università degli Studi di Salerno  
rsavy@unisa.it

**Antonio Sorgente**  
Istituto di Scienze Applicate e Sistemi Intelligenti  
del CNR – Pozzuoli  
antonio.sorgente@cnr.it

## Abstract

**English.** The CHROME Project aims at collecting a wide portfolio of digital resources oriented to technological application in Cultural Heritage (henceforth CH). The contributions for the realisation of such objective come from the efforts of computer scientists, psychologists, architects, and computational linguists, who constitute an interdisciplinary equipe. We are collecting and analyzing texts, spoken materials, architectural surveys, and human motion videos, attempting the integration of these data in a multidimensional platform based on multi-level annotation systems, game engines importing, and virtualization techniques. As case of study we choose to work on the magic travel along three Charterhouses located in Campania region: S. Martino in Naples, S. Lorenzo in Padula (Salerno) and S. Giacomo, in Capri.

**Italiano.** *Il progetto CHROME (Cultural Heritage Resources Orienting Multimodal Experiences – PRIN 2015 MIUR) si pone come scopo la raccolta di una ampia gamma di risorse digitali da utilizzare in applicazione tecnologiche per il miglioramento della fruizione dei beni culturali (CH). A questo obiettivo concorrono interdisciplinariamente informatici, psicologi, architetti, linguisti che collezionano testi, registrazioni di parlato, rilievi architettonici, video e human motion capture. Questi dati sono poi in-*

*tegrati in una piattaforma nella quale è possibile effettuare una annotazione multidimensionale, sono anche utilizzati per la virtualizzazione di ambienti tridimensionali e il porting in ambienti di gaming.*

## 1 Introduction

The CHROME project was born with the intention of creating a framework and methodology to collect, represent and analyze cultural heritage contents and present them through artificial agents whose behavior is inspired by accurate analysis of expert guides, museum curators and tour operators. These gatekeepers are those professional figures possessing a significant amount of knowledge concerning how people should be guided in the exploration of cultural contents. In this sense, they act as mediators between cultural heritage and visitors by using a set of communication strategies, both verbal and non-verbal, aimed at maintaining a high level of engagement and delivering high-quality content.

The overall experience of accessing cultural heritage is greatly enriched by these professional figures: their knowledge and experience, therefore, should not be overlooked when designing artificial agents oriented to cultural heritage presentation. As this knowledge is primarily based on experience collected on the field, the CHROME project aims at recording the performance of gatekeepers in a sensible environment so that formal analysis of their behavior can be documented and studied. The result of this process (see Fig. 1), conducted jointly by humanities and computer scientists, will lead to the formal-

zation of a model describing the behaviors adopted by gatekeepers when presenting cultural heritage. This will then be used to control a humanoid robot designed to follow similar presentation strategies. Taking in account this aim, the main goals of the project are to: **collect** and provide the scientific community with reference datasets to study human-human interaction during the presentation of cultural heritage by professionals; **investigate** the structure of the texts contained in the collected corpus in order to produce automatic approaches supporting text generation for oral presentations in cultural heritage domain; **provide** a reference computational model to support development of artificial agents exhibiting coherent and engaging behavioural strategies. In addition to the orality degree of the assembled presentations, special attention will be attributed to non-verbal aspects. Specifically, CHROME will concentrate on enriching the presentation with consistent prosody and gestures. Finally, another goal is to **evaluate** the impact of these agents in simplifying access to cultural heritage and attract visitors in cultural sites.

For the realization of such goals, five research groups are involved in the CHROME projects covering different scientific and humanistic disciplines that complement each other. The equipe is highly interdisciplinary and is formed of linguists (with specific competences in prosody, pragmatics, paralinguistics, and non-verbal behavior analysis), computational linguists and computer scientists (with skills in Artificial Intelligence and Human Machine Interaction) The teams involved in the project are:

- **UrbanEco** (Naples – Federico II) an interdisciplinary team formed by computer scientists, architects, linguists, aiming at collecting 3D architectural surveys and speech and gesture corpora. UrbanEco is also designing multimodal interaction systems; sub-partner linked to this unit is the “**Polo Museale della Campania - MiBaCT**” the local section of the Italian Cultural Ministry managing more than 30 museums in our region;
- **ILC** (Pisa – CNR) will develop systems for automatically extracting and organizing linguistic and domain knowledge from domain-specific corpora;
- **UniSa** (University of Salerno) will analyze texts and will afford the theme of prosodic analysis of spoken material finalized at speech synthesis issue;
- **ISASI** (Pozzuoli, CNR) will afford the challenge of CH question answering and language

generation for the realization of interaction models in natural language;

- **RomaTre** (Roma, University RomaTre), will confront the theme of multimodal communication and gesture analysis.

As case of study we choose to work on the magic travel along three Charterhouses located in Campania region: S. Martino in Naples, S. Lorenzo in Padula (Salerno) and S. Giacomo, in Capri. All the texts, the architectural surveys and the audio-video recordings, in other words, all the digital resources that we have and will collect and that we describe in the next sections, concern with these wonderful sites.

## 2 The Challenge

An interesting aspect of the CHROME project is to tackle some methodological and technological challenges.

A first challenge regards the role of gatekeepers in shaping visitors’ experience. In fact, the communication in museums is considered an important issue even if museum specialists have been reproached to not do enough in this field (Antinucci, 2014), with some exceptions. Many advancements have been obtained concerning the attempt to understand museum visitors needs and to look for new ways of communication to improve the experience of visiting museums. Investigations about visitors psychological approach (Dufresne-Tassé C. & Lefebvre A., 1995) helped museologists to develop possible methods not only to exhibit artefacts but also to give them sense, providing further explanations. So museum experts may better know visitors, and they are ready to be helped by technology (Cataldo L., 2011).

Moreover, another important aim regards the extraction of concepts and expressive forms from texts. Natural Language Processing technologies are crucial in the process of converting textual documents into knowledge resources. New techniques for the automatic acquisition of linguistic knowledge from texts are needed. Terminology extraction is a central field of research for a number of applications, such as Ontology Learning and Text Mining. Different methodologies have been proposed so far to automatically extract domain terminology from texts. Term extraction systems make use of various degrees of linguistic filtering and of statistical measures ranging from raw frequency to Information Retrieval measures such as TF-IDF (Salton et al.,

1988), up to more sophisticated methods such as the C-NC Value method (Frantzi et al., 1999) or contrastive approach (Bonin et al., 2010).

Another important issue we are going to manage is the analysis of social behaviors in dissemination contexts. The specificities of guided tours have been investigated in (Mondada, 2013), who studies the distribution of knowledge among guides. This stresses the need to adapt to different people during visits; while the relevance of a

user model is pointed out by literature in gesture and Conversational Analysis. Concerning the use of words and iconic gestures in didactic explanations to children and expert and novice adults, their adaptation to the Speaker's Recipient Design and their efficacy for comprehension, (Campisi & Özyürek, 2013) show that people use more words when addressing to adults, but wider and more informative gestures for children. Also, precision was defined as providing details on the

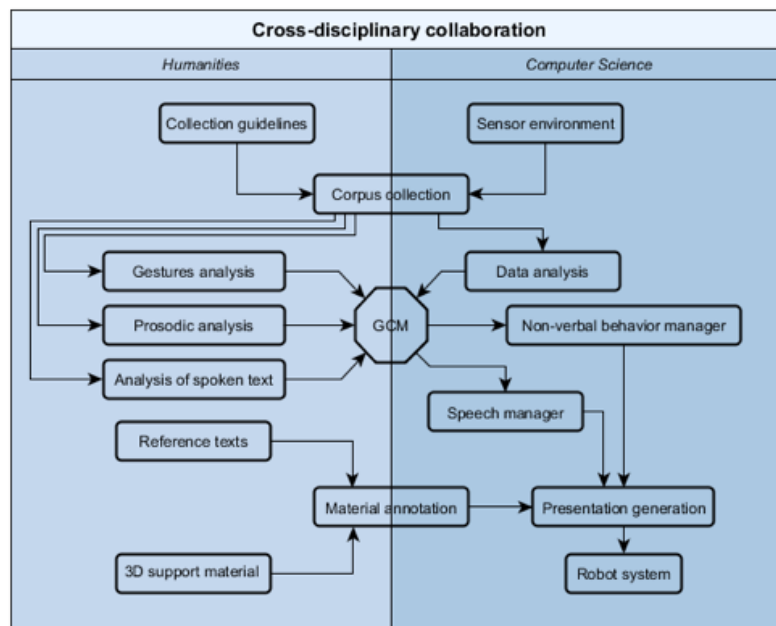


Fig. 1 The CHROME interdisciplinary chart

topic of one's discourse (Vincze et al., 2014), while vagueness is how blurred are the boundaries of one's ideas or discourse.

Spoken text analysis and, prosodic analysis and synthesis will also be addressed. Advanced use of parametric speech synthesis, such as focus/prominence generation by prosodic modification or expressive prosody modelling, has been tested in some research projects (i.e. ALIZ-E). Pushing forward prosodic analysis on gatekeepers' performance can improve the knowledge needed to synthesize natural specialized speech.

Finally, the technologies to mediate the access to digital cultural heritage will be considered. In order to dynamically assemble and present narratives, a formalism to represent different aspects of cultural stories (i.e. (Mele & Sorgente, 2013)) as reported by gatekeepers is necessary. By providing semantically annotated multimedia materials and contents obtained collecting a documental basis, it is possible to use mash-up techniques to dynamically assemble contents and synchronize them with the available media.

### 3 CHROME methodology

CHROME is a cross-disciplinary project focused on combining computational linguistics and behavior analysis methods with expertise in museology to formalise computational models of gatekeepers (see Fig. 1). The main result of this research will be the Gatekeeper Computational Model (GCM) to generate engaging presentations of cultural heritage. The project is organized in three main phases. The data collection phase foresees recording of gatekeepers presenting cultural contents and surveying activities to collect reference texts and annotated 3D models. During data analysis, these resources will be annotated and examined to obtain the GCM. Activities will compare oral expressions with expressions found in texts to automatically select fragments that can compose the final presentation together with gestures and prosody synthesis. 3D models annotation will allow to connect presentation to automatic selection of auxiliary material. Demonstrator implementation will

serve for the validation of the GCM, to disseminate the research results and estimate the impact of the approach in a real environment.

The methodology proposed in the CHROME project targets the following objectives:

- O1. Provide reference datasets to study human-human interaction during the presentation of cultural heritage.
- O2. Survey written contents for cultural heritage dissemination and compare these with the multimodal materials collected in the framework of the CHROME project.
- O3. Provide a reference Gatekeeper Computational Model (GCM) to support development of artificial agents mimicking the ability of expert guides to select and organize contents and applying proper verbal and non-verbal behaviour
- O4. Evaluate the impact of dissemination oriented, multimodal behavioral models on the capability of artificial agents to simplify access to digital cultural heritage and attract visitors in cultural sites

## 4 The present status

At the time we are writing this paper (July 2018) we are at month 16 of 36. Up to now we have collected and analysed many data on Campania Charterhouses: texts, audio, video and 3D reconstructions.

### 4.1 Charterhouses Text

For the three Campania Charterhouses (S. Martino, S. Lorenzo and S. Giacomo), we have collected 102 texts that belong to different document types. In particular, such texts are divided among the following categories: Scientific texts; Specialized catalogues; Dissemination catalogues; Specialized guides; Certified web material; Dissemination kits.

### 4.2 Textual Analysis

Starting from these texts, some lexical and semantic analyses have already been conducted on part of them. The main ones concerned: *i*) Domain vocabulary extraction; *ii*) Event annotation: some texts are annotated added semantic information with respect to reference formalism event based. In particular, the formalism adopted is CSWL (Cultural Story Web Language) (Sorgente et al., 2016). The purpose of this approach is to have a semantic level that will allow us to define an information retrieval not only based on text search; *iii*) AAT concepts recognition: the Art & Architecture Thesaurus (AAT) (Getty,

2018) is a structured vocabulary containing around 40,000 concepts and descriptions related to fine art, architecture, decorative arts, archival materials and material culture. In this step the aim is to link the concepts inside charterhouses texts to such vocabulary.

### 4.3 Digital photogrammetry

The architects group have completed the activity of aerial photogrammetry digital survey performed by UAV and laser scanner on the 3 main charterhouses buildings and on many interiors.

### 4.4 Video recording of touristic guide

Three of four touristic guides have been video recorded during tours in the S. Martino Charterhouse while describing the artistic features, and each one is followed by a public of four visitors. Cameras are pointed on the guide and on the public, speech sounds are recorded with three microphones, one headset worn by the guide and two on field at about one meter equidistant from the guide and pointing to the visitors, too. Speech analyses on these material consists of:

- Orthographic level: Transcription of words, pauses, filled pauses, false starts;
- Phonetic level: Phonetic transcription and annotation of coarticulation phenomena, Speech quality analysis;
- Syllabic level: Annotation of syllables, Speech fluency and speech rate analysis;
- Intonation level: Pitch movements in relationship with the segmental level, Emphasizing patterns, speech style.
- Textual level: analysis of sentences, text structure, and communicative goals.
- Multimodal behavior level: annotation of gestures, face and gaze, including physical description, semantic analysis, classification in terms of textual, emotional and interactional functions.

The tool chosen for annotating the speech and video material is ELAN<sup>1</sup>. In each video portion the guide's gestures and body communication will be annotated in terms of the communicative functions they serve. Thus the annotation will allow to distinguish the styles of the guides: e.g. a very "technical" guide will use gestures and body communication more frequently aimed at describing the artwork or the author, while a "friendly" guide's body behaviors will be often aimed at creating syntony with tourists.

---

<sup>1</sup> <https://tla.mpi.nl/tools/tla-tools/elan/>

## 5 Summarizing

CHROME aims at formalizing data collection and annotation paradigms for architectural heritage, in particular the annotation regards texts, video, audio and gestures. From the annotated data, we will: *i*) perform correlation analysis to identify cross-domain patterns and link them to communicative goals; *ii*) describe how an expert presenter relates to the physical environment while she describing it; *iii*) identify which communicative strategies can be mimicked by an artificial agent with the available technology. Possible domains of simulation will be the deictic and iconic gestures, face and gaze behaviour; *iv*) implement a final demonstrator adopting the formalized strategies to generate dynamic presentations for the attending visitors.

## 6 Acknowledgments

This work is funded by the Italian PRIN project Cultural Heritage Resources Orienting Multimodal Experience (CHROME) #B52F15000450001.

## Reference

- Antinucci F. (2014). *Comunicare nel museo*. Laterza Milano
- Bonin F., Dell’Orletta F., Montemagni S., Venturi G. (2010). *A Contrastive Approach to Multi-word Extraction from Domain-specific Corpora*. In: LREC’10 – Seventh International Conference on Language Resources and Evaluation (Valletta, Malta, 17-23 May 2010). Proceedings, pp. 3222 – 3229.
- Campisi E. and Ozyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics* (47), pp. 14-27
- Cataldo L. (2011). *Dal Museum Theatre al Digital Storytelling*. Franco Angeli Milano
- Dufresne-Tassé C., Lefebvre A. (1995). *Psychologie du visiteur du musée*. Hurtubise Montréal
- Mele, F., Sorgente, A. (2013). OntoTimeFL – A Formalism for Temporal Annotation and Reasoning for Natural Language Text. *New Challenges in Distributed Information Filtering and Retrieval, Studies in Computational Intelligence* 439, pp. 151-170
- Mondada, L. (2013). Displaying, contesting and negotiating epistemic authority in social interaction: Descriptions and questions in guided visits. *Discourse Studies* 15, pp. 597-626
- Frantzi, K., Ananiadou, S. (1999). The C-value / NC Value domain independent method for multi-word term extraction.
- Getty AAT: About the AAT. <http://www.getty.edu/research/tools/vocabularies/aat/>. Accessed April 2018
- Sorgente A., Calabrese A., Coda G., Vanacore P., and Mele F. (2016). Building multimedia dialogues annotating heterogeneous resources. In *Artificial Intelligence for Cultural Heritage*, chapter 3, pages 49–82. Cambridge Scholars Publishing.
- Salton G., Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, 24(5), pp. 513-523
- Vincze L., Poggi I., D’Errico F. (2014). Precision in Gestures and Words. *Ricerche di Pedagogia e Didattica – Journal of Theories and Research in Education* 9, 1. Communicating certainty and uncertainty: Multidisciplinary perspectives on epistemicity in everyday life