UNIVERSITÀ DELLA CALABRIA

UNIVERSITA' DELLA CALABRIA

Dipartimento di *Ingegneria Informatica, Modellistica, Elettronica e Sistemistica*

Dottorato di Ricerca in

Information and Communication Technologies

CICLO

XXXI

**DISCOVERING THE "WORLD CITY":
FROM TEXTS' ANALYSIS TO 3D SCENES VISUALIZATION**

Settore Scientifico Disciplinare M-STO/08

Coordinatore: Ch.mo *Prof. Felice CRUPI*

Firma _____

Supervisore/Tutor: Ch.mo *Prof. Roberto GUARASCI*

Firma_____

Dottorando: Dott./ssa *Valentina BOVA*
Firma _____

i

Dedicated to life, which never ceases to surprise us…

Dedicated to my family, infinite source of happiness…

Contents

Abbreviations

Symbol	Description
ICT	Information and Communication Technologies
IE	Information extraction
POS	Part-of-speech tagging
NLP	Natural Language Processing
DT	Determiner
NN	Noun, singular or mass
VBZ	Verb, 3rd person singular present
IN	Preposition or subordinating conjunction
$PRP	Possessive pronoun
AI	Artificial Intelligence
CG	Computer Graphic
AR	Augmented Reality
VR	Virtual Reality
tr	Trajector
lm	Landmark
s_i	Spatial Indicator

List of Figures

List of Tables

Introduction

I. Three-dimensional Computer Graphics in the Cultural Heritage field: Objectives and Potentialities

Computational science and informatics are strongly linked to the representation of information in the way of texts, images and 3D models, which are able to increase user's knowledge in diversified sectors. Among these types of data,3D models are certainly among the most complete and, consequently, suitable instruments to spread information. Moreover, the improvements made in 3D technology have not only revolutionized the way we see reality, but they have also broken down methodological and historical barriers.

3D representations can be considered containers of information of different nature and origin that, if isolated, have little value but can become complex and exhaustive if observed and analysed jointly. At the same time, the three-dimensional models may be "exploded" or decomposed in all their parts giving the possibility to observe and analyse all the components separately. That means that it is a study that goes from the more general to specific detail, and vice versa, trying to cover all aspects of knowledge.

The "3D" datum is now present in daily life both in terms of leisure and in many professional sectors. In fact, the 3D modeling technique is used to create virtual models of existing or non-existing objects, related to several fields such as mechanical, industrial, architectural, design or medical. Moreover, three-dimensional reconstructions, rendering, videos and applications of Augmented Reality and Serious Game have seen, in the last three decades, a large development in the field of Cultural Heritage [1]. The number of experimentations and case studies in this context, in fact, are quickly growing and a new discipline is being defined under the name of Virtual Cultural Heritage. It concerns the 3D reconstruction of historical places, either existing, no longer existing or never realised, and it allows users, scholars and historians to illustrate the synthesis of their studies and to discover the characteristics of

artefacts. In other words, all the works that in the past were shown to visitors through drawings or physical scale models are now made available to the users by three-dimensional elements located in a virtual world. The discipline that makes this possible is Computer Graphics, the science or the art of visual communication, which has several specific advantages such us:

- the possibility to offer three-dimensionality to a place or a project of the past;
- the chance to make changes according to new research and discoveries;
- the opportunity to replicate, infinite times, the three-dimensional elements in order to use them for additional scientific, educational or communicative purposes.

II.   Work Aims and Motivations

In this thesis work, the principal issue is not the "realism" of the three-dimensional representations or the simple transposition in 3D of works of the past, but the presentation of an emerging methodology, the tools and the applications used to create the 3D reconstructions together with all the difficulties associated with this aim. The abovementioned methodology, which will be discussed in Chapter 6, will be applied to a specific case study concerning the Cultural Heritage field in order to "give life" to a complex and incomplete project of the past and to spread its precious knowledge. The case study is the project of the "World City" [2], a city that, despite its historical and architectural importance, linked to the events and characters of the past that have affected it, is underestimated or, even, not recognized as a work forming part of the world cultural heritage. Therefore, this thesis work aims to provide an innovative tool able to transfer, through the use of 3D models corresponding to textual spatial descriptions, the knowledge of this complex project of high cultural value.

To conceive his project Otlet collaborated with important exponents of architecture, such as Hendrik Christian Andersen and Le Corbusier and, with each of them, he conceived a different project for composition and characteristics. Among the various projects, in this work, it has been decided to deepen more the one of Otlet and Le Corbusier because particularly documented through numerous written and graphic sources. The project, even though it was completely designed, was never realized and remained an utopian project.

The World City was conceived by the architect as a very complex set of full and empty spaces able to respond to the emerging needs of the post-war period. Its planimetric organization aims, not only, to satisfy the functional criteria but also to face, for the first time, a political problem. The architectural project, in fact, aims to be a "turning point",

capable of making significant changes to a society full of imbalances and instability. The World City, therefore, identifies itself as an utopian city based on peace, on the exchange of knowledge and on the collaboration between men. In other words, a new alternative and better reality, able to distance humanity from a past full of discord, suffering and disparity. This is why the city is called to be, at the same time, a political and administrative center, an intellectual center (Mundaneum), a residential center (cité hotelier), a sports center (stadium) and an economic center, as well as a city with administrative autonomy.

The choice of using a three-dimensional virtual representation is based on the strong communicative connotation of 3D models, which facilitates a more rapid assimilation of the historical concepts with reference to both a specialist and not specialist public. Indeed, the power of the digital simulation of historical-cultural contents lies in the possibility of involving ordinary people, without historical or informatics skills, in successful learning processes that lead to a deep and lasting understanding of the information being communicated. The knowledge transfer takes place through new informative environments in which the users are stimulated thanks to the use of visualization and interfacing technologies. The methods or the tools used to create virtual worlds are numerous and aim to generate complex cognitive dynamics based on the spatial exploration of the data and the consequent comprehension of the information contained. The technique used and described in this work is the "text-to-scene generation", a methodology which aims to create 3D scene reconstructions by interpreting natural language. This research area is not yet widespread or fully developed, but several studies are being carried out in this direction.

III. Text-to-3D-Scene Application Cases

Text-to-3D-scene-generation is an emerging methodology, which allows the creation of 3D environments without being extremely time consuming or requiring an enormous effort from the user, but directly from natural language descriptions. It permits to build three-dimensional scenes without learning specific graphical software, which require special skills and knowledge in the field of 3D modeling, and without spending an enormous amount of time in modeling objects. It is a versatile application, in fact, it can be useful in several areas and contexts, namely, to express designers' creative aims, to show existing projects or artworks of the past, to facilitate scholarly learning and much more.

For instance, the methodology could be an interesting instrument to shape mental images and visual ideas or to materialise imaginary scenes born from writers' fantasy [3]. The

advantage is that only in some occasions does the described setting really exist and the association between textual language and three-dimensional image is moderately simple, however, other times, the setting is the result of the writer's fantasy and the reader can comprehend what the author writes only by using his imagination. Moreover, the comprehension is even more difficult if the text refers to a specific domain and, consequently, the three-dimensional visualization could be very problematic for non-expert readers. On these bases, the text-to-3D-scene-generation approach may simplify and explicit this association by assigning a shape to textual descriptions and generating the corresponding three-dimensional representation in a virtual space.

At the same time, the technique can be used to synthesize basic 3D object arrangements from user's input examples. Precisely, once given an input descriptive sentence, the system can generate a plausible corresponding scene by learning from a larger scene database [4].The methodology could be a useful starting point to create new, different virtual scenarios which can be suitable for the most varied recreational or professional needs. For example, it could provide help to designers in the presentation of new ideas or projects because they could instantaneously create three-dimensional virtual environments starting from a textual description and they could modify them, according to the interlocutor's preferences, by adding or removing descriptive details. In addition, the technique could also be intended as a way to allow the hypothetical buyers to materialize their project idea by themselves exploiting natural language descriptions and without any type of graphical software skills.

Another remarkable applicative example of the 3D-scene-generation methodology concerns the knowledge acquisition process of children with intellectual deficits which cause thinking, communication and socialization problems [5]. Specifically, these children encounter serious issues in handling linguistic structures and, consequently, face problems in communication and socialization. On these premises, a text-to-scene conversion system can be a special support for children's learning processes because it can be used as a computer-assisted teaching approach in the education of intellectually challenged children. The latter, in fact, cannot generally connect words to their meanings and shapes because they consider them abstract entities. Hence, the usage of a virtual representation may simplify the link between the two elements in order to give sense to the words present in the sentence and to provide a positive influence on children's learning processes.

A further way to exploit the text-to-scene generation methodology, maybe the most ambitious, refers, once again, to the scholastic context where the students can have the possibility to mature and increase their algorithmic thinking [6]. Indeed, the rapid

development of information technologies in daily life implies the necessity of increasing people's ability to think algorithmically, namely, to solve problems through a stepwise procedure. In spite of this, school education is principally aimed at teaching numeric computation, so, students are generally not able to: a) identify the problem exactly, b) subdivide the problem into sub-problems, c) plan a step-by-step solution. On this basis, a software tool able to help educators to teach algorithmic thinking may be an interesting solution [7]. In this scenario, the 3D-scene-generation technique can allow students to seethe three-dimensional representation deriving from a textual description and to study, simultaneously, the source code which generates the scene. In addition, they can change the location of the objects which appear in the scene and see how the code changes in order to understand the main programming language constructs and functions.

A last example of use of the text-to-3D-scene generation technique concerns the process of investigation and analysis of natural language visualization from a computational point of view. The main challenge is that machines do not have any cognitive ability and they need precise and complete instructions to understand human language, rich in ambiguity, polysemy and implicit information, in order to automatically generate a correspondent virtual three-dimensional representation [8]. An additional difficulty of the automatic visualization process is that the objects, existing in a described scene, are not often explicitly mentioned in the description because much knowledge about the world around us is obvious for men and for their mental abilities. For instance, if one says "*The knife and the fork are, respectively, to the right and to the left of the dish*" he takes for granted that the interlocutor understands that the dish and the cutlery are not put on the floor or on a bed but that they are located on a table. This means that, to create a system able to interact with people, it is necessary to investigate: a) how people use natural language to describe spaces and situations; b) which are the objects involved; c) which are object characteristics and locations and d) how to transfer human knowledge to the computer. On these bases, the text-to-scene generation technique could be seen as the result of a process of analysis of these aspects. Furthermore, it may be a connecting element between one-dimensional and three-dimensional information capable of "giving shape" to the words by transforming them into more or less complex 3D models. In this way, it would be possible to remove linguistic and domain barriers connected to the knowledge of word meanings or to the characteristics of objects referring to a particular field. Finally, it could be intended just as a form of learning about specific domain objects in their formal and functional characteristics and a way to enrich users' knowledge with a strong visual impact.

Among the various application cases of the methodology, it has been decided to exploit the text-to-3Dscene-generation technique in the field of the Cultural Heritage. More specifically, virtual scenes have been created starting from descriptive sentences extrapolated from primary and secondary textual sources related to the World City project. This last aspect, that is the "nature" and origin of the texts, represents the innovative point of this thesis work, since the creation of 3D scenes did not start from sentences invented by the user or from documents without artistic or cultural value, but from texts referred to the World City. In this way, the aim is to spread the knowledge of a project which is a "piece of history" marked by an extraordinary historical, cultural and architectural importance. This topic will be dealt with in more detail in paragraph 6.2.1.

IV Outline of the Thesis

This thesis is divided into four main sections, each of them is characterised by a specific goal:

The first objective of the work was to collect all the sources regarding the "World City" (Chapter 1) in its two principal versions, which are Andersen's original plan and subsequently Le Corbusier's project. Once the relevant writings and drawings were collected a meticulous analysis was carried out in order to acquire a sound knowledge of the topic of study (§ 2.2). Consequently, all the buildings of Le Corbusier's project have been modelled in three dimensions with the purpose of creating the 3D virtual models of the principal subjects of the utopian city (§ 6.4.2).

The second task was to create a link between the previously modelled three-dimensional elements, and the textual information describing the characteristics of the buildings and their locations in the urban space. This information has been automatically extracted from the original texts (§ 2.3) and simple 3D representations of the described scenes have been projected in a virtual environment. For this purpose, a rudimentary but efficient system able to transform a text written in natural language into a three-dimensional scene has been implemented (Chapter 5, § 6.3, § 6.5) and it has been used to virtually rebuild a masterpiece of the Cultural Heritage of the XX century: the World City Project. As already said, the city, due to its complexity and its highly utopian nature, was never built and it remained an unrealised dream, thus, this thesis work aims to provide an automatic reconstruction of it with the purpose of spreading its knowledge to all the users interested in world Cultural Heritage. Indeed, according to the Principles for the Conservation and Restauration of Built

Heritage it is known that "In the protection and public preservation of archaeological sites, the use of modern technologies, databanks, information systems and virtual presentation techniques should be promoted" [9]. Moreover, the relationship between the field of Cultural Heritage and Information and Communication Technologies (ICTs) could be exploited not only to represent sites and artworks of the past, or no longer existing, but also to illustrate projects which were designed and never realised. Therefore, the purpose here was to create a system able to generate, in an unusual way, virtual models useful to researchers to formulate conclusions about their studies or to general users to discover and understand the characteristics of their cultural heritage.

The third aim of the work was to generate an immersive three-dimensional reconstruction of the World City, aware that a 3D representation of the urban spaces, described and drawn by the authors, allows the users to have a better perception and knowledge of the project plan. People, in fact, could have the possibility to live personally in that "ideal piece of world" and to explore the City in all its parts discovering the marvellous buildings and the green areas. They could live an immersive experience and see the materialization of several descriptions contained in the historical books, publications and articles (§ 6.6). Furthermore, they could compare the original aim of the authors with their real impressions and suggestions and allow themselves to get wrapped up in the utopian atmosphere.

The fourth objective was to create a web page containing a good part of the material produced or collected concerning the project of the World City (Chapter 7). This page, although it is not yet present on the network due to problems related to copyrights with reference to downloadable texts, is intended as a large container ready to host an increasing number of sources, reconstructions, studies and everything that can spread the knowledge of the Otlet project to the most heterogeneous users. At present, it contains a series of references to graphic and textual sources, brief descriptions of the project, three-dimensional models and links to other Web pages or documents. However, the idea is to increase the content and, therefore, the importance of the page turning it into an instrument of knowledge and enhancement of an extraordinary work of our cultural heritage.

The ideal final task which, due to the lack of time in a doctoral process, has not yet been reached, should be the correlation between the three-dimensional models and all the documents and sources related to them. Only in this way will it be possible to have a global vision of the works in question in their geometrical, historical, urbanistic and symbolic aspects.

1. Case Study: "The World City"

1.1 Introduction

This chapter will present the case study of the work, the project of the "World City", starting from an excursus of the precedent projects related to the concept of ideal cities. The "World City", in fact, is an utopian city, born from an idea of a symbolic and considerable figure in the world of documentation: Paul Otlet[1] [10]. He planned to create a City where people could live in harmony and in a universal cooperation; a place where enclose, at a global level, the great institutions of intellectual work: libraries, museums and universities. The evolution of Paul Otlet's concept of the *World City* (*Cité Mondiale*) was developed in the period from 1910 till 1941 and, to promote his idea, he collaborated, on different occasions, with several famous architects like Hans Christian Andersen and Ernest Hébrard, Octave van Rysselberghe, Louis Van der Swaelmen, Le Corbusier, Victor Bourgeois, Maurice Heymans and Stanislas Jassinski [11].

1.2 The ideal Cities

The ideal city is a planned, real or utopian city whose urban organization responds to precise philosophical, scientific and rational criteria.
In the ancient Greek era the philosopher Plato theorized an ideal city which was able to mirror the kosmos, on the one hand, and the individual on the other. He described the ideal polis as a city based on justice and human virtue, a form of social and political organization that allowed individuals to maximize their potentialities, serve their fellow citizens, and live in accordance with universal laws and truths. Plato theorized the ideal polis from a political

---

[1] Paul Otlet (1868-1944), was an author, a lawyer, a socialist and pacifist militant. His will, as a visionary man, was to live in a World characterized by peace and scientific progress. Moreover, he aimed to map out the world's knowledge and store it, through an archival system, into wooden cabinets. His idea and writings anticipated the advent of the World Wide Web, in fact, together with this friend Henri La Fontaine he laid the foundations of the Mundaneum, a sort of paper Google with pacifist principles. He founded the Union of International Association in 1910 (now located in Brussels) and he spent several years in the implementation of projects regarding the construction of an ideal society and an ideal city. In particular, in the early 1900s, he conceived an unrealised project of an "perfect" city, called, the World City.

point of view, indeed, he set a five-fold classification to describe how the city ought to be governed. In his idea, the best form of government was an aristocratic model based on the rulership of philosopher kings; the second form was the timocracy, founded on the leadership of an elite of guardians or strong men; the third type was the oligarchy, consisting of rules "by the few" and, the last two forms were the democracy and the tyranny which represented the rules "by the many". The philosopher believed that the ideal polis had to be an enlightened city, based on the highest universal principles, that he called the "City of God": a place where humans could aspire for the best and could look for inspiration and guidance in carrying out their worldly affairs [12].

Conversely to the theoretical Greek urbanization, the Roman one was founded on practical functionality, in fact, it was set on precise practical criteria, which included the organization of the settlement model [13]. In particular it was based on a checkered planimetry crisscrossed by two large streets, the "cardo" and the "decumano", which converged into the centre of the city: the Forum.

The design of the ideal city mainly spread in the Renaissance when urban settlement assumed a new role linked to the centrality of man.

An urbanistic and architectural theorization of the ideal city characterized the Renaissance era in which the revaluation of classical art and the attention to the harmony and proportionality of the forms coexisted [14]. In this context Architecture became an instrument of centralization of man that triggered a process of transformation of the city role from a simple settlement or monumental function to an identity role. However, ideal cities could not be formed in reality, without political and economic support. Therefore, regular geometry and straight streets with a strong perspective of cities represented the power of early capitalism and the central state.

The concept of ideal city was re proposed, in the XX century, by the famous architect Le Corbusier[2] who, not only elaborated, together with Paul Otlet, the project of the World City, which will be illustrated below, but also after the non-realization of it, continued to design utopian cities. Indeed, in 1935 published a treaty, named "*La Ville Radieuse*" in which he described a utopian city, functional and well organized. It was composed of both an

---

[2] Swiss architect, town planner and designer. Le Corbusier is considered one of the founders of contemporary architecture, as well as an exponent of the Modern Movement and Brutalism. He was the designer of several French private residences, such as the Ville Savoye, and of public buildings throughout the world. He was also the inventor of modern housing solutions, such as the Unitées d'Habitation, and was the creator of cities inspired by utopian visions. Finally, he invented furniture that became part of Western popular culture.

industrial and residential area, which occupied only 12% of the total soil, leaving the rest of the land to green areas. The street for pedestrians and the motorways would have been diversified and the public transport would have taken place underground. Some years later Le Corbusier was tasked to plan the city of Chandigarh, an Indian City, by the prime minister Nehru. The city should have been a symbol of harmony and hope and its urban structure should have contained several green areas among the buildings. The city was designed for half a million people, it should have been self-sufficient and its plant resembles the human body because the architect dreamed of a city on "human scale". Its most famous monument is the big hand stretched out to the sky that was located in the centre of the city and represents a symbol of peace [15].

A last notable example of an ideal city project was based on an idea of a Belgian bibliographer, Paul Otlet, who thought about a city able to be the symbol of the feeling of unity between people. It should have been a centre of knowledge in which to preserve the memory of the entire world. The idea of the "World City" creation, from its implementation to the last project, received both approvals and controversies and it was changed or revisited several times.

1.3 The World City Project

The early 20th century was marked by rapid industrial, economic, social, and cultural changes. It was characterized by prosperity and economic growth on the one hand, and on the other by a rise in poverty and slums in major cities. The markers of the period were urbanization, architectural and scientific advances and the spread of goods and information. After these years of splendour a period of competition began between nations and the rivalry between European powers led to the outbreak of the first World War. In this way, from 1914 to 1918 the majority of the world's states were characterised by destruction and death and the world became a site of terror. Thus, in 1919, soon after the end of World War I, the idea to create an "ideal place" in the middle of the Europe, which took the name of "World City" [2] arose. It should have been a city based on harmony and scientific progress and a self-sufficient city from a political, economic, social and intellectual point of view. The idea was to create a centre for professionals, philosophers, artists and scientists, in other words, a "place of world culture" in which gather to the traditional institutions of intellectual creativity and enhance the peaceful use of knowledge.

This new city presents itself as a collection of elements scattered in an immense green park near the lake. The project, which never had a precise location, even though, one of the places favored by Otlet was the city of Geneva, was unsuccessful because it was rejected by the Geneva authorities for reasons of too high costs, both because of the Nazi invasion in the 1940s, which reduced the Otlet project to nothing, because it was considered harmful.

The "international city", also known as the 'new Geneva', was called to materialize the dream of building a place that must accommodate all world knowledge and in which humanity could recognize itself. The idea was to create a symbol of unity and brotherhood and a place of the globe where the image and the meaning of the world may be perceived and understood. It should have been a "home for men" after the war, construed as an experiment in organizing and introducing order and permanent peace to a world-wide society [16]. In addition, it should have been a monument to contemporary man and an extension of the present highest institutions of intellectual and cultural life.

As cited before, Paul Otlet collaborated with numerous famous architects in order to design the complex project of the World City and, each of them, elaborated their own spatial representation of it. All the projects, in fact, are characterised by different aspects from an architectonical, urbanistic and geographical point of view.

1.3.1 Andersen Project

In 1913 Hendrik Christian Andersen, a Norwegian sculptor, published a book titled "*Creation of a World Center of Communication*" [17] where he set up his idea regard the construction of an international city (fig.1,2).

Figure1: Illustration of Andersen World Centre, in H. C. Andersen, and E. M. Hébrard., "Creation of a World Centre of Communication", Paris, 1913, p.I.
.
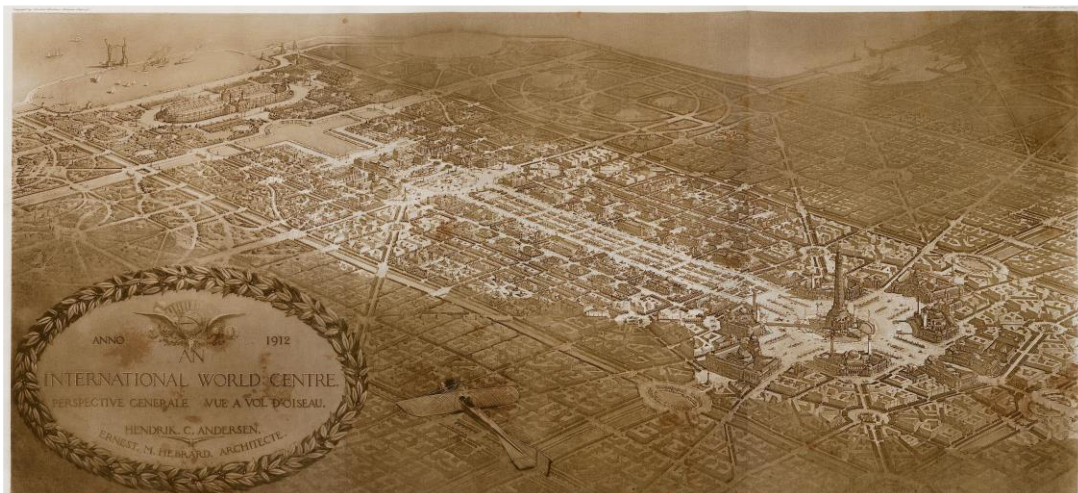
Figure 2: Illustration of Andersen International World Centre, H. C. Andersen, and E. M. Hébrard, "Creation of a World Centre of Communication", Paris, 1913, p. XIIX.

In the first part of the book he talked about the history of the most important architectural complexes of all ages, from prehistoric to modern times, and he found in the past the best and most comprehensive organization. Andersen structured his book alternating drawings and theoretical reflections and, most likely, his idea was inspired by the American transcendentalist philosophy. Indeed, he believed that the human's path goes through the attainment of peace and harmony between men. The sculptor established a collaboration with the French architect Ernest Hébrard to whom, at first, he gave the task to design a fountain to place in a public square in the centre of the city. Several pages of "*Creation of a World Center of Communication*" are dedicated to the "fountain of life" which should have been one of the most important symbols of the City, in fact, the flow of the water represents the continuous flow of human's life that, day after day, goes from sunrise to sunset. It should be noted, however, that in the book the link between the fountain and the idea of the World City remains undefined.

The complex project of Andersen and Hébrard, which lasted about nine years, was known to an architecture historian, Urbain Ledoux, who, aware of the will of Paul Otlet and Henri La Fontaine to build an international center for international associations, decided to inform them. Therefore, when the latter went to Paris in 1912, the drawings were almost completed and they had the possibility to see them. Their impressions were positive and they expressed the appreciation for the idea and to the structure of the city, in fact, although the project was not yet finished, it was presented in a meeting of the Union of International Associations in April 1913.

The World City did not have a precise geographical location because, in this way, the architect would have been able to express his ideas freely, without any types of restrains. The only certain thing was that the City should have been built near the sea in order to be a point of convergence of all the sea-routes of the globe.

Hébrard's city is structured along an axis of symmetry that aligns a monumental centre, some directional areas and a big station while, all the other elements of the plan, are distributed with respect to this axis in a specular way. The residences and the health care establishments are located on concentric zones around a monumental area where a symbol of the city the *Tower of the Progress* (fig.3) is placed.

Figure 3: Elevation of the Tower of Progress, in H. C. Andersen, and E. M. Hébrard., "Creation of a World Centre of Communication", Paris, 1913.

It is the dominant element located in the middle of the "*Congress Square*" (fig.4,5), a circular square inscribed in a square-shaped one, where big boulevards converge and where, in each corner, one of the four Congress Buildings are set.
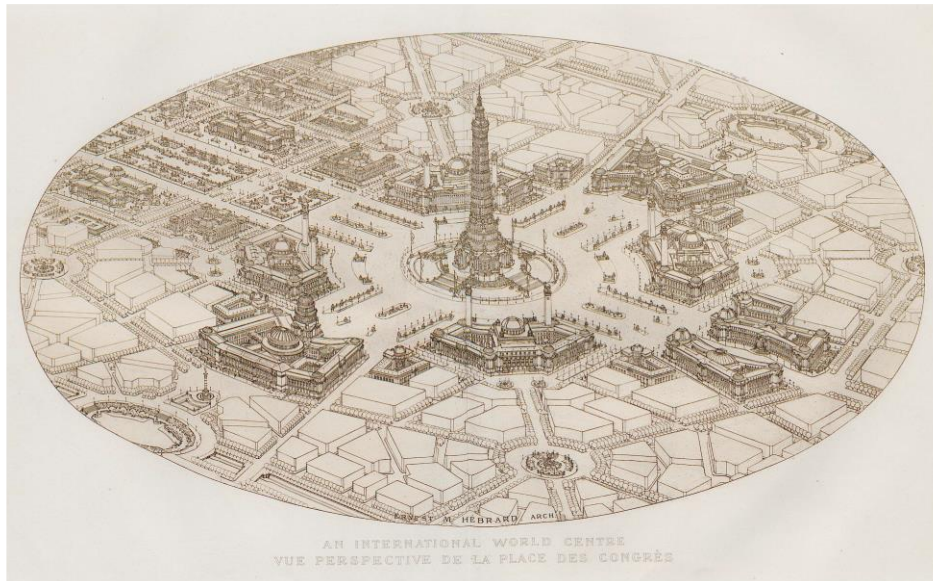
*Figure 4: Perspective of the Congress Square, in H. C. Andersen, E. M. Hébrard,
"Creation of a World Centre of Communication", Paris, 1913.*
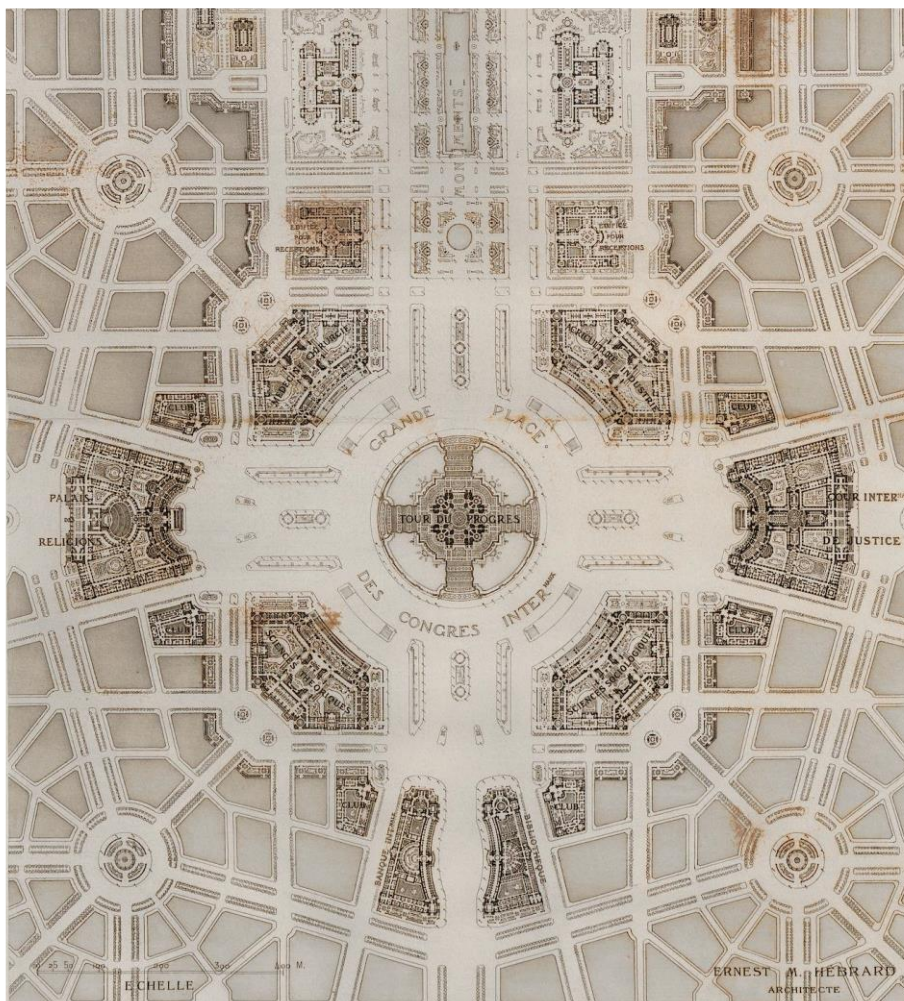
*Figure 5: Plan of the Congress Square in H. C. Andersen, and E. M. Hébrard.,
"Creation of a World Centre of Communication", Paris, 1913.*

A bit further away from the centre two parks and several sports facilities overlook on the seafront while, on the opposite area with respect to the axis of symmetry, some industries and services are placed. To conclude, a navigable channel surrounds the urban area.

Andersen, aware of the interest of the regimes in his project established a relationships with Benito Mussolini who, in 1926, after a meeting with the sculptor, offered for free a land near Rome for the construction of the World City. Paul Otlet, as a great pacifist, was not enthusiastic about the news, furthermore, he did not agree with his megalomaniac and idealistic attitude, so, he decided to abandon the project and to start a new relationship with a reknowned architect: Le Corbusier.

1.3.2 Le Corbusier Project

During the first months of 1928, the Belgian bibliographer communicated with the Le Corbusier and Jeanneret architectural firm, sited in Paris, to propose the design of a new World City to be built in Geneva. Otlet already knew Le Corbusier and he shared his architectural style and inspiring principles. Furthermore, he appreciated the ability to assert his ideas. These are the reasons why he decided to contact him to present the project on which he worked for many years, together with Andersen, without ever seeing it realized. The great project should have been managed by an association of friends belonging to several European capitals and numerous buildings such as a library, a museum, scientific associations should have been included. The core of an even vaster project should have been the Mundaneum and its planning took place between 1928 and 1930, within a complex set of related events. Le Corbusier's work could be distinguished in two different consecutive periods, which lead to the definitive formulation of the project. The first phase ran from March to September 1928 and ended with the publication of a brochure published by Otlet [18], which contained the motivations of the idea, a description of the architectural project and drafted by Le Corbusier. The second phase, instead, lasted from September 1928 to August 1929 and was characterised by a large revision of the project itself.

The internationalist ideology of Otlet and the ambitions of its program forced Le Corbusier to put a particularly intense effort on the project. Indeed, already in 1926, during the preparatory phase of the project, he had studied the entire area of the hill drawing the highest site on the Morillon plateau under the village of Saconnex (fig.6).
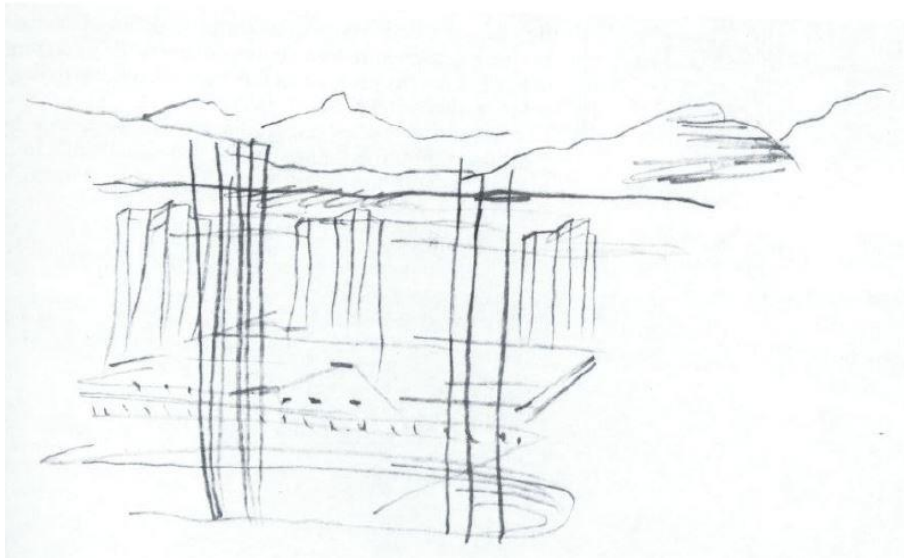
Figure 6: Mundaneum lands seen from Saconnex; in the background Le Corbusier traces the outline of Mont Blanc; March 1928, (©AFLC 2461).

There are not any documents evidencing an inspection made specifically in the site of Geneva, therefore, Le Corbusier probably worked "from memory" on photographs and drawings.

During a meeting in Brussels, Paul Otlet presented to Le Corbusier and Pierre Jeanneret, the results of many years of work with the architect Hébrard and the sculptor Hendrik Andersen on the construction of the "World Center"; subsequently, Le Corbusier was asked to give shape to a large museum, a library, a centre of the International Associations, an international university and, finally, a large natural park with sports facilities, a residential center, a railway station and an airport.

The institution of the Mundaneum, according to Otlet's idea, should not have been a traditional complex of buildings but an immense emblematic space. To better describe his idea, he made a sort of synoptic table, called "plan forms" (fig.7), which was used by Le Corbusier as guidelines to develop his future project.
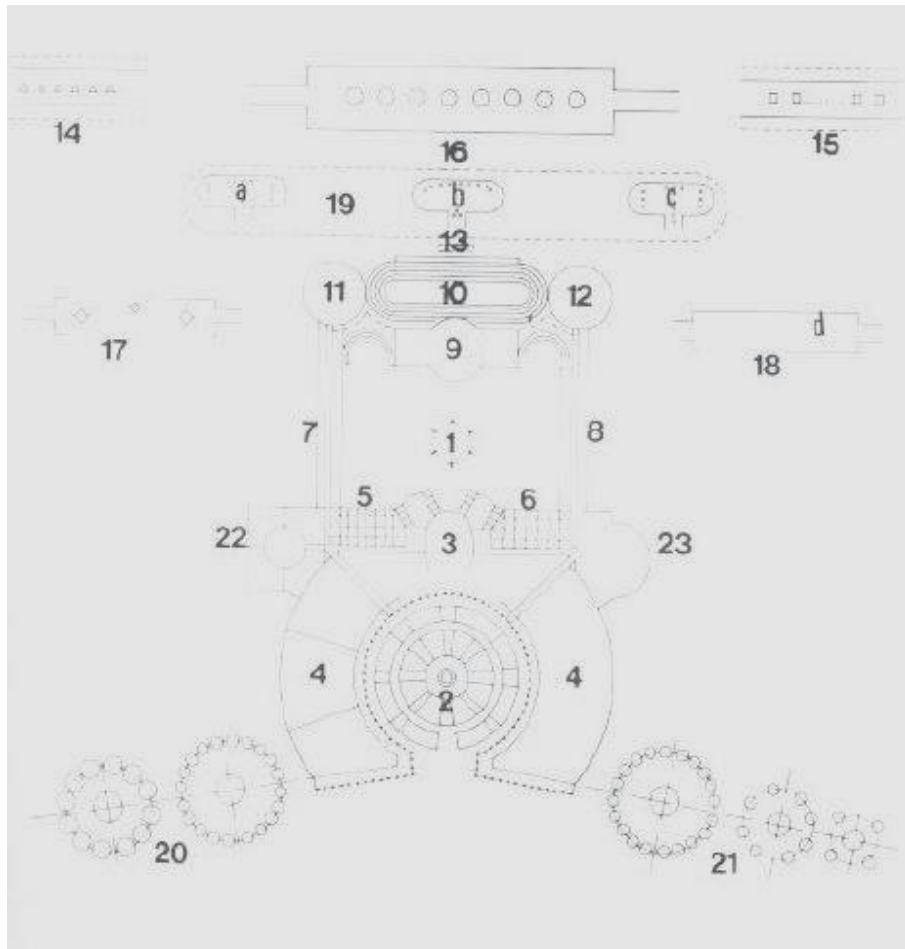
Figure 7: Schematic plan of relations between the individual parts; December 1927, (©AFCL 24574).
1. Sacrarium; 2. Globe; 3. Hall and centre; 4. Museum; 5. Associations Offices; 6. Nations Offices; 7. Library stores; 8. Temporary expositions; 9. Peace building; 10. Stadium; 11. Cinema; 12. Diorama; 13. Station; 14. Business centre; 15. School centre; 16. City pavilions; 17. Autonomous institutes; 18. University residences; 19. Natural garden; 20. Museum: Asia and Europe sections; 21. Museum: Americas Asia and Oceania sections.

The forms, represented in the drawing, can be easily reduced to the figures of the circle, the square and the triangle, arranged inside a grid, which summarize the meaning of precise functions. Le Corbusier, then, made a three-dimensional transposition of these shapes which became a cylinder, a cube and a pyramid. The buildings of the Mundaneum were characterised by complex ideological contents which explains the articulated and structured genealogy of Le Corbusier's last project (fig.8,9,10,11).

Figure 8: Axonometric perspective on the entire project with numeration; Chinese ink; Leather trasing; 1,36 x 1,07 ; Archives of the "Fondation Le Corbusier"; Paris (©AFCL 24525).

For example, the Museum of human creation, which is the most important building of the complex, had the shape of a spiralling pyramid to respond to "the absolute continuity of events in history" [19].
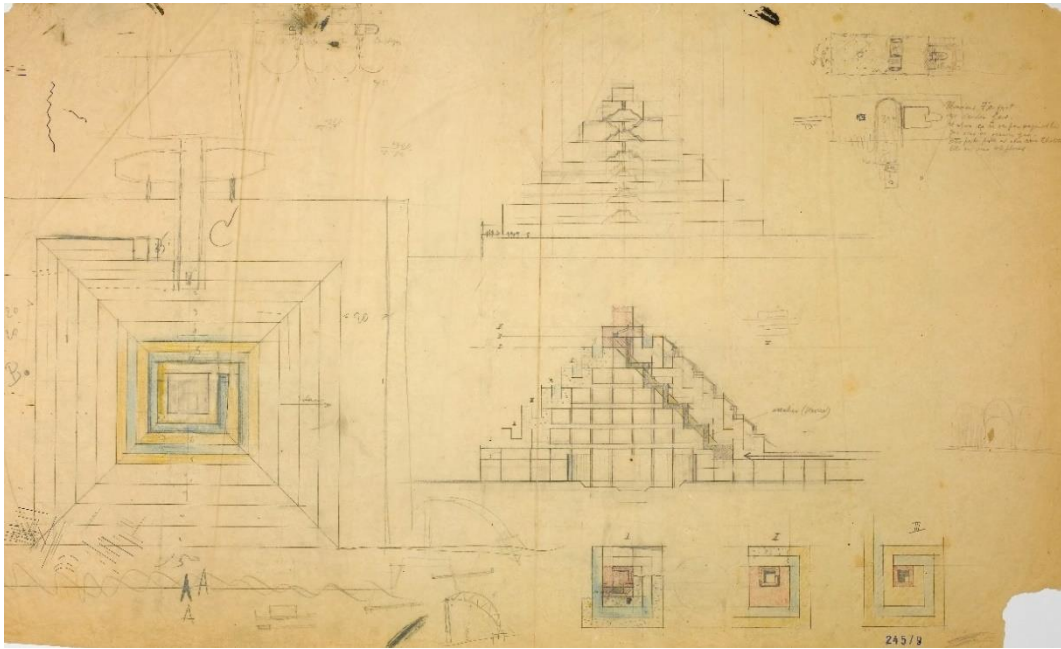
Figure 9: Study drawing, plan and a couple of elevations of the Museum; drafts, calculation and notes; coloured and black pencil; 0,67 x 1,09 ; Archives of the "Fondation Le Corbusier"; Paris (©AFCL 24579).
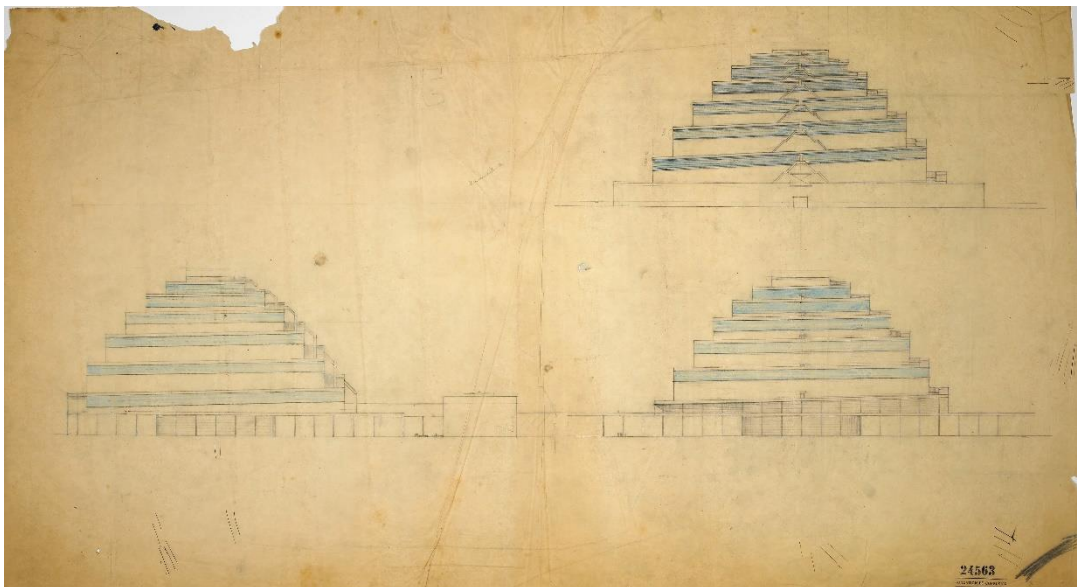
Figure 10: Three elevations of the Museum with shape; study calculations; coloured and black pencil; 0,67 x 1,09 ; Archives of the "Fondation Le Corbusier"; Paris (©AFCL 24563).
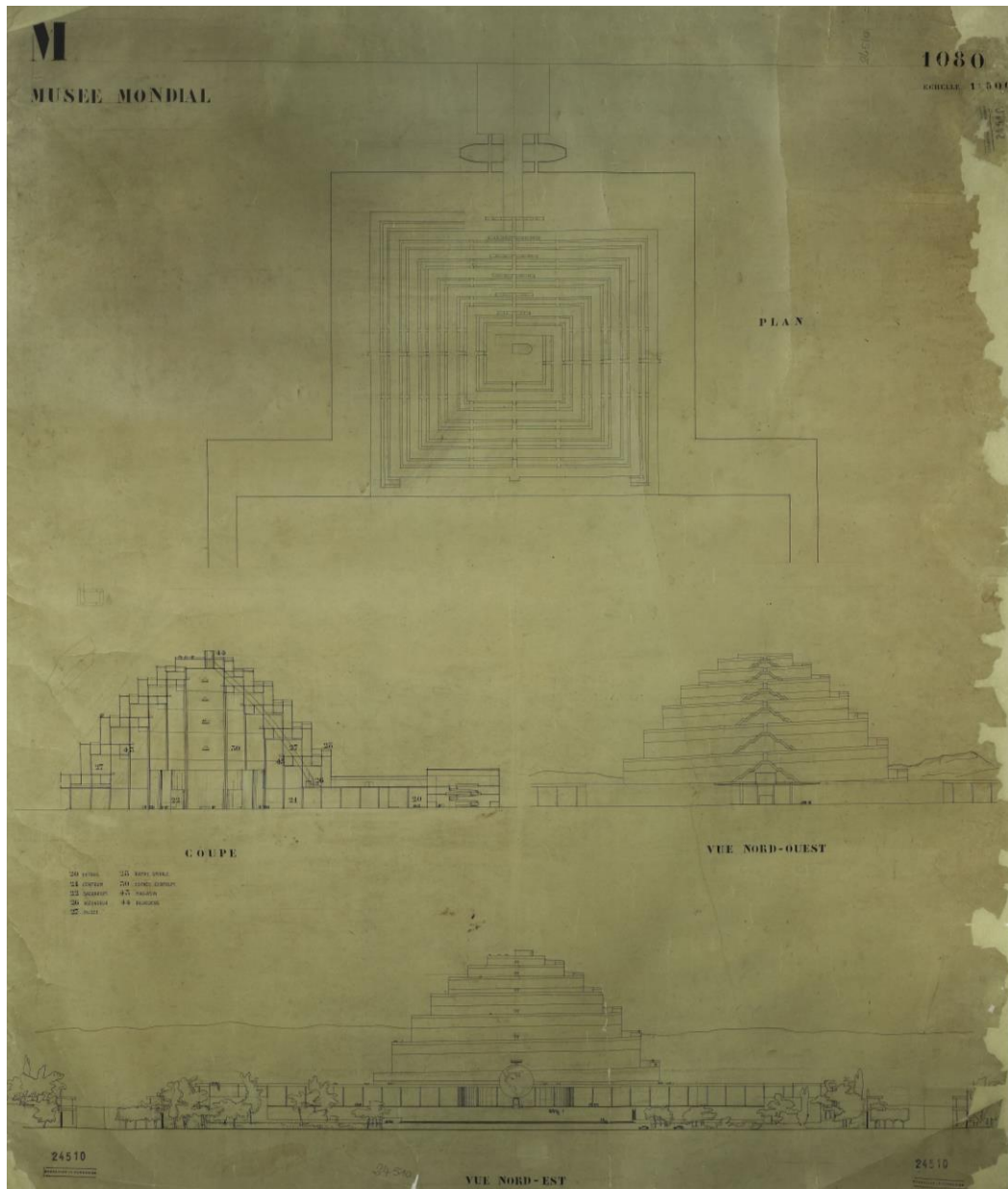
Figure 11: Study drawing, plan, section and elevation showing the museum with legend; coloured pencil; scale 1:500; Chinese ink; 1,11 x 1,40 ; Archives of the "Fondation Le Corbusier"; Paris (©AFCL 24510).

Furthermore, inside the Museum, Le Corbusier designed the construction of a Sacrarium, a symbolic monument which gave a place to the Spirit of History [20].

A precise location was assigned to the complex and articulated project. In fact, it was to be integrated into an already built environment characterised by many consolidated elements such as Saconnex, Pregny and Morilon (fig.12).
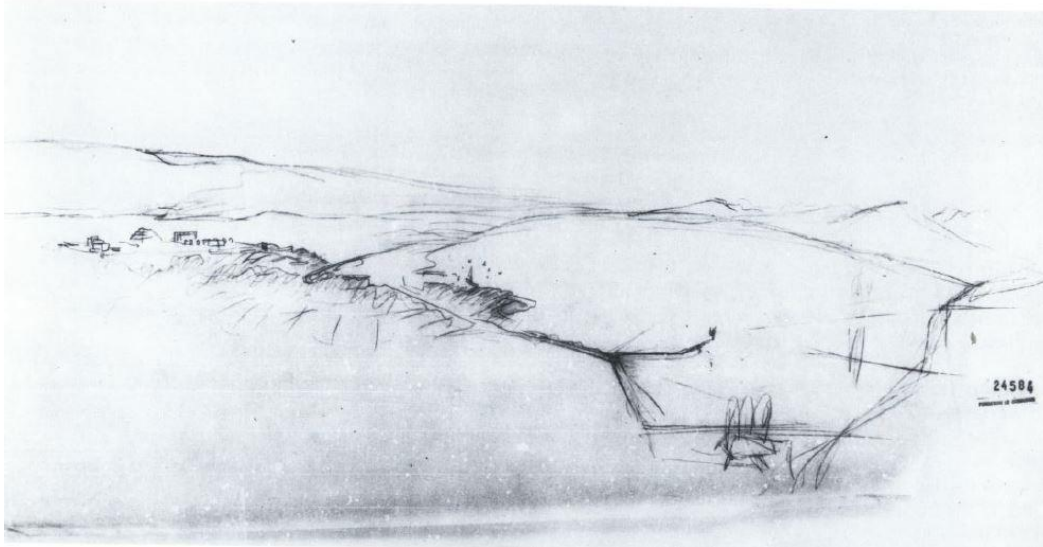
Figure 12: Perspective of the site; July 1928, (©AFCL 24584).

 Its construction was initially financed by donation from wealthy individuals, governments, municipalities and institutions that were firmly convinced of the validity of a world centre for the accumulation, organization, and dissemination of knowledge.

2. Corpus Creation and Analysis

2.1 What is a corpus?

J. Sinclair [21] defined a corpus as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research". Ideally, a corpus is designed to be representative of a language through careful selection, indeed, it is not simply a set of data but an organic and systematic collection of both oral and written material. In other words, the corpus can be formed of simple sentences, found in various contexts, or whole texts, recorded or transcribed, and this enormous amount of material, coming from heterogeneous sources, is provided in electronic form and it is incorporated into computerized databases. The corpora could be monolingual, bilingual or multilingual. The first one represents only one language while, the other ones, represent two or more idioms. A corpus may be open or closed with the difference that, the open corpora do not claim to contain all data from a specific area while, the closed corpora contain all or nearly all data from a particular domain. The choice of using one or the other model depends on the objective of the specific research project. In particular, if the corpus is to be used as an instrument to study the characteristics of a given language in a specific period of time, then the model to be preferred is the closed one, as it is able to offer a "snapshot of the state of a language" at a given time "[22]. On the contrary, if the goal of the corpus is to analyze the evolution of a language over time, the model to be used is the open one, as it able to reflect "the changing state of a language" [22].

A classic research question is how representative a corpus is and the answer depends on how capable it is of describing relevant aspects of the language. This means that corpus selection is a very important aspect both for corpus builders and corpus users, indeed, the knowledge obtainable from its use depends on the set of questions that can be investigated and on the composition of the corpus itself. The choice and creation of the corpus are essential steps because making mistakes in the choices of documents can affect all the subsequent work.

2.2 Corpus Construction and Contents

In this section, we will describe the corpus construction, which will be used to extrapolate sentences containing spatial information belonging to the architectural and urbanistic domain.

In particular, we focus on the following aspects [22]:

- Definition of the characteristics of the reference population from which a significant sample was extracted;
- Definition of qualitative and quantitative criteria for determining the representativeness of the corpus;
- Presentation of the composition of the corpus (types of documents, sources, etc.) and its dimension.

The corpus in question is a specialist multilingual corpus[3] of written texts, formed by four sub-corpora, one for each language (French, English, Italian, Spanish), belonging to the same domain. The existing studies regarding the construction of document collections, present in the literature, are based on the key principle of corpus representativeness [24] with respect to the reference population. The corpus, in fact, must fulfil the role of a representative sample, in the statistical sense of the term, because all the observations obtained from its analysis must be valid and extendable to all the individuals of the population. In particular, the extracted terminology must be as representative as possible of the reference domain. This means that the collected material must form a significant set of the total documents of interest; in other words, the corpus must represent the population but on a smaller scale. One of the criteria that contributes to ensure the representativeness of the corpus is its size, however, there are not yet precise directives concerning the right dimension.

Given the large amount of information and documents in all areas of knowledge, determining the number of all the available sources, for a precise domain, is a difficult and significant issue.

With respect to the criterion of quality, in order to carry out proper terminology work, the texts have to meet several requirements [25]:

I. "The texts should be representative of the period taken into consideration";

II. "The texts should be written by different authors, experts of the domain";

---

3 Multilingual corpora can be parallel and comparable. In the first case, the texts in their native language, are also available as translations into another language, while in the second case, the corpora do not include any translated texts.

III. "The texts should be original and not translations because, in this way, it is possible to ensure that the used terminology is appropriate and specialist";

IV. "The texts, that compose the corpus, should be complete and not fragments of the whole documents".

All these criteria should guarantee the maximum reliability of all the texts which form the entire corpus.

As previously mentioned, it aims to represent and analyse a particular domain, never analysed before for purposes similar to mine: the architectural field of the "World City" project. The collected documents concern the two most important and structured works of the utopian City, which are the Andersen and Le Corbusier's projects. The two architects created their own version of the city, in fact, the former designed a more sumptuous and complex project, while, the latter, planned a smaller city based on a human scale. The two projects, despite their diversity, have some elements in common such as the ideological aims of the city construction or some buildings and institutions; Le Corbusier, in fact, analysed the prior Andersen project re-proposing some aspects and changing others.

2.2.1 Corpus Composition from Sources

Regardless of the unique characteristics of the two projects, both were the result of numerous years of study and research in which they implemented, modified and improved their idea. The work process, as well as the final project, is attested by several documents of different nature and origin.

Primary and secondary sources have been collected after a laborious and meticulous research. They are original books and publications by the authors, letters of correspondence between Paul Otlet and the architects, drafts and drawings, but also books, scientific articles or writings about the topic. The primary sources are stored principally in the archives of the Mundaneum[4] sited in Mons, in Belgium, and in the archives of the "Fondation Le Corbusier"[5], sited in Paris. Conversely, the secondary sources have been found in particular libraries, other archives or on the Net.

The set of these documents composed our corpus, which final purpose is to extrapolate from the texts all the spatial information regarding the setting and the configuration of the city.

---

4 <http://www.mundaneum.org/>

5 <http://www.fondationlecorbusier.fr/>

This permitted us to obtain numerous descriptive sentences which have been subsequently used to create a test and training dataset.

The corpus documents, as already mentioned, could be divided into primary and secondary sources; the first one are letters, books, publications written by Paul Otlet or the architects while, the second ones, are papers, books, journal articles or web documents related to the referent domain.

At the end of the analysis, we collected 29 texts in total, for an amount of 411.401 tokens[6]. The small number of documents depends on the fact that not much has been written about the spatial composition of the World City and increases their importance for the purposes of a spatial reconstruction of the city itself. These documents, originally of different formats (.doc, .pdf, .jpeg) have all been transformed in .txt files; in particular, in the last two cases, has been used the software ABBYY[7], able to provide optical character recognition and document capture.

In the follow, a more specific description of the 4 sub-corpora is proposed.

2.2.2 Construction of the four Sub-Corpora

The English sub-corpus is the most numerous, in fact, the total number of texts amounts to17 units of different length, author and origin. In particular, the population is composed by:

a) an original publication by Andersen;

b) one book;

c) six book chapters;

d) five papers;

e) one journal article;

f) one letter, the only one that describes the spatial configuration of the project;

g) two web documents

for a total of *201,747* tokens.

The French sub-corpus is formed by 6 documents which are:

a) two original publications by Paul Otlet;

---

6 A token is a linguistic unit or even a minimal unit of analysis. Thus, it can be represented by a word, a punctuation element, a date, a number or an acronym.

7 < https://www.abbyy.com/it-it/>

b) one book;

c) one book chapter;

d) one journal;

e) one paper

for an amount of *85,738* tokens.

The Italian sub-corpus has a total number of texts equal to 5:

a) one book;

b) two book chapters;

c) two web documents;

for a total of *105,930* tokens

The last sub-corpus, the Spanish one, is formed by only 2 texts which are:

a) one book chapter;

b) one paper

with an amount of *17,986* tokens.

We believe that the construction of a multilingual corpus is an added value for the representation of a domain knowledge and for information retrieval. The choice to build a comparable[8] corpus, in fact, ensures that we can work on "authentic" words and not on translations of the texts. This aspect responds to one of the qualitative criteria defined in [25].

2.3   Sentence Extraction

One of the most important functions of natural language is to describe spatial relationships between objects through linguistic constructs, which contain numerous references to spatial information. This type of information is easily understandable by the human mind but machines, on the contrary, do not have the same cognitive capacities. The main difficulty for a computer is to identify and extract only the right information from a

---

8 A comparable corpus is a collection of "similar" texts in different languages or in different varieties of a language.

text, leaving out all the other content that, despite enriching the description, is not useful for the 3D scene generation.

In this thesis work, we generate a sentence extractor, written in Python programming language, capable of parsing large data to automatically extract specific sentences from the aforementioned corpus. The main aim is to allow people to automatically obtain the required data instead of reading the entire text to select precise information.

Spatial information is generally provided by the use of prepositions that establish a relationship between two or more objects. However, the same preposition could also be used to talk about events or situations which not involve spatial descriptions. For example, the preposition *on* is used to describe spatial configurations like "the picture is *on* the wall" or "the bottle is *on* the table" but, the same preposition, can be used additionally in completely different contexts such us "They need to concentrate *on* their studies" or "The discussion will be *on* a topic you have studied recently". The first two sentences express spatial concepts, concepts of verticality or objects overlapping, while the last two sentences do not have any kind of reference to entities location. This means that the usage of prepositions depends on several aspects like the entities involved in the scene or the general situation of speech. In other words, prepositions are often used in front of nouns or pronouns to show the relationship between them and other words in the sentence, but they can also be used to describe the time when something happens (*They arrived **on** Sunday*), the way in which something is done (*We went **by** train*) and even more. On this basis, the identification and the extraction of particular sentences from a large number of texts characterised by heterogeneous information is a big and interesting challenge [26].

In this work, a script able to read and analyse a text with the purpose of extracting all the sentences containing spatial descriptions has been created. The latter are composed by three central concepts belonging to the Holistic Spatial Semantic Theory [27] in which the three main spatial roles are defined. They are trajector(s), landmark(s) and spatial indicator(s) that, linked together, generate a spatial triplet. In particular:

- the trajector is a spatial role label assigned to a word that denotes a central object of a spatial scene;
- the landmark is a spatial role label given to a word that signals a secondary object of a spatial scene (to which a possible spatial relation between two objects can be established);
- the spatial indicator is a spatial role label assigned to a word that indicates a spatial relation between objects (trajectors and landmarks) of a spatial scene.

To better understand, let us show an example:

The [car]TRAJECTOR is [under] SPATIAL_INDICATOR the [tree] LANDMARK.

where the 'car' is a trajector object, the 'tree' is a landmark object and 'under' is a spatial indicator.

A key element of the system is a set of text extraction rules that identify relevant information to be extracted from the input text. In particular, the adopted extraction process is based on some word lists belonging to a "spatial domain" and on specific rules which guide the system to identify only the sought information.

The system is subdivided in four sections which, by successive and interconnected steps, lead to a gradually more and more accurate result. The general idea is to start from an entire document, written in one of the four languages, and to parse it in order to identify the possible spatial indicators and labels (trajectors and landmarks) present in it. Of course, the interest is not to find these entities randomly located in the text but to find the sentences in which they are connected to each other. For this purpose, the system firstly divides the entire text in single sentences and, subsequently, analyses them to discard, step by step, those that are not useful for the final goal. The four parts of the system are the following:

1. *Sentence_Split.py* = the system opens a document, in .txt format, and separates strings using two specific delimiters, a dot ('.') or a long sequence of blank spaces. Then, it creates a .txt file (*Sentences.txt*) that collects all the sentences and writes them on individual lines.

2. *Search.py* = the system reads as input a .txt file containing all the possible prepositions or expressions that may be spatial indicators (ex. *on*, *in the centre of*, *to the right, …*).Then, it opens a *Sentences.txt* file and it analyses if each sentence contains or does not contain one or more spatial indicator(s). Finally, it creates, as output, a .txt file (*outfile.txt*) including only the phrases which contain spatial indicator(s).

3. *CountOccurrences.py* = the system reads two .txt files, namely, the previous *outfile.txt* and *domainNouns.txt* which stores terms or multi-words that may be trajector(s) or landmark(s) of the sentences. The latter are all written in lowercase therefore, the next step of the system is to replace all the capital letters, present in

outfile.txt file, with lowercase letters in order to be case-insensitive. In the end, the system counts the number of labels occurrences (number of time trajector(s) and/or landmark(s) appear in each sentence) and, if it is greater than two, it extrapolates the sentences from the *outfile.txt* file and writes them into another .txt file (*labelSentences.txt)*. In this way, it is hoped that at least one label is the trajector of a sentence and another one is its landmark which are connected through a spatial indicator detected in the previous step.

4. *Last.py* = in this phase the system improves the results establishing which are the positions that can be occupied by the labels in order to be possible trajector(s) and landmark(s). The locations are set up with reference to the positions of the spatial indicators taking into account some sentence structure rules. In the following table (tab. 1), for example, the main rules adopted to identify a hypothetical landmark in a sentence are illustrated:

a)	**Spatial Indicator**	landmark		
b)	**Spatial Indicator**	Article	landmark	
c)	**Spatial Indicator**	Article	adjective	landmark
	Spatial_indicator Position: s_i	**Landmark position: s_i + 1**	**Landmark position: s_i + 2**	**Landmark position: s_i + 3**

Table 1: Landmark possible positions.

More specifically, by referring to the location of the spatial indicator (signed as *s_i)* the landmark may be located in the next position *(s_i + 1)* or in *(s_i + 2) / (s_i + 3)* position. Sample sentences are:

*a)*A [zoo] ᴛʀᴀᴊᴇᴄᴛᴏʀdesigned for children is located [in the south of] ꜱᴘᴀᴛɪᴀʟ ɪɴᴅɪᴄᴀᴛᴏʀ[London] ʟᴀɴᴅᴍᴀʀᴋ;

b) The [fork] ᴛʀᴀᴊᴇᴄᴛᴏʀis [to the left of] ꜱᴘᴀᴛɪᴀʟ ɪɴᴅɪᴄᴀᴛᴏʀ[the] ₐᵣₜᵢ𝒸ₗₑ [plate] ʟᴀɴᴅᴍᴀʀᴋ;

c) Put the [box] ᴛʀᴀᴊᴇᴄᴛᴏʀ[on] ꜱᴘᴀᴛɪᴀʟ ɪɴᴅɪᴄᴀᴛᴏʀ [the] ₐᵣₜᵢ𝒸ₗₑ [wooden]ₐ𝒹ⱼₑ𝒸ₜᵢᵥₑ[table] ʟᴀɴᴅᴍᴀʀᴋ!

A further case regards the absence of the landmark in the sentence; this happens when the description is characterised by the presence of an implicit landmark like in the sentence: The [balloon] ᴛʀᴀᴊᴇᴄᴛᴏʀflew [up] ꜱᴘᴀᴛɪᴀʟ ɪɴᴅɪᴄᴀᴛᴏʀ. The landmark, in this case, is NIL (void).

Once these rules have been taken into account, the system screens the input file (*labelSentences.txt*) and generates a final file (*Last.txt*) as output. This file shall collect all the sentences in which spatial descriptions and spatial relationships appear.

The four abovementioned steps are shown in the table below (tab. 2):

	SCRIPTS	ACTION	INPUT file(s)	OUTPUT file(s)
1-	**Sentence_Split**	**Subdivision** of the whole documentinto **singlesentences.**	*Originaldocument (File.txt)*	*Sentences.txt*
2-	**Search**	**Selection** of the sentences which **contain** one or more **spatial indicators.**	*Spatial_indicator.txt Sentences.txt*	*Outfile.txt*
3-	**CountOccurrences**	**Count** of the number of times that **words** representing trajetors or landmarks appear in each sentence. If num>2 the sentence is extracted.	*DomainNouns.txt Outfile.txt*	*LabelSentences.txt*
4-	**Last**	**Application of rules**: assignment of the **positions** that can be occupied by the labels in order to be possible trajector(s) and landmark(s)	*LabelSentences.txt*	*Last.txt*

Table 2: Summarization of the four script steps.

2.3.1 Evaluation of the Results

The system worked on the abovementioned corpus and extrapolated 705 sentences that may be classified into two different categories: "useful extractions" and "useless extractions". In particular, the sentences belonging to the first category are those that have been extracted by the system and that really contain spatial information, while, the ones belonging to the second category, are the sentences extracted even if they do not contain spatial data. The number of "useful extraction" sentences is 591, while, the amount of "useless extraction" sentences is equal to 127 units. In addition, there is a third category that includes all the sentences not detected by the system which contain spatial information; in other words, the missing data. They are, in total, only 75, as depicted in the following figure (fig.13).

The reasons why the system has not been able to extrapolate these sentences from the corpus are various, but the main ones are basically two:

- the presence of landmarks or trajectors implied within the sentence;

- the use of personal pronouns (e.g. it/ they), indefinite pronouns (e.g. one/ ones/ both), object pronouns (e. g. it/them) demonstrative adjectives (e.g. the latter/ this/ that/ these/ those) in place of nouns indicating a landmark or trajector.
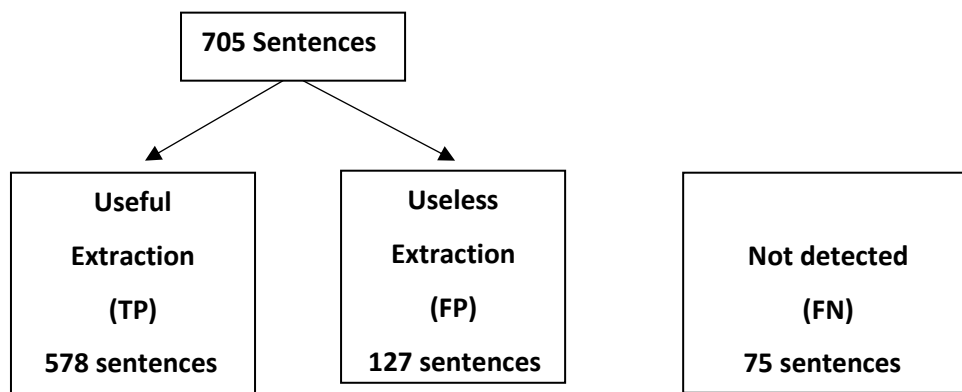
Figure 13: Schematization of the results.

The results have been estimated by the evaluation metrics of precision and recall defined as:

$$precision = \frac{TP}{TP+FP} \qquad\qquad recall = \frac{TP}{TP+FN}$$

where:

TP = is the number of system-extracted sentences that contain spatial information;

FP = is the number of system-extracted sentences that do not contain spatial information;

FN = is the number of sentences containing spatial information that the system does not extract from the texts.

The following table (tab. 3) shows the results of the system run:

PRECISION	**0,820**
RECALL	**0,885**

Table 3: Evaluation metrics of precision and recall.

These results certify that the used system has an adequate level of reliability because it is capable of responding to the task with a limited margin of error.

2.4   Sentences-corpus compilation

Once the sentences containing spatial information were extrapolated from the corpus, they have been used to create our *sentences-corpus,* which is composed by the above mentioned 705 sentences. The latter have been useful to train and test a probabilistic programming language that is able to automatically detect and extract, from an input text, trajector(s), landmark(s) and spatial indicators(s), together with the spatial relationships. In particular, the sentences-corpus has been subdivided into two different datasets: the training set, made up of 80% of the total (564 sentences), and the test set, made up of the remaining 20% (141 sentences).

The automatic extraction of these elements will be described in the fifth chapter (§ 5.5 Evaluation of the Results) of this thesis.

3. Spatial Information Extraction

3.1 Introduction

One of the most important goals of artificial intelligence in Natural Language understanding is to describe the location of objects in space and their transformations. Several applications, in fact, necessitate spatial knowledge to work, such as geographical information systems, prevision of navigational instructions to robots or text to scene conversion systems.

To our knowledge, there have been few studies based on spatial machine learning in the sense of the extraction of spatial information from natural texts. As we know, natural language is highly complex, thus, the task of allowing machines to understand it is a big challenge to face. But, what type of linguistic knowledge should a machine have to be able to elaborate our language?

Certainly machines must have a complete knowledge of language structure and organization (its lexicon and morphology), they must recognize linguistic expressions like sentences or phrases (syntax), they must assign knowledge to simple or complex linguistic expressions (semantic) and they should use sentences in the right contexts and in the correct way according to a precise communicative task (pragmatics) [28].

3.2 Motivation and Challenges

When we talk or write to someone, we use our natural language which, for people, is the easier and more intuitive way of communication. Many times the sentences contain spatial information that help to describe places or situations in common life, information that the human mind can immediately understand, while, machines cannot do so. On these basis, the task of this work consists in transferring human knowledge to the machine in order to allow an automatic extraction of spatial information from natural language. The following example better illustrates our purpose:

E1:   *In the evening put my magazine on architecture on the bedside table.*

Let us analyse the sentence that might be a command given to a robot; to understand it we have to know that a sentence could contain different types of information such as:

- spatial information;

- spatial and temporal information;

- spatial, temporal and context information;

- etc…

Hence, allowing the machine to recognize and to extract from the text only spatial information is a big challenge for machine learning techniques.

A useful approach to face this problem is based on the definition of spatial primitive and notions that are *trajector* and *landmark* and the spatial indicator that links them, as described in the previous paragraph. In the above example "magazine" is a trajector and "bedside table" is a landmark while the preposition "on" is the spatial indicator that defines the region of space in which the magazine should be placed with respect to the bedside table. An important aspect to underline is that we have to find a methodology able to automatically map only the parts of the sentence which express spatial information by excluding the other ones. In our examples the machine has to understand that "evening" and "architecture" are not landmarks and that neither "In" nor "on" [architecture] are spatial indicators.

Once having recognized the spatial elements and their roles (for example "magazine" is the trajector and not the landmark) we could represent them through an abstract scheme like *on(magazine, bedside table)*. The further step is to connect them via a particular formal spatial meaning representation, like *Externally-connected (magazine, bedside table)*, which suggests that the magazine is directly connected to the bedside table (it rested on it) and it is not located near/above or under it. Following these steps, we assume that spatial notions will be managed by machines.

The next section is dedicated to a brief explanation of the main challenges that emerge from this methodology:

1. The first problem of spatial primitive and notions extraction from natural language regards the application field. As we easily understand there is a large difference between the textual description of buildings and monuments, for example in a square, objects in a room or entities in a fable, because the domain changes and, therefore, also the trajectors and the landmarks to be recognized by the machine. For this

reason, a detailed knowledge of the domain of interest is needed because, only in this way, could a computer be able to extract the correct information instead of erroneous entities.

2. A univocal representation of the knowledge does not exist because it depends on the characteristics of the space. If we consider, for example, the following sentences "*the vase is on the carpet*" and "*the vase is on a windowsill on the carpet*" certainly we expect a completely different representation of the scene. Indeed, in the first case, our mind would imagine that the vase is in contact with the carpet, while, in the second one it would suppose that it is in contact with the windowsill and, the latter, is on the carpet. Consequently, in the first example, the vase should be connected to the carpet whereas, in the second one, it should be disconnected from it. In the sentence "*the vase is on a windowsill **on** the carpet*" the spatial entity *above* would be more appropriated, compared to the preposition "on", to express an *externally connected* relation between the two objects. In common language, however, people often do not pay attention to the most correct way to describe a spatial situation, but they use everyday language which could cause several misunderstandings.

3. This simple example shows that we are in need of a spatial formalism to deal with semantic problems.

4. One of the main problems of natural language is certainly its ambiguity and polysemy because a single element of a phrase could have many different meanings according to the context and to the domain. In the previous example (E1), indeed, the preposition "on" followed by "architecture" is not a spatial preposition while, the one followed by "the bedside table" is the link between the "magazine" trajector and the "bedside table" landmark.

5. This is a crucial aspect of information extraction techniques, so, to skip the obstacle, we have to give some clues to the machine in order to identify and label the right elements.

6. Nowadays there are some annotated data relating to spatial notions but they belong to generic fields and not to an architectonical domain or, even more, to the "World City" domain. For this matter we need to create our own dataset formed by annotated sentences in which elements related to our field appear.

3.3   Spatial Information Extraction Methodologies

The strategies and instruments used in this work to face the above-mentioned difficulties are the usage of a spatial annotation scheme together with a detailed study of the properties of natural language. The two methodologies are briefly described below:

- **M1:** A **spatial annotation scheme** able to cover both dynamic and static spatial notions has been used, even if, in this work, we only focus on the latter.

- **M2**: An in-depth study on the **characteristics of spatial language** and on its **features** has been carried out in order to comprehend them and to highlight the generic error types. Furthermore, we performed the information extraction by enriching background knowledge with architectonic data.

3.4   Human Language VS Computer Code

Natural language is undoubtedly the main instrument and the most suitable to communicate contents among people. Nowadays, however, the overwhelming role of computers in our daily life is evident because they are able to store and process large amounts of structured data and make them available in the form of information. The society in which we live is often defined as a knowledge society because we have a daily need to locate and manage large amounts of information contained in texts on the web, in our computers, in digital libraries, but also in e-mail, clinical records, minutes, reports, and more. The large amount of information is mostly in the form of unstructured data, inevitably difficult to access, so, we need of tools able to extract interesting data through a "sifting" process of coded information in the text.

The development of automated systems that can understand natural language is a topic of great relevance and interest due to the complexity of human language which makes it extremely challenging to model. The main difficulty is the fact that computers do not have

the ability to analyse human texts, but they simply analyse the code. This aspect implies the need to transfer appropriate knowledge to the machine in order to allow them to interpret human language in a specific context and to interact and communicate with people. Computers can store and process data only in the form of two symbols, 0 and 1 (bits), that are aggregated into 8-digit (bytes) sequences. For this reason, all types of text, as well as images, sounds or numbers, must necessarily have a binary representation. Unfortunately, text encoding is not a simple bit sequence (zero-level encoding) because it also envisages the enrichment of the text with information about the linguistic-textual structure to permit a better text interpretation (high level encoding). Indeed, in the absence of a proper structure of the text (subdivision into chapters, sections, paragraphs, etc.), linguistic structures (morphological, syntactic and semantic information) and relationship between words, one would inevitably encounter a loss of information. The texts, in fact, hide a series of linguistic relationships which, in order to be made accessible to the computer and subsequently used for further research, must be explicitly encoded [29].

On these assumptions, the annotation process has acquired a predominant role because it allows to make explicit an implicit linguistic structure rendering it computable. Semantic annotation is the process attributing meaning to a text through which it is possible to assign a specific sense to the phrases, excluding the possibility of their multiple interpretations. Natural language, in fact, in the absence of specific techniques, is an extremely complex form of communication, rich in ambiguity and difficult to interpret through automated processes.

3.5   Natural Language Processing

The expression *Natural Language Processing* (NLP) refers to a particular research field in which a computer has the ability of analysing and understanding natural speech. It is a very challenging task because, as we know, a machine requires a programming language to comprehend human language without misunderstanding.

NLP was born in the years of World War II, with the electronic calculator and with the first automatic translation studies. Over the following years, it was a topic addressed by cybernetic scholars and, later, by *Artificial Intelligence* (*AI*) researchers, which aimed at a specialization on "communicative" intelligence by creating instruments able to analyse

language. In other words, they focused on the creation of automatic grammatical, syntactic and semantic analysers [30].

The NLP literature shows syntactic analysis systems that refer to different grammatical theories; there are several solutions for semantic analysis, while, there are still no pragmatic language processing tools. Nowadays, by utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation.

According to John Rehling, an NLP expert at Meltwater Group, "Apart from common word processor operations that treat text like a mere sequence of symbols, NLP considers the hierarchical structure of language: several words make a phrase, several phrases make a sentence and, ultimately, sentences convey ideas"[9] .

NLP is used to analyse text, allowing machines to understand how humans speak but it has to face a difficult problem in computer science: human language is rarely precise, or plainly spoken. To understand it, in fact, it is necessary to understand not only the words, but also the concepts and how they are linked together to create meaning.

The majority of NLP approaches are based on machine learning algorithms, instead of hand-coding large sets of rules, that try to automatically learn these rules by analysing input data and making a statistical inference. Generally, in structured learning models, the output variables depend on the application of some constraints techniques which are able to optimize the results [31]. In the last years, for example, high importance was attached to relational learning [32] [33] [34] [35] where the words of the sentences are considered like entities linked to each other and labelled as person, location, facility, etc. [36] [37] and are stored in a relational database.

Generally, the model will be more accurate the more data are analysed. Common tasks of NLP software programs are:

- Text Summarization to extract the most important and central ideas from block of texts ignoring irrelevant information;
- Sentence segmentation, part-of-speech tagging and parsing [38];
- Named entity recognition and extraction [39];

---

[9] <http://mashable.com/2011/11/08/natural-language-processing-social-media/#5OkOoLaLMEqC>

- Automatic keyword tags generation from content;
- Web document classification based on the relationship between webpages via hyperlinks [40];
- Co-reference resolution;
- Sentiment Analysis to identify the sentiment of a string of text, from very negative to neutral to very positive.

Broadly speaking, NLP can be divided into two distinct areas: generation of text (synthesis) and comprehension of it (analysis). For *text generation* we mean the creation of a text, based on a set of concepts to express, that respects the laws of the language in which it is created. An example of application can be the generation of responses in man-machine dialog, translation from another language, or the creation of a summary [41]. Understanding the text, in this context, is interpreted as annotating or extracting its conceptual content based on phonetic, grammatical, syntactic, semantic and pragmatic or contextual rules, or even based on statistical processes.

NLP, as well as for the above applications, is used in information retrieval and information extraction techniques and in text correctors (lexical, grammatical, syntactic and stylistic). Indeed, it is used in writing scripts for word correction, in automatic translation or in the creation of human-machine interface systems based on dialogue.

The NLP process is generally divided into four main phases [42]:

1. *text pre-processing;*
2. *lexical and morphological analysis;*
3. *information extraction;*
4. *text analysis.*

The first stage of *pre-processing* is aimed at the correct scanning of the text according to the chosen type of analysis. The preliminary tasks are to "clean" the text and to normalize it in order to standardize spaces, apostrophes or accents and to recognize a priori particular entities in the text (dates, numbers, currencies, abbreviations, names or expressions). This section also includes a text annotation consisting in associating the meta-information to the words present in the document, namely, the grammar category, the lemma, the number of occurrences of a given word in the corpus, some morphological characteristics and so on.

The *lexical and morpho-syntactic analysis* provides a pragmatic representation of the corpus, studying its lexicon. At this stage, the development of the speech is not taken into account, but words are extracted from an urn, which in this context is called the "Bag of Words". In this phase the "full" words in the text are defined, such as verbs, adverbs, nouns, adjectives, and the most frequent are highlighted. Subsequently, the "empty" words are also analysed such as prepositions, conjunctions, determinants, etc. Thanks to this type of analysis, using statistical techniques, it is possible to describe some constants of the language like the incidence of some word classes in the corpus.

The third phase, *information extraction*, allows us to focus our attention on the part of language that is particularly significant. It is a very important moment in textual analysis, useful to select a small percentage of the reference vocabulary in order to carry out textual analysis. At this stage Information Retrieval techniques are applied with the purpose of providing the users answers in response to specific requests (such as search engines, for example).

The final step is the *textual analysis* phase which concerns all the operations directly linked to the corpus such as: identification of lexical structures, disambiguation, information extraction, research of words of interest and so on.

3.5.1   Language Analysis

The above-mentioned four steps, as already said, allow machines to handle natural language in order to give answers to several users' needs. In this section, we focus particularly on the lexical, syntactic and semantic analysis which are briefly overviewed.

Morphological and Lexical Analysis

A vocabulary, composed of numerous expressions, words and tokens, is nothing more than the lexicon of a language. In turn, a word is a linguistic unit with a specific meaning which is formed by a set of characters called tokens. The latter are the smallest units of analysis [43].
The purpose of lexical analysis is to split the text in paragraphs, sentences and words taking care to separate the non-words tokens, like punctuation marks or blank spaces.

Syntactic Analysis

A complete syntactic analysis is based on three different processes which are *Part-of-Speech Tagging, Parsing* and *Dependency-based Parsing.*

- Part-of-speech tagging (POS)-

Language is generally characterised by two different types of lexical classes which are open and closed classes. The latter contains a fixed number of words like articles, pronouns or auxiliary verbs while, the former is not composed of a known number of words because it contains adjectives, nouns, verbs and adverbs. This said, the task of POS tagging is to assign the right lexical category to each word that appears in a text. This is not trivial matter, as natural language is full of ambiguities and the same words, depending on the context, can have more than one meaning.

A POS tag example is shown in the following simple sentence:

The car is on the street on my right.

In this case the part-of-speech tags are the following:

DT/The NN/car VBZ/is IN/on DT/the NN/street IN/on PRP$/my NN/right./.

The tags have been assigned using Penn Treebank standard tag set[10].

- Parsing -

The process consists in translating a flat input into a hierarchical structure, which represents the meaning elements in an input sentence. The two principal results are a p*arse tree,* formed by the units identified during the part-of-speech process, and a *dependency structure representation*. The parse tree of the above sentence is shown in the following figure (fig.14). It is composed by a root, the sentence, and some leaves the nodes, which are the various tokens. Moreover, abbreviations, like DT (determiner), NP (noun phrase), VB (verb phrase) or AUX (auxiliary verb) and so on, appear in the branch nodes.

---

[10] <http://cogcomp.cs.illinois.edu/demo/pos/results.php>
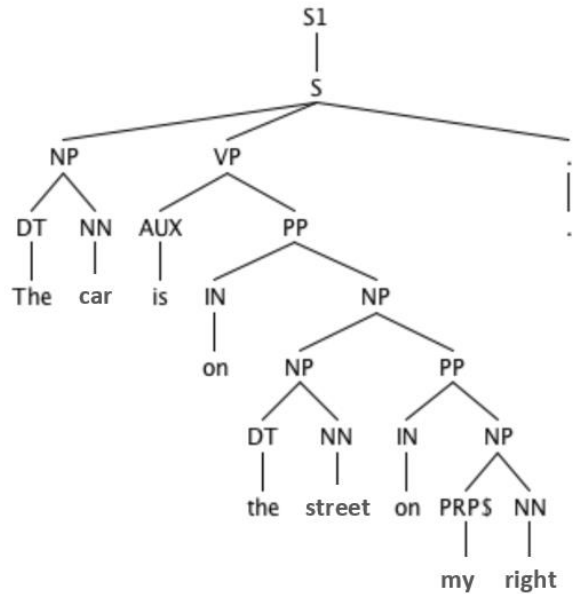
Figure 14: Parse tree representation including POS tags.

- Dependency-based parsing -

In the dependency-based parsing (fig.15) all the links between the elements or items in the sentence are shown. These types of relations, called dependencies, are binary asymmetric relations that connect a head and a dependent. The principal difficulty of the process is to identify the head and, subsequently, to differentiate it from the dependent. There are several methods that can be used to achieve this, for example, it is possible to group all the words belonging to the same syntactic and semantic category and exchange it with the head of the group itself.
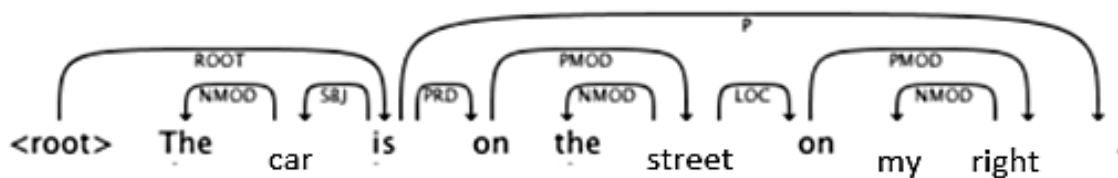
Figure 15: Dependency-based parsing representation with dependency relations.

Semantic Analysis

Semantic analysis is the process of establishing the meaning of the words contained in a sentence, in an oral discussion or in a written text. According to Charles W. Kreidler[11], semantics is the systematic study of meaning and it differs from linguistic semantics that is the study of how languages organize and express meaning [44].

It is based on the study of the relationship between language and meaning with reference to the entire vocabulary of the language, that is, all words that make up a language. Very often, however, a smaller vocabulary is taken into account as compared to the vocabulary of the whole language. It is a smaller subset of the lexical semantics, a knowledge base, in which the individual terms are associated with semantic fields, sets of themes or topics.

Semantic analysis is the search for the meaning of a term or phrase and it is a process of interpretation that men use when they communicate with others. It tries to make sense of the words, giving it the best possible meaning but it is a challenging task due to different reasons:

- some words have more than one meaning, depending on the reference domain; for example, the word "bass" can indicate both a fish and an instrument, as shown in the figure below (fig.16).

Bass: fish   Bass: instrument

Figure 16: Words' different meaning, depending on the domain.

This means that, to understand the correct meaning of the term in the sentence, we need to receive additional contextual information. The process of selecting the meanings is called *Word Sense Disambiguation* and it tries to erase the semantic error that is committed when people use one word instead of another, associating an incorrect meaning to the term;

---

[11] Charles W. Kreidler is Professor Emeritus of Linguistics at Georgetown University.

- the same concept can be defined with multiple synonyms while maintaining the same meaning. For example, the word "building" can be easily replaced by terms like "construction", "edifice", "structure" or also "home";

- sometimes the words do not indicate precise or physical objects but abstract concepts such as "fantasy", "imagination", "solitude", "beauty", "goodness" and so on.

An important process used to understand natural language is **Semantic Role Labeling** [45] which, through the comprehension of events and participant in a sentence, helps to answer the question "Who did what to whom …and where". Indeed, it is based on the detection of the Agent, the Predicate, the Theme and, sometimes, the Location in a sentence by exploiting the fact that, generally, the *agents* tend to be the subject of an active sentence, the *themes* the direct objects and so on. The relationships between these entities are codified in databases like PropBank[12] and FrameNet[13] [46], which uses frame-specific semantic roles.

Each verb in the sentence is annotated with a structure that takes the name of semantic frame and consists in a predicate label, indicating the sense of the verb, and a number of arguments called semantic roles. The latter designate the roles of the items present in the sentence with reference to the predicate.

Two different annotated sentences are shown below:

<div align="center">

The bottle   *is*   *on*   *the table*
Agent[A0]   V:be   patient[A1]

The conference   *is*   *on*   *February*
Agent[A0]   V:be   temporal[AM-TMP]

</div>

There are two important approaches for semantic analysis that are: *meaning representation* and *lexical semantics* [47]. In the first approach, the meaning of an expression is enclosed in formal structures, while, lexical semantics aims to detect what individual lexical items mean and their correlations with the structure of the language.

---

[12] <https://propbank.github.io/>
[13] <https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>

3.6   Spatial Annotation Scheme

The principal difficulties in using machine learning to extract spatial information from natural language are the following:

1. the absence of a common semantic model for spatial information;
2. the existence of several formal spatial representation models;
3. the difference between natural language expressiveness and formal spatial representation models;
4. the absence of annotated data, useful to machines for learning and extracting the spatial entities and spatial relations.

This thesis section and the following fifth chapter have been developed in collaboration with the KU University of Leuven[14] and the Tulane University[15] in order to define a spatial annotation scheme, that will help us to support the main spatial semantics aspects, and to use a recent declarative learning-based programming language useful to design complex joint inference and learn models [48] The scheme takes into consideration both static and dynamic spatial relations, even if, in this thesis work, we only faced the former . The annotation scheme is based on *holistic spatial semantics* together with *qualitative spatial reasoning* models and is used to tag all the spatial roles and their relations in a text. The final goal is to generate a large corpus related to the spatial language which is useful to train the machine learning techniques, to connect language to formal spatial representation models and to estimate its performance.

The annotation scheme takes into account the two most important aspects of spatial information which are:

- how people express spatial concepts in natural language;
- how the formal models are designed in order to represent spatial knowledge representation.

Moreover, it is divided in two different parts based, respectively, on *holistic spatial semantic theory* and on *qualitative spatial reasoning* (SQR) *models.*

---

[14] Prof. Marie-Francine Moens
[15] Prof. Parisa Kordjamshidi and her research group

3.6.1 Holistic Spatial Semantics

The Holistic Spatial Semantics theory is an approach to spatial semantics that has the utterance of a discourse as its main unit of analysis, rather than the isolated words. The main spatial roles considered in the scheme are the following [49], some of them already described in the previous paragraph:

- **Trajector (TR):** it is the entity characterized by a location description. It can be a person, an object or an event and it could be static or dynamic;

- **Landmark (LM):** it is the reference entity because the position or the motion of the trajector is referred to it. Like the trajector, it can be a person, an object or an event;

- **Spatial indicators (SP):** it is the element that defines spatial properties or constraints of the trajector with respect to the landmark;

- **Region:** it indicates a region of space referred to a landmark and it can assume two different values: interior and exterior;

- **Path:** it is a characterization of the trajector motion with respect to a region defined by the landmark. The motion can be real or virtual and rich or minimal. In particular, the minimal path is characterized by its beginning, middle and end and it is enriched once it is combined with region or place information;

- **Motion:** the concept indicates if there is or there is not a perceived motion;

- **Direction:** it represents a direction along an axis;

- **Frame of Reference (FoR):** it denotes one or more reference point(s) or, even better, a coordinate system characterised by axes and angles which is used to identify the location of an object. There are three different reference types which are: intrinsic (object-centered), relative (object-centered) and absolute (geo-centered) [50]. In particular:

  - *Intrinsic frame of reference* is a binary spatial relation in which the position of an object is defined in relation to another object or to a part of it;
  - *Absolute frame of reference* is a binary spatial relation in which the location of an object is defined in relation to an absolute system (for example the cardinal points);

- *Relative frame of reference* is a ternary system where the position of an object is expressed in relation to the observer's viewpoint and the location of another object. The three FoR can be represented as shown in the following figure (fig.17):
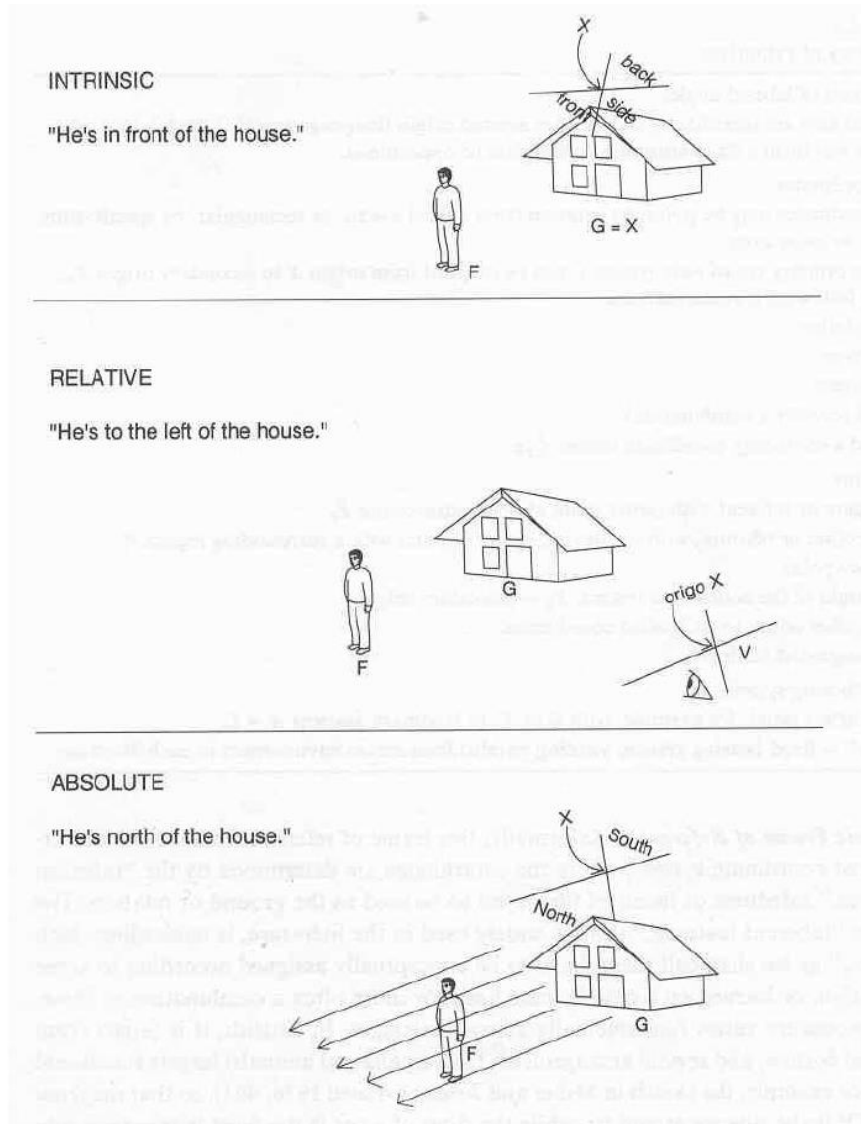
Figure 17: Representation of the three linguistic FoR.

The link between the first three roles (trajector(s), landmark(s) and spatial indicator(s)), defines a spatial relation that, when it is formed by only three arguments, is called *spatial triplet*.

The application of these theoretical concepts to linguistic description is an interesting challenge due to the words' polysemy and the many-to-many mapping between semantic concepts and form classes [27]. This means that the same word can be a trajector or a landmark or can be used to express the concept or a region or a path.

In the table below (tab. 4) the relational representation of the annotation scheme is illustrated together with an example:

-1-	TRAJECTOR(idT,token)
-2-	LANDMARK(idL,token,path)
-3-	SPATIAL_INDICATOR(idI,token)
-4-	MOTION_INDICATOR(idM,token)
-5-	SR(idS,idT,idL,idI,idM, gtype, stype, sp_value,f_o_ref)

Table 4: Relational representation of the annotation scheme.

where a *token* can be a word or multi-words.

Every token identifies a precise spatial role and is characterised by a unique key. A token can play more than one role, in fact, it can be, in the same sentence, a trajector and a landmark by taking part in different spatial relations. In the above table (tab. 4):

- the identifier idT, in relation -1- , represents a token that plays the role of a trajector;

- the identifier idT, in relation -2-, plays the role of a landmark and it is characterised by an additional element: the path. Its value can be {BEGIN, MIDDLE, END, ZERO} in base of "the point of the route"; in particular ZERO means that the path is not significant and it is often omitted;

- in relation -3-, the identifier indicates a spatial relation called *spatial indicator* which is generally represented by an English preposition and establishes a link between the trajector and the landmark which is independent from its semantics;

- in relation -4-, the identifier denotes the presence of any kind of motion described in the sentence;

- the last relation, -5-, shows the connection between all the elements mentioned above which compose the spatial configuration. It is characterised by its own identifier, idS, trajector, landmark and spatial_indicator identifiers, and additional attributes such as general_type (gtype), specific_type (stype), spatial_value (sp_value) and frame_of_reference (f_o_ref).

Ex.	*At 7.00 am Maria was already at the airport, so, she left Rome very early in the morning.*

<TRAJECTOR id='1'> *Maria* </TRAJECTOR>
<TRAJECTOR id='2'> *she* </TRAJECTOR>
<LANDMARK id='1' > *airport* </LANDMARK>
<LANDMARK id='1' path='BEGIN'> *Rome* </LANDMARK>
<SPATIAL-INDICATOR id='1'> *at* </SPATIAL-INDICATOR>
<MOTION-INDICATOR id=`1'> left </MOTION-INDICATOR>
<SR id=`1' trajector=`1' landmark=`1' spatial-indicator=`1' motionindicator=`
  NIL'/ general-type=`Region' specific-type=`RCC-8 spatialvalue=`
  EC' frame-of-reference=`intrinsic/>
<SR id=`2' trajector=`2' landmark=`2' spatial-indicator=`NIL' motionindicator=`
  1'/ general-type=`Direction' specific-type=`Relative' spatialvalue=
  NTPP' frame-of-reference=`absolute/>

For a better understanding of each component of the annotation scheme, it is important to know the meaning of *general-type*, *specific-type* and *spatialvalue* which will be briefly described in the following sections.

3.6.2   Qualitative Spatial Representation

The second and last part of the annotation scheme is founded on the qualitative spatial reasoning (SQR) model which have the characteristic of being independent from natural language. Indeed, they are based on logical, algebraic or geometrical spatial semantics. The adopted qualitative approach has two main points of strength:

It is a flexible model, able to deal with incomplete knowledge [51];

It has a computational inference over relationships and concepts that is very close to human mental conceptualization.

The method is characterized by three relevant aspects of spatial information which are topological, directional and distal [52]. The latter, in addition to the shapes, the morphologies, the sizes and the motions of the objects, are able to specify the positions and the characteristics of the entities in the space.

In the following, we briefly describe the foundations of spatial meaning representation and qualitative calculi models already existing and we propose a link between the spatial triplets and the type of spatial relationships, called *general-type* of the spatial relations.

Topological Relations

An essential aspect of spatial knowledge is the distinction between topological relationships and spatial entities. In the context of reasoning models each spatial entity, involved in a space, is not considered a simple point, but it is valuated as a region that constitutes a subspace of a topological space [51]. The eight most important relations, defined in all topological models and based on *parthood* and *connectedness* relationships, are: DC, EC, PO, EQ, TPP, NTPP, TPP^{-1} and NTPP^{-1} (fig.18).
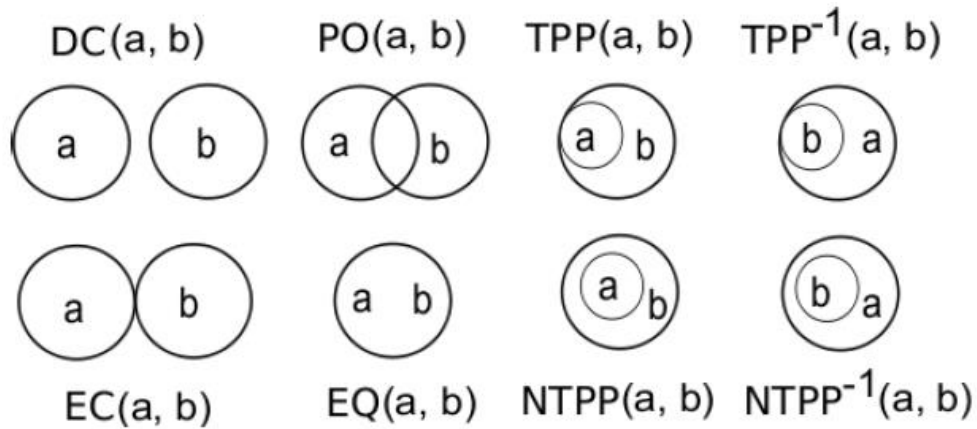
Figure 18: The RCC-8 relations.

The Region Connection Calculus by Randell et al. [53], called RCC-8 model, is the main approach in this domain which represents the topological relations used in natural language. In the above figure (fig.18) some relation symbols that are the synthesis of their spatial meanings are shown: disconnected DC(a; b), externally connected EC(a; b), partial overlap PO(a; b), equal EQ(a; b), tangential proper-part TPP(a; b), non-tangential proper-part NTPP(a; b), tangential proper-part inverse TPP^{-1}(a; b), and non-tangential proper-part inverse NTPP^{-1}(a; b). These types of relations are able to describe, in a simple way, overlap and touching relationships between two different spatial entities in a space.

Directional Relations

Linguistic descriptions of spatial relations between entities usually contain the direction or the orientation of the objects involved in the scene. For example, very often we hear expressions like *to the right of* or *in the south of* which are two different types of qualitative descriptions. In particular, the first is an absolute qualitative description ({S(south), W(west), N(north), E(east), NE(northeast), SE(southeast), NW(northwest), SW(southwest)}), while, the second, is a relative one ({Left, Right, Front, Behind, Above, Below}).

Distal Relations

Distance is a very useful scalar entity used in spatial descriptions; it is a relevant aspect of space, in fact, it can qualitatively and quantitatively represent a relationship between objects. In particular, an example of qualitative expression is *near to*, while, a quantitative one is *a metre above*. We commonly classify the distance in *absolute* and *relative* with the difference that the former describes the distance between two entities, while, the latter defines the distance between two objects with reference to a third entity.

Qualitative Calculi Models: a way to decrease the Gap between Natural Language and Formal Models

As already mentioned, natural language is an ambiguous and polysemous instrument used by humans to express spatial concepts which involve many special aspects at the same time. Conversely, formal spatial models handle only one spatial aspect at a time and specify its spatial logic in detail [54]. This aspect underlines the large difference of expressivity between the two approaches [55], difference that is an interesting issue to be analysed. To reduce this gap we mapped the linguistic expressions to multiple calculi by using the above mentioned annotation scheme which define for every spatial relation and configuration one or more *general types* (gtype), its *specific type* (stype), its *spatial value* and its *frame of reference*. In particular, the gtype can be {REGION, DIRECTION, DISTANCE} which, in turn, could assume different values. If the gtype is REGION the stype is one of the topological relations expressed in the RCC8 formalism [53]; when the gtype is DIRECTION, the stype can be {ABSOLUTE, RELATIVE}, while, if the gtype is DISTANCE the

correspondent stype can be categorised as {QUALITATIVE, QUANTITATIVE}. The following figure (fig.19) shows an illustration of the described scheme:
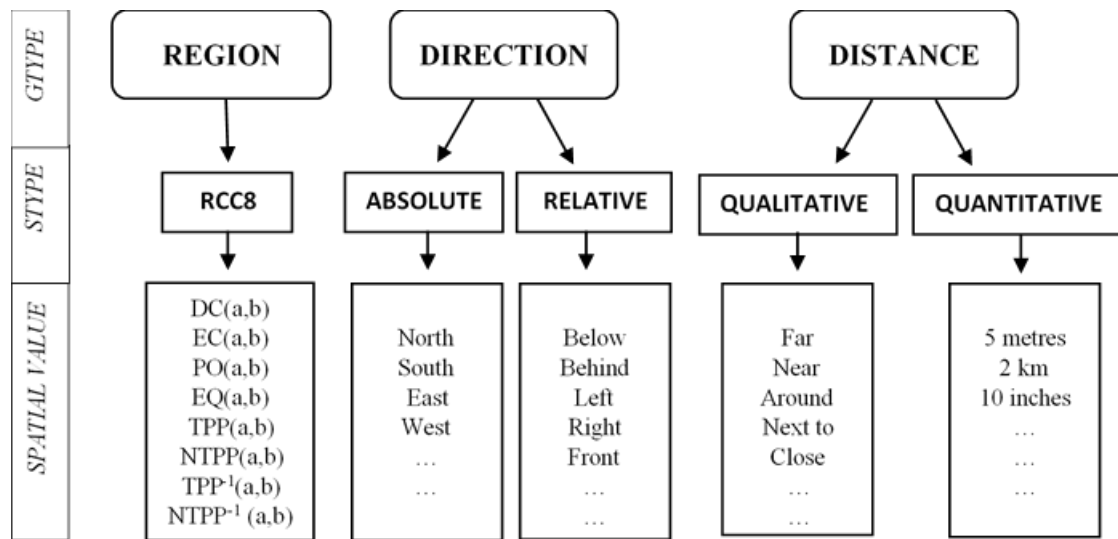
Figure 19: Representation of the Annotation Scheme.

3.6.3 Spatial Descriptions

In this section, we analyse two different types of descriptions: simple and complex. We start from simple spatial descriptions, in order to simplify the comprehension of all the elements of the schematisation, then, we pass to complex descriptions by analysing more articulated sentences, in which more trajectors, landmarks and spatial indicators appear.

3.6.3.1 Simple Spatial Descriptions

A simple spatial description is a sentence that includes only a trajector, a landmark and a spatial indicator. They are three elements that compose a sentence with a structure similar to a temporal expression, with the difference that they describe the locations of entities in the space. As noted many times, a machine does not have the ability to understand the difference between spatial and temporal descriptions, so, in order not to confuse spatial and non-spatial information, the components of the sentence have to be tagged according to the above described annotation scheme. In this way, we can "teach" the system that in the sentences *There is a party on Saturday* and *There is a chair on the floor* the expression "on Saturday" identifies a temporal condition, while, the expression "on the floor" identifies a spatial one.

With that assumption, in the following, we show several examples of simple spatial descriptions, specifically, we firstly illustrate how to annotate a *trajector*, a *landmark* and a *spatial indicator* and, finally, we propose some sentences which include all three labels:

Trajector

> **He waits on the stairs.**
> <TRAJECTOR id='TR1'> He </TRAJECTOR>
> **The alarm clock is on the bedside table.**
> <TRAJECTOR id='TR1'> The alarm clock </TRAJECTOR>
> **She is singing in her room.**
> <TRAJECTOR id='TR1'> She </TRAJECTOR>
> **Stop!**
> <TRAJECTOR id='0'> NIL </TRAJECTOR>

Trajector annotation does not change according to the verb tense, as shown in examples *a.*, *b.* and *c.*, while when the trajector is implicit "NIL" is added (ex. *d.*).

Landmark

Conversely to a trajector, a landmark is not always needed in a sentence description, or even, it is implicitly expressed. In these cases, we annotate it by "NIL".

> **The car is in front of the traffic light .**
> <LANDMARK id='LM1' path='ZERO'>the traffic light</LANDMARK>
> **The kite flew over the river.**
> <LANDMARK id='LM1' path='ZERO'>the river</LANDMARK>
> **The kite flew over.**
> <LANDMARK id='LM1' path='ZERO'>NIL</LANDMARK>
> **The students went out of the school.**
> <LANDMARK id='LM1' path='BEGINNING'>the school</LANDMARK>
> **You will find my house on the left!**
> <LANDMARK id='LM1' path='ZERO'>NIL</LANDMARK>

Spatial Indicator

Spatial indicators are usually represented by prepositions but can also be verbs, adverbs, nouns or multi-words. In the following, we analyse some examples:

Carl is waiting outside the church.

<SPATIAL-INDICATOR id='SP1' > outside </SPATIAL-INDICATOR>

 The dog hid the bone under the ground.

 <SPATIAL-INDICATOR id='SP1' > under </SPATIAL-INDICATOR>

My house is in front of the park.

<SPATIAL-INDICATOR id='SP1' > in front of </SPATIAL-INDICATOR>

There is a wonderful fountain in the middle of the square.

<SPATIAL-INDICATOR id='SP1' > in the middle of </SPATIAL-INDICATOR>

The land is surrounded by a high mountain range.

<SPATIAL-INDICATOR id='SP1' > surrounded by </SPATIAL-INDICATOR>

A circular dome covers the temple.

<SPATIAL-INDICATOR id='SP1' > covers </SPATIAL-INDICATOR>

Two columns support the large terrace.

<SPATIAL-INDICATOR id='SP1' > support </SPATIAL-INDICATOR>

Motion Indicator

Finally, several spatial scenes are not static representations but dynamic and they are generally described by verbs which express motion.

A man is crossing the street.

<MOTION-INDICATOR id='M1' > is crossing </MOTION-INDICATOR>

Wait, I am walking towards you!

<MOTION-INDICATOR id='M1'> walking towards </MOTION-INDICATOR>

The funicular leads up the mountain.

<MOTION-INDICATOR id='M1' > leads up </MOTION-INDICATOR>

Spatial Relations

Each previously described spatial component, is linked to the others in order to generate *spatial relations* (SR). They are characterised by an *id.number* and, in the case of simple descriptions, they are composed by a trajector, a spatial indicator and one/zero landmark. In addition, they contain the semantic attributes (gtype, stype, spatial-value, FoR) which represent the greatest difficulty of annotation. Below, we show some examples of spatial relations concerning simple sentences:

EX. 1: The car is at the car wash.

<TRAJECTOR id='TR1' > The car </TRAJECTOR>

<LANDMARK id='LM1' path='ZERO'> the car wash </LANDMARK>

<SPATIAL-INDICATOR id='SP1' > at </SPATIAL-INDICATOR>

<SR id='S1' trajector='TR1' landmark='LM1' spatial-indicator='SP1' motion-indicator='NIL' gtype='REGION' specific-type='RCC8' spatialvalue='TPP' FoR='INTRINSIC' />

EX. 2: Every Saturday we go to the church.

<TRAJECTOR id='TR1' > we </TRAJECTOR>

<LANDMARK id='LM1' path='END'> the church </LANDMARK>

<SPATIAL-INDICATOR id='SP1' > to </SPATIAL-INDICATOR>

<MOTION-INDICATOR id='M1' > go to </MOTION-INDICATOR>

<SR id='S1' trajector='TR1' landmark='LM1' spatial-indicator='SP1' motion-indicator='M1' gtype='REGION' specific-type='RCC8' spatialvalue='TPP' FoR='INTRINSIC' />

EX. 3: The cat is on the armchair.

<TRAJECTOR id='TR1' > The cat </TRAJECTOR>

<LANDMARK id='LM1' path='ZERO'> the armchair </LANDMARK>

<SPATIAL-INDICATOR id='SP1' > on </SPATIAL-INDICATOR>

<SR id='S1' trajector='TR1' landmark='LM1' spatial-indicator='SP1' motion-indicator='NIL' gtype='REGION' specific-type='RCC8' spatialvalue='EC' FoR='INTRINSIC' />

EX. 4: Take the road on the left!

```
<TRAJECTOR id='0' > NIL </TRAJECTOR>
<LANDMARK id='0' > NIL </LANDMARK>
<SPATIAL-INDICATOR id='SP1' > on the left </SPATIAL-INDICATOR>
<SR   id='S1'   trajector='0'   landmark='0'   spatial-indicator='SP1'   motion-
indicator='NIL'    gtype='DIRECTION'  specific-type='RCC8'  spatialvalue='left'
FoR='INTRINSIC' />
```

3.6.3.2 Complex Spatial Descriptions

The annotation scheme is also able to handle complex spatial descriptions which differ from the simple ones for several reasons:

They have more than one landmark;

EX. 1:   The picture is in the dining room, on the wall near the table.

```
<TRAJECTOR id='TR1'> The picture </TRAJECTOR>
<LANDMARK id='LM1' path='ZERO'> the dining room </LANDMARK>
<LANDMARK id='LM2' path='ZERO'> the wall </LANDMARK>
<LANDMARK id='LM3' path='ZERO'>the table </LANDMARK>
<SPATIAL-INDICATOR id='SP1' > in </SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id='SP2' > on </SPATIAL-INDICATOR >
<SPATIAL-INDICATOR id='SP3' > near  </SPATIAL-INDICATOR>
<SR  id='S1'  trajector='TR1'  landmark='LM1'  spatial-indicator='SP1'  motion-
  indicator='NIL'  general-type='REGION'  specific-type='RCC8'  spatialvalue='
  NTPP' FoR='INTRINSIC' />
<SR  id='S2'  trajector='TR1'  landmark='LM2'  spatial-indicator='SP2'  motion-
  indicator='NIL'  general-type='REGION'  specific-type='RCC8'  spatialvalue='
  EC' FoR='INTRINSIC' />
<SR  id='S3'  trajector='TR1'  landmark='LM3'  spatial-indicator='SP3'  motion-
  indicator='NIL'   general-type='DIRECTION'  specific-type='RELATIVE'
  spatial-value='CLOSE' FoR='INTRINSIC' />
```

They describe **sequential scenes** in which, for example, an entity can be, at first, a trajector of the sentence and, secondly, a landmark.

EX. 2: In the drawer there is a grey box, behind the box there is a wallet, inside the wallet there is my passport.

```
 <TRAJECTOR id= 'TR1'> a grey box </TRAJECTOR>
<LANDMARK id= 'LM1' path='ZERO'> the drawer </LANDMARK>
<SPATIAL-INDICATOR id= 'SP1' > in </SPATIAL-INDICATOR >
<TRAJECTOR id= 'TR2'> a wallet </TRAJECTOR>
<LANDMARK id= 'LM2' path='ZERO'> the box </LANDMARK>
<SPATIAL-INDICATOR id= 'SP2' > behind </SPATIAL-INDICATOR >
<TRAJECTOR id= 'TR3'> my passport </TRAJECTOR>
<LANDMARK id=  'LM3' path='ZERO'> the wallet </LANDMARK>
<SPATIAL-INDICATOR id= 'SP3' > inside </SPATIAL-INDICATOR >
<SR  id='S1'  trajector='TR1'  landmark='LM1'  spatial-indicator='SP1'  motion-
  indicator='NIL'  general-type='REGION' specific-type='RCC8' spatialvalue='
  TPP' FoR='INTRINSIC' />
<SR  id='S2'  trajector='TR2'  landmark= 'LM2'  spatial-indicator='SP2'  motion-
  indicator=  'NIL'      general-type='DIRECTION'    specific-type='RELATIVE'
  spatialvalue= 'BEHIND'  FoR= 'INTRINSIC' />
<SR  id='S3'  trajector='TR3'  landmark='LM3'  spatial-indicator='SP3'  motion-
  indicator='NIL'   general-type='REGION' specific-type='RCC8'
  spatial-value= 'TPP' FoR='INTRINSIC' />
```

EX.3: The ball is under the table on the right.

<TRAJECTOR id='TR1'>The ball </TRAJECTOR >

<TRAJECTOR id='2' >the table </TRAJECTOR >

<LANDMARK id='LM1' path='ZERO'> the table </LANDMARK >

<SPATIAL-INDICATOR id='SP1' > under </SPATIAL-INDICATOR >

<SPATIAL-INDICATOR id='SP2' > on the right </SPATIAL-INDICATOR >

<SR id='S1' trajector='TR1' landmark='LM1' spatial-indicator='SP1' motion-indicator='NIL' general-type='REGION' specific-type='RCC8' spatialvalue='
TPP' FoR='INTRINSIC' motion-indicator='NIL'/>

<SR id='S2' trajector='TR1' landmark='NIL' spatial-indicator='SP2' motion-indicator='NIL' gtype= 'DIRECTION' specific-type='Relative' spatialvalue= 'RIGHT' FoR= 'RELATIVE' />

<SR id='S3' trajector='TR2' landmark='NIL' spatial-indicator='SP2' motion-indicator='NIL' gtype= 'DIRECTION' specific-type='Relative' spatialvalue= 'RIGHT' FoR= 'RELATIVE' />

4. Text Annotation

4.1 Annotation task

Our aim is to allow machines to interpret and use data provided by the users in natural language starting from the awareness that the main problem is that human text and speech are understandable only by people and not by computers. Indeed, despite the fact that in the last decades computers have become part of our life as "intelligent machines", able to memorize and analyse big amount of information, their "communicative abilities" are still rudimentary. This means that they are not capable of understanding human language because they cannot interpret the meaning of our words without executable computational models (programs). They do not comprehend the language but they only read codes. Computers can store and process data exclusively in the form of two symbols 0 and 1 (bits) that are aggregated into 8-digit sequences (bytes), therefore, all types of text, as well as images, sounds or numbers, must necessarily have a binary representation. However, it should be noted that the encoding of a text cannot be confused with a simple succession of bits (zero level coding), because it provides information related to the linguistic-textual structure and allows the interpretation of the text (encoding of high level). In fact, in the absence of a correct structure of the text (subdivision into chapters, sections, paragraphs ...), of linguistic structures (morphological, syntactic and semantic information) and of relations between words, many information would inevitably be lost. This aspect highlights the necessity to use a methodology able to work as a bridge between people and machines: the semantic annotation technique. The task of annotation is to assign a specific meaning to words in a natural text in such a way as to avoid any possibility of misunderstanding due to the ambiguity and complexity of our language. In this way, machines can interpret the input data correctly, execute them and answer with pertinent information. Annotation involves attaching data to some other piece of data in order to link the element with the right domain and to establish a relation between the annotated and annotating data [32]. There are three different types of annotations: *informal*, *formal* and *ontological*. The difference between them is that the first is not machine-understandable because it does not use a formal language, the second is machine-understandable but it does not use ontological terms and,

the last, is both machine-understandable and linked to an ontology. The latter, in artificial intelligence and computer science, is "a specification of a conceptualization" [57], namely, a set of representational primitives able to model a domain of knowledge. This means that the ontological annotation is the most complete and shared conceptualisation and it ensures a common, understood meaning. Nowadays several tools exist to create annotations of data, both manually, semi-automatically and fully automatically. Specifically, manual annotation tools allow users to add annotations to documents while automatic tools do not need manual intervention.

4.2   MAE (Multi-document Annotation Environment)

In this work we used MAE [58][77][16], a general-purpose natural language annotation tool, that allows users to create their annotation tasks, to use non-consuming tags and to define connections between annotated elements. It is a graphical annotation tool written in Java which permits good flexibility in defining the annotations but, at the same time, users have to define and formalize their task specification in a Document Type Definition (DTD) file. It reads .txt and .dtd files, containing respectively the text to be annotated and some rules for the annotation scheme and, in the end, it provides output annotations in XML format.

4.2.1 Document Type Definition (DTD) file

A DTD is a set of declarations that define a *markup language,* in particular, it outlines a set of rules that control the structure of any XML file. A DTD contains all the declarations that specify name, elements, attributes and entities for the output XML file and it also establishes constraints for how they can be used.

Task Name

The task name is defined with !ENTITY tag followed by the word *name* and a personal name of the task enclosed into double quotes:

<p style="text-align:center"><!ENTITY name "personalTask"></p>

---

[16] < https://github.com/keighrim/mae-annotation/wiki >

Task Elements

MAE handles two different types of tag elements that are:

- **Extent tag** = useful to label spans of text belonging to a document;
- **Link tags** = used to detect a relation between extent tags.

The extent tag and the link tag are defined, respectively, in this way:

<div align="center">

<!ELEMENT ExtentTagName (#PCDATA) >
<!ELEMENT LinkTagName EMPTY >

</div>

In the first case, the *#PCDATA* string appears, between brackets, which indicates that the *ExtentTagName* will be associated to a span of text while, in the second case, the word *EMPTY* appears, that means that there will not be any associations with a span of text.
All the different extent tags are characterised by their own colour, which the system automatically generates; conversely, no colours are attributed to the link tags.

Attributes

Attributes are additional information given to the element types. They are defined by !ATTLIST tag and they include all the information related to the tags. MAE provides some pre-defined attributes which are:

- id;
- start;
- end;
- text.

for Extent Tags, and

- id;
- to;
- from;
- toText;
- fromText.

for Link Tags.

Below are examples of two different ways in which it is possible to define an attribute:

<!ATTLIST Extent/LinkTagName   attributeName1 (YES | NO) #IMPLIED >

<!ATTLIST Extent/LinkTagName   attributeName2 CDATA #IMPLIED >

In the first case the name of the element (ExtentTag or LinkTag) is associated to an attribute name that can have pre-fixed values (YES or NO) while, in the second example, the attribute can be a simple text (CDATA) provided by the user.

-id Attribute-

By default, MAE assigns to each tag an identifier that is characterised by the first letter of the tag name plus a progressive number (for ex. t1, t2, t3…). However, it is possible to define a single id attribute in this way:

<!ATTLIST Type id   ID prefix="VB" #REQUIRED >

By doing so we will obtain VB1, VB2 and so on.

-start and end Attributes-

Start and end Attributes characterise the initial and final characters of an element in the text. However, it is also possible that an extent tag is "non-consuming" because it is hidden or omitted in the sentence, therefore, it is necessary to define an #IMPLIED attribute in such a way:

<!ATTLIST Tag1 start/end #IMPLIED >

-Attribute Types-

MAE provides 4 different types of attributes that are:

- **CDATA** = it is a free text, a stand of characters;
- **ID** = a unique identifier of the attribute;
- **closed value set** = a set of options for attributes;
- **IDREF** = an attribute type that refers to another tag.

The last attribute is used to connect two or more elements present in the document, in fact, it points to an element which has an ID attribute previously declared in the text. To better understand, let us show an easy example:

Declaration of IDREF Attribute in a DTD:

<!ELEMENT artist EMPTY>

   <!ATTLIST artist name CDATA #REQUIRED>

   <!ATTLIST artist artistID ID #REQUIRED>

<!ELEMENT album EMPTY>

   <!ATTLIST album name CDATA #REQUIRED>

   <!ATTLIST album album ArtistID IDREF #IMPLIED>

A sample of XML:

<artist name="Nick Cave" artistID="NC"/>

<album name="Murder Ballads" **albumArtistID="NC**"/>

4.3 DTD scheme

The DTD scheme used to annotate the sentences, previously extracted from the corpus (§ 2.3 Sentence Extraction) referred to the "World City" domain has been structured in the following way:

- A Task Name called SCENE has been defined which contains five tag elements that are SENTENCE, TRAJECTOR, LANDMARK, SPATIALINDICATOR and RELATION.

- The default attributes: id, start, end and text have been assigned to the first four Extent Tags;

- Since the last element is a Link Tag, the four pre-defined attributes have been assigned to it, along with specific attributes concerning the above-described annotation scheme (§ 3.6   Spatial Annotation Scheme): *spatial_indicator_id, spatial_indicator_text, general_type, specific_type, spatial_value* and *FoR*.

The following table (tab. 5) illustrates the framework of the DTD file:

```
<!ENTITY name "scene">
<!ELEMENT SENTENCE ( #PCDATA ) >
<!ELEMENT TRAJECTOR ( #PCDATA ) >
<!ELEMENT LANDMARK ( #PCDATA ) >
<!ELEMENT SPATIALINDICATOR ( #PCDATA ) >

<!ELEMENT RELATION EMPTY >
<!ATTLIST RELATION spatial_indicator_id " ">
<!ATTLIST RELATION spatial_indicator_text ( above
                                            | along
                                            | among
                                            | around
                                            | at
                                            | at each side
                                            | at the corner of
                                            | at the end
                                            | at the extremity of
                                            | at the side of
                                            | at the sides
                                            | at the summit of
                                            | at the top of
                                            | behind
                                            | between
                                            | beyond
                                            | in
                                            | in front of
                                            | inside
                                            | in the centre
                                            | in the centre of
                                            | near
                                            | next to
```

```
                                          | on
                                          | on one side of
                                          | on the left
                                          | on the left of
                                          | on the right
                                          | on the right of
                                          | on the top of
                                          | outside
                                          | over
                                          | perpendicular
                                          | supported by
                                          | surrounded by
                                          | to right and left
                                          | to right and left of
                                          | to the left of
                                          | to the right of
                                          | toward
                                          | traversed by
                                          | under
                                          | upon
                                          | other )
        #IMPLIED "other" >
          <!ATTLIST RELATION general_type ( region
                                          | direction
                                          | distance) >


          <!ATTLIST RELATION specific_type (RCC8
                                          | ABSOLUTE
                                          | RELATIVE
                                          | QUALITATIVE
                                          | QUANTITATIVE) >


          <!ATTLIST RELATION spatial_value (DC
                                          | PO
                                          | TPP
                                          | TPPi
                                          | EC
                                          | EQ
```

```
                                        | NTPP
                                        | NTPPi
                                        | close
                                        | front
                                        | right
                                        | right and left
                                        | left
                                        | below
                                        | above
                                        | middle
                                        | behind) >
            <!ATTLIST RELATION FoR (intrinsic
                                | relative
                                | absolute )>
```

Table 5: DTD file representation.

4.4 From the annotation to the automatic spatial role labels and relationships detection

The annotation phase lays the basis to the next step in which a recent declarative learning-based programming language (DeLBP), called Saul, has been used. This system allows us to automatically detect the spatial role labels and the relationships between them given an input text.

The annotation scheme, described in this section, has been followed to annotate several sentences of our sentences-corpus (the training set), which is composed of the spatial descriptive sentences extrapolated from the corpus, as described in Chapter 4. The latter have been divided into two different sets which are the training dataset and the test dataset in such a way that the first group is formed by the largest number of sentences (80% of the entire sentences-corpus) while, the second one, is composed by the remaining sentences (20% of the entire sentences-corpus). The training dataset is a set of examples used by the system for learning: to fit the parameter values of the classifier. Conversely, the test dataset is a set of examples used to assess the performance of a fully-trained classifier, hence, to know how the system responds to our final task.

5.   Spatial Relationships Detection

5.1 Introduction

One of the main challenges of natural language processing is to map from an input structure to an output structure which represents lexical, syntactical and semantic aspects of the text. For example, Part-of-speech (POS) tagging is able to provide a syntactic representation of textual data while Semantic Role Labeling (SRL) and Entity-Relation (ER) extraction are semantic lightweight representations. The difficulty arises from the fact that the only way to solve these problems is to consider various kinds of relationships between linguistic elements at multiple levels of granularity.

In this chapter, we talk about semantics extrapolation carried out by Saul [59], a probabilistic object-functional programming language written in Scala [60], which aims to:

Allow a programmer to learn and manipulate named abstraction over relational data;
Upload and read trainable probabilistic or discriminative components into the program;
Provide inferences over the trainable components in order to support decision making with respect to the domain and the application constrains.

5.2 Saul structure

Saul programming language is an intelligent problem-solving system which interacts with a large amount of information; it was developed to interpret natural language statements and to support the required level of understanding [36]. This means that the main challenging task is to resolve context-sensitive ambiguities and to support knowledge-based inferences in order to interact with real world data.

Saul's models are general factor graphs [61] which, extended to relational representations [62], are able to express constrained factors [63].

It allows the programmer to define his data model through which he can establish a set of variables of interest, that will be assigned to the given data, and a set of models to be learned,

with the purpose of identifying *named entities* of types (person, location, etc.) and relations of types (works-for, lives-in, etc.) in a given free form text.

A part of Saul code is dedicated to *training* models and to teach it how to make inferences (decisions). In other words, the system is trained to "comprehend" how to recognise, in an input text, the entities involved and how to identify the relations between them.

5.2.1 Data Representation

To facilitate the communication between the programmer and the heterogeneous real world data, Saul provides a graphical representation. In particular, it uses two different types of graphs which are the data graph and the model graph [64].

Data Graph

The data graph encodes the information given to the system in the form of annotated instances. The representation is formed by nodes and edges and, in particular, the nodes are the entities, the relationships and the properties while the edges are the link between them. The following figure (fig.20) shows an example of data graph referred to the sentence "W. Dale Nelson covers the White House for the Associated Press". In the graph, the entities are represented by rectangles, the relationships by diamonds and the properties by ovals.
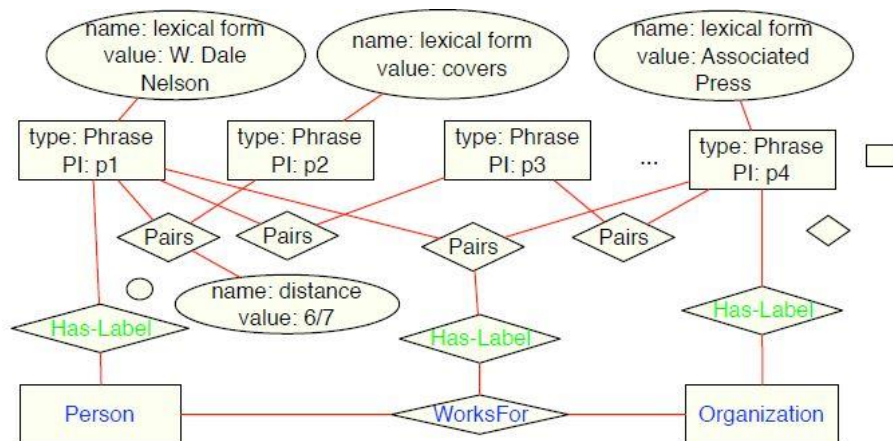
Figure 20: Representation of the data graph referred to the sentence "W.Dale Nelson covers the White House for the Associated Press".

Model Graph

The second type of graph is the model graph (fig. 21), which shows the relational schema of the data. It is structured in such a way that the types of entities and relations are represented by nodes while the connection between them are symbolised by edges. The latter are characterised by a degree that specify a relation property such us *1:1*, *1:n* or *n:1*. Figure 21 illustrates an example of model graph in which the entities and the relationships are indicated with a double line in order to make a distinction with the abovementioned data graph.
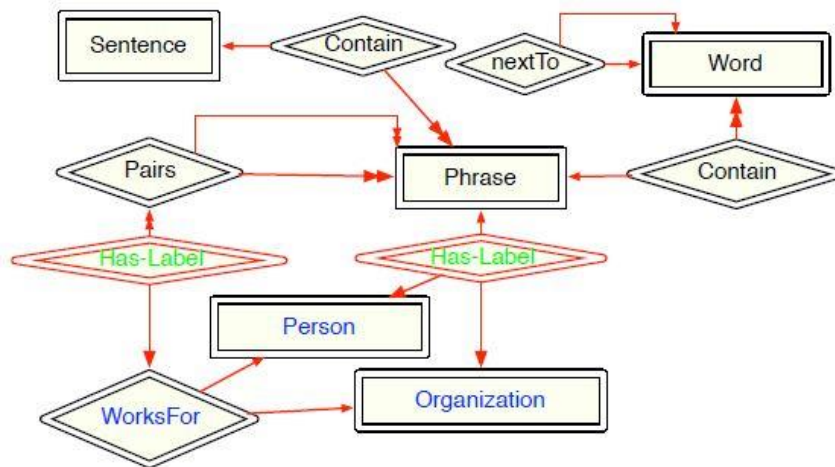
Figure 21: Representation of a part of model graph.

5.3 Knowledge Acquisition

A necessary step to allow machines to acquire knowledge about a specific domain is the link between natural language and spatial ontology, which has the purpose of assigning to each element of the sentence a concept contained in the reference ontology. The importance of the link is due to the complexity, ambiguity and polysemy of our language, in fact, spatial language contains numerous terms which express complex spatial semantic aspects, while, formal spatial models managed by machines, have a lower level of expressivity. This means, that not only can they handle one aspect at a time, but also that they need many spatial logic details.

Currently an ontology already exists, called GUM-Space [65], that specifies detailed semantics for linguistic spatial expressions. It covers languages linked to space, spatial relationship and actions in space, thus, it offers a high simplification of the problem of connecting natural spatial language to its machine interpretation. Its purpose is to be an exhaustive description of categories in spatial language, indeed, it was used by a huge number of users interested in analysing and annotating data. GUM-Space contains 73 spatial modalities that are well distinguished in the ontology thanks to their hierarchical dependencies.

We utilise the aforementioned ontology to partially resolve the gap between natural spatial language and machine spatial language. In particular, the problem has been faced through two different abstract representations:

- A **linguistic conceptual representation,** called Spatial Role Labelling (SpRL), that aims to identify, in the sentence, all the words that play a spatial role or a spatial relationship [66].

- A **formal semantic representation**, called Spatial Qualitative Labelling (SpQL), that aims to assign spatial relations through semantic attribute values based on qualitative spatial representation models (QSR) [67] [68].

5.3.1   Mapping sentences

Given an unknown amount of sentences X composed by x_i number of words ($x_i = x_1 + x_2 + ... + x_n$) the task has been to map each sentence to the spatial ontology. Hence, at the beginning, some words of the sentences have been labelled with spatial roles (SpRL) and the spatial relations have been identified, after, the spatial relations have also been labelled with their own spatial semantics.

When words have multiple meanings, automatically, they have also multiple roles and multiple semantic assignments.

5.3.2 Spatial Role Labelling (SpRL)

The goal of SpRL is to find the words that play a spatial role in the sentence in order to assign them the right roles and to link them and to extract special relations. Let us show an example (fig.22, 23):

Sentence: *There is a church with a high bell tower on a hill.*
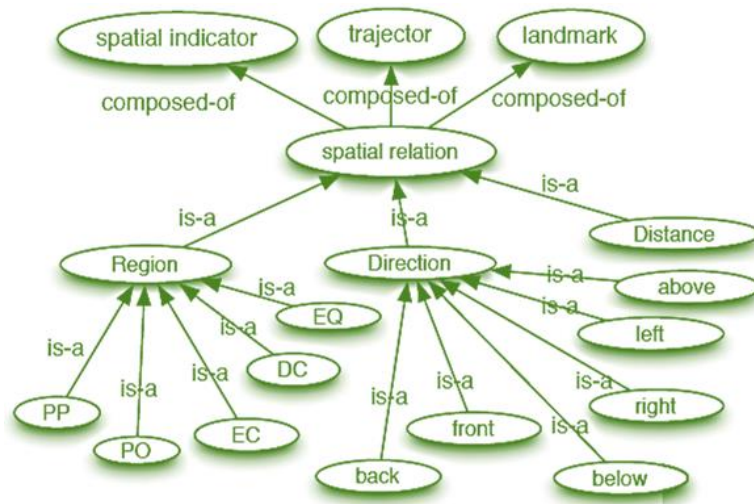
Figure 22: Spatial ontology.
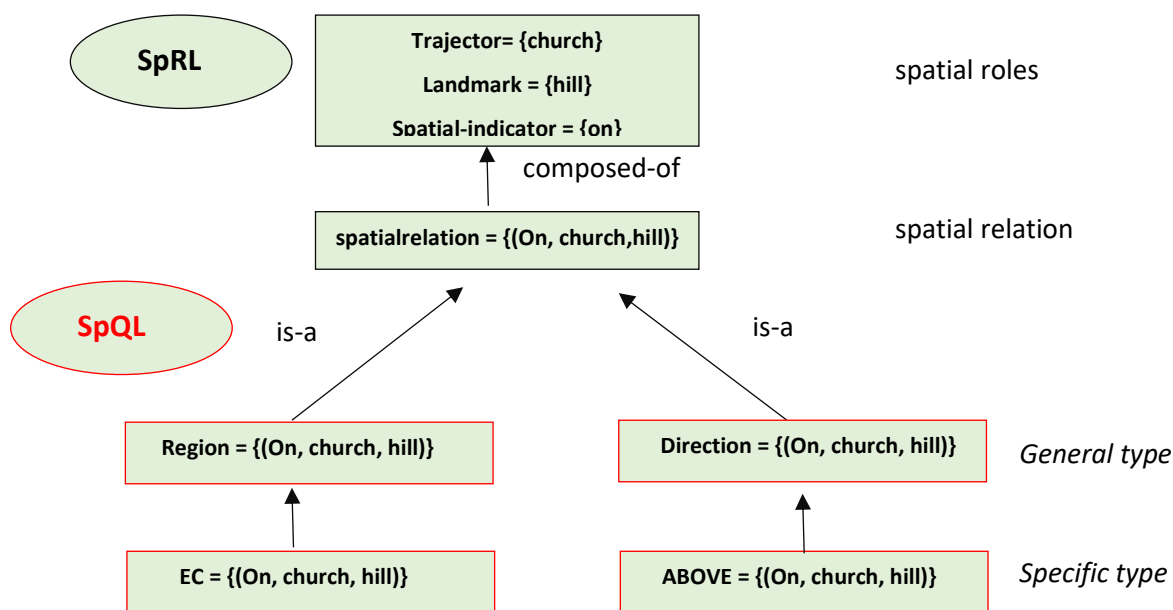
Figure 23: Example of SemEval-1 benchmark.

In the sentence the three spatial entities that we need to extract are: *on, church* and *hill*. In particular, the word *church* plays the role of trajector **(tr)**, the word *hill* plays the role of landmark **(lm)** and the preposition *on*, that holds a position between the two words, is a spatial indicator. The three special roles are the three main nodes of the ontology, called *single labels*, while their connection constitutes a spatial relation, or *spatial triplet*. In addition, the three nodes, connected one to each other, take the name of *linked labels*.

Spatial relations, that can be one or more than one in the sentence, are generally provided by the use of prepositions that establish a relationship between two or more objects. However, many times, some prepositions are not special indicators because they could also be used to talk about events or situations which do not involve spatial descriptions. For example, the preposition *on* could be used to describe spatial configurations like "the picture is *on* the wall" or "the bottle is *on* the table" but, the same preposition is also used in different contexts such as in the sentences "They need to concentrate *on* their studies" or "The discussion will be *on* a topic you have studied recently". The first two sentences express spatial concepts, concepts of verticality or objects overlapping, while the last two sentences do not have any kind of reference to entities locations. This means that the usage of prepositions depends on several aspects like the entities involved in the scene or the general situation of speech. In

other words, prepositions are often used in front of nouns or pronouns to show the relationship between them and other words in the sentence, but they can also be used to describe the time when something happens (*They arrived* **on** *Sunday*), the way in which something is done (*We went* **by** *train*) and even more. On this basis, the identification and the extraction of the spatial indicators from a texts, characterised by heterogeneous information, is a big and interesting challenge.

Another interesting and difficult aspect to face, related to the identification of labels, is that sometimes trajector and/or landmarks are not expressly indicated in the sentence but they are implicit. In natural language, in fact, it can happen that we do not need to indicate all the spatial entities [69] because the situation can in any case be understood from the context. The result is that there are no words representing the spatial label(s), so, the machine cannot detect them. An example of this situation is exposed in the sentence "*Turn on the left!*", where the trajector *you* is implicitly expressed in the exclamation. In this case, it can be tagged with the term *undefined* and, by so doing, the spatial triplet becomes $<on_{sp}undefined_{tr}left_{lm}>$.

The same problem can also be found with reference to the spatial relations which can be inferred by spatial reasoning. Sometimes, in fact, they are not specified in the sentence but they can only be understood by humans thanks to their mental ability. For example, if we say "*The bookshelf is in front of the wall behind the desk*" there are, not only, two immediately recognizable relations, $<in\ front\ of_{sp}bookshelf_{tr}wall_{lm}>$ and $<behind_{sp}wall_{tr}desk_{lm}>$, but there is also another one: $<behind_{sp}bookshelf_{tr}desk_{lm}>$.

In view of the difficulty of this type of relations to be managed, we decided not to consider them in the tasks of this thesis.

5.3.3 Spatial Qualitative Labelling (SpQL)

Once the spatial configuration from the text has been extracted, the next step forward is to automatically map the spatial roles to qualitative spatial representations, or spatial calculi, which generally focus on only one aspect like the distance, the shape or the direction [51]. On the contrary, as already mentioned, spatial language could refer to different meanings with a single expression [70]. This means that SpQL has to be able to cover multiple aspects by maintaining a good level of generality.

In the previous chapters we have talked about the sentences-corpus compilation (§ 2. Corpus Creation and Analysis) and the annotation scheme (§ 4. Text Annotation) that is inspired by SpatialML [71] and another framework that was presented in [72]. On these bases, we have annotated the sentence-corpus in order to support Saul in the detection of the spatial roles and the relations. In other words, the mapping lets the machine visualize the spatial semantic content of the sentence and, in some cases, to draw out the inferences. To better understand we show an example below:

"The lamp is to the right of the radio on the bedside table".

The system can comprehend that the two objects, the lamp and the radio, are both put on the bedside table, namely, that they have, not only, a connection between them but also a spatial relation with the bedside table.

5.4 Constraints and Features

Constraints and features are two important pieces of information for machine learning models [73] because they are able to capture, respectively, the global and local characteristics. The two elements, in fact, in addition to numerous training sentences, lead machine to understand the properties of the problem. In this section, we discuss them separately.

5.4.1 Constraints

The automatic extraction of spatial roles and relations can be carried out by exploiting spatial language characteristics. Let us show the most important ones belonging to the SpRL and the SpQL representations [49]:

SpRL constraints

I. Considered that a spatial relation is formed, at least, by a trajector and a landmark linked by a spatial indicator:
    - If the sentence contains a trajector and a landmark it consequently also contains a spatial indicator;
    - If the machine detects a spatial indicator in the sentence it necessarily has to extract a spatial triplet.

II.   If a word representing a trajector or a landmark changes its role in the same sentence, it could be linked to different spatial indicators. For instance, in the sentence "*The picture is hung on the wall in front of the television*" the word "wall" is, at first, a landmark connected to the spatial indicator "on", secondly, a trajector linked to the spatial indicator "in front of".

III.   A word can assume only a role with respect to a spatial indicator: it can be a trajector or a landmark.

SpQL constraints

I.   *is-a* constraint: taken into account two nodes (or labels), l_1 and l_2 , belonging to the ontology, if a relationship such as l_1 *is-a* l_2 exists, then all the instances of l_1 should be also instances of l_2 ;

II.   *assignment* constraint*:* a general triplet can be assigned to zero, one or more semantic labels, while a spatial triplet should be assigned at least to one semantic label;

III.   *input* constraint*:* to reduce the amount of output we can exploit the background knowledge. For instance:

- in most cases the trajector of a sentence is a singular (NNS) or plural noun (NN) or, with respect to the dependency tree, is the subject (SBJ) of it. Equally, a landmark is generally a singular (NNS), plural (NN) or proper noun (PRN). This imply that we can start to search our spatial labels in the feature extracted by the syntactic and dependency parser. However, we must remember that trajectors and/or landmarks could also be implicit, therefore, undefined;

- in the English language, generally, prepositions play the role of spatial indicators, so, we can benefit from POS tags by analysing all the words tagged as IN and TO by parsers. Furthermore, we can also collect them in order to create a lexicon of prepositions linked to the corpus.

5.4.2 Features

Features are local characteristics, referred to spatial language, that can be used to train the learning models in order to allow machine learning models to better recognize and extract spatial labels from the input sentences.

The features describe one or more words (multi-words) present in a sentence; the most important local features referred to a single word are:

I. **Word-form $\Phi_{wf}(x_i)$:** lexical information can be extrapolated from a word which can be used as a feature. For example, in the two sentences *"A walk in the park"* and "*A walk in the morning",* the only difference is that the last word, in the first case, is a landmark while, in the second case, is temporal information;

II. **Part-of-speech tag $\Phi_{pos}(x_i)$:** As already said a POS tagger aims to read a text and to assign a part-of-speech tag to each word (noun, verb, adjective, etc.). This assignment could be exploited to detect the spatial role of a word or a group of words, in fact, for example, the IN tag, which represents a preposition, increases the possibilities of a word being a spatial indicator. Equally, the NN tag, which represents a noun, increases the chance of being a trajector or a landmark. To better understand, we can consider two sentences containing the same word (left):

<div align="center">

He <u>left</u> the train.

He turned on the <u>left</u>.

</div>

The POS tag helps the system understand that in the first sentence the word is a verb (VB), so it cannot have a spatial meaning while, in the second sentence, it is a noun and it can be a spatial label;

III. **Semantic role $\Phi_{srl}(x_i)$:** semantic role can be an aid element for spatial role labelling process because it can suggest the semantic label of a term; for example if a word is labelled as locative (LOC) by a semantic role labeller its chance to be a spatial label is very high;

IV. **Spatial context of the preposition $\Phi_{spc}(x_i)$:** despite the fact that natural language is composed of many different verbs and nouns, spatial terms such as *left/right/front*, are a limited number and they can easily be collected in a list. Starting from this list, if a spatial indicator has one of the spatial terms in its neighborhood we can use this aspect as a feature. This feature can help the machine recognize the spatial sense of the prepositions in a spatial context and to detect the undefined landmarks. To better understand, in a directional phrase like "o*n the right"* the landmark is generally undefined but, if it is followed by the preposition "*of*" the situation is the opposite because the landmark is explicit (*"On the right of [the lake]$_{LM}$"*). This means that,

when the preposition "of" follows a directional phrase, there is a high chance to have an explicit landmark while, when there is only a directional phrase, it is common that the landmark is implicit. To handle this aspect, we can create a feature with three dimensions to distinguish if the spatial indicator has a spatial context with a second preposition. (*on_theright_of*), if it has a spatial context without a second preposition (*on_the right*) or if it has no spatial context (*on*);

V. **Dependency relation $\Phi_{dprl}(x_i)$:** dependency parsers could be useful to suggest a correct spatial role assignation to a word. For instance, if a word is labelled as a subject (SBJ) it may be, with a high probability, a trajector. In the same way, if it is labelled as noun modifier (NMOD) certainly it has no spatial role.

5.5 Evaluation of the Results

In this thesis work, we used Saul to process the sentences previously extrapolated from the "World City" corpus (§ 2.3 Sentence Extraction) and we evaluated how it is able to detect labels (trajector(s), landmark(s) and spatial indicator(s)), and relations. For the sake of completeness, a relation is correct when the three components (each triplet) are properly extracted and connected to each other, thus, a prediction is true when all the single labels are according to the ground truth.

The outputs have been estimated according to the evaluation metrics of precision, recall and F1-measure and they have been calculated for each label separately. Recall, Precision and F1-measures are so defined:

$$Recall = \frac{TP}{TP+FN} \qquad Precision = \frac{TP}{TP+FP} \qquad F1 = \frac{2*recall*precision}{(recall+precision)}$$

where:

TP= the number of tags[17] produced by the system which match an annotated tag;

FN= the number of annotated tags that do not match a tag produced by the system;

FP= the number of tags produced by the system which do not match an annotated tag.

---

[17] an example of a tag, referring to the identification of a spatial indicator, provided by Saul is the following:
<SPATIALINDICATOR id = "SP_S0 (0, 1)" start = "2" end = "4" text = "In" />

When a preposition is wrongly classified as spatial, the number of FPs increases and the rate of precision decreases, while, when a preposition is wrongly classified as non-spatial, the number of FNs increases and the rate of recall decreases.

5.5.1 Experimental Results

We previously said that Saul can extract spatial semantics independently from the reference domain, however, it is evident that the results depend on the lexical features used to train the software. In fact, its performances decrease if we train in one domain and we later test in another domain. To understand the difference in results, we decided to carry out two different experiments:

1. In the first one, we used a model trained on a large but general domain corpus. It contains an amount of 9,409 example sentences belonging to two different resources. The largest one is **TPP dataset** [74], which aims to solve the problem of preposition sense disambiguation, in fact, it contains 8,096 training sentences in which the most common prepositions used in natural language to talk about spatial configuration appear. The second and the third datasets are, respectively, a subset of **IAPR TC-12 image Benchmark** (known as **CLEF**) [75], which stores 1,213 sentences, and a subset of **GUM dataset** [76], which contains only 100 sentences. They are composed of short texts, of about 15 words, which describe entities and their relative or absolute location in the space.

All these sentences are indexed and annotated in order to be the training set of our experiment, while, the train-data is a portion of our sentences-corpus (§ 2.4 Sentences-corpus). The results, shown below (tab. 6), are reasonably good and they have been separately estimated for the trajector, landmark, spatial indicator and relation.

LABEL	PRECISION	RECALL	F1
Trajector	69.017	64.028	66.429
Landmark	82.988	97.396	89.617
Spatial Indicator	81.818	66.832	73.569
Relation	54.465	51.638	54.396

Table 6: Results based on a general training domain corpus.

Nature of Errors

We decided to manually analyse part of the output annotated sentences in order to comprehend, in broad terms, the types of the errors made by the system. We classified the principal mistakes in 5 different classes which are:

- **C1 = <u>Transitivity Property Error:</u>** This error is encountered when the system is not able to handle the transitivity property, so, it does not comprehend when, in a sentence with more relations, an entity has more than one role and it changes it in the sentence itself. For instance, sometimes the landmark of the first relation begins the trajector of the following relation like in the sentence "Both Bank and Library are under a dome which cover the chief room in each building". In the latter the two different relations are:

  Correct relation R1: Both [Bank]$_{tr1}$ and [Library]$_{tr}$ are [under]$_{si}$ [a dome]$_{lm}$ which covers the chief room in each building.

  Correct relation R2: Both Bank and Library are under [a dome]$_{tr}$ which [covers]$_{si}$ [the chief room]$_{lm}$ in each building

  but the system generated the following output:

  System relation R1: Both [Bank]$_{tr1}$ and [Library]$_{tr1}$ are [under]$_{si}$ [a dome]$_{lm}$ which covers the chief room in each building.

  System relation R2: Both Bank and Library are under a dome which covers [the chief room]$_{tr}$ [in]$_{si}$ [each building]$_{lm}$.

  It is evident that the second relation is totally incorrect because the system assigned the trajector role to "the chief room" instead of "dome" and, consequently, also erred the spatial indicator and the landmark.

- **C2 = <u>Nested Relation Error:</u>** This error could occur when the sentence is more complex, in particular, when it contains a relation embedded into another relation. The consequence is an incorrect role assignment from a semantic point of view, namely, a wrong connection between label and spatial indicator. The example below clarifies the concept through a distinction between the wrong system annotation and the correct one.

  System annotation: [The square]$_{tr1}$, with a green area [on]$_{si1}$ [both sides]$_{lm1}$, slightly enters with a staircase [in]$_{si2}$ a square called Forum.

Correct annotation: [The square]$_{tr1}$, with [a green area]$_{tr2}$ [on]$_{si2}$[both sides]$_{lm2}$, slightly enters with a staircase [in]$_{si1}$[a square]$_{lm1}$ called Forum.

The square is actually a trajector but it is not linked to the spatial indicator "on" and to the landmark "both sides" but it is related to the spatial indicator "in" and to the landmark "a square".

- **C3 = <u>None Role Error</u>:** The error consists in the assignment of a none label to a word in the sentence that, on the contrary, plays a spatial role. The mistake generally derived from an incomplete system detection of the spatial indicator, as shown in the example below:

  System annotation: [The Hall]$_{tr}$ is [on the left]$_{si}$ of the Temple of Religion, dedicated to public lectures.

  In this automatic wrong annotation there is no landmark associated to the spatial indicator.

  Correct annotation: [The Hall]$_{tr}$ is [on the left of]$_{si}$[the Temple of Religion]$_{lm}$, dedicated to public lectures.

- **C4 = <u>Preposition Sense Disambiguation Error:</u>** This type of error occurs when a general preposition, used in a sentence, is considered a spatial indicator by the system although it is not it. An example is the following:

  System annotation: [The hall]$_{tr}$ that introduced the universe [in]$_{si1}$ its different aspects is[near]$_{si2}$[a Planetarium]$_{lm.}$

  In the sentence an additional spatial indicator appears which is not linked with any label (both trajector and landmark).

  Correct annotation: [The hall]$_{tr}$ that introduced the universe [in]$_{si1}$ its different aspects is[near]$_{si2}$[a Planetarium]$_{lm.}$

- **C5 = <u>Missing Domain Words:</u>** As mentioned before, the operation of the system should be independent from the reference domain, however, the absence of specific words into the training-set, which correspond to possible trajectors or landmarks, could bring about mistakes. For example, the next sentence illustrates how the system was not able to associate to the multi-

word "International Association" the role of trajector and, in its place, it detected "A peculiar aspect".

System Annotation: [A peculiar aspect]$_{tr}$ of the initiative is the idea of concentrating the International Associations [in]$_{si}$ [a land]$_{lm}$ of the Hague.

Correct annotation: A peculiar aspect of the initiative is the idea of concentrating the [International Associations]$_{tr}$ [in]$_{si}$ [aland]$_{lm}$of The Hague. International Associations, in fact, is a specific composite name referred to the "World City" project which is not present in a general or architectural domain.

2. In the second experiment, we merged the abovementioned general corpus with a smaller but specific architectural domain corpus concerning the "World City" project. The latter, is formed of 705 sentences extrapolated from the Corpus, as described in Chapter 2, which contain one or more spatial relationships each. Furthermore, these relations have the particularity of including specific words or multi-words (e.g. Planetarium, Hall of Modern Times, World City, etc.…) used in the description of the project. These are the reasons why we had an improvement of the system results, as shown in the table below (tab. 7).

LABEL	PRECISION	RECALL	F1
Trajector	92.086	94.118	93.091
Landmark	97.143	96.454	96.797
Spatial Indicator	85.417	80.000	81.299
Relation	85.030	94.040	89.308

Table 7: Results based on a more specific training domain corpus.

The improvement of the results, in percentage, are the following:

- **3.751%** for the trajectors;
- **13.168%** for the landmarks;
- **7.730 %** for the spatial indicators.

In addition, we also tried to assign the right *general type* (direction, region, distance) to each spatial indicator and the output are the follow (tab. 8):

GENERAL TYPE			
Direction	44.74	94.44	60.71
Region	83.76	77.78	80.66
Distance	58.33	100	73.68

Table 8: Results concerning the general type assignation.

The performance of the system illustrates that the extraction of trajectors, landmarks and spatial indicators as well as the identification of spatial relations is highly accurate. We report the accuracy of the model in classifying the type of relations in three categories of direction, region and distance as well and obtained a reasonable performance. The trained model of this stage can be used later to extract relations from text and provide input to our graphical visualization component.

6. The three-dimensional representation

6.1 Introduction

"3D datum" is present in our everyday life both in terms of entertainment [78] and in many professional fields. Indeed, the films that we watch at the cinema or at home are in 3D, several images or animations are in 3D, many applications used by modern electronic devises like laptops, I-phones, I-pads, etc. are in 3D, but also numerous images useful to sponsor projects, products or work are in 3D. In addition, the three-dimensional images are also used in the medical field to show clinical tests and radiographs or in the area of cultural heritage. Here, in fact, the three-dimensionality can make a visit to a museum or an archaeological site more exiting and immersive.

The term "3D", as mentioned, is used in several contexts and, despite the fact that it is adopted with the same meaning, it is connected to different sectors of the 3D world like Virtual Reality (VR), Augmented Reality (AR) or Computer Graphic (CG). In particular, CG comprehends all the manipulation techniques of the images through the computer like CAD drawing or photo re-touching. Thanks to the 3D graphic, it is possible to obtain static images, called rendering, and dynamic rendering, named virtual animations [79]. The latter form parts of the VR which consists in the virtual simulation of an object and that can be immersive or not immersive. The difference between the two types of simulations is simple because the former is based on the immersion of the user in the virtual world through innovative devises (special gloves, glasses or suits), while, the latter is founded on a distant observation of the three-dimensional modes through a screen or a desktop. An additional kind of the 3D representation is AR which consists in the enrichment of a real scene with virtual elements. This means that, in VR the scenery can be totally invented, without links to reality, while in AR the base set is a real scene to which virtual models are overlapped [80].

6.2 Innovative Computer Graphic Applications

From what has been said, it is simple to comprehend that the Computer Graphic is a key element of our life but, at the same time, that three-dimensional images or videos do not astound users anymore. Indeed, we need new improvements or innovative methodologies that can capture people's interest and attention.

With regards to the application of CG to professional sectors, such as the Cultural Heritage fields, big efforts are being made to create new systems or methods with educational and scientific purposes [81]. For example, many applications like educational school tools, virtual museums [82] or virtual guides are based on 3D graphic because it is the easiest way to learn, and play. However, building virtual environments is time-consuming and it could be very difficult for non-expert graphic modelers because they need to spend a lot of time trying to comprehend the tool and to create their own models. A revolutionary solution could be the creation of a system able to generate 3D scenes directly from natural language, switching the difficult step of modelling the virtual artefacts. The system should be able to interpret users' goals and to build the virtual scene automatically, just by reading one or more sentences. The principal challenge of the purpose is natural language mapping due to the complex structure and the ambiguity of the texts and because, in common speech, people are used to not mention common sense facts about the environment. This aspect emphasises that natural language is not the best instrument to define positions and spatial relationships of the objects in the space. To better comprehend it we can think, for example, that nobody, when describing a tree, says that it is planted in the ground but he supposes that the listener already knows and imagines it. Hence, it is very important that the system will be able to infer implicit information that is not contained in the document by translating the semantic sense of the text, expressed in common language, into a virtual graphic representation of the description.

The automatic three-dimensional reconstruction is an emerging methodology used in several areas such as art, education or robotics and, nowadays, it is also developing in the Cultural Heritage field [83] in order to promote and represent works of the past or project ideas never realized.

6.2.1 Related works

The amount of literature which has been published in the field of natural language input for creating virtual environments is so large that it is almost impossible to cite all relevant works. Then, in the following, the works most similar to this thesis project are presented.

A text-to-scene conversion system, frequently cited as the first in this field, is SHRDLU [84] which allows users to insert, as input, commands written in natural language with the purpose of using them to move virtual objects around in a small "blocks world". Likewise, in [85] a very simple 3D dynamic graphic visualisation is shown, starting from verbal scene descriptions spoken by a human user. In addition, the output 3D representation can be modified as the user provides a more detailed description of the scene or cites new objects. The successor to Spoken Image, is the SONAS system [86], an intelligent multi-user multimedia system which creates three-dimensional virtual scenes exploiting a synergic combination of inputs, like spoken natural language and gesture. Indeed, the system allows the user to navigate in a three-dimensional virtual city and to interact with it through multiple modes. For example, by providing the instruction "Move the car behind the wall", the user should see the car moving from the current place to the position behind the wall.

In NALIG [87] (Natural Language Image Generation) the authors focused on the analysis of the Italian language and developed a system able to generate scenes by interpreting simple Italian phrases of the form <subject> <preposition> <object> [<reference>] and to define relationships between objects like SUPPORT(A,B) or NEAR(A,B). Another older system is Put [88] which focused on making graphic creations starting from the expressive power of natural language; it is an object-placement system that, once it has analysed the natural use of spatial relations, it uses a combination of linguistic commands and direct manipulations to prototype a complex scene.

Zeng et all. [89] describe a prototype system, called 3D Story Visualiser (3DSV) that generates three-dimensional virtual scenes taking in input simplified descriptions of scenes related to books stories. The authors describe a methodology based on the parameterization of the visual words present in the text, on the creation of a data structure in XML format and on the final creation of a virtual interactive 3D representation in real time. The work [90] presents a further method able to automatically construct three-dimensional scenes starting from narrative books but it introduces in the 3D representations an innovative element: the audio streaming. The latter is used to correctly resolve the dynamic interactions between all

3D representations, so, it represents the base timeline against which all the virtual representations are aligned to provide a multimodal version of the original text.

An interesting storytelling project is CONFICIUS [92] a multimodal animation system which takes as input a single sentence, containing an action verb, and generates a 3D animation with speeches and sound effects. It is a complete framework of language visualization that uses computer graphic techniques with NLP to obtain high-level dynamic scenes. Similarly, in [92] the problem of converting fiction texts into 3D animations is addressed. The strategy described in the work focuses on the annotation, within the text, of a series of elements such as avatars, objects or relationships, and on the subsequent conversion of them into models that populate the 3D environment. The innovation of this work compared to those previously exposed is the ability of the system to create 3D animations starting from texts that are not simplified from a linguistic point of view.

A performant constraint based placement system is presented in the work [93]. This system allows the user to quickly and easily generate complex scenes by using a combination of automatically– generated placement constraints, pseudo–physics, and a semantic database to guide the automatic placement of objects. A similar, but more advanced system is described in the paper [94] in which the authors illustrate an innovative automatic scene generation. The main difference from the previous work is that the 3D scenes derive not only from textual description but also from language voice input and that the constructed scenes are not static images but they are completely interactive. Another interactive system is [95], which generates virtual three-dimensional scenes starting from a cooperative game. The two players are the speaker, who provides an utterance, and a listener, who responds with an action, and, obviously, the more detailed the description of the speaker is, the more accurate the spatial scene will be.

A domain-specific prototype system is CarSim [96] [97][98], a tool able to process accident reports, written in French, to extract information and animate 3D scenes. In other words, it analyses linguistic descriptions of car accidents and it provides an output 3D representation. The pioneering WordsEye [99] is one of the main works centered on this theme, in fact, it creates three-dimensional scenes from input descriptions by converting a parse tree to a dependency representation that, in turn, is transformed into a semantic one. In other words, the tool allows users to describe a scenario through a short textual description and, after a syntactic and semantic analysis of it, it builds a virtual three-dimensional representation of the scene. An enhancement of the original version of WordsEye is *Frame semantics in Text-to-Scene Generation* [100], a system containing 10,000 images, 2,000 3D objects manually

annotated with geometries information and a lexicon of 15,000 nouns. The tool is able to depict scenes starting from the interpretation of natural language and to do little inferring of background locations and poses of characters. A further implementation of WordsEye tool has been presented in the paper [101] where the authors describe a methodology to represent unstated facts in 3D environments. More specifically, the system tries to infer the presence of the "missing objects" which are the ones that are not mentioned in the text but that should appear in the scene. Their placement will depend on the knowledge of the objects described in the textual description and on their explicit constraints. An analogous but more challenging work [3] regards the development of a system able to infer the presence of implicit objects in a scene, or constraints on them, starting from textual descriptions contained in fiction books. The necessity of inferring the presence of implicit objects comes from the fact that the writers generally do not provide detailed explanations of the scenes but they lead the readers to use their imagination to "construct" the settings. This means that the system has to hypothesize the most precise spatial configurations possible by exploiting the information provided by the text. Another work, which concerns the capability of a system to infer the environment of textual descriptions is [102]. It tries to create plausible three-dimensional settings taking into account the objects or the actions involved in the textual descriptions; for example, if we consider the sentence "Carl is having a shower" the aim of the system is to comprehend that the scene is set in a bathroom and not in a bedroom or in a kitchen.

A further interesting system concerning the text-to-scene generation approach is [103]. It is based on manual links between language and objects which are used to create three-dimensional scenes in a virtual space. Conversely, the paper [104] describes an advanced 3D generation approach able to automatically learn from data, without human help. Indeed, it is capable of mapping textual terms to objects and of creating a final 3D representation of the described scene. A recent work [105] proposed the creation of 3D scenes supported by the interaction of the user who is involved in two different steps of the process: at the beginning and the end. In particular, initially, he writes a descriptive sentence to be translated in 3D and, once the system builds the scene, he can modify it by moving the objects into the virtual environment.

Another relevant recent project is MUSE [106] that integrates natural language processing with the development of virtual immersive environments. Its main purpose is to "bring text to life" through 3D interactive storytelling, in fact, MUSE takes as input natural language

texts, like children's stories, and generates a corresponding virtual and interactive three-dimensional world.

Two last works that is worth mentioning are [107] and [108]. The first one describes a system able to receive, as input, a natural language text and to show, as output, images retrieved from an online photo collection. It is not specifically focused on the creation of a 3D scene but it represents, in any case, an illustration of a textual description.

Differently from the abovementioned projects the original contribution of this work lies in three distinct points:

- The "nature" and the provenance of the texts: the proposed system is able to realize virtual three-dimensional scenes starting from already existing texts while, the previous works mainly focused on the generation of 3D scenes starting from sentences invented by the users which the system can easily handle. These sentences, in fact, are most commonly formulated ad hoc for a three-dimensional representation and involve trajectors and landmarks of simple identification, connected to each other by unique spatial indicators. In particular, it is usual for users to use generic names of objects such as *tree*, *chair*, *house* or, at most, proper names of famous works or monuments, like the Eiffel Tower or the Colosseum which, logically, are easily recognizable by the system as entities (subject or object) of spatial descriptions. For the sake of clarity, an example of a sentence written by the user could be "The machine is parked under the tree", in which, in addition to the spaced triplet, there is only the verb of the descriptive sentence. On the contrary, as already said, our system takes into account already existing sentences, more specifically, the sentences extrapolated from the corpus mentioned in chapter 2, which often contain several additional information with respect to the simple triplet <trajector> <spatial indicator> and <landmark>. Such context information are responsible for potential errors because the system could confuse a noun, not belonging to the spatial description, for a landmark or a trajector or a generic preposition for a spatial indicator. An example is shown in the following sentence: *"The Mundaneum itself is located on top of a plateau and has in addition to the main buildings, a large open area for future expansion"*. In the latter, the trajector is "The Mundaneum", the landmark is "plateau" and the spatial indicator is "on top of", while, all other words, specifically, "the main building", " large open area "," future expansion "and" in "are not spatial entities since they do not generate spatial relationships;

- Aim of the work: the second point of innovation is closely connected to the first and concerns the final goal of the work. In fact, compared to the already existing tools, our system does not have a playful or entertainment purpose, but aims to become an instrument able to spread the knowledge of works of our cultural heritage that have never been realized, but which have been described by the authors in written texts. Specifically, the system developed in this thesis work, as mentioned several times, aims to spread the project of the World City;

- Accessibility: the third point concerns the creation of a system starting from the use of an open source tool, easily available on the network and installable, and not of a proprietary software. This means that all users will be able to use the system for free and without any restrictions.

In addition to the three points mentioned above, it must be said that the implementation of the "text-to3Dscene-generation" system has determined, in its preliminary phases, a further contribution: the expansion of the application domain of the tool used for the identification of spatial relationships present in the descriptive sentences (provided as input to the software). In fact, after the annotation of sentences describing the World City project and several training operations, it has been increased its performance in the automatic learning and in the extraction of relational features from sentences referred to this specific architectural domain (testing operation). In the absence of training actions, the tool could not correctly identify the entities (trajector and landmark) present in the text because they are neither common architectural terms (e.g. column, scale) nor basic terms already present in its knowledge domain (e.g. house, tree ...). On the contrary, they refer to the World City domain, where the entities are characterized by specific and univocal names (e.g. Mundaneum, Hall of modern times ...).

We will talk more in detail about this tool and the training and testing operations in Chapter 5.

6.2.2 Task Definition

The main task of this work is to analyse the "World City" project (§ 1.3 The World City Project) and to spread its knowledge to all the users interested in the Cultural Heritage field. In particular, we worked on the generation, as output, of plausible 3D virtual scenes described in some input sentences. These sentences are selected from the sentences-corpus (§ 2.4 Sentences-corpus) and, specifically from the test-dataset (§ 4.4 From the annotation to the automatic spatial role labels and relationships detection).

However, a basic spatial knowledge is the foundation of the automatic 3D scene reconstruction, human knowledge that the machine has to acquire in order to generate a correct three-dimensional representation.

6.2.2.1 Basic Spatial Knowledge

When people describe a scene they usually avoid several details that are not important in the story and could be boring the listen. This information, however, is essential to the machines to "understand" the scene and to give it consistency. People in fact, as opposed to a computer, have a default knowledge that permits them to comprehend and imagine the spatial composition of the space [109][110]. For instance, when we say "the book is near the light", we immediately imagine a table with a lamp and, near it, a book, even if a light could be also hung from the ceiling. This happens because a very strong assumption used in common-sense reasoning is that when an object cannot fly (in this case the book) it must be supported by another object (for example a table). Consequently, the light, near it, cannot be hung from the ceiling but placed on the table too. Given that, to build a realistic representation of the scene, it is important to transfer human reasoning to the machine in the form of instructions or commands that can help the system understand the correct position of the objects in the virtual environment. Very often, in fact, we know a priori the location of the objects in the space, for example, a double bed is generally leaning with its head against the wall and in the centre of the room while a picture is usually on a wall. These rules take into account both general knowledge and the characteristics of the 3D objects and are useful for the machine to generate a plausible three-dimensional scene. They generally depend on the prepositions and on their meaning and take different forms; the main ones are HORIZONTAL_SUPP(x,y) or VERTICAL_SUPP(x,y), if an object is supported by another one, or EMBEDDED_IN(x,y), if it is incorporated into another object.

6.2.2.2 Spatial Preposition Meaning

According to its definition, "a preposition is a word placed before a noun or noun-equivalent to show in what relation the person or thing stands to something else"[18]. However, in general speech, the choice of the most appropriate preposition is not univocal, but depends on the object's lexical category and on several context factors. At the same time, the same preposition may be used to describe different spatial configurations. For example, in the sentence "A is on B" the relationship is given by the preposition "on" which indicates that the two elements are in physical contact because the object A is located on the surface of the object B, against the force of gravity [111]. However, the preposition could have other meanings and uses like in the sentence "the mirror is on the wall" which introduces the concept of verticality. The mirror, in fact, is never put horizontally but is hung vertically on the wall or on another vertical surface. This means that preposition meaning depends on many factors related to the objects involved in the scene, to the general situation of speech and to the geometrical, physical and functional characteristics of the objects [112]. In combination, we have to take into account three other aspects which can influence the final scene configurations: the similarity, the tolerance and the use of prepositions in natural language [113].

The concept of similarity is easily understandable from the following figure (fig.24) which shows two different possible layouts connected to the sentence "there are some apples in the wicker basket".

A   B

Figure 24: Representation of the concept of similarity.

---

[18] < https://www.itsoky.com/2018/08/28/what-is-preposition/>

In the first image (image A), in fact, the wicker basket contains only apples while, in the second one (image B) it also contains other types of fruits. From this simple example we can understand that, without any specifications about the objects or the scene contexts, the "in" preposition may refer to two, or more, different possible 3D representations.

Tolerance, instead, is a concept that provides a variable rate of indeterminacy of the object's position in the scene and, for this reason, it is a big challenge in automatic scene generation. The figure below (fig. 25) illustrates an example of indeterminacy given by the sentence "*the plant is in the corner of the room*" that might refer to two different situations: a) or b). In the first (a), the corner is empty and the indeterminacy rate is low, while, in the second one (b) the corner is occupied by another object, so, the indeterminacy rate is higher.
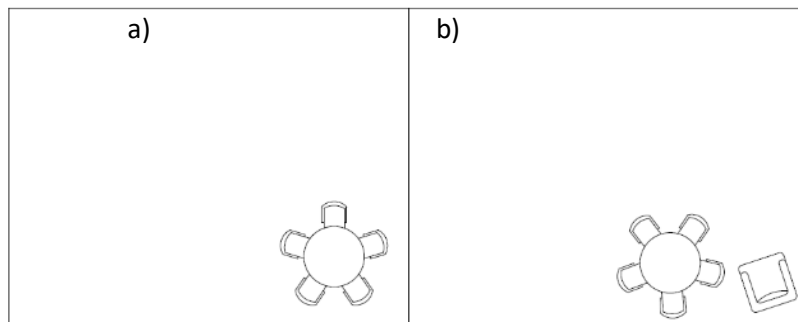
Figure 25: Representation of the rate of indeterminacy in a 3D scene generation.

This time too, a lack of knowledge about the spatial context may cause the existence of more plausible and correct 3D representations.

Finally, the use of prepositions in natural language is also a big problem for machine understanding because, sometimes, the choice of a preposition is given for conventional reasons. For example, in general speech, we say "*in the car*" or "*on the bus*" to indicate the same spatial situation.

According to the above statements, the use of prepositions involves in-depth knowledge of different domains and it is very difficult to formalize. A possible way to create an automatic methodology, able to comprehend and to represent natural language, is to define some primitives with basic characteristics [114].

These primitives could be:

- *object* with *cavity*, that offers the possibility to host other objects;
- *geographical objects* (state, lake, mountains, etc.);

- *fluids*;
- *solid objects,* which can support other objects or stay near/on them.

To each of them, it is possible to assign a "set" of spatial indicators that will have a specific meaning and usage linked to the domain of belonging. With this kind of classification, it could be easier to represent a sentence like "water in the glass" or "an apple in the basket" by assigning to the preposition "in" the correct sense that depends on the objects' structures.

6.3 Proposed Framework

This work shows a representation of spatial knowledge, in fact, a code able to generate a plausible three-dimensional scene starting from an input sentence selected by the user has been developed. More concretely, the operator has to choose, from a given list of sentences describing different scenarios, the one that he prefers. Subsequently, the system is capable of materializing, as output, the textual spatial description in a three-dimensional representation. In particular, it identifies a link between the words that represent the objects involved in the scene and three-dimensional models, previously modelled and stored in a separate 3D database (fig. 26).
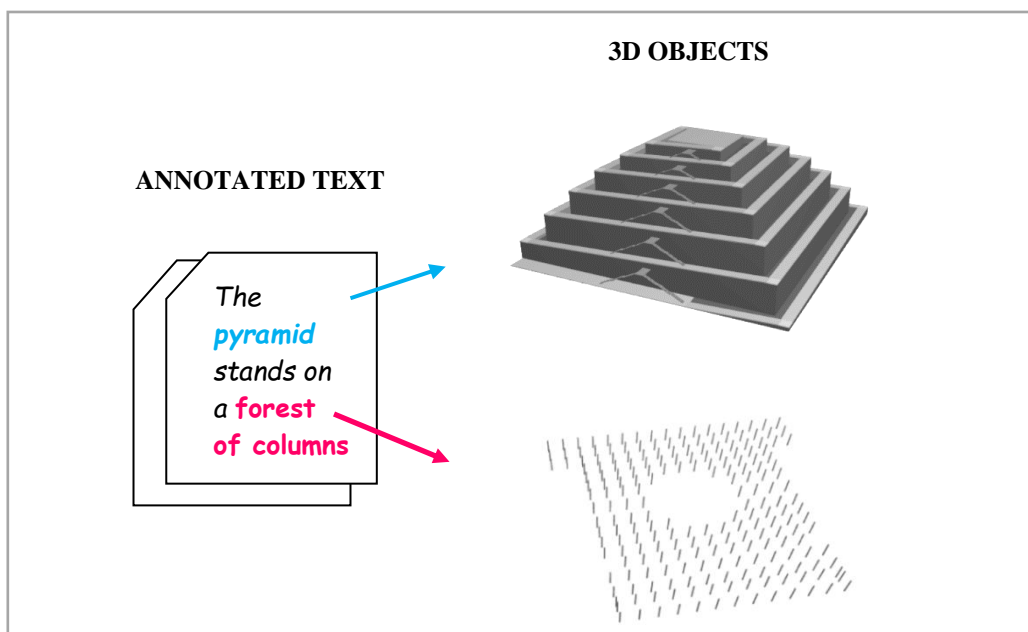

Figure 26: Connection between words and three-dimensional models.

Through a series of consecutive steps, it analyses and parses the text, it identifies the objects of the virtual scene with its relationships, and, in the end, it generates a candidate representation (fig. 27).
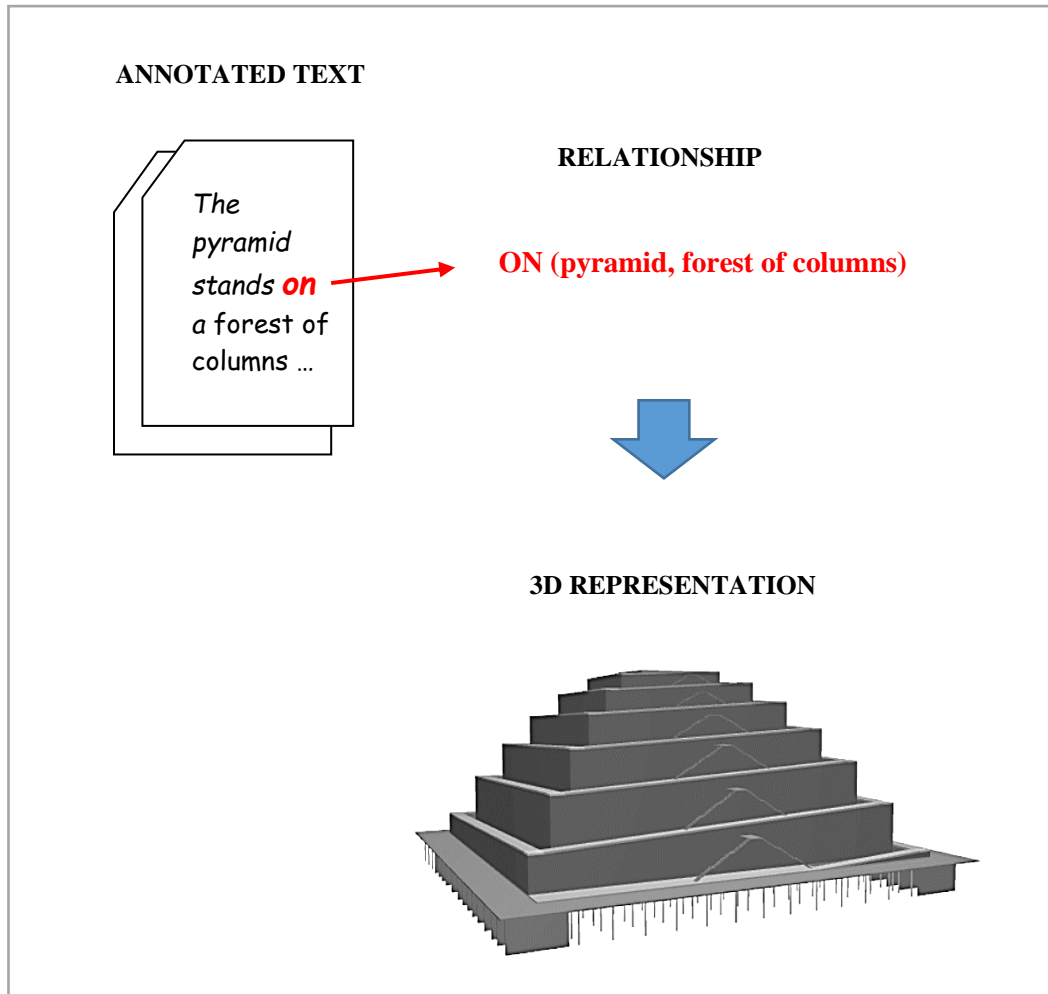

Figure 27: 3D Scene Generation.

The latter is shown as a static scene and, the user, can both see it from any position (top, front, right/left, perspective view) and modify it, changing the objects' location by direct control. The following figure (fig. 28), shows the architecture of the system which handles, simultaneously, textual information and three-dimensional models to provide a final 3D virtual reconstruction of the descriptive selected sentence(s).
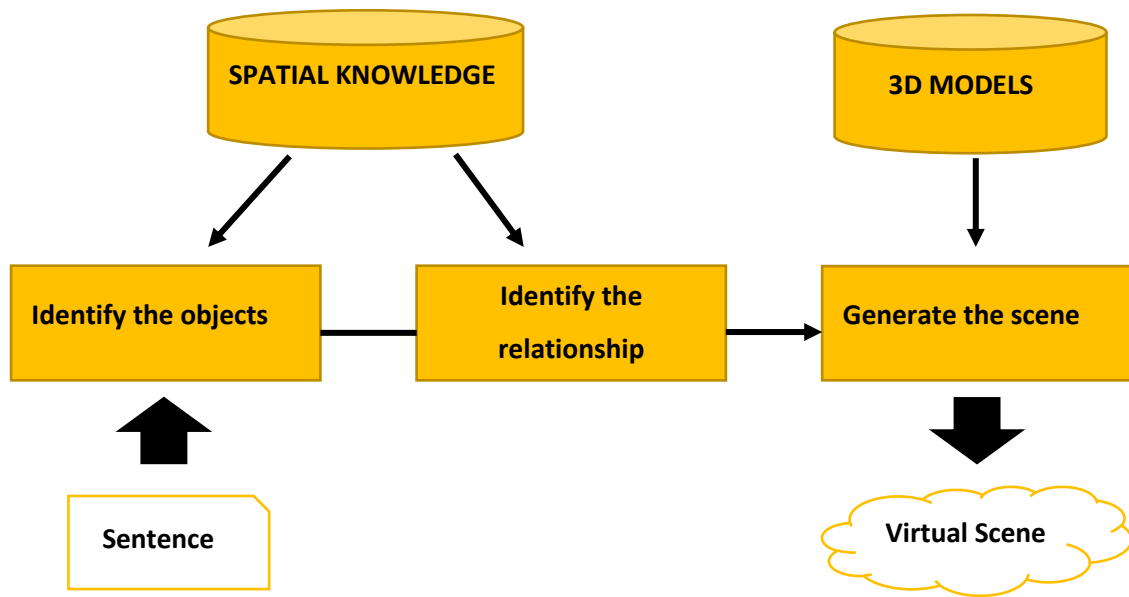
Figure 28: Illustration of system architecture.

6.4 Virtual Reconstruction and Graphical Tools

Virtual reconstructions, namely the creation of three-dimensional virtual models of artefacts, generally derive from a scrupulous and detailed 3D modelling which is performed by means of specific graphical software. Moreover, object modelling is always preceded by an in-depth study of all the available data (texts, drawings and videos) which describe the characteristics of the objects and allow the users to generate the most accurate and truthful models possible, as fast as possible [115]. These software are called photorealistic because they are able to create models practically similar to reality that can be analysed from all points of view. The digital model, in fact, contains all the information concerning its historical, artistic, geometric and dimensional characteristics. Starting from these considerations, it is evident that the three-dimensional digital model, sometimes also connected to an ontology, represents an instrument of knowledge and documentation of artefacts of the past, of modern buildings or of future constructions [116].

There are many types of software which allow the creation of 3D models and, substantially, they can be divided into two large groups: open source software, which can be downloaded for free from the network, and commercial software, which require a license to purchase. The former generally have reduced capacities in contrast to commercial software which often have a higher performance and more user-friendly interfaces. In this scenario, it is

crucially important to talk about the difference between the 3D CADs and the BIM[19] which are two different instruments used to create three-dimensional virtual models. The choice of one software, compared to the others, depends on the final result that the user intends to achieve and on his predispositions and abilities.

It must be said, however, that one single software is not sufficient to perform a "complete work" because the modelling phase, despite being the basic and the most important step, is simply the first of a series of operations to be carried out to obtain a good final result. Subsequently, in fact, we need other programs that allow the decimation of the model, photo editing, rendering and so forth.

6.4.1 BIM: Building Information Modeling

The BIM acronym (which means "Building Information Modeling")[20] refers to the new generation 3D architectural CADs which are not simply limited to the drawing of simple graphic elements (lines, polylines, circles, arcs, etc.), but they offer the possibility to realize one's own projects through the design of technical components of the building like land, walls, floors, pillars, windows, doors and so on.

BIM cannot be simply considered an evolution of CAD, but it is an extremely useful "key tool" in several areas, such as the Cultural Heritage field [117], or the construction one, because it is able to offer a high design support that CAD, on the contrary, cannot guarantee. It represents a new design instrument, more technical and extremely quicker, in fact it allows:

- **a reduction of the difficulty**: the designer is not only assisted during the design phase, but once the 3D model has been created, the tool is able to automatically generate plans, elevations and sections;
- **a reduction in errors**: plans, elevations and sections are not independent drawings but bi-dimensional representations of the same 3D object. This means that any changes made to the BIM model are automatically applied to all the other drawings;

---

[19] The U.S. National BIM Standard Project Committee uses this definition: "Building Information Modeling (BIM) is a digital representation of physical and functional characteristics of a facility. A BIM is a shared knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle; defined as existing from earliest conception to demolition."
[20] < https://www.acca.it/bim-building-information-modeling>

- **a reduction in time and costs**: the designer no longer has to draw a huge amount of lines and polylines, but he has to use and place the objects, with their specific properties and information, inside the project.

Furthermore, BIM technology is based on the concept of integration in the sense that it allows to combine, in a single model, information of different nature concerning the geographical location, the geometry, materials and technical installation properties but also the phases of realization and maintenance of the future [118]. Therefore, the difference between BIM and CAD programs does not concern only the object's 3D modeling technique, but is based on a completely different design approach [119].

Closely related to the concept of BIM is that of IFC (Industry Foundation Classes). It is an open data format, designed to facilitate interoperability between the various users as it allows the interchange of an information model without loss or distortion of data or information.
It must be said, however, that in this thesis project, we decided not to use the IFC format but the .obj one. The latter, in fact, after a series of tests, turns out to be the format that the graphics tool, used for the final creation of the three-dimensional scene, can manage in the best way.

Among the most famous BIM software we have Bentley[21], Revit[22], ArchiCAD[23], Allplan[24] or Vectorworks[25]: all designed to model buildings or architectural objects and, for this reason, equipped with specific tools that can speed up the work and allow instant changes. Although they are BIM software, they also have "pure" modeling tools that allow users to model design objects or more advanced architectural forms through a few basic commands such as extrusion, Boolean operations, lofts, etc.
Among the various software mentioned, the three most used are Bentley, Autodesk Revit and Graphisoft ArchiCAD, all three with high performance. Bentley is an important software in the market that offers the user the opportunity to model, document and display projects of any size or complexity. It is able to realize "smart projects", not simple CAD graphic representations, since it generates models rich in information to support professionals of any discipline.

---

[21] < https://www.bentley.com/en>
[22] < https://www.autodesk.it/products/revit/overview>
[23] < https://www.graphisoft.com/it/soluzioni/>
[24] < https://www.allplan.com/it/>
[25] < https://www.vectorworks.net/bim>

It is used in many sectors, such as architecture, construction, engineering and offers extremely versatile, powerful and easy-to-use modeling tools.

. Revit has had a large and rapid spread in recent years and, with good probability, has also become more powerful than ArchiCAD especially on the modeling of not regular shapes. At the same time, it has a fairly complex interface and usage mode that often make it difficult to learn, especially for users with low computer graphic knowledge. ArchiCAD, on the contrary, is easier to use and, being a program that has existed for several years, presents a large amount of free resources on the web (already modeled objects, tutorials, libraries and so forth) and sharing forums. An additional feature to underline is that it can be used on all the existing operating systems while Revit is limited to the Windows operating system only. For the reasons mentioned above, it was decided to use ArchiCAD for modeling the 3D objects concerning the case study of the World City.

6.4.2 ArchiCAD 3D Modeling

We exploited ArchiCAD functionalities to model the architectural elements of the World City according to the last project realised by Le Corbusier. We decided to realize the virtual three-dimensional representation of Le Corbusier's project, rather than Andersen's project, because, the latter, did not focus very much on the architectural features of the single buildings that would have compose the city, but designed a large-scale urban composition. This aspect would have prevented a detailed 3D modeling of the World City.

At the base of the three-dimensional modeling it was necessary to use an element that could act as a constraint to the development of the entire city, namely, an element able to ensure the correct size of the buildings and an exact location of them with respect to the landscape and to the building themselves. For this purpose, one of Le Corbusier's plan was used. It is one of the numerous drawings designed by the architect in support of his idea in the period around 1928-1929. These are original scale technical plans, elevations and sections, but also perspectives, axonometries or drafts that Le Corbusier realised to describe the characteristics of his complex project. They are drawings of different sizes, with a high historical and architectural value, and they are mainly stored in the archives of the "Fondation Le Corbusier", in Paris.

Below are some examples of the drawings which I used, as guide line, to set the World City arrangement (figures 29, 30, 31, 32, 33).

Figure 29: Drawing, situation plan with legend and orientation; Indian ink; Leather tracing; 0,63 x 0,90 ; Archives of the "Fondation Le Corbusier"; Paris (©AFLC 24521).
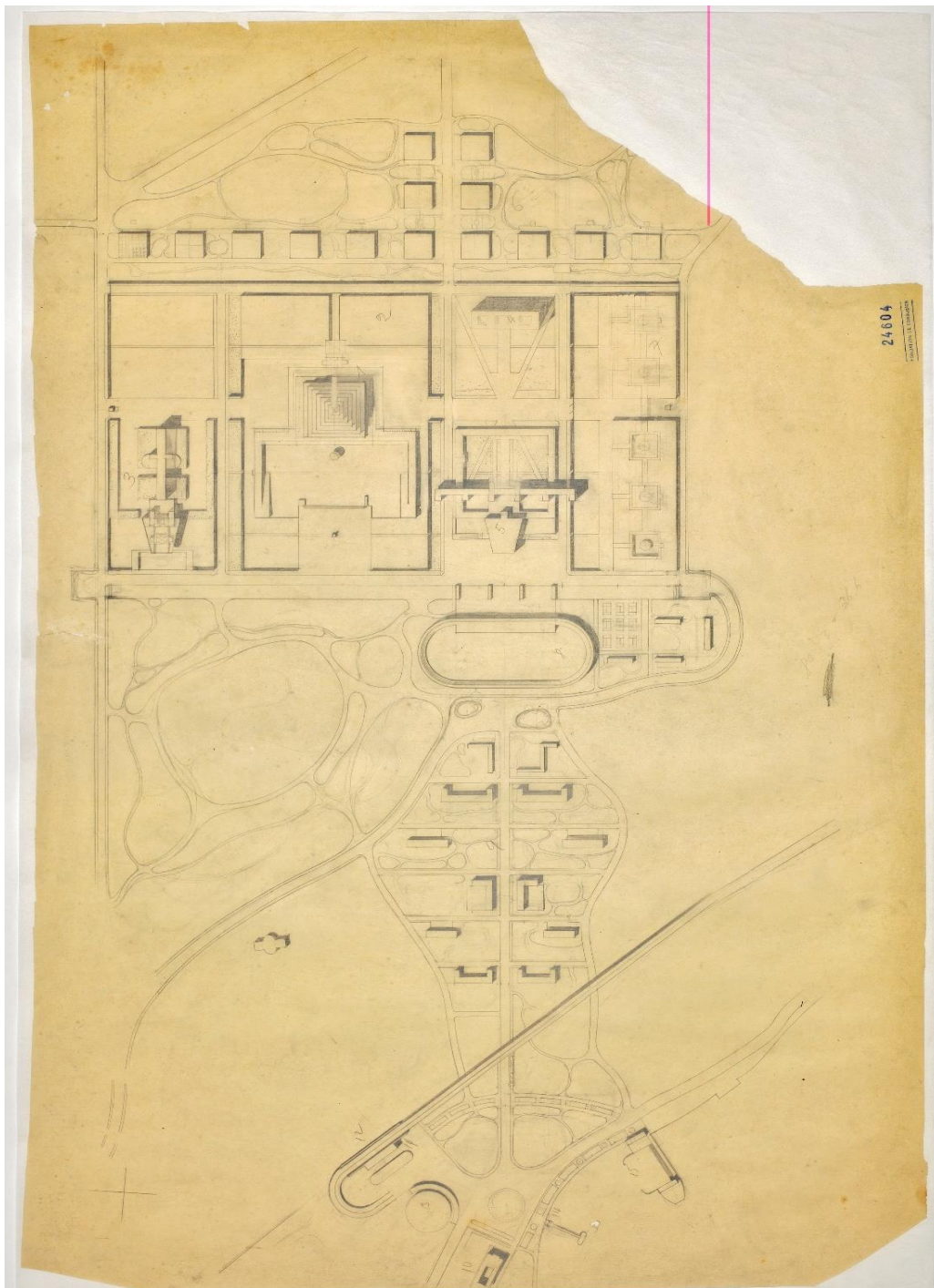
Figure 30: Drawing, plan of the project with museum; Black pencil; Study drawing; 0,74 x 1,11 ; Archives of the "Fondation Le Corbusier"; Paris (©AFLC 24604).
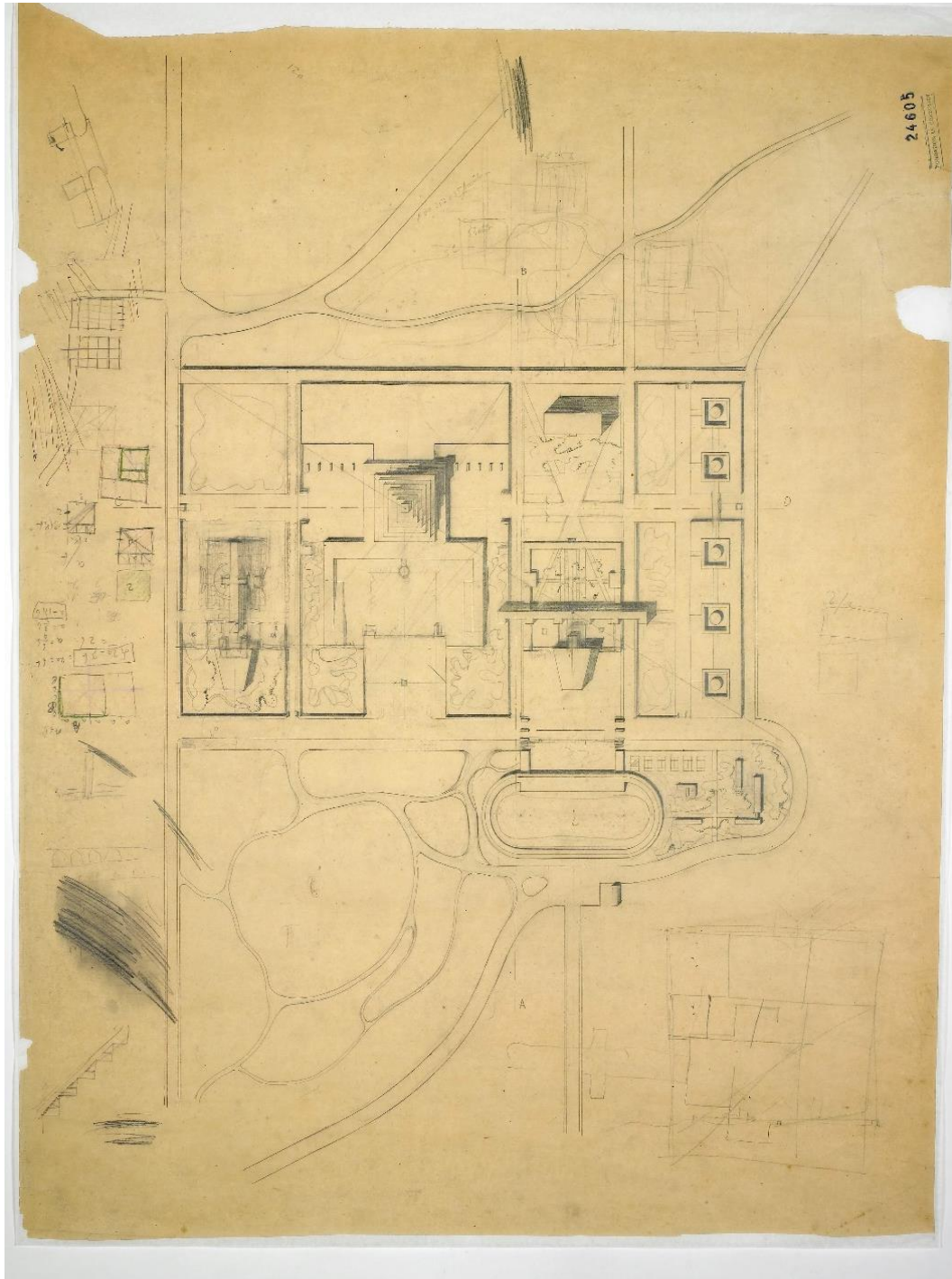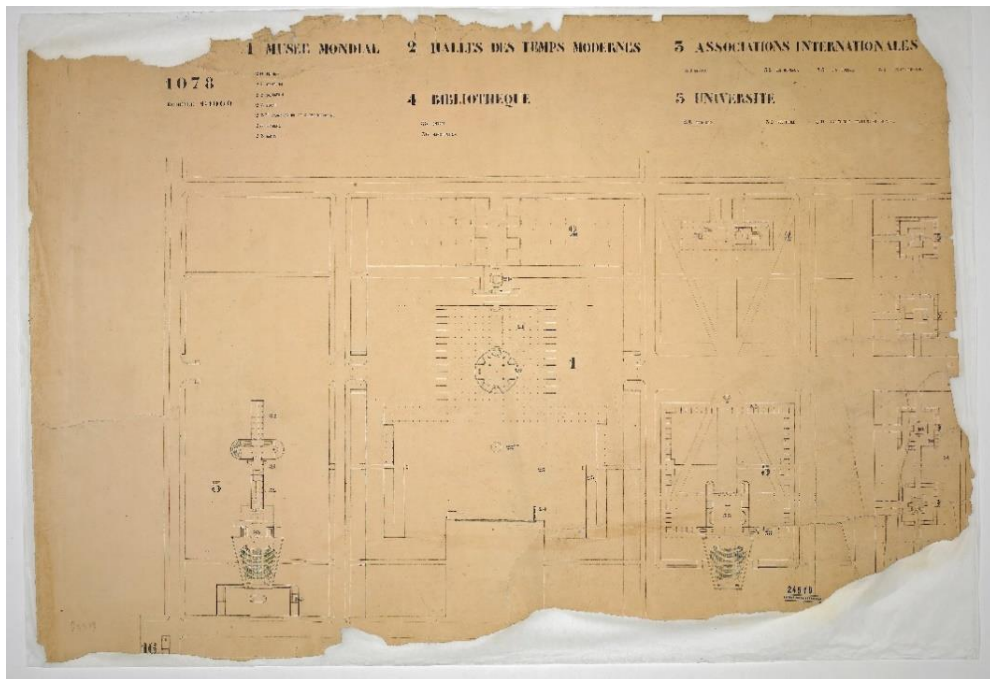
Figure 31: Drawing, Plan of the project with drafts and calculation; Black pencil; Study drawing; 0,73 x 0,8 ; Archives of the "Fondation Le Corbusier"; Paris (©AFLC 24605).

Figure 32: Ground floor: plan of the museum, library, university, with numeration and legend; scale 1:1000; drafts and calculation; Black pencil; Canson paper; 0,93 x 1,40 ; Archives of the "Fondation Le Corbusier"; Paris (©AFLC 24519).
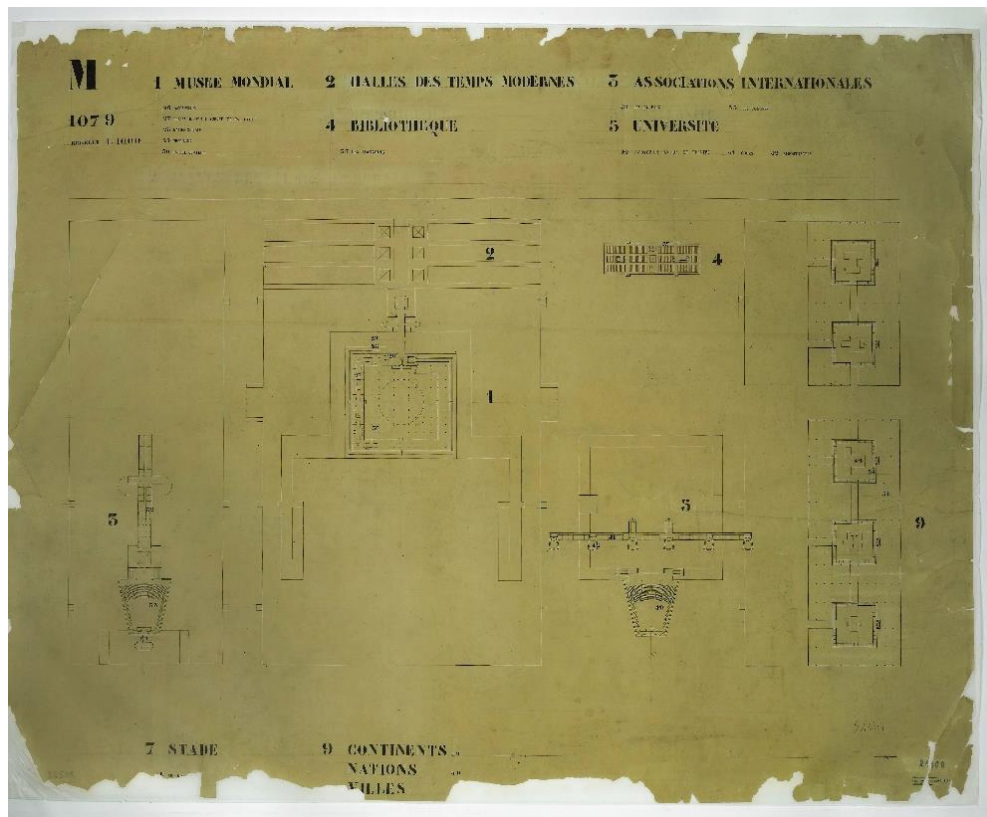
Figure 33: First Floor: study drawing, plan with legend; scale 1:1000; drafts and calculation; Chinese ink; Leather tracing; 1,07 x 1,38 ; Archives of the "Fondation Le Corbusier"; Paris (©AFLC 24509).

103

Immediately, the main problems related to the three-dimensional reproduction of the World City were identified. A first difficulty is attributed to the file size: the City, in fact, is characterised by complex buildings, full of architectural details, which could not be easily handled by a traditional computer, without high graphical performance. A possibility that could solve this problem is to subdivide the entire model into parts, namely, to handle one building at a time. On these bases, we decided to model each edifice in a separate file and, in the end, to merge all of them in a final file, which shows the World City in its entirety. Secondly, we tried to define the highest possible LOD[26] of each building, which is the best balance between the level of detail and the file size. A first operation was to detect all the elements that could play a decisive and characterising role in the three-dimensional representation of the World City buildings. These are, for example:

- the *Sacrarium*: a hidden place of the museum in which Le Corbusier imagined to store stones engraved with the names of men representing human intelligence;

- the *pilotis*: the columns used by the architect to elevate the buildings from the ground in order to give lightness to the structure and to exploit the area below as a garden or passage for pedestrians;

- the *roof garden*: a flat roof with plants used to allow men to contemplate nature and to be surrounded by it;

- the *hollow pyramid*: a museum designed to enclose cultural, mythological and sacral functions just according to the classical tradition. It is a symbolic building characterised by its cavity, which transforms a static and heavy construction into a casing designed to contain world science and knowledge.

- *The stepped pyramid*: the museum has the shape of a stepped ascending pyramid, which recalls the figure of a spiral. The latter is characterised by a strong symbolism, in fact, it indicates men's history and life: something that goes further, with continuity, upwards. According to Otlet's idea, this growing path comes from scientific evolution that brings men towards a deeper knowledge and world progress.

In addition, also the location of the buildings in the space is established based on a precise architectural rule. The constructions, in fact, are placed in order to respond to the *Golden*

---

[26] The term LOD represents the "Definition level" that the objects of the BIM model must assume on the basis of the desired final result.

Ratio[27] that means that the figure which contains the buildings has the golden rectangle proportion (fig. 34).
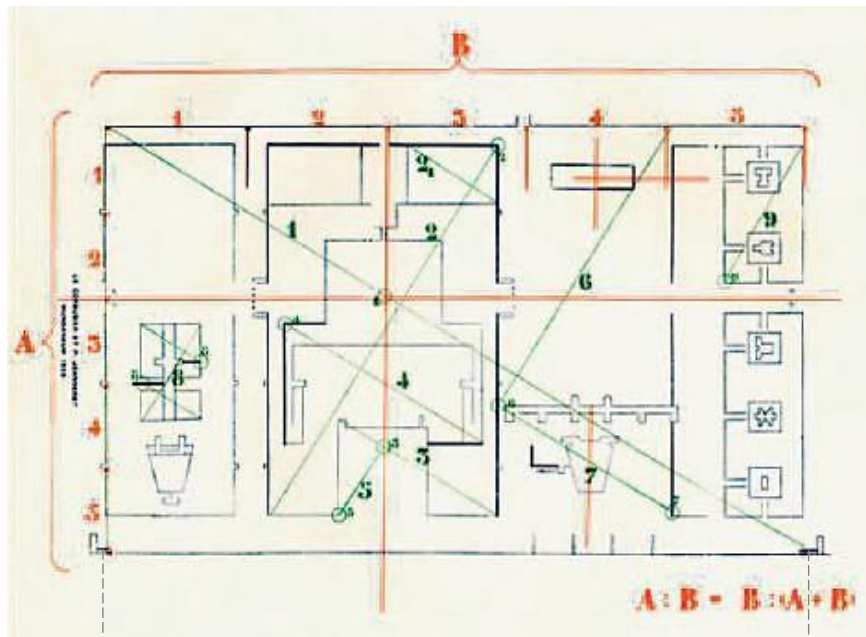

Figure 34: Building distribution according to the Golden Ratio; the explication of the relation A:B=B:(A+B).

In nature, art and architecture the golden rectangle is constructed by drawing an arc with the line drawn from half of the square of the opposite corner as shown in the figure below fig. 35).
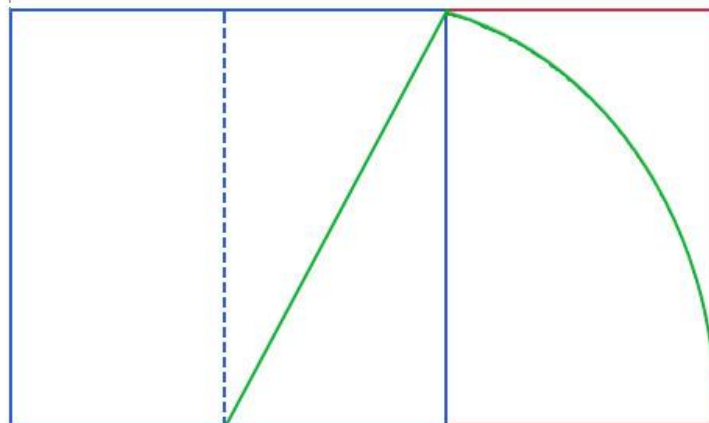

Figure 35: Way to construct the golden rectangle.

---

[27] Golden Ratio: it is a relationship between two numbers which results in the irrational number 1.618033. Since ancient times, the golden section (or divine proportion) was used to create sculptures and works that had the right proportions. The formula is based on the following mathematical calculation: *(a+b) : a = a : b = b : (a-b)*.

The modeling operations followed numerous phases during which different techniques were experimented to solve the various problems encountered on each occasion. The building modeling started from the basic elements provided by ArchiCAD's toolbox palettes (fig. 36): wall, door, window, column, beam, slab, stairs, roof, shell, skylight, curtain wall, morph, object and mesh.
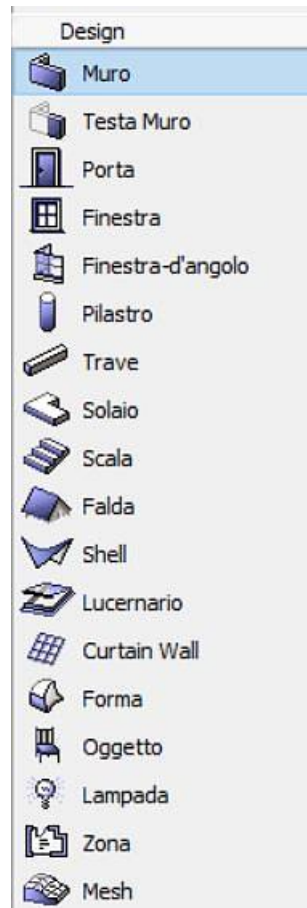

Figure 36: ArchiCAD's toolbox palettes.

The latter, in fact, connected together and subjected to shape and size modification or to Boolean operations, like union, difference, intersection and subtraction, are able to give rise to endless 3D elements.

An interesting instrument, present in the toolbox palette, is the morph tool (fig. 37) which is useful to create freeform elements.

Figure 37: ArchiCAD's morph tool.

Indeed, thanks to this tool, without importing shapes from other software or from the Net, we created all the "special objects", namely, all the 3D elements which do not have common shapes but a particular morphology that characterises them as univocal objects related to the World City project.

6.4.2.1 World City 3D Models

The 3D modeling concerned the main buildings of the World City, which should at first have included the five traditional institutions of intellectual creativity. In particular, only the outer shells of the buildings have been modeled because all the textual descriptions extrapolated from the corpus refer to spatial configurations of external spaces.
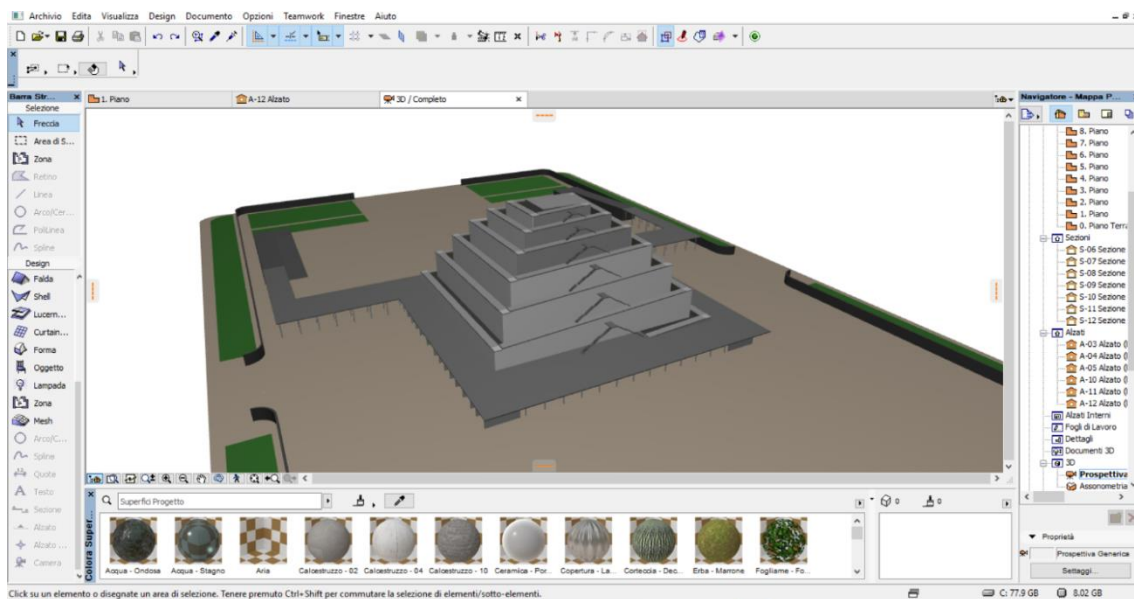
1. The *World Museum* (fig. 38);

Figure 38: Perspective view of the World Museum.

The Museum was designed to show the complex mechanism of the world and the human phenomenon of life, full of events, discoveries or problems. It should have been divided in three different sections, which are: the National and Geographical sections, the Scientific sections and the Historical ones. Museum aim was a regular exposition of simple and cheap things (copies, reproductions or facsimiles), in fact, it should have not had a preservation function of important or ancient objects [16].
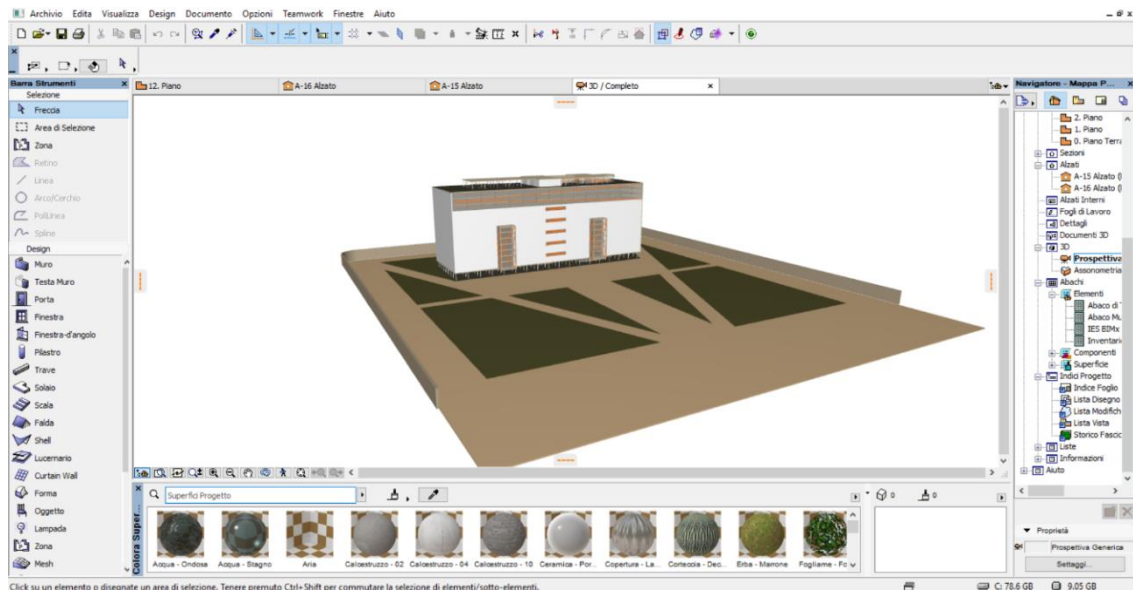
2. The *Library* (fig. 39);

Figure 39: Perspective view of the Library.

The library is intended as the world centre of books and a modern documentary encyclopaedia. It should have collected books from all over the world with the purpose of storing them in a safe place, far from the wars, and of making them freely available to everyone. In addition, it should have also contained: journals, newspapers, official state publications and scientific, social and pedagogical publications, record of laws, manuscripts, inventions, statistics, sheet music photography, photograph records and so on [16].
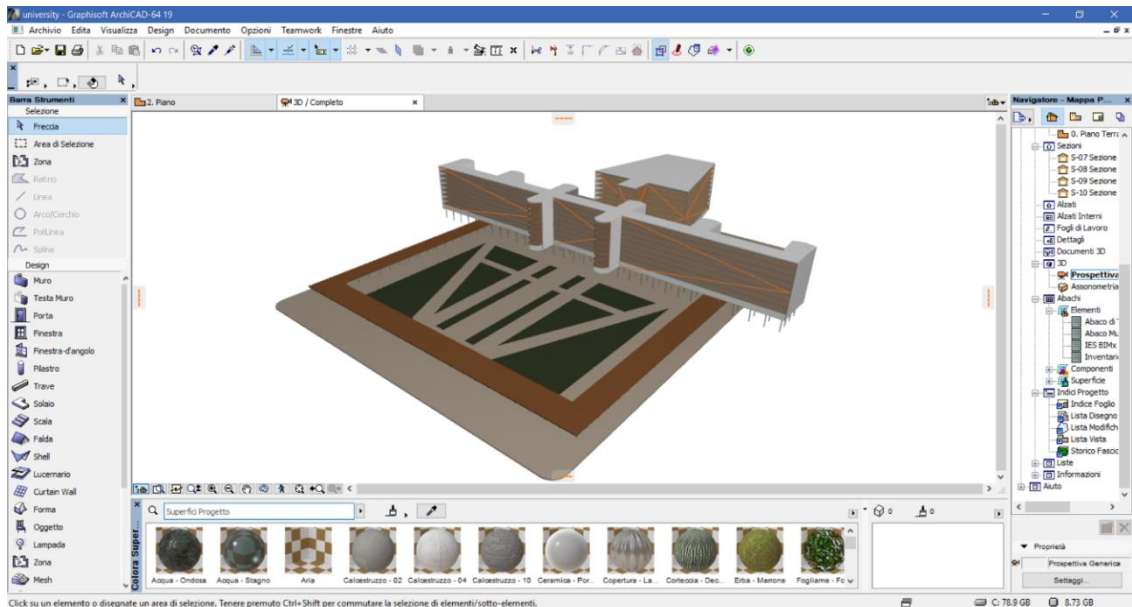
3. The *World University* (fig. 40);

Figure 40: Perspective view of the World University.

The University was planned to be the most important educational centre addressed to students coming from all over the world. It should have been a place to deal with themes like diplomacy, economics, sociology, labour relations, journalism, and welfare [16].

4. The *Building for International Associations* (fig. 41);
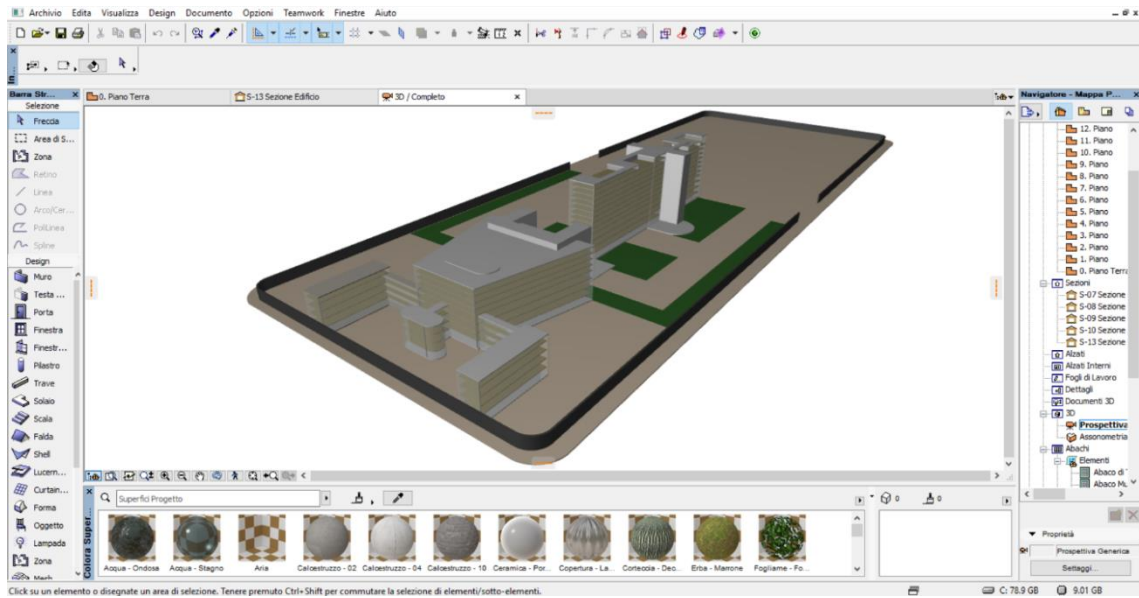

Figure 41: Perspective view of the International Associations.

The aim of the building was to hold the representative offices of several international associations. It should have been a building of peace and social work for artists, scientists and educators [16] in which to hold meetings, conferences and congresses.

5. The *Exhibitions of Continents*, *Nations and Cities* (fig. 42);
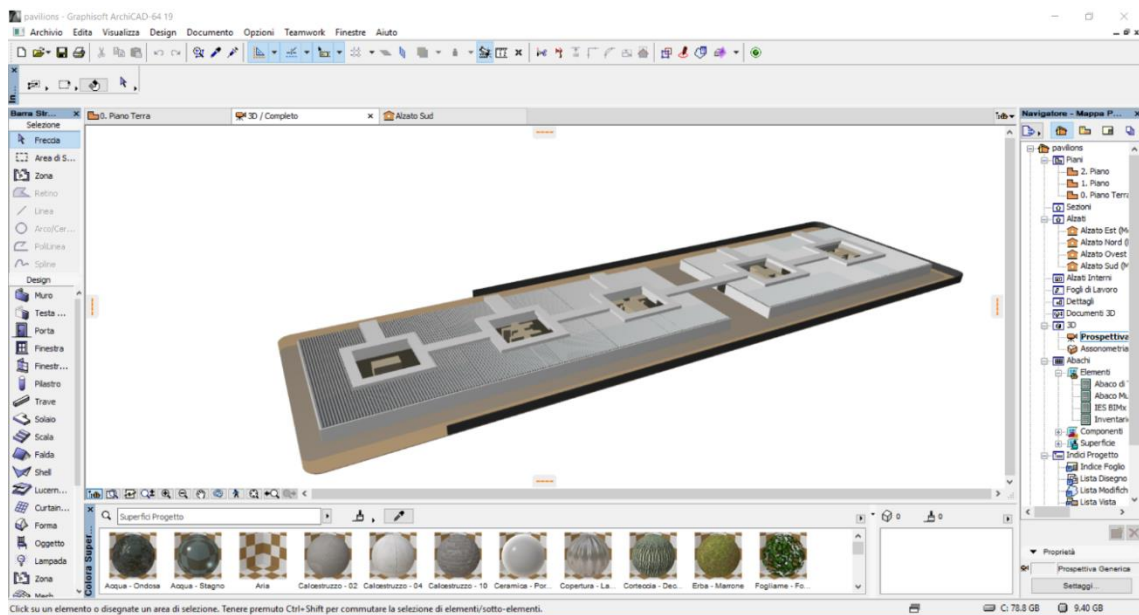

Figure 42: Perspective view of the Exhibitions of Continents.

The building was finalized to contain study rooms and offices and to enclose the exhibitions of continents, nations and cities. The exhibitions, housed in five different pavilions, should have concerned cultural activities like architecture, art or urban planning [16].

In addition, there are another two relevant constructions which are:
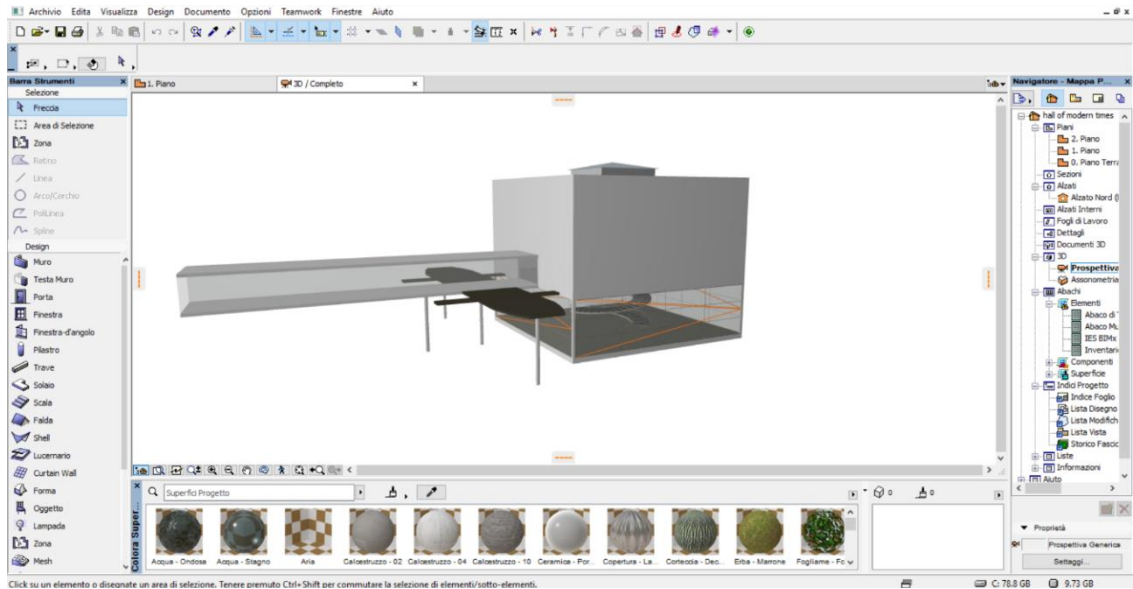
6. The *Hall of Modern Times* (fig. 43);

Figure 43: Perspective view of the Hall of Modern Times.

It was designed connected to the pyramid, in fact, it should have performed the function of access to the museum. Furthermore, it was planned to contain exhibitions of contemporary and cultural creativity.
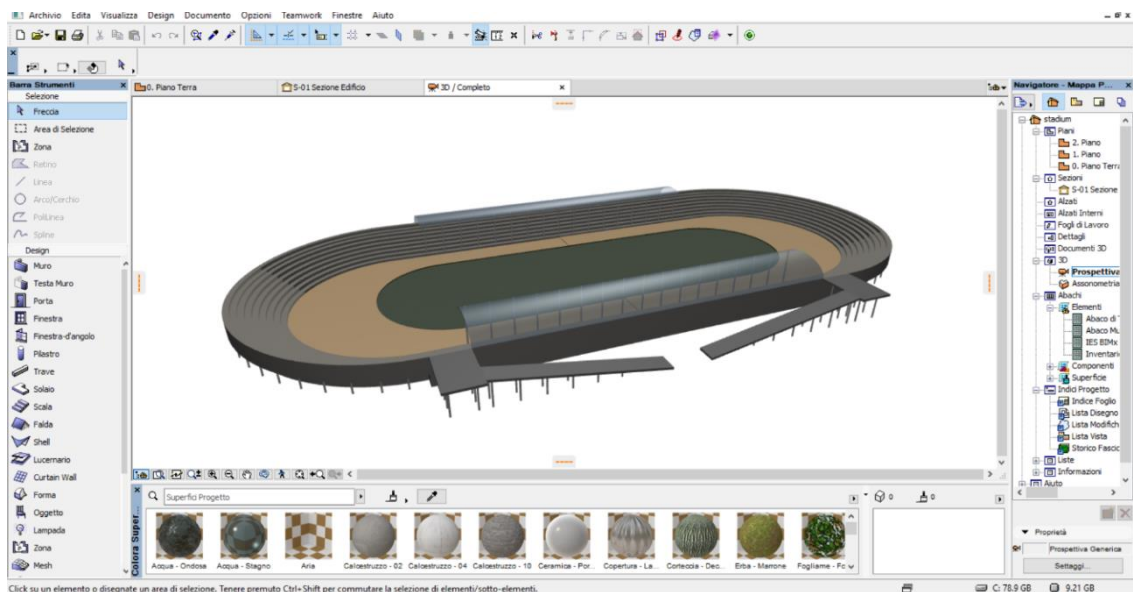
7. *The Stadium* (fig. 44);

Figure 44: Perspective view of the Stadium.

It was located outside the precinct containing the buildings and it should have been the core of the physical culture, in fact, it should have been equipped for all the eighteen Olympic games. Not distant from it, the offices of the Olympic committee were planned.

All these buildings, assembled together and located in the site in order to respond to golden ratio rules, give rise to the World City. The following figures (fig. 45, 46) show two views of the complex project and highlight the relationship between the buildings and the open spaces. Furthermore, the images show the integration between constructions and nature which was one of the cardinal points of Le Corbusier's architectural idea.
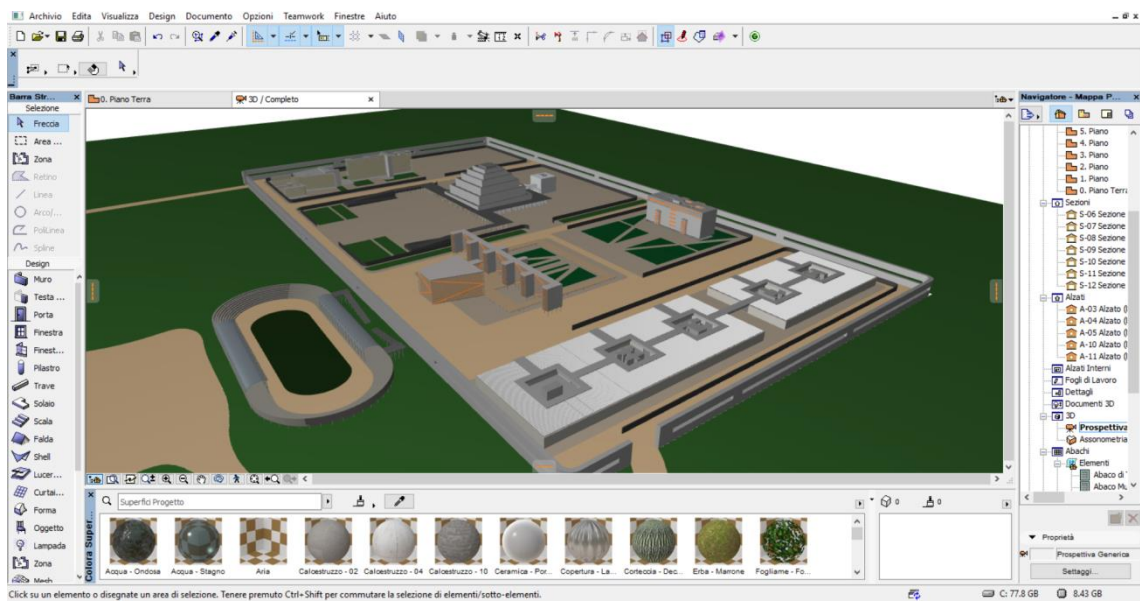

Figure 45: Perspective view of the World City.
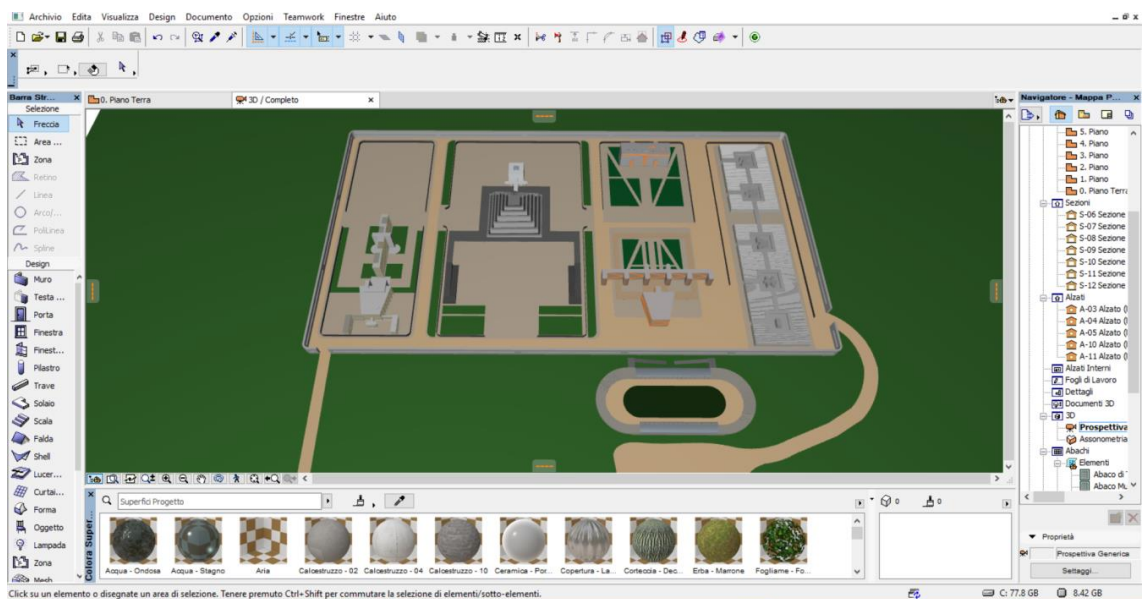

Figure 46: Top view of the World City.

6.4.3 3D Scenes Composition in Blender

Once the 3D objects were modelled with ArchiCAD, the next step was to export each model in a .3ds format and to import them in another software: Blender[28].

Blender is a professional 3D computer graphic software, which is increasingly used to model 3D elements, generate renderings of scenes, create animations, video games and so forth. Its creator was the Dutch animation studio Neo Geo that, in 1995, decided to develop a new in-house application that, in the following years, enlarged its function achieving a high level of performance. In addition, through the open source Python language it is possible to expand the functionalities offered by this software, developing new ones or automating them.

The majority of graphical software users agree that, like other software of its kind, the wide range of commands and instructions provided by Blender, makes it a "difficult" program to use and, in addition, its tricky interface gave it a bad reputation despite its numerous strengths. For example, it requires little space for installation and provides a great variety of geometric primitives, which allow users to generate any type of object, useful for creating photorealistic renderings and animations or even for 3D prints.

That said, it is important to clarify the reason why we did not use Blender from the beginning and we preferred ArchiCAD. The simple motivation is that, contrary to Blender, ArchiCAD permits to draw in plan view with three-dimensional elements which, consequently, give rise to virtual 3D complex objects. In other words, it is possible to create objects with precise dimensions length, height, angles, thickness, and so on and shapes and to locate them to a specific position in the three-dimensional virtual world (x,y,z). These aspects enable to create detailed 3D models related to each other through exact distances or proportions. These reasons led us to divide the process of 3D elements generation in two steps, characterised by the use of the two different and powerful graphical tools.

With respect to Blender software, in this thesis work, we selected and used just a few functions which are the ones that allow to:

- import external 3D objects;
- modify 3D objects by changing their shapes, sizes, locations, etc.;
- apply texture to the models;
- make photorealistic renders;
- handle the creation of three-dimensional scenes writing scripts in Python.

---

[28] <https://www.blender.org/>

6.4.3.1 Why use Blender

Despite the large range of existing 3D modeling software, we decided to generate the three-dimensional scenes concerning the sentences descriptions (§ 2.3 Sentence Extraction) into the Blender virtual environment.

The choice of using Blender, instead of other graphical software, was dictated by several reasons. First of all, it is an Open Source software so it can be downloaded for free from the official site and can be used without any particular restriction. Moreover:

- It is a multiplatform software available for Windows operating systems (both 32 and 64 bit), for Mac OSX 64 bit and for Linux;

- It allows to import and export 3D objects in several formats without losing information;

- Its popularity is increasing and very often, once people have created their own models, they share them on the net;

- It is maintained by a passionate group of users that update it constantly and it benefits from an active community in which people can ask questions or share opinions about its functionalities;

- It could be used at different levels based on users' necessities and skills;

- It makes the "source codes" available;

- It offers the possibility to create or modify a scene by scripts: Blender exploits the power of Python in order to automate time-consuming modelling and animation tasks. It uses a built-in Python console window and a .bpy library (blender python library) thanks to which it gives the chance to create custom panels and menus, build and package add-ons and generate virtual scenes.

6.4.3.2 Blender Reference System

The best way to properly locate the objects in the virtual space is to set a reference system composed of a horizontal plane at the ground level, and a grid composed of three axes where:

- The X-axis represents "front-to-back" movements;
- The Y-axis represents "side-to-side" movements;
- The Z-axis is the direction of "top-to-bottom" motions.

The grid refers to the 3D Cartesian coordinate system in which the origin, or the centre, is located at the (0, 0, 0) coordinate and the X, Y and Z axes have a positive and a negative orientation (fig. 47). Each point in the virtual space can be represented in function of its three coordinates (X, Y, Z) which allow to easily locate the 3D objects into the virtual space.
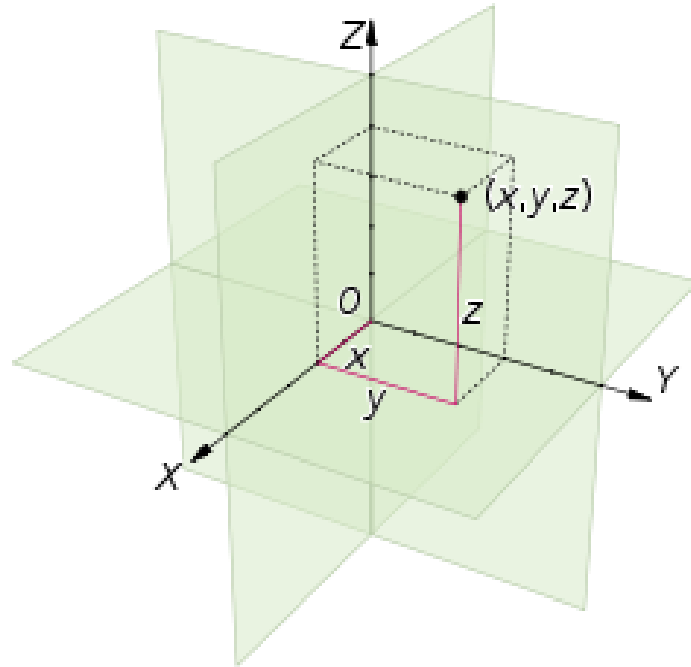

Figure 47: Three-dimensional Cartesian coordinate system.

Analysing the problem from a psyco-linguistic point of view, the positive and negative directions of the axes give us the possibility to assign different meanings to the spatial indicators. For example, the axis oriented according to the direction of the force of gravity is a very good reference to represent concepts like ABOVE or BELOW, while, the positive and negative X-axes are a good way to discriminate spatial indicators such as IN_FRONT and BEHIND. In some cases, like the last one, the reference is associated with an implicit observer, in fact, the word BEHIND is referred to an observing direction which cannot be deduced directly. In other cases, the directional aspect is not relevant, for example, if we consider the sentences "*the car is outside the parking*" or "*the sofa is near the television*" the direction is not important because the position of the objects could be assigned by means of spherical coordinates.

The analysis of prepositions is a very sensitive issue because they can completely change the three-dimensional representation of the virtual scene. Moreover, the main challenge is that spatial configurations often depend more on the objects involved in the scene than on the prepositions. For instance, let us consider the preposition IN; it generally describes a

relation between the volumes of the objects because one of them, the trajector, is totally or partially contained in the other one, the landmark. However, sometimes there is not a volume relationship because the two objects are completely independent. For example, in the sentences "*the camel is in the desert*" or "*the bird is in the sky*", the desert and the sky could be seen like an environment of the scene and not a geometric volume.

6.4.3.3 Spatial Indicators Definition

We defined a set of formal spatial indicators which are mapped to natural language terms. These are the most common prepositions used in natural language and they are measured using axis-aligned bounding boxes from the viewer's perspective. They are used to assign to the trajector a movement in the virtual scene, in fact, each of them, according to its meaning, is connected to a specific command in Blender graphical tool.

The construction of the scene is based on some hypotheses and benchmarks:

1. In a simple scene, trajectors and landmarks are linked by **spatial indicators** that specify the relative position of the objects in the virtual world;

2. The initial position of each object is given in (0, 0, 0) coordinates that is the centre of the scene. Furthermore, the centre of the object is, by default, its barycentre, as shown in the figure below (fig. 48);
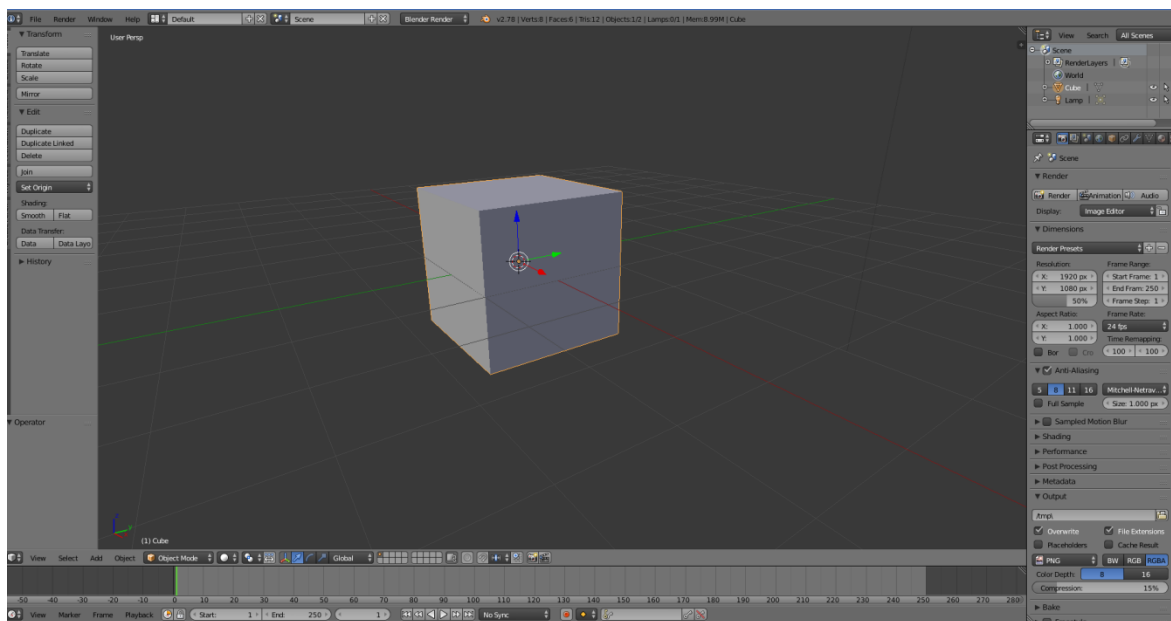

Figure 48: Representation of an object located in the centre of the virtual scene.

3. The landmark remains in the original position (except when the spatial indicator involved is *between*) while the trajector moves compared to the fixed landmark;

4. Trajector movements depend on the spatial indicators that, in turn, are expressed with reference to the landmark dimensions.

Let us analyse the principal spatial indicators used in natural language and the corresponding Blender command that permit to move a selected object in the virtual space. With regard to the latter aspect, as previously mentioned, Blender has an internal Python interpreter that allows users to add functionality by writing scripts in Python. These scripts are powerful, versatile and easy to understand and they can be written in a specific text window. The interaction between the graphical tool and the programming language takes place by means of the ***bpy module*** that Blender provides to the Python interpreter. Once it is imported in a script, it gives access to Blender data, classes, and functions.

Prepositions and graphical tool commands are the following:

- **ON:** the preposition "on" indicates that the trajector is put in direct contact with the landmark that supports it (fig. 49).
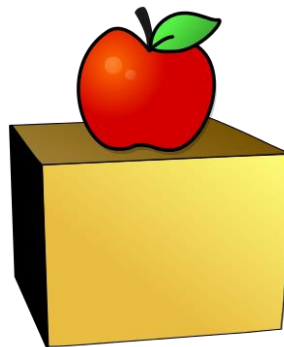

Figure 49: Representation of "on" preposition where the apple is the trajector and the box is the landmark.

For this reason, the trajector has to move upward until it reaches landmarks Z_{lm}. Its final position is:

$$(\, 0 \, , \, 0 \, , \, z_{lm} \,)$$

while Blender's command is:

bpy.ops.transform.translate (value= (0 , 0 , float*z_{lm}))

In the latter instruction appears:

- the call to a translation operator (bpy.*ops.transform.translate*);

- the amount of the translation *(value= (_ , _ , _))*;

- the *float* term which indicates a particular type of number with floating point and real values. Specifically, they are real numbers written with a decimal point that divides the integer and fractional parts. The decision to use float numbers is due to the necessity to deal with much more than integer numbers in order to ensure a greater accuracy.

- **ON THE LEFT:** the spatial indicator "on the left" implies that the trajector has to move to the left (along y-axis), with respect to the user's point of view, so, referring to the landmark, to its right. In this case we are in front of a *relative frame of Reference* because there is a ternary relation between the arguments that are: the landmark, object_1, the trajector, object_2, and the viewpoint, the user. The image below (fig. 50) better explains the concept.
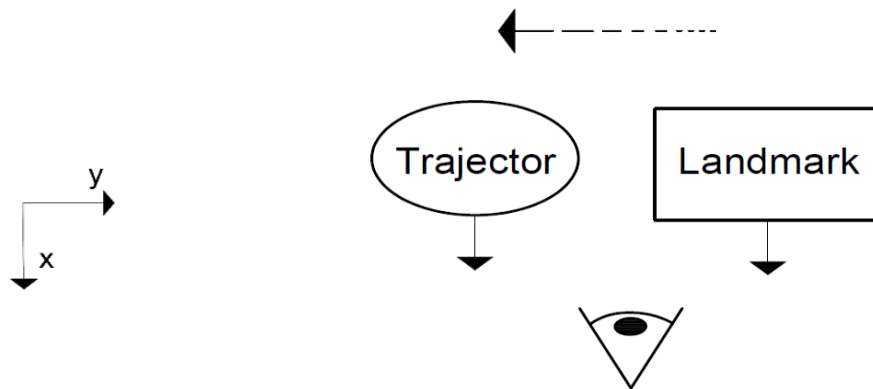
Figure 50: Bi-dimensional representation of the spatial indicator "on the left".

The arrows represent the normals to the frontal surfaces of the objects; for example, in a church, the normal is located perpendicular to the side that contains the portal. The movement of the trajector is fixed equal to the y-dimensions of the landmark in order to assure a proportionate distance between the two elements:

$$(\, 0 \, , \, - y_{lm} \, , \, 0 \,)$$

117

and Blender's command is:

bpy.ops.transform.translate(value= (0 , - float* y_{lm} *random.uniform(a,b) ,0))

The random function is a Python module that allows to generate random numbers within a defined range. In particular, I used the *uniform* random module function which permits to generate a random floating number between *a* and *b*. On these basis, in the absence of specific distance information (for example 1m, 80 cm, etc.), the result of Blender command is not a fixed movement but a more generic one.

The painted and two-dimensional representation of the final scene is shown in the following figure (fig. 51).
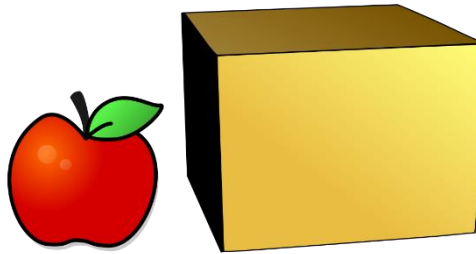
Figure 51: Representation of "on/to the left" preposition where the apple is the trajector and the box is the landmark.

- **ON/TO THE LEFT OF:** conversely to the previous spatial indicator, in this case, there is no ternary relationship between elements but a binary one. Indeed, the point of view of the user is not involved, rather, the object moves only with respect to landmark position (fig. 52).

Figure 52: Bi-dimensional representation of the spatial indicator "on the left of".

Trajector translation is:

$$(0 , + y_{lm}, 0)$$

while the graphical tool's command is:

bpy.ops.transform.translate(value= (0 , + float* y$_{lm}$ *random.uniform(a,b), 0))

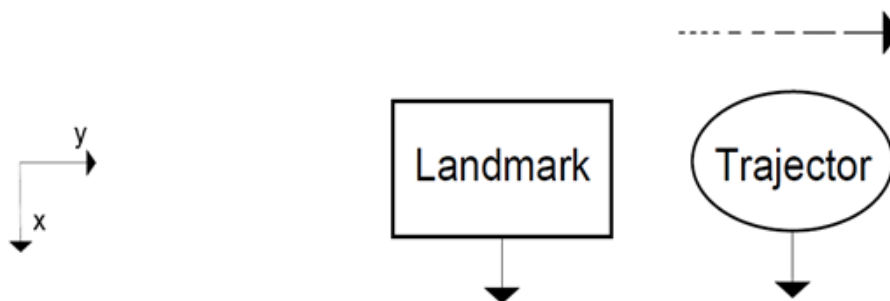- **ON THE RIGHT**= A reasoning similar to that of the spatial indicator "on the left" can be done for the spatial indicator "on the right". The difference is that, in this case, the trajector moves to the right of the landmark (along y-axis), with respect to the user's point of view, and to the left with respect to the landmark.

- **ON/TO THE RIGHT OF**= also in this case, the trajector moves in the opposite direction with respect to the movement performed in the previous example, that is, to the left with respect to the user's point of view, and to the right with respect to the landmark.

- **IN= INSIDE:** the prepositions "in" and "inside" depict situations in which an element, the trajector, is contained in another one, the landmark. There is no specification about the exact position of the trajector, in fact, it can be in the centre of the axes or not, but it certainly touches the landmark with its base (horizontal surfaces) and it could or could not be tangential to the landmark profile. Its final location is:

$$(\alpha , \beta, z_{lm})$$

where *α* and *β* are two random rational numbers needed to drag the object away from the centre of the landmark.

Blender's instruction, instead, is:

> *bpy.ops.transform.translate(value=(float*random.uniform(a,b),*
>
> *float*random.uniform(a,b), float*z$_{lm}$)*

- **IN THE CENTRE OF = IN THE MIDDLE OF:** The main difference between the spatial indicators "in" and "in the centre of" is that, in the first case, we do not know trajector's position, but we only understand that it is located inside landmark's borders. Conversely, the spatial indicator "in the centre of" indicates that the trajector is situated exactly in the geometrical centre of landmark's shape (fig. 53). In other words, we can consider it a more specific spatial indicator which provides additional information, useful in spatial descriptions.
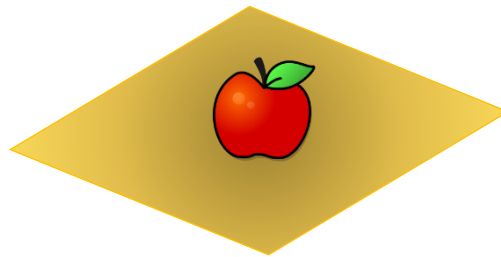
Figure 53: Representation of "in the centre of" preposition where the apple is a trajector and the plane is the landmark.

Hence, the position of the trajector with respect to the landmark is:

$$(0 , \ 0, \ z_{lm})$$

while, the graphical tool command is:

bpy.ops.transform.translate(value=(0, 0, float* z_{lm}))

- **IN FRONT OF**= when an object is in front of another one it is necessary that their normals point toward each other (fig. 54).
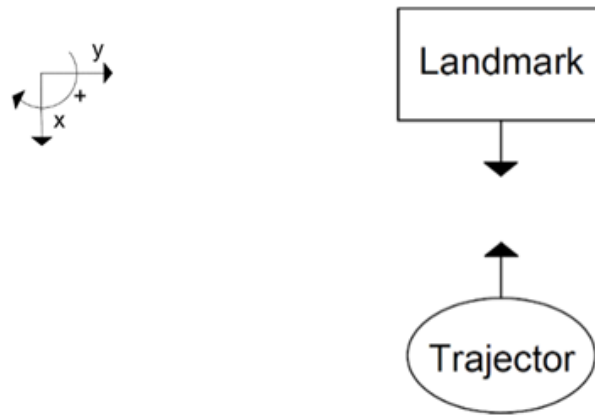
Figure 54: Bi-dimensional representation of the spatial indicator "in front of".

For this reason, the trajector has to move twice: it has to translate forward and to rotate 180 degrees. The following two consecutive commands have been set up:

bpy.ops.transform.translate(value=(float*((x_{tr} + x_{lm})/2 , 0 , 0)))

and

bpy.ops.transform.rotate(value= π, axis= (0 , 0 , 1))

In the first command, two aliquots, x_{tr} + x_{lm}, which correspond to the x-dimension of the trajector and the landmark, have been added and divided by two in order to ensure a reasonable distance between the two objects.

The second formula, instead, means that the rotation value is equal to π and it happens around the z-axis.

- **BEFORE:** the difference between the prepositions "in front of" and "before" is founded on the trajector's orientation. Indeed, while in the first case its normal points to landmark's one, in the second case the two normals have the same directions (fig. 55).
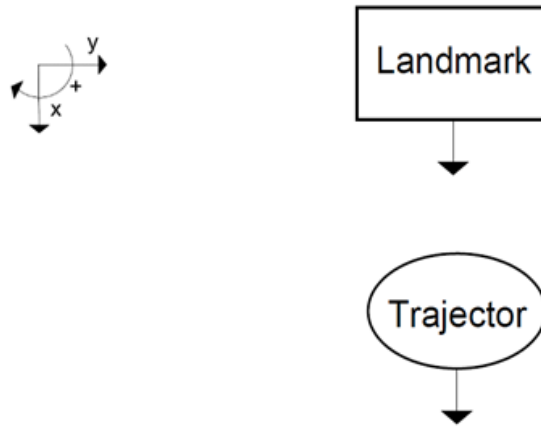
Figure 55: Bi-dimensional representation of the spatial indicator "before".

The landmark, as in the previous examples, stays still in the centre at the origin, whereas the trajector moves forwards to reach its final position:

$$(x_{tr}, 0, 0)$$

The instruction given to Blender to translate the object forward, at a proportional distance from the landmark, is:

bpy.ops.transform.translate(value=(float*((x_{tr} + x_{lm})/2 , 0 , 0)))

In other words, conversely to the "in front of" commands, the object does not rotate at all.

- **BEHIND**: an element is considered behind another element when it is on the back of it and it is partially or totally hidden by its shape. The spatial indicator does not specify the distance between the two objects but it only indicates that the trajector is situated in a rearward position with respect to the landmark (fig. 56).
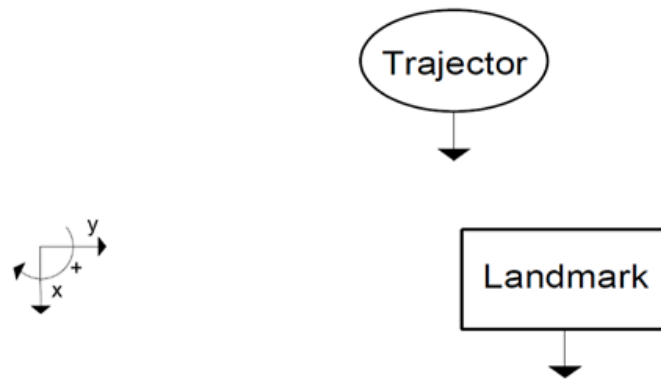
Figure 56: Bi-dimensional representation of the spatial indicator "behind".

This means that trajector's final position is:

$$(\alpha, -\beta * y_{lm}, 0)$$

where:

- α is a random rational number, within an assigned range.

- β is a random rational number, greater than one, within an assigned range.

The corresponding command to give to Blender is:

bpy.ops.transform.translate(value=(float* random.uniform(c,d),

- float*y_{lm}*random.uniform(a,b), 0))

- **IN THE BACKGROUND**: Conversely to the case mentioned before, when we say that an object, a trajector, is in the background we point out that it is located at the back of another element, a landmark, at a considerable distance from it (fig. 57).

Thus, the last location of the trajector is:

$$(0, \ - \ \alpha \ *y_{lm}, \ 0)$$

where**:**

- α is a random rational number much greater than one.
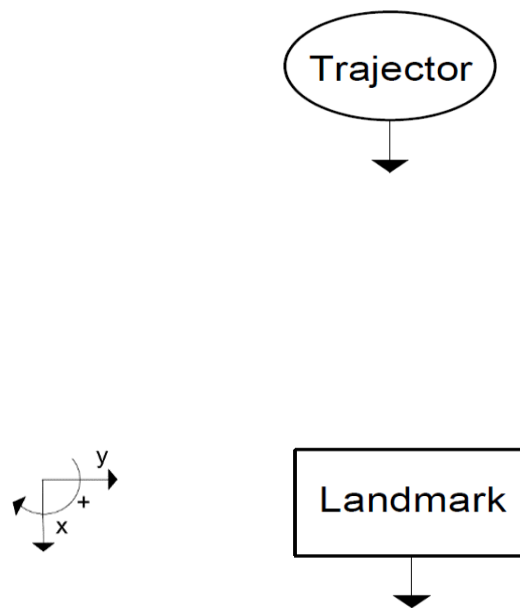
Figure 57: Bi-dimensional representation of the spatial indicator "in the background".

and the command to assign to Blender is:

bpy.ops.transform.translate(value=(0, - float* random.uniform(a,b), 0))

- **IN THE FOREGROUND**: The difference between the spatial prepositions "in the background" and "in the foreground" is based on the opposite location of the trajector with respect to the landmark. In fact, an object is in the foreground if it is the element located as far as possible in the front of the scene (fig. 58).

Figure 58: Bi-dimensional representation of the spatial indicator "in the foreground".

Trajector's final position is:

$$(0, \ +\alpha \ *y_{lm}, \ 0 \ )$$

where:

- α is a random rational number much greater than one, in order to accurately distance the two objects.
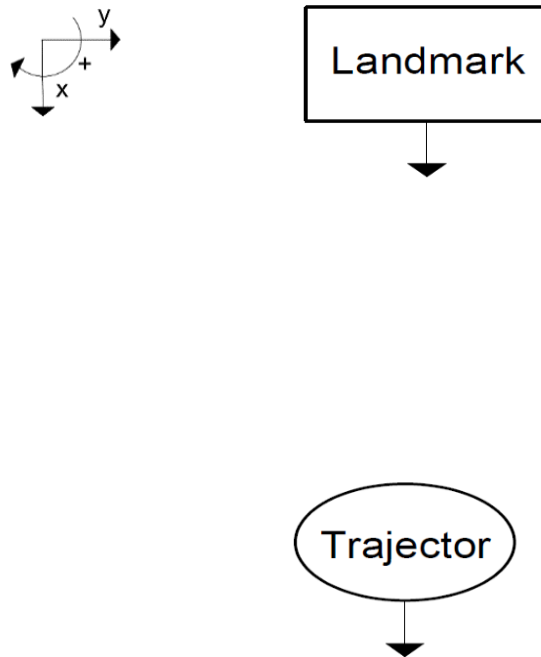
while Blender's command is:

bpy.ops.transform.translate(value=(0, + float* random.uniform(a,b), 0))

- **NEXT TO = NEAR:** the words "next to" and "near", which could appear in a sentence, do not provide a clear and precise indication of the final location of the trajector with respect to the landmark. The object, in fact, could be *to the left* or to *the right of* the landmark, but also *in front of* or *to the back of* it. This means that it is necessary to use, once again, a random function (*random.uniform(a, b)*) that permits us to generate a casual position of the 3D element into the scene.

All eight possible translations of the trajector in the x-y plane have been collected in the following list:

$$\textbf{\textit{motion = bpy.ops.transform.translate(value= [m\_1, m\_2, m\_3, m\_4, m\_5,}}$$
$$\textbf{\textit{m\_6, m\_7, m\_8])}}$$

where:

$m_1 = $ *(+ float *x_{lm}* random.uniform(a, b) , 0, 0)*

$m_2 = $ *(- float*x_{lm}* random.uniform(a, b) , 0, 0)*

$m_3 = $ *(0 , + float *yl_m * random.uniform(a, b), 0)*

$m_4 = $ *(0 , - float *yl_m * random.uniform(a, b) , 0)*

$m_5 = $ *(+ float *x_{lm} * random.uniform(a, b), + float * yl_m * random.uniform(a, b) , 0)*

$m_6 = $ *(+ float*x_{lm}*random.uniform(a, b), - float * yl_m*random.uniform (a, b) , 0)*

$m_7 = $ *(-float * x_{lm}*random.uniform(a,b),+float*yl_m*random.uniform (a, b) , 0)*

$m_8 = $ *(-float *x_{lm}* random.uniform(a, b), - float * yl_m *random.uniform (a, b),0)*

They are schematized in the following figure (fig. 59).

Figure 59: Bi-dimensional representation of the eight possible configurations of the scene after a trajector random translation.

The casual translation of the trajector can be set up by applying the command:

bpy.ops.transform.translate(value=(random.choice(motions)))

Another aspect to take into account is that the trajector could be rotated, compared to the landmark local reference system, therefore, after its translation, a possible rotation in the range of 0 degree to 360 degrees (2π) needs to be introduced. The instruction is:

bpy.ops.transform.rotate(value=random.uniform(0,6.28), axis=(0,0,1))

Example final configurations are shown in the following figure (fig. 60).

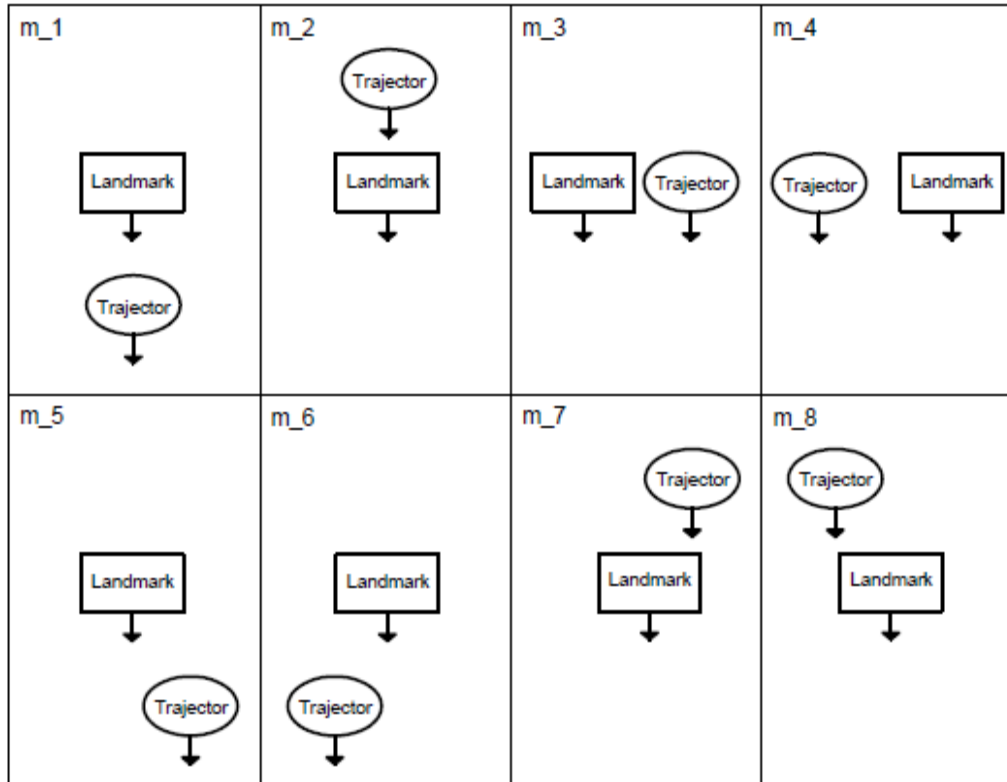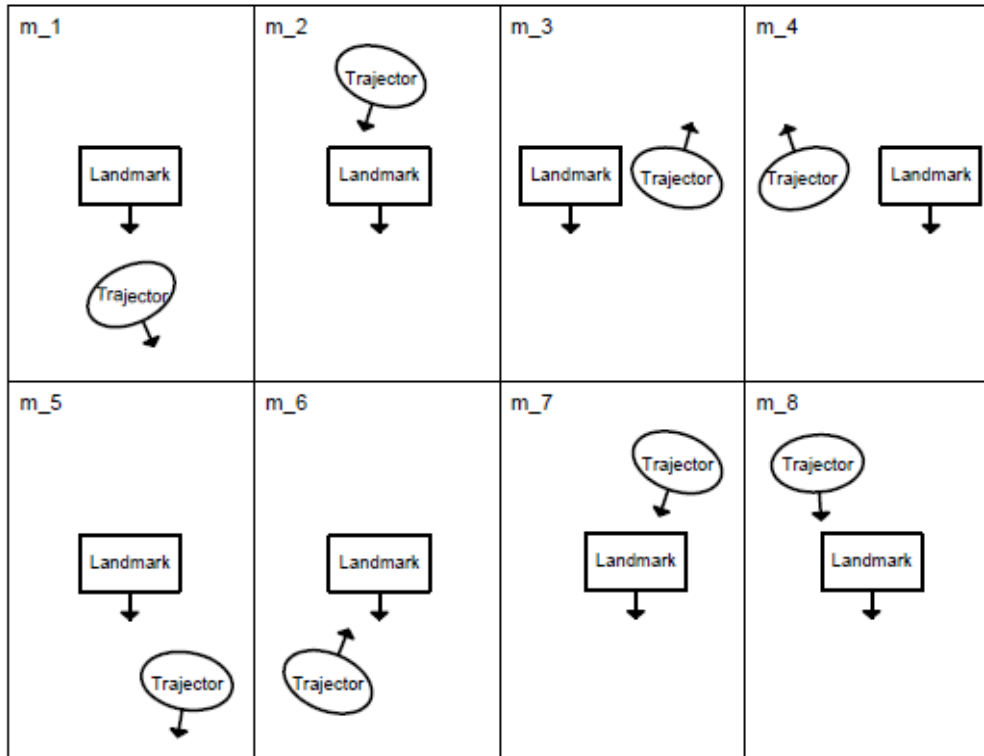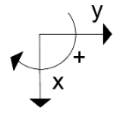

Figure 60: Bi-dimensional representation of the eight possible configurations of the scene after a trajector random rotation.

- **BETWEEN:** the spatial indicator "between" implies the presence of, at least, three objects in the scene: a trajector and two landmarks. A landmark, in fact, is located on the left side of the trajector and, a second one, is located on its right (fig. 61).
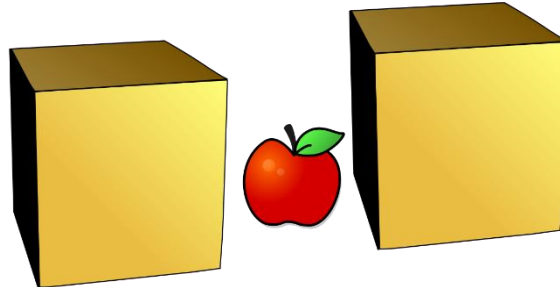
Figure 61: Representation of "between" preposition where the apple is the trajector and the two boxes are the landmarks.

In general speech, the distances between each trajector and the landmark are not specified and they can be the same or different. This aspect has led us to use, also in this case, a random function to assign casual distances between each pair of objects. Unlike all the other cases, the system receives by the spatial indicator a different instruction, in fact, the trajector must stay still in the centre of the scene while the landmarks must change their positions.

For the sake of simplicity, let us assume that the objects stay on the x-axis or on the y-axis, so, the final positions of the two landmarks are:

$$\begin{cases} (+\,xtr\,,0,0) \\ (-\,xtr\,,0,0) \end{cases} \quad \text{or} \quad \begin{cases} (0,+\,ytr\,,0) \\ (0,-\,ytr\,,0) \end{cases}$$

So, the command to assign to the objects are:

$$\begin{cases} \mathbf{bpy.\,ops.\,transform.\,translate(value = ((+(float * Xtr * random.\,uniform(a,b)),0,0))} \\ \mathbf{bpy.\,ops.\,transform.\,translate(value = ((-(float * Xtr * random.\,uniform(a,b)),0\,,0))} \end{cases}$$

$$\text{or}$$

$$\begin{cases} \mathbf{bpy.\,ops.\,transform.\,translate(value = (0,(+(float * ytr * random.\,uniform(a,b)),0))} \\ \mathbf{bpy.\,ops.\,transform.\,translate(value = (0,(-(float * ytr * random.\,uniform(a,b)),0))} \end{cases}$$

In the two couple of formulas, the difference of the sign $(+\;$ or $\;-)$ indicates that the two landmarks move to opposite directions along x-axis, in the first couple, and along y-axis, in the second.

- **ABOVE**: Even though the words "on" and "above" are generally used as synonyms in common speech, there is a large difference between them. Indeed, in the first case, the trajector is in contact with the landmark while, in the second, there is no physical interaction. For example, if we say "*the shelf is above the table*" we would like to indicate that a shelf (trajector) and a table (landmark) are distant from each other, specifically, we describe a scene in which the shelf's z-position is higher than the table's. Conversely, if we say *"the glass is on the table"* we refer to a situation in which the lower surface of the glass is in contact with the upper surface of the table.

  To formalize the "above" spatial indicator in graphical software it is necessary to modify the z-value of the trajector leaving the z-value of the landmark at zero. The amount of the translation depends on the types of objects that are involved in the scene but, to set a standard value, it has been assumed that it is equal to the sum of landmark's z-dimension and trajector's z-dimension in order to ensure an adequate distance between the two objects.

  $$(0, 0, (z_{lm} + z_{tr}))$$

  Blender's corresponding command is:

  $$bpy.ops.transform.translate \ (value= (\ 0 \ , \ 0 \ , \ float*(z_{lm} + z_{tr}) \ ))$$

- **BELOW**: As in the case of the spatial indicator "above", an object is "below" another object if they do not touch because one of them stays in a lower position. Given that the landmark does not change its location, the trajector has to move downwards and its final location is:

  $$(0, 0, - (z_{lm} + z_{tr}))$$

  And Blender's instruction is:

  $$bpy.ops.transform.translate \ (value= (0 \ , \ 0 \ , \ - \ float*(z_{lm} + z_{tr})))$$

- **OUTSIDE:** Let us consider two objects, A and B. When we say that A is outside B it means that B does not contain A and that the two objects are not even tangential to each other. Hence, to locate the trajector in the right position, it is necessary to take into account both dimensions of the landmark and trajector, in the XY plane, together with an additional rate. The latter ensures that the trajector completely exits from the circumference of radius *(R+r)* on which there could be some contact points between the two elements. The increment is assigned by using a random function *random.uniform(a,b),* where *a* and *b* are respectively the start-point value and the end-point value of the range. In this way the barycentre of the trajector is translated from the point P (fig. 62) to a further position respect to the centre of the circle(s): to a point of the circumference of radius *[(R+r) + random.uniform(a,b)]*.

To sum up, the two objects, landmark and trajector, are both imported in the centre of the scene but, while the first remains in the original position, the trajector moves outside the profile of the landmark according to this formula:

$$x_p{}^2 + y_p{}^2 = (R+r)^2 \quad \rightarrow \quad (\tfrac{X_{lm}}{2} + \tfrac{X_{tr}}{2})^2 + (\tfrac{Y_{lm}}{2} + \tfrac{Y_{tr}}{2})^2 = (R+r)^2$$

R: landmark radius
r: trajector radius
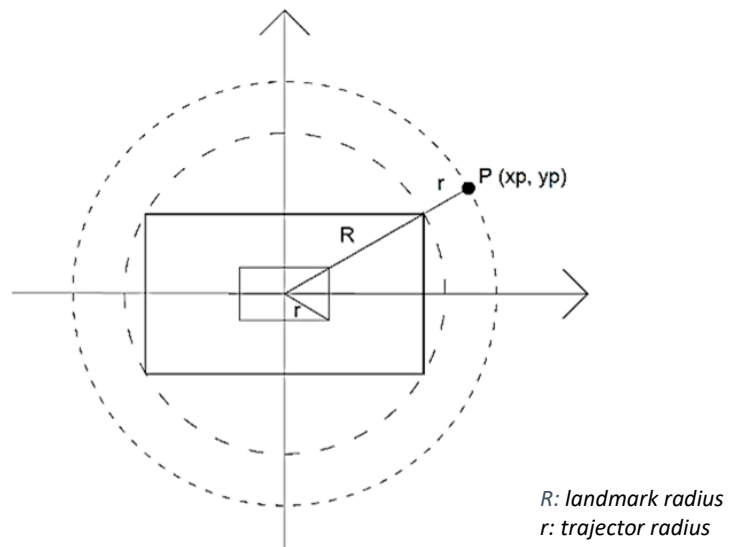
Figure 62: Illustration of the circumferences that circumscribe trajector's and landmark's shapes.

In the figure:

- the smallest rectangle (A) circumscribes the trajector shape;
- the biggest rectangle (B) circumscribes the landmark shapes;
- P(x_p, y_p) is a point on the circumference of ray (R+r).

- **TOWARD:** Contrary to the previous cases, "toward" is a different type of spatial indicator because it defines a trajector's movement. The latter, in fact, does not remain in a fixed position but it moves and its motion is in the direction of the landmark. For the sake of simplicity, we can hypothesize that it runs on a linear path, not on a curved one, and that, from a distant location, it reaches the landmark or it surpasses it (fig. 63).
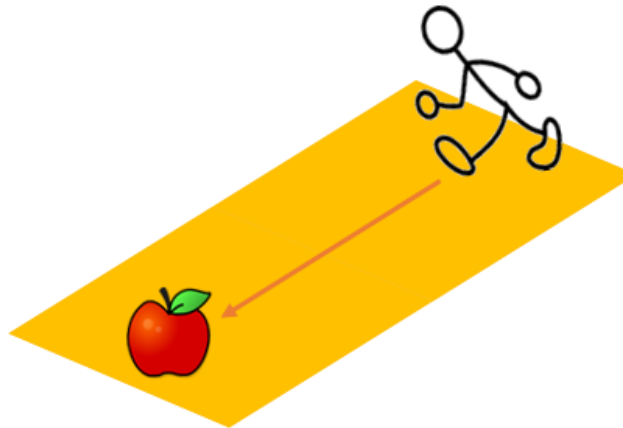
Figure 63: Representation of "toward" preposition where the man is the trajector and the apple is the landmark.

Trajector's motion takes place along a straight line passing through the origin centre which is the position of the landmark. According to the simple rules of the Cartesian plane, the trajector runs on a line defined by the formula $y = m*x$ where m is the slope and it is equal to: $m = \frac{y_{tr} - y_{lm}}{x_{tr} - x_{lm}}$ (fig. 64).
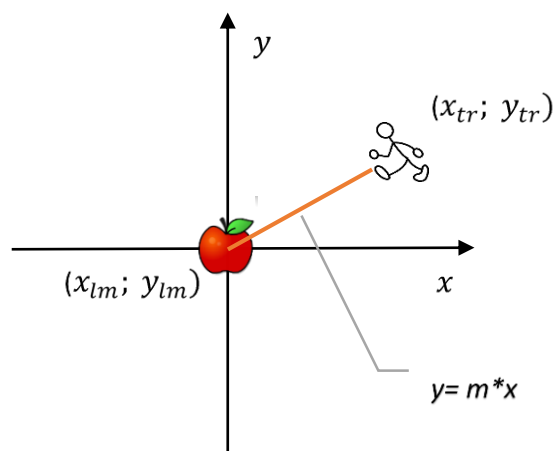
Figure 64: Illustration of the straight line passing through the X- and Y-axis origin.

Considered that landmark's coordinates are (0; 0) the slope and the "path equation" are:

$$m = \frac{y_{tr} - 0}{x_{tr} - 0} \quad \rightarrow \quad y = \frac{y_{tr}}{x_{tr}} * x$$

Thus, the command to assign to the graphical tool is:

bpy.ops.transform.translate(value=

(±random.randint(a,b),±random.randint(a,b),0))

In the latter formula a different random function appears: *random.randint(a,b)* which does not deal with float number but only with positive and negative integers (used in the Cartesian plane) between a fixed range. Moreover, the ± sign indicates that the initial location of the trajector can be in the first, second, third or fourth quadrant. Thanks to this command, the trajector moves from the X- and Y-axis origin to a generic point in the X-Y plane (with Z=0) and, from this position, it starts its motion according to the above-mentioned equation: $y = \frac{y_{tr}}{x_{tr}} * x$.

All the above mentioned spatial indicators represent trajector's movements in the virtual environment. They are managed by graphical tool instructions that "translate" a textual preposition, or a combination of words, into a set of commands to assign to the objects.

Each spatial indicator, so defined, has a connection with one of the eight topological relations (RCC-8), the directional relations or the distal relations described in section 3.6.2. In particular, we can show some examples in the following table (tab. 9):

ON	REGION	RCC-8	EC(a,b)	INTRINSIC
ON THE LEFT = ON THE LEFT OF	DIRECTION	RELATIVE	LEFT	INTRINSIC
IN = INSIDE	REGION	RCC-8	DC(a,b)	INTRINSIC
IN THE CENTRE OF	REGION	QUALITAVIVE	MIDDLE	INTRINSIC
IN FRONT OF	DIRECTION	RELATIVE	FRONT	INTRINSIC
BEHIND	DIRECTION	RELATIVE	BEHIND	INTRINSIC
NEXT TO = NEAR	DISTANCE	QUALITATIVE	CLOSE	INTRINSIC
BETWEEN	REGION	RCC-8	DC	INTRINSIC
ABOVE	REGION	QUALITAVIVE	CLOSE	INTRINSIC
BELOW	REGION	QUALITAVIVE	CLOSE	INTRINSIC
OUTSIDE = BEYOND	REGION	RCC-8	DC	INTRINSIC
TOWARD	DIRECTION	RELATIVE	FRONT	INTRINSIC

Table 9: Correspondence between the spatial relations and the topological, directional and distal relations.

6.5 Scene Generation System

The three-dimensional representations are the result of a combination of 3D objects and rules. In particular, to generate a virtual scene we need both a collection of 3D models, representing physical objects, and a program able to link all the textual information with the stored virtual elements. Accordingly, we modelled and collected the three-dimensional entities and we wrote two different scripts: one in a Python Shell and the other in the graphical tool. The two scripts are connected to each other and, with an automatic procedure, are capable of transforming a sentence, written in natural language, into a 3D representation of the described scene. In the upper part of the code there are all the import commands that allow Python to read and use specific files and methods (import *os, xlrd, xml.etree.cElementTree, time, csv, random*) while, below, is the real script that could be divided into 6 different sections:

1. A file containing sentences extracted from the primary and secondary sources of the "World City" project (§ 2.4   Sentences-corpus - § 5.5.1 Experimental Results) is given as input. At this stage the system, using a random algorithm, selects one of these sentences and shows it to the user. The latter will be transformed in a 3D representation.

2. Once the sentence is selected, a "screening step" begins: the name(s) of the trajector(s) and of the landmark(s) present in the text are shown to the user.

3. Subsequently, another file containing the main characteristics of the 3D objects is given as input. These characteristics are:
   - the ID number;
   - the object NAME;
   - the PATH of the 3D dataset;
   - the OBJECTNAME = name of the object in the graphical software file;
   - the TR_X_DIM/ TR_Y_DIM/ TR_Z_DIM = the "virtual dimensions" of the object along X, Y and Z axis.

4. In the fourth step a XML file, containing all the precedent information, is generated.

5. The aforementioned XML file is read and a .csv file, the format that the graphical software understands, is created. The file has the following format:

   LABEL;NAME;OBJECT_PATH;OBJECT_NAME;LM_X-DIM;LM_Y-DIM;LM_Z-DIM;TR_X-DIM;TR_Y-DIM;TR_Z-DIM

   where:

   - *LABEL:* tr (=trajector) or lm(=landmark);
   - *NAME:* the word corresponding to a trajector or a landmark in the sentence;
   - *OBJECT_PATH:* the 3D dataset path;
   - *LM_X-DIM/ LM_Y-DIM/ LM_Z-DIM:* the dimensions of the landmark along X/Y/Z axes;
   - *TR_X-DIM/ TR_Y-DIM/ TR_Z-DIM:* the dimensions of the trajector along X/Y/Z axes.

6. The sentence, chosen by the user at the beginning, is saved in a text file.

7. After all these steps, a second program, written in Blender's script window, automatically starts. It contains all the instructions required by Blender to place the three-dimensional objects into the virtual environment, indeed, according to the above-mentioned spatial indicator rules (§ 6.4.3.3 Spatial Indicators Definition), the software allows to move the trajector(s) or the landmark(s) in the virtual space.

8. The final result is the three-dimensional reconstruction of the textual spatial description contained in the selected sentence.

6.6 Visual Results

The system has been tested with about 30 sentences and we can consider the results widely acceptable, in fact, it is able to create simple 3D scenes which correspond to the input descriptive sentences.

Below, some examples of automatic scene reconstruction are shown:

Ex. 1. Text: *The Sacrarium is in the centre of a forest of columns.*

Ex.2.	Text: *A globe, in a scale 1 = 1000000 with the planetarium inside, is situated in front of the museum*

Ex.3.   Text: *The hall of modern times is a small building located behind the Museum connected to it through a gallery*

Python Shell

File Edit Shell Debug Options Windows Help

Python 2.7   [MSC v.1500 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
>>>
>>> |

Read the following sentence:

14) The hall of modern times is a small building located behind the Museum connected to it through a gallery

L E T   U S   A N A L Y S E   I T ‼

The TRAJECTOR in the sentence is:

HALL_OF_MODERN_TIMES

While . . .**The LANDMARK in the sentence is:**

MUSEUM

W E L L   D O N E ‼

Now SEE THE **3D SCENE REPRESENTATION** ON BLENDER !!!

Once started click on Run Script.

Ex.4.	Text: *The architect plans a composition where the Library is in front of the University*

Ex.5.	Text:	*Le Corbusier is pointing out the idea of building a platform on a forest of columns.*

Read the following sentence:

01) Le Corbusier is pointing out the idea of building a platform on a forest of columns

L E T   U S   A N A L Y S E   I T ‼

The TRAJECTOR in the sentence is:

PLATFORM

While . . .The LANDMARK in the sentence is:

FOREST_OF_COLUMNS

W E L L   D O N E ‼

Now SEE THE **3D SCENE REPRESENTATION** ON BLENDER !!!

Once started click on Run Script.
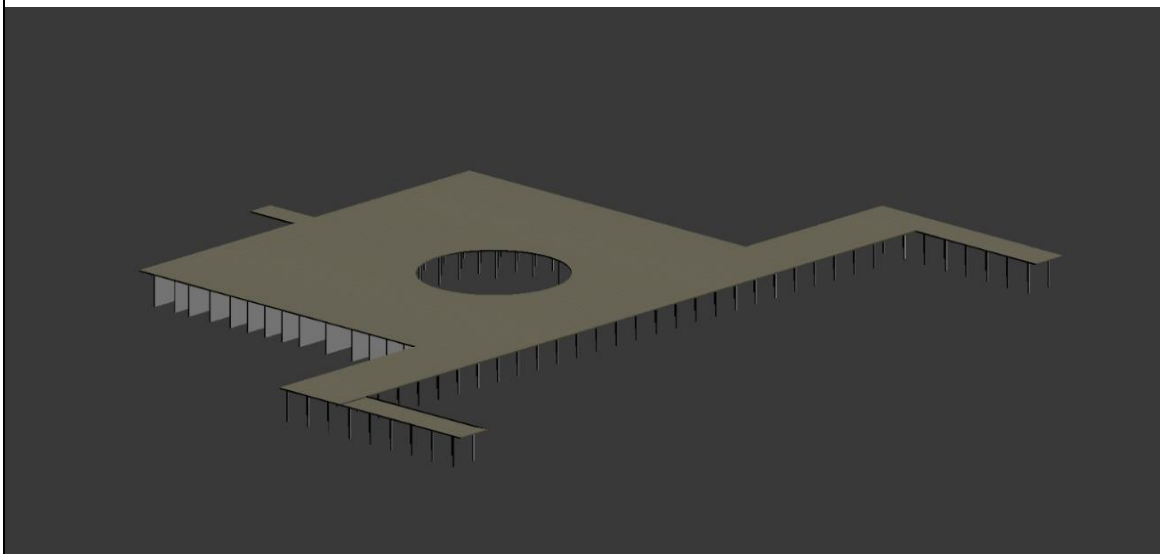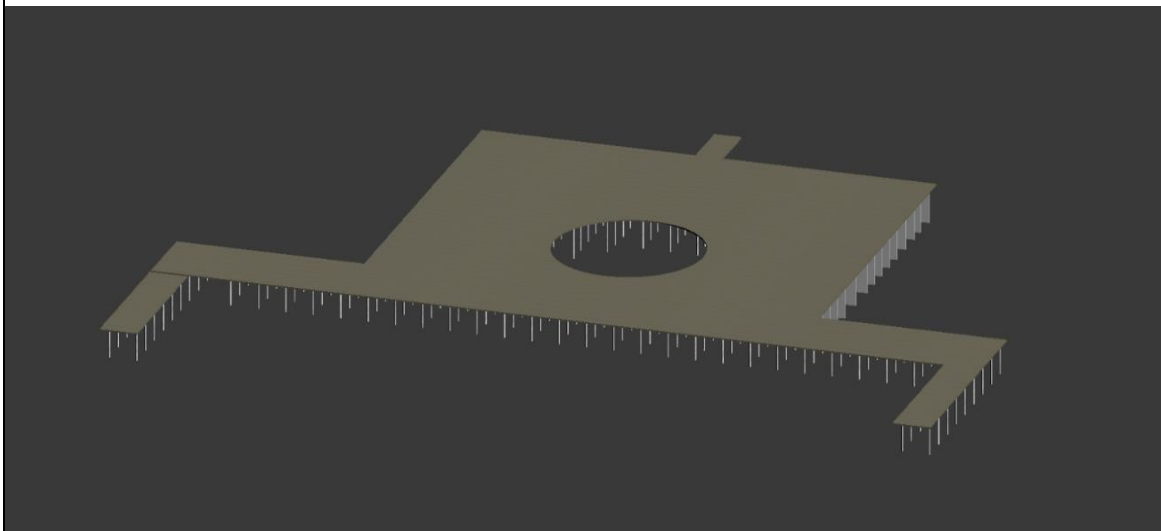
Ex.6. Text: *Four cruciform columns,70 metres tall, are located in the centre of the Sacrarium.*

Ex.7.	Text: *Two symmetrical ramps, strictly linked to the platform, are located in the middle of the Mundaneum main square.*

Ex.8. Text: *The Museum would be included in the enclosure of the Mundaneum*

Ex.9.	Text:	*The Museum rises between the International Associations and the University, quite firmly aligned*

Ex.10. Text: *The Pavilions, which accommodates international exhibitions of various human cultural activities, is a Mundaneum's building located on the left of the Library*

6.6.1 Evaluation of the Visual Results

The evaluation of a language visualization system could be done through two different techniques: the subjective technique and the objective one. The difference between them is that the subjective technique, unlike the other, necessitates the participation of people, more or less expert in the sector, to express a judgment about the output. On the contrary, the objective technique can be based on comparisons or on graphic evaluating parameters. It can be divided into two additional categories which are the diagnostic evaluation and the adequacy evaluation [120]. The diagnostic evaluation aims to estimate the performance of the system based on the initial inputs, given by the user, and it requires a large amount of data to understand the system's coverage and errors. On the contrary, the task of adequacy evaluation is to estimate the fitness of the system with respect to a given objective.

As easily understood, the evaluation of the results is a very complex procedure that requires a lot of time and effort so, in this chapter, we provide simplified examples of subjective and objective evaluation techniques, useful to provide an evaluation, but still not sufficiently detailed and exhaustive.

In particular, we designed a subjective evaluation based on a questionnaire experiment, and an objective evaluation, founded on a comparison between the original position of the buildings of the World City, the one designed by Le Corbusier, and the relative position assigned to them by the system inside the virtual space.

In addition, we also provided a relative evaluation by comparing our system outputs with one of the most common and used system: WordsEyes [99].

6.6.1.1 Subjective Evaluation: Evaluation Questionnaire

It was decided to evaluate the system through a questionnaire that we administered to three totally different groups of participants:

- students from three classes of a middle school: 34 people from 11 to 13 years;
- students of the civil engineering and architecture department: 31 people from 19 to 26 years;
- workers of an IT company: 25 people from 25 to 40 years.

The sample of participants was selected in order to engage competent people in the architectural field and in the IT sector, but also schoolchildren who are approaching the world of information technology and cultural heritage.

The questionnaire, designed to estimate the quality of our system, contains a list of 10 pairs of images to be compared. The two images are: a bi-dimensional technical drawing - one of the original plans of Le Corbusier (or other graphical sources), and a three-dimensional representation - one of the scenes generated automatically by the system. Near each image, five evaluations are reported (1. Very Bad, 2. Acceptable, 3. Average, 4. Satisfactory, 5. Excellent) and the participants are called to express a personal judgment based on how much the 3D representation satisfies them and how much, according to their point of view, it corresponds to the textual description provided as input.

The objective of the questionnaire (appendix C), structured in this way, is to analyse, thanks to a direct comparison between the two images, if and how much a three-dimensional representation is more communicative and understandable than a simple two-dimensional drawing.

The results of the questionnaire are shown in the following table (tab. 10):

Sentence	Participants#	2D image score	Abstentions#	3D image score	Abstentions#
a)	90	197/450	0	326/450	0
b)	90	266/450	12	239/450	4
c)	90	243/450	4	391/450	0
d)	90	202/450	8	237/450	6
e)	90	259/450	2	281/450	0
f)	90	181/450	17	383/450	0
g)	90	254/450	2	397/450	0
h)	90	367/450	3	394/450	0
i)	90	247/450	5	297/450	0
j)	90	351/450	2	369/450	2
Tot.		**2567**/4500		**3314**/4500	

Table 10: Questionnaire results.

The outcome shows that three-dimensional representations are more easily understood by users, in fact, they recorded a level of appreciation equal to *74.64%,* in contrast to 2D drawings that obtained *57.04%.* More specifically, the two-dimensional images obtained greater evaluations from the II group of participants, that is, from the Building and Architectural Engineering students and lower ratings from the remaining two groups. The reason is certainly the fact that the students of Building and Architectural Engineering are probably the only ones able to read and interpret, in the correct way, a technical design

such as a plant, a prospect or a section. However, at the same time, in some sentences there are terms not specifically of architectural or common use (e.g. *Sacrarium, Planetarium, Cité Hôtelière*, etc.) which may have caused some misunderstandings. The evaluations of the 3D representations, on the other hand, did not vary greatly according to the group of voters but, on average, they were very similar. This aspect highlights that the three-dimensional data has a greater communicative and informative power because it does not depend on users' skills and age.

The table also shows the number of abstainers, that is, those participants who decided not to assign any evaluation to some images. In the notes, they generally wrote that they did not understand the representation.

6.6.1.2 Objective Evaluation: Position Discrepancy

The three-dimensional representations, provided by the system, reproduce, in an indicative way, "portions of space" of the World City project. These virtual scenes give an idea of the buildings and the architectural elements that compose the city and show their spatial relationships but, at the same time, they do not illustrate the correct reciprocal position of the entities or their right orientation in the space. This is because the textual descriptions do not provide distal or directional indications but only topological ones (§ 3.6.2 Qualitative Spatial Representation). An addition of descriptive details in the initial sentences could solve the problem, but this was not done because it would lead to a falsification of the original sentences extracted from World City's historical sources. For this reason, we can consider the three-dimensional representation a satisfactory output able to show the characteristics of the complex structures that compose the entire project and to allow users to analyse them by exploring the virtual space.

However, in the following, comparisons between some system graphical outputs and the original project have been carried out. In particular, we overlapped some top views, extrapolated from the three-dimensional scenes generated by the system, to the plan drawn by Le Corbusier to show the divergences between the two representations.

Obviously, to estimate the differences in positions of the objects we overlaid the two images, in the XY plane, taking as a common reference point the position of the landmark. Then, the divergences are evaluated with respect to trajector's position which is the element that moves into the scene contrary to the landmark that stays still.

It was decided to evaluate the error from a graphical point of view because it is the most effective way to show the imprecisions of the system. In particular, in the images below, the correct original position of the buildings (or architectural elements) is highlighted with a broken line around their profiles, while, the abovementioned divergence is shown through an arrow that runs from the virtual model's location, seen from the top view, to the buildings' right position, shown in the plan of Le Corbusier.

Text: *A globe [tr], in a scale 1 = 1000000 with the planetarium inside, is situated in front of the Museum [lm].*
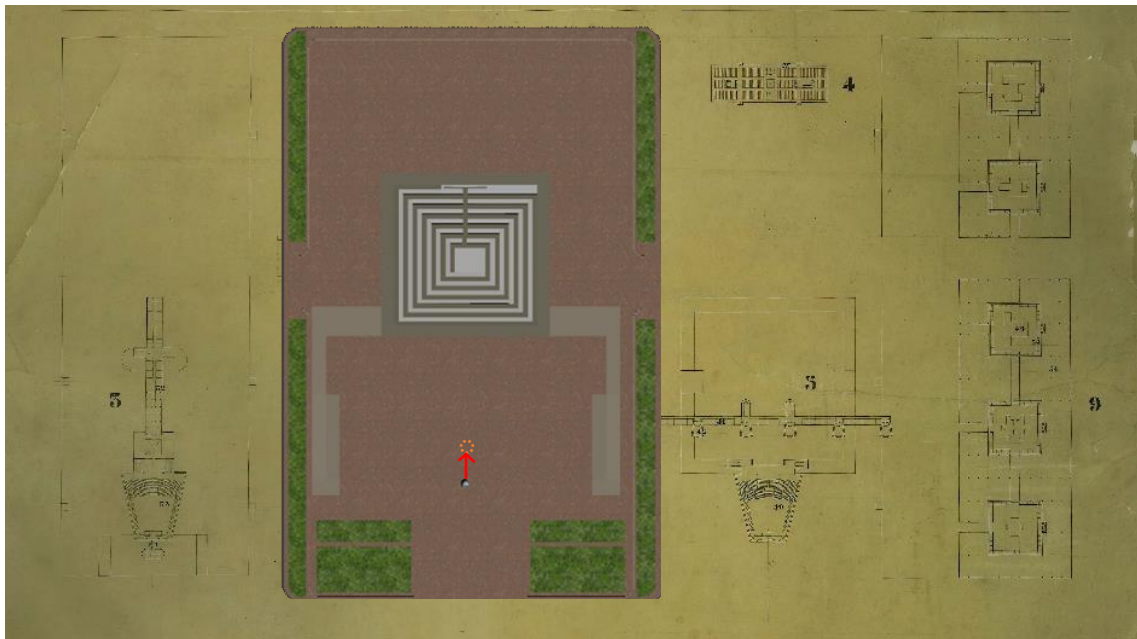
Text: *The Museum [tr] rises between the International Associations [lm] and the University [lm], quite firmly aligned.*
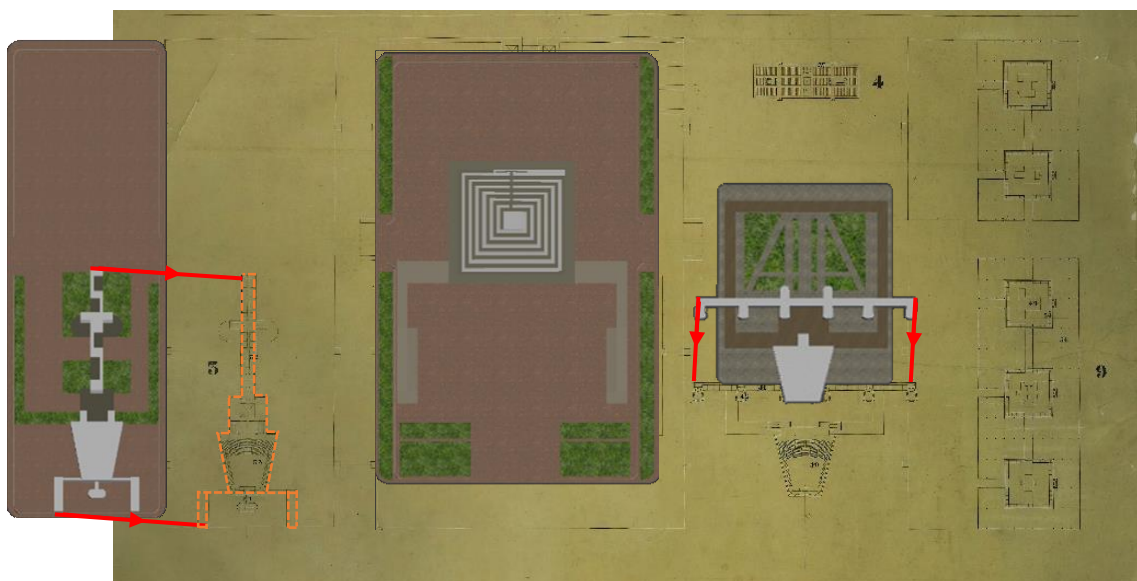
Text: *The Museum [tr] would be included in the enclosure [lm] of the Mundaneum.*
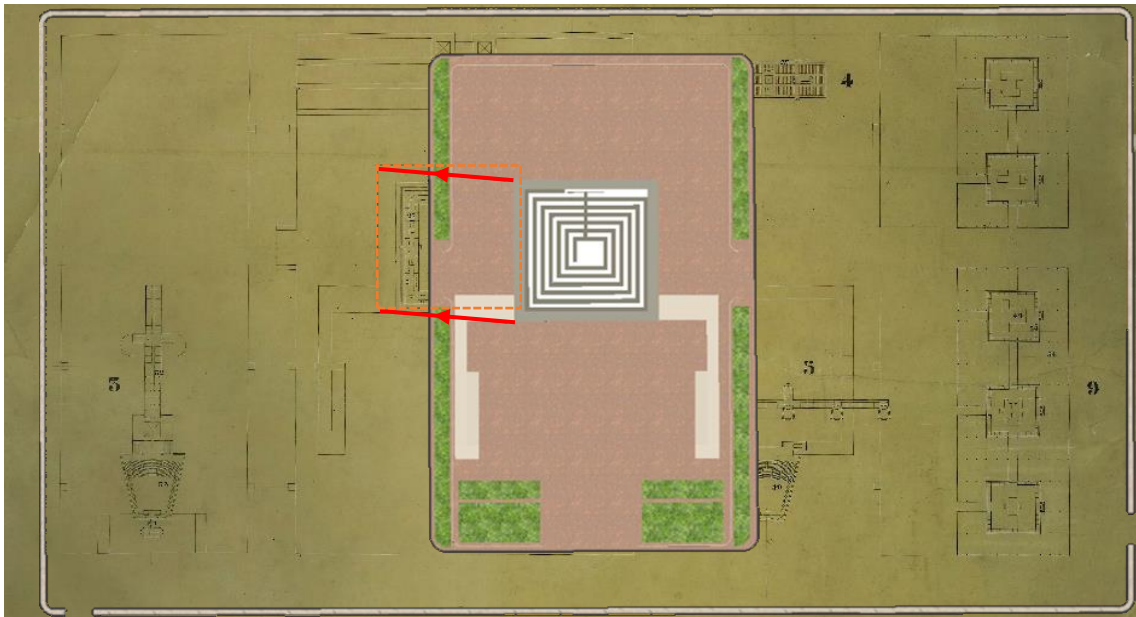
Text: *The architect plans a composition where the Library [tr] is in front of the University [lm].*
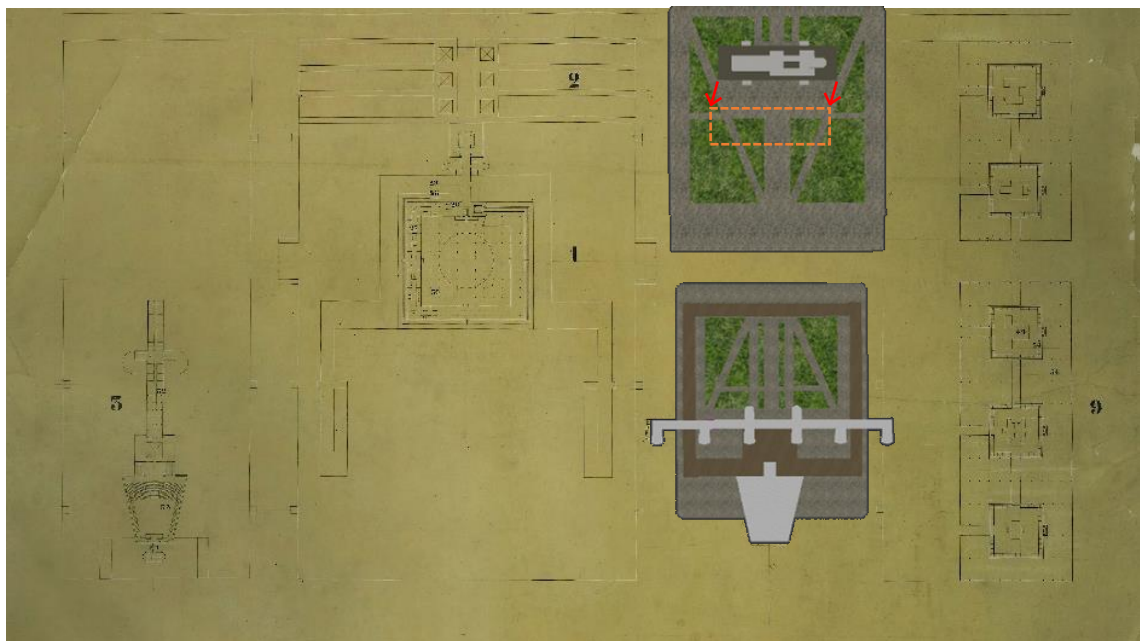
Text: *The Pavilions, which accommodate international exhibitions of various human cultural activities, is a Mundaneum building located on the right of the Library.*
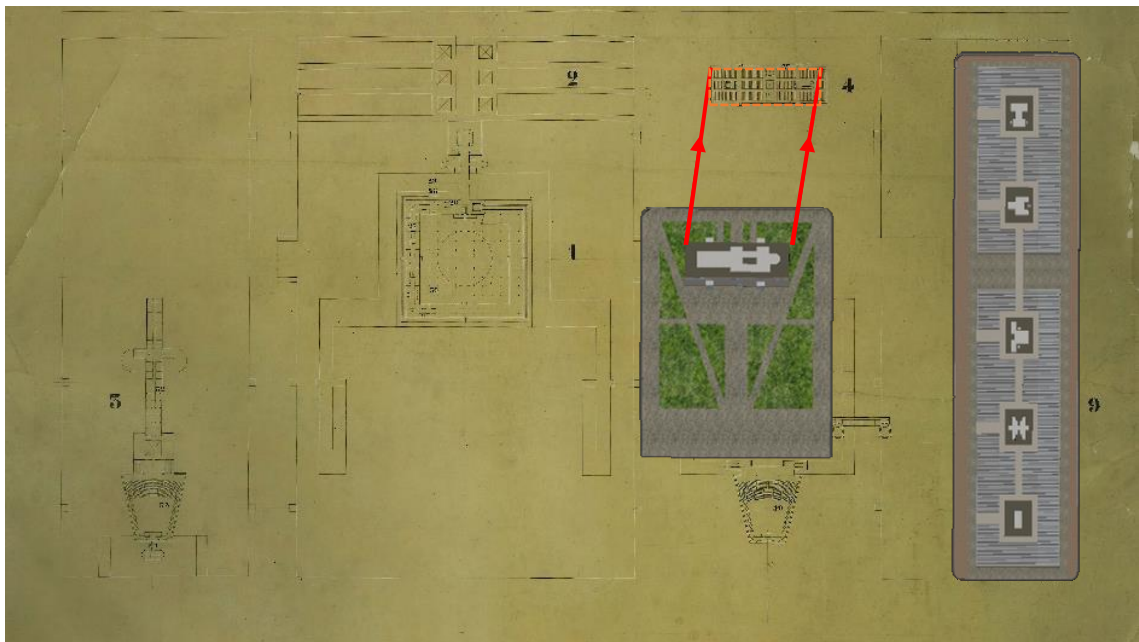
Text: *Le Corbusier located a Stadium [tr] before the enclosure [lm], to be the centre of physical culture.*
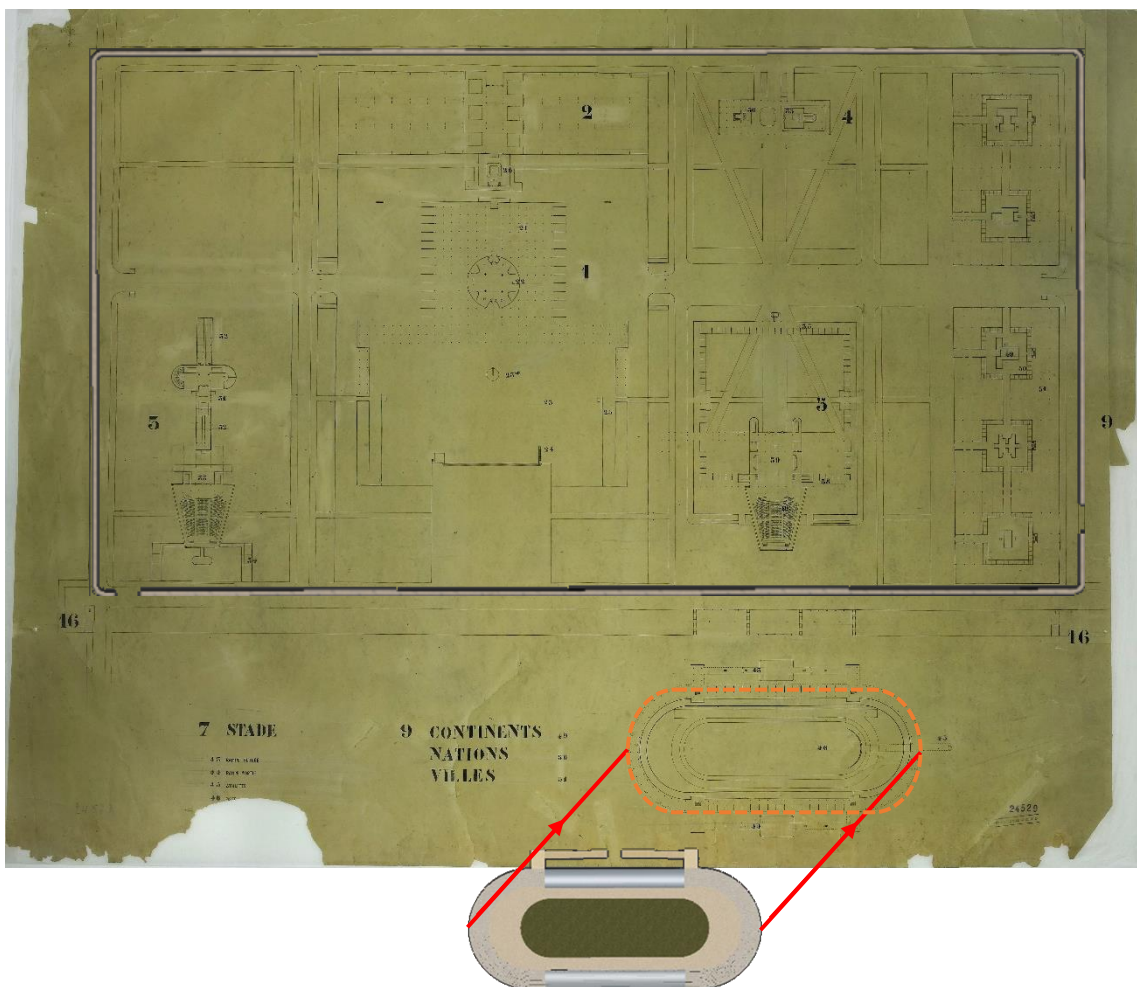
6.6.2. Comparison with WordsEye outputs

In order to demonstrate the utility of our system in the World City domain, we randomly selected two of the sentences previously analysed and we tried to generate the corresponding virtual scene by using *WordsEye* tool. We decided to use it for two main reasons:

- It is the most popular text-to-3D-scene generation system;
- It is an on-line system with a simple and intuitive graphical interface.

The comparison is useful to see how a powerful and well developed tool, designed for a generic domain, is able to reconstruct scenes concerning the World City project. What has emerged is that *WordsEye* system generates scenes only if in the input sentences are present generic objects (such us column, tree, man, etc.), not closely connected with Otlet's project. Indeed, the 3D database does not include three-dimensional elements which refer to the World City like, for example, "the Hall of Modern Times", "the Pavilions" or "the Sacrarium". On the contrary, the system is capable of reconstructing three-dimensional scenes starting from simple sentences containing words like "Stadium", "Pyramid" or "University". However, also in these cases, the 3D models refer to a generic architectural domain and not to the subject of study. It is clear that this type of output cannot satisfy the aim of this thesis work because it would not provide a 3D representation of the World City project and it would not allow and simplify the discovery of its main characteristics. In the following some examples are shown to clarify the abovementioned concepts:

Example 1:

We tested *WordsEye* with the sentence "*The Sacrarium is in the centre of a forest of columns*" (fig.65) and the two principal problems detected are the following:

- The word "Sacrarium" has not a corresponding three-dimensional element in the 3D database;
- It exchanged the spatial indicator "in the centre of" with a 3D object, in fact, it tries to locate in the scene a three-dimensional model called "centre".
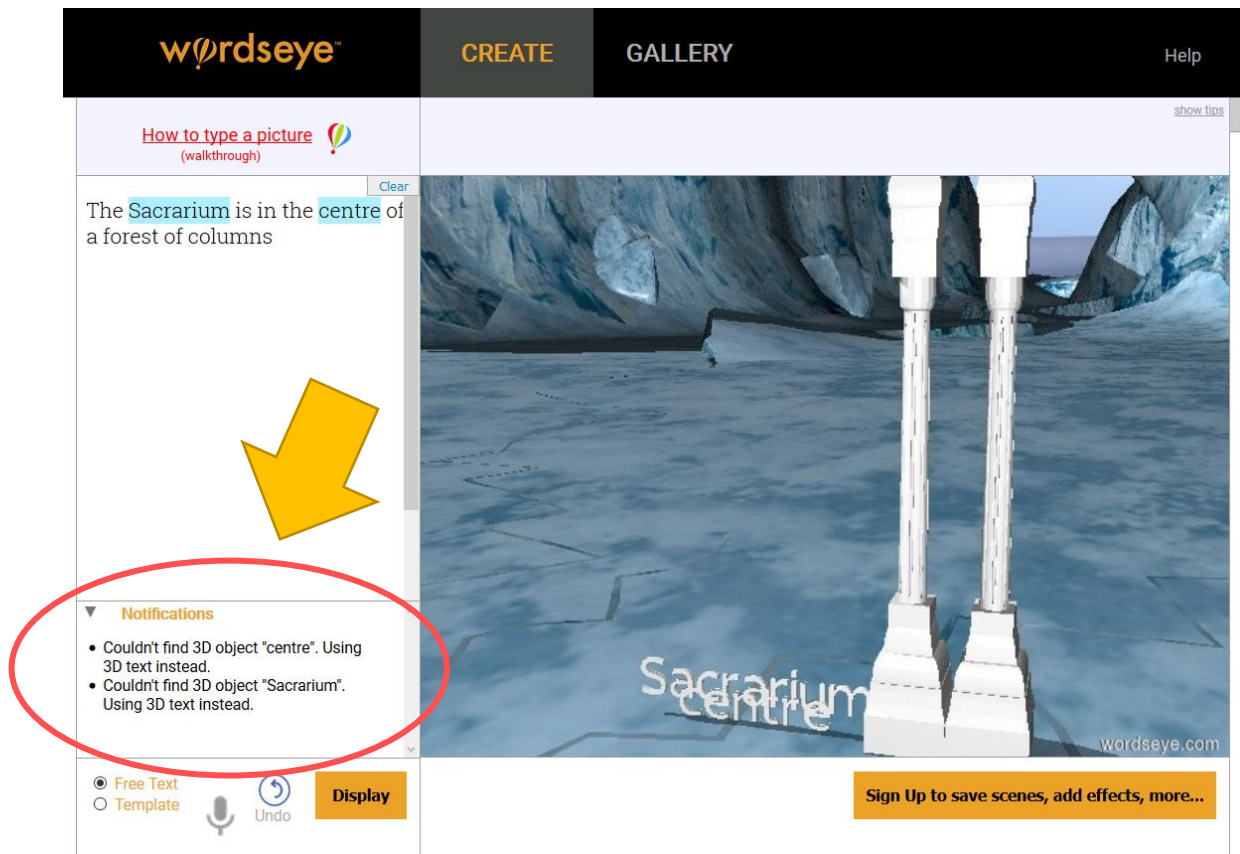
Figure 65: WordsEye 3D representation of the sentence "The Sacrarium is in the centre of a forest of columns".

Example 2a:

The image shows that *WordsEye* cannot understand the input sentence "*The architect plans a composition where the Library is in front of the University*", so, it does not reconstruct anything (fig. 66).
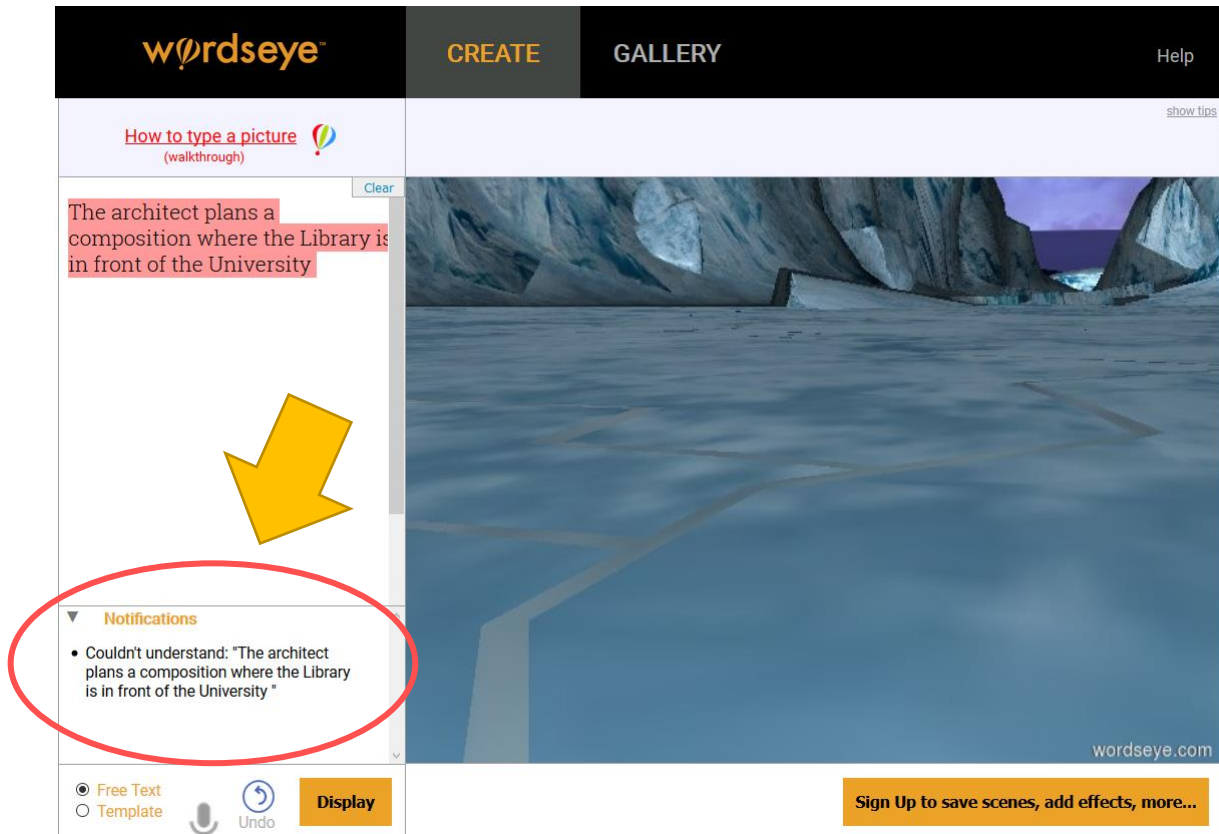
Figure 66: WordsEye 3D representation of the sentence "The architect plans a composition where the Library is in front of the University".

Example 2b:

We tried with a simpler sentence: "*The Library is in front of the University*". In this case the system generates a 3D scene but the three-dimensional objects that it locates in the virtual space do not correspond to the World City buildings, but refers to a general domain (fig. 67).
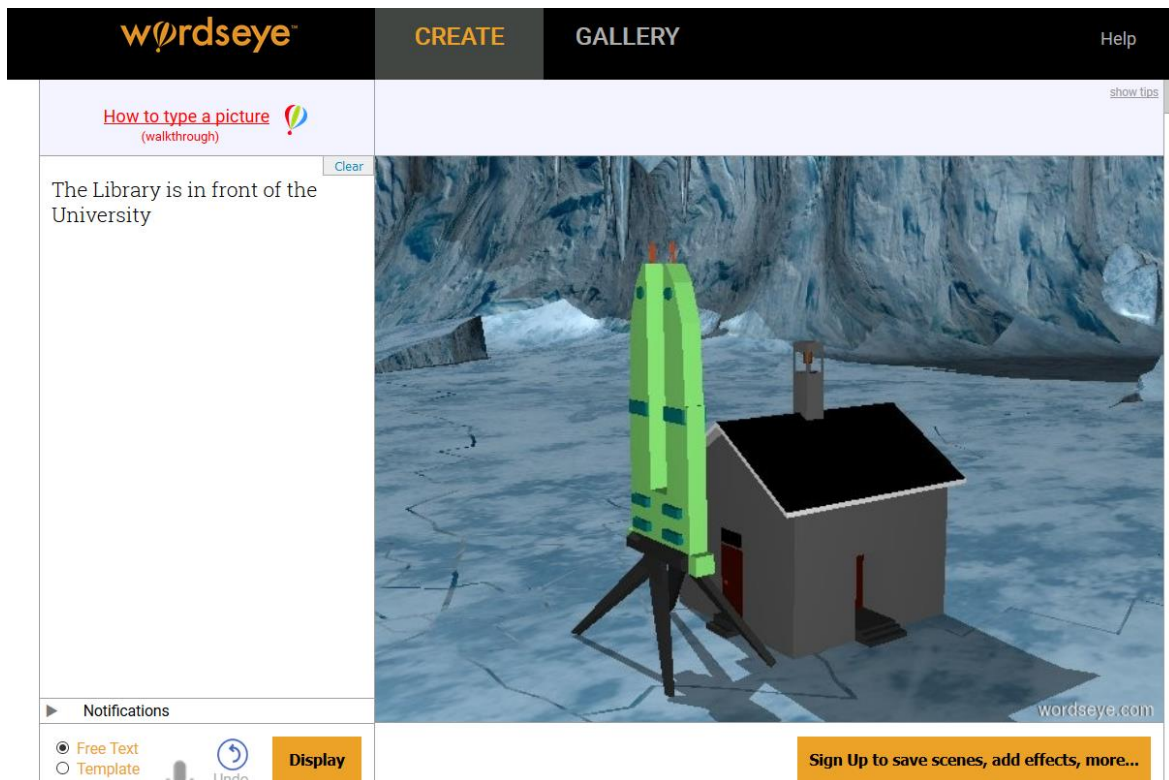
Figure 67: WordsEye 3D representation of the sentence "The Library is in front of the University.

In this case, 3D objects are known to both frameworks with the difference that, since Wordseye operates in a generic domain and has no knowledge of the World City, it is not able to recreate 3D scenes related to the project. On the contrary, it produces a virtual representation in which two objects appear, which are recognized and tagged by the tool as "Library" and "University", but which have nothing to do with the buildings designed by Le Corbusier for the World City.

Moreover, the hypothesis of inserting in WordsEye's 3D database such models would fail for two distinct reasons. In particular, because:

- a 3D modeler should know, model and insert these three-dimensional models into WordsEye's 3D library;

- the tool, in front of the words "Library" and "University", present in a sentence provided as input by the user, would not be able to establish the element to be inserted in the virtual scene, if the generic model or the one referred to the World City.

Example 3a:

Like in the previous example (Example 2a), the original sentence "*Two symmetrical ramps, strictly linked to the platform, are located in the middle of the Mundaneum main square*", which may be too complex, has not provided a three-dimensional virtual reconstruction because the system cannot find neither the trajector/landmark nor the spatial indicator (fig. 68).
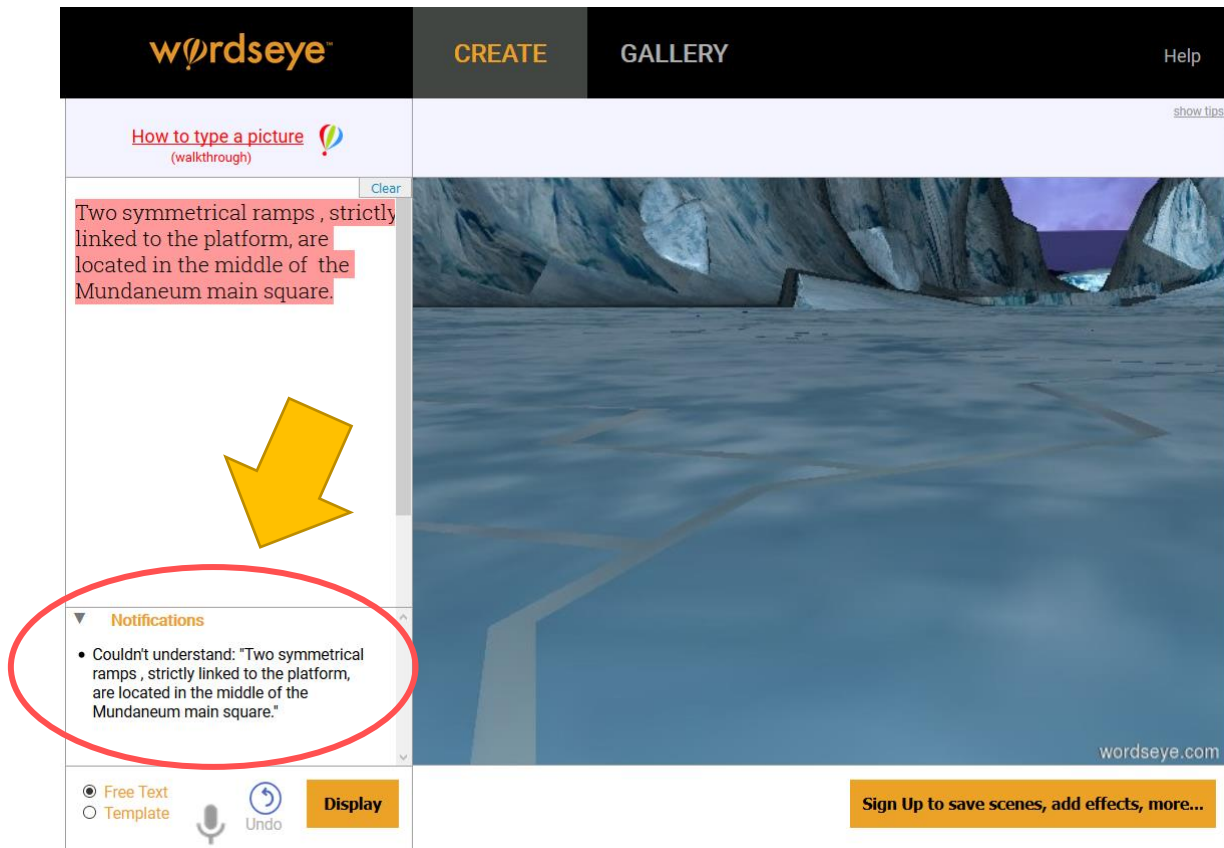
Figure 68: WordsEye 3D representation of the sentence "Two symmetrical ramps, strictly linked to the platform, are located in the middle of the Mundaneum main square".

Example 3b:

We again modified the input sentence in order to simplify it and to obtain a plausible 3D scene. However, the system has not provided a satisfactory output because it is not able to detect neither the landmark "square", nor the spatial indicator "in the middle of". Moreover, it showed two skating ramps, one aligned to the other, which are not related to the architectural field but to the sports one (fig. 69).
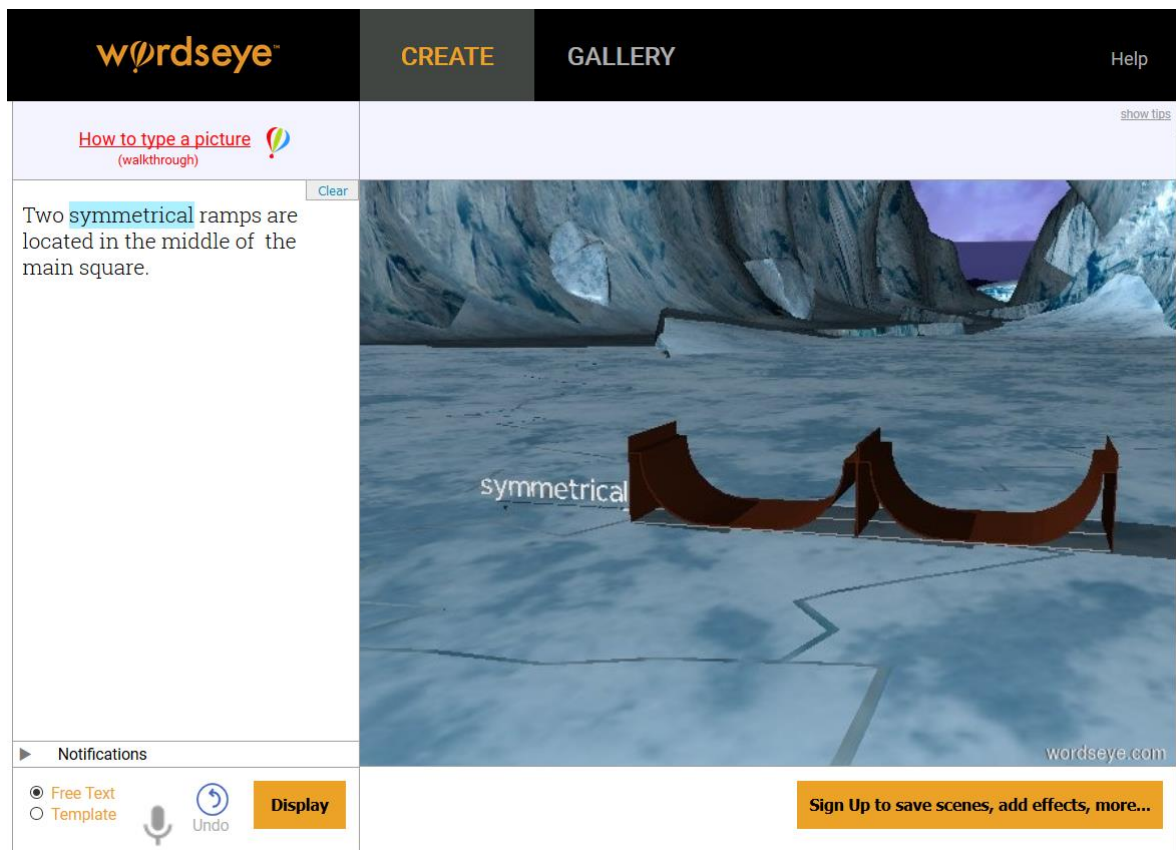
Figure 69: WordsEye 3D representation of the sentence "Two symmetrical ramps are located in the middle of the Mundaneum main square".

In addition to the previous considerations, it is important to underline that *WordsEye* outputs are extremely simple and unrealistic. Furthermore, the system allows to explore the virtual space through rotations, translations and zoom operations but, at the same time, the response times are very slow and, sometimes, the system even locks. This means that when it is capable of showing a 3D scene the visualization is not smooth and this determines an unpleasant virtual exploration process.

6.7 Limitations and Highlights of the System

Although the system shows satisfactory results, there are still many challenges regarding the creation of detailed and complex scenes. Some of these limits are listed below and, in the future, we aim to resolve them by improving the functioning of the system.

6.7.1 Limitations

A case of failure of our system is the inability to break down a complex object into components, so, sentences like "There is a coat on the back of the chair" could not be handled. In this case, in fact, the system would place the coat on the seat of the chair and not in the correct position. The goal is that, in the future, the system will be able to separately manage the different parts of the virtual objects in order to avoid this type of error.

A limitation of the visual system (Blender) is that it is not able to handle the co-reference resolution. This means that phrases like "There is a Museum with a globe in front of it" cannot be transformed into virtual scenes because it is not capable of associating a three-dimensional object to the word "it", even if Saul system is able to extract the spatial relationship.

The system makes an evident imprecision in the presence of the "outside" space indicator. This preposition, in fact, does not specify in which direction the trajector must move with reference to the landmark: if it must advance, go backwards, or if it has to move to the right or to the left. This lack of information, in most cases, causes an error in the production of the 3D scene. An example is shown in the figures below (fig. 70,71).
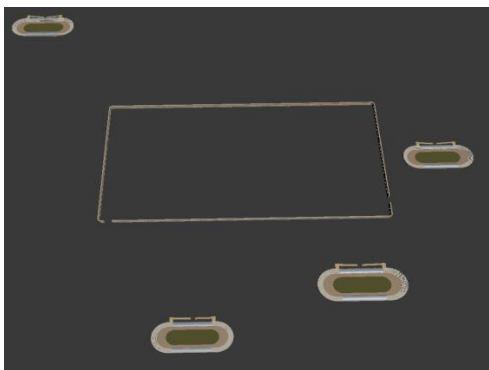
Figure 70: Error in "outside" usage.
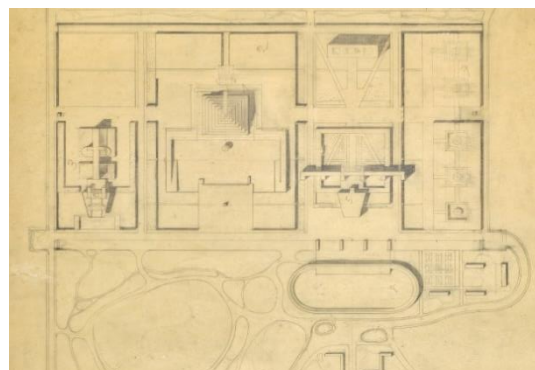
Figure 71: Le Corbusier's plan.

The code was run four consecutive times and, although the system in all four cases placed the trajector out of the enclosure, only at the fourth try did it place the stadium in a position that is close to that designed by Le Corbusier in the project. Indeed, in the first three attempts, the discrepancy between the right position of the stadium (established according to the drawings) and the actual location, provided by the system, is excessive. This inaccuracy could not be considered an error in the absence of a reference architectural project.

The same problem exists for the spatial indicator "near" because, also in this case, the system does not know where to move the trajector, if on the right or left of the landmark.

6.7.2 Highlights

In addition to the limitations that, as mentioned, we will try to solve, the system has points in its favour; we list the main ones below:

One of the central difficulties of "text-to 3Dscene generation" systems, given a specific setting, is to be able to extract the right model from the 3D database and place it in the virtual space. The database, in fact, can contain several three-dimensional objects with the same function, but with different characteristics. This aspect could mislead the system because it could extrapolate an object not suitable for the scene to be generated. For example, in [73], the authors show the difficulty of contextualization of their system with the following example: "*There is a bottle of wine on the table in the kitchen*" (fig. 72).

Figure 72: Wrong 3D scene generated by another system.

As is easily understood, the selected table is inappropriate for a kitchen, moreover, the bottle selected by the system is not a bottle of wine but a bottle of water.

Our system does not have this kind of problem because, not only do all the objects in the database refer to the sole domain of the World City, but they are also distinguished by a unique name that does not cause an incorrect interpretation of the system during the construction of the scene.

A well-known challenge for a "text-to-3D scene generation" system is to disambiguate the "on" preposition. In the English language, in fact, this preposition is generally used to describe a spatial relationship in which an object, the trajector, overlaps with another object, the landmark. However, in some cases, when the landmark surface is not horizontal, but vertical, the "on" preposition it is used to indicate a contact between the vertical surfaces of the two objects (trajector and landmark). To better understand let us show two examples:

- " [The statue]$_{tr}$ is located [*on*]$_{si}$ [the large basement]$_{lm}$ "
- " [The picture]$_{tr}$ is [*on*]$_{si}$ [the large wall] $_{lm}$ of the Museum ".

In the first case, the "on" preposition indicates a supported by relation, while, in the second case, there is an adherence between the vertical surfaces of the two objects.

We can assert, that the system, presented in this work, is able to manage this kind of disambiguation, hence, correctly interpreting the "on" preposition according to the geometric characteristics of the landmark presents in the scene.

An additional ability of the system is to manage the geometric and dimensional characteristics of the three-dimensional objects. Indeed, it analyses the adjectives, present in the textual descriptions that precede a trajector and/or a landmark. For example, if the name of an object (trajector or landmark) follows the adjective "tall", the system automatically changes the z-dimension of the three-dimensional model and locates it in the virtual scene. Similarly, by reading the adjective "green", it changes the colour of the 3D object and places it in the virtual environment. These functions may be particularly useful to analyse and visualise the changes that the World City project underwent during the design phase. In fact, some textual descriptions, as well as some of Le Corbusier's drawings, show variations in the artefacts in terms of size, shape, materials and so forth.

7. The World City on the Web

7.1 Introduction

The historical and cultural importance of the World City project is undoubted, however, it is little known due to the lack of diffusion of the primary and secondary sources concerning it. As mentioned before, the archives of the Fondation Le Corbusier (Paris) and those of the Mundaneum (Mons) scrupulously preserve numerous written documents and drawings, which are added to documents or books stored in libraries or disseminated on the web. The dispersion of these documents can hinder a thorough knowledge of the project or, even worse, completely avoid its knowledge, therefore, this thesis work aims to group several sources already present, to analyse, order and rework them in order to provide a global vision of the project and to establish a connection between them. In particular, it was decided to provide a three-dimensional representation of the World City, aware of all the information that a 3D model can provide the user.

7.2 Web Page Creation

Based on the above, it was decided to create a web page, written in HTML, CSS and Javascript languages, able to illustrate the project to the users and to stimulate their curiosity. The page shows a brief description of the project, some written sources, several original drawings and a few three-dimensional representations that depict scenes described in the texts and that show, briefly outlining the composition of the World City.

The web page could be consulted for educational purposes or simply for pleasure and the user, according to his interest and knowledge, could explore the contents either superficially or in depth. In particular, once having installed the free software Blender, he could open .blend files in order to navigate in the three-dimensional space of the World City.

It is still a work in progress that can be improved and implemented in the future through the addition of new information and functionalities, but it is certainly a starting point for the

promotion of the project in its many facets, in compliance with the copyright rules regarding, in particular, letters and drawings.

In the following, some screenshots of the Web page are shown.

At the top of the page, a navigation bar allows the user to browse the platform, in fact, it is basically a list of links.

By clicking on an element, the user can reach the corresponding section on the page and satisfy his curiosity. In total, the sections are 7:

1) the *HOME section*, which only contains the name of the project and, as background, a drawing of Le Corbusier (Drawing, aerial perspective on the entire project, ©AFCL 24530) that shows, in a perspective view, the whole project included in the natural context (fig. 73).

Figure 73: HOMEPAGE representation.

2) the *ABOUT* section, in which a concise description of the project is given. This section has been enriched by additional information that further clarifies or specifies some concepts. For example, a tooltip has been used to give the user a brief biography of the bibliographer Paul Otlet who is the creator of the project (fig. 74).

Figure74: ABOUT section representation.

The most important buildings of the World City, instead, are presented to the users through three-dimensional models (fig. 75).

Museum

Figure 75: Museum 3D representation on Sketchfab.

They can be recalled through a simple click on the name of the building; by doing so, a direct link to a Sketchfab web page is established. Sketchfab is a platform shared by millions of people that allows you to publish and search for the best 3D models.

Sketchfab supports several 3D formats, in total 28, therefore, the user can create his own model with his favourite graphical software and upload it on the platform. This is a way to display them in a browser in real-time in order to spread and share the knowledge about the object of interest.

For the sake of coherence, I uploaded my models in the Architecture and Cultural Heritage & History section and I enriched them with a description and some tags. The latter, can help the other users to find them and to know their main characteristics.

3) the *GRAPHICAL SOURCES* section, that stores several original drawings made by Le Corbusier; these drawings regard the general composition of the World City and the characteristics of some of its most important buildings. Starting from a portfolio of images, organised in a grid (fig. 77), the user can select the one he prefers and zoom it. A caption under the image provides a synthetic description of it and its source (fig. 78).

Figure 77: GRAPHIC SOURCES section representation

Map of the museum, library, halls, university, numeration and legend / 1: 1000 / Gelatin print / Canson paper / 0.93 x 1.40 / 24519

Three elevations of facades of the museum with silhouettes / Black pencil, blue chalk / Study layer / 0,85 x 0,90 / 24563

Study drawing, elevation, section and plan of a part of the project with sketches, calculations, numbering and notes / Black and colored pencil / Study Layer / 0.67 x 1.09/ 24579

Figure 78: Zoom of the images with details.

167

4) the *TEXTUAL SOURCES* section, which holds the two main publications of Paul Otlet (publication n. 128 and publication n. 133) along with some documents, written in different languages (English, French and Spanish), which talk about the project of the World City from an architectural, political, historical and economic point of view (fig. 79).
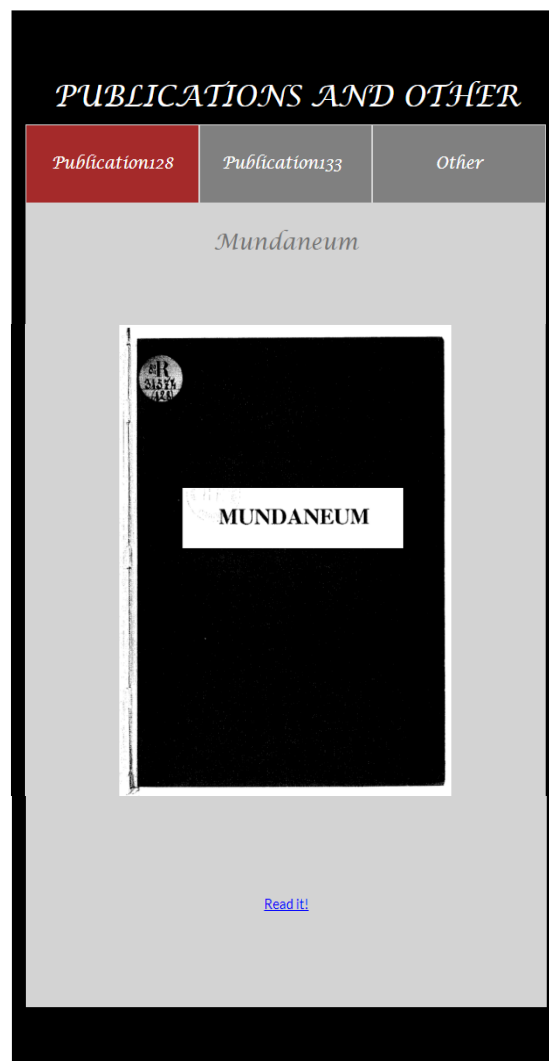
Figure 79: TEXTUAL SOURCES section representation.

5) the *LETTERS* section, that contains extracts of letters, written by Paul Otlet and the architects Le Corbusier and Andersen, during the World City design phases. By positioning the cursor above each letter, a tooltip provides a short text which briefly   contextualises the correspondence (fig. 80).

Figure 80: LETTER section representation.

6) the *3D SCENE* section, which shows some renderings of three-dimensional scenes (fig. 81). The latter are the scenes that the system, described in the sixth chapter of this thesis, has automatically generated starting from some sentences extrapolated from the textual sources. In addition, the user can directly interact with the three-dimensional models by opening a .blend file downloadable from the 3D SCENE section through a click on the descriptive sentences.
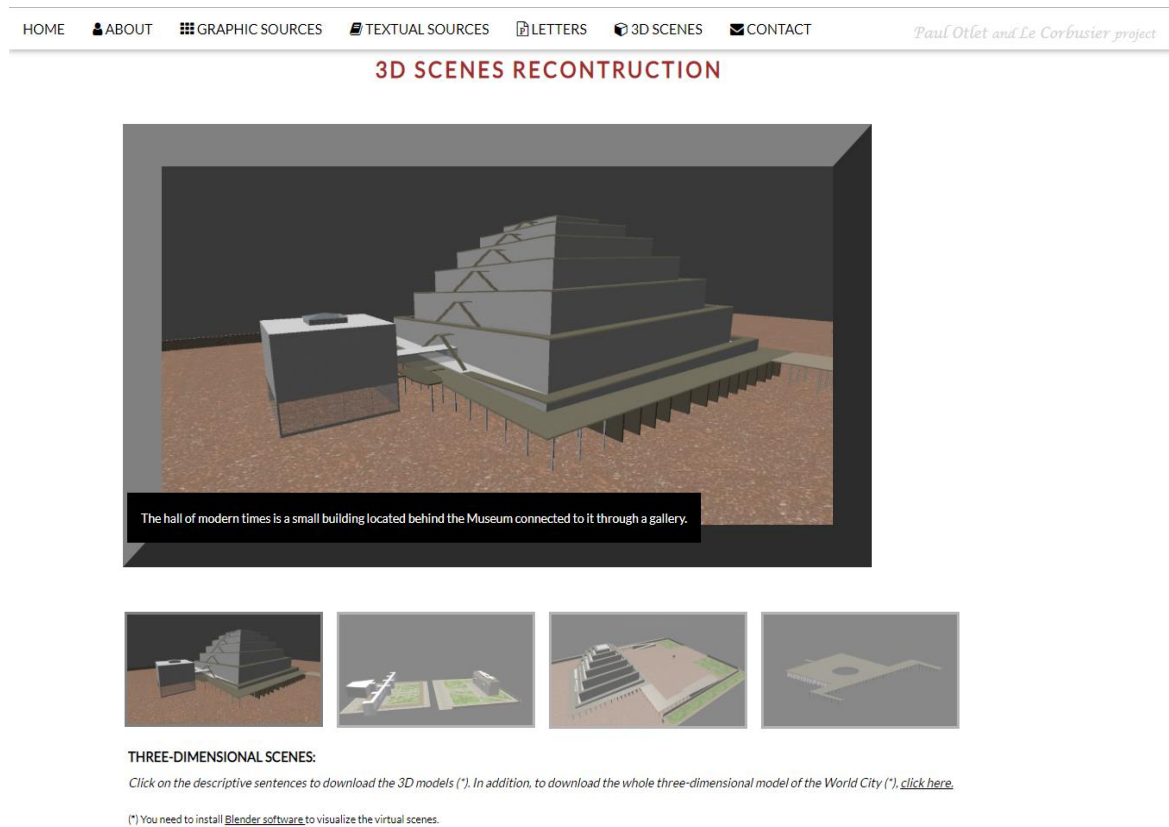
Figure 81: 3D SCENE section representation.

7) the last section is the *CONTACT* one. It contains a direct link to my e-mail for any possible and welcome comments, reflections or requests for additional information (fig. 82).
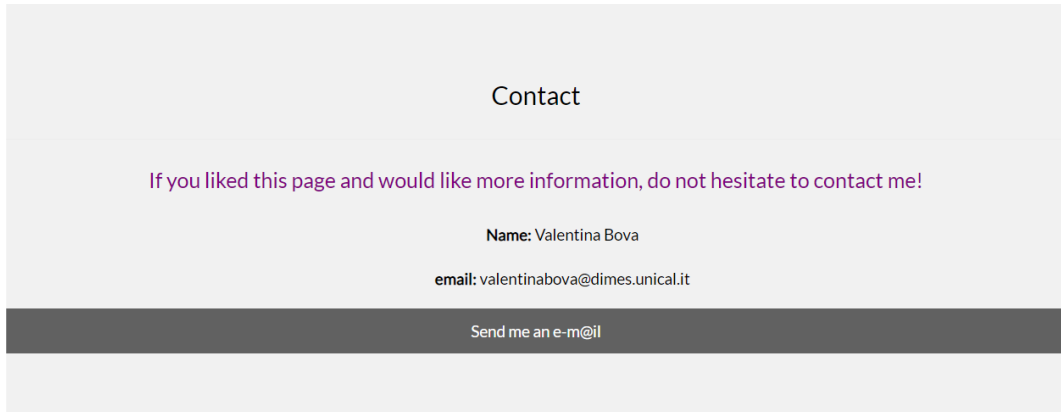
Figure 82: CONTACT section representation.

8. Conclusions

8.1 Summary of the Study

In this thesis work we presented a prototype system designed to generate three-dimensional visualizations of described scenes starting from sentences written in natural language and concerning the "World City" project. The system uses a natural language processing pipeline, which is able to analyse the text from a syntactic and a semantic point of view. In particular, at a semantic level, it provides information about the 3D objects and their relationships and translates semantic annotations into commands in the visualization tool. These rules are important information that the graphical software exploits to define the final location of the objects in the virtual environment. Hence, according to these commands, the three-dimensional elements, starting from the same position, which is the origin of the axis (0,0,0), can move in the space and compose the scene described in the sentence. The system extracts the three-dimensional models from a 3D database which contains several objects previously modelled in accordance with Le Corbusier's technical drawings.

The correct extraction of the 3D models is made possible by a direct connection between a word (or multi-word), present in the text, and a three-dimensional element, stored in the "3D library" and marked by specific keywords. Thanks to them, the system is able to materialize the objects in the virtual space and place them in the right position (fig. 84), as described above.
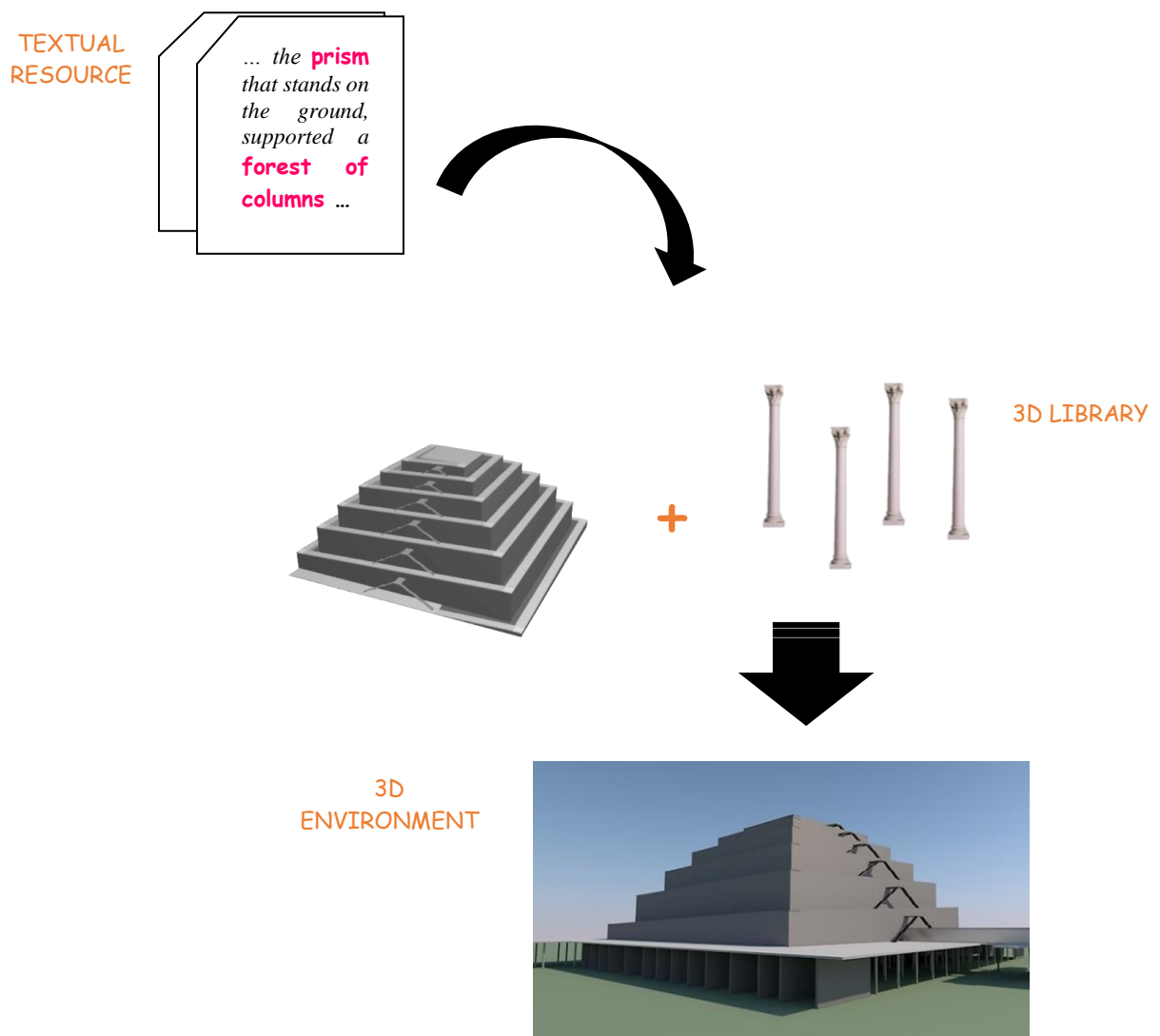
Figure 84: *Association between a textual resource and an element of the 3D library in order to create the three-dimensional scene.*

In this phase, user interaction is desirable, in fact, human collaboration is a good solution to refine the output results and to involve people in the scene generation. In addition, it is a promising way to acquire more implicit spatial knowledge which is useful to improve system performances.

We believe that a text-to-scene generation system provides a stimulating way to interact with intelligent machines and it is addressed to a large range of people who can play, learn and work by building 3D scenes. We tested it on a specific domain, the "World City" project, but the work can be extended in order to cover other different fields. Indeed, the

methodology used is extremely versatile and it could be used in a variety of working, disciplinary and research fields. More specifically, it can be exploited in numerous applications in which it is useful to transfer spatial ideas and configurations into a visual three-dimensional communication. The reason why it was used with reference to the abovementioned project, is based on the awareness that the difficulty of dissemination and enhancement of the beauty of our cultural heritage depends, not only, on the cultural products themselves, but also, and above all, on the strategies used to promote their knowledge.

The case study is a clear example of how the new computer technologies, properly implemented, can, after preliminary studies and analysis, aid and support the dissemination of cultural content.

8.2 Contribution

We consider our work an innovative contribution which adds on to previous research in the field.

At first, we enlarged the Saul software domain by evaluating the learning-based programming language on some domain-specific texts regarding the "World City" project. Indeed, the recognition of objects and their spatial relationships in language (Spatial Role Labelling) is heavily influenced by the application field, and our project is an important first step to integrate perceptual world knowledge in machine understanding of natural language. A second main contribution regards the 3D scene generation. Previous systems, in fact, rely on complex rule-based methods to map object references to concrete 3D objects. On the contrary, our work aims to generate virtual representations of textual descriptions by using a simple approach based on an open source tool and a continuously developing programming language. We illustrated a system able to "give shape and life" to words passing from a mono-dimensional representation (text) to a three-dimensional one (virtual scene) and we do not build the scene in a common way, by drag and drop commands, but we use scripts which automatically select the 3D objects from a database and locate them in the right position in the virtual environment.

8.3 Future Works

The described system is a work that can be further developed in the future because many improvements remain to be done both at the language and graphical levels. In particular, our future task is to increase its knowledge representation and utilization: we aim to make the system able to handle - simultaneously - more linked sentences in order to create complex scenes with many trajectors, landmarks and spatial indicators. Each triplet will be connected by a relation and, in turn, also the relations in the text will be linked together. In this way, the system will be able to infer both explicit and implicit spatial constraints which are not directly specified in the description, but left up to the reader's imagination. Let us consider an example sentence: "*On a square, between the obelisk and the halls of modern times, is located a prism that rests on pillars*". In the text three different relations appear, which are:

- *On (prism; square),* where "prism" is the trajector, "square" is the landmark and "on" is the spatial indicator;
- *Between (prism; obelisk, halls of modern times),* where "prism" is the trajector, "obelisk" and "halls of modern times" are the landmarks and "between" is the spatial indicator;
- *On (prism; pillars),* where "prism" is the trajector, "pillars" are the landmark and "on" is the spatial indicator.

Firstly, our aim is to extract the labels and the three relations and, secondly, to connect them to create a single complete scene. Hence, the system shall understand the hierarchy of the correct elements and generate a square, place some pillars on it and locate the pyramid on them. In addition, it shall place on the right of the "pillar+pyramid" block the obelisk and on its left the halls of modern times.

Another challenge is to improve the system by incorporating more performant algorithms that will be able to create virtual scenarios taking into account the scenes settings and the domains. The purpose is to assign the correct size of the objects and the right distance between them according to their nature and configuration. For instance, the machine should be able to understand the difference between the following two sentences:

"*the chair* [tr]" *is in front of* [si] *the desk* [ln]"   and

"*the sofa* [trajector]" *is in front of* [spatial indicator] *the television* [landmark]"

The latter represent two very different scenarios, in fact, in the first sentence the distance between the two objects is small (it must only allow people to sit comfortably), while, in the second one, it is longer (the seating distance must ensure a good view of the screen and it depends on the television size). This means that, in the calculation of the distance between trajector(s) and landmark(s), the characteristics of the objects (shapes, sizes and functions) are also involved and these must be taken into account by our system.

We also plan to make the machine able to understand numerical distances between entities during the definition of the spatial indicators (§ 5.   Spatial Relationships). In other words, the system should establish the difference between two kinds of relations that are "A is near B" (generic distance) and "A is 2 metres away from B" (precise distance) in order to replace the random functions, used to define the spatial indicators by specific distance assignments between trajector(s) and landmark(s).

A further objective is not to limit the system to static representations but, on the contrary, we aim to work on dynamic relations. This means that the system must be improved from a linguistic and a graphical point of view. At the moment, in fact, CLEF dataset (§ 5.5.1 Experimental Results) contains only 25 spatial indicators annotated as dynamic and the remaining are static. This means that we cannot create an appropriate *training* dataset able to allow the system to the correct automatic determination of the dynamic spatial indicators during the *test* phase. Consequently, the graphic rules for creating moving scenes in the virtual space have not been implemented, in fact, the system only detects the initial and the final locations of the 3D models in the virtual environment. On these bases, a future development of the work could be to enrich the database with numerous dynamic relations in order to allow Saul system to annotate the "animated" spatial indicators correctly with the purpose of showing, in the 3D world, the motion from the start point to the end point. In other words, the future goal is to make the system capable of reconstructing a virtual scene corresponding to the sentence: "Starting from the peristyle the visitor, coming through the entrance, can see the globe in the hall".

This would be a big step forward because the animations are much more explanatory than words or images. Indeed, if a picture can replace, indicatively, a thousand words, the animated scenes could replace a million. In other words, the dynamic representations could provide not only information about the entities' properties and locations, but also temporal information.

Finally, we plan to improve the recognition of the spatial objects and their relations by exploiting bi-dimensional images. Indeed, when humans describe objects and scenarios, very often they use ambiguous and polysemous words, which could make machine understanding of language difficult. This means that relying only on the texts is often insufficient to ensure the generation of a truthful and univocal three-dimensional scene, however we need more specific information which can be extracted from images. It is precisely for that reason that we plan to integrate visual and textual knowledge by mapping images and words in order to assist language tasks and to help the system create an optimal 3D scene.

Appendix A

Sentence_Split

```
import re
import codecs

s_i=open("englishSpatialIndicatorsSearch.txt").read().splitlines()
Frasi1= open('Frasi1.txt', 'w')
text=open('Frasi.txt', 'r')
for line in text:
    NewLine= line.replace(',', ' ,').replace(';', ' ;').replace('«', '« ').replace('»' ,
            ' »').replace(':', ' :').replace(' "', ' " ').replace('" ', ' " ')
    Frasi1.writelines(NewLine)
Frasi1.close()
text1=open('Frasi1.txt', 'r')
infile= open('infile.txt', 'w')
for line in text1:
    Line = line.replace('A','a').replace('B','b').replace('C','c').replace('D','d')

.replace('E','e').replace('F','f').replace('G','g').replace('H','h').replace('J','j')
        .replace('K','k')
.replace('L','l').replace('M','m').replace('N','n').replace('O','o')

.replace('P','p').replace('Q','q').replace('R','r').replace('S','s').replace('T','t')

.replace('U','u').replace('V','v').replace('W','w').replace('X','x').replace('Y','y')
        .replace('Z','z')
    for i in s_i:
        if i in Line:
            infile.writelines(line + '\n')
infile.close()
lines_seen = set()
outfile= open('outfile.txt', 'w')
for line in open('infile.txt', 'r'):
    if line not in lines_seen: # not a duplicate
        outfile.write(line + "\n")
        lines_seen.add(line)
outfile.close()
new_lines = re.compile('\n{2,9}')
with open("outfile.txt") as f:
    contents = f.read()
contents = re.sub(new_lines, '\n\n', contents.strip())
with open("outfile.txt", 'w') as f:
    f.write(contents)
```

Sentence_Split

```
from nltk import sent_tokenize, word_tokenize
import re

testo ="path/text.txt"
infile = open(testo, 'r')
data = infile.read()
pos = ()
for i in range (len(data)):
     if data[i]== '.' :
          pos+= (i, )
frasi= data.split('.')
f = open('Frasi.txt', 'w')
frasi2=[]
for i in frasi:
    if len((i))>2 and (i[1].isupper()==True or i[0].isupper()==True):
         f.writelines(i + " " + "\n")
         frasi2= frasi2 + [i]
f.close()
```

CountOccurrences

```
import string

labels=open("englishNounsS.txt").read().splitlines()
l=len(labels)
labelSentences= open('labelSentences.txt', 'w')
for line in open("outfile.txt", "r"):
    if not line.strip():  # if line is blank ignore it
       continue
    else:
        Line= line.replace('A','a').replace('B','b').replace('C','c').replace('D','d')
            .replace('E','e').replace('F','f').replace('G','g').replace('H','h')
            .replace('J','j').replace('K','k') .replace('L','l').replace('M','m')
            .replace('N','n').replace('O','o').replace('P','p') .replace('Q','q')
            .replace('R','r').replace('S','s').replace('T','t').replace('U','u')
            .replace('V','v').replace('W','w').replace('X','x').replace('Y','y')
            .replace('Z','z')
        count_occurrences=0
        tot=0
        for j in labels:
             count_occurrences=Line.count(j)
             tot=tot+count_occurrences

        number=0
        if tot>1:
             labelSentences.writelines(line + '\n')
labelSentences.close()
```

Last

```python
import codecs
import re

s_i=open("englishSpatialIndicatorsLast.txt").read().splitlines()
labelSentences=open("labelSentences.txt", "r")
labelList=open("englishNounsLast.txt").read().splitlines()
final1= open("final1.txt", "w")
final2= open("final2.txt", "w")
final3= open("final3.txt", "w")

for line in labelSentences:
     parole_=re.findall(r"[\w']+", line)
     parole=[]
     for P in parole_:
         p=P.lower()
         parole=parole+[p]
     print(parole)
     if not line.strip():  # if line is blank ignore it
        continue
     else:
         paroleNew=[]
         for i in range(len(parole)-5):
             if parole[i]=='on' and parole[i+1]=='either' and parole[i+2]=='side' and
                parole[i+3]=='of':
                     parole[ parole.index(parole[i]):parole.index(parole[i+5]) ] =
                     [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+5])
                      ])]
             elif parole[i]=='on' and parole[i+1]=='one' and parole[i+2]=='side' and
                  parole[i+3]=='of':
                     parole[ parole.index(parole[i]):parole.index(parole[i+4])] =
                     [' '.join(parole[parole.index(parole[i]): parole.index(parole[i+4])
                      ])]
             elif parole[i]=='at' and parole[i+1]=='the' and parole[i+2]=='side':
                     parole[ parole.index(parole[i]):parole.index(parole[i+3])] =
                     [' '.join(parole[parole.index(parole[i]) :parole.index
                     (parole[i+3])])]
             elif parole[i]=='the' and parole[i+1]=='summit' and parole[i+2]=='of':
                      parole[ parole.index(parole[i]):parole.index(parole[i+3])]=
                      [' '.join(parole[parole.index(parole[i]):parole.index
                      (parole[i+3])])]
             elif parole[i]=='face' and parole[i+1]=='one' and parole[i+2]=='another':
                      parole[ parole.index(parole[i]):parole.index(parole[i+3]) ] =
                      [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                      ])]
elif parole[i]=='at' and parole[i+1]=='the' and parole[i+2]=='extremities'
     and parole[i+3]=='of':
                      parole[ parole.index(parole[i]):parole.index(parole[i+4])]=
                      [' '.join(parole[parole.index(parole[i]): parole.index(parole[i+4])
                      ])]
             elif parole[i]=='in' and parole[i+1]=='the' and parole[i+2]=='centre':
                     parole [ parole.index(parole[i]):parole.index(parole[i+3]) ]=
                     [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                     ])]
             elif parole[i]=='in' and parole[i+1]=='the' and parole[i+2]=='midst'and
                  parole[i+3]=='of':
                     parole[ parole.index(parole[i]):parole.index(parole[i+4]) ] =
                     [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+4])
                     ])]
             elif parole[i]=='the' and parole[i+1]=='opposite' and parole[i+2]=='side':
                     parole[ parole.index(parole[i]):parole.index(parole[i+3])] =
                     [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                     ])]
             elif parole[i]=='the' and parole[i+1]=='opposite' and parole[i+2]=='sides':
```

```
                    parole[ parole.index(parole[i]):parole.index(parole[i+3])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                        ])]
            elif parole[i]=='side' and parole[i+1]=='by' and parole[i+2]=='side':
                    parole[ parole.index(parole[i]):parole.index(parole[i+3])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                        ])]
            elif parole[i]=='to' and parole[i+1]=='the' and parole[i+2]=='left':
                    parole[ parole.index(parole[i]):parole.index(parole[i+3])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                        ])]
            elif parole[i]=='to' and parole[i+1]=='the' and parole[i+2]=='right':
                    parole[ parole.index(parole[i]):parole.index(parole[i+3])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                        ])]
            elif parole[i]=='to' and parole[i+1]=='the' and parole[i+2]=='back':
                    parole[ parole.index(parole[i]):parole.index(parole[i+3])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+3])
                        ])]
            elif parole[i]=='to' and parole[i+1]=='right' and parole[i+2]=='and' and
                    parole[i+3]=='left':
                    parole[ parole.index(parole[i]):parole.index(parole[i+4])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+4])
                        ])]
            elif parole[i]=='separated' and parole[i+1]=='from' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+(2)])
                        ])]
            elif parole[i]=='spring' and parole[i+1]=='up' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+(2)])
                        ])]
            elif parole[i]=='supported' and parole[i+1]=='by' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+(2)])
                        ])]
            elif parole[i]=='occupied' and parole[i+1]=='by' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+(2)])
                        ])]
            elif parole[i]=='in' and parole[i+1]=='front' and parole[i+2]=='of':
                    parole[ parole.index(parole[i]):parole.index(parole[i+3])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+(3)])
                        ])]
            elif parole[i]=='far' and parole[i+1]=='from' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+(2)])
                        ])]
            elif parole[i]=='surrended' and parole[i+1]=='with' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+2])
                        ])]
            elif parole[i]=='surrounded' and parole[i+1]=='by' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+2])
                        ])]
            elif parole[i]=='bordered' and parole[i+1]=='by' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [' '.join(parole[parole.index(parole[i]):parole.index(parole[i+2])
                        ])]
            elif parole[i]=='symmetrically' and parole[i+1]=='placed' :
                    parole[ parole.index(parole[i]):parole.index(parole[i+2])]=
                        [''.join(parole[parole.index(parole[i]):parole.index(parole[i+2])])]
    paroleNew=paroleNew+parole
    print(paroleNew)
    print('\n')
    for i in range(len(paroleNew)-4):
```

```
                    if paroleNew[i] in s_i:
                          indexS_I=(paroleNew.index(paroleNew[i]))
    if (((paroleNew[indexS_I+1])in labelList) or
        ((paroleNew[indexS_I+2])in labelList) or
        ((paroleNew[indexS_I+3]) in labelList) or
        ((paroleNew[indexS_I+4]) in labelList) or
        (paroleNew[i] == 'in front') or (paroleNew[i] == 'in the centre' and
         paroleNew[indexS_I-1] in labelList) :
                                  final2.writelines(line + '\n')

             for i in range(len(paroleNew)-1):
                    if paroleNew[len(paroleNew)-4] in s_i:
                        if (((paroleNew[len(paroleNew)-3])in labelList) or
                            ((paroleNew[len(paroleNew)-2])in labelList)or
                            ((paroleNew[len(paroleNew)-1])in labelList) ):
                                  final1.writelines(line + '\n')
                      elif paroleNew[len(paroleNew)-2] in s_i or paroleNew[len(paroleNew)-3]:
                          if  ((paroleNew[len(paroleNew)-1])in labelList) :
                                  final1.writelines(line + '\n')
             for i in range(len(paroleNew)-1):
                    if (paroleNew[i] == 'in' and paroleNew[i+1] == 'front')  or (paroleNew[i] ==
                       'surrounding' and paroleNew[i+1] =='it') or (paroleNew[i] == 'to' and
                       paroleNew[i+1] =='right' and paroleNew[i+2] =='and' and paroleNew[i+3]
                       =='left'):
                            final3.writelines(line + '\n')
final1.close()
final2.close()
final3.close()


File = ['final1.txt', 'final2.txt', 'final3.txt']
with open('Path/File.txt', 'w') as outfile:
    for file in File:
        with open(file) as infile:   # file concatenation
            for line in infile:
                 outfile.write(line)


lines_single = set()
Last_= open('Last_.txt', 'w')
for line in open('File.txt', 'r'):
     if line not in lines_single: # not a duplicate
          Last_.write(line + "\n")
          lines_single.add(line)
Last_.close()


Last= open('Last.txt', 'w')
Last2= open('Last2.txt', 'w')
for line in open ('Last_.txt', 'r'):
    words=line.split()
     for i in words:
         if i != ('In') or i!=('in' ) :
              Last2.writelines(line + '\n')
         else:
              posIn= words.index(i)
              if words[posIn+1].isdigit() == False:
                   Last2.writelines(line + '\n')
Last.close()
Last2.close()
l_single = set()
TheLast= open('TheLast.txt', 'w')
for line in open('Last.txt', 'r'):
    if line not in l_single: # not a duplicate
         TheLast.write(line + "\n")
         l_single.add(line)
TheLast.close()
```

Appendix B

```python
import bpy
import csv
import mathutils
import random


file = open('Sentence.txt', 'r')
Sentence = file.read()
print(Sentence)
SentenceWords = Sentence.split()
si_position= SentenceWords.index("[si]")
spatial_indicator= (SentenceWords[si_position-1])

tr_position= SentenceWords.index("[tr]")
tr_adj= (SentenceWords[tr_position-2])

lm_position= SentenceWords.index("[lm]")
lm_adj= (SentenceWords[lm_position-2])



textFile = path\\forBlender.csv'

with open(textFile, 'r') as file:
    reader = csv.reader (file, delimiter=',')
    n=0
    for row in reader:
        if row[0] =='lm' :
            n=n+1
            tot_lm= n
            if tot_lm==1:
                bpy.ops.wm.append(directory= row[2], filename =row[3])

            elif tot_lm==2:
                bpy.ops.wm.append(directory= row[2], filename =row[3])
                bpy.ops.transform.translate(value=(0, (-1.7*(float(row[5])+ float(row[8]))
                                        *random.uniform(0.5,0.6)),0))

        elif row[0]=='tr'and spatial_indicator == "between" :
            bpy.ops.wm.append(directory= row[2], filename =row[3])
            bpy.ops.transform.translate(value=(0, (-random.uniform
                                    (0.4,1.0)*float(row[5])) , 0))
            #tr = bpy.context.active_object
            if tr_adj == 'long':
                bpy.ops.transform.resize(value=(5,1,1))
            elif tr_adj == 'narrow':
                bpy.ops.transform.resize(value=(1,0.5,1))
            elif tr_adj == 'large':
                bpy.ops.transform.resize(value=(1,3,1))
            elif tr_adj == 'narrow':
                bpy.ops.transform.resize(value=(1,0.5,1))
            elif tr_adj == 'short':
                bpy.ops.transform.resize(value=(1,1,0.5))
            elif tr_adj == 'tall':
                bpy.ops.transform.resize(value=(1,1,3))



        elif row[0]=='tr' and (spatial_indicator== "on" ):
            bpy.ops.wm.append(directory = row[2], filename = row[3])
```

```python
        bpy.ops.transform.translate(value=(0,0,float(row[9])))
        if tr_adj == 'long':
            bpy.ops.transform.resize(value=(5,1,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'large':
            bpy.ops.transform.resize(value=(1,3,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'short':
            bpy.ops.transform.resize(value=(1,1,0.5))
        elif tr_adj == 'tall':
            bpy.ops.transform.resize(value=(1,1,3))



    elif row[0]=='tr' and spatial_indicator == "before":
        bpy.ops.wm.append(directory = row[2], filename = row[3])
        bpy.ops.transform.translate(value=((float(row[4])+ float(row[7]))/2,0,0))
        if tr_adj == 'long':
            bpy.ops.transform.resize(value=(5,1,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'large':
            bpy.ops.transform.resize(value=(1,3,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'short':
            bpy.ops.transform.resize(value=(1,1,0.5))
        elif tr_adj == 'tall':
            bpy.ops.transform.resize(value=(1,1,3))



    elif row[0]=='tr' and spatial_indicator == "on_the_left_of":
        bpy.ops.wm.append(directory = row[2], filename = row[3])
        bpy.ops.transform.translate(value=( 0,  float(row[5])*random.uniform(1,2),0))
        if tr_adj == 'long':
            bpy.ops.transform.resize(value=(5,1,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'large':
            bpy.ops.transform.resize(value=(1,3,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'short':
            bpy.ops.transform.resize(value=(1,1,0.5))
        elif tr_adj == 'tall':
            bpy.ops.transform.resize(value=(1,1,3))



    elif row[0]=='tr' and spatial_indicator == "in_front_of":
        bpy.ops.wm.append(directory = row[2], filename = row[3])
        bpy.ops.transform.translate(value=(((float(row[4])+ float(row[7]))/2,0,0)))
        bpy.ops.transform.rotate(value= 3.14,  axis= (0 , 0 , 1))
        if tr_adj == 'long':
            bpy.ops.transform.resize(value=(5,1,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'large':
            bpy.ops.transform.resize(value=(1,3,1))
        elif tr_adj == 'narrow':
            bpy.ops.transform.resize(value=(1,0.5,1))
        elif tr_adj == 'short':
            bpy.ops.transform.resize(value=(1,1,0.5))
        elif tr_adj == 'tall':
```

```python
            bpy.ops.transform.resize(value=(1,1,3))


        elif row[0]=='tr' and (spatial_indicator == "in" or spatial_indicator==
                  "in_the_centre_of"):
            bpy.ops.wm.append(directory = row[2], filename = row[3])
            bpy.ops.transform.translate(value=(0,0,0))
            if tr_adj == 'long':
                bpy.ops.transform.resize(value=(5,1,1))
            elif tr_adj == 'narrow':
                bpy.ops.transform.resize(value=(1,0.5,1))
            elif tr_adj == 'large':
                bpy.ops.transform.resize(value=(1,3,1))
            elif tr_adj == 'narrow':
                bpy.ops.transform.resize(value=(1,0.5,1))
            elif tr_adj == 'short':
                bpy.ops.transform.resize(value=(1,1,0.5))
            elif tr_adj == 'tall':
                bpy.ops.transform.resize(value=(1,1,3))


        elif row[0]=='tr' and (spatial_indicator == "next_to"or spatial_indicator==
            "near"):
            bpy.ops.wm.append(directory = row[2], filename = row[3])

            m_1= (float(row[7])*random.uniform(1,3),0,0)
            m_2= (-float(row[7])*random.uniform(1,3),0,0)
            m_3= (0, float(row[8])*random.uniform(1,3),0)
            m_4= (0, -float(row[8])*random.uniform(1,3),0)
            m_5= (float(row[7])*random.uniform(1,3),float(row[8])*random.uniform(1,3),0)
            m_6= (-float(row[7])*random.uniform(1,3),float(row[8])*random.uniform(1,3,0)
            m_7= (float(row[7])*random.uniform(1,3),-float(row[8])*random.uniform(1,3,0)
            m_8= (-float(row[7])*random.uniform(1,3),-float(row[8])*random.uniform(1,3,0)

            motions= [m_1,m_2,m_3,m_4,m_5,m_6,m_7,m_8]

            bpy.ops.transform.translate(value=(random.choice(motions)))
            bpy.ops.transform.rotate(value=random.uniform(0,6.28), axis=(0,0,1))
            if tr_adj == 'long':
                bpy.ops.transform.resize(value=(5,1,1))
            elif tr_adj == 'narrow':
                bpy.ops.transform.resize(value=(1,0.5,1))
            elif tr_adj == 'large':
                bpy.ops.transform.resize(value=(1,3,1))
            elif tr_adj == 'narrow':
                bpy.ops.transform.resize(value=(1,0.5,1))
            elif tr_adj == 'short':
                bpy.ops.transform.resize(value=(1,1,0.5))
            elif tr_adj == 'tall':
                bpy.ops.transform.resize(value=(1,1,3))



#Not used, but a possible usefull command
#elif lm_adj == 'red':
    #r = (1, 0, 0) # Set color RGB value
    #for o in bpy.context.selected_objects:
    #      o.active_material.diffuse_color = (r)
```

185

Appendix C

Questionnaire

Thank you for your participation in the questionnaire!

We ask you to express a judgment by clicking on the checkbox you deem most appropriate. You have to evaluate the level of understanding of each image, that is, how much the graphic representation is explanatory of the descriptive sentence, shown above, and how much it enriches it with information. More specifically, each sentence is followed by two pictures that show a three-dimensional and a bi-dimensional representation of the textual description.

Which one do you understand better? Give your rating and have fun!

---

a) Le Corbusier is pointing out the idea of building a platform on a forest of columns
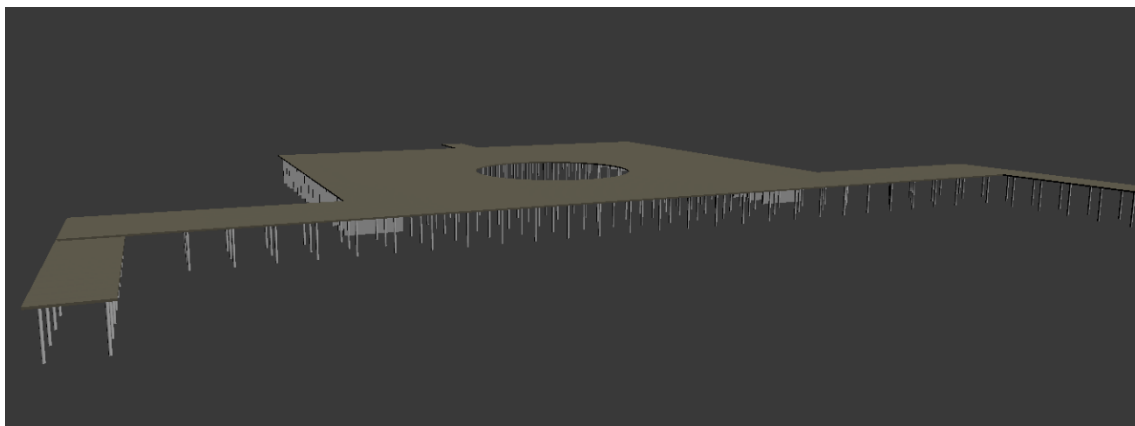
☐ Excellent ☐ Satisfactory ☐ Average ☐ Acceptable ☐ Very Bad

Note: _____

☐ Excellent

☐ Satisfactory

☐ Average

☐ Acceptable

☐ Very Bad

Note:

_____

_____
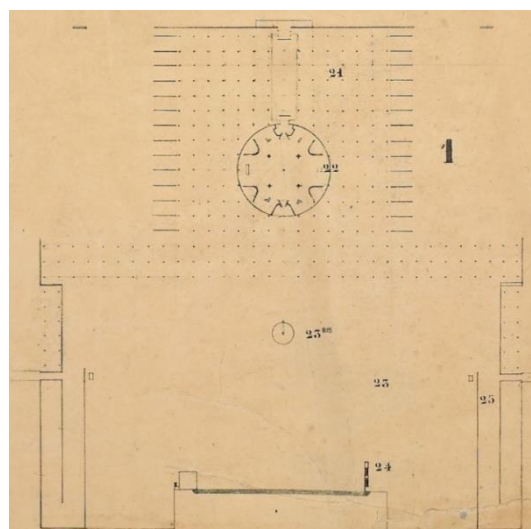
b) The Sacrarium is in the centre of a forest of columns
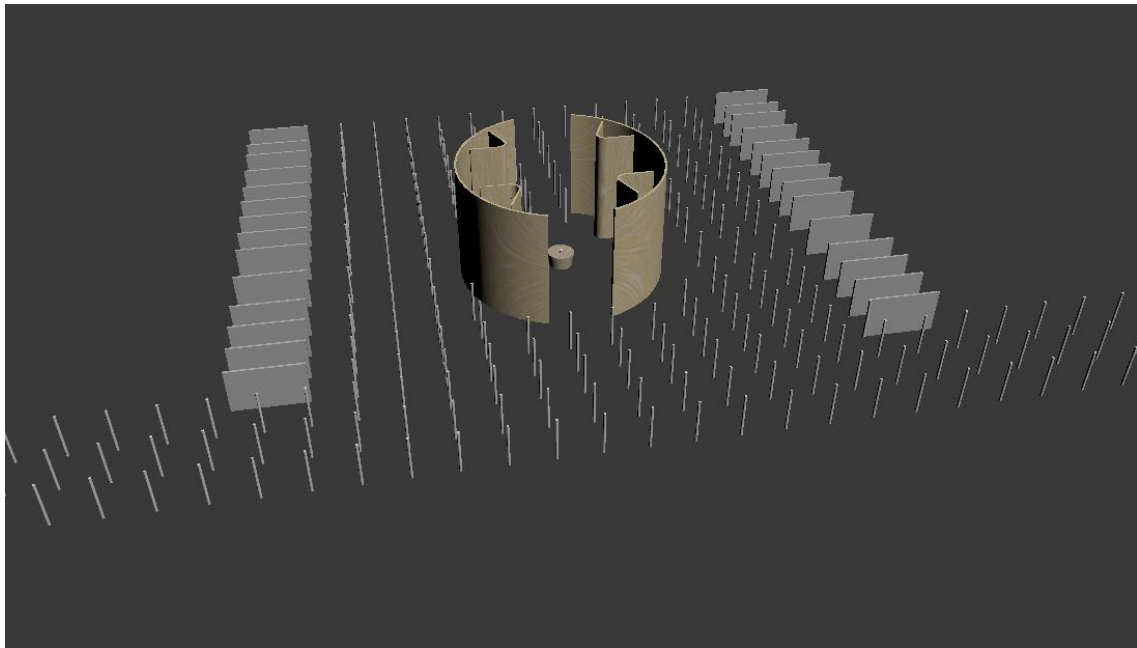
☐ Excellent ☐ Satisfactory ☐ Average ☐ Acceptable ☐ Very Bad
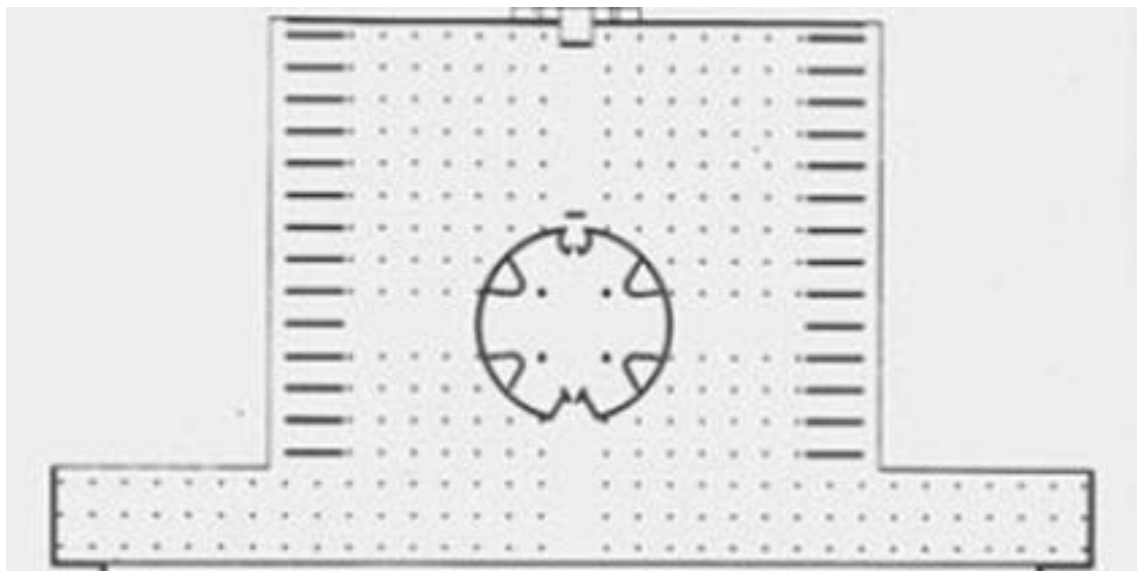
Note: _____

☐ Excellent ☐ Satisfactory ☐ Average ☐ Acceptable ☐ Very Bad

Note: _____

c) *The architect plans a composition where the Library is in front of the University*
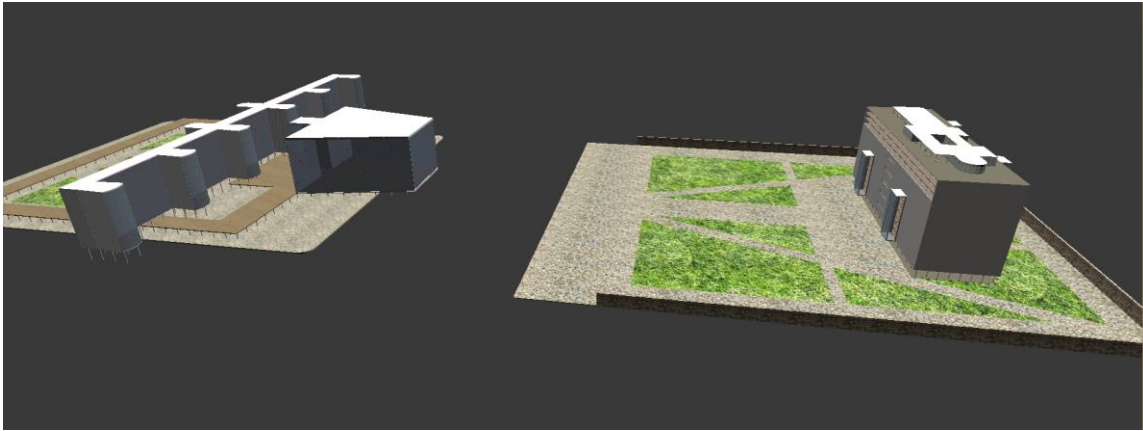
☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad
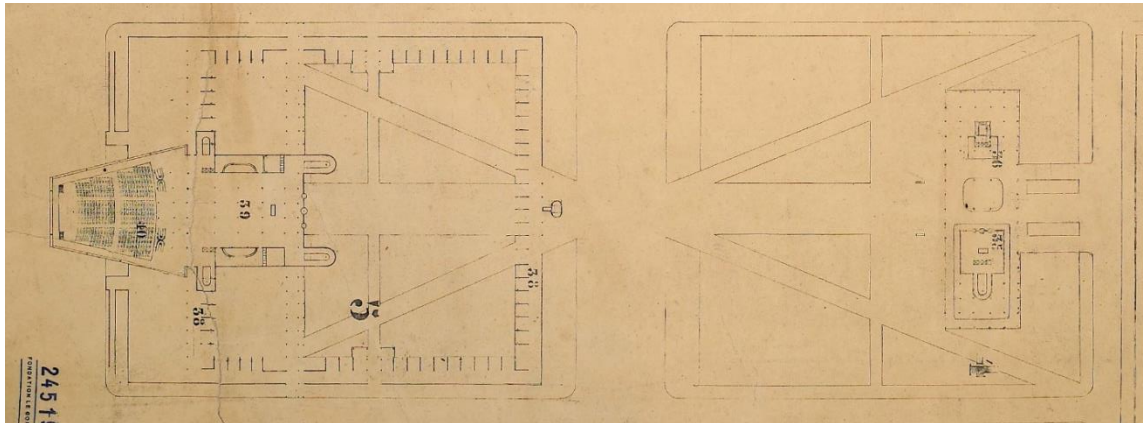
Note: _____

☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad

Note: _____

*d) The planetarium is in front of the Museum, in the same arrangement as in Otlet's 1927
    schema of the Mundaneum*
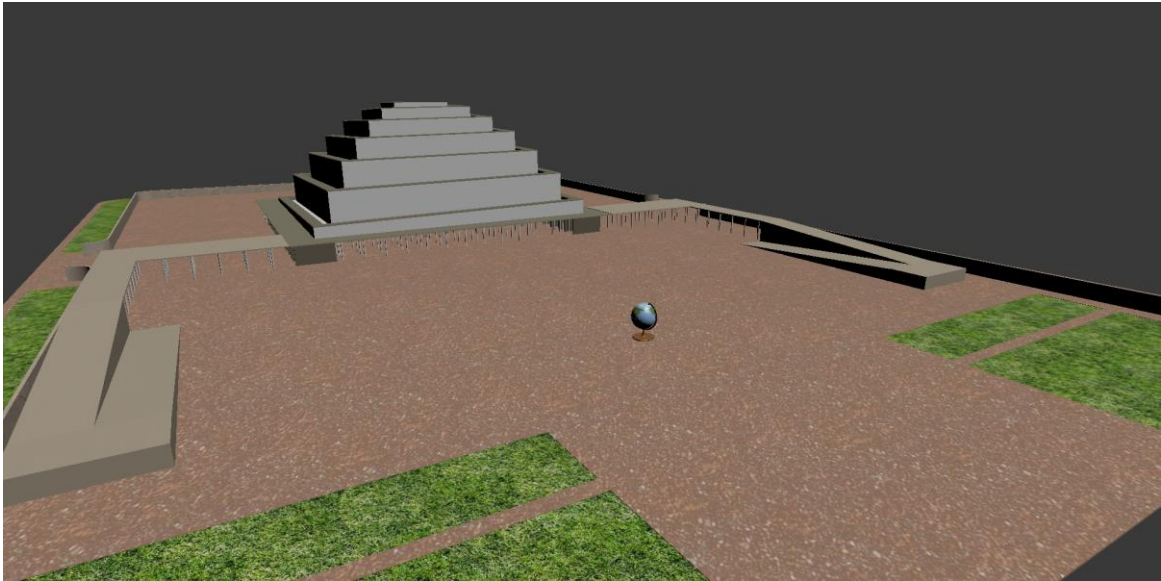
☐ Excellent ☐ Satisfactory ☐ Average ☐ Acceptable ☐ Very Bad

Note: _____

☐ Excellent

☐ Satisfactory

☐ Average

☐ Acceptable

☐ Very Bad

Note:

_____

_____

e) The cité hôtelière, which includes the residential buildings, is planned before the Stadium, the main building of the sport centre.

☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad
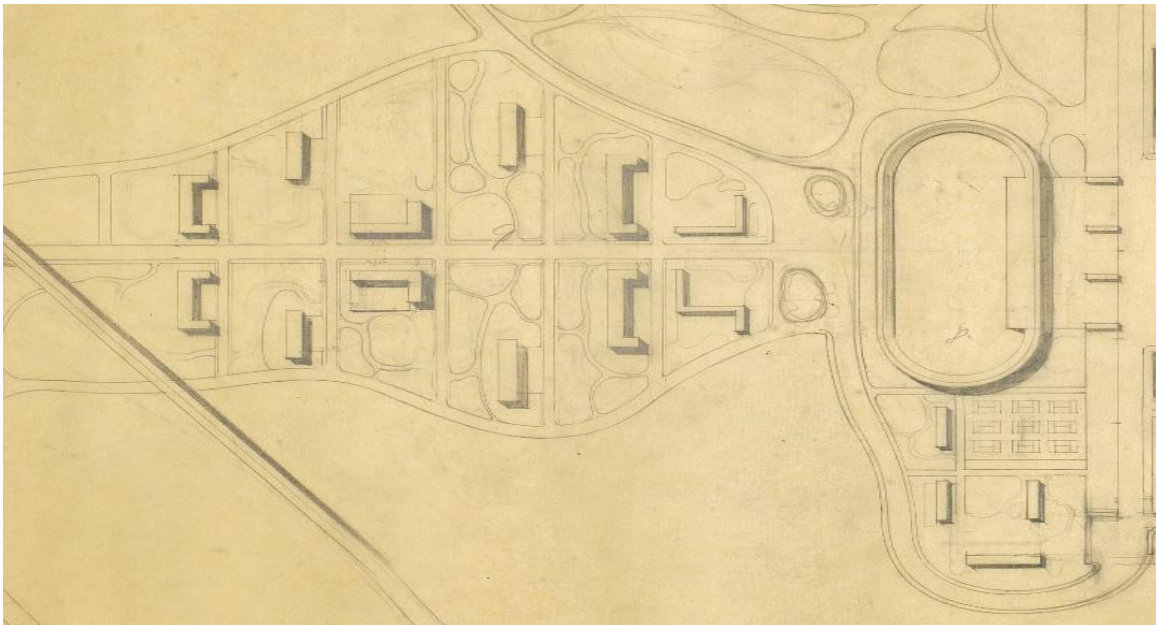
Note: _____

☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad

Note: _____

f) Four cruciform columns, 70 metres tall, are located in the centre of the Sacrarium.
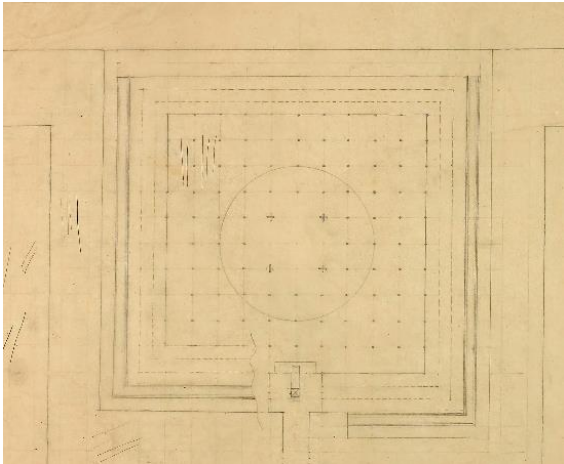
☐ Excellent

☐ Satisfactory

☐ Average

☐ Acceptable

☐ Very Bad

Note:

_____

_____

☐ Excellent

☐ Satisfactory

☐ Average

☐ Acceptable

☐ Very Bad

Note:

_____

_____
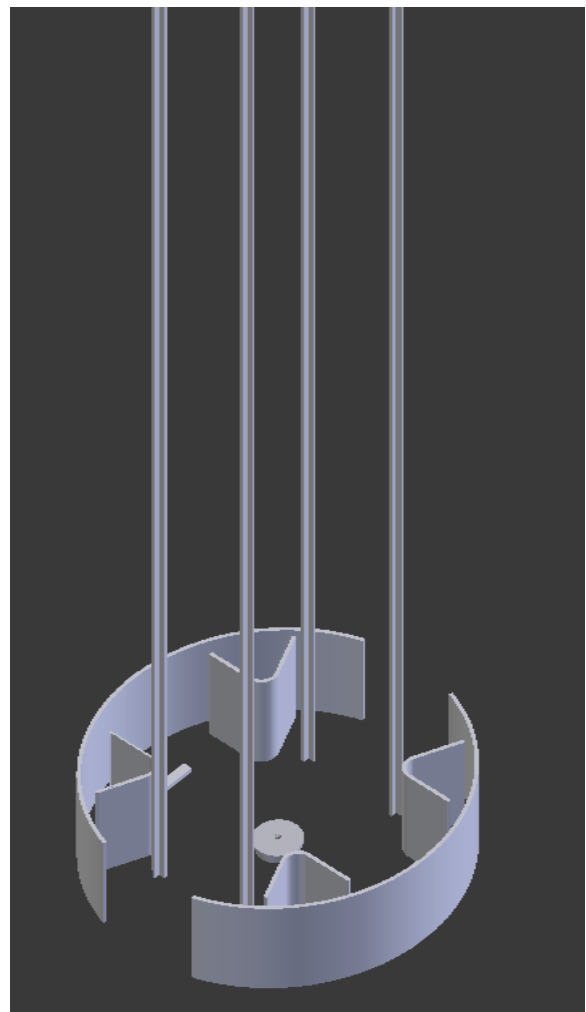
g) The Hall of modern times is a small building located behind the Museum connected to it through a gallery.
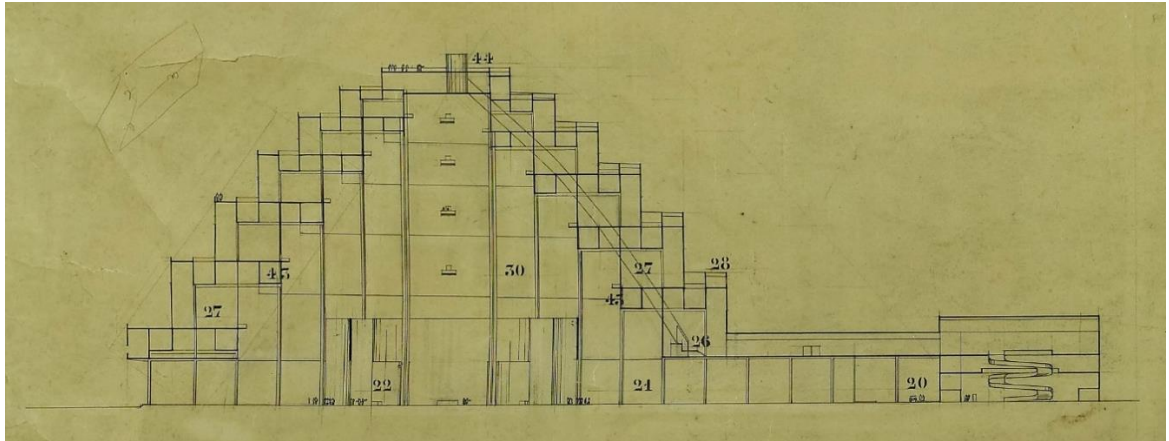
☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad
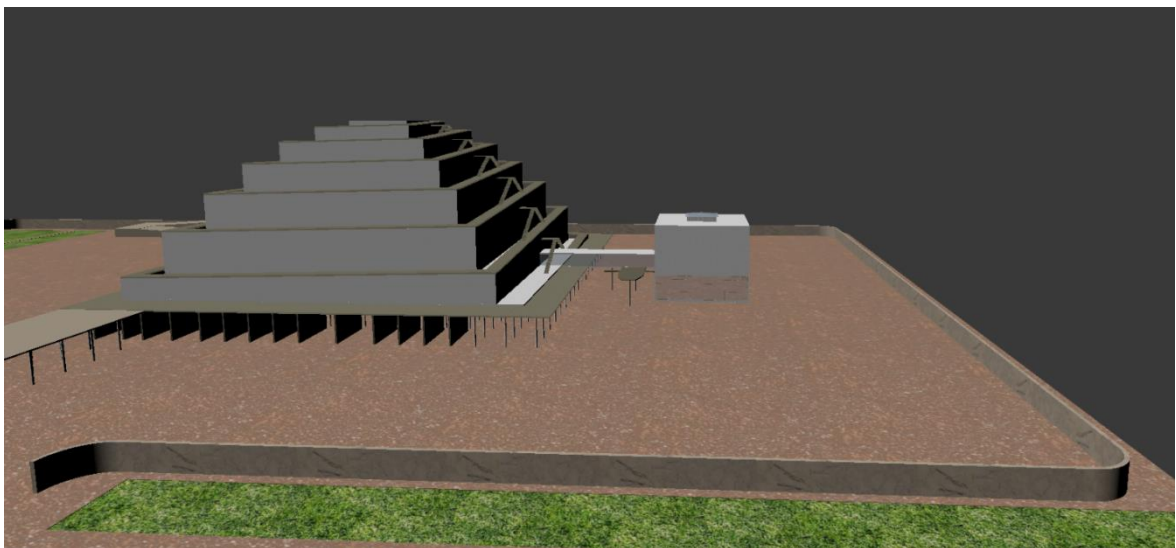
Note: _____

☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad

Note: _____

h) Two symmetrical ramps, strictly linked to the platform, are located in the middle of the Mundaneum main square.
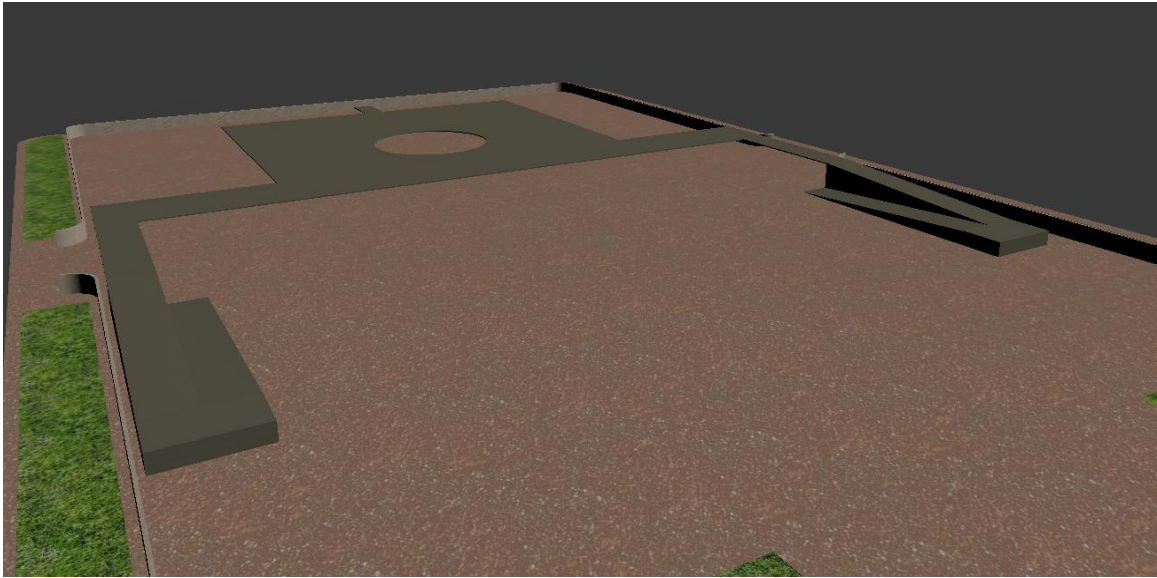
☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad

Note: _____

☐   Excellent

☐   Satisfactory

☐   Average

☐   Acceptable

☐   Very Bad

Note:

_____

_____
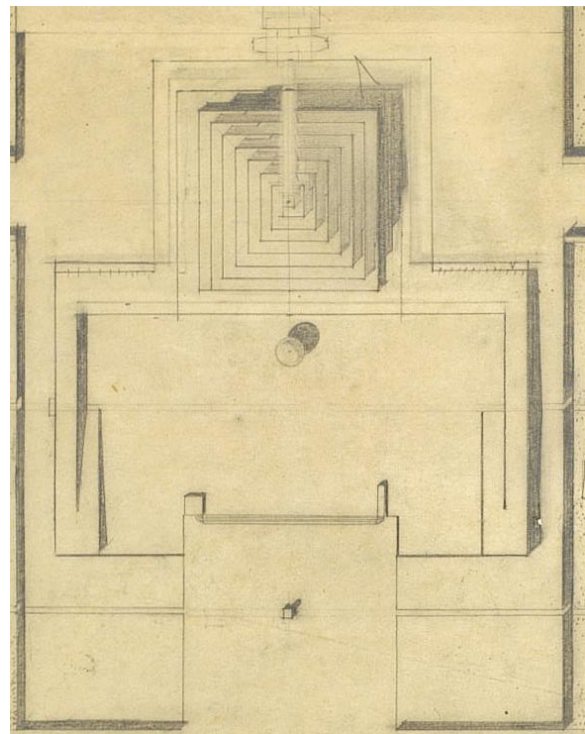
i) The Pavilions of continents, which accommodates international exhibitions of various human cultural activities, is a Mundaneum's building located on the left of the Library.

☐ Excellent ☐ Satisfactory ☐ Average ☐ Acceptable ☐ Very Bad

Note: _____

| 1 MUSEE MONDIAL | 2 HALLES DES TEMPS MODERNES | 5 ASSOCIATIONS INTERNATIONALES |
| | 4 BIBLIOTHEQUE | 5 UNIVERSITE |

7 STADE	9 CONTINENTS
	NATIONS
	VILLES
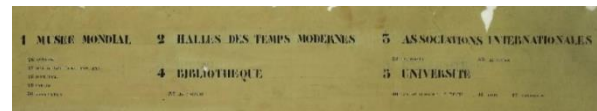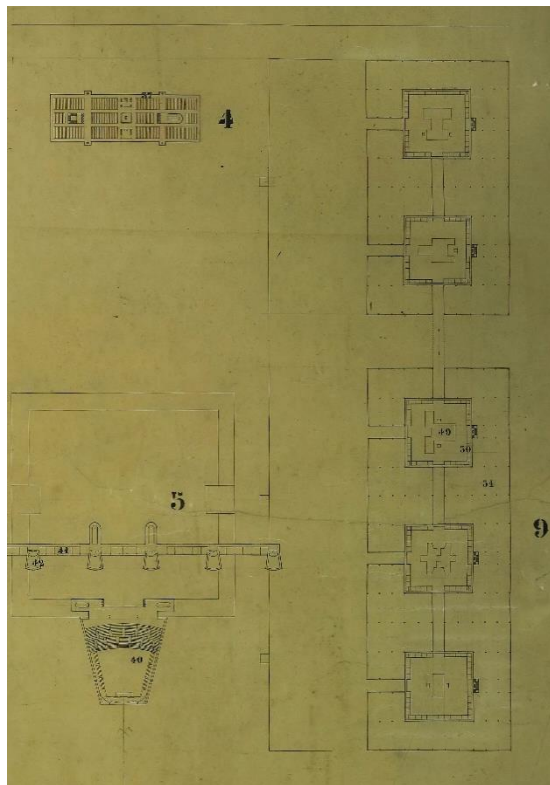
☐ Excellent

☐ Satisfactory

☐ Average

☐ Acceptable

☐ Very Bad

Note:

_____

_____

l) According to Le Corbusier's principle, the library stands on pilotis, distancing itself from the ground
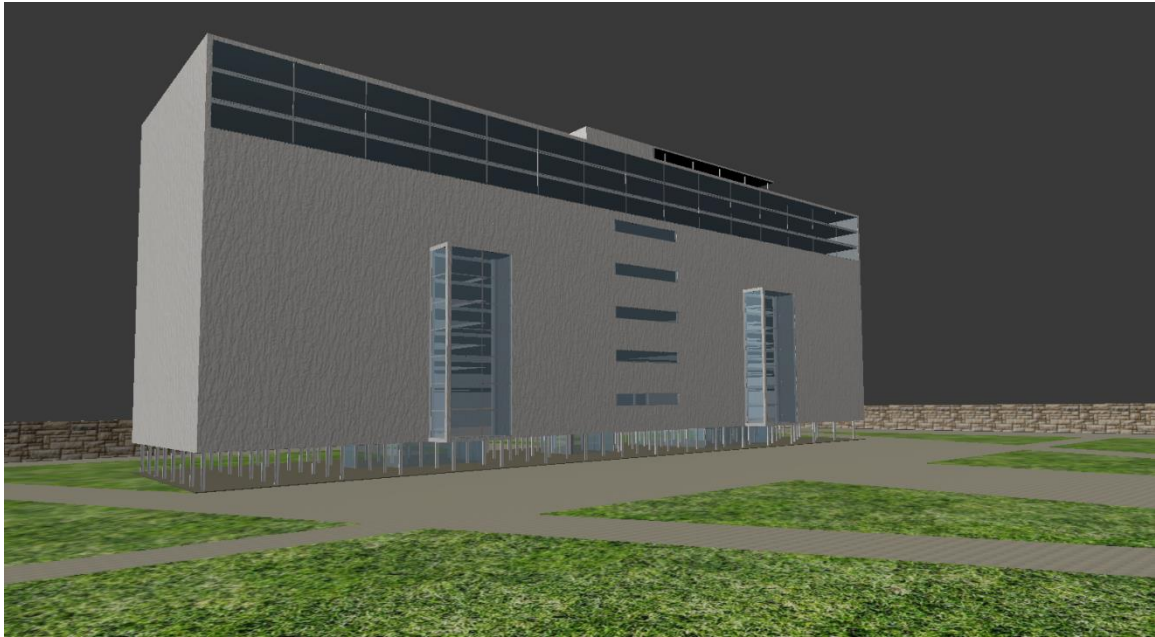
☐ Excellent   ☐ Satisfactory   ☐ Average   ☐ Acceptable   ☐ Very Bad
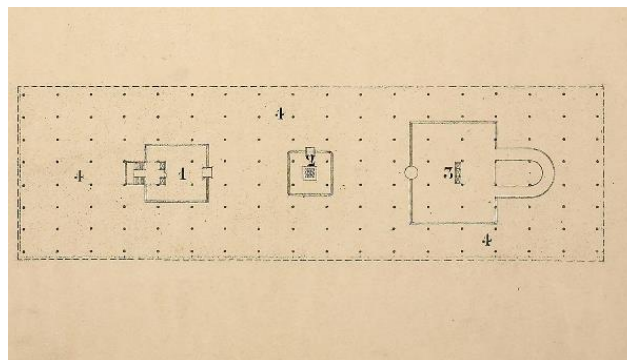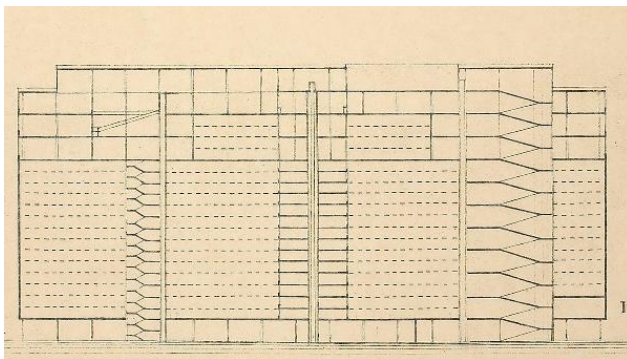
Note: _____

☐ Excellent

☐ Satisfactory

☐ Average

☐ Acceptable

☐ Very Bad

Note:

_____

_____

Appendix D

```
<!DOCTYPE html>
<html>
<title>The World City</title>
<meta charset="ISO-8859-1">
<meta name="viewport" content="width=device-width, initial-scale=1">
<link rel="stylesheet" href="https://www.w3schools.com/w3css/4/w3.css">
<link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Lato">
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-
awesome/4.7.0/css/font-awesome.min.css">

<style>
.centro {display:block; margin:0px auto;text-align:center;}
.tablelink {background-color: grey; color: white;float: left; border: solid 1px lightgrey;
outline:
 none; cursor: pointer;padding: 25px 25px;font-size: 20px; width: 33.33%;}
.tablelink:hover {background-color: lightgrey;}
.tabcontent {color: white;display: none;padding: 80px 40px;height: 100%;}
#Publication128 {border: solid 1px lightgrey; background-color:lightgrey;}
#Publication133 {border: solid 1px lightgrey; background-color:lightgrey;}
#Other {border: solid 1px lightgrey; background-color:lightgrey;}
.tooltip {position: relative;display: inline-block;border-bottom: 1px dotted orange;}
.tooltip .tooltiptext {visibility: hidden;width: 320px;background-color: orange;color:
black;text-
 align: center;border-radius: 6px;    padding: 5px 5px;    position: absolute;    z-index:
1;
 bottom: 150%;    left: 50%;    margin-left: -60px;}
.tooltip .tooltiptext::after {content: "";position: absolute; top: 100%; left: 10%; margin-
left: -
 7px;  border-width: 10px;border-style:solid; border-color: orange transparent transparent
 transparent;}
.tooltip:hover .tooltiptext {visibility: visible;}
.Lettertooltip {position: relative;    display: inline-block;    border-bottom: 0px dotted
orange;}
.Lettertooltip .tooltiptext {visibility: hidden; width: 380px; background-color: orange;
 color: white; text-align: center; border-radius: 6px; padding: 5px 5px; position:
 absolute; z-index: 1; bottom: 105%;left: 18%; margin-left: -60px;}
.Lettertooltip .tooltiptext::after {content: "";position: absolute; top: 100%; left: 10%;
 margin-left: -7px;  border-width: 10px; border-style: solid; border-color: orange
 transparent transparent transparent;}
.Lettertooltip:hover .tooltiptext {    visibility: visible;}
.Letterimage {  display: block;  width: 100%;  height: auto;}
.Letteroverlay {  position: absolute; top: 0; bottom: 0; left: 0; right: 0; height: 100%;
 width: 100%;  opacity: 0;  transition: .5s ease;  background-color: grey;}
.Lettercontainer {  position: relative;  width: 100%;}
.Lettercontainer:hover .Letteroverlay {opacity: 1;}
.Lettertext{color: white;font-size:15px;position: absolute;top:50%;left: 20%;background-
 color:grey;}
.w3-container{margin-top:80px;margin-bottom:20px; margin-left:200px}
.w3-black,.w3-hover-black:hover{color:yellow;background-color:#000}
.content{max-width:800px;margin:auto}
.w3-jumbo{font-size:32px!important}

body,h1,h2,h3,h4,h5,h6 {font-family: "Lato", sans-serif;}
body, html {height: 100%;color: #777;line-height: 1.8;}

/* Create a Parallax Effect */
.bgimg-1, .bgimg-2, .bgimg-3 {    background-attachment: fixed;    background-position:
center;
 background-repeat: no-repeat;    background-size: cover;}
/* First image (Logo. Full height) */
.bgimg-1 {    background-image: url("Immagine1.jpg");    min-height: 100%;}
/* Second image (Portfolio) */
.bgimg-2 {    background-image: url("Immagine2.png");    min-height: 400px;}
/* Third image (Contact) */
.bgimg-3 {    background-image: url("/w3images/parallax3.jpg");    min-height: 400px;}
.w3-wide {letter-spacing: 10px;}
.w3-hover-opacity {cursor: pointer;}

/* Turn off parallax scrolling for tablets and phones */
@media only screen and (max-device-width: 1024px) {.bgimg-1, .bgimg-2, .bgimg-3
{background-attachment: scroll;}}
```

```
</style>

<body>
<!-- Navbar (sit on top) -->
<div class="w3-top">
<div class="w3-bar" id="myNavbar" >
<a class="w3-bar-item w3-button w3-hover-black w3-hide-medium w3-hide-large w3-right"
href="javascript:void(0);" onclick="toggleFunction()" title="Toggle Navigation Menu">
<i class="fa fa-bars"></i>
</a>
<a href="#home" class="w3-bar-item w3-button" style= "font-size:16px;
color:black;">HOME</a>
<a href="#about" class="w3-bar-item w3-button w3-hide-small" style= "font-size:16px;
color:black;"> <i class="fa fa-user"></i> ABOUT</a>
<a href="#portfolio" class="w3-bar-item w3-button w3-hide-small" style= "font-size:16px;
color:black;" ><i class="fa fa-th"></i> GRAPHIC SOURCES </a>
<a href="#textual sources" class="w3-bar-item w3-button w3-hide-small" style= "font-
size:16px;
color:black;"><i class="fa fa-book"></i> TEXTUAL SOURCES</a>
<a href="#letters" class="w3-bar-item w3-button w3-hide-small" style= "font-size:16px;
color:black;"><i class="fa fa-file-powerpoint-o"></i> LETTERS</a>
<a href="#3Dscenes" class="w3-bar-item w3-button w3-hide-small" style= "font-size:16px;
color:black;"><i class="fa fa-cube"></i> 3D SCENES</a>
<a href="#contact" class="w3-bar-item w3-button w3-hide-small" style= "font-size:16px;
color:black;"><i class="fa fa-envelope"></i> CONTACT</a>
<a  class=" w3-right" style="font-family:Lucida Calligraphy; font-size:16px;
color:lightgrey;w3-hover:yellow; padding-right:20px; padding-top:13px;"> Paul Otlet <span
style=font-size:13px;>  and  </span> <span> Le Corbusier <span style=font-size:13px;>
project </a>
</div>

<!-- Navbar on small screens -->
<div id="navDemo" class="w3-bar-block w3-white w3-hide w3-hide-large w3-hide-medium">
<a href="#about" class="w3-bar-item w3-button" onclick="toggleFunction()">ABOUT</a>
<a href="#graphical sources" class="w3-bar-item w3-button"
onclick="toggleFunction()">GRAPHICAL SOURCES</a>
<a href="#textual sources" class="w3-bar-item w3-button" onclick="toggleFunction()">TEXTUAL
SOURCES</a>
<a href="#letters" class="w3-bar-item w3-button" onclick="toggleFunction()">LETTERS</a>
<a href="#3Dscenes" class="w3-bar-item w3-button" onclick="toggleFunction()">3D SCENES</a>
<a href="#contact" class="w3-bar-item w3-button" onclick="toggleFunction()">CONTACT</a>
</div>
</div>

<!-- First Parallax Image with Logo Text -->
<div class="bgimg-1 w3-display-container w3-opacity-min" id="home">
<div class="w3-display-middle" style="white-space:nowrap;">
<span class="w3-center w3-padding-large w3-black w3-xlarge w3-wide w3-animate-opacity"
style="font-family:Lucida Calligraphy;">
<b> THE <span class="w3-hide-small">WORLD</span> <span class="w3-hide-small">CITY</span>
PROJECT</span> </b>
</div>
</div>

<!-- Container (About Section) -->
<div class="w3-content w3-container w3-padding-64 centro"  id="about">
<h3 class="w3-center"> <b style="letter-spacing: 3px; color:brown;">  ABOUT THE PROJECT
</b>  </h3>
<p class="w3-center"><em>Characteristics and Purposes</em></p>
<p style="text-align: justify; font-size:13px;">The <i> "Cité Mondiale" </i> is a project,
never realized, conceived by <span class="tooltip">
<b style="font-size:14px"> Paul Otlet </b> <span class="tooltiptext" style="text-
align:justify; padding:10px; font-size:10px;"> Paul Otlet (1868-1944),
was an author, a lawyer, a socialist and pacifist militant. His will, as a visionary man,
was to live in a world characterized by peace and scientific progress.
Moreover, he aimed to map out the world's knowledge and store it, through an archival
system, into wooden cabinets. His idea and writings anticipated the advent of
the World Wide Web, in fact,  together with this friend Henri La Fontaine he laid the
foundations of the Mundaneum, a sort of paper Google with pacifist principles.
Furthermore, he spent several years in the implementation of a project regarding the
construction of an ideal society.</span> </span> since 1910.
It should have been a utopian city based on harmony and scientific progress and a self-
sufficient city from a politic, economic, social and intellectual point of view.
The idea was to create, in the middle of the Europe, a "place of the world culture" in
which gather the five traditional institutions of intellectual creativity and enhance
the peaceful use of knowledge. This new city presents itself as a collection of elements
scattered in an immense green park near the lake.
```

```
</p>
<div class="w3-row">
<div class="w3-col m5 w3-center style="padding:0px">
<img src="otlet.jpg" class="w3-round w3-image" alt="Plan of the World City" style="border-
radius:13px; box-shadow: 10px 10px lightgrey;" width="330" height="190">
</div>

<div class="w3-col m7 w3-hide-small w3-padding-large">
<p style="text-align:justify; font-size:13px;"> Paul Otlet's project was unsuccessful
because it was rejected by the Geneva authorities for reasons of too high costs,
both because of the Nazi invasion in the 1940s, which reduced the Otlet project to nothing,
because it was considered harmful. The "International City", also known as the
"New Geneva", was called to materialize the dream of building a place that must accommodate
all world knowledge and in which humanity can recognize itself.
A site where to gather people from all over the world to encourage them to communicate and
share knowledge. It should have been a spatial non-space, a "home for men" after World War
I
and a world centre for the accumulation, organization, and dissemination of knowledge.
<br>The core of "World City" is the Mundaneum which includes numerous buildings; they are:
the <b style="font-size:14px"> <a href="https://skfb.ly/6AMEP" target="_blank"><i> World
Museum </i> </a> </b>,
the <b style="font-size:14px">  Hall of Modern Times </b>, the <b style="font-size:14px">
<a href="https://skfb.ly/6AMZv" target="_blank"> <i> International Associations </i> </a>
</b>, the <b style="font-size:14px"> <a href="https://skfb.ly/6AMYQ" target="_blank"> <i>
Library</i></a> </b>,
the  <b style="font-size:14px"> <a href="https://skfb.ly/6AMFP" target="_blank"> <i>
University </i></a> </b> and the <b style="font-size:14px"> <a href="https://skfb.ly/6AMZY"
target="_blank"> <i>Pavilions</i></a> </b>.
Moreover, outside the enclosure of the Mundaneum, the project envisaged the creation of <b
style="font-size:14px"> stadium </b> and of the <b style="font-size:14px"> cité hôtelière
</b>.
<br>The institution of the Mundaneum, according to Otlet's idea, should not have been a
traditional complex of buildings, but an immense emblematic space.</p>
</div>
</div>
</div>
<div class="w3-row w3-center w3-dark-grey w3-padding-32" > </div>

<!-- Second Parallax Image with Portfolio Text -->
<div class="bgimg-2 w3-display-container w3-opacity-min">
<div class="w3-display-middle">
<span class="w3-xxlarge w3-text-white w3-wide">PORTFOLIO</span>
</div>
</div>

<!-- Container (Portfolio Section) -->
<hr color=" red" size="15"  width="300px" align="center";>
<div class=" w3-container w3-padding-48" id="portfolio" style="max-width:830px; margin-
left:180px;">
<h3 class="w3-center " > <b style="letter-spacing: 3px; color:brown;   margin-left:140px
">GRAPHIC SOURCES </b> </h3>
<p class="w3-center " ><em><span style="font-size:18px;   margin-left:140px            ">
In the following some drawings about the project </span> </em><br> <span style="margin-
left:140px"    > (stored in the archives of the "Fondation Le Corbusier", in Paris)
</p><br> </span>

<!-- Responsive Grid. Four columns on tablets, laptops and desktops. Will stack on mobile
devices/small screens (100% width) -->

<div class="w3-row-padding w3-center w3-section"  >
<div class="w3-col m3 " >
<img src="tavole/24521B.jpg" style="width:100%; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Drawing, plan of situation with legend and orientation / 1:
12000 / Indian ink, blue pencil / L leather / 0.63 x 0.90/ 24521">
</div>
<div class="w3-col m3 " >
<img src="tavole/24604.jpg" style="width:100%; height:250px;"  onclick="onClick(this)"
class="w3-hover-opacity" alt="Drawing, shaded outline of the project with museum / Black
pencil / Study layer / 0.74 x 1.11/ 24604">
</div>
<div class="w3-col m3">
<img src="tavole/24510_22.jpg" style="width:100%; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Study drawing, plan, section and elevation showing the museum
with legend / 1: 500 / Indian ink / Leather layer / 1.11 x 1.40/ 24510">
</div>
<div class="w3-col m3 " >
```

```html
<img src="tavole/24579_25.jpg" style="width:375px; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Study drawing, elevation, section and plan of a part of the
project with sketches, calculations, numbering and notes / Black and colored pencil / Study
Layer / 0.67 x 1.09/ 24579">
</div>
</div>

<div class="w3-row-padding w3-center w3-section">
<div class="w3-col m3">
<img src="tavole/24525.jpg" style="width:100%; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Drawing, axonometric perspective on the whole project with
numeration / Indian ink / 0.76 x 1.50 / 24.525">
</div>
<div class="w3-col m3">
<img src="tavole/24519.jpg" style="width:375px; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Map of the museum, library, halls, university, numeration and
legend / 1: 1000 / Gelatin print / Canson paper / 0.93 x 1.40 / 24519">
</div>
<div class="w3-col m3">
<img src="" style="width:100%; height:250px;" onclick="onClick(this)" class="w3-hover-
opacity" alt="">
</div>
<div class="w3-col m3">
<img src="tavole/24509.jpg" style="width:375px;  height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Study drawing, plan with legend / 1: 1000 / Indian ink /
Leather layer / 1.07 x 1.38 / 24509">
</div>
</div>

<div id="demo" style="display:none"  >
<div class="w3-row-padding w3-center w3-section" >
<div class="w3-col m3">
<img src="tavole/24583.jpg" style="width:375px;  height:240px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Two Study Drawing, Library Plans / Black and Red Pencil /
Study Layer / 0.51 x 0.59/ 24583">
</div>
<div class="w3-col m3">
<img src="" style="width:450px; height:250px;"  class="w3-hover-opacity" alt="">
</div>
<div class="w3-col m3">
<img src="tavole/24605.jpg" style="width:100%; height:240px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Drawing, shaded mass outline of the project with sketches and
calculations / Pencil / 0.73 x 0.87/ 24605">
</div>
<div class="w3-col m3">
<img src="tavole/24563.jpg" style="width:375px;  height:240px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="Three elevations of facades of the museum with silhouettes /
Black pencil, blue chalk / Study layer / 0,85 x 0,90 / 24563">
</div>
</div>

<div class="w3-row-padding w3-center w3-section" >
<div class="w3-col m3  ">
<img src="tavole/24512.jpg" style="width:100%; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="World Library / Floor plans A, B, C with room numbers, legend
tables / Graphic scale / 13-6-29 / Gelatin print / Canson paper / 0.76 x 0.63 / 24512">
</div>
<div class="w3-col m3">
<img src="tavole/24513.jpg" style="width:100%; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="World Library / Floor plans D, E, F with room numbers, tables
of legends / Graphic scale / 13-6-29 / Indian ink / Leather layer / 0,76 x 0,71/ 24513">
</div>
<div class="w3-col m3">
<img src="tavole/24514.jpg" style="width:100%; height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="World Library / Level Plans, Floor G and Cutting with Room
Numbers (Books, Public, Staff) / Graphic Scale / 13-6-29 / Gelatin Print / Canson Paper /
0.67 x 0.63 / 24514">
</div>
<div class="w3-col m3">
<img src="tavole/24581.jpg" style="width:375px;  height:250px;" onclick="onClick(this)"
class="w3-hover-opacity" alt="World Library / Four Drawings, Facade Elevations of the
Bibioteque, Various Sketch Front and Back / Black and Yellow Pencil / Study Layer / 0.62 x
1.10/ 24581">
</div>
</div>
</div>
```

```
<button id="myBtn" type="button" style="border-radius: 15px; margin-left: 53%; width:120px;
height:40px; background-color:lightgrey; cursor: pointer;"  onclick="disableBtn()" > <b>
LOAD MORE!</b> </button>

<!-- Modal for full size images on click-->
<div id="modal01" class="w3-modal w3-black" onclick="this.style.display='none'">
<span class="w3-button w3-large w3-black w3-display-topright" title="Close Modal Image"><i
class="fa fa-remove"></i></span>
<div class="w3-modal-content w3-animate-zoom w3-center w3-transparent w3-padding-64">
<img id="img01" class="w3-image">
<p id="caption" class="w3-opacity w3-large"></p>
</div>
</div>


<!-- Container Textual sources-->
<div class="w3-content w3-container w3-padding-32" style="max-width:830px; margin-
left:220px;"  id="textual sources">
<h3 class="w3-center " > <b style="letter-spacing: 3px; color:brown;  MARGIN-LEFT:-70PX;
">TEXTUAL SOURCHES </b> </h3>


<div class="w3-container w3-black  w3-xlarge w3-padding-64" style="margin-left: -15%" >
<div >
 <h2 class="centro w3-jumbo" style="font-family:Lucida Calligraphy;">PUBLICATIONS AND OTHER
</h2>
<button class="tablelink" style="font-family:Lucida Calligraphy; text-align:center; font-
size:18px;" onclick="openPage('Publication128',this, 'brown')"
id="defaultOpen">Publication128</button>
<button class="tablelink" style="font-family:Lucida Calligraphy; text-align:left; font-
size:18px;" onclick="openPage('Publication133',this, 'brown')">Publication133</button>
<button class="tablelink" style="font-family:Lucida Calligraphy; text-align:center; font-
size:18px;" onclick="openPage('Other',this, 'brown')">Other</button>
</DIV>
<div id="Publication128" class="tabcontent" >
<p class="w3-text-grey" style="text-align:center; font-family:Lucida Calligraphy;">
<i>Mundaneum  </a> </i></p> <br>
<img src="pubb128.PNG" width="380" height="500" class="centro">
<br><br> <p style="text-align:center; font-size:15px; color:blue;"> <a href=
"texts/p_128.pdf" target="_blank"> Read it! </a> </p>
</div>
<div id="Publication133" class="tabcontent ">
<p class="w3-text-grey" style="text-align:center; font-family:Lucida Calligraphy;"> <i>
Cité Mondiale. <br> Geneva: world Civic Center   </i></p>
<img src="pubb133.PNG" alt="Pubblication n.133" width="380" height="500" class="centro">
<br><br> <p style="text-align:center; font-size:15px; color:blue;"> <a href=
"texts/p_133.pdf " target="_blank"> Read it! </a> </p>
</div>

<div id="Other" class="tabcontent ">
 <p style=" font-family:Lucida Calligraphy;  color: grey;font-size:17px;margin-
left:25px;"><a href= "texts/A.pdf " target="_blank"><i>Karel Teige's "Mundaneum" (1929)
</a> </i> </p> <br>
<p  style="font-family:Lucida Calligraphy;  color: green;font-size:17px;"><a href=
"texts/B.pdf " target="_blank"><i>Opening the Shrine of the Mundaneum   </a> </i> </p> <br>
<p  style=" font-family:Lucida Calligraphy;  color: red; margin-left:90px; font-
size:17px;"><a href= "texts/C.pdf " target="_blank"><i>Museo del Mundaneum (1928): la
génesis de un prototipo   </a> </i> </p> <br>
<p  style="font-family:Lucida Calligraphy; color: blue;  margin-left:0px; font-
size:17px;"><a href= "texts/D.pdf " target="_blank"><i>Architectural Metaphors of
Knowledge: The Mundaneum Designs of Maurice Heymans, Paul Otlet, and Le Corbusier   </a>
</i> </p> <br>
<p  style="font-family:Lucida Calligraphy;  color: purple; font-size:17px;"><a href=
"texts/E.pdf " target="_blank"><i>Transnational Associations Journal. Otlet et Le Corbusier
</a> </i> </p> <br>
<p  style="font-family:Lucida Calligraphy;  color: hsla(9, 100%, 64%, 0.5); margin-
left:25px;font-size:17px;"><a href= "texts/F.pdf " target="_blank"><i>Connaissance totale
et Cité mondiale   </a> </i> </p> <br>
<p  style="font-family:Lucida Calligraphy;  color: tomate;font-size:17px; margin-
left:140px;"><i>La città mondiale. <br>Andersen, Hébrand, Otlet, Le Corbusier  </i> </p>
</div>
</div>
</div>


<script>
function openPage(pageName,elmnt,color) {
    var i, tabcontent, tablelinks;
    tabcontent = document.getElementsByClassName("tabcontent");
    for (i = 0; i < tabcontent.length; i++) {
```

```
        tabcontent[i].style.display = "none";
    }
    tablelinks = document.getElementsByClassName("tablelink");
    for (i = 0; i < tablelinks.length; i++) {
        tablelinks[i].style.backgroundColor = "";
    }
    document.getElementById(pageName).style.display = "block";
    elmnt.style.backgroundColor = color;}

// Get the element with id="defaultOpen" and click on it
document.getElementById("defaultOpen").click();
</script>

 <!—Letter grid -->
<div class="w3-padding-32" style="width:1000px; margin-left:-90px;" id="letters">
<h3> <b style="letter-spacing: 3px; color:brown; margin-left:310px; ">OTLET'S CITATIONS AND
LETTERS </b> </h3>

<div class="w3-row-padding" style=" padding:25px;  BACKGROUND-COLOR: BLACK; margin-
left:125px">
<div class="w3-half" >
<div> <span class="Lettertooltip"> <img src="lettere/2.jpg" style="width:100%; margin-
bottom:16px; float:left;"  >
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px; ">
Presentation of the motivations and characteristics of the project.</div>
<div> <span class="Lettertooltip" >         <img src="lettere/lettOr_.png"
style="width:100%; margin-bottom:16px;  ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:102%;"> The collaboration between Otlet and Le Corbusier seems to become "official"
with the letter of 2 April 1928. It was written after some meetings in Paris and
Brussels.</div>
<div> <span class="Lettertooltip" > <img  src="lettere/1.png" style="width:100%; margin-
bottom:50px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:106%;"> Exaltation of the idea of the World City. </div>
<div> <span class="Lettertooltip" > <img src="lettere/21.PNG" style="width:100%; margin-
bottom:18px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:102%;"> Letter sent to Andersen by Otlet, in which he expresses his impressions,
illustrates the points of convergence identified and the methods of collaboration. It
refers, in particular, to the desire to provide - through the construction of a center of
global character "un point de ralliement aux organismes internationaux". </div>
</div>

<div class="w3-half">
<div> <span class="Lettertooltip" > <img src="lettere/13.PNG" style="width:100%; margin-
bottom:20px;margin-top:20px;"   >
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:96%;"> Outlet highlights the importance of the work and encourages collaboration in
an atmosphere of trust and optimism.</div>
<div> <span class="Lettertooltip" >         <img src="lettere/14.PNG" style="width:100%;
margin-bottom:20px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:114%;"> Willing to materialize the design idea by merging, in one thing, the idea
and the building. </div>
<div> <span class="Lettertooltip" >        <img src="lettere/16.PNG" style="width:100%;
margin-bottom:20px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:106%;"> Architecture as an answer to all needs. </div>
<div> <span class="Lettertooltip" >         <img src="lettere/15.PNG" style="width:100%;
margin-bottom:20px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:108%;"> Perfect combination of the idea and the architecture. </div>

<div> <span class="Lettertooltip" ><img src="lettere/17.PNG" style="width:100%; margin-
bottom:18px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:104%;"> The ideas of Otlet and Le Corbusier assume a character of "internationalism"
and are oriented towards the construction of the World City in its entirety. </div>
<div> <span class="Lettertooltip" > <img src="lettere/20.PNG" style="width:100%;
HEIGHT:212PX; margin-bottom:50px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:105%;"> Otlet, in the aftermath of the First World War, considers the World City as
the instrument to peacefully resolve conflicts between nations.
The structure of the project highlights the centrality of the interaction between people
and that of knowledge as founding elements of a new and different brotherhood among people.
</div>
```

```
<div> <span class="Lettertooltip" >  <img src="lettere/24.PNG" style="width:100%; margin-
bottom:18px; ">
<span class="tooltiptext" style="text-align:justify; padding:10px; font-size:12px;
bottom:102%;"> Andersen confirms its willingness to cooperate.
However, he never intended to design a real city capable of becoming a permanent seat of
international associations but rather to demonstrate, through the elaboration of
architectural drawings, the possibility of creating a physical place in which art could
represent the element capable to establish peace and harmony among peoples. </div>
</div>
</div> <br> <br> <br>
<!—grid end-->

<!-- Overlay effect when opening sidebar on small screens -->
<div class="w3-overlay w3-hide-large" onclick="w3_close()" style="cursor:pointer"
title="close side menu" id="myOverlay"></div>

  <!-- Slideshow Header -->
  <div class="w3-container w3-content centro w3-padding-32"  style="width:90%";
id="3Dscenes">
<h3 style="letter-spacing: 3px; color:brown; MARGIN-LEFT: 40PX";> <B >3D SCENES
RECONTRUCTION </B></h3> <BR>
<div class="w3-display-container mySlides">
<img src="1.png" style="width:100%;margin-bottom:-6px; MARGIN-LEFT: -80PX; border: solid
grey 45px;  border-style: outset; height: 470px;">
<div class="w3-display-bottomleft w3-container w3-black" STYLE="MARGIN-LEFT: -35PX">
<p  style="font-size:14px"> <a href="halls_pyramid.blend" style="text-decoration:none;">
The hall of modern times is a small building located behind the Museum connected to it
through a gallery. </a> </p>
</div>
</div>
<div class="w3-display-container mySlides">
<img src="2.png" style="width:100%;margin-bottom:-6px; MARGIN-LEFT: -100PX; border: solid
grey 45px;  border-style: outset; height: 470px;">
<div class="w3-display-bottomleft w3-container w3-black" STYLE="MARGIN-LEFT: -55PX";">
<p  style="font-size:14px"> <a href="library_university.blend" style="text-
decoration:none;"> The architect plans a composition where the Library is in front of the
University. </a></p>
</div>
</div>
<div class="w3-display-container mySlides">
<img src="3.png" style="width:100%;margin-bottom:-6px; MARGIN-LEFT: -100PX; border: solid
grey 45px;  border-style: outset; height: 470px;">
<div class="w3-display-bottomleft w3-container w3-black" STYLE="MARGIN-LEFT: -55PX">
<p  style="font-size:14px"> <a href="globe_pyramid.blend" style="text-decoration:none;"> A
globe, in a scale 1 = 1000000, with the planetarium inside, is situated in front of the
museum. </a> </p>
</div>
</div>
<div class="w3-display-container mySlides">
<img src="4.png" style="width:100%;margin-bottom:-6px; MARGIN-LEFT: -100PX; border: solid
grey 45px ;  border-style: outset;  height: 470px;">
<div class="w3-display-bottomleft w3-container w3-black" STYLE="MARGIN-LEFT: -55PX">
<p  style="font-size:14px; "> <a href="platform_columns.blend" style="text-
decoration:none;">  Le Corbusier is pointing out the idea of building a platform on a
forest of columns. </a> </p>
</div>
</div>
</div>
<div class="w3-row-padding w3-section" style="width:100%">
<div class="w3-col s3" >
<img class="demo w3-opacity w3-hover-opacity-off" src="1.png"
style="width:100%;cursor:pointer; border:solid grey 3px;" onclick="currentDiv(1)" >
</div>
<div class="w3-col s3">
<img class="demo w3-opacity w3-hover-opacity-off" src="2.png"
style="width:100%;cursor:pointer; border:solid grey 3px;" onclick="currentDiv(2)" >
</div>
<div class="w3-col s3">
<img class="demo w3-opacity w3-hover-opacity-off" src="3.png" style="width:100%; height:
124px; cursor:pointer; border:solid grey 3px;" onclick="currentDiv(3)">
</div>
<div class="w3-col s3">
<img class="demo w3-opacity w3-hover-opacity-off" src="4.png" style="width:100% ; height:
124px; cursor:pointer; border:solid grey 3px;" onclick="currentDiv(4)" >
</div>
</div>
```

```
<p> <b style="font-size:15px; MARGIN-LEFT: 20PX; color: black;"> THREE-DIMENSIONAL SCENES:
</b> <BR> <i style="font-size:14px; MARGIN-LEFT: 20PX; color: black;"> Click on the
descriptive sentences to download the 3D models (*).
In addition, to download the whole three-dimensional model of the World City (*), <a
href=" world_city.blend">  click here.</i></a></p>
<p style="font-size:12px; MARGIN-LEFT: 20PX; color: black;"> (*) You need to install <a
href=" https://www.blender.org/download/" target="_blank"> Blender software </a> to
visualize the virtual scenes. </p>


<!-- Contact -->
<div class="w3-light-grey  w3-padding-64 centro" style=" margin-top: 100px" id="contact">
<h3 class="w3-center centro ">Contact</h3>
<hr>
<p style="font-size:20px; color: purple; text-align:center;">If you liked this page and
would like more information, do not hesitate to contact me!</p>
<div style= "margin-left: 50px;">
<p> <b> Name:</b>  Valentina Bova </p>
<p> <b> email: </b>valentinabova@dimes.unical.it</p>
</div>
<button type="submit" class="w3-button w3-block w3-dark-grey"><A
HREF="mailto:valentinabova@dimes.unical.it" target="_blank" style=" text-decoration:none;">
Send me an e-m@il </a></button>
</div>


<script>
function disableBtn() {
    document.getElementById('demo').style.display='block';
    document.getElementById("myBtn").disabled = true;}

// Script to open and close sidebar when on tablets and phones
function w3_open() {
    document.getElementById("mySidebar").style.display = "block";
    document.getElementById("myOverlay").style.display = "block";
}

function w3_close() {
    document.getElementById("mySidebar").style.display = "none";
    document.getElementById("myOverlay").style.display = "none";
}

// Slideshow Images
var slideIndex = 1;
showDivs(slideIndex);

function plusDivs(n) {
  showDivs(slideIndex += n);
}

function currentDiv(n) {
  showDivs(slideIndex = n);
}

function showDivs(n) {
  var i;
  var x = document.getElementsByClassName("mySlides");
  var dots = document.getElementsByClassName("demo");
  if (n > x.length) {slideIndex = 1}
  if (n < 1) {slideIndex = x.length}
  for (i = 0; i < x.length; i++) {
     x[i].style.display = "none";
  }
  for (i = 0; i < dots.length; i++) {
     dots[i].className = dots[i].className.replace(" w3-opacity-off", "");
  }
  x[slideIndex-1].style.display = "block";
  dots[slideIndex-1].className += " w3-opacity-off";
}

// Modal Image Gallery
function onClick(element) {
  document.getElementById("img01").src = element.src;
  document.getElementById("modal01").style.display = "block";
  var captionText = document.getElementById("caption");
  captionText.innerHTML = element.alt;
}

// Change style of navbar on scroll
```

```
window.onscroll = function() {myFunction()};
function myFunction() {
    var navbar = document.getElementById("myNavbar");
    if (document.body.scrollTop > 100 || document.documentElement.scrollTop > 100) {
        navbar.className = "w3-bar" + " w3-card" + " w3-animate-top" + " w3-white";
    } else {
        navbar.className = navbar.className.replace(" w3-card w3-animate-top w3-white",
"");
    }
}
// Used to toggle the menu on small screens when clicking on the menu button
function toggleFunction() {
    var x = document.getElementById("navDemo");
    if (x.className.indexOf("w3-show") == -1) {
        x.className += " w3-show";
    } else {
        x.className = x.className.replace(" w3-show", "");
    }
}
</script>
</body>
</html>
</html>
```

Bibliography

[1] G. Sansoni, M. Trebeschi, and F. Docchio, "State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation", Sensors, 9(1), 2009, pp. 568-601.

[2] P. Otlet, Cité Mondiale. Geneva : World Civic Center : Mundaneum, Publication n. 133 de l'Union des Associations Internationales, Palais Mondial, Bruxelles, 1929.

[3] A. Cropper, "Identifying and inferring objects from textual descriptions of scenes from books," in OASIcs- OpenAccess Series in Informatics, Vol. 43, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014, pp. 19-26.

[4] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, and P. Hanrahan, "Example-based synthesis of 3D object arrangements", ACM Transactions on Graphics (TOG), Vol.6, 2012.

[5] R. Rugma and S. Sreeram, "Text-to-Scene Conversion System for Assisting the Education of Children with Intellectual Challenges", in International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 8, August, 2016.

[6] S. Cooper, W. Dann, and R. Pausch, "Developing algorithmic thinking with Alice", in proceedings of ISECON. 2000. pp. 506-539.

[7] S. Cooper, W. Dann, and R. and Pausch, "Developing algorithmic thinking with Alice", in proceedings of ISECON, Vol. 17, November, 2000, pp. 506-539.

[8] A. Chang, M. Savva, and C. D. Manning, "Learning Spatial Knowledge for Text to 3D Scene Generation," in Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 2028– 2038.

[9] K. Charter, "Principles for conservation and restoration of built heritage", Marsilio, Venice, 2000.

[10] E. Ranfa, "Paul Otlet: una vita per la documentazione", AIB, studi 53.1, 2013.

[11] G. Gresleri, and D. Matteoni, "La città mondiale". Andersen, Hébrand, Otlet, Le Corbusier, first edition, Polis/Marsilio Editori, 1982.

[12] Y. Charbit,and V. Arundhati, "The Platonic city: history and utopia.", Population, Vol.2, 2002, pp. 207-235.

[13] H. Blumenfeld, "Theory of city form, past and present.", Journal of the society of architectural historians, Vol.3/4, 1949, pp. 7-16.

[14] M. Osti, "Le città ideali del Rinascimento tra sogno e realtà. The ideal cities of the Renaissance between dream and reality", 2018.

[15] Congress Internationaux d'Architecture moderne (CIAM), La Charte d'Athenes or The Athens Charter, 1933. Trans J.Tyrwhitt. Paris, France: The Library of the Graduate School of Design, Harvard University, 1946.

[16] Teige, Karel. "Mundaneum", na, 1929.

[17] H. C. Andersen, and E. M. Hébrard, "Creation of a World Centre of Communication", Paris, 1913.

[18] P. Otlet, "Mundaneum", Publication n. 128 de l'Union des Associations Internationales, Palais Mondial, Bruxelles, 1928.

[19] W. Van Acker, "Architectural Metaphors of Knowledge: The Mundaneum Designs of Maurice Heymans, Paul Otlet, and Le Corbusier.", library trends, 2012, pp. 371-396.

[20] W. Van Acker, "Opening the Shrine of the Mundaneum: The Positivist Spirit" in the Architecture of Le Corbusier and his Belgian 'Idolators', in A. Brown & A. Leach (Eds.), in Proceedings of the Society of Architectural Historians, 2013.

[21] J. Sinclair, "Corpus and Text - Basic Principles" in Developing Linguistic Corpora: a Guide to Good Practice, ed. M. Wynne, Oxford: Oxbow Books, 2005, pp.1-16.

[22] L. Bowker, and J. Pearson, "Working with Specialized Language: a practical guide to using corpora", Routledge, London e New York, 2002.

[23] A. Caruso, and A. Folino, "Corpus-based knowledge representation in specialized domains," in Corpus-based studies on language varieties, Peter Lang, 2016, pp. 11-35.

[24] D. Biber, "Representativeness in corpus design", Literary and linguistic computing 8.4, pp. 243-257, 1993.

[25] J. Pearson, "Terms in Context," Amsterdam-Philadelphia: John Benjamins Publishing Company, 1998, pp. 58-62.

[26] V. Bova, and E. Cardillo, "Descriptive Sentence Extraction for Text to 3D scene Generation", The International Workshop on Knowledge Extraction and Semantic Annotation, KESA 2018, April 22-26 Athens, Greece, 2018.

[27] J. Zlate, "Holistic spatial semantics of Thai, Cognitive Linguistics and Non-Indo-European Languages", 2003, pp. 305–336.

[28] Marie Francine Moens. Information Extraction: Algorithms and Prospects in a Retrieval Context. The Information Retrieval Series 21. New York: Springer, 2006.

[29] A. Lenci, S. Montemagni, and V. Pirrelli, "Testo e computer: elementi di linguistica computazionale", Carocci, 2005.

[30] A. Paolani, "Dal parlato allo scritto: le machine che parlano e quelle che leggono", *in I Quaderni di* Telema, Le machine che capiscono, Mucci A. (a cura di), n. 208 di Media Duemila, 2006, pp89-90.

[31] M. W. Chang, L. Ratinov, and D.Roth, "Structured learning with constrained conditional models", Machine learning, Vol.3, 2012, pp. 399-431.

[32] P. Domingos, and M. Richardson. Markov logic: "A unifying framework for statistical relational learning", in ICML'04 Workshop on Statistical Relational Learning and its Connections to Other Fields, 2004, pp. 49–54.

[33] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, "Learning probabilistic relational models", in Proceedings of the 16th International Joint Conference on Artificial Intelligence, Vol. 2, 1999.

[34] B. Taskar, P. Abbeel, and D. Koller, "Discriminative probabilistic models for relational data", in Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, UAI'02, Morgan Kaufmann Publishers Inc., 2002, pp. 485–492.

[35] T. G. Dietterich, P. Domingos, L. Getoor, S. Muggleton, and P. Tadepalli, "Structured machine learning: the next ten years". Machine Learning, Vol. 1, 2008, pp. 3–23.

[36] C. Sutton, and A. McCallum, "Collective segmentation and labeling of distant entities in information extraction", in ICML Workshop on Statistical Relational Learning and Its Connections, 2004.

[37] D. Roth, and T. W. Wen, "Global inference for entity and relation identification via a linear programming formulation.", Introduction to statistical relational learning, 2007, pp. 553-580.

[38] M. Collin, "Parameter estimation for statistical parsing models: theory and practice of distribution-free methods", in New Developments in Parsing Technology, Kluwer, 2004, pp. 19–55.

[39] C. Sutton, and A. MacCallum, "Introduction to conditional random fields for relational learning", in L. Getoor and B. Taskar, editors, Introduction to Statistical Relational Learning, MIT Press, 2006, pp 108–143.

[40] B. Taskar, M. F. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data", in Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, Advances in Neural Information Processing Systems 16, 2004.

[41] A. Pollani, and A. Di Carlo, "Il Libro Bianco sul trattamento automatico della lingua", Fondazione Ugo Bordoni, Roma, 2009.

[42] S. Bolasco, "Statistica testuale e text mining: alcuni paradigmi applicativi", Quaderni di Statistica, 7, 2005.

[43] D. Jurafsky, and James H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing", Computational Linguistics and Speech Recognition, Prentice Hall, second edition, 2008.

[44] C. W. Kreidler, "Introducing english semantics", Routledge, 2002.

[45] L. Màrquez, X. Carreras, K. C. Litkowski, and S. Stevenson, "Semantic role labeling: an introduction to the special issue", Computational Linguistics, Vo1.2, 2008, pp.45–159.

[46] T. Fontenelle, "FrameNet and frame semantics: a special issue", International Journal of Lexicography, Vol.3, 2003.

[47] D. Jurafsky, and J. H. Martin, "Speech and Language Processing: an Introduction to Natural Language Processing", Computational Linguistics and Speech Recognition. Prentice Hall, second edition, 2008.

[48] P. Kordjamshidi, D. Khashabi, C. Christodoulopoulos, B. Mangipudi, S. Singh, and D.Roth, "Better call saul: Flexible programming for learning and inference in nlp", in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 3030-3040.

[49] P. Kordjamshidi, Structured Machine Learning for Mapping Natural Language to Spatial Ontologies, Doctoral thesis, KU Leuven University– Faculty of Engineering, 2013.

[50] J. Cussens, "Issues in learning language in logic", in A. Kakas and F. Sadri, editors, Computational Logic: Logic Programming and Beyond, volume 2408 of LNCS, 2002, pp. 171–177.

[51] J. Renz, and B. Nebel, "Qualitative spatial reasoning using constraint calculi", in "Handbook of spatial logics", 2007, pp.161-215.

[52] P. Kordjamshidi, and al. "Spatial role labeling annotation scheme, in: Handbook of linguistic annotation." Handbook of linguistic annotation, 2015.

[53] A. David, Z. C. Randell, and G.C. Anthony, "A spatial logic based on regions and connection", in Proceedings of the 3rd International Conferance on the Principles of Knowledge Representation and Reasoning, KR'92, 1992, pp. 165–176.

[54] J. Hois, and O. Kutz, "Natural language meets spatial calculi" in C. Freksa, N. S. Newcombe, P. Gärdenfors, and S. Wölfl, editors, Spatial Cognition VI. Learning, Reasoning, and Talking about Space, Vol. 5248 of LNCS, Springer, 2008, pp. 266–282.

[55] J. A. Bateman, "Language and space: A two-level semantic approach based on principles of ontological engineering", in International Journal of Speech Technology, Vol.1, 2010, pp. 29–48.

[56] O. Stock, editor. "Spatial and Temporal Reasoning", Kluwer Academic Publishers, 1997.

[57] T. Gruber, "Encyclopedia of database systems", in Ling Liu And M. Tamer Özsu (Eds.), Springer-Verlag, 2009.

[58] A. Stubbs, "MAE and MAI: lightweight annotation and adjudication tools", in Proceedings of the 5th Linguistic Annotation Workshop, association for Computational Linguistics, 2011.

[59] P. Kordjamshidi, R. Dan, and W. Hao "Saul: Towards Declarative Learning Based Programming", IJCAI, 2015.

[60] M. Odersky, L. Spoon, and B. Venners, "Programming in Scala: A Comprehensive Step-by-step Guide", Artima Incorporation, USA, 1st edition, 2008.

[61] F.R. Kschischang, B.J. Frey, and H.A. Loeliger, "Factor graphs and the sum-product algorithm", Information Theory, IEEE Transactions on, 47 Vol.2, 2001, pp.498–519.

[62] B. Taskar, P. Abbeel, and D. Koller, "Discriminative probabilistic models for relational data", in Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, pp. 485–492, 2002.

[63] A. Cunningham, M. Paluri, and F. Dellaert, "Ddf-sam: Fully distributed slam using constrained factor graphs", in Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, 2010, pp. 3025–3030.

[64] A.Ghrab, S.Skhiri, S.Jouili, and Esteban Zimnyi, "An analytics-aware conceptual model for evolving graphs" in Data Warehousing and Knowledge Discovery, volume 8057 of LNCS, Springer Berlin Heidelberg, 2013, pp. 1–12.

[65] J. Hois, T. Tenbrink, R. Ross, and J. Bateman, "GUM-Space-The Generalized Upper Model spatial extension: A linguistically-motivated ontology for the semantics of spatial language", version 3.0, SFB/TR8 Technical report, Collaborative Research Center for Spatial Cognition, University of Bremen, Germany, 2009.

[66] P. Kordjamshidi, M. van Otterlo, and M. F. Moens. "Spatial role labeling: towards extraction of spatial relations from natural language", ACM -Transactions on Speech and Language Processing, 8, 2011, pp.1–36.

[67] A. Galton, "Spatial and temporal knowledge representation", Journal of Earth Science Informatics, Vol.3, 2009, pp. 169–187.

[68] P. Kordjamshidi, J. Hois, M. van Otterlo, and M. F. Moens, "Learning to interpret spatial natural language in terms of qualitative spatial relations", Oxford University, 2013.

[69] J. latev, "Spatial semantics", in D. Geeraerts and H. Cuyckens, editors, The Oxford Handbook of Cognitive Linguistics, 2007, pp. 318–350.

[70] L.A. Carlson, and S.R. Van Deman, "The space in spatial language", J. of Memory and Language, 51, 2004, pp. 418–436.

[71] I. Mani,J. Hitzeman, J. Richer, D. Harris, R. Quimby, and B. Wellner, "SpatialML:Annotation Scheme, Corpora, and Tools", in 6th Int. Language Resources and Evaluation, LREC'08, 2008.

[72] Q. Shen, X. Zhang, and W. Jiang, W., "Annotation of Spatial Relations in Natural Language", in Int. Conf. on Environmental Science and Information Application Technology, 2009.

[73] M. W. Chang, L. ArieRatinov, and D. Roth, "Structured learning with constrained conditional models", Machine Learning, Vol.3, 2012, pp. 399–431.

[74] K. Litkowski, and O. Hargraves, "SemEval-2007 Task 06: Word-sense disambiguation of prepositions", in Proceedings of the 4th International Workshop on Semantic Evaluations, ACL, 2007, pp. 24–29.

[75] M. Grubinger, P. Clough, H. Müller, and T. Deselaers, "The IAPR benchmark: a new evaluation resource for visual information systems", in Proceedings of International Conference on Language Resources and Evaluation (LREC), 2006, pp. 13–23.

[76] J. Bateman, T. Tenbrink, and S. Farrar, "The role of conceptual and linguistic ontologies in discourse", Discourse Processes, Vol.3, 2007, pp.175–213.

[77] R. Obradovic, "Computer Animation as a Medium Used for Education and for Entertainment", Summer school on image processing (SSIP) 2017, Novi Sad, 2017.

[78] W. J. Schroeder, B. Lorensen, and K. Martin, "The visualization toolkit: an object-oriented approach to 3D graphics" Kitware, 2004.

[79] S. Y. Oh, and J. Bailenson, "Virtual and Augmented Reality", The International Encyclopedia of Media Effects, 2017, pp.1-16.

[80] R. Jayakanthan, "Application of computer games in the field of education", The electronic library, Vol.2, 2002, pp. 98-102.

[81] M. Papastergiou, "Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation" Computers & Education, Vol.1, 2009, pp. 1-12.

[82] G. Lepouras, and C. Vassilakis, "Virtual museums for all: employing game technology for edutainment", *Virtual reality*, Vol.2, 2004, pp. 96-106.

[83] G. Guidi, M. Russo, "Reality-based and reconstructive models: digital media for cultural heritage valorization", SCIRES-IT-SCIentific RESearch and Information Technology, Vol.2, 2011, pp. 71-86.

[84] T. Winograd, "Procedures as a representation for data in a computer program for understanding natural language (No. MAC-TR-84)," MASSACHUSETTS INST OF TECH CAMBRIDGE PROJECT MAC, 1971.

[85] S. O. Nuallain, and A. G. Smith, "An investigation into the common semantics of language and vision.", Integration of Natural Language and Vision Processing, Springer, Dordrecht, 1994, pp. 21-30.

[86] J. Kelleher, T. Doris, Q. Hussain, and S.O. Nualláin, "SONAS: Multimodal, Multi-User Interaction with a Modelled Environment", In Spatial Cognition, 2000, pp. 171-184.

[87] G. Adorni, M. Di Manzo, and F. Giunchiglia, "Natural language driven image generation," in Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics, association for Computational Linguistics, 1984, pp. 495-500.

[88] S. R. Clay, and J. Wilhelms, "Put: Language-Based Interactive Manipulation of Objects", in IEEE Computer Graphics and Applications, 1996, pp. 31–39.

[89] X. Zeng, N. E. Gough, and Q. H. Mehdi, "From Visual Semantic Parameterization to Graphic Visualization," Ninth International Conference on Information Visualisation (IV'05)(IV), London, England, 2005, pp. 488-493.

[90] K. Glass, et al., "Mechanisms for multimodality: taking fiction to another dimension", in AFRIGRAPH '07: Proceedings of the 5th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa, New York, USA, 2007, pp. 135–144.

[91] M. Ma, "Automatic Conversion of Natural Language to 3D Animation", Ph.D. thesis, University of Ulster, 2006.

[92] K. Glass, and S. Bangay, "Automating the creation of 3D animation from annotated fiction text", in Proceedings of the IADIS International Conference on Computer Graphics and Visualization, Vol. 3, 2008.

[93] K. Xu, J. Stewart, and E. Fiume, "Constraint-based automatic placement for scene composition", in Graphics Interface Vol. 2, 2002, pp. 25-34.

[94] L. M . Seversky, and L. Yin, "Real-time automatic 3D scene generation from natural language voice and text descriptions", in Proceedings of the 14th ACM international conference on Multimedia , ACM, 2006, pp. 61-64.

[95] D. Golland, P. Liang, and D. Klein, "A game-theoretic approach to generating spatial descriptions", in Proceedings of the 2010 conference on empirical methods in natural language processing, Association for Computational Linguistics, 2010, pp. 410-419.

[96] S. Dupuy, A. Egge, V. Legendre, and P. Nugues, "Generating a 3D simulation of a car accident from a written description in natural language: The Carsim system", in Proceedings of the workshop on Temporal and spatial information processing, vol. 13, Association for Computational Linguistics, 2001, p. 1.

[97] A. Egges, P. Nugues, and A. Nijholt, "Generating a 3D Simulation of a Car Accident from a Formal Description: the CarSim System", 2001.

[98] R. Johansson , D. Williams, A. Berglund, and P. Nugues, "Carsim: a system to visualize written road accident reports as animated 3D scenes", in Proceedings of the 2nd Workshop on Text Meaning and Interpretation, Association for Computational Linguistics, 2004, pp. 57-64.

[99] B. Coyne, and R. Sproat, "WordsEye: An Automatic Text-to- Scene Conversion System," in Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH „01), ACM, New York, 2001, pp. 487–496.

[100]   B. Coyne, O. Rambow, J. Hirschberg, and R. Sproat, "Frame Semantics in Text-to-Scene Generation, in KES (4)", 2010, pp. 375-384.

[101]   A. Chang, M. Savva, and C. D. Manning, "Learning Spatial Knowledge for Text to 3D Scene Generation," in Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 2028– 2038.

[102]   R. Sproat, "Inferring the environment in a text-to-scene conversion system," in Proceedings of the 1st international conference on Knowledge capture, ACM, 2001, pp. 147-154.

[103]   M. Savva, A. X. Chang, G. Bernstein, C. D. Manning, and P. Hanrahan, "On being the right scale: Sizing large collections of 3D models," in SIGGRAPH Asia 2014 Workshop on Indoor Scene Understanding: Where Graphics meets Vision, 2014.

[104]   A. Chang, W. Monroe, M. Savva, C. Potts, and C.D. Manning, "Text to 3d scene generation with rich lexical grounding," arXiv preprint arXiv: 1505.06289, 2015.

[105]   S. Dessai, and R. Dhanaraj, "Text to 3D Scene Generation", in International Journal of Latest Trends in Engineering and Technology (IJLTET), vol. 6, n. 3, 2016, pp. 255-258.

[106]   W. De Mulder, N. Q. Do Thi, P. van den Broek, and M. F. Moens, "Machine understanding for interactive storytelling", in Proceedings of KICSS 2013: 8th international conference on knowledge, information, and creativity support systems, Springer, 2013, pp. 73-80.

[107]   K. Schwarz., P. Rojtberg, J. Caspar, I. Gurevych, M. Goesele, and H.P. Lensch, "Text-to-Video: Story Illustration from Online Photo Collections", in KES, Vol.4, 2010, pp. 402-409.

[108]   D. Joshi, J. Z. Wang, and J. Li, "The story picturing engine: finding elite images to illustrate a story using mutual reinforcement.", in Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, ACM, 2004.

[109]   T. Regier, "A Model of the Human Capacity for Categorizing Spatial Relations", Cognitive Linguistics, Vol.1, 1995, pp.63-88.

[110] K. Lockwood, K. Forbus, D. Halstead, and J. Usher, "Automatic categorization of spatial prepositions", in Proceedings of the 28th annual conference of the cognitive science society, 2006,pp. 1705-1710.

[111] L. C. Boggess, "Computational Interpretation English Spatial Prepositions",Vol. 75, ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB, 1979.

[112] M. Feist M, "On in and on: an investigation into the linguistic encoding of spatial scenes", Doctoral dissertation, Northwestern University, 2000.

[113] A. Herskovits, "On the spatial uses of prepositions", in Proceedings of the 18th annual meeting on Association for Computational Linguistics, Association for Computational Linguistics, 1980, pp.1-5.

[114] M. Di Manzo, G. Adorni, and F. Giunchiglia, "Reasoning about scene descriptions", in Proceedings of the IEEE, 1986, pp. 1013- 1025.

[115] C. Hansen, "Known and potential high performance computing applications in computer graphics and visualization", High Performance Computing for Computer Graphics and Visualisation, Springer, London, 1996, pp. 23-29.

[116] D. Di Mascio, P. Pauwels, and R. De Meyer, "Improving the knowledge and management of the historical built environment with BIM and ontologies: the case study of the Book Tower", 13th International Conference on Construction Applications of Virtual Reality, 2013.

[117] S. Logothetis, A. Delinasiou, and E. Stylianidis, "Building information modelling for cultural heritage: a review.", ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015.

[118] C. Eastman et all., "BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors", John Wiley & Sons, 2011.

[119] L. Hirschman, and Thompson, H.S. (1995) Chapter 13, Evaluation: Overview of Evaluation in Speech and Natural Language Processing. In R.A. Cole, J. Mariani, H. Uszkoreit, A.Zaenen, and V.Zue (eds.), Survey of the State of the Art in Human Language Technology, Cambridge University Press.

[120] L. Hirschman, and H S. Thompson, "Overview of evaluation in speech and natural language processing.", 1997.

Ackowledgments

Il momento dei ringraziamenti è sempre quello più complesso in quanto le persone a cui vorresti dir grazie sono veramente tante.

La prima che mi sento di ringraziare è il Prof. Roberto Guarasci, che mi ha dato la possibilità di intercorrere il percorso di dottorato. È stata dura, non posso dire il contrario, ma durante la salita ho imparato tante cose, viaggiando anche molto … e viaggiare, si sa, fa cresce, e ci permette di conoscere nuove persone e di aprire gli occhi a nuovi orizzonti. Grazie per avermi dato sempre fiducia, per aver scherzato nei momenti in cui la mia ansia cresceva, per avermi lasciata libera pur controllando che tutto andasse per il verso giusto. In questo dottorato mi sono addentrata in campi totalmente distanti da me e, da ingegnere edile, mi sono dovuta cimentare in nuovi domini mettendomi sempre alla prova e apprendendo giorno dopo giorno cose nuove. A volte il cuore batteva, a volte mi sono sentita persa… ma alla fine ho raggiunto il mio traguardo, e ne sono fiera.

Un altro grazie speciale va a tutto il LabDoc che in questi tre anni è stata una vera e propria famiglia. Sono stata accolta fin da subito a braccia aperte e, col tempo, l'affetto e l'amicizia che provo per ciascuno di loro sono cresciuti in maniera incontrollata. Grazie perché ogni mattina era un piacere alzarmi sapendo che avrei visto i vostri sorrisi: li porterò sempre nel cuore! Un bacio particolare va alla dolce Susina, che è stata sempre pronta ad ascoltarmi e a consigliarmi. Non è solo una amica…è di più! Un grazie speciale anche alla dolce Anto, sempre affettuosa e rassicurante che ha sempre trovato il tempo per essere una grande guida e una grande amica. Grazie ad Erika, Francesca P., Elisa, Valeria, Francesca M., Maria, Claudia, Elena, Federica, Anna, Toni, Pino e Camillo. Per ciascuno di voi avrei voluto spendere tante belle parole, parole che ho nel cuore, ma siete tanti per cui preferisco mandare a tutti un grande bacio e stringervi forte, ringraziandovi per l'affetto che mi avete dimostrato in ogni occasione. Grazie anche al grande Guerino, che con la sua simpatia, professionalità e dolcezza mi ha sempre dimostrato la sua stima e la sua amicizia.

Un immenso ringraziamento va, poi, alla mia famiglia che è sempre stata, e sempre sarà, il mio punto di riferimento, la mia luce nel buio e la mia acqua nel deserto. Ringrazio mia madre, che ha sopportato le mie preoccupazioni e i miei dubbi e mi ha saputa rassicurare con quei suggerimenti che solo una mamma dal cuore grande riesce a dare. Grazie per essermi stata accanto, soffrendo con me nei momenti più difficili e gioendo in quelli più belli. Ringrazio mio padre, che mi ha sempre dato sicurezza, anche quando mi sentivo disorientata

e mi ha indicato la retta via pur facendomi procedere da sola, consentendomi di crescere.

Grazie a mio fratello Marco che con sua pacatezza e gentilezza è riuscito sempre a trasmettermi serenità e grazie anche ad Ida per la dolcezza e sensibilità che ha sempre avuto nei miei riguardi.

Come non ringraziare anche i miei zii Aldo e Luisa che mi hanno costantemente ascoltata e consigliata, senza mai stancarsi, e mi hanno dato quei suggerimenti di vita che porterò sempre con me. Grazie anche ai miei zii Nenella e Oreste che, tra una battuta e uno scherzo, hanno dimostrato grande dolcezza e sensibilità nei miei riguardi. Grazie anche agli zii Carlo e Anna e, chiaramente, grazie a nonna Angela che ha sempre creduto in me, lodandomi ed essendo fiera di me. Tutto ciò non può che riempirmi il cuore di gioia.

Grazie anche a tutti miei amici, a quelli di vecchia data ma anche a quelli nuovi. Un grazie particolare a Rita, amica da sempre, ma anche alla mia coinquilina Andrea e ai miei colleghi della Pc cube, soprattutto alla cara Martina.

Mi fermo qui, nonostante le persone da ringraziare siano molte altre, ma è veramente impossibile citarle tutte. Concludo, quindi, dedicando a voi questo mio traguardo e ribadendo il mio Grazie