POLITECNICO DI MILANO

DIPARTIMENTO DI INGEGNERIA IDRAULICA, INFRASTRUTTURE VIARIE,
AMBIENTALE E DEL RILEVAMENTO – SEZIONE GEOMATICA

DOTTORATO DI RICERCA IN INGEGNERIA AMBIENTALE E DELLE INFRASTRUTTURE

XXV CICLO

Connections among images, texts and maps

Relatore: prof. Luigi Mussio
Politecnico di Milano

Coordinatore del Corso di Dottorato: prof. Alberto Guadagnini
Politecnico di Milano

dott. ing. Valentina Dante

matricola 753331

Se voi storiografi o poeti o altri matematici non avessi coll'occhio viste le cose,
male le potresti riferire per le scritture.
E se tu, poeta, figurerai una storia con la pittura della penna,
e 'l pittore col pennello la farà di più facile sadisfazione e men tediosa a essere complesa.

Leonardo da Vinci, Manoscritto A (foglio 99 recto)

La langue est alors le vaisseau à la mer, non plus en chantier...
Il n'y a que le vaisseau sur mer qui soit un objet à étudier dans l'espèce vaisseau

de Saussure, *Course de linguistique générale*

Both the doctoral curriculum and writing this thesis required strenuous research and absolute dedication, that I couldn't achieve without the support of several loving people.

My gratitude goes to Lorenzo, for encouraging me when needed and even spending some nights awake with me, working on this topic.

Thanks to Gregorio, for conscientiously sleeping in the afternoon, from time to time.

Thanks to Marta, for waiting till Christmas.

Many thanks to mum and dad, for their words and help during the writing of this thesis and along the doctoral course.

Thanks to Lorenzo, for managing improved images (my tries were far worse).

Finally, my heartfelt thanks to:

- prof. Luigi Mussio, for strong and effective support during these years;
- prof.ssa Carla Nardinocchi (Roma - La Sapienza) for help, once again;
- Chiara Zonda and Katia Marino, for image improvement.

General index

Figure index

Table index

Introduction

All the several disciplines concerning survey constituted in a single science with the foundation of geodesy and cartography. Over more than four centuries, they both contributed to the development of major branches of applied mathematics: 3D space geometry, calculus, statistics, treatment of spatial observations.

After the second half of the twentieth century, information technologies made new observations available, while data had very different quality and size of databases grew, requiring the use of new methodologies and procedures for data analysis.

Within this picture, geometry, mathematical analysis and discrete mathematics show some limits which prevent the construction of robust models that Statistics could easily handle. In fact, there is now the need for the complete automation of procedures for both management and processing of data: size of available databases requires considerable computing power and should undergo a drastic downsizing, in order to allow a manual analysis or processing. Naturally, then, the technique developed and searched the most effective solution for calculation automation: in particular, some operational problems are substantially solved yet, by the use of dedicated software.

The focus of this thesis is to develop further strategies of elaboration, looking for analysis simplification by increasing the effectiveness of the automation: the aim is to reduce the number of computational steps, that actually force the use of extremely sophisticated software and hardware.
We will study some principles of Linguistics, in order to find processing strategies that may be borrowed to geomatic analysis (study of information coming from maps and images).
We aim to demonstrate that the centring (or matching) required for databases comparison can be successfully improved by using the connections analysis (since the high probability of intersections can be a parameter of recognition reliability) and even some direct comparisons between the structures of the graphs.

Starting from the contrast between the rationalist approach and the empirical point of view, we are looking for integrated tools of abstract rules and statistical modules: so this is a new science, the so-called Human Computer Sciences, or Knowledge Engineering. The main goal of this modern discipline is image processing, assisted by machine learning and robot vision. However, the linguistic competence computer technology can simulate is not yet comparable to that in humans, mainly for two reasons: there’s a technical limit to feigning and even the understanding of language processing is incomplete up to now.

The thesis project focuses on a possible new challenge of artificial intelligence, suggesting a theoretical proposal about image matching in the geomatic field: my study deals with the chance to integrate some concepts, borrowed by liberal arts in the humanistic field, into the context of automated data. The final target is to measure the reliability of the rules of the Universal Grammar by Chomsky, integrated into the automatic reading of maps and images: we'll perform a syntactic recognition of comparison models among maps, images or 3D models, by using archetypes.

Indeed, there is an analogy between the hierarchical structure (tree-shaped) of models and languages syntax: in fact, the combination rules-representations perfectly corresponds to the representation of syntactic structures in mind. Please note that:

- models are built by sub-models, just as words compose sentences;
- while sub-models consist of primitives, just as combinations of letters and syllables make up words.

Grammars are then defined as combination rules for objects primitives and they directly derive from the language that can describe the model. Given that a valid syntactic approach can successfully describe complex objects, I studied its expressive capacity through:

- analysis of the number of primitives in the chosen and adopted model. Naturally, a modest amount of atomic elements makes easier both computation both management of an automated system;
- and analysis of rules, to be applied iteratively according to re-writing instructions.

Therefore, the best possible syntactic representation establishes an effective compromise between the number of existing relationships and the endless amount of primitives necessary for model description, taking into account the technical constraints (the available computational power) and final result of the study.

Finally, the aim of this thesis is to prove the bijective correspondence between an image and the text relevant to its description, of course considering that quality of correspondence depends on both image definition and text complexity. The text shall perfectly correspond to visible structures in the image, without referring to single objects: the key point is that text isn't a map.

chapter 1

State of the art in image processing

Each data set refers to an object and contains the observed values of statistical variables, which typically are samples extracted from random variables, and can be alternately discrete or continuous.

Anyone can conduct statistical analysis on acquired information, through several and different approaches:

- by comparisons via parameterized models, whose unknowns are estimated during the analysis process,
- by processing through a descriptive analysis, athwart the study of variability and interdependence of attributes of elements, within a class of objects,
- by checks for validation, together on data and on models.

Therefore, Statistics works on different issues, related to data analysis: research in this field is focusing towards calculations full automation. Indeed, as long as estimates are based on a few data, everybody can rely on personal memory, on manual dexterity of the operator, or even on a rough order of the information. Conversely, if the amount of data grows, it is desirable that all procedures are automated, so that a computer can manage the counts on its own, without requiring human intervention: the calculus automation, in effect, eliminates the editorial work on data and in the meanwhile it allows limiting the preparation time needed to manage information during processing.

However, the implementation of a fully automated mechanism, which does not require the intervention of an operator, imposes the provision of effective procedural constraints: in fact, the analysis of all possibilities is due, in order to forecast and arrange every required position within the algorithm.

The generality of application conditions in the algorithm is in fact the tool that allows obtaining reliable results from raw data processing, provided that information enters the system as a regulated flow.

Often, too, it is required that the analysis results are available within a short period after entering the data: speed of processing is a goal nowadays, therefore we'll use in automated system all necessary artifices in order to minimize calculation time.

1.1. Automation in digital photogrammetry.

The disciplines related to land surveying became sciences with the foundation of geodesy and cartography: both subjects are contiguous to astronomy and far beyond the cultural limits of traditional land surveying. In more than four centuries, those disciplines have contributed to the development of several branches within applied mathematics: geometry in a 3D space, calculus, statistics, treatment of observations with spatial reference.

After the second half of the twentieth century, the conquest of space and information technologies have made available new observations, which are different in quality and according to size of databases: information now requires the use of new methodologies and procedures for data analysis and computational statistics. In this perspective, the fundamentals of geometry and mathematical analysis show clear limits to build models that statistics can handle.

Instead, discrete mathematics provides several contributions to the necessary procedures, but the weakness of models leads to look for other cultural supports (not certainly methodologies and alternative procedures), seeking outside the areas of mathematics and physical sciences.

Numerical photogrammetry is an extensive example of statistics application, due to:

- the simplicity of analytical models, borrowed from 3D geometry (they come for the image-and-object representations),
- and the amount of data, required by the image processing: this element is typical for digital photogrammetry.

The processing, which is carried out according to a descriptive analysis on acquired information, provides the parameters estimation for used models and - in the end - the validation of both data and models.

The introduction of digital techniques in conducting photogrammetric processes, started in the Nineties, revolutionized instrumentation and methods of relief: the potential offered by the new image management determined the research towards the highest degree of automation in the process of interpretation and processing.

Initially, the classic photogrammetric techniques were transferred into digital systems, setting the photogrammetric analytical processes on digital images: today, all stages of the traditional sequence are operational on computerized systems, provided they happen under the supervision of an operator, which still works according to the traditional process.
Subsequently, objectives and fundamentals of digital photogrammetry were integrated with techniques related to computer vision, which actually deals with the automation of vision-related phenomena. The correspondence is possible because both areas pursue the digital

image processing, in order to improve visual and automatic interpretation, with the use of empirical or stochastic filtering techniques. Moreover, the aim of both studies is common, reaching for complete automation of search for known elements on the images.

Then, there is now the need to come to the complete automation of management procedures and data processing: the size of available databases requires considerable computing power and it should undergo a drastic reduction, in case of a manual analysis or processing. Naturally, techniques search the most effective solution for calculation automation: in particular, some operational problems are substantially solved yet, thanks to use of dedicated software.

The digital image processing consists of analysis and frame manipulation, performed by a computer. The images are represented in a mathematical form as continuous functions $f(x,y)$ of two variables xy, which represent the spatial coordinates of points in the image plane: the luminance function f represents the grey level in every point, namely the intensity of the sensor response to light energy collection.

The image, scanned with digital sensors, is then recorded on the mass memory and it is processed by computers. In order to provide adequate efficiency, computers must comply with certain minimum requirements:

- memory, to bear the image size,
- and display capabilities, because we need to work with the highest possible geometric and radiometric resolution.

Within the geomatic discipline, we mainly work on data sets with spatial reference: each one relates to an object category, which contains the observed values of statistical variables that usually are samples extracted from random variables.

Statistical analysis allows verification and validation of results, obtained from the application of the usual procedure: in particular, the multivariate statistical inference allows the testing of hypotheses, ending with their acceptance - or decline.

Since the gathered data has uneven spatial distribution and irregular temporal trends, it is often necessary to regularize information, through:

- mapping techniques and grilling,
- thresholding, to purify the abnormal data.

Moreover, we can interpret the spatial reference information in the field of Geomatics. Indeed, the topological (geometrical) structure of collected data provides knowledge:

- which could be useful for further clusterization,
- or even for structuring the graphs.

Finally, if analysis refers to many databases, the centring is a must performance, in order to allow comparison. The matching, in fact, solves the correspondence problem among data structures: in vision-related issues, it looks for the correspondence between the image of an object and a model, or among multiple images referring to the same single object.

In digital photogrammetry, matching mainly happens with two common situations:

- in the automatic location of features, signalized points or items of cartographic interest, which will be integrated into the databases of a geographic information system (GIS);
- during the search for homologous points (which is part of a process for relative orientation of images), it intervenes in the choice of tie points and in the stereographic or multi-image restitution.

The classification of matching strategies is based on the abstraction level of the scene description; in other words, it follows the shape of data structure:

- an area-based matching uses a representation of data through the item area (pixel described by coordinates) and its intensity of grey,
- a feature-based matching gives a representation based on points and lines;
- a relational matching provides the description of the data according to the set of geometric and topological connections within objects taken from the scene (primitive).

It is uncommon to obtain a perfect match between relational descriptions of data: it is therefore necessary to select, according to certain criteria, the most likely representation among possible combinations. In fact, the size of the problem itself prevents the evaluation of all solutions: luckily, the search space of possible matches is limited by both number and completeness of topological and geometric relationships.

The problem related to automatic identification of objects can be solved through a system for the recognition of homologous features in images: in order to simplify the process from the computational point of view, it is possible to identify a technique for relational description construction based on the principles of perceptual organization, by assessing together the level of abstraction and the type of available data.
The relational matching thus leads to the possibility of working without a priori information on images; on the other hand, it bears higher computational complexity, which comes from the size of the search space.

The construction of a relational description, based on the principles of perceptual grouping, comes from studies on the perception process for the human visual system: the aim is to implement the ability to detect in a scene the structures which have simultaneously characteristics of continuity, regularity, repetition and symmetry, since these features are rarely accidentally together.

In order to obtain syntactic models recognition, the standard procedure is as follows:

- a selection of primitives comes from the analysis of the ideal model: since they are the basic elements for objects description, they properly constitute the model;
- then, there's the chance to proceed with grammar construction, corresponding to the description of a given class of objects (model);
- pattern recognition occurs through the pre-processing of the object. It typically has two distinct phases:
  - encoding[1] and approximation of the message, because of information compression;
  - filtering and improving information quality: the target is to reduce noise and to diminish damages due to models. During the filtering phase, the most common techniques for segmentation are commonly used:
    - a thresholding mechanism (clustering) operates through the definition of a threshold for a feature (be it statistical or structural);
    - the techniques of edge detection work through the detection of discontinuities: this is a widely spread practice, because region boundaries often collect lots of different information;
    - the recognition of areas (region extraction) comes with a particular study method about sizes, shapes and textures;
- the representation of the model proceeds then through segmentation and subsequent extraction of primitives, in order to represent the model by sub-models and sub-models in terms of primitives;
- and - finally - a syntactic analysis (so-called parsing) is performed: during this phase, a syntax analyser (parser) decides on the membership of a representation in a given class of objects (model), which are described by a certain common grammar.

The identification of an object on the stage happens by researching the correspondence between the image description and the relational model description: all the candidates in the image that resemble the model are taken into account, while solutions acceptance concerns the application of rules and geometric constraints. Moreover, within the hypothesis of invariant relations under a change of the observation point, it is possible to strengthen the search for a solution with the simultaneous use of multiple images.

[1] For instance, binary information is encoded with a sequence of 0 and 1, while a wave is approximated by an expansion in Fourier series (wavelet).

1.2. Cultural foundations of the Computer Science.

Shannon is considered to be the founder of the Theory of Information:

- he analysed the chance to measure and compress information, eventually eliminating unnecessary parts from data,
- he noted the similarities between Boolean algebra and the structure of electrical circuits,
- and he finally developed the Artificial Intelligence.

Information is a measure of available freedom of choice, when choosing a message: for instance, in an elementary situation, when the choice only refers to two possible messages, information equals to one and the quantity of information is defined by the logarithm of the number of possible choices.

According to this approach, complexity is due to the missing information to get the full explanation about system operation. This figure is properly expressed by the *H* function by Shannon and its reliability is inversely proportional to the probability that the system reached its complete organization and operation in a purely coincidental way.
In fact, complexity of a system comes from the number of different elements composing it: thus, the missing information is predominant, while it is unlikely that constitution is due to random assembly of components.
Anyway, it is still unclear how to recognize a complex system from a random one, or how to distinguish information by a complex system from entropy by a disordered set. Nevertheless, every piece of information should add something to the amount of available knowledge.

Information is therefore a process: through knowledge acquisition practice, everybody can recognize utility and necessity of interdisciplinary work. According to Whitehead, the clash of opinions (or theories) is an opportunity and uncertainty is an asset. No language can fully catch the description of the real world: it is too rich to be lit only by a spotlight, nor grabbed from a single point of view.

The Information Sciences were born in the twentieth century, from the cultural gathering of different disciplines, whose common goal is the study of the human mind: Logic, Linguistics, Philosophy, Psychology, Cybernetics, Natural and Artificial Languages, Cognitive Science, Artificial Intelligence, Anthropology and Neuroscience.
The fundamental issue is the nature of knowledge and its representation within the human mind. Cognitive scientists believe in mental representations and consider the computer as an effective model of human mind functioning.

1.2.1. History of Logic.

For most of its history, Logic was traditionally represented by the Aristotelian and Stoic theory; as a result, while Mathematics could decrypt Physics, Logic cannot escape the Aristotelian plant.

Modern logic roots on the idea that the whole thought is expressed through signs and the true language is a set of symbols and combination rules.

Raymond Lully was the first trying to reduce logical thought to calculations, nevertheless results in the conception of an artificial rational language are related to Leibniz' work: in order to allow a symbolic language (*characteristica universalis*) expressing any concept, the alphabet of human thought (*mathematische Schriften*) converts reasoning into calculation and eventually the language in an algebra of thought.

Anyway, the founders of Mathematical Logic were Boole (1847, *The Mathematical Analysis of Logic*; 1854, *Laws of Thought*) and Frege (1884, *Die Grundlagen der Arithmetik*; 1903, *Grundgesetze der Arithmetik*).

Boole assumes, as the fundamental feature of human thinking, the ability to isolate objects' classes and relate them to their names and symbols. The objects of Logic are relationships between classes and their modes of perception: therefore, the laws for generating classes only depend on the structure of intellect, they determine the shape of the argument and they can be mathematically expressed.

Boole expresses the propositions of Logic in a mathematical form and he achieves the mathematical formulation of the non-contradiction principle ($x^2 = x$): therefore, we assume a direction to analysis and classification, based on divisions in pairs of opposites.

The starting point is the analogy between:

- propositional connectives *and* and *or*;
- arithmetic operators + and –;
- and algebra, considering it only allows numbers *0* and *1* (which exactly represent *false* and *true* in a logical key).

Boole detects the analogy between the classes' calculation and the propositional calculus and he derives all the theorems related to traditional Logic.

Frege then constructs Logic in an axiomatic form.

Further development of Mathematical Logic is due to works by:

- Peano (1895, *Formulario mathematico*);
- Russel (1903, *Principia mathematica*);

- Carnap (1934, *Logische Syntax der Sprache*);
- Tarski (1950, *Logic, Semantics, Mathematics*);
- and Quine (1960, *World and Object*).

In his *Tractatus logico - philosophicus*, Wittgenstein argues for the existence of a relationship between the structure of a proposition and the structure of a fact. However, there is no coincidence between proposition and fact, because language cannot fully grasp the world[2].

Every formalization shows its limits in understanding and expressing problems in ordinary language, while there would be no restrictions within a formalized language. The proof of this theorem lies in the Epimenides paradox[3]. In fact, Gödel replaced truth with the concept of provability: this proposition cannot be proved, unless it is false, within a specific formal system. Then:

- provability applies to a false proposition, opposing the assumptions about non - contradictory nature of the formal system;
- or, vice versa, the proposition is true and it's not provable - in this case, the formal system is incomplete.

As a result, we know that not all true arithmetical propositions can be proved within a single formal system:

- Gödel, convinced that the primitive recursive functions are not adequate to the concept of mechanical procedure, develops recursive functions in a broader sense, according to the thesis of the Vienna Circle;
- in the computer field, Turing shows the limits of theorems borrowed from formal systems: in particular, the *halting problem* arises from the inability of the software to decide a priori when to stop.

Hilbert poses three problems:

- presence of contradictions between the axioms of Mathematics (consistency);
- existence of true statements, that are not provable in Mathematics (completeness);
- existence of a rigorous method to determine truth or falsity of a statement (decidability).

---

[2] The thesis of indefinability of reality captivated the Vienna Circle, as it called sensible the only propositions which can provide a defined prescription (logical positivism). Subsequently, Gödel expressed a similar concept in a mathematical form: Mathematics cannot fully express the logical truth.

[3] *All Cretans are liars*. (Epimenides of Crete)

The statement can only be true if inherently false, and vice versa it is false also representing a truth.

In 1931, Gödel replied to the first two, while five years later both Church and Turing independently solved the third.

Actually, Turing did not solve the problem from the logical - mathematical point of view, because he imagined an ideal universal machine, consisting of:

- a sliding paper tape, covered with boxes;
- a pen;
- a memory, which allows the machine to remember a symbol at a time;
- a tool that decides how to intervene in the operation of the machine itself:
  - by sliding the tape towards the right or left;
  - writing on the box;
  - deleting a number of already imprinted previous instructions.

According to Turing, one single machine is enough for each type of operation, provided that it can learn its operating rules: this imaginary mechanism represents the foundation of the idea of the computer. However, even an imaginary machine meets these expectations: Turing proved the existence of the *halting problem*, related to the risk of blocking the computer in an endless process[4].

Thus the era of Computer Science and Artificial Intelligence began, fitting into the story of the dream about creating a thinking machine: it's an idea, concerned with the desire to analyse the human mind in its mechanisms, as if it were a machine (somehow formalizing a correspondence with hardware and software in the computer field).
Even the creation of automata and artificial creatures similar to humans is an ancient dream: it can be found in the Jewish and Greek myths, in the myth of the Golem and in the legend of Pygmalion. From the middle Ages, these myths are spreading throughout Europe, influencing Bacon, Albert the Great, Leonardo da Vinci, the astronomer Müller and all alchemists.

[4] For example, there is the assumption that every even number is the sum of two primes:

- if it is false, the computer it will stop, at some point;
- but if it is true, the computer will never stop.

Turing proved the existence of a group of mathematical statements for which it is impossible to know, a priori, if the computer would have stopped or not. In this way, he gives an original answer to the problem of Hilbert (*Entscheidungsproblem*): it is not possible to decide in advance the truth of all mathematical statements.

Turing ambitiously studied the mechanisms of the human mind, through the construction of a machine: verifying that the new machine has not a soul, he anticipated Artificial Intelligence challenge and fitted both the history of Logic and the alchemists' dream.

In 1642, Pascal built a machine performing additions and subtractions.

In 1671, Leibniz designed a machine performing multiplications.

The real breakthrough is the universal computer:

- in 1837, Babbage imagined a machine to calculate the logarithmic tables, using the technique of finite differences[5]. The project failed, despite the funding raised from the British government;
- Babbage then dedicated to the creation of a universal and programmable machine[6]: the Analytical Engine could do whatever it's ordered to do. Nevertheless, Babbage failed again;
- design only became effective with the project by Turing, in 1936.

The "weak" theory about Artificial Intelligence believes that the computer is a tool able to formulate hypotheses about the workings of mind, in a rigorous and precise way. However, computers do not think: computers are programmed to perform duties of human beings, without reasoning during procedures.
Conversely, the "strong" theory about Artificial Intelligence believes that a suitably programmed computer could understand how programs explain workings of mind. The human psyche would especially express the ability to manipulate symbols, as a computer does: therefore, the chance to program conscious thinking computers would have no theoretical limits, but only time-related problems.

The Turing test is a game of imitation. A man (A) and a woman (B), hidden, should suggest an interrogator (C) that A is B, while C must discover who among A and B is man and

[5] Later, he realizes that the function of squaring natural numbers produces odd numbers of distance 2. A similar situation also occurs for the polynomials: after a certain number of differences, there is a constant value, for which the values of the polynomial are calculated in reverse, as a sum of finite differences.
Since the polynomials allow the representation of many mathematical functions of practical interest, this type of adding machine would be able to calculate most features.

[6] Lady Ada Lovelace, Lord Byron's daughter, founded programming, designing conditional and iterative statements: Babbage's machine would have done different calculations, according to the different conditions, because the universal machine would be able to work with generic symbols in variable combinations, without being limited to numbers.
Although the machine manipulates symbols, it is not intended as a model of human thought, because human mind isn't just a mechanism.

woman, by asking questions of any type. If a car substitutes A and C are unable to deceive, it means that C has capacity for thought.

Conversely, the philosopher Searle argues that if a program, simulating the presence of a person and responding in Chinese, deceives a native Chinese language, it does not mean that the computer understands Chinese, but it only demonstrate the availability of instructions and procedures suitable for controlling the Chinese symbols.
In *Minds, brains and programs*, Searle describes the Chinese room experiment[7]: the computer processes and manipulates data according to a program of instructions supplied from the outside, and even if it completed reasoning, it would end the procedure without understanding it.

However, this example generated an extensive debate, between those who think that an artificial semantics is impossible and those who do not take the answer for granted.
In fact, you do not learn a language according to the instructions coming from the outside, but vice versa representations of an artificial system are not necessarily devoid of meaning: for example, some scholars suggest an interaction between perceptual reality and motor reality, let's say between a system and its environment.

By the way, Searle says the problem does not change even for a robot with an internal computer that controls its movements: in this case, instructions are stimuli, reaching the computer from a system of artificial vision – and they shall be classified as symbols.

Conversely, according to the philosopher Marconi, the distinction between the processing of light signals on the retina (operated by the visual cortex) and the one referred to light signals on a camera (operated by the computer) is critical.

There are many objections to the Turing's hypothesis:

- there are theoretical limitations to what the discrete state machines can compute (mathematical objection);
- the functioning of the brain is continuous, so the imitation by a discrete state machine would be very unlikely (biological objection);
- to think is equivalent to having self-awareness, as a thinking being, and this is hardly deducible from Turing's experiment (philosophical objection).

[7] In a room with two windows, sheets with Chinese characters come in: knowing Chinese language would be the only chance to decrypt them. From the other window, incoming sheets have instructions in a known language.
Then, when it's possible to match the new information (from the second window) to unintelligible characters (on the first sheets), you can simulate the response, by using the process suggested from the outside, even without mastering the mechanism.

The famous (though disputed) Weizenbaum's Eliza program, while not claiming to deceive his audience, is able to converse with them, provoking reactions.

According to Turing, the critical assumption to build thinking machines are born from a sense of discomfort (likely, a legacy of faith), which gives human beings a privileged place in the universe. When you question whether a machine could think, the judgment is the result of the computation, rather than the creation of the machine. In fact, if the machine behaves in a certain way, it actually thinks: that is, the behaviour of a machine can be influenced by external parameters.

Subsequently, automata developed to robots[8].

However, instead of being afraid of the machines, the human being is often ashamed of his humanity, which is less considered than the perfection of things[9].

In 1997, even the chess player Kasparov was defeated by a machine, called Deep Blue: the shame is scorching.

[8] The term *robot*, introduced by Karel Capek (1920), derives from a Slavic root that indicates the work related to physical effort, often forced (*rabota* is the Russian word for work, while *rob* means slave in various Slavic languages).

[9] The philosopher Gunther Anders announces the birth of a new feeling: shyness toward the power of the machines.

1.2.2. Cognitive Sciences.

The definition of cognitive science is still inaccurate and, in fact, it corresponds to the actual instability of the discipline: there is not yet a complete summary of specific and original techniques in this area of Information Technologies.

Without demonstrating that the Artificial Intelligence is impossible, Gödel believes that the mind doesn't exclusively resides in the brain: this situation resembles what is in a machine, such as a computer.

The Cognitive Sciences arise when imagination, design and practical realization of computers make sense, dream or hope for the possibility of Artificial Intelligence: they are formed by an array of disciplines (Computer Science, Linguistics, Psychology and Neuroscience), which try to cope with the problem of the nature of knowledge, through the history of Philosophy.

The problems faced by the cognitive sciences are many and the answers have not (yet) a unitary form:

- can operations of machines and biological systems be traced back to common patterns, aimed at a purpose?
- when a scientific explanation is acceptable?

The answers come from Cybernetics and by Artificial Intelligence: the recurring names are those of Wiener and Turing, with the concepts of feedback and computation, algorithm and formal system.

The concept of reasoning as computation[10] shows issues concerning three types:

---

[10] The idea that reasoning is a calculation is ancient:

- reasoning is calculus, as catching things at a glance, or know a picture by subtracting one thing to another (Hobbes, 1651);
- then Leibniz extended the concept, through the analysis of the reasoning mechanisms and symbolical description of the universe: out of formalization, he called the universal characteristic a philosophical calculation. Leibniz defined calculation as the deductive process that infers, without facts, such as the proceedings of Logic and Mathematics: it's a calculation similar to that relevant to common Algebra, in which reasoning is deduced in from axioms, through definition of logical operations and relations, by symbols;
- Descartes, Pascal and Locke believed that the rules of syllogism, without heuristic value, were useless in the discovery of truth. The syllogistic reasoning is artificial - and in fact all human beings whom are capable of reasoning reason correctly.

- interaction between matter and mind (philosophical problem);
- relevance of meanings (semantic problem);
- and empirical verifiability of explanations (epistemological problem).

Descartes relies on the doctrine of metaphysical dualism: body and entity are different substances, while the problem is the interaction between the two.

The alternatives to dualism were:

- idealism, according to which ideas, spirit and mind are actually essential, while material objects are illusory or derivatives;
- and materialism, according to which everything (the mind, too) is matter[11].

The research of cognitive science investigated the correctness of human reasoning: so far, the cognitive experiences have not definitively refuted the claims of Descartes, nor completely accepted Leibniz, nor Mill, which supports education to Logic and familiarity with the principles of correct reasoning.

Hume, in his *Treatise of Human Nature* (1734-1737), describes the laws that govern thoughts and feelings in a mirror to what Newton did for the motion of physical bodies.
According to Hume, ideas are similar to physical particles: their interactions are determined by natural laws and there's no need to explain how they become thinking, as they're rational.

For the first time, he doubts that observations of objects and events (either in the past, either present) can suggest anything about objects and events in the future. The problem of induction only has a sceptical response: between *to be* and *ought to be*, or between a descriptive and a prescriptive proposition, there is a leap of logic.
Even some experiences of the cognitive sciences scratched the presumption of a natural ability to inference.
For example, the thesis of cognitive penetrability identifies vision and, more generally, any kind of perception with tendency towards categorization (propensity to establish membership in a class for each object seen): only a part of the data really comes from the senses, while the rest is due to the remaining cognitive system.

In fact, for centuries, the test of any scientific theory lied in the same facts and observations, which should be neutral with respect to theories; however, since the last century, it is

---

[11] Even in this case, however, there is a contradiction: the reasoning is a manipulation of symbols with sense, according to rational rules. If the manipulator of symbols considers the meanings of the symbols, the reasoning is not just mechanical, because the meanings do not respect physical laws. Conversely, if it does not refer to their meanings, the manipulator does not provide an argument, because the reasoning cannot be separated from the meanings.

observed that the observations are steeped in theory (*theory-laden*, according to studies by Popper, Kuhn and Feyerabend).
According to Kuhn, the type of observations of each paradigm is internal to the paradigm itself: in other words, an observer makes use of data provided by the paradigm in which it operates, without considering other kinds of information.
Hence, for Popper, scientific theories cannot be verified, but refuted only.

Finally, the cognitive sciences report to a universalistic conception of the human being, against cultural relativism, with a significant impact on the level of knowledge and that of human relationships.

However, the cultural divide between cultural man and the natural man is not yet fully absorbed: it’s because the cognitive science properly deals with the mind (the seat of algorithmic processes), while the operating model has been proposed for both the human being and the computer.

1.3. Connections between Geomatics and Cognitive Sciences.

Geomatics is the name of a new science, that deals with features and structure of spatially referenced information, either time-varying or not. It's inclusive of the methods of acquisition, organization, classification, processing, analysis, management, return and dissemination (availability to knowledge).

The metageomatics is devoted to the study of Geomatics and refers to the disciplines of epistemology and cognitive science.

With the birth of Geomatics, the disciplines related to survey and detection successfully met the human sciences: various communication problems highlight a possible bridge between the disciplines of surveying and human sciences, in particular in the field of quality measurement and management.
The Linguistic discipline, for example, contributes to the definition of the fundamentals of statistical and numerical analysis upon geo-referenced data (images, maps and 3D models).

The human sciences, despite the extent of the description (replaced by processing, borrowed from the characteristics of mathematics), pioneered the study of language, communication, social relations and media. The path followed development and progress of both the world and the society, allowing the disciplines to overcome the technological determinism of an engineering theory about information.

In fact, language is a human characteristic that allowed open communication about each content type: in different ways and situations, the ability to speak encouraged the construction of diverse social relations, in which information is always essential.

In this context, there are:

- information systems with spatial reference data (information are both geographic and territorial, and can be either dependant or even independent from time variable);
- images (from digital images processing, to computer graphics, including 3D models and multi-dimensional colour maps);
- the development of information infrastructures (Internet), including networks of mobile phones and system GNSS (Global Navigation Satellite Systems)

Observations management (which is an elaboration technique near to computational statistics and data analysis) points out that part of the open issues in the disciplines related to survey and detection does not belong to mathematical problems, but they're rather logic (according to Boole, Peano, Tarski and Quine).
It's important to take into due account the limits of original theories and especially Gödel's incompleteness theorems, in response to the claims of completeness of Hilbert.

Between cognitive sciences, both Linguistics and Psychology (*Gestalt*) offer interesting contributions to problems about segmentation, classification, clustering and interpretation. In fact, the study of human language and also studies about natural languages are setting the pace to analysis of maps, images and 3D models.
Starting from the comparative method of Von Humboldt, throughout the distinction between signifier and signified (De Saussure), via the discovery of phonemes and morphemes (Trubetzkoy), the study of syntagmatic grammars (Bloomfield) or transformational and generative grammars (Chomsky) is fundamental for trying to define primitives and their models towards grammatical and syntactic association.

A brief digression on some particular aspects of information science and technology is shown below: an interesting case of this cultural fusion is the recognition of syntactic patterns, which realizes the comparison of parts of maps, images and 3D models with archetypes (parsers).

The methods of Cluster Analysis, the use of Graph Theory, techniques of Operational Research (Mathematical Programming, Grammars, Syntax and Algebras, Decision Theory), the shape descriptors: they all constitute attempts at validation, modelling and interpretation of data. The identification of a single central nucleus is complex: any procedure is applied to increase the *a posteriori* probability, with respect to the probability (either conditioned) that was assigned *a priori*.
In terms of informational variation, the other dispersion indicators must prevail on the residual variation. Bayes' theorem sets these principles, taking shape as an instrument for control, without giving any operational guidance.

As parts of the Information Technologies, both Data Analysis and Processing of Information circumscribe and define the message. Subsequently, they interpret it, including:

- in the first phase, outliers exclusion, by means of robust procedures which delete gross errors and irrelevant information;
- then, in the process of understanding the data, methods and procedures aiming to group, sort, dissect, construct vectors and finally formalize the information, which must be as much as possible clean and proper.

Where the model is clearly defined, to analyse and to process the data means to refer to the Approximation Theory or to least squares techniques.

On the other hand, within the Information Technologies field, practices are different: procedures are coming from Discrete Mathematics, Artificial Intelligence and Linguistics.
The same techniques are supporting Texture Analysis and Pattern Recognition, which are methods for interpreting and understanding images, maps, drawings and 3D models.

The Geomatic discipline is a patchwork of methods, strategies and processes: unfortunately, its premises are still poorly structured from the strictly theoretical point of view.

The focus of all these techniques (that are mathematical, statistical, numerical and computer-related) seems to be on the ability to see the world around them from a certain point of view, such as catching it with a look (as it appears in the vision field): instead, we should understand that the very potential of these practices is to connect data which, under certain assumptions, respond to the concepts of proximity or similarity.

1.3.1. Mechanisms for vision.

Seeing means to grasp the relationship between the objects in a context, eventually appreciating the connections, with reference to the structure[12]. The Psychology of Form (*Gestalt*) observes the interactions between the two main functions of the mind:

- collection of information;
- data processing.

According to the Gestalt, perception of form is the basis for formation of concepts, because structural elements are identified in the raw data, through the distinctive features.
Perception is then an adaptation of rough information, as caught by the senses, to a model (pattern) of a simple form, which Arnheim called visual concept (or visual class).

Moreover, according to cognitive scientists, understanding only exists when it is possible to reproduce the perceptual phenomena: this concept incorporates the observations of Bacon, about knowledge providing the ability to produce or reproduce objects.
In fact, research in the cognitive sciences assigns paramount importance to the eye, between other perceptual skills: visual activity is considered equivalent to cognitive activity[13].

Actually, the ability to perceive objects is a particular form of troubleshooting, based on two kinds of mental processes:

- top - down processes, using existing knowledge and subjective expectations to direct attention to the significant qualities of the stimulus, to interpret information and finally to recognize the object;
- bottom - up processes, which identify and organize the quality of the stimulus while recording sensory information. In this case, the processing must incorporate all the features of the perceived object, to allow viewing of the whole.

---

[12] The Greek philosophy admits the separation between senses and thinking (perception and reasoning), even not categorically: also Aristotle stated the soul never thinks without a picture.

[13] Perhaps, the paramount importance assigned to the eye is typical of Western culture:

- the art historian Stoichita assigns this responsibility to the Platonic myth of the cave: if men would have wanted to eat or drink, instead of know or see, other perceptual skills (touch, taste or smell) would have been preferred;
- also, the myth about the origin of the art, by Pliny the Elder, defines a human being with a shadow line: the painting is a symbol of the art and the copy from the truth is the founding act of representation, completing the equivalence between vision and cognitive act.

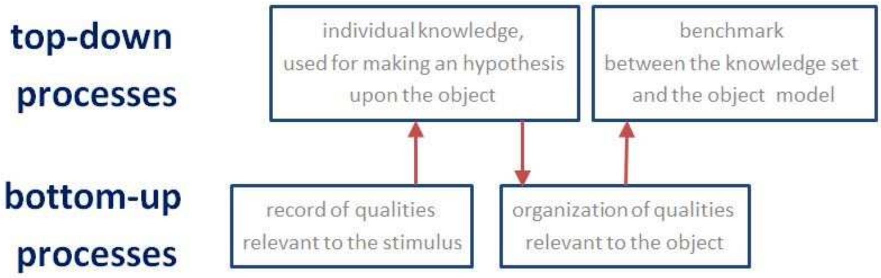

Figure 1.1. Model for elaboration processes.

The scientists, wanting to build computers that are really able to recognize objects, soon discovered that bottom - up processes cannot work alone. In order to ensure the correctness of their conclusions, in fact, it is necessary to provide the system with additional information of at least two types:

- we must allow the machine to decide what features to combine,
- it's fundamental to allow the software distinguishing objects from their background.

Therefore, it is necessary to integrate the bottom - up perception processes with elaborations of data proceeding top - down, which compare the integrated features with the information previously available (the model, stored in memory), eventually allowing confirmation of the hypothesis.

In addition, please consider that the conscious perception of an object begins when it is visible as a complete whole: the earlier stages of registration and integration of features must be managed and completed before.
The conscious perception is thus a holistic process, because the perception of the whole prevails over its parts: in other words, there is the tendency to perceive patterns, objects and scenes in their entirety, eventually ignoring their specific components.

Finally, along top - down processing, the holistic perception of an object or a scene contributes to identify the individual components.

1.3.2. Gestalt principles.

Historically, the concept of holistic perception stems from the research of a school of thought (*Gestalt*[14]) founded in Germany in the early twentieth century, as a reaction to the current of structuralism, then dominant in psychology.

Structural scientists sought to identify the constituent elements of sensory experience and believed that any perception could be interpreted as a combination of its components.

According to the followers of Gestalt, the perception does not arise from the combination of distinct elements, but just resides in the immediate response to complex patterns, caught in their entirety. In fact, the entire object is not perceived as the sum of its constituent parts, but as a complete and autonomous structure: even if it is composed of different parts, the perceived object is made up by the specific way in which the parties are mutually organized.

The operational principle of Gestalt psychology is that the brain is holistic, parallel and analogic, with self-organizing tendencies.
The human eye sees objects in their entirety before perceiving their individual parts, suggesting that the whole is different from the sum of its parts[15].
Further, the whole is anticipated, when the parts are not integrated or complete.

Therefore, the useful information lies precisely in the organization: *Gestalt* properly means configuration and, according to its proponents, the basic unit for the study of perception is the structure of the sensory stimulus and it’s not composed by its individual elements.

Gestalt psychology tries to understand the laws of human ability to acquire and maintain stable perceptions in a noisy world.
Gestalt psychologists stipulate that perception is the product of complex interactions among various stimuli. In particular, the gestalt effect is the form-generating capability of human senses, with special respect to the visual recognition of figures and whole forms, instead of just a collection of simple lines and curves.

---

[14] Gestalt psychology or Gestaltism is a theory of mind and brain of the Berlin School: in German, *Gestalt* stands for "essence or shape of an entity's complete form".

[15] *The whole is other than the sum of the parts*. [Kurt Koffka]

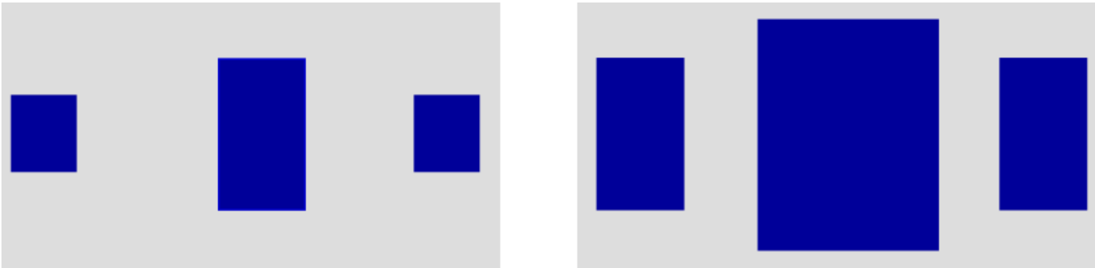

Figure 1.2. Example: the importance of contest (three rectangles are identical).

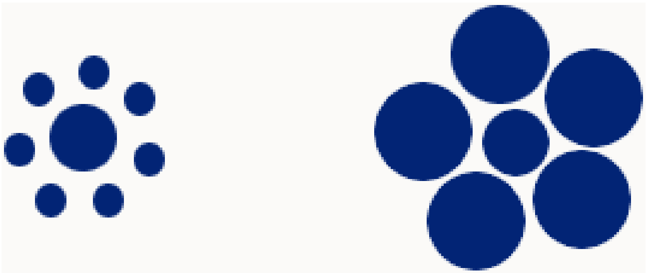

Figure 1.3. Example: the importance of contest (two circles are identical).

The nervous system is designed to collect the building blocks of sensory stimuli, throughout innate mechanisms, based on some basic rules which are referred to as the principles of perceptual organization.

The perceptual organization was summarized by the Gestalt principle of meaningfulness or pithiness (*Prägnanz Gesetz*), which says that we tend to order our experience in a manner that is regular, orderly, symmetric and simple. In other words, it sums up the three phases of the theoretical evolution of the human vision phenomenon:

- the base is the ability to proceed from analysis of a set to inspection of its parts;
- before alternatives, each individual chooses the simplest interpretation;
- the measure of the significance of a grouping relates to the degree of non-randomness of the structure.

A major aspect of Gestalt psychology is that it implies that the mind understands external stimuli as a whole, rather than perceiving the sum of their parts; the wholes are structured and organized using grouping laws.

Gestalt psychologists attempt to discover refinements of the law of *Prägnanz*, writing down laws that theoretically allow prediction of sensation interpretation.
The various laws are also called principles, depending on the paper where they appear: however, they all deal with the sensory modality vision.
The visual Gestalt principles of grouping were first introduced in Wertheimer (1923).

In fact, according to the Gestalt psychology, there are six factors influencing the grouping of objects during visual perception process:

1. the law of proximity states that an individual tends to perceive an assortment of objects close to each other as a group. This mechanism is often used in advertising logos, to emphasize which aspects of events are associated.
   The principle allows the distinction of objects groups in a set.

   For example, in Figure 1.4, you see three points-set, instead of 36 identical elements;

Figure 1.4. Example: law of Proximity.

2. the law of similarity states that elements within an assortment of objects are perceptually grouped together if they are similar to each other. Similarity can occur in the form of shape, colour, shading or other qualities.
   This principle allow distinguishing two objects that are adjacent or overlapping, according to the differences of texture in the image.

   For example, the following Figure 1.5 illustrates circles and squares: in this depiction, 20 circles are shaded blue, 4 circles are shaded grey and 4 squares are blue. We perceive the grey circles and the squares as grouped together, forming two horizontal lines within the square of circles. This perception of lines is due to the law of similarity.

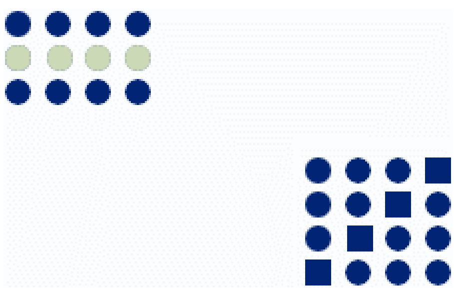

Figure 1.5. Example: law of Similarity.

3. the law of closure states that individuals perceive objects such as shapes, letters, pictures, etc., as being whole even when they are not complete. Specifically, when parts of a whole picture are missing, our perception fills in the visual gap. A research shows that the reason the mind completes a regular figure that is not perceived through sensation is to increase the regularity of the surrounding stimuli.
   This principle helps to perceive shapes as complete, even when they are hidden.

   For example, in Figure 1.7, we perceive a closed circle; however, there's a gap in the shape. If the law of closure did not exist, the image would depict an assortment of lines with different lengths, rotations, and curvatures — but with the law of closure, we perceptually combine the lines into whole shapes.

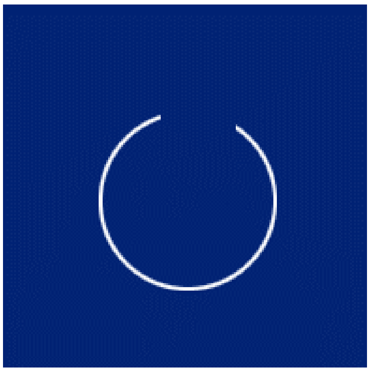

Figure 1.6. Example: law of Closure.

4. the law of continuity states that elements of objects tend to be grouped together and therefore integrated into perceptual wholes if they are aligned within an object.
   In cases where there is an intersection between objects, individuals tend to perceive the two objects as two single uninterrupted entities. Therefore, stimuli remain distinct even with overlap. We are less likely to group elements with sharp abrupt directional changes as being one object.
   This principle allows deciding to which object a certain line belongs, in case of overlapping objects.

   In Figure 1.8, we see two continuous lines, *ab* e *cd*, instead of four segments or two broken lines.

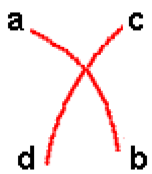

Figure 1.7. Example: law of Continuity.

5. the law of common fate states that objects are perceived as lines that move along the smoothest path. Experiments using the visual sensory modality found that movement of elements of an object produce paths that individuals perceive that the objects are on. We perceive elements of objects to have trends of motion, which indicate the path that the object is on.

   The law of continuity implies the grouping together of objects that have the same trend of motion and are therefore on the same path; this principle allow distinguish a moving object on its background.

   For example, if there is an array of dots and half the dots are moving upward while the others are moving downward, we would perceive the upward moving dots and the downward moving dots as two distinct units.

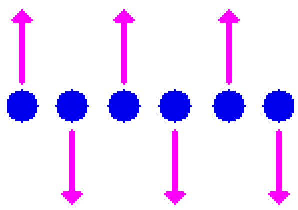

Figure 1.8. Example: law of Common Fate.

6. the law of good gestalt explains that elements of objects tend to be perceptually grouped together if they form a pattern that is regular, simple, and orderly.

   This principle is less specific than the others and includes the complex rules according to which the perceptual system organizes the stimuli. In detail, the norm implies that as individuals perceive the world, they eliminate complexity and unfamiliarity so they can observe a reality in its most simplistic form. Eliminating extraneous stimuli helps the mind create meaning, implying global regularity, which is often mentally prioritized over spatial relations. The law of good gestalt focuses on the idea of conciseness, which is what gestalt theory is based on.

   As an example, please find Figure 1.10: we perceive a triangle on a three-circle base, as the superposition of two figures.

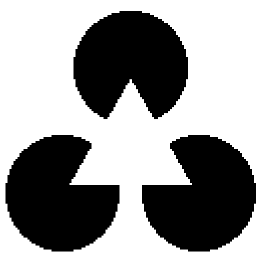

Figure 1.9. Example: law of good gestalt.

In 1958, Hocberg added a general principle of simplicity (the so-called minimum principle[16]): the perceptual response to a stimulus just requires the minimum amount of necessary information for its full description.

In addition to formulating the principles of perceptual organization, Gestalt psychologists noticed the automatic tendency to differentiate into any scene:

- the figure, identified as the object that attracts attention,
- the background, which is the field on which the figure stands.

The reading of the image is allowed by the distinction between figure and background, which is not arbitrary, but depends on the characteristics of the visual stimulus: under equal conditions, the background is perceived as the form that tends to circumscribe another.

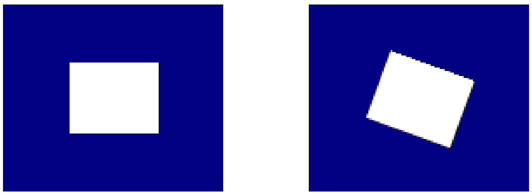

Figure 1.10. Example: figure and its background.

However, in the scenes when the clues are scarce or ambiguous, the mind may find it difficult to decide what form give the meaning of the figure (and the same happens to the background), because the organization to be assigned does not appear clearly defined.
This kind of phenomenon is illustrated by the so-called reversible figures: given that the same portion of the scene cannot be perceived simultaneously as background and as the figure, we perceive the object alternately in one form or another.

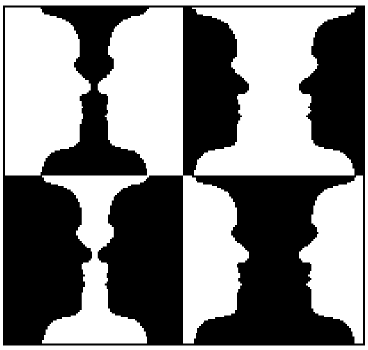

Figure 1.11. Example: reversible image.

[16] The idea of simplicity and essentiality was linked to the principle of minimum length for encoding in the data transmission field, which was formulated by the information theory, during the same years.

Perceptual illusions arise from the context in which the items appear unrealistic and they can be explained by the intervention of top - down process.

The scientists are particularly interested in illusions: sometimes, the context influences the kind of features that are perceived in the object. Furthermore, some illusions suggest seeing something that is not actually present in the stimulus.

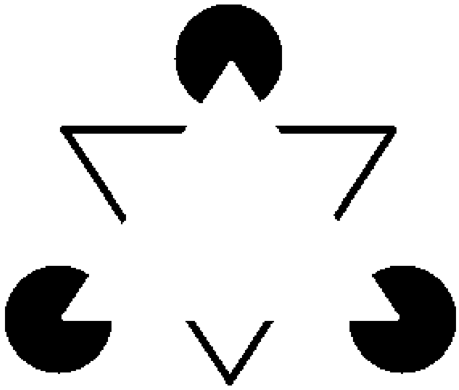

Figure 1.12. Example: illusory contours (Kanizsa triangle).

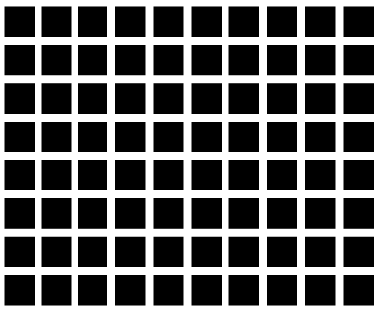

Figure 1.13. Example: grid.

1.3.3. Object modelling.

The object modelling is an innovative and technologically advanced technique in the study of phenomena and their evolutionary dynamics, which allows extensive mode of representation for bodies and figures.

The relationships between the characteristic elements (points, lines, surfaces and 3D objects) determine the validity of the modelling, while their sets have correspondences with the groups of symmetries in the formal spaces of algebra.

A methodology for object modelling runs through four levels, reflecting a top - down approach, as it proceeds by successive refinements:

1. the external level describes the entities involved, by using a natural language;
2. the conceptual level corresponds to the synthesis of all external models: since it is difficult to integrate them all in a single conceptual standard, it is appropriate to choose a unique (geometrical) reference and then transform other representations according to the chosen geometry.
   Moreover, even if the conceptual model could be described by sentences in a natural language, it is preferable to adopt the formal instruments of the approach entity – relationship, in order to simplify the transition to the next level;
3. the logical model corresponds to the implementation of the conceptual standard, because it organizes the data according to hierarchical structures, space frames or relational arrangements: databases are built with strings and tables, while their properties are specified by creating an entity dictionary, indicating attributes and relationships;
4. the physical model actually is the implementation of databases and their structures: it defines the pointers and all information needed, in order to effectively manage the data.

A paradigm of this type provides a high level of abstraction from original data physical structure, because it models the system according to objects definitions, rather than using attributes or relationships. Consequently, such a representation would allow the creation of complex structures (objects of objects), which nevertheless are designed as individual: this feature eliminates the need to explain all geographical and semantic features of an object.

Please find below some other features of the object modelling approach:

- ease of description of new types of data,
- possibility of defining operations on new objects,
- ability to structure the objects in a hierarchical manner.

Furthermore, the introduction of specific procedures for processing the data also resolves the main problems related to information management:

- storage, intended as archiving,
- and retrieval, as to say querying the system.

Moreover, for management of data with spatial reference, it is also necessary to create more complex entities and to define manipulation methods. In this context, we should perform the typical operations for processing space - time information:

- georeferencing (transformation, projection, centring),
- representation (interpolation, reconstruction, description),
- selection (classification, aggregation, generalization).

An object can be described as a conceptual entity, with its own context and state, and it's represented by values of some local variables, plus a set composed of operations and methods.

Each object can be classified and grouped in a class, which defines the type of the object as a function of the characteristics (described by specific variables): then the classes form sets of objects with common geometrical or thematic aspects, identified by a name.
The attributes list of a class shows the names of the features for which the objects have a value: then, each object is associated with a table that contains a value for each attribute in the list.

If multiple classes have variables and methods in common, you can define a super - class that groups them: by convention, the universal object is the root of the system and it's the first level of a hierarchical rooted level structure.

Classes and objects are defined so that their correspondence is bijective: this request is consistent with the convention of mutual exclusion between classes.
Such condition appears to be far restrictive, but it allows the construction of a simple data structure, by requiring a many-to-one relationship between objects and classes.

An object-oriented data structure can be obtained by introducing a link between the elements with spatial reference and their morphological characteristics: a three-dimensional body, in fact, can be identified by its geometric characteristics and themes.
In accordance with this condition, therefore, it is necessary to impose two requests for the construction of the formal data structure:

- each object must be associated with an identifier (name or number),
- each identifier must have a connection to the attributes of the object.

1.3.4. Automatic pattern recognition.

Alternatively to the vision, another type of strategy (also, derived from the cognitive sciences) recurs to the language and to the use of grammars: this approach can also be applied in the geomatic field. In fact, when you define a grammar, the recognition mechanism (parsing) establishes the membership of a string in a given language.

The recognition of a scene (or a model) derives from the availability of:

- a grammar, that can describe images with a visual language,
- and a procedure of syntactic analysis, which can determine the membership of visual objects to a model.

As a linguistic phrase can be formed by verb and noun, so a visual object (or a model) can be broken down into parts (or sub-models), further divided into primitives, arranged in two-dimensional or three-dimensional space. Therefore, there is an analogy between the hierarchical structure of the models and the syntax of the language: the focus may be just a typical hierarchical structure of language and, according to some researchers, even the thought underlying the linguistic ability.

In both approaches, respectively based on the vision or the language, the concept of difference (or similarity) is present at different levels within the geomatic process: it involves the Cluster Analysis, relational matching and parsing, when phenomena of ambiguity occur.

1. Clustering means dividing a certain set of objects in a number of groups, with internal consistency and isolated from the outside. The data (objects) are similar to each other if they belong to the same cluster, whereas dissimilar when they belong to different clusters. The grouping of points is often linked to the classification that the human eye naturally applies when considering areas with a high density of information, separated by regions of low density.

2. In the matching process, the concept of similarity identifies the similarities between an image (or a model, or a part thereof) and its elements. According to the conventional classification, matching strategies differ according to the level of abstraction into compared data.

   The relational matching compares complex descriptions of the object and the relationships within it, replicating the strategy of a human eye with respect to similarity.
   Relationships are represented by an associated graph and the similarity between graphs is measured by construction and comparison of labels associated with its nodes.
   Analyses proceed through:

   - number and type of reports (collinearity, parallelism);

- and rank of the node: the parameter identifies the number of sides converging in a single node. It's a reliable characteristic of complexity, since both the concept of distance and the hierarchical tree are functional to the purpose.

The distance between objects inserted in a given space must obey certain rules:

- defined positivity: $D_{ij} > 0$
- symmetry: $D_{ij} > D_{ji}$
- validity of the triangle inequality: $D_{ij} + D_{jk} > D_{ik}$ e $D_{ij} - D_{jk} < D_{ik}$

The common choices are:

- the Euclidean distance (L_2 norm);
- and the Manhattan distance (L_1 norm).

The cognitive operations, related to the perception of visual patterns, have higher order of connection for resemblance: therefore, the hierarchy of the composition order determines the details of the groups in the total pattern, because the pure resemblance is a cohesion factor only if it's supported by the structure of the context.

In fact, the act of perception is related to the classification of a phenomenon under a visual concept: the operation is intellectual (or cognitive), because perception and recognition are intimately connected. Thus, when the perceived object is not clearly defined nor it resembles a mnemonic picture, the ambiguity in the perceived information suggests several formal patterns to the observer.

Many problems, associated with ambiguities and noise, are typical methods of Pattern Recognition. The highest stage of understanding the image covers complex grammars and syntaxes, such as context-sensitive and constraint-free grammars.

1.4. From Linguistics to image analysis.

There are seven levels of reality:

1. the living reality, that is the directly perceptible world;
2. the imagined reality, which is not directly perceptible, but derives from personal experience of each one;
3. the dreamed reality, with its own dreamlike and abstract dimension;
4. the represented reality, through many different styles (mime, music, etc.);
5. the narrated reality, through in any form (either orally or in writing);

Tre casettine
dai tetti aguzzi,
un verde praticello,
un esiguo ruscello: rio Bo,
un vigile cipresso.
Microscopico paese, è vero,
paese da nulla, ma però...
c'è sempre disopra una stella,
una grande, magnifica stella,
che a un dipresso...
occhieggia con la punta del cipresso
di rio Bo.
Una stella innamorata?
Chi sa
se nemmeno ce l'ha
una grande città.

Figure 1.14. *Rio Bo*, by the Italian writer Aldo Palazzeschi.

6. the depicted reality, in any form (painted, carved, built);

Figure 1.15. Illustration, inspired by *Rio Bo*, by an anonymous painter.

7. the recorded reality (for example, through images), that is the portion of reality described with the maximum realism.

Similarly, images can have different forms:

- official images, for ceremonies;

Figure 1.16. Example of an official image: the class photograph.

- technical images, that are devoid of spurious elements or vital ones;

Figure 1.17. Example of a technical image: a picture of the class.

- images that are typical about everyday life.

Figure 1.18. Pictures taken in a class (amateur images).

Finally, it is necessary to build a conceptual parallel between technical images, maps and texts (they all are referred to as written productions, in general):

- the correspondence between technical images and maps (either geographical, either topographical) is given by the closed legend of used symbols: the title block suggests the association between symbols and concepts. It therefore determines the ability to decrypt and interpret all information, that's visible in the representations.
- vice versa, the correspondence between technical images (or maps) and text lies in the formal description of representations: usually, we have a direct impersonal speech.
  In particular, please note the main use of verbal expressions of the same type:
  - we use a present simple, or a past simple[17]
  - typically in the intransitive form,
  - without using modal auxiliary verbs, neither applying inchoative forms, nor frequentative ones, etc.

[17] In fact, we exclude the chance to describe an image in the future, unless for a fictitious and artificial translation of present.

1.4.1. Syntactic analysis.

Parsing is a tool created in the linguistic environment and borrowed from the Cognitive Sciences. It is an interesting case of cultural contamination into the Information Science and Technology: comparison of parts of maps, images or 3D models can happen through archetypes (parsers).

Literally, it is the collection of methodologies of Linguistics, for grammatical and logical analysis on language and texts. The framework is applied in the context of artificial intelligence and it is being integrated in image processing, expert systems, machine learning and vision in robotics engineering.

The syntactic analysis, as to say the process of recognition of parts in a sentence, string or sequence within a given grammar, searches a derivation for a given string. It can be interpreted as an array of non-terminal symbols (hidden states) of the grammar, combined together with terminal symbols of the sequence (or string or phrase): in other words, it could be a Viterbi alignment between hidden states and observations.

The process leads therefore to a syntax-based pattern recognition system, which runs through successive stages:

1. analysis of the ideal model, that involves:
   - selection of primitives, constituting the basic model, which are fundamental for the objects description,
   - and the construction of a grammar, which must be perfectly suitable to describe a given class of objects (model)
2. the creation of the corresponding grammar, that shall be able to describe a particular class of objects (model);
3. actual pattern recognition, in three steps:
   - pre-processing of the object: the information is encoded and approximate, sometimes even compressed. Subsequently, a filtering allows to improve their quality, reducing noise and degradation of the models;
   - representation of the model: the process consists in segmentation and extraction of primitives, to describe the model through primitives;
   - syntactic analysis (parsing), to decide the membership of a representation of a class of objects (model) described by a grammar, or to determine the correctness of the proposed syntax.

Implicitly, we already compared parsing to the analysis of a scene. In fact, the syntactic process of pattern recognition can be regarded to as the transfer function of a linguistic method into the field of vision, because it detects the eventual existence of an analogy between the hierarchical structure (tree) of the models and the syntax of the languages:

- the models are made up of sub-models, variously combined, as the phrases are formed by words,
- and primitives make up the sub-models, such as words are formed by the concatenation of syllables and letters.

A linguistic phrase is formed by verb and noun. In the same manner, an image (or picture) or a visual object (or 3D model) can be broken down into parts (or sub-models), which are further divided into primitives, variously arranged in space (2D or 3D).

The key issue is the combination rules - representations, built as similar as possible to how the mind represents syntactic structures. Consequently, a syntactic structure is hierarchical and it resembles a tree: therefore, the entire method is very similar to the process of vision.

The combination rules for objects primitive are grammars, derived from the language that can describe the model. Naturally, the syntactic approach is effective when it handles complex objects with a few primitives and re-writing rules: in other words, the descriptive power of a language is inversely proportional to the complexity of the recognition process.

In order to recognize a scene (or a model), we shall manage:

- a grammar, which describes a visual language,
- and a method of syntactic analysis, able to determine the membership of a certain visual objects in a model. The procedure is described by the grammar, in order to build a formal representation of objects in a scene (or a model) and their relationships.

The parsing checks if a sentence (or string) belongs to a particular language; in other words, it checks the chance of derivability from a specific grammar. At the same time, the parsing generates the derivation tree of the string.
Such a syntax-based system for pattern recognition is evidently a comprehensive methodology of many procedures:

- recognition of the closer neighbourhood,
- robust processes of separation and union,
- sequential cycle of procedures repetition,
- processes based on the hierarchy of grammars and their corresponding languages.

Several algorithms are classified in different types, according to the needs of grammars to which they relate:

- one possible approach relies on the matching between the parsed string and the reference strings,
- one further extension deals with the comparison between graphs of the string and the graphs related to some reference strings,
- for type 3 grammars, we can usefully apply finite state automata mechanisms,
- while for context-free grammars, we usually use pushdown automata.

The parsing methods may be:

- top-down type, when the starting point is in an initial symbol; through replacements, we eventually obtain terminal symbols only. In this way, we achieve the fitting of the phrase with an operation of good fit,
- or vice versa, bottom-up type, when the starting point is a given string and the aim is to return back to the start symbol.

The first step in the construction of grammars is to determine the structure of a single object: in the syntactic description of the image (or picture) or object (or 3D model), you must choose a set of primitives.
Let's express some examples:

- for language, primitives are phonemes;
- in writing, primitives are characters;
- for models and visual languages, the choice of atomic elements is arbitrary and strictly dependant on the figures that must necessarily belong to the unit of information, with respect to specific needs[18].

[18] For example, if the size (or the shape, or position) is important in the recognition problem, the primitives must contain information about these parameters.

For problems related to the automatic detection of rectangles, it's convenient to choose as primitives:

- an horizontal oriented segment,
- and a vertical perpendicular (turned 90°) segment,
- then an horizontal segment (turned 180°, therefore oppositely oriented)
- and finally a vertical segment (turned 270°).

Naturally, all sections shall be used with standardized length (they all shall be equal to one single unit).

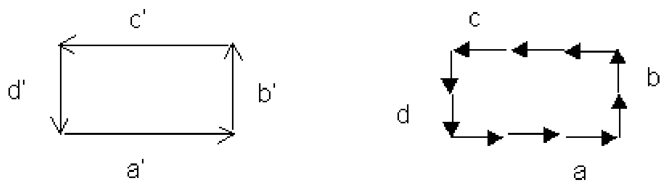

Figure 1.19. Example: automatic recognition of rectangles.

Lots of primitives have been proposed, for different application fields.
In general, the concatenation is a sufficient relationship for one-dimensional use – and it's just the case that is usually applied in Linguistics.

<phrase> = <noun phrase> + <verb phrase>

Figure 1.20. Concatenation, in the linguistic field.

Use of grammars in two-dimensional scope leads to levels without a natural order: in fact, relationships are relational-type (*at the top*, *at the bottom*, *on the left*, etc.).
By the way, complexity grows accordingly in the 3D space, too.

There are different ways to define operators, suitable for expressing relations:

1. for instance, an object could be analysed through its boundaries:

quadrangle = line + line + line + line

Figure 1.21. Representation of a quadrangle, by concatenation.

2. else, the same object could be described considering two points (head and tail) and then applying two dedicated operators:
   - an operator called +, to identify the concatenation between head and tail;
   - and an operator called ≈, which can swap head and tail.

cylinder → ≈v + b + v + ≈t + b

Figure 1.22. Representation of a cylinder, by concatenation and swap.

3. otherwise, the same object could be described by introducing a dedicated operator *, which catches contiguity:

$$\text{cylinder} \rightarrow \text{side} * \text{roof} \rightarrow (\approx v + b + v) * (t * b)$$

Figure 1.23. Representation of a cylinder, by concatenation, swap and contiguity.

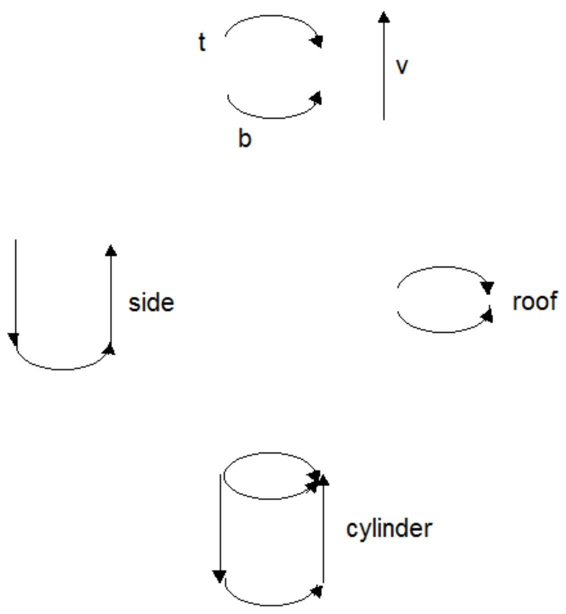

Figure 1.24. Reconstruction of a cylinder.

Scientists in the fields of both artificial intelligence and computer vision have been working for years to the simulation of the human perception system, in order to equip the computer with perceptual abilities.

In fact, for a computer, an image is an array of numbers, indicating the values of grayscale or RGB parameters. The Gestalt principles rule many strategies adopted in computer vision techniques: the examination of an image determines the significant components in it and allows their classification into known objects.

segmentation strategy	**reference to Gestalt laws**
Thresholding	figure / background
region growing	proximity, homogeneity
morphological methods	continuity, closure
statistical methods	homogeneity
search and association of edges	continuity, closure

Table 1.1. Segmentation strategy.

Usually, we take advantage of the segmentation technique: automatic image analysis provides the extraction of features about colour, size or intensity.

Below, there's a brief description of most commonly used segmentation techniques:

- the definition of a threshold is a non-contextual technique, generally based on the mapping of grey values in the image, with respect to two references (black and white).
  The procedure represents a way to split the image between background and figure; its sensitivity directly corresponds to the number of levels considered in the analysis;
- the edge detection is a method based on recognition of discontinuities: it is a widely used practice, because borders of regions often host information.
  The technique is contextual type, because it derives significant regions by applying the Hough transform;
- the region extraction is a method focusing on recognition of areas, which properly studies sizes, shapes and textures.
  The technique is non-contextual type, because it's based on dividing the image into regions: all pixels must be assigned and each one can only belong to one single zone; in addition, each region must be connected to the others.

1.4.2. Relational matching.

The relational matching is a technique developed to solve the problem of the correspondence between data descriptions. It is a high level process, which uses the most complex data representation; therefore, it's a multifaceted problem, connected to photogrammetric analysis and even to the development of techniques of computer vision and artificial intelligence.

In order to compare two objects, the relational matching benchmarks their descriptions: it searches a correspondence between the elements of the first object and the ones belonging to the second, by comparing attributes and relationships.

According to von Helmholtz[19], the visual perception is a form of unconscious inference. Vision derives from a likely interpretation of incomplete data (then it requires *a priori* assumptions on the world); the recognition of constituent elements of an object leads to the association with the complete object, through integration of its characteristics.
The unconscious interpretation hypothesis was taken on by studies on perception: it is assumed that the visual system uses a kind of Bayesian inference, in order to acquire awareness from sensory information.

The matching is exact (perfect matching) when each object of the first description finds its counterpart into the other, and vice versa, since the relationship is two-way.
Using a system implementing a locally exact matching, we can recognize in the image all the perfect correspondences between relationships representations (through graphs): software automatically finds the association among points with the same encoding and compatible geometry, according to the provided information.
The definition of local accuracy comes from the recognition of perfect matches between the structures, identified by the relational description, thus creating a one-to-one association between the elements of examined graphs.

Otherwise, when it is accepted that some objects in a description do not find their correspondents into the other, we have an imperfect matching. In this case, the process has higher complexity: in my previous thesis, I already created software dealing with it.
In detail, the aim is to search similarities between graphs, controlling all structures and only storing the correspondences that are believed to be correct, without employing a unique criterion for choice. For myself, I suggest preferring a semi-automatic approach, because systems not requiring human intervention proved to be excessively fallacious; instead, if the user directly controls the software, indicating which matches should be assumed as correct

[19] Hermann von Helmholtz is the founder of scientific studies on visual perception.

amongst those found, the program can search other associations by exploiting the topological information provided by representations.

According to Shapiro and Haralick, comparison of structures proceeds with stages:

1. assignment of a relational description to the structures found in the image;
2. choice of a matching function between one description and another, through the definition of a relational distance between corresponding representations;
3. use of a search algorithm for the minimum value of relational distance.

Application of the matching procedure reflects the evolution accomplished in the study of correspondence between descriptions:

- the first application of the theory comes out the chemical field, for molecules classification: a comparison between relational descriptions of atoms (considered as primitives) and reciprocal links (relationships) allowed automatic storage for structures that were not included in the database yet;
- later, the same method has been implemented in the context of computer vision, because interpretation of a scene enhances with study of correspondences between a relational description extrapolated from an image and models about objects;
- else, a completely automatic vision system proceeds through the reconstruction of a surface, over the research of correspondences between two relational descriptions extracted from images – and my thesis deals with this case.

Supposed that *A* is the set of parts of an object O_A, while *B* represents the set of the parts of an object O_B, relationships between *A* and *B* have grade *N*.
They correspond to:

$$R \subseteq A^N$$
$$S \subseteq B^N$$

Then the correspondence between elements of a set *A* and the ones of the set *B* is described by the function $h: A \rightarrow B$.

The composition $R \circ h$ by *R*, with *h* as the primitives set of *B*, correspondent to the primitives of *A* that are in the set R_i, is defined according to the following relationship:

$$R_{ik} \circ h = \{(b_1, \dots, b_i) \in B : \exists (a_{k1}, \dots, a_{kj}) \in A\}$$

Thus, relationships reported below are both valid:

$$R_{ik}(a_{k1}, \ldots, a_{ki})$$
$$h(a_i) = b_i \quad \forall i = 1, \ldots, N$$

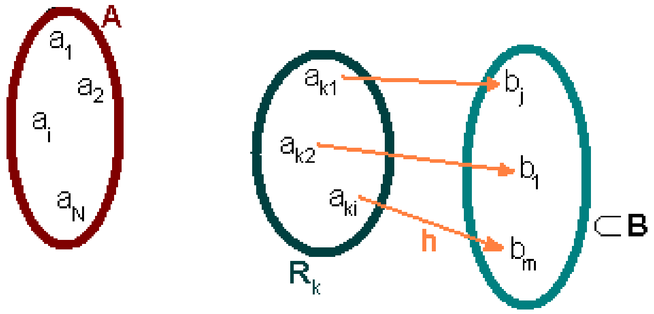

Figure 1.25. Representation of composition $R_1 \circ h$.

Through the correspondence function *h*, it's possible to compose a binary relationship R_2.

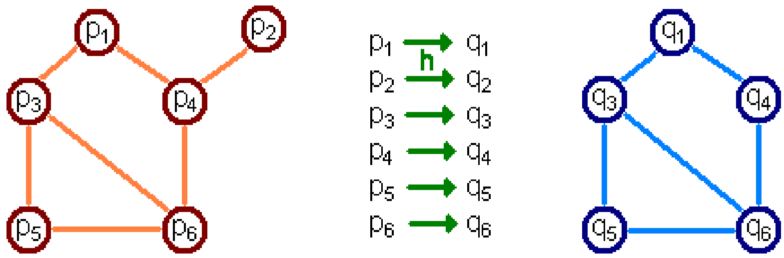

Figure 1.26. Representation of composition $R_2 \circ h$.

The perfect matching is a correspondence associating every relationship of grade *N* in *A* to another one in *B*: when it's applied to each component of a single relationship *R*, with grade *N*, the outcome is yet a relationship in *S*, with grade *N*.

This kind of transformation is called homomorphism, because it preserves the structure of sets it works on.

When $S \subseteq B^N$ is a relationship in *B* with grade *N*, a homomorphism from $R \subseteq A^N$ to *S* is a correspondence, according to the formula:

$$h: A \to B \,|\, R \circ h \subseteq S$$

In other words, it creates a set belonging to *S*:

$$\{h(a_1), \ldots, h(a_i)\} \subseteq S_i$$

A relational homomorphism finds a correspondence between all primitives in *A* to a sub-set of primitives in *B*, specifically dealing with the same relationships working in *A,* too.
If *A* is far smaller of *B*, the search of a relational homomorphism resolves in the quest of an object copy, within a bigger space.
When the two structures *A* and *B* approximately have the same dimensions, then the search of a homomorphism equals to determine the similarity between two objects.

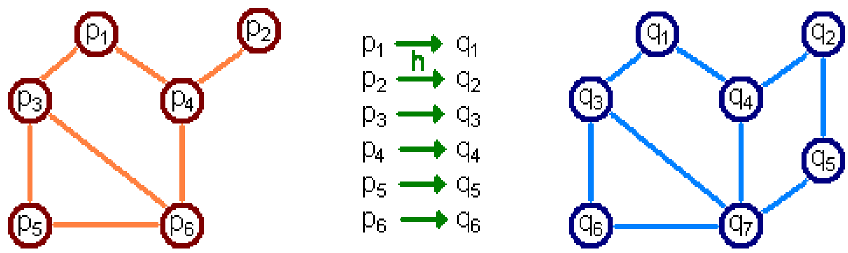

Figure 1.27. Representation of a relational homomorphism.

A homomorphism type one-to-one is called a monomorphism.
The correspondence function, in this case, links every primitive in A with a single primitive in B: therefore, it's a stronger correspondence with respect to homomorphism.

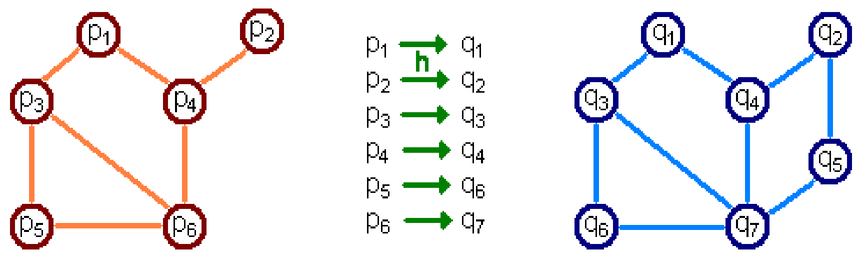

Figure 1.28. Representation of a relational monomorphism.

An isomorphism, finally, is the strongest kind of correspondence: a relational isomorphism *f*, from a relationship *R* with grade *N* to another relationship *S* with grade *N* is a monomorphism from *R* to *S*, while f^{-1} is a monomorphism from *S* to *R*.

In this situation, *A* and *B* just have the same number of elements and one primitive in *B* only corresponds to a single primitive in *A*; vice versa, every relationship in *B* only corresponds to a single relationship in *A*.

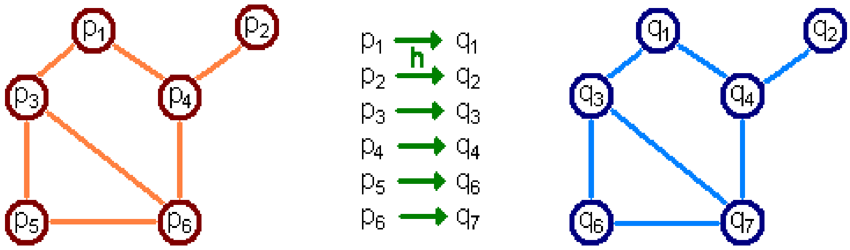

Figure 1.29. Representation of a relational isomorphism.

However, because of segmentation faults, it is practically impossible to achieve a perfect matching, neither between images, neither between an image and a model: there is always the chance that some relations or primitives in a description do not find a correspondence into the other. In fact, among the descriptions extracted from real images, there's no guarantee about nor surjectivity nor injectivity of identified relationships.

Furthermore, within the photogrammetric field, we shall consider:

- the variation of the pinch point, which generates perspective deformations and occlusions into images;
- the noise, introduced between the observations in the acquisition phase, which affects the chance to extract relevant and reliable information.

Therefore, the concept of relational matching requires an extension, which includes the imperfect matching case, in order to allow the possibility of having both relationships and individual primitives.

In the imperfect relational matching, we look for the less different correspondence between available relational descriptions: in this case, we associate elements that correspond one to each other in an inexact way, of course verifying the value of each attribute, with respect to that of the characteristic corresponding to the candidate point.
The quality of every match is judged by means of its relational distance.

Two relational descriptions of parts set in *A* and *B*, about objects called O_A and O_B, which are respectively composed by *I* and *L* elements, correspond by an isomorphism *f* from *A* to *B*:

$$D_A = (R_1, R_2, \dots, R_i)$$
$$D_B = (S_1, S_2, \dots, S_i)$$

Then the relational distance (as defined by Shapiro and Haralick) represents the minimum total gap of the best correspondence *f* from *A* to *B*. It comes from the formula:

$$E(f) = |R \circ f - S| + |S \circ f - 1 - R|$$

The total error on *f* is the sum of structural errors for every relationships couple.
In other words, it's composed by:

- the number of relationships $R \subseteq A^N$ which cannot find a correspondence with *f* within the relationships $S \subseteq B^N$;
- and the number of relationships $S \subseteq B^N$ which cannot find a correspondence with f^{1} within the relationships $R \subseteq A^N$.

Boyer and Kak use instead a probabilistic approach to evaluate simultaneously geometric and semantic attributes, working in the Bayesian framework.

According to the principles of information theory, the search of best matching equals to the use of the best transmission channel, which carries the slightest noise.
The information contained in a relational description *R* coincides with the sum of data belonging to the relationships attributes; it's proportional to the probability that the attribute of the report assumes that value.

Using the signals' theory, the information contained in a selected symbol *a*, with probability of occurrence *P(a)*, can be defined as:

$$l(a) = -\, log_b\, P(a)$$

Definition is due to the fact that the logarithm is the single functional relationship that can be set between information and its associated likelihood.

The relational distance $D_h(D_A, D_B)$ of the correspondence function *h* is defined as the sum of conditional information between relationships *R* and *S*.
The conditional information $I_h(R/S)$ is a function of the logarithm of correspondent conditional likelihood, which actually represents the probability that the attribute of a relationship $R \subseteq A^N$, defined on a set of parts *A* relevant to the relational description D_A, takes a certain value (when the value of the attribute in the correspondent relationship in the set *B* of relational description D_B is known).

For two relational descriptions of sets of parts *A* and *B* about objects O_A and O_B:

$$D_A = (R_1, R_2, \dots, R_i)$$
$$D_B = (S_1, S_2, \dots, S_i)$$

the following relationships are valid:

$$R \subseteq A^N$$
$$S \subseteq B^N$$

Then, by introduction of a threshold value *w(r)* for each attribute *r* belonging to *R*, the correspondence $h: A \rightarrow B$ (as obtained by minimization of relational distance) is only accepted when:

- the sum of relationships that are still without a correspondence, weighted with respect to importance and reliability, assesses a value inferior to a global threshold;
- the value related to attributes difference for each relationship $R \subseteq A^N$, corresponding to the relationship $S \subseteq B^N$, proves to be minor of a certain threshold.

The chance to perform an imperfect matching between assigned descriptions really complicates the computational aspect of the correspondence problem: in order to achieve a solution, it's fundamental to reduce the search space, by using primitives that satisfy the principles of rarity and distinctiveness.

The use of data descriptions with high level structuring meets this condition, but together with the disadvantage due to the complexity of the representation construction: the use of a process of semantic type requires the integration of *a priori* knowledge in the system.
According to the Gestalt theories, it is possible to build meaningful data groups, with low probability of accidental formation in the image, without *a priori* information on the scene. In fact, perceptual organization of data refers to the human ability to infer relevant groupings and structures in an image without prior knowledge about its content.
Through the principles of perceptual organization, we can thus solve the process of object recognition and also reduce the size of the search space for correspondences for an automatic vision system. Despite its high capacity, in fact, the human perception system can fail in grasping relevant structures in the same situations in which, however, an automatic vision system achieves success.

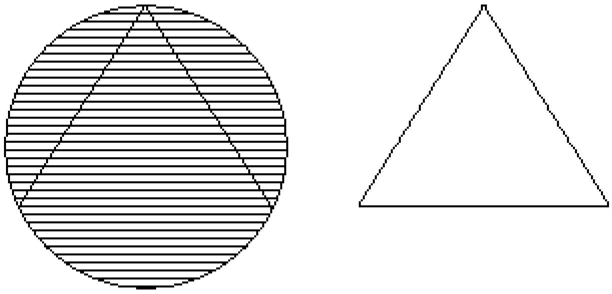

Figure 1.30. Example: presence and visibility of the same element in two images.

Furthermore, the human visual system is subject to visual illusions, too.
When the inference process is not successful, it's necessary to formulate a set of assumptions accepted by the visual system.

In cases like the one shown in Figure 1.31, computers are far stronger than the human beings, because they can determine the scene geometry without affecting evaluations with incorrect optical perceptions.

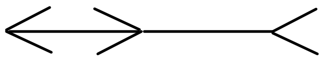

Figure 1.31. Example: optical illusion (the two segments have the same length).

Finally, we have full knowledge of the process for the creation of a relational description on an image. The procedure is described below.

The first part of the description (primal sketch) is divided into two different tasks:

- description of the changes in the image intensity, by using as primitives both the boundaries and the region (raw primal sketch)
  In order to derive it, Marr and Hildreth designed an optimal filter, simulating the operation of the physiological processes of the human vision system;
- explanation of geometrical relationships, by grouping the selected primitives (full primal sketch), with the aim to reduce the problem of false matches.

The issue related to spatial correspondence is finally resolved by the use of stereo-vision, through the construction of a representation level called 2.5D, which symbolizes a map of orientations and depths of surfaces.
In detail, it includes:

- the values referring to distances from the surface;
- the values of orientation;

- the boundaries, along which orientation usually varies – because rapid changes in values lead to edge detection;
- and the boundaries where depth is discontinuous (this feature is called subjective edge).

In order to further reduce the ambiguities of the correspondence problem, it is also possible to apply the Gestalt rules since the pre-processing phase: this way, since groupings are built according to the principles of perceptual organization, we're ultimately limiting the search space for matches.

We'll start from the image description based on primitives, which identify changes in picture intensity (raw primal sketch); we use perceptual groupings in full primal sketch in order to ease the correspondence problem: this trick allows removal of ambiguities and the construction of the 2.5D representation.
As a result, we can say that the description based on the features study coincides with the raw primal sketch, while the description obtained following a perceptual grouping corresponds to the full primal sketch.

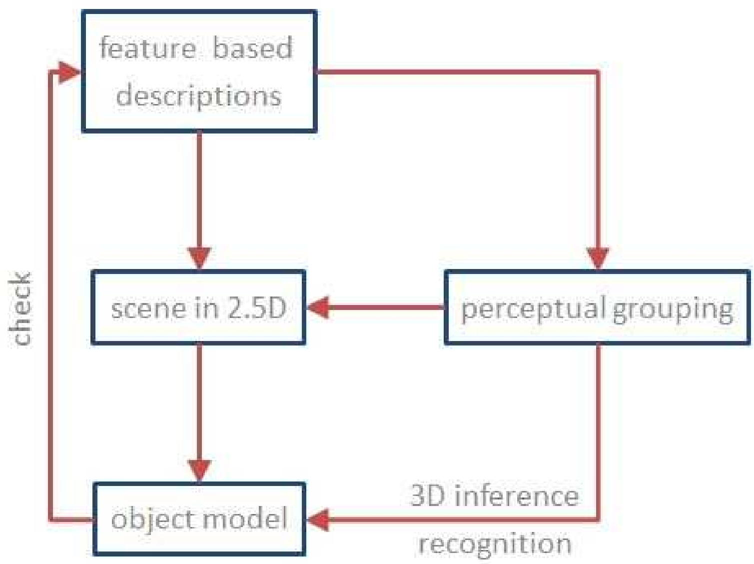

Figure 1.32. Perceptual grouping of data.

chapter 2

Linguistics

The thesis project focuses on a new challenge about artificial intelligence.

Starting from the contrast between the rationalist approach and the empirical point of view, we are looking for integrated tools of abstract rules and statistical modules: this is actually a brand new science, called the Human Computer Sciences, or even Knowledge Engineering.
The main goal of this modern discipline is in image processing, assisted by machine learning and robot vision. However, the linguistic competence simulated with computer technology is yet not comparable to that of human beings:

- due to technical reasons, of course because the state of the art has not yet achieved the complete mastery of all possible linguistic exceptions;
- and due to incomplete understanding of language processing, by man.

The objective of research is to bring the rules of Universal Grammar in the automatic reading of maps and images, through the recognition of syntactic models for the comparison of map parts, images and 3D models, using archetypes. Indeed, there is an analogy between the hierarchical structure (tree-shaped) of models and languages syntax: in fact, for some linguists, the hierarchical structure would be just inherent in the structure of the human mind.

The combination of rules-representations perfectly corresponds to the representation of syntactic structures within the mind:

- because even a syntactic structure is a hierarchical structure and it’s intuitively displayed as a graph or a tree;
- and the whole procedure related to the syntactic method is similar to that already used for the development of techniques for automatic vision.

Please note that:

- models are built from sub-models, just as sentences are formed by words;
- and sub-models are built from primitives, just as the words are a combination of letters and syllables.

In addition, the number of models and rules is probably finite, because the structure is likely similar among endless options.
Conversely, a superficial resemblance is excluded, because the lack of distance would easily generate confusion during communication between different parties.

The grammars are then defined as the rules of primitive combination into objects and they derive themselves from the language that can describe the model.

Given that the syntactic approach is valid if it's powerful (in other words, it proves to be able to describe complex objects), we must study the expressive capacity through:

- the analysis of the primitives number in the model we chose and adopted.

  Naturally, a modest amount of atomic elements allows easier computation and management for the automated system;

- and the analysis of the rules, to be applied iteratively, according to some re-writing instructions.

Therefore, the best possible syntactic representation establishes a compromise between the number of endless existing relationships and the amount of primitives that are strictly needed for the description of the model: in this analysis, we must take into account the technical constraints (computational available power) and the final result of the study.

2.1. Language[1].

The language constitutes the communication system working between individuals: information is transmitted via symbols and the system is finite, arbitrary and combined according to the rules of grammar.
The transmitted information is a part of the end product of a process that manages:

- sensory perception,
- concepts,
- feelings and emotions,
- ideas and thoughts
- through content, involving temporal succession.

Communication exists whenever there is an information transfer from a sender to a recipient; virtually, the message in its original form coincides with the information decoded by the receiver, without loss due to noise.

The language is the human ability to communicate, by means of verbal systems.
In fact, language and thought are connected all along: the search for a universal language stems from the desire to have a rational tool for expressing concepts. In other words, language is a logical organizer of experience and thought, because it is the most powerful tool providing symbolic representation; thus, it constitutes the very basis of conceptual functions communication.
Furthermore, language also represents the most economical and appropriate way to participate in the community life, indeed allowing an interchange between every individual and the group.

The ability of language was developed by humans as a result of structural changes of the oral cavity:

- the retreat of the uvula has made human beings capable of expressing a diverse range of sound,
- thus being capable of guaranteeing a specific (instead of generic) naming of the world.

However, man is not the only animal to use conventional signals: nevertheless, the language is prerogative of man. In animals, information is shared without forms of verbal thought, nor do speech and action interact: in fact, no other living specie has a similar language regarding complexity and level of processing.

---

[1] The word *language*, that is borrowed from Latin in English, French and Italian, probably comes from the Indo-European root **dang –va*.

The language is the concrete way in which the ability of human speech manifests itself: the verbal potential of an individual is carried out in an historical, geographical and social contest. In fact, all languages in the world are defined historical-natural languages:

- because they are born from the historical development of a region of the world
- and also they are opposed to artificial languages, which usually have simpler structures.

The common traits that identify a language are:

- vocabulary,
- phonemic system,
- grammar and syntax,
- style and pragmatic,
- and, eventually, a common system made of signs.

2.1.1. History of the study of language.

The issue related to the origins of human language is controversial:

- the language may be innate, in other words it could have a biological basis in the genetic makeup of the species,
- or it might be a learned skill.

However, the discovery of the FOXP2 gene[2] seems to completely define the correctness of the first hypothesis, because it is the possession of both word and syntax to progressively build the mind, integrating experiences through the environmental influence and perceptions, up to full self-consciousness.

In the following, we give a brief overview of the fascinating history of the language, in order to find out all the reasons to think that languages belong and are part of the modern disciplines of surveying.

All the languages that are known in the world up to now are approximately 5,000 and can be grouped into ten super families:

- three are in the sub-Saharan Africa zone;
- one occupied, 100,000 years ago, the Middle East and the Mediterranean basin - and it still survives in the Basque, Caucasian and Tibetan languages;
- two spread in Oceania 40,000 years ago – they are called the Australian family and Indo-Pacific family;
- other two covered Europe and Asia, including the Middle East and the North African area – they are called Euro-Asian family and Afro-Asian family.
- another Amerindian language (spread between 25,000 and 10,000 years ago) belongs to the people of all American area;
- two areas related to successive migrations, on the northern edge of the American continent, constitute the last super-family.

---

[2] FOXP2 is the gene responsible for the skill of language.

Since mutations in this gene are proved to be associated with peripheral disorders of speech, it seems not to be involved in grammatical-type processes.

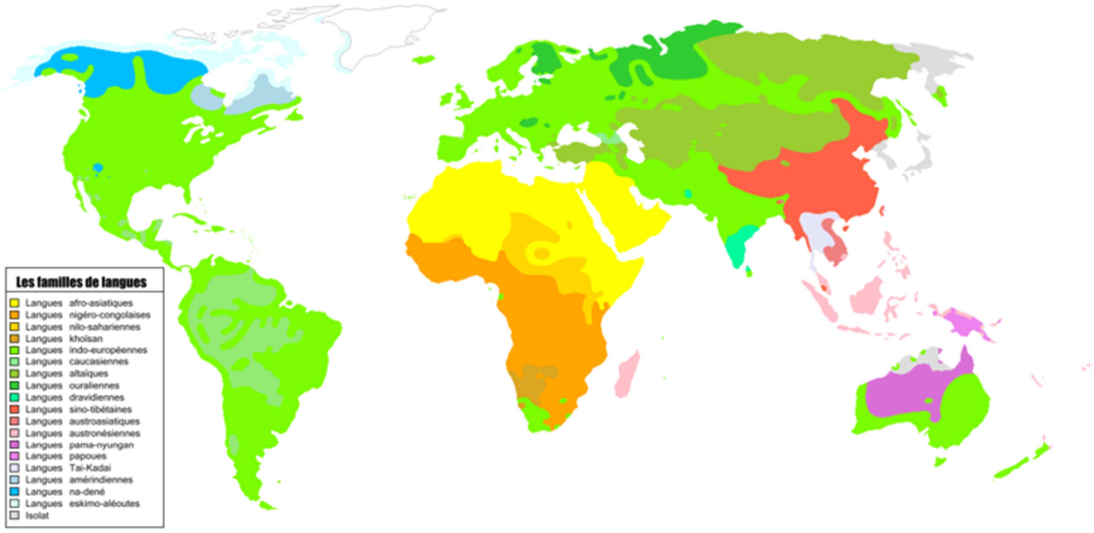

Figure 2.1. Principal language families (or groups of families) of the world.

Interest in the language began in antiquity: the first grammatical studies took place in ancient Greece, alongside discussions on linguistic philosophy and about the concept of λόγος, which is the word intended in its various manifestations (expression, speech, literature, literary and scientific works).

The conventionalist theory asserts the arbitrariness of the linguistic reference:

- Parmenides saw the words as labels for illusory things,
- Empedocles and Democritus tried to demonstrate the accuracy of the correspondence between label (word) and thing (concept),
- finally, the Sophists believed and asserted that the names have different in nature with respect to things[3].

Aristotle, in his *De interpretatione*, says the naming of things happens is by convention: sounds and signs which are associated to concepts may vary, but the objects and images which they refer are the same for all, within the mind.

Plato, in the *Cratylus*, makes Hermogenes say that language is conventional. Socrates would instead claim that the language is to be performed according to nature, like every other human fact: even appointing is a perfectly natural action, because it is a component of the mean.

Plato does not take sides in favour of the thesis of language as a convention or nature (whether in diversity, nor through similarity), but he states that it is using to determine the

---

[3] *The language cannot show existing things, just as a non-existing thing doesn't manifest its nature to another one.* [Gorgias]

meaning of words. In fact, language is a tool coming from repeatable and repeated choices; therefore it can be judged by the adequacy of its purposes.

The Stoics intervened, supporting the division of the philosophy between several disciplines: ethics, physics and logic. The language falls within the logic field, which deals with dialectic and rhetoric, sounds and meanings.

In their naturalistic vision of the world, the study of meanings includes the *representation* and the *expressions*, thus vindicating the instrumental role of language.

During the Roman times, the route taken by the Greek scholars was continued: in fact, even the structural similarity between the two languages Greek and Roman allowed one easy metalinguistic step.

In *De lingua latina*, Varro says the development of the language from a small original nucleus of primary words, which later change through productive work.

In addition, anticipating modern theories, the study of language starts to be divided into distinct disciplines of linguistics (etymology, morphology and syntax).

During the Middle Ages, Scholasticism focused on speculative grammar, as a synthesis of thought, language and objective reality: the concepts of genus and universal species are categories of thought and language, which cannot show a real response.

Dante Alighieri, in his *De vulgari eloquentia*, takes grammar apart from the vernacular language:

- the first speech is spontaneously spoken, as it's learned by a child,
- while the grammar was then invented and formalized by scientists and requires you have to study and to learn how to use it.

The language is intended as a characteristic of man: it is a necessity, with material and rational substance.

Through reading the Bible, Alighieri says that God gives language directly to Adam: Hebrew is then the first language of humanity. Therefore, idioms change in space and time as a result of divine punishment (the destruction of the Babel tower).

Figure 2.2. *The Tower of Babel* (1583), by Pieter Brügel the Elder.

The poet interprets that from that event three distinct linguistic stocks have originated:

- the Mallard-Slavic strain, located in the northern part of Europe;
- a strain related to Hispanic, French and Italian languages, geographically located in the southern European area;
- and a further strain between Eastern Europe and Asia.

In fact, the origin of modern languages is not known yet:

- they may derive from a common original language (and this possibility is commonly called the *monogenetic hypothesis*),
- or they could come from primordial strains (and this is called the polygenetic hypothesis).

In any case, the existing languages are the result of a continuous differentiation process, which occurred over the millennia.

The Modistae subsequently claimed that the grammar reflects the structure of reality in the way it's perceived by the human mind.

They even distinguished between:

- *modi essendi*, literally modes of being, understood as the ability to exist, proper of things and objects;
- *modi intelligendi*, literally modes of understanding, properly intended as the ability of understanding events and facts, things and objects;
- *modi significandi*, literally modes of signifying, as the proper form of speech parts.

During the Renaissance, the geographical discoveries encouraged the study of new languages, then the rediscovery of classical cultures and, ultimately, the acknowledgment of the importance of vernacular languages in use in Europe.
Moreover, it's the time in which the contamination of non-European language on the European speeches begins, especially due to the Arab presence in the Mediterranean.

Francis Bacon criticized traditional philosophy: in his view, the function of language lies in the ability to acquire and transmit knowledge. The grammar, understood as the language conventionality, has been given by God to man after the punishment of Babel, in order to restore the relationship between words and things and ultimately allow communication between individuals.

According to Hobbes, the language would rather have more functions:

- to stabilize the thoughts, protecting them against the risk of dissipation;
- to facilitate communication;
- to encourage both the creation and the development of secondary skills (typically, mathematics and reasoning).

Locke criticized the doctrine of nativism and the original language: in his point of view, knowledge is purely nominalist. In fact, the language creates and determines the social nature of man, while signs designate ideas in the mind of those who use them, according to an arbitrary association:

- you must have the specific character to know,
- but the general character is sufficient to call, or to assign a name.

According to Locke, language is a tool constitutionally ambiguous and it's composed by three elements:

- names of simple ideas, directly drawn from experience and therefore natural;
- names of mixed modes or complex ideas, which derive from subjective elaboration: every individual selects the ideas, makes them up, then assigns them a name;
- names of natural substances.

His theory results in a pessimistic conception, due incommunicableness that just lies in the nature of language:

- in fact, the phonation occurs before the understanding of the ideas which the words correspond to;
- and it is also possible to allocate the same name in different ideas.

Herder and von Humboldt argue that language is the formulation of connections, abstractions and reflections on an object, whose entirety and distinctive qualities constitute our perceptions.

According to Leibniz, the man talks driven by needs and passions: then he resumes the Platonic position, according to which the meanings of words are determined by:

- natural reasons, in which the case has its share
- or moral reasons, where a given choice enters the reasoning, determining the result.

His criticism towards Locke regards both the sign arbitrariness and the nominalist character of knowledge: the sign is autonomous in the knowledge process, because it does not translate into words a ready content, while it’s already involved in the formation of thought.

Finally, Leibniz promotes the German language and makes etymological studies in an attempt to reconstruct the history of primitive peoples and build a mathematical language, which could allow universal peace.

According to Blumenfeldj, the phoneme is the smallest unit with distinctive sonic characteristics: each language chooses his phonemes, but the decision is not conventional (nor arbitrary), nor natural (or necessary), because sounds strictly depend on the functionality of the speech.

Port Royal secured the existence of a general and universal grammar, which underlies every historical language: a perfect and primitive language would give rise to all the historical languages, for detachment and subsequent degeneration.
All linguistic facts have a match in thought: however, the language proves sometimes to be insufficient to perfectly translate ideas.

Ivan Pavlov was the first to demonstrate that the language is an important resource in the development of intelligence. With the aim to recognize the mediating function between the environment and the human being, he made studies and experiments on perceptions, mental representations and signal processing: in practice, these data form the basis of the original formation of concepts.

Jean Piaget supported the presence of two basic stages of development:

- egocentric language (0-6 years), consisting of echolalia and monologues, animism and attribution of names to even unreal objects;
- while, in the second phase, the social language expands, providing dialogues and bilateral communications.

Bernstein then elaborated a theory that indicates the relationship between the type of language developed by individuals and the relationship between family environment and occupational class and type.
In fact, Sociolinguistics[4] takes care of the analysis of social determinants of language and identifies the stratification of the social structure: the language is not only practical use, but also interaction - and social control, too.

McLuhan, in his general theory of systems, defines a system as a set of different elements, reciprocally related through a structure. It regulates the processes within the system, in order to exert a constant influence on the structure: please note, whenever the processes evolves anyhow, the structure tends to change too, gradually.

The School of Palo Alto (California), linking to the information theory concepts coming from other disciplines (cybernetics, psychology, sociology, anthropology and ethology, in particular), formulated then some communication axioms:

- each communication has an aspect of content and one related to the relationship – and, namely, it classifies the content;
- the nature of a relationship depends on the communication sequences between the communicating individuals involved.

According to Wiener, founder of Cybernetics, the mechanism of self-regulation (feedback) is defined by:

- knowledge, which always is limited;
- time, as it's an unavoidable factor of knowledge, because the phenomena can be described as time-varying stochastic processes;

[4] Sociolinguistics began as a specialized field of Linguistics: its research subjects are language, communication and interplay between the two with society and culture.
Empirical data are concrete linguistic messages or texts: the choice of words, along with other clues, can be attributed to different levels of education or to the type of relationship (both social and psychological) between speakers.

In mathematics, Kolmogorov introduced the algorithmic complexity, in order to define a sequence of N symbols according to the length $L(N)$ of the instructions which are strictly necessary to generate it.

Complexity and randomness are linked together according to the concept of entropy: the maximum is found in correspondence of a completely random distribution[5].

[5] As an example, the sequence of the first N digits of π has low complexity, because the number is definable with a rule (a program of instructions), whose calculation is resolved in a lower number of bits than the content in the sequence, when N is sufficiently big.

On the contrary, a completely random sequence has maximum complexity, because its description can only be made by the sequence itself.

2.2. Linguistics.

The contributions of Linguistics concern the nature of human language and the relationship with other mental faculties: by function, the language is use of inter-subjective signs, to be used aiming to communication.

In detail, Linguistics is the discipline that studies the language, as the innate human potential to produce speech, and it is responsible for:

- language, as the power of the human species to use symbolic communication tools.
  Linguistics studies the human expressive capability logically pre-existing on its concrete implementation (phonation or speech, writing, gestures): in fact, up to two years, the infant possesses the language, but it still cannot manage his tongue;
- and the historical language, which is the product of this faculty.
  Languages are complex systems for human communication, consisting of minimum segments, which convey meaning (morphemes and phonemes), reciprocally articulated in a complex hierarchical system (sentence).

The general Linguistics deals with the development of categories and concepts with which describe this innate ability.

It can be divided between different areas:

- grammar, which is traditionally composed of four disciplines (phonology[6], prosody, morphology and syntax);
- semantics;
- pragmatic;
- lexicology[7] and etymology.

---

[6] The Phonology is the study of sounds, in relation to their function as differentiators of meanings, from the Greek words φωνή λόγος.

The Phonemic is the branch of linguistics that studies the functioning of sounds in a language, or – in other words - the organization of distinct units of sound (phonemes). This discipline belongs organically to the grammar of any language.

In contrast, Phonetics is the general study of speech sounds (phones): the phoneme is the smallest phonological units of a language.

[7] Lexicology is the discipline that deals with the structure of the lexicon: it's an independent subject by the Sixties. It researches the basic components of a language and determines rules and relationships between lexical components (morpheme, word and groups of words).

In the ancient world, the first attempts of Linguistics are related to translations, during the Hellenistic period: the Modern Linguistics was founded in the fifth century.

At the beginning of the nineteenth century, interest in Sanskrit and the work of Von Humboldt developed the study of language, through the comparative grammar.

Then we owe to De Saussure (1916), Trubetzkoy (1926) and Chomsky (1956) the modern concepts organization.

The birth of Linguistics as a science is recent and it's conventionally due to Ferdinand de Saussure (*Cours de linguistique générale*).
He revolutionized the study of language in a functional sense, finding the fundamental units of *chaine – parlée*, through the introduction of structural analysis.
In contrast to the traditional original names, he argued that the linguistic sign unites a concept and an acoustic image: nevertheless, this is not the physical sound, but the psychic trace leaved by sound. Thus, the linguistic sign is a psychic entity with two faces, summarized by the pair meaning (*signifié*) and signifier (*signifiant*): the substitution between concept and meaning, between image and signifier, finally leads to an arbitrary ratio.

Then, the School of Prague introduced the term *structure*, because each language (as a system) requires a structural analysis.
However, it is still uncertain whether we have the opportunity to study the structures semantics in a systematic way:

- Bloomfield called impossible the scientific study of meanings;
- while European scholars are more confident in the ability to define structural models for the lexicon.

vehicle = car + engine + wheel + wheel + wheel + wheel

Figure 2.3. Example of decomposition of meaning minimum units (Hjelmseev's Theory).

Then, Trubetzkoy defined a subdivision of the *chaine – parlée* into elementary units, from the phonetic point of view: the phonemes are signs and work oppositely, in order to allow recognition of different meanings.

Natural languages have a structure with a double feature of code; their double articulation allows greater variety of expression:

- the units in the first joint (morphemes) actually compose the message;

- but, at the same time, they are composed of phonemes without a common meaning[8].

Finally, the grammar of a language is a diagram, which identifies the permitted sentences in that speech and shows the rules that allow you to combine words into sentences (Chomsky, *Theory of Formal Languages)*.

In fact, there are interesting intersections between Logic and Linguistics:

- human thinking places objects into classes. The purpose of Logic is to analyse and to establish relationships between classes of objects, because the human mind acts on the basis of the criterion of concepts opposition: the thought is based on an underlying network of meanings (Boole);
- the language too is characterized by phonemes, which are entities identified just in opposition (De Saussure, Trubetzkoy). The language consists of expressions in transformation, while maintaining a significant deep structure (Chomsky).
  Therefore, the Logic is tied to the theory of natural language and the natural language can be reduced to a set of symbols and combination rules;
- and the development of Information Technology Science clearly demonstrates productivity of the interaction between language, logic and observation of the real world.

The statistical approach to language began when scholars are aware of the value of statistical methods in sounds recognition:

- the hidden Markov models (HMM) are used to build the language of Eugene Onegin (Markov, 1913);
- the Theory of Formal Languages (Chomsky, 1957) arises from the need to understand the fundamental properties of natural languages, by using a mathematical tool. It deals with phrase-structure grammars, characterized by a set of re-writing rules which describe the language types.
  In fact, the human mind represents syntactic structures[9] of sentences;

---

[8] Conversely, Mathematics and Logic are communication systems in which the minimum units of the message have its own form and meaning and they cannot be further divided: in fact, they are the units of first articulation.

If a similar language should express the totality of human experience, the codes should be much larger in number, thus complicating storage of information.

On the other hand, a clearer language would have excessive complexity, when it was composed of a single articulation.

[9] A syntactic structure is hierarchical type, so it can be displayed with multiple representations.

According to a statistical study of language:

- each individual statement (De Saussure's *parole*) becomes a sample of a population,
- and each phoneme in a language has a frequency (which is usually constant),
- if a different frequency is just due to accidental fluctuation (Herdan, Shannon), the generative grammar rules can however assign a structural description to various sentences (Chomsky).

Various grammars have been proposed up to now: everyone was investigated in its properties in order to identify a number of recognition mechanisms, called automata.

Chomsky works on the combination rules – representations.
The rules operate on the representations, eventually producing other representations.
There are four representation types, arranged in a hierarchical scheme:

- deep,
- shallow,
- phonetic
- and conceptual.

In the past, geodetic and cartographic experts have been the necessary link between the world of ideas (composed of mathematicians, statisticians, physicists) and the explorers of sky and earth (made by astronomers, geographers, navigators).

Today, the bridge function proposes an even more complex and high level between the theoretical experts (cognitive scientists, computer scientists, specialists in artificial intelligence) and end users:

- complexity lies in the fact that the two worlds are almost completely disjoint and the community of experts in the disciplines of detection is a small niche, substantially separated from both worlds;
- the higher elevation, instead, is in the ambition to move from modelling and interpretation to calculation and learning: this passage would be certainly challenging and require understanding about work of representations in the human mind.

2.2.1. Structuralism.

Structuralism was born through Psychology and Linguistics (Jakobson and Trubetzkoy).

De Saussure defined the linguistic sign as the union of a signifier and a meaning: in detail, the signifier is the set of sounds (speech production) that invokes a certain meaning for speakers. Conversely, the meaning has a more complex definition, because it is related to the concept, the object, the phenomenon, or the indication offered by the signifier.

In addition, the meaning of a word depends on both the psychological subject and the language itself: the object is not in itself, but depends on the subject that takes consciousness or knowledge. Indeed, the subject is affected by his emotional and cognitive structures, while the language is determined by the choices of person and community and it finally determines the logical organization of the conceptual world.

It is therefore correct, in Linguistics, to define the meaning as "verbally elaborated", rather than refer to a concept, an object, an action or relationship: therefore, the meaning may be defined as the part of extra-linguistic reality, referred to a signifier.
Thus, the link between signifier and signified is arbitrary and has a historical reason (please note that the cause is not logic, anyhow): it depends on a convention[10].

Language is a system for connecting a set of signifiers to the universe of meanings. The relationship is not strictly two-way, because:

- multiple meanings can correspond to a single signifier (polysomic words);
- several signifiers can correspond to only one meaning (this is the case for synonyms);
- a set of signifiers may indicate a different meaning than the sum of the meanings (for example, take the expression *it's raining cats and dogs*);
- some signifiers include meaning areas which are actually belonging to other concepts.

Each linguistic sign has:

- an acoustic image - the signifier is made up and constituted by the succession of sounds that compose it;
- and the expressed concept (meaning).

---

[10] Let's make an example, for better understanding of concept. A linguistic sign is like to a banknote:

- the signifier is the rectangle of paper, with size and color set,
- while the meaning is the value attributed to the rectangle.

The link between the rectangle of paper and its value is arbitrary: it has no rational meaning, but it depends on a convention.

In order to have a signs system operating, signs must be attributed to the language by the social community.

Please note the characteristics of the linguistic signs, considering each one carries a concept with infinite meanings:

- duplicity, because within the linguistic sign signified and signifier enter into a relationship;
- arbitrariness, because there is no obvious or concrete relationship between meaning and significance[11];
- conventionality, because between sender and receiver in the same linguistic community there is a communicative agreement.

[11] In fact, onomatopoeia demonstrates how the sound characteristics of the same subject are expressed differently, even through direct comparison between different languages.

2.2.2. Minimalism.

Chomsky is a U.S. linguist, a philosopher and a communication theorist.
His work is characterized by the search for the innate structures of natural language, which is a distinctive feature of man as a species, overcoming the traditional conception of linguistics (focused on the study of the peculiarities of spoken languages). Chomsky offers a formalized description, with almost mathematical level and structure, of grammatical and syntactical structures of our language.

The new theory starts off with a critique of structuralism: Chomsky argues that structuralism isn’t considering the creativity of language.
In order to understand how a language works, it’s not enough to discover the structure or even just describe, analyse and classify the components: according to Chomsky, structuralism omits the study of the ability of native speakers to produce and understand an indefinite number of sentences, which they have never heard before.
Creativity is instead a key feature of language using: to speak is to create something new, compared to the limited number of words and existing rules. It is neither a mere play of words, and it’s not just a mechanical implementation of grammar rules (however, the speech is dependent on grammar).
Then, grammar generates statements, but it does not produce them in a mechanical way: according to Chomsky, knowledge of a language is in the ability to produce and understand an eventually infinite number of sentences. Chomsky asserts the existence of creativity, which is governed by rules that can generate sentences, and therefore the language level of each speaker is a set of rules and principles.

Chomsky refutes the behaviourist theories of language acquisition: language is not learned, because there is a form of training or you are immersed in a cultural context, but it develops just as the human body. The language does not determine thought. There are not huge differences between languages.

Chomsky reassesses also the philosophical idea of the existence of mental objects (concepts), integrating it in the explanation of cognitive processes.
The cognitive psychologist Kosslyn experimentally proved the existence of mental images: they are figures, since they respect spatial rules[12].

---

[12] For example, the arrangement of parts in a real object is maintained in the mental representation: if there are given distances between certain points, proportionality is maintained even in the mental representation of the figure.

In an experiment on a map (an object with details located at different distances), the human being focuses especially on objects placed near the object, as in a mental scanning, actually comparable to scanning with optical sensors.

The theory of Chomsky shows the language characteristics in general principles, highlighting the importance of rules and representations: the human mind produces mental images about syntactic structures built as a hierarchical tree.
In addition, the mental rules can produce representations from some of them, the so-called *production rules*.

The grammar is a mental competence possessed by the speaker. In other words, it is the system of rules that lie in the speaker mind and constitute the linguistic knowledge: the theory of Chomsky is therefore based on innate knowledge of universal principles, which govern the language creation.

Finally, Chomsky supports the innateness of language: the expressive variety of language use implies that the brain of speakers contains grammatical unconscious principles.
Chomsky argues that the structural similarity in the languages justifies the existence of an innate universal grammar, made up of rules that can connect the limited number of phonemes that the vocal organs of the human species produce.
Evolutionary biologists have put forward a theory, based on two concepts:

- considering the evolutionary advantages, it is assumed a natural selection of the human species capable of communicating, at the expense of earlier hominids;
- in addition, we consider the inheritance of certain grammar disorders, giving them also a genetic component.

2.2.3. Universal Grammar.

The theory called Universal Grammar is characterized by the search for the innate structures of natural language, intended as a distinctive feature of man as a species. Obviously, this overcomes the traditional conception of Linguistics, focused on the study of peculiarities in spoken languages.

The idea that all languages have common fundamental aspects is not new in Chomsky's theory: philosophers of several centuries ago shared his basic idea, but the theory of Universal Grammar rehabilitates nativism, after a period of behavioural approaches to the language. Therefore, the universal elements that should form the Universal Grammar are those already discussed in the observations by Bacon and other speculative grammarians.

Chomsky believes that the human brain contains a limited set of innate rules for organizing the language. In addition, the theory postulates striking similarities between the deep structures of language, which have common properties hidden by the surface structures: therefore, under all languages there's a basic common structure.

The Universal Grammar is a linguistic theory postulating that the principles of grammar are shared by all languages and they are innate to human beings: this hypothesis was created to describe the acquisition of language and address the argument of the poverty of the stimulus[13]. Indeed, who is fluent in a language know what expressions are acceptable: the enigma key is to understand how the speaker can understand the restrictions of his own language, since expressions which violate the restrictions are neither perceived nor advisable during learning.
Chomsky claimed that the infant does not need to learn any specific rule of a language: rather, we must expect that all languages follow the same set of rules, although effects and interactions may vary with some linguistic universals parameters.

Of course, Chomsky does not describe a specific language, nor postulates that all languages share the same grammar. The theory seeks to identify a set of rules that explain the innate language acquisition by children and their approach to the construction of valid sentences. In particular, the theory research the innate rules that allow the construction of valid sentences,

[13] The argument of the poverty of the stimulus is contained in the question: how can a child learn his mother tongue so well, so quickly?

The lack of negative evidence is the core of the argument: the unfair terms are not presented to the language learner because they are grammatically incorrect for the speakers: who speaks the language does not consider the use of these expressions and does not to submit them to those who must learn.

For example, in Italian, you cannot associate the word some in a plural noun.

placing some restrictions as universal features of human language, in order to explain a very effective acquisition and solve the problem of poverty of the stimulus.

The grammaticality of a sentence is added to the question of the meaning and the ability to understand it. It's possible to build a sentence that is grammatical and meaningless[14]: nevertheless the linguistic problem shown by these expressions is different from that posed by (non-) sentences that are meaningful but not grammatical[15].

Recent experiments suggest that some portions of the human brain (crucially involving Broca's area, a portion of the left inferior frontal gyros) are selectively activated in the learning of languages that meet the requirements of Universal Grammar, while it does not activate when the languages grammar is artificially manipulated.

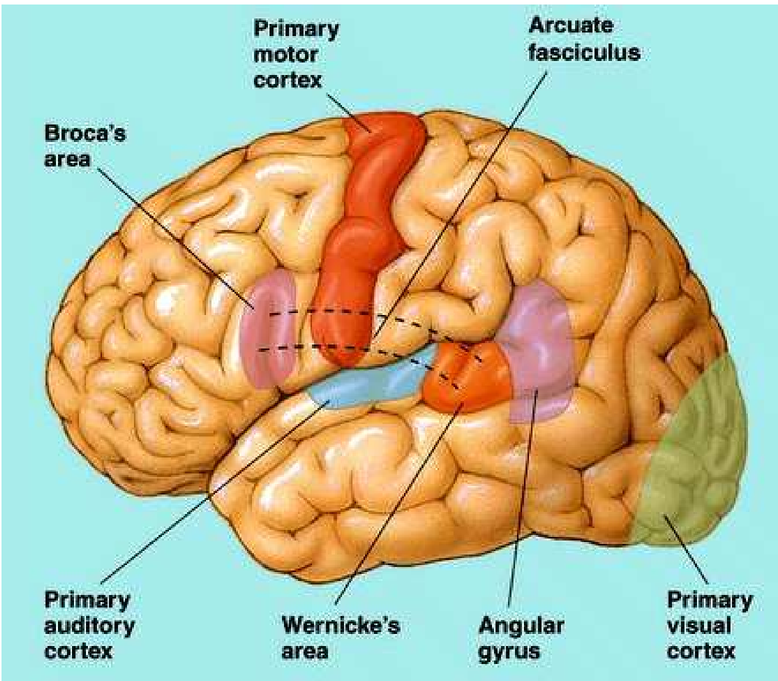

Figure 2.4. Brain areas activated by language (Broca's area).

---

[14] *Colorless green ideas sleep furiously.* [Chomsky]

This sentence was composed by Noam Chomsky in his 1957 *Syntactic Structures*, as an example of a sentence that is grammatically correct, but semantically nonsensical. Although the sentence is grammatically correct, no obvious understandable meaning can be derived from it, and thus it demonstrates the distinction between syntax and semantics.
As an example of a category mistake, it was used to show inadequacy of the then-popular probabilistic models of grammar, and the need for more structured models.

[15] A sentence with no sequential structure can be significant, but it cannot be grammatical: *am sandwich I the biting*.

These findings show the link between biology and the rules of the language.

The idea of a universal grammar is supported by the existence of creole languages[16], because they all share certain characteristics:

- syntactically, each one uses participles to form verbs in the future and in the past;
- each language allows the use of multiple negatives;
- a single change in the inflection of a sentence may be sufficient to create direct interrogative sentences, without further modify the construction of the sentence.

Of course, some linguists oppose the theory of the universal grammar.

For example, it is totally rejected by Sampson, who considers it a grammatical generalization: it would come from speculation on existing languages, rather than from predictive assessments about the possibilities of a language.

Some perceive the assumptions underlying the universal grammar as unfounded: in analogy with Bayesian inference, a language learner can assume the grammatical restrictions on his own, noting the absence of a certain class of expressions.

Finally, although the majority of the studied languages seem to share common basic standards, research is hampered by sampling bias, because linguistic death mainly hit those areas in which there were multiple examples of not-conventional languages.

[16] The presence of creole languages is cited as a support: these languages form and develop by the union of several societies, through the sharing of different language systems.
Originally, these only formed pidgin languages, while later they became mature languages, resulting in the development of a system of grammatical rules, with the emergence of native speaking individuals.

2.3. Transformational grammars.

The transformational-generative grammar is considered to be the most significant contribution to theoretical linguistics of the twentieth century.

A generative grammar is a set of rules that recursively specify the well-formed expressions of a language, by using a rewriting system. In detail, this is an algorithm to be used in order to decide the grammatical correctness of a sentence.
In most cases, a generative grammar can generate an infinite number of strings, starting from a finite set of rules: these properties are very desirable for a model of natural language, because also the human brain has a finite capacity, yet humans can understand and generate a huge amount of separate sentences.

The generative grammar should be distinguished from traditional grammar, because:

- the traditional grammar is highly prescriptive, rather than purely descriptive;
- it is not mathematically explicit;
- historically, it only analysed a restricted set of syntactic phenomena.

When the generative grammar was first proposed, it was a formalization of the implicit rules set known in a native language, which are responsible for producing expressions in it.

However, Chomsky rejected that interpretation: the grammar of a language is related to what a person needs to know to recognize an expression as grammatical, while it cannot be a hypothesis about the process involved in understanding or in the production of language.

In any case, please note that most of the native speaker would reject many sentences produced by a phrase-structure grammar: for example, although the grammar permits deep nesting, deep-nested sentences are not accepted by those who hear them. The limit of acceptability is empirical and variable, yet not catchable in a formal grammar.

Formally, a grammar G is a quadruple (X,V,S,P), consisting of:

- an alphabet X of terminal symbols;
- an alphabet V composed of nonterminal variables;
- a distinctive symbol S (also called purpose, or axiom of the grammar), which belongs to the alphabet V;
- a set P of strings pairs (v,w), which are called production rules, built on the union of the two alphabets, also denoted by $\alpha \rightarrow \beta$. The string α cannot be empty, while β can be.

The language generated by the grammar consists of all strings of terminals, which can be obtained starting from the symbol S and applying a production at a time to the forms of sentence gradually produced.

Mathematically, the language *L* produced by the grammar of *X*, *V* and *P* is:

$$L = \{a^n b^n, n > 0\} \Leftrightarrow \begin{cases} X = \{a, b\} \\ V = \{S\} \\ P = \{S \to aSb, S \to ab\} \end{cases}$$

The generative grammars can be described and compared by means of the Chomsky hierarchy, proposed during the Fifties. The classification defines types of formal grammars with increasing expressive power:

- the more general grammar is type 0. It defines no limits to the rules of re-writing, while constraints increase with type;
- the first two grammars can describe natural languages;
- while the last two, much more easy to handle by the computational point of view, are suitable to analysis of programming languages.

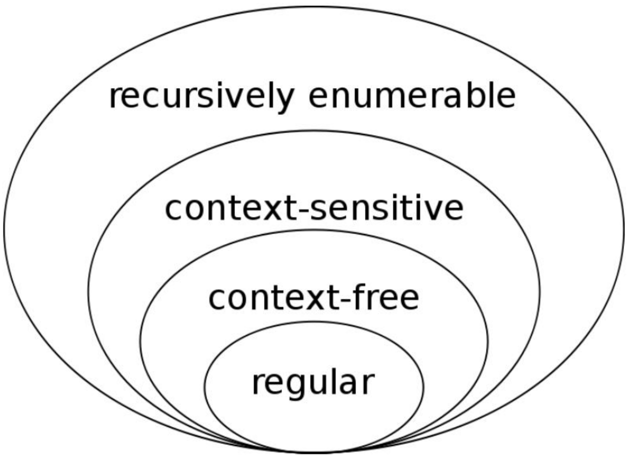

Figure 2.5. Chomsky hierarchy.

Starting from the hierarchical classification of Chomsky, other grammars were then introduced and defined.
They all are deterministic, in order to overcome difficulties in the field of automatic recognition of models:

- planned context-free grammars, for which there are constraints in the rules application by an intermediate state to another. In fact, every rule has a label and two fields that indicate the possible rules:
  - the first state (*success field*) contains the rules to be chosen in advance;

    - while the second (*failure field*) recollects the rules to choose from, if the rules contained in the first field don't fit the intermediate string;

- transformational grammars,

- specific grammars for building three-dimensional models (web and graph).

Chomsky said that even the phrase-structure grammars are adequate to the description of natural languages: then, he formulated the system of transformational grammar.
In this case, he tried a formalization of implicit rules known in a native language, which are responsible for the production of expressions, by creating a tool for the grammaticality evaluation of an expression (instead of mere checking of a hypothesis on the process of comprehension or production of language).
Therefore, all the transformational grammars are more powerful than context-free grammars

During the Sixties, Chomsky introduced two ideas for construction and evaluation of grammar theories:

- the distinction between competence and performance.
  People make mistakes in the real world language, but errors in linguistic performance were irrelevant to the study of linguistic competence (assumed as the knowledge which allows you to create and understand grammatical expressions). Therefore, the linguist can study an idealized version of the language, thus simplifying the linguistic analysis;

- furthermore, Chomsky distinguished between:

    - grammars with descriptive adequacy, which define for each language the (infinite) class of grammatical expression, thus describing the language in its entirety;

    - grammars with explanatory adequacy, which allow understanding the mental structures of the language, by describing the language grammar and by developing hypotheses about mental representation of linguistic knowledge.
      According to Chomsky, the nature of mental representations is largely innate, so if a grammatical theory is explanatorily adequate it must explain the grammatical nuances of languages in the world as minor variations of the universal model of human language.

Although linguists are far from forming adequate grammars for description, progress in terms of descriptive adequacy will occur only if the target will be the explanatory adequacy: the understanding of individual languages structures can only be achieved through the comparative study of a range of languages (called family).

Since the mid-Nineties, research on transformational grammar is inspired by Chomsky's linguistic minimalism. The *Minimalist Program* aims to develop concepts related to derivation and representation, which are peripheral aspects of the transformational theory:

- the economy of derivation is a principle which postulates that movements (in other words, changes in languages) occur in order to improve interpretability of obscure language emergencies;

interpretable emergency	*dogs bite*	the inflection gives meaning to the sentence, making it interpretable
interpretable emergency	*dog bites*	the inflection duplicates the information expressed by the noun subject, therefore it is interpretable

Table 2.1. Example: interpretable and interpretable emergencies.

- the economy of representation is the principle for which grammatical structures only appear when they're needed: as to say, the structure of a sentence should not be more complex than what is strictly necessary to fulfil the obligations of grammar;
- the derivation of syntactic structures should be uniform, that is, the rules should not be set for application in arbitrarily selected points of a derivation, but they should be applied through leads.

Chomsky (*Syntactic Structures*, 1957) believes that each expression has its deep structure, representing the sentence meaning, and it can be transformed into other distinct forms, with the same primary sense.

Therefore:

- a grammar generates sentences, which are characterized by a derivation tree;
- a set of transformation rules can change sentences, in order to allow and form the variety of known expressions;
- a series of morphemic rules reports any representation to a phonemes string.

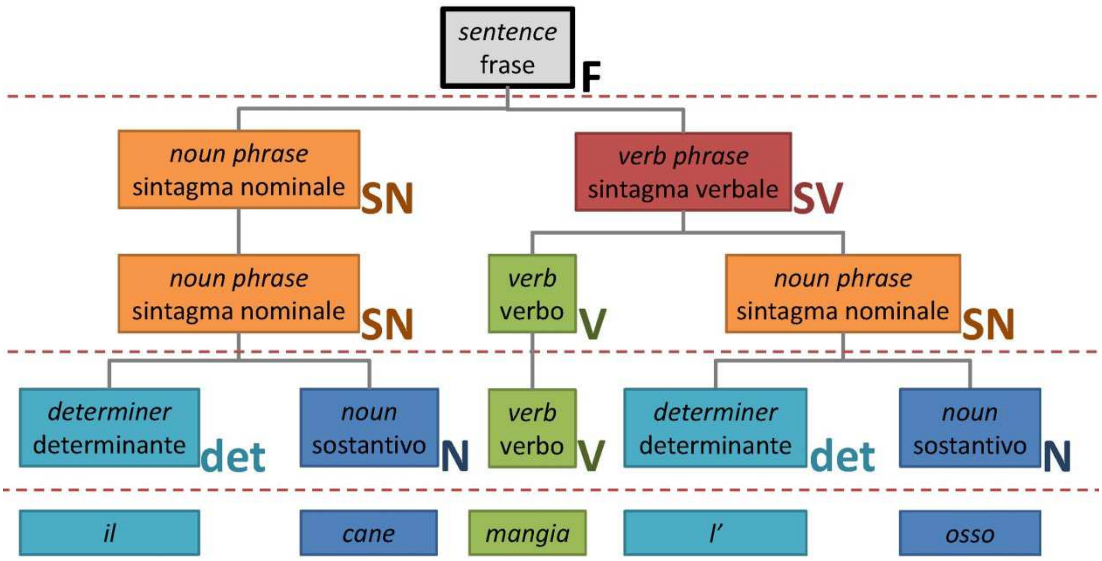

Figure 2.6. Example of syntagmatic indicator (phrase maker), represented by a tree.

$$[[\ [Il]_{det}[cane]_{N}]_{SN}[[mangia]_{V}[[l']_{det}[osso]_{N}]_{SN}]_{SV}]_{F}$$

Figure 2.7. Example of syntagmatic indicator (phrase maker), represented in textual form.

For the grammar that produced the derivation tree shown in Figure 2.6, generation of a sentence involves:

- a transformation, which combines the concept of time to describe the action;
- and a morphemic rule, which transforms the selected verb form in the correct expression of the verbal concept.

Conversely, extraction of a sentence belonging to a language with a grammar allows following two different paths:

- the sentential shape (succession of rules), from the starting point, up to the string of non-terminal elements;
- the derivation tree, where each node is classified by a symbol taken from both non-terminal and terminal elements: if an internal node is classified with *A* and its direct derivatives are X_1, X_2, … X_n, then A → X_1, X_2, … X_n is a rule within *P*.

2.3.1. Theory of Formal Languages.

The Theory of Formal Languages is part of the so-called Cognitive Science: its origins are in computer science, linguistics and psychology, although the nature of knowledge is certainly a philosophical problem. In fact, cognitive science refers to the teachings of Bacon and Galileo, because it combines understanding with the ability to produce: to understand is to build an artificial system that performs processes, whose practical results are identical to those achieved by a human being.

We use an underlying hypothesis: in some ways, reasoning is similar to calculation[17].

Mathematicians proposed several theories that converge in Gödel’s recursion theory, in which only the so-called recursive functions are computable.
In this special case, the algorithm concept is differentiable from that of steps in a procedure:

- a procedure is a finite set of instructions that can be performed mechanically, in a given time interval, with a certain amount of work.
  The arrest is imposed by external conditions:
  - if there are no further instructions at the end of a series of elementary instructions (definitely, they have finite number),
  - or if an instruction imposes a stop.
- an algorithm is a special procedure which stops for every input.

In general, a procedure defines a mapping (partial recursive function) from the set of all possible inputs to the set composed of all outputs.
If the procedure is an algorithm, the mapping is called a recursive function.

Of course, a procedure can be used to define a language.

The theory of recursion considers some basic functions (called initials), inherently simple:

- the zero function *Z*, which assigns zero to any number:

$$Z(x) = 0$$

[17] Leibniz and Spinoza imagined the mind as a system:

- the first thought it is governed by laws, logical rules and formal instructions,
- while for other, the brain is properly structured like a machine.

- the successor function *S*, which assigns to any number its successor:

$$S(x) = x + 1$$

- the projection function which, within a series of *n* numbers, selects the *m*-th:

$$U_m^n(x_1 \quad \cdots \quad x_m \quad \cdots \quad x_n) = x_m$$

A function is called primitive recursive whether it is an initial function, or if it is defined from the initial functions through correct combinations:

- through primitive recursion, which defines two properties:
  - the value of the function at *0*
  - the value of the function at *n+1*, which is solely dependent on the value of the function itself in *n*
- through the composition of two functions:

$$g(x) = f_1(x)f_2(x)$$

The commonly known mathematical functions are all primitive recursive, so that Hilbert imagined that the concept of computable function is coincident with that of the primitive recursive function.
When, intuitively, other functions appeared to be computable, the concept of primitive recursive function expanded, by introducing the more general recursive function.

Of course, there is a more rigorous way to define procedures and algorithms. In particular, please consider:

- the Turing machine, proposed in 1937, is an ideal machine and a mathematical entity. It's composed as:
  - a tape, which is divided into boxes with infinite length,
  - and a slider that moves along the tape, a box at a time, which reads and writes symbols in the tape boxes.

  The machine performs quintuple instruction: therefore, substantially, it can calculate any recursive function.

$$
\begin{matrix}
[* & 1(q_1) & 1 & 1 & * & *] \\
[* & 1 & 2(q_1) & 1 & * & *] \\
[* & 1 & 2 & 3(q_1) & * & *] \\
[* & 1 & 2 & 3 & 4(q_H) & *]
\end{matrix}
$$

Figure 2.8. Example: Adding Machine Turing algorithm[18].

- the grammar of Chomsky type 0;
- the algorithm of Markov;
- the lambda calculus, by Church.

If it is possible to simulate a procedure with one of these formalisms, then you can do it with any other in the list.

Given a Turing machine computes any recursive function, the theory of Turing machines is analogous to the theory of recursion.
Even Turing argued that any effective procedure can be carried out by one of his machines, but this claim is not provable (and it does not constitute a theorem).

In any case, a Turing machine, despite being ideal, however, is similar to a real machine (or even to a computer), as it coincides with its instructions.

In the Theory of Formal Languages, a language is defined as a set of strings.
A string is a finite sequence of symbols chosen from a finite vocabulary. In particular, in natural languages, a string matches a sentence, while sentences are sequences of words chosen from the vocabulary of possible words.

---

[18] In the example, it is assumed that the cursor is located on the first available box to the left and the input is what is written on the tape, in turn, to the left of the cursor.
The Turing machine adds *1* to any number, starting from zero:

- it only contains two steps: the one on current state q_1 and the one on final state q_H;
- the cursor is in the first box on the left (q_1) in which it is written *1*. After reading, it executes the first instruction: it overwrites *1*, stays in the state q_1 and moves to the right;
- again, it applies the first statement, because it is still in q_1 and it reads *1*. The cursor sums *1* to the previous value, records the result and it moves to the right;
- in the next step, the cursor is in an empty box and applies the second statement: the machine sums *1* to the previous value, records the new result and goes into the state q_H.

The machine no longer has any statement to execute, since no instruction begins with that state.

The output is the sum of four *1*, successively, obtaining number *4* in the end.

A grammar is defined as the quadruple

$$G = (N, T, P, S)$$

- *N* is the vocabulary of nonterminal symbols (general syntactic symbols);
- *T* is the vocabulary of terminal symbols (special, specific symbols);
- *P* is the set of re-writing rules;
- and *S* is the start symbol.

Chomsky divides languages into classes, creating a hierarchy, based on the difference in the type of constraints the grammars are subjected:

- unrestricted grammars (grammars of type 0), are more general and do not provide limits on re-writing rules.
  In general, the problem of determining whether a string is generated by this grammar cannot be solved;
- context-sensitive grammars (grammars of type 1). In this case, the limits of re-writing foresee that the left side of each rule (phrase) contains maximum the same number of symbols than the right part.

rule	$\alpha \to \beta$ $\lvert\alpha\rvert \leq \lvert\beta\rvert$
grammar	$S \to aSBC, S \to abC$ $CB \to BC$ $BB \to bb$ $bC \to bc$ $cC \to cc$
derivation (string)	$S \to aSBC \to aabCBC \to aabBCC \to aabbCC$ $\to aabbcC \to aabbcc$
language	$L(G) = [(a^n b^n c)^n n^n]$ with n≥1.

Table 2.2. Example: Type 1 grammar, derivation and language match.

- context-free grammars (grammars of type 2): the re-writing limits require that every rule should have only a single non-terminal symbol on the left side.

  The derivation of a sentence can be described as a tree of derivations (parse tree): according to this view, a sentence is not a string of words, but rather a tree with subordinate and superordinate branches connected at nodes.

  However, a description of natural languages structure with context-free grammars is actually impossible;

rule	$A \rightarrow \alpha$
grammar	$S \rightarrow S+T, S \rightarrow T$ $T \rightarrow T \times F, T \rightarrow F$ $F \rightarrow (E), F \rightarrow a$
derivation (string)	$S \rightarrow S+T \rightarrow T+T \rightarrow F+T \rightarrow a+T$ $\rightarrow a+T \times F \rightarrow a+F \times F$ $\rightarrow a+a \times F \rightarrow a+a \times a$
language	arithmetic expressions formed by symbols a, +, (…, e)

Table 2.3. Example: Type 2 grammar, derivation and language match.

- regular (finite state) grammars (grammars of type 3), the most simple: the right side contains only a single terminal symbol, or eventually a terminal symbol and a non-terminal for each production rule:

$$A \rightarrow aB \quad \text{or} \quad A \rightarrow b$$

  This kind of grammars corresponds to the mathematical algorithms invented by Markov (hidden Markov models).

Through the theory of discrete stochastic processes, a Markov chain require that for the system the transition probability from one state to another in a given time is independent of the sequence of previous states and only depends on the state in which the system instantly is. In fact, chosen and set two consecutive times k and $k+1$, the two vectors are random variables, with probabilities:

$$X_k = \begin{cases} x_1 & x_2 & \dots & x_i & \dots & x_n \\ p_{1k} & p_{2k} & \dots & p_{ik} & \dots & p_{nk} \end{cases}$$

$$\underset{A_k}{\rightarrow} \quad X_{k+1} = \begin{cases} x_1 & x_2 & \dots & x_j & \dots & x_n \\ p_{1k+1} & p_{2k+1} & \dots & p_{jk+1} & \dots & p_{nk+1} \end{cases}$$

In this case, the transition matrix A_k allows switching from one state to another in the system:

$$P_{k+1} = A_k P_k = \begin{array}{c} \\ x_1 \\ \dots \\ x_j \\ \dots \\ x_n \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & \dots & x_j & \dots & x_n \end{array} \\ \begin{bmatrix} a_{11} & \dots & a_{1i} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{j1} & \dots & a_{ji} & \dots & a_{jn} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{ni} & \dots & a_{nn} \end{bmatrix} \end{array}$$

In other words:

$$p_{jk+1} = \sum_{i=1}^{n} a_{ji} p_{ji} \qquad \forall\, j = 1, n$$

The system evolves, from a starting state *0*, in *m* steps, to a final state *U*:

$$X_0 = \begin{cases} x_1 & x_2 & \dots & x_i & \dots & x_n \\ p_{10} & p_{20} & \dots & p_{i0} & \dots & p_{n0} \end{cases} \quad \underset{A}{\rightarrow} \quad X_U = \begin{cases} x_1 & x_2 & \dots & x_j & \dots & x_n \\ p_{1U} & p_{2U} & \dots & p_{jU} & \dots & p_{nU} \end{cases}$$

where the total transition matrix is given by the total product of all matrices of transitions from each state to its next. Therefore:

$$P_U = AP_0 = \left(\prod_{k=0}^{m-1} A_k \right) P_0 = A^m P_0$$

The second equalities are possible if and only if all the transition matrices are equal: in this case, the (discrete) stochastic process is stationary, in the strong sense.

The same condition, combined with the knowledge of a number of states at least equal to

The same condition, combined with the knowledge of a number of states at least equal to *(m+1)* ≥ *n*[19], allows estimation of the transition matrix to be applied to the calculation of all the other states of the system, thanks to the algebraic system:

$$P_1 = A_e P_0 = diag^{(-1)}(matP_0^T)vecA^T$$
$$P_2 = A_e P_1 = diag^{(-1)}(matP_1^T)$$
$$\dots$$
$$P_U = A_e P_{m-1} = diag^{(-1)}(matP_{m-1}^T)$$

- the operator *vec* disposes in a vector the elements of a matrix[20];
- the operator *mat* repeats a given vector, forming a matrix;
- the operator $diag^{(-1)}$ disposes the elements of a vector in the main diagonal of a matrix.

The redundancy of information, in order to perform the least squares estimation of the transition matrix, is required to minimize the inevitable errors related to relief or to the model. The single alternative, in case the databases contain abnormal data, is to employ robust procedures to reject outliers.

In a Markov chain, there is a bijective correspondence between symbols and corresponding output. Conversely, in a hidden Markov model, states are hidden, that is, there is only one set of symbols and observations that allow probabilistic inference with respect to the sequence of states.
Therefore, a hidden Markov model is defined by:

- the set of states *S*:

$$S = (S_1, S_2, \dots, S_n)$$

- the observations sequence *Q*:

$$Q = (q_1, q_2, \dots, q_k)$$

[19] In fact, the optimum condition expects that the number of states is *(m+1)* ≥ *3n*, for reasons related to control of the reliability of the estimates.

[20] If the items are rows (or columns) of a matrix, they will be arranged in a bigger matrix, which will maintain the same arrangement.

- the probability of the states at the beginning of the observation period Π;
- the transition probability $p(S_j|S_i)$ related to the transition from the state i to state j;
- the emission probability $p(q|S_i)$, which is the probability of observing q when the system is in the state S_i.

From the model, we can calculate:

- the probability of an observations sequence, by means of the forward algorithm;
- the most likely sequence of hidden states, corresponding to the sequence of observations, by means of the Viterbi algorithm;
- the estimation of the model parameters, using the algorithm forward – backward.

2.4. The contribution of Linguistics.

The information related to geomatic engineering belongs to two different types, with many specificities of equal importance:

- the characteristics (or the entity), represented by computational or quantitative information, obtained mathematically and appropriately processed in order to derive an accurate local output.
  Spatial information (data quantity, location coordinates, graphs, etc.) belongs to this feature group;
- and attributes, which are qualitative data, which may be semantic, non-spatial, textual, disconnected from position or graphics, or metadata, fingerprints, descriptions.

Within this scheme, the object of study is provided by the attributes, while its geographical position is essentially constituted by the spatial information related to characteristic.

Without dwelling on the ability of used terminology, the crucial point is related to the perception of information by machines and humans:

- the attributes are easily thinkable for the human mind, because the brain optimally operates with semantic information. On the other hand, this kind of information is a real challenge to the interpretative ability of a machine;
- in the same way, to handle numerical information and graphics is very complex for the human brain. The same thing is instead perfectly simple to a machine, which manages data with simplicity and efficiency.

Historically, science elaborated complex computational strategies for the management of spatial information, untying automation from the semantic ability. In detail, there are machines especially programmed to handle graphical and numerical information with the highest efficiency, which are also able to process semantic attributes of the data.

The problem comes when you try to incorporate both of these processes into a single automated system: the historical separation of the two processes does not seem surmountable up to now (Egenhofer and Frank, 1992; Algarni, 1994; Worboys, 1994, Smith and Mark, 1998).

The study of natural languages seems therefore essential to the development of geomatic engineering techniques. In particular, we discuss about interpretation, integration and automation, in order to improve the mapping process. Unfortunately, perhaps, it will never be possible to achieve full automation in geomatic engineering, just because natural language is not complete.

The essence of language cannot be easily understood, nor defined (Jenkinson, 1967).

However, the language is the preferred vehicle of thoughts and this main communication instrument represents an embodiment of civilization and human culture, expressing verbal creativity. In fact, anyone can use language to optimize their own expressive ability; just as a double feature, at the same time it represents:

- a strength indicator, since it allows use and control of what is in the world
- and a weak characteristic, because there is not a language quote which, alone, allow government of objects.

The difficulty of capturing the essence of natural language - and, consequently, to optimize its use through technology - seems to leave a preferential use to mathematical and numerical techniques, born for the development of scientific fields.

Thus, the Geomatics (focusing on surveying and control) is an excellent example:

- interpretation is the process of identifying the characteristics of an image and communicating results to other users (Lillisand, 1994). In particular, the second part requires you to express in a graphical or textual form the recognition modes for details in the image: hence, natural language is the cornerstone of the process of interpretation.
  However, interpretation is constantly influenced by two factors, which are yet widely studied and well-known:
  - image resolution;
  - and scale[21].

  Often, however, we neglect the instrument techniques of interpretation: the natural language, or its resolution.

  It will be possible to provide computer-based systems a sufficient knowledge to manage categorizations, when it will be possible to properly appoint every specific detail, with an appropriate vocabulary. Names are the main part of the attribute information: in fact, the best description provides an unambiguous definition of the feature[22].
- specialization follows capacity and performance by users and it demonstrates that knowledge is integrated: a deep understanding of the discipline requires parallel information from other sectors, in order to improve knowledge.

[21] With digital systems, you can automatically switch to the selected reference (Agurous, 2000): therefore, users can locate and represent specific features autonomously.

[22] Men need names, to conclude the visual interpretation. In order to simplify the task to both humans and computers, the characteristics nomenclature should be clear, precise, unambiguous and concise.

In the field of geomatic engineering, the integration between sectors is evident within GIS[23]: the objective of obtaining consistent data necessarily requires the use of a uniform vocabulary for an accurate description of spatial information coming from different sources. However, only rarely non-spatial data are included in the systems[24].

- automation comes from the integration of different sciences and technologies, in order to solve a particular problem.
  In addition, the automation depends on the database design, created with clear and precise terminology for attributes and objects.

- language limitations, mainly due to human (in)capacity toward language optimization.

  In a theoretical way, every language allows generation of any vocabulary or definition[25]; however, no language is complete, or sufficient, or adequate in itself.

  We can also add a list of problems, which are strictly related to the computerized management of semantic data:

  - words which indicate verbs, nouns or adjectives with just the same formal structure;
  - or words having different meanings, with the same form;
  - long words, which are memory-consuming and worsen software management;
  - synonyms, which are to be associated to the original concept;
  - generic words (such as *data*, *information*, *model*, *concept*), which have different meaning depending on the user's perception.

The cognitive sciences cover intermediate topics between technology and human sciences: sometimes, in this regard, we talk about a new discipline, called cognitive engineering.

Science, up to half of the twentieth century, sought universal laws.
Subsequently, it abandons the goal of unifying different, irregular and complex phenomena, according to the identification of common elements. Systems with the same structure are classified according to different behaviours, or epistemological categories of simplicity; characters of order and regularity are opposed to complexity, chaos and disorder, with the uncertainty principle.

---

[23] *Geographic information systems (GIS) integrate data from various resources into a single, homogeneous system.* […] [Egenhofer and Frank, 1992]

[24] There are many examples about lack of integration between the attribute information, which often are difficulties related to use and understanding of the reference nomenclature.

[25] Associating a name is to indicate description and function of objects, thus building attributes.

When the system of science fell into a crisis, the desire changed to description of reality in its variety and changeability (relativity theory), whereas all forms of knowledge is incomplete and limited.

The description of reality, extended in space, cannot change over time and every phenomenon or process is connected to the rest.

In order to describe this system, the fields of Logic, Mathematics (Cybernetics mainly), Physical Science, Philosophy (structuralism, linguistics and *Gestalt Psychologie*) were combined in the Information Science: in fact, some terms (*status*, *stage*, *relationship*, *signal*, *noise*, *feedback*) belong to both the information technology sciences and the human science (in detail, Semiotics).

During elaboration of a scanned image, some possible matching involves:

- geometrical aspects, as they operate in the image space, in order to search suitable information for the formation of the corresponding model,
- and topological aspects, carried out in the object space, with the adoption of coordinate systems which must be congruent with object reconstruction.

Other kinds of matching can also involve:

- geographical aspects, in order to define the object position in the object space (this process is called geotagging),
- and even syntactic and semantic aspects, in order to interpret the scene in the image space and analyse the possible sequence.

chapter 3

Case study: an image representation

It's now the time to prove the eventual reliability of the proposed thesis.
This chapter just deals with further representations of a single image, taken by a common digital camera, aiming to a textual description that could be eventually traced back to the original scene.

The image we chose basically respects two conditions:

- it should be unpublished and hardly recognizable, in order to reduce the impact of previous experience during the reconstruction phase. This requirement directly refers to the Gestalt principles, aiming to limit the effects of known features;
- and it could come from an non-professional glimpse, in order not to align to formal features of centring, proportions and others.

Detailing what follows, we'll come to an image description, targeting to build a conceptual map that streamlines the further development of (semi) automatic software: it should prove to be a valid tool for interpretation and classification of representations, under unambiguous and reliable rules.

In this first work, the expert system that performs the reconstruction is actually a human being, with good skills at drawing: at this stage, the primary focus of the thesis is to prove the reliability of connections. Besides, it's fundamental to check the procedure in order to define the whole process of image recognition.

3.1. Case study: the image, as a photograph.

As an example, to be used as a case study, we'll refer to the photograph in Figure 3.1, on the landscape at the summit of Mount San Giorgio (Canton Ticino, Switzerland).

Figure 3.1. Picture on the landscape at the summit of Mount San Giorgio [L. Mussio].

In what follows, therefore, we'll try to come to the image description, targeting to build a conceptual map that streamlines the further development of (semi) automatic software: this should offer interpretation and classification of representations, as much as possible unambiguous and reliable, which can be eventually traced back to the original scene.

3.1.1. Histograms.

A histogram is an indicator of some key factors in a digital image: it's a graphic that represents, in relative terms, the number of pixels referring to a certain tonal value in a single image. Therefore, it's an indicator of the statistical distribution about brightness in the picture:

- on the horizontal axis, we usually represent values related to shadows, to midtones and to highlights[1]. In other words, the histogram shows a set of 255 columns, side by side:
  - on the left of chart, there's full black (where values of three colours[2] are all equal to 0),
  - while, on the right, there's full white (and values RGB are equal to 255).
- finally, height of chart is directly proportional to the number of pixels concentrated in a tonal range.

Range could be classified as:

- tonal range, when it describes amplitude of the colour shades
- or dynamic range, if it refers to amplitude of brightness levels caught by camera sensor.

In other words, value of range amplitude (either tonal or dynamic) is directly proportional to precision and accuracy of the photographic reproduction, from the point of view of both colour shades and brightness gradations. For technical images, too, it's important to maximize the range value: it's a matter of trying to get rich, blended colours or tones, from deep black to bright white, avoiding as much as possible to work with washed-out images (that appears to be flat).

Figure 3.2. Example: the same image, with low-to-high dynamic range.

---

[1] In order to refer to different zones of both the tonal range and the dynamic range, we traditionally use terms as follows:

- shadows are the darkest areas of an image
- highlights are, on the contrary, the brightest parts of the picture
- and midtones are in the middle

[2] Naturally, the three colors are the fundamental ones: Red – Green – Blue.
In fact, this definition comes up with the acronym RGB.

In a properly exposed photography, all shadows, all midtones and all highlights contain data. For picture in Figure 3.1, the image shows a special concentration of midtones pixels, but too has information in both the very dark and very bright areas, without having completely black zones, or completely white ones (these areas would have been pointless, since they're not showing data).

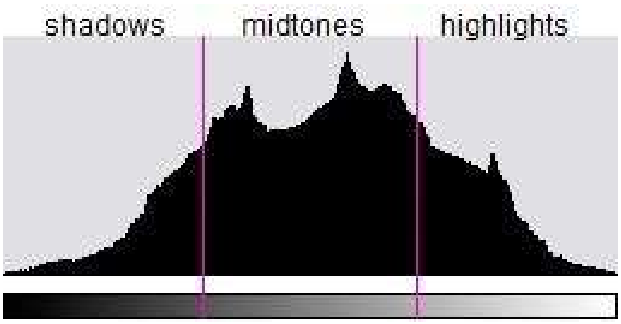

Figure 3.3. Example: histogram for a properly exposed photography.

Vice versa, in an underexposed photography, some pixels are so dark they appear black: they're on the first column of the chart. In a case like this, the histogram features are:

- total absence of bright whites on the scene (the effect is shown in the pink area of Figure 3.4)
- and a phenomenon called clipping: some pixels are out of chart, in other words the image contains areas that are full black and do not convey information. Black pixels are represented in the first column of chart, on the left.

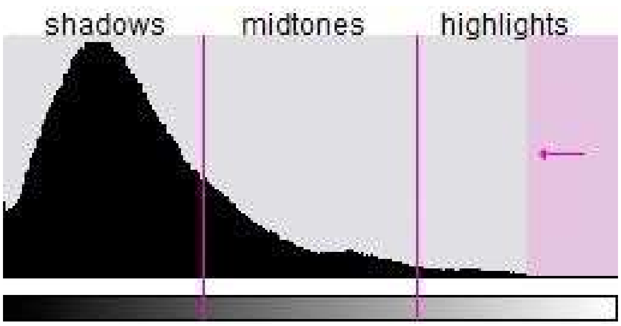

Figure 3.4. Example: histogram for an underexposed photography.

Finally, in an overexposed photography, some pixels are white and appear on the last column on the chart. In a case like this, the histogram features are:

- total absence of deep blacks on the scene (the effect is shown in the pink area of Figure 3.5)
- and a phenomenon called clipping: some pixels are out of chart, in other words the image contains areas that are full white and do not convey information. When the right column is extremely high, there are many completely white pixels in the image.

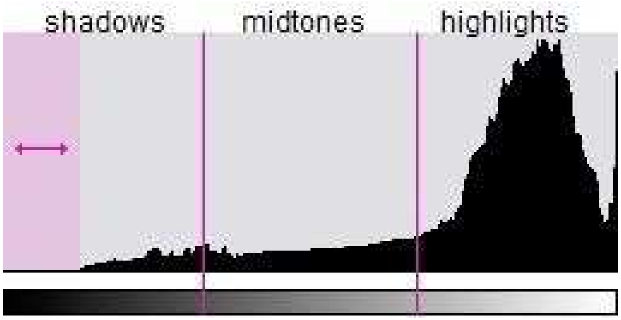

Figure 3.5. Example: histogram for an overexposed photography.

There are several methods to calculate the value of a pixel; therefore there are several types of histograms, representing the same image among different criteria:

- the RGB histogram is the most widespread one, as it calculates the value of every pixel overlapping the three colour channels. It can also show a single dye at once, performing a decomposition of signal. It's a purely mathematical representation, therefore abstract, that suitably fit effective identification of any clipping phenomena;
- the brightness histogram is based on the analysis of the three RGB channels and calculates a weighty average of RGB values for every pixel. Naturally, the values correspondent to the three colors do not have the same weight:
  - green has percentage fixed to 59%, therefore the chart is built to let green prevail on other colors: this trick allows generating a representation that is as close as possible to the human sensation of brightness;
  - red weights 30%;
  - while blue has percentage to 11%;

- finally, colors histogram allows the visualization of all the three colors (with all possible combinations) on a single chart:
  - grey zones represent mix of colours;
  - while, where a colour channel exits the chart, there are some areas on the original image in which information has been lost (it's a limited loss when unbalance occurred to a single fundamental dye).

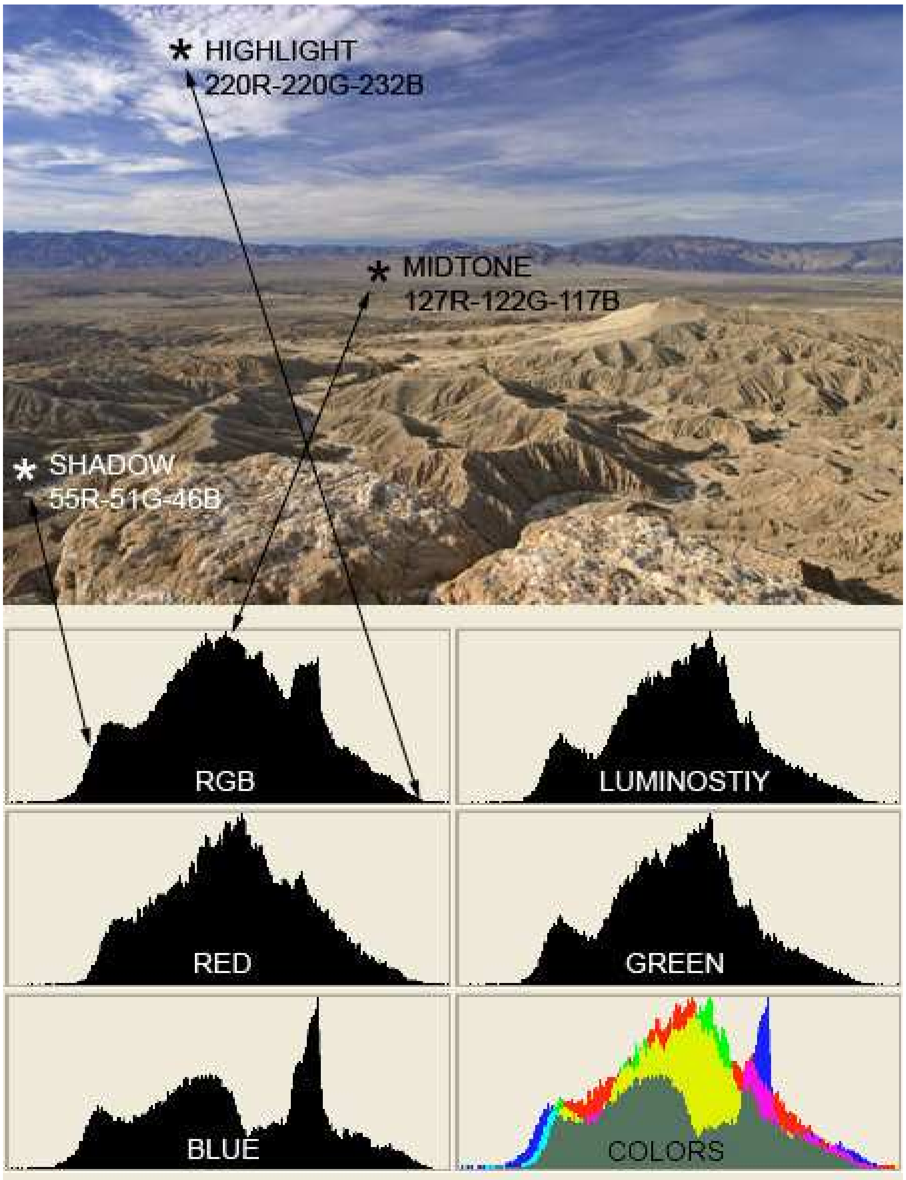

Figure 3.6. Example: RGB histogram, brightness histogram and histogram of the colours, all referring to a single image.

It’s clear that every type of histogram has its own features to represent pictures of reality.

In order to evaluate the contrast of an image, through the reading of its histogram, we can refer to the dimension of the area in which tones concentrate the most: contrast is good when the area is extended and contains many tones, from brighter to darker. Naturally, final effect mostly depends on the subject caught in the picture.

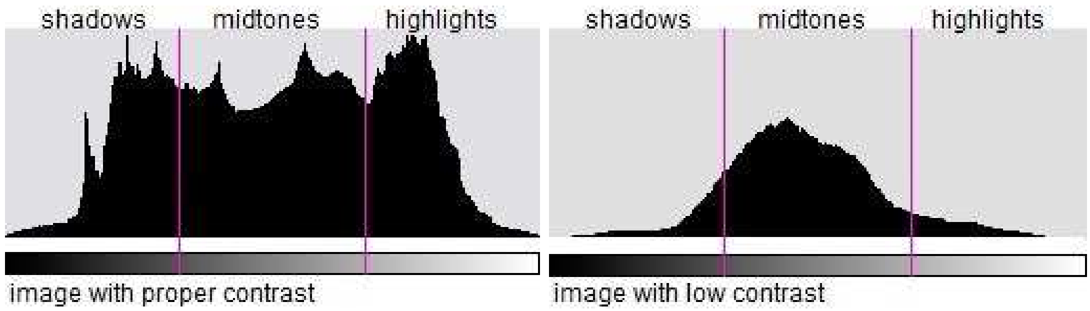

Figure 3.7. Example: evaluation of image contrast.

The histogram is a tool that can only offer an approximate information in the image: in fact, histogram width is equal to 256 pixels, therefore it can show 256 tonal values maximum (while an RGB image in 8 bits could bring 16,7 million tonal values). Thus, the use of histograms strictly depends on quality, correctness and reliability of interpretation[3] brought to available information.

Even starting from the original image in Figure 3.1, it’s possible to calculate the histogram: the chart indicates the chromatic composition of picture, nevertheless without information on area distribution.

[3] For example, when picturing a snowy panorama, there’s no point strictly referring to previous indications: the image could have a technically perfect, balanced histogram but show grey snow. In fact, snow generates reflections and a digital camera would compensate the light excess with underexposing the picture (but the snow will never appear white and clean).
Vice versa, on a properly balanced image, with bright snow, the histogram will be typical for an overexposed picture (concentrated on the right, with most data in highlights).

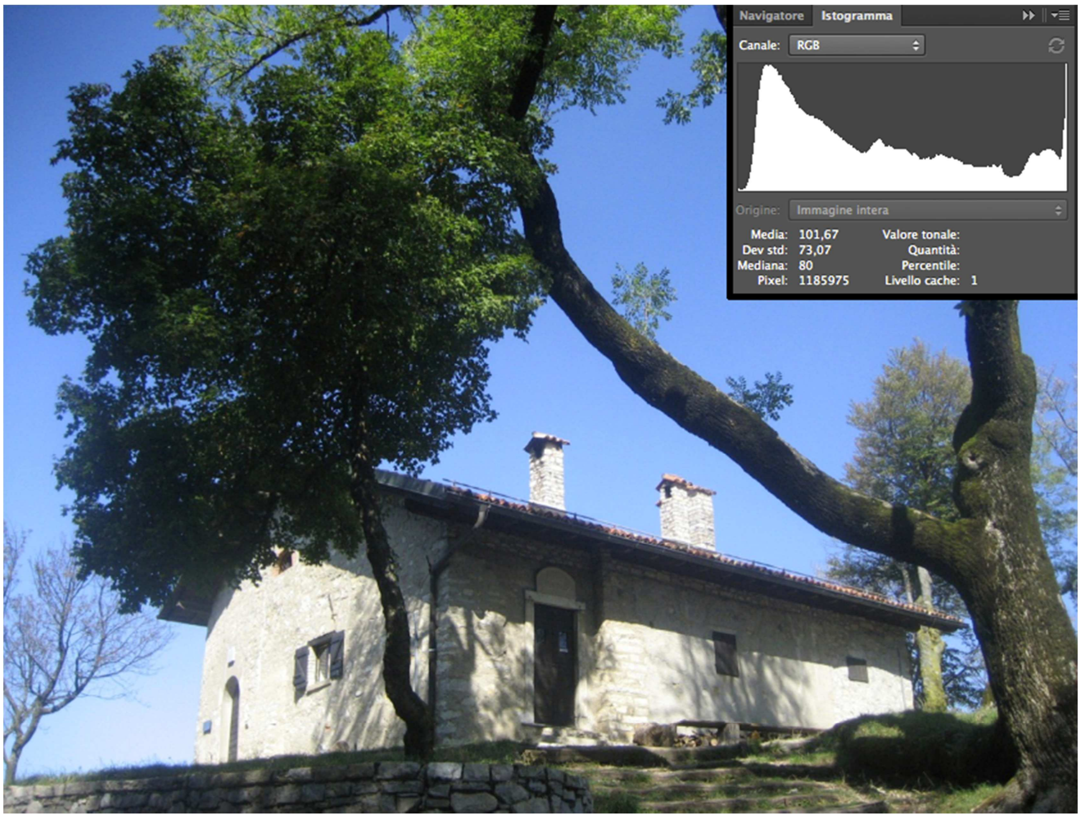

Figure 3.8. RGB histogram for Figure 3.1.

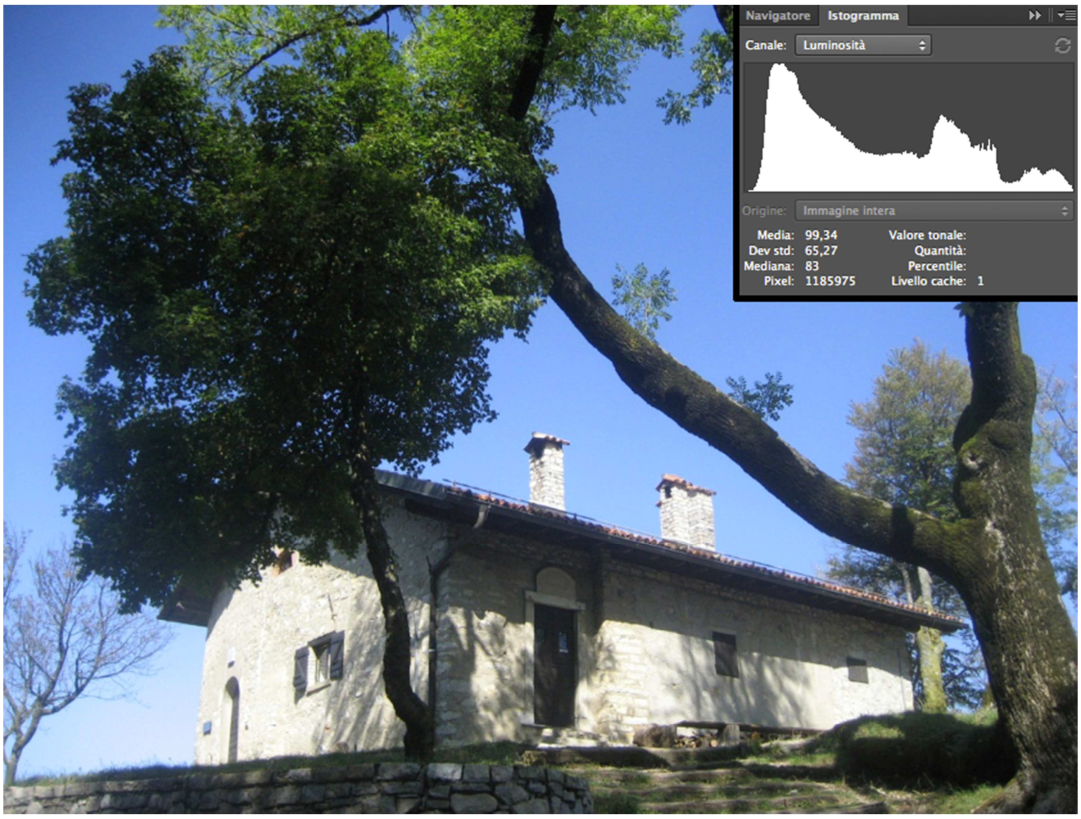

Figure 3.9. Brightness histogram for Figure 3.1.

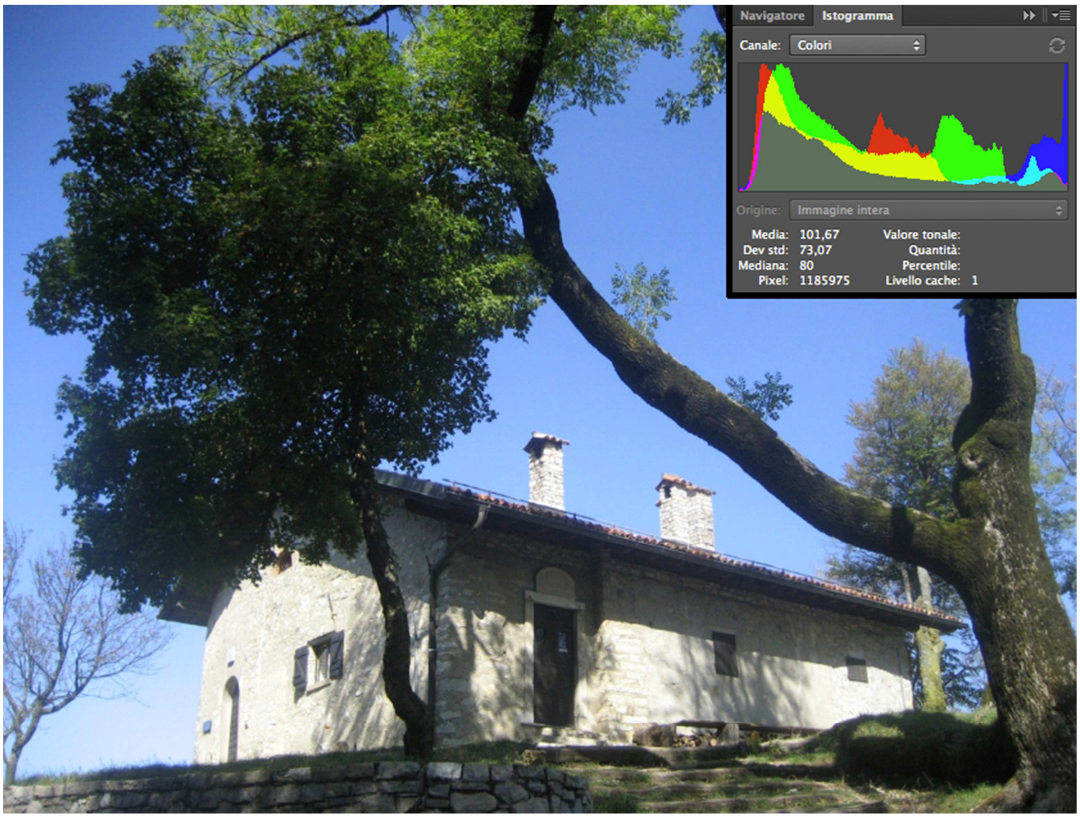

Figure 3.10. Colors histogram for Figure 3.1.

In fact, this information doesn't convey knowledge, as interpretation of a picture from its histograms generates many ambiguities.

Finally, readings shall consider that every histogram is sensitive and strictly related to the analysed portion of image: if there isn't a selection, histogram refers to the whole image; on the contrary, if a selection appears on the image, histogram will show values relevant to that selected portion. This hint suggests that histogram technique is especially useful for acknowledgement of closed surfaces in images, to ease interpretation and improve association to the relevant pattern or label.

Please note that, up to now, there is not any information regarding dimension, nor number of objects visible on the image.

3.1.2. Lines' graph.

Back to the image shown in Figure 3.1, we'll proceed trying to find all discontinuity points: the aim is identifying all pixels where colour difference is beyond a certain value, through building a suitable 0/1 matrix.

Then, it's possible to elaborate the image building up the lines, through edges recognition techniques: at first, we're going to work on safe lines, connecting contiguous point (alignment of discontinuity points, as identified at previous step).

Figure 3.11. Discontinuity matrix on image in Figure 3.1 (Laplacian image).

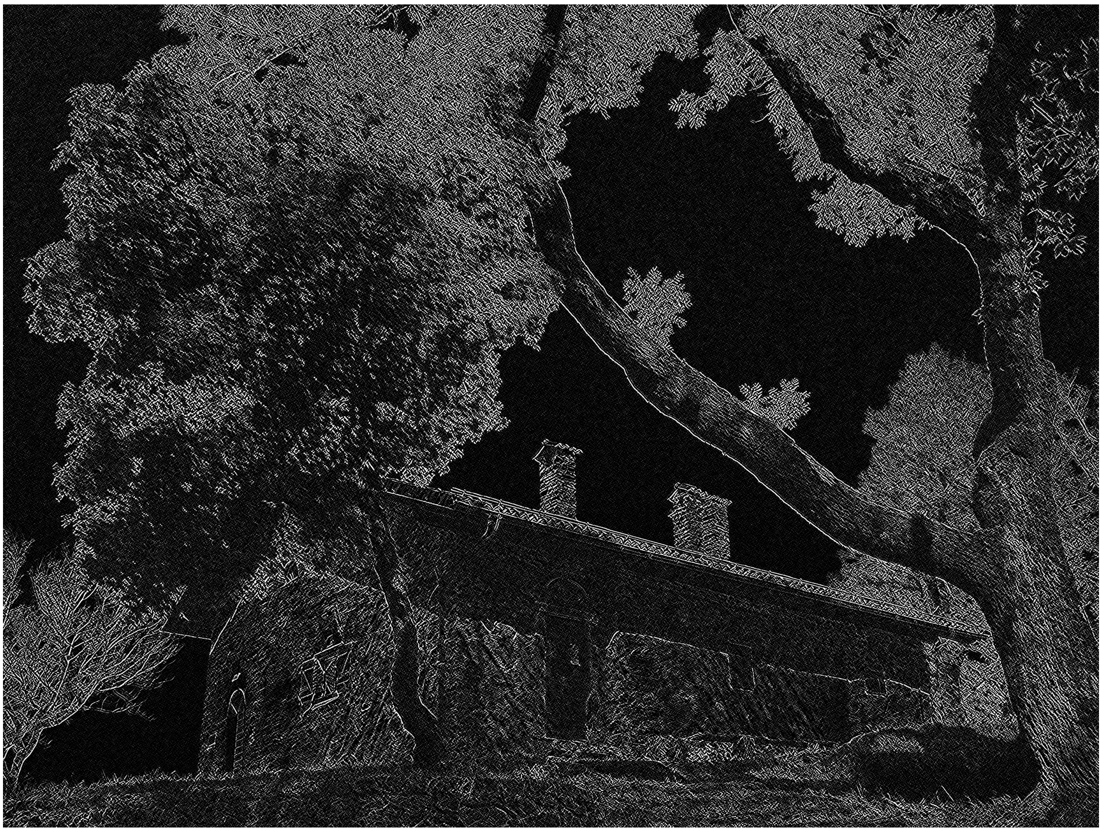

Figure 3.12. Lines reconstruction, on image in Figure 3.1 (gradient image).

Nevertheless, in the image we're considering paradigmatic (Figure 3.1), it's possible to locate a line in which the contrast of the overall image suddenly changes: the expert system we're looking for should know how to handle this information and maybe suggest for the presence of a shadow in the panorama. In detail:

- the shadow comes from the presence of a physical object, that shall be located with respect to trigonometrical calculations and photogrammetry principles
- and lines that are broken by this "contrast line" shall therefore be reconnected, with the aim to improve reading and understanding of the overall image.

Then, we would add to the lines' graph other "suggested lines", which are coming from forced alignment of neighbouring lines: the continuation condition being thus replaced to a weaker one (sole contiguity), as a function of possible intervention of a (human) operator.
This process is deliberately thought type semi-automatic, by virtue of intervention complexity:

- two contiguous lines could be associated for their closeness, thus establishing a minimum distance to allow automatic forced alignment;
- but please consider that an incorrect association of two close lines, that shouldn't have been connected, will be very harmful to interpretation of the overall image;

- therefore, a slight improvement of automatic interpretation capability could be possible through implementation of a system able to learn on its own experience: in this case, operator interventions are limited to the first period of software operations, while they will be reducing progressively when occurrences get repeated associations.

Solution of engineering issues requests construction of data in a formal structure: it's therefore necessary to identify geometrical data and their relationships with features.

In order to describe a formal structure for data, there's the chance to apply the graphs' theory: this allows association of thematic elements and geometric ones, forcing topological relationships between components in a complex object.

3.1.3. Relational description.

In order to finalize the relational description for data in an image, the key elements in descriptions (primitives) shall be indicated by features (attributes) defined in name and value, both numeric or symbolic.

Each primitive represents the description of a portion of the image: it's independent from any other and it's defined by its attributes list, as stated in the following expression.

$$primitive =$$
$$= \{(name_{attribute\ 1}: value_{attribute\ 1})(\dots : \dots)(name_{attribute\ n}: value_{attribute\ n})\}$$

The whole set of primitives is shown in the relevant list *P*, as follows.

$$P = \{p_1, \dots, p_n\}$$

Data relevant to images, models or maps can be described at different abstraction levels:

- at the lowest level, in a very realistic approximation, the image is described through area elements (area-based description), using grey values shown by pixels;
- at an higher level, we use features like points, lines and regions extracted from image (feature-based description): the correspondence between vector descriptions of attributes supply approximate values to simple local methods and therefore this technique proves suitable for surfaces and objects reconstruction;
- at the highest abstraction grade, data description is relational type (relational description or structural description) and includes relationships between primitives features, tracing the theories related to perceptual organization designed by the Gestalt psychology[4].

Area-based matching names data representation based on an area element (pixel), which is described by its coordinates and grey intensity.

Every point that has been analysed in order to find a correspondence is the centre of a window in the first image (reference image); then a statistical comparison occurs with masks

[4] This is a theory of mind and brain from the Berlin school.

The principle maintains that the human eye sees objects in their entirety before perceiving their individual parts: Gestalt psychology tries to understand the laws of our ability to acquire and maintain stable percepts in a noisy world.

in another frame (target image). Thus, similarity is refined by measuring disparity between pixels in the target image and those in the reference image.
Therefore, the area-based methods are suitable to supply a set of points where the object surface could be considered regular.

Amongst area-based methods, the Least Squares Matching is considered to be the most complete technique.
The algorithm considers pixels $f_1(i,j)$ in a reference window in an image: with varying a parameters' set, it tries to find another area in another image where pixels $f_2(i,j)$ are as similar as possible to the $f_1(i,j)$ ones. The "best fit" position supplies the minimum sum of squares between differences of radiometric values for correspondent pixels in the two masks: the following relationship perfectly fits the condition.

$$\sum_i \sum_j [f_1(i,j) - f_2(i,j)]^2 = min$$

The search for the minimum position occurs through sampling of the floating window, made according to the parameters of a mathematical model that interprets the effect of perspective deformations.
The Least Squares Matching method obtains estimations with sub-pixel accuracy; nevertheless, the convergence radius is scarce and this technique can prove to be unreliable, as it rigorously should only have local validity.
Therefore, it becomes appropriate to introduce additional constraints between information[5], in order to improve algorithm performances.

The feature-based matching uses data representations based on points, lines or regions identifying discontinuity lines and edges of every object in the scene.
The number of possible correspondences between n features of a couple of images is equal to $n!$: thus, it's impossible to evaluate the whole series, which also include a number of false correspondences. In this case, too, it becomes appropriate to limit the search space, with the introduction of constraints.
For these reasons, the feature-based methods are complementary to area-based techniques. In fact, applying a combination of the two procedures can identify and select

[5] The matching operated between frames with strong geometric relationships (Multiphoto Geometrically Constrained Matching) represents a robust version of the correspondent algorithm for least squares: improvement of performances comes thus from both the application of geometrical constraints and the chance to use images redundancy.

correspondent points on images: after verifying the hypotheses, localization of points comes out to have a sub-pixel accuracy.

Nevertheless, all methods using data descriptions at the lowest level request the use of supplementary information to apply the epipolar constraint, while use of relational descriptions only allows to solve the issue of correspondence, without requiring a priori knowledge of an object shape.
The matching can only be solved with high level data descriptions, through application of constraints that limit the search space for possible correspondences.

A relational description can supply the model for every structure.

The elements for a structural description for processes of image elaboration are:

- a set of features (points, edges, regions);
- values of attributes associated to features (coordinates, length, orientation, area, grey value);
- spatial relationship between features.

The basic unit of description is the relationship R_i, that associates elements contained in a set called A. Then, taken an object O, the relationship R_i with grade $i = 1, \dots N$ is defined as the set of the Cartesian product of the set of its primitives:

$$A_i = \underbrace{A \times \dots \times A}_{I} = \{a_1, a_2, \dots, a_n\} \times \dots \times \{a_1, a_2, \dots, a_n\}$$

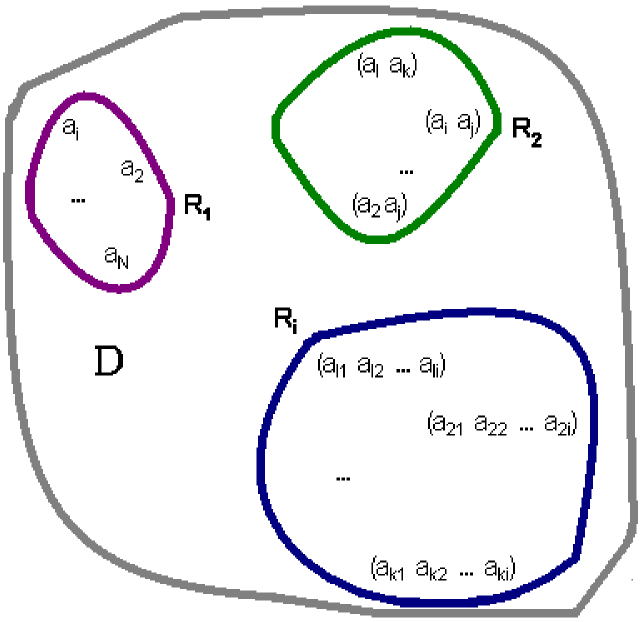

Figure 3.13. Relational description of the object O.

3.1.4. Search methods for the relational matching.

The contribution by the Gestalt theories to the process of visual interpretation is then involved into the development of knowledge related to grouping phenomena; actually, it derives from the study of perceptual organization, born in open contrast to dominant atomistic explanation. Groupings made according to perceptual organization principles occur without having any a priori information on contents in a scene: we define relationships between elements in the image, building them up trying to maximize their likelihood to maintain its own features up to the end of interpretation process.
In fact, the success rate in identifying the fundamental structures in the scene just derives from the ability to extract clusters characterized by low non-accidental chance to occur: it's important to isolate this kind of clusters with respect to a random distribution of elements.

The perceptual organization represents the association process of a significance level for a potential group of elements in the image.
Measurement of relationship significance between elements in the image comes from statistical information on grade of non-accidentality of occurrences.

There are many combinations for structures to be considered in an image, nevertheless there only some relevant ones, which can be useful for interpretation.
According to Lowe, it's important to study the factors that limit potential relationships classes to a smallest set of relevant structures:

- selecting a class of significant relationships;
- introducing a measurement to select obtained structures, with respect to their accuracy and their deviation from ideal relationship;
- applying recursively the principles of perceptual organization.

The core issue is to limit the computational complexity of correspondence problem, reducing the search space; perceptual organization schemes reach the target searching for rare structures that allow eliminating problem ambiguities (false target).

The perceptual grouping reveals relationships between image elements, that could likely stand till the end of scene interpretation process.

The number of relationships to be considered raises with the square of the number of elements in the scene: thus, for n elements, the maximum number of couples to be examined is equal to $n\frac{n-1}{2}$.

Also assuming that the camera point of view (fully comparable to the eyes) is independent from the object on the scene, only some relationships are likely to maintain through all the phases of image elaboration: properly, they are stable relationships under many points of view. For instance, see the three points in Figure 3.17, which have two types of significant relationships: with collinearity and constant spacing, they're invariant with respect to a wide range of points of view.

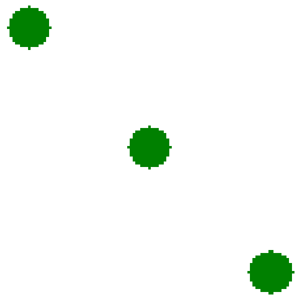

Figure 3.14. Example: collinearity and constant spacing.

On the other side, the three points displaced in the shape of an equilateral triangle (shown in following Figure 3.18) could refer to a generic triangle, as they can vary depending on the point of view. Silhouettes are 3D objects and their 2D representation is a projection onto the retina: thus, perception of three points as an equilateral triangle depends on the position of this page with respect to the line of sight.

Figure 3.15. Example: three points displaced in the vertices of an equilateral triangle.

There are very few relationships that preserve itself under every point of view (they're so called invariant): the simplest ones are collinearity and proximity.
Furthermore, the use of a proximity relationship (that can project elements that are close in the object space into close ones in the image space, too) offers a double advantage:

- it supplies an additional measurement to express a judgment on the significance of localized structure,
- and it allows reduction of connections number to be analysed.

There are other relationships that can resist to most points of views, as parallelism and constant spacing between elements in the scene.

Two parallel lines, even in a context rich with information, represent a stronger relationship when they're close (when they're not, we could miss the information).

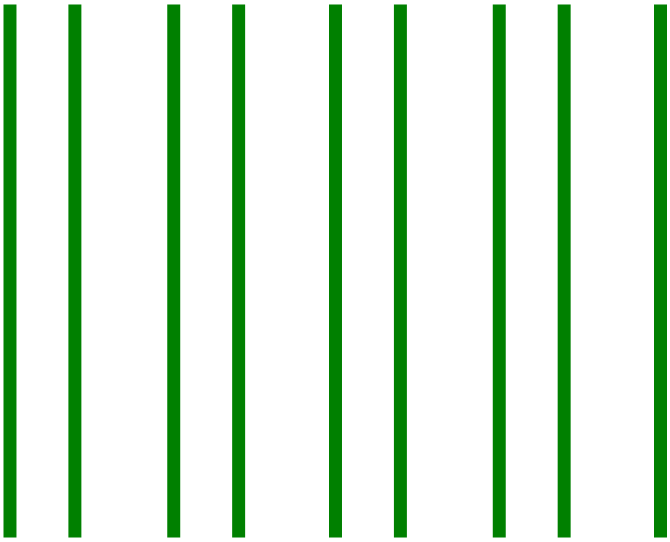

Figure 3.16. Example: parallelism.

Analysis of relationships with low likelihood for accidental occurrences represents an interesting matter to start the recognition process: recursively applying the perceptual organization to a set of relationships, it's possible to create structures with scarce probability of being randomly generated.

The simplest way to combine relationships into new structures is to treat each and every new connection as an original element in the scene. With this trick, the built up primitives are:

- virtual type,
- with higher significance than the original ones
- and they can compose structure with even higher significance.

3.1.5. Areas tree.

Back again to the image of Mount San Giorgio (Figure 3.1), it's possible to proceed with identification and localization of all areas with closed boundaries: during this phase, all closed lines are to be selected as pattern elements in the image.

Then, we must associate a prevailing colour to every area, through a dedicated histogram built for each pattern (inside the zone). In fact, we already considered the chance to get back to a histogram analysis after recognition of closed surfaces in the image, to ease interpretation and association to correspondent label.

Every object with homogeneous features (closed boundaries and colour) can thus be associated to a group c_n, con $n \geq 0$. Each cluster shall be suitably described within a cluster database, specifically built for the case study image. Please note that:

- in order to automatically manage the process, it's fundamental to have a unique key to access and read the database
- the specific texture of each object is defined by its own colors histogram, but of course it can only be "informative", as every occurrence will present different chromatic percentages. Therefore, yet also in this case, the association between objects and clusters has a probabilistic component
- the specific geometry of an object could even be assigned a priori, because human artefacts have defined and identifiable shape, unless perspective. It can be useful to exclude the image original perspective, through orthogonalization techniques, geometrical calculations and photogrammetry principles: in order to use full available information, we should always work on orthorectified images.

key	texture	geometry
P1	sky	not assigned
P2	tile	not assigned
P3	trunk	not assigned
P4	green parts	not assigned
P5	fixtures	cuboid, unless perspective
P6	details	cuboid, unless perspective
P7	gutter	cuboid, with contiguous sections
P8	stone	not assigned

Table 3.1. Cluster database for Figure 3.1.

It's therefore clear that, for every future occurrence of a cluster c_x in an image, a computer could automatically recognize the object, performing the association to best fit description.

The procedure of the *nearest neighbour* allows connection of nearest points to an emanation point. When transferring the function of an emanation point to each one of connected points, the procedure can be repeated to generate new connections until the set of points has been completely analysed.

Then, we can use robust procedures for separation and union:

- in order to interrupt the chain of connections when it's too long (over an arbitrary limit, fixed by user)
- or with the aim to connect different chains, when they're too close (under an arbitrarily fixed distance)

The results obtained since now can be further modified, corrected and finally improved with applying these procedures repeatedly and sequentially. When a reproduction point has been found, we'll stop the process: this will happen anyhow, even though the point could be a relative optimum and not the absolute one.

Then we may complete the Table 3.1, by adding the association to the model related to cluster, with respect to colour identified as prevailing. The result comes up with a new table, which also include the labels for recognized patterns.

key	texture	geometry	label
P1	blue	not assigned	sky
P2	tile	not assigned	wall
P3	tile	circular portion	roof
P4	trunk	not assigned	trunk (of a tree)
P5	green parts	not assigned	leaf / grass
P6	fixtures	cuboid, unless perspective	door / window / persian
P7	details	cuboid, unless perspective	poster
P8	details	cuboid, unless perspective	mailbox
P9	gutter	cuboid, with contiguous sections	gutter
P10	stone	not assigned	rock
P11	stone	not assigned	bench

Table 3.2. Association between labels and clusters, completing Table 3.1.

Patterns compose elements (*tree*, *building*, *bench*), just as sentences are made by words.

tree = n x trunk + n x leaf, con $n \geq 0$

window = n x glass + n x persian, con $n \geq 0$

Figure 3.17. Representation of composition of elements *Tree* and *Window*.

The bench, instead, has a pattern with a hole in it; therefore it "includes" a portion of pattern that belongs to the item that's behind the bench, actually. In such a case, the expert system shall recognize a cluster of similar and contiguous patterns and be able to associate them without connecting them to foreign (but neighbouring) textures. The aim is to correctly interpret the image, by excluding from the *bench* pattern the texture that's inside the boundaries but is nevertheless foreign.
Please also note that this "other texture" will be different on every occurrence.

Figure 3.18. Occurrence of *bench*.

Therefore, the elements database shall contain all objects E_n.

key	composition	label
E1	trunk, leaf	tree
E2	glass, persian	window
E3	wall, tile	chimney

Table 3.3. Elements database, for Figure 3.1.

Finally, some super-elements are composed by both patterns and elements: their complexity is merely conceptual, but they will not overload the software. Each object shall be added to the elements database, in order not to add computational issues to automatic elaboration.

$$\text{building} = n \ x \ \text{wall} + n \ x \ \text{window} + n \ x \ \text{door}, \qquad \text{con } n \geq 0$$

Figure 3.19. Representation of composition of super-element *building*.

key	composition	label
SE1	wall, window, door	building

Table 3.4. Super-elements database: for example, the addendum to Table 3.3.

Therefore, the elements database shall include all objects E_n plus the super-elements, composed by both patterns and elements.

key	composition	label
E1	trunk, leaf	tree
E2	glass, persian	window
E3	wall, tile	chimney
SE1	wall, window, door	building

Table 3.5. Complete elements database, for Figure 3.1.

3.1.6. Syntactic method.

At this time, it's useful to get back to the similitude between grammar and lexicon, in other words between syntax and taxonomy:

- points, lines and surfaces directly correspond to the alphabet:
  - points are like letters,
  - while lines (from straight ones to conic ones) and surfaces (from triangles to quadrilaterals) are often many, thus corresponding to names, in the lexicon;
- the section (for lines) and the colour (for surfaces) correspond to adjectives in the lexicon;
- the relationships, both topological or geometrical, are verbs in the lexicon;
- any eventual gradation (corresponding to weight) of relationships correspond to adverbs in the lexicon;
- the articles, the prepositions and the conjunctions belong, within literature, to nominal or verbal syntagms (phrases): thus, we expect they'll maintain their identity even through interpretation of maps and images;

The description can produce sentences, then periods, then texts: if principles application was completely developed up to now in the geomatic field, we could completely interpret a text that fully describes an image or a map.
Naturally, as in literature we could find ancient words, or incorrect ones, during the interpretation process of a map or an image we could have voids, errors and outliers.

Therefore, we aim to build an expert system that could correctly code available information and translate them into a suitable convention. In order to extend the same concept to the reconstruction of a 3D space[6], we'll apply an extrusion, proceeding from reading (a map or an image) back to the original scene.

Finally, please note a dynamic model is composed of:

- interpretation of changes between maps or images, that were taken in different moments in time,
- and the relative transcription

The syntactic method for model recognition is applied to several problems, all referring to different grammars. Some researchers are working on characters automatic recognition

[6] In fact, interpretation and transcription of a solid model are quite rare: on the other side, even androids have a sight and this is due to images interpretation, because tools for acquiring ambient information are often small in dimension and fixed per position, therefore preventing 3D data storage.

(even the Chinese ones); some others are focusing on the shape analysis issue (for waves and boundaries), typically in the medical or geological field: they're studying seismograms, electrocardiograms, electroencephalograms, etc.

In the geomatic field, we could successfully use this procedure for the analysis of textures for digital images.

Some results are already available, from past experiences borrowed by automatic soil classification (from remote sensing images); in fact, stochastic grammars create models for soil separation, according to the use they're dedicated.

Another application comes from automatic recognition of roofs on top of buildings[7], within the methods for DTM construction and city models, through the laser scanning techniques. We all know that there are some effective procedures dealing with aerial information for the laser scanning method: unfortunately, they're not applicable to ground measurements. In this case, we expect the parsing technique to prove useful.

At first sight, the issue regarding the syntactic method application to automatic recognition of objects may appear to be seamless, because it requires an interpretation procedure with infinite language, which can suitably fit the chance for distorted variables. Nevertheless, the contextualized problem manages and organizes data collection, easing both the analysis and the interpretation, as it allows to switch to finite models with limited complexity, that finally reduce the number of distorted variables.

Another chance is in using the error correcting parsing methods, which are especially effective in automatic characters recognition and are often used within procedures that switch the minimum distance concept to tree grammars. Algorithms like this are called error correcting tree automata: they refer to the original concept for distance between trees and allow the recognition (for example) of a random version of an alphabet character. This is the reason why the most widespread *CAPTCHA*[8] test had to improve their algorithm in time:

---

[7] For automatic roofs recognition, for example, we could use both stochastic and expanded grammars, to allow classification of roofs with irregular geometry. In fact, context-free grammars are used to represent rectangles, but they can expand to include irregular shapes.

Besides, further knowledge on probability about occurrences of some types of roofs could ease the classification, according to the maximum likelihood criterion or even according to the Bayes' theorem.

[8] English acronym CAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart) is a public Turing test and it's completely automated, in order to distinguish between computers and human beings. The requirement is to write a sequence of letters and numbers, reading them through distortion and obfuscation.

The term was born in 2000 by Luis von Ahn, Manuel Blum and Nicholas J. Hopper (Carnegie Mellon University) and by John Langford (IBM).

- the first production technology has been abandoned, since automatic reading was available: test potential was completely lost.

Figure 3.20. Example for an original CAPTCHA.

- the new CAPTCHAs, rather than creating a distorted background over letters twisting, try to get difficult to find the solution through the segmentation technique, by adding an uncertain line over the letters.

Figure 3.21. Example for a modern CAPTCHA, with line.

- or even, we could worsen the segmentation capability by approaching the symbols one on another, as we could see on the Yahoo pages now.

Figure 3.22. Example for a modern, crowded CAPTHA.

Through all these concepts, we may conclude that use of stochastic languages actually allows control of noisy and distorted models, through availability of probabilistic information or likelihood measurements.

3.2. Image description, with object-based structure.

The image size is 13.5cm x 10cm and shows a slight areal magnification (exactly, 6.25 times), for it has been taken with a non-professional digital camera with a display size 5.4cm x 4cm.

Given the image area equal to 135cm^2, we choose to compose a description with 1.350 words: thus corresponding to 10 words for square centimetre[9]. Of course, the original text was written in Italian language, for our convenience and ease of use: that's reported integrally in the right column of Table 3.5.
Therefore, it's pacific that the description translated in English language (below, in the left column of table) is not exactly 1.350 words long.

Another note: again with reference to the areal scale, every square millimetre on the image corresponds to 0.16 square millimetre on the camera: due to the (non-professional) quality of the digital camera we used, these values are not far from highest definition details on the original image.

The image (13.5cm per 10cm) shows a landscape, without panorama. Background is light blue, tending to azure in the lowest part of image. There's a "basement" from left to right, that raises from 1/20 to 1/10 of image full height: - on the left, till middle of image width, there're two lines, composed with about 20 small grey elements that long in the horizontal direction. They become three lines in the second quarter; - the right part shows, on half, four rectangular elements turning high (inclined 45° on the left and raising a level); - a thin green line covers the left part; - the last quarter being formed by a flat green/brown cupola raising another level and getting down to the image bottom (on the left, a brown rectangular element covers over half of previous quarter).	*L'immagine (13,5 cm per 10 cm) mostra un paesaggio, senza panorama. Lo sfondo è azzurro, tendente al celeste pallido nella parte inferiore.* *Il "basamento", da sinistra a destra, passa da un ventesimo a un decimo dell'altezza dell'immagine:* - *da sinistra, per metà della larghezza dell'immagine, due file, composte da una ventina di piccoli elementi grigio/marroni (allungati nella direzione orizzontale), diventano tre file nel secondo quarto;* - *la parte destra mostra, per metà, quattro elementi rettangolari che arretrano verso l'alto (piegando 45° a sinistra e salendo di un "livello");* - *una striscia sottile verde sormonta la parte sinistra;* - *l'ultimo quarto è formato da un cupola appiattita verde/marrone che s'innalza di un altro "livello" e scende fino al fondo (a sinistra, un elemento rettangolare marrone si estende sopra la metà del quarto precedente).*

[9] On the other hand, another possible choice would apply 100 words for every square centimetre (one for square millimetre), thus corresponding to a punctual indication of color for every square millimetre. This would actually be a (rough) raster description of the whole image.

In the foreground, there are two brown linear elements, irregularly screened (their width is respectively 1/10 and 1/20 of the whole width of the image, tapered upward). The second shows a bipartition on the left, after the first third, and then another on the left after the second third, on the right part:	*In primo piano, stanno due elementi lineari marroni, irregolarmente retinati (larghi rispettivamente un 1/10 e 1/20 della larghezza dell'immagine, rastremati verso l'alto), di cui il secondo bipartito a sinistra, dopo il primo terzo, e ulteriormente bipartito a sinistra, dopo il secondo terzo, nella bipartizione di destra:*
- the first element is located above the "base", after the first third on the left of the image. It raises vertically, up to half, leaning to the left, after 1/4 of its height;	- *il primo elemento si erge, sopra il "basamento", dopo il primo terzo sinistro dell'immagine e sale verticale, fino a metà, piegandosi a sinistra, dopo un quarto dell'altezza;*
- the upper 2/3 of this element are covered by a dark green stain (lighter in its upper right part), composed of small lanceolate parts, partially overlapping (size of a postage stamp).	- *i due terzi superiori di questo elemento sono coperti da una macchia verde scuro (più chiara nella parte in alto a destra), composta da piccole parti lanceolate, parzialmente sovrapposte (aventi le dimensioni di un francobollo).*
The gross form of the spot (though undulating in outline) is:	*La forma grossolana della macchia (comunque ondulante nel contorno) è:*
- an isosceles trapezoid (with an oblique vertical side equal to 6/10 of image height and the two bases 30° descendent on the right over the horizontal axis; one doubles the other and the longest is higher, as long as the oblique side);	- *un trapezio isoscele (con un lato obliquo verticale, pari a 6/10 dell'altezza dell'immagine, e le due basi discendenti verso destra di 30° sull'orizzontale, una doppia dell'altra e con la maggiore in alto, lunga quanto il lato obliquo);*
- surmounted by a scalene trapezium that climbs up to the top (the foundations of which originate from the larger base of the isosceles trapezoid, returning to the left and right about 1/10 of its base and ascending to the right of 60° to the horizontal).	- *sormontato da un trapezio scaleno che sale fino alla sommità (le cui basi originano dalla base maggiore del trapezio isoscele, rientrando a sinistra e a destra circa 1/10 della sua base ed ascendendo verso destra di 60° sull'orizzontale).*
The background is glimpsed on the left, for an extension between 1/10 and 1/20 the width of the image (the shape has two humps, near to the margin, separated by a creek, up to 2/3 of the bush, and thence it has a few undulations):	*Lo sfondo s'intravvede a sinistra, per una estensione tra 1/10 e 1/20 della larghezza dell'immagine (la sagomatura forma due gobbe, verso il margine, separate da un'insenatura, fino a due terzi della macchia, donde prosegue con poche ondulazioni):*
- in the upper part, shattering the scalene trapezoid (where an irregular thin brown element stems to the left, 30° to the horizontal);	- *nella parte più alta, frantumando il trapezio scaleno (dove un elemento sottile irregolare marrone discende verso sinistra, a 30° sull'orizzontale);*
- centrally, between bipartitions of the second element;	- *al centro, tra le bipartizioni del secondo elemento;*

- on the left and on the right of the second element, after the first third.
The second element is born in the bottom right corner (covering the "base") and it rises almost vertically, leaving a space at the top right, equal to 1/15 of the width of the image. The two bipartitions form a 45° angle to the horizontal, straightening before reaching the summit (both at half the height):
- the last third is covered by a dark green stain (lighter in its upper part and on the right), composed of small lanceolate parts, partially overlapping (size of a postage stamp), that reveal this element, as well as the background in the central part;
- to the right of the first bifurcation, and to the left of the second, respectively protrude two and three shaves spots of the same nature (equally spaced).
The gross form of the spot (though undulating in outline) is an angled trapezium (with the larger base on the top, length 1/3 of the width of the image, and the oblique side descending towards the right with a 60° angle of to the horizontal, up to 1/4 of the height of the image).
In the background, in the lower half of the image, there is a central body. It leaves, on both right and left, two spaces each equal to 1/6 of the width of the image).
The colors are different:
- milky white, with a rectangular texture (in staggered rows of half the rectangle), is the lower part of the central body. Two parts form:
  - to the left, a pentagon (with two right angles, compared to the "base", with a slight uphill. They also point upwards at the intersection of two sides, at 45° to the horizontal);
  - to the right, an isosceles trapezium with vertical bases (the greater on the left and the minor on right) and the
	- *sulla sinistra e sulla destra del secondo elemento, dopo il primo terzo.*
	Il secondo elemento si erge dall'angolo in basso a destra (coprendo il "basamento") e sale quasi verticale, lasciando alla sommità uno spazio a destra, pari a 1/15 della larghezza dell'immagine. Le due bipartizioni formano un angolo di 45° sull'orizzontale, raddrizzandosi prima di raggiungere la sommità (entrambe a metà dell'altezza della seconda):
	- *l'ultimo terzo è coperto da una macchia verde scuro (più chiara verso l'alto e da destra verso sinistra) composta da piccole parti lanceolate, parzialmente sovrapposte (aventi le dimensioni di un francobollo) che lasciano intravedere questo elemento, oltre lo sfondo nella parte centrale;*
	- *a destra della prima bipartizione, e a sinistra della seconda, fuoriescono rispettivamente due e tre macchie rade della stessa natura (ugualmente distanziate tra loro).*
	La forma grossolana della macchia (comunque ondulante nel contorno) è un trapezio rettangolo (con la base maggiore sulla sommità, lunga 1/3 della larghezza dell'immagine, e il lato obliquo discendente verso destra con un angolo di 60° sull'orizzontale, fino a 1/4 dell'altezza dell'immagine).
	In secondo piano, sta un corpo centrale fino a metà inferiore dell'immagine (lasciando, a destra e sinistra, spazi pari ciascuno a 1/6 della larghezza dell'immagine).
	
	I colori sono vari:
	- *bianco latte, con una tessitura rettangolare (a righe sfalsate di metà rettangolo), è la parte bassa del corpo centrale. Due parti formano:*
	  - *a sinistra, un pentagono (con due angoli retti, rispetto al "basamento", se non per la leggera salita, e la punta verso l'alto all'incrocio di due lati a 45° sull'orizzontale);*
	  - *a destra, un trapezio isoscele con le basi verticali (la maggiore a sinistra e la minore a destra) e il lato obliquo inferiore nascosto*

lower oblique side hidden by the flattened dome and by the rectangular brown element.

- there are three elements grey, spotted with milky white:
  - two respectively extend the pentagon on the right and the isosceles trapezoid on the left;
  - the third stops in the centre of the isosceles trapezoid.
- there are the two elements in an homogeneous grey, which extend upwards, the isosceles trapeze and the pentagon. All elements have a parallelogram shape (that of the trapezoid is slightly tapered, in the same taper direction). The area ratio is:
  - 1:1 milky white and grey spotted;
  - as between milky white and grey uniform, for the isosceles trapezoid;
  - 2:1 between milky white and grey uniform, on the pentagon.
- brown, shiny and wavy in the highest part, is a profile that rolls along the upper side of the uniform grey, above the isosceles trapezoid and the pentagon, up to the first linear element (in the foreground);
- six small elements (equally spaced) are dark green:
  - three are inside the pentagon;
  - and three are inside of the isosceles trapezoid.

In both cases:

- the first on the left are both rectangular and elongated;
- the other rectangular, inside the pentagon (they're connected, from left to right, by two rectangular elements in milky white and uniform grey, decreased in height by two thin grey horizontal elements, underlying and overlying);

dalla cupola appiattita e dall'elemento rettangolare marrone.

- *grigio maculato bianco latte sono tre elementi:*
  - *due estendono rispettivamente il pentagono a destra e il trapezio isoscele a sinistra;*
  - *il terzo interrompe il trapezio isoscele, al centro.*
- *grigio omogeneo sono due elementi che estendono, verso l'alto, il trapezio isoscele e il pentagono. Tutti gli elementi hanno una forma a parallelogramma (leggermente rastremato quello del trapezio, nella stessa direzione di rastremazione). Il rapporto d'area è:*
  - *1:1 tra bianco latte e grigio maculato;*
  - *come tra bianco latte e grigio omogeneo, per il trapezio isoscele;*
  - *ma 2:1 tra bianco latte e grigio omogeneo, per il pentagono.*
- *marrone, brillante e ondulato nella parte più alta, è un profilo corrente lungo il lato superiore del grigio omogeneo, soprastante il trapezio isoscele e il pentagono, fino al primo elemento lineare (in primo piano);*
- *verde scuro sono sei piccoli elementi (ugualmente distanziati tra loro):*
  - *tre all'interno del pentagono;*
  - *tre all'interno del trapezio isoscele.*

In entrambi i casi:

- *i primi a sinistra rettangolari allungati;*
- *gli altri rettangolari, all'interno del pentagono (congiunti, da sinistra verso destra, da due elementi rettangolari bianco latte e grigio omogeneo, diminuiti in altezza da due elementi sottili orizzontali, sottostanti e soprastanti, grigi);*

- but they're square, inside the isosceles trapezoid.	- *ma quadrati, all'interno del trapezio isoscele.*
Overlying the brown profile, there are three vertical elements:	*Soprastanti il profilo marrone, ci sono tre elementi verticali:*
- the first block is over the elongated rectangular dark green element, on the isosceles trapezoid;	- *il primo isolato è sopra l'elemento rettangolare allungato verde scuro del trapezio isoscele;*
- the second two, partially overlapping, are above the square dark green piece on the left (of isosceles trapezoid).	- *i secondi due, parzialmente sovrapposti, sono sopra l'elemento quadrato di sinistra verde scuro (del trapezio isoscele).*
They all have a height equal to the homogeneous grey below and are topped with a thin rectangular profile, shiny brown, with slightly sloping gable.	*Tutti hanno un'altezza uguale al grigio omogeneo sottostante e sono sormontati da un sottile profilo rettangolare, marrone brillante, a due falde poco spioventi.*
A third level consists of two grey/green elements, respectively in the lower left corner (above the "base") and just behind the second linear element (in the foreground):	*Un terzo piano è costituito da due elementi grigio/verdi, siti rispettivamente all'angolo in basso a sinistra (ma sopra il "basamento") e proprio dietro il secondo elemento lineare (in primo piano):*
- infrequent and tangled, the part on the left rises up to 1/3 of the image. The approximate form (though undulating in outline) is a kite with one side along the left edge, the next side horizontal and the major diagonal upward to the right by 60° to the horizontal, intersecting with the minor one in a ratio of 3:1);	- *rado e aggrovigliato, quello di sinistra s'innalza fino a 1/3 dell'immagine. La forma approssimata (comunque ondulante nel contorno) è un aquilone con un lato lungo il margine sinistro, il lato successivo orizzontale e la diagonale maggiore ascendente verso destra di 60° sull'orizzontale che si interseca con la minore in un rapporto di 3:1);*
- the right part is a spot (which underlies a vertical bright brown element). It comes 1/12 more than half of the image. The approximate form (however undulating in outline) is a rectangular trapeze with horizontal bases, where the angle between the minor base and the oblique side is on the extension of the oblique side of the rectangle trapezoid belonging to the stain overlying the second linear element (in the foreground).	- *a macchia quello di destra (cui sottostà un elemento verticale marrone brillante). Arriva 1/12 oltre la metà dell'immagine. La forma approssimata (comunque ondulante nel contorno) è un trapezio rettangolo con le basi orizzontali, dove l'angolo tra la base minore e il lato obliquo è sul prolungamento del lato obliquo del trapezio rettangolo appartenente alla macchia sovrastante il secondo elemento lineare (in primo piano).*
Other details:	*Alcuni piccoli dettagli:*
- in the pentagon, to the left and above the dark green elongated rectangular element, there are two grey square elements. Around them, there is a thin element (milk white), consisting of two vertical elements,	- *nel pentagono, a sinistra e sopra l'elemento rettangolare allungato verde scuro, sono posti due elementi quadrati grigi. Attorno ad essi, è posto un elemento sottile (bianco latte), composto di due elementi verticali, congiunti da*

joined by a semi-circular element;	*un elemento semicircolare;*
✚ between the pentagon and the isosceles trapezoid, starting from the base to the brown profile, a thin brown linear element stands vertical for the height of the milky white and inclined at 45° to the height of the homogeneous grey (connected, before and after, with short curved sections);	✚ *tra il pentagono e il trapezio isoscele, a partire dal basamento fino al profilo marrone, si erge un sottile elemento lineare marrone, verticale per l'altezza del bianco latte e inclinato a 45° per l'altezza del grigio omogeneo (raccordato, prima e dopo, con brevi tratti curvilinei);*
✚ in the isosceles trapezoid, over the elongated rectangular dark green element, there is a thin horizontal element with a full semi-circular element above. Both are milk white. To the right of these elements, there is a linear brown element;	✚ *nel trapezio isoscele, sopra l'elemento rettangolare allungato verde scuro, è posto un elemento sottile orizzontale e sopra questo un elemento semicircolare pieno (entrambi bianco latte). A destra di questi elementi, è posto un elemento lineare marrone;*
✚ in the top of each of the three vertical elements (in the central body) there is a black square element;	✚ *nella sommità di ciascuno dei tre elementi verticali (del corpo centrale) è posto un elemento quadrato nero;*
✚ in the first half of the first bifurcation of the second linear element (in the foreground), a linear black element is superimposed (partially, it overlaps also the vertical brown element in the third level), with an orange tip at the top left;	✚ *nella prima metà della prima bipartizione del secondo elemento lineare (in primo piano) è sovrapposto un elemento lineare nero (in parte, sovrapposto anche all'elemento verticale marrone in terzo piano), con una punta arancione in alto a sinistra;*
✚ to the left of the same linear element, above the first bifurcation, there is a circular yellow element, with a black dot at the centre;	✚ *a sinistra dello stesso elemento lineare, sopra la prima bipartizione, è posto un elemento circolare giallo, con un punto nero al centro;*
✚ in the centre of the flattened dome and of the brown rectangular element, there are two rectangular elements, respectively, yellow/green and milky white.	✚ *nel centro della cupola appiattita e dell'elemento rettangolare marrone sono posti due elementi rettangolari, rispettivamente giallo/verde e bianco latte.*

Table 3.6. Image description, with object-based structure, for picture in Figure 3.1.

3.3. Evaluation of image reconstruction.

Through the conceptual analysis we discussed, it's possible to recompose all the information in a single elaboration: we will build a new assembly image, with systematic procedures representation. Naturally, there's no full recognisability of original image, due to further elaborations that inevitably cause loss of information or noise in the reconstruction.

Thus, we're reporting an image that directly descends from the original picture through logical passages, but recognition isn't simple as both description and representation are very different from "frame zero". Every detail had its own analysis and got a classification to a specific pattern (identified as the most likely), unfortunately however it figures slightly different from the original one, due to uncertainty and approximation given by the software used for automatic recognition.

An architect helped us recovering the whole image, critically reading the description and trying to reproduce the original picture: he worked as our automatic expert system, the software which would have been used instead. He delivered two different elaborations, with two outputs at the end:

- he produced a first image, which was the result of a preliminary and unsupervised analysis, working on his own and with no access to who produced the description. This procedure compelled him to interpret the description trying to avoid misleading and false interpretations. Results are shown in the following Figures 3.28, 3.29 and 3.30
- then the second drawings series comes from reading again the description, together with the writer, in order to avoid misleading and mistakes in input information. Naturally, there was no discussion, not to add any data to contents of text: however, this procedure is normally classified as a supervised operation. Double-check allows the definition of punctual details, that eventually where not (enough) clear through the description.

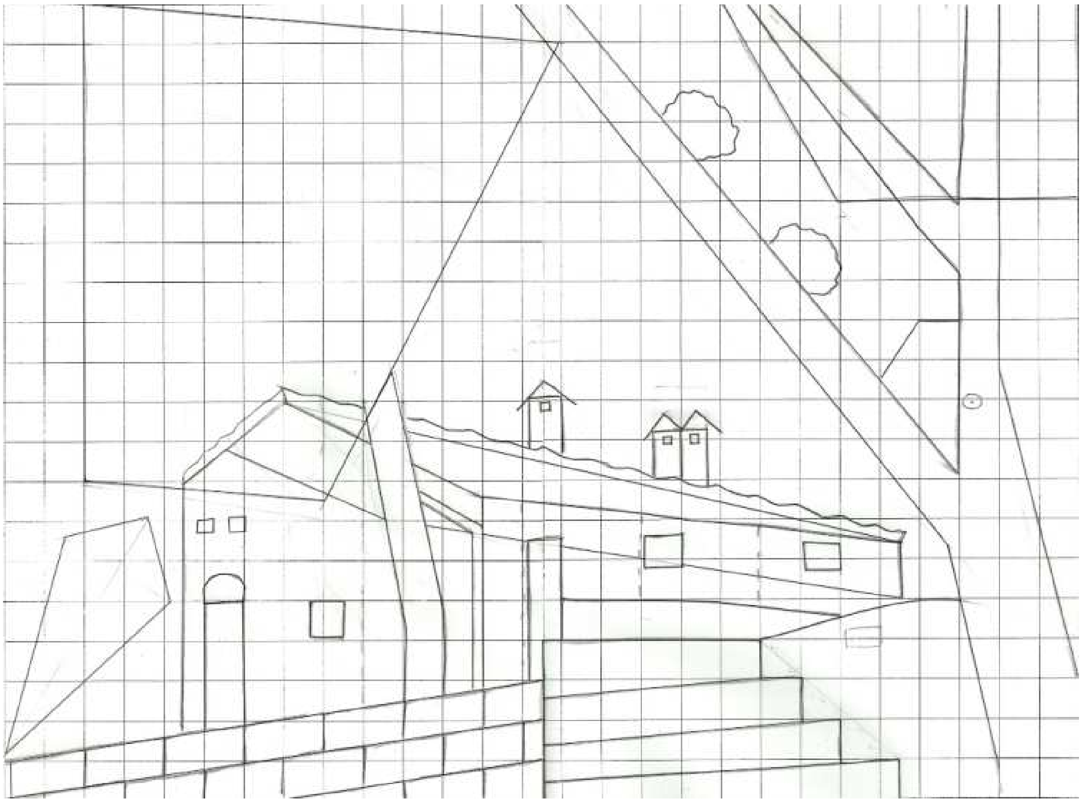

Figure 3.23. Preliminary sketch (unsupervised procedure) of picture in Figure 3.1.

Figure 3.24. Preliminary colourful sketch (unsupervised procedure) of picture in Figure 3.1.

Figure 3.25. Preliminary reconstruction (unsupervised procedure) of picture in Figure 3.1.

At this time, we had the first evaluation of likelihood between the original image (shown in Figure 3.1) and the reconstruction shown in Figure 3.26.

Figure 3.26. Superposition of reconstruction (unsupervised procedure) and original picture.

Please note that pixels were considered wrong whereas neither colour, neither geometry correspond to the original picture: therefore, for every mistake in the reconstruction, the matrix doubles the error. This circumstance allows us to use a half parameter when considering the error percentage.

Finally, we can compare the number of black pixels on total pixels value: the result for the likelihood assessment shown in Figure 3.27 values 58,58%.

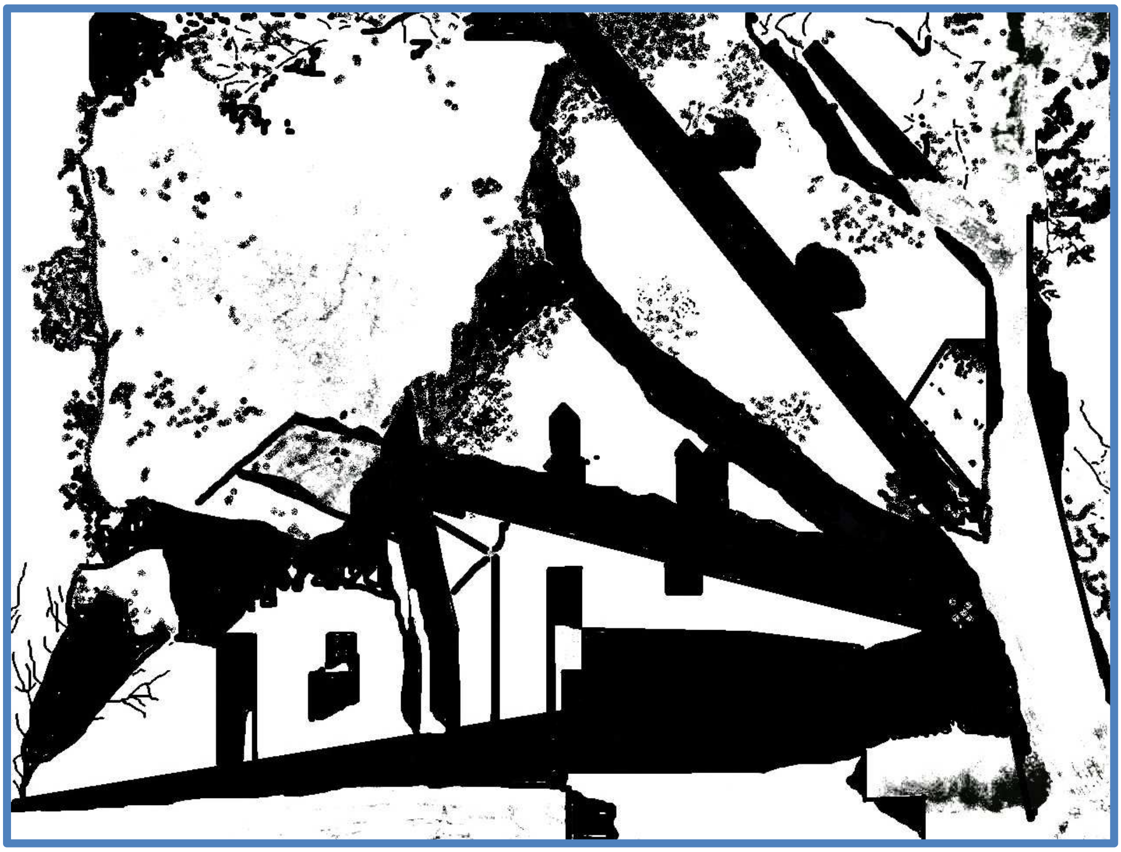

Figure 3.27. Matrix relevant to likelihood assessment between the reconstruction drawing (unsupervised procedure) and the original picture.

We had some misleading details in the description causing false interpretations: even though the description was correct and adequate, not having any wrong or imprecise word, some details had been mistaken.

It was clear that this kind of errors could have been eliminated through another step of the procedure: a critic reading of description would have been an operation type supervised, that would eventually improve the overall quality of reconstruction.

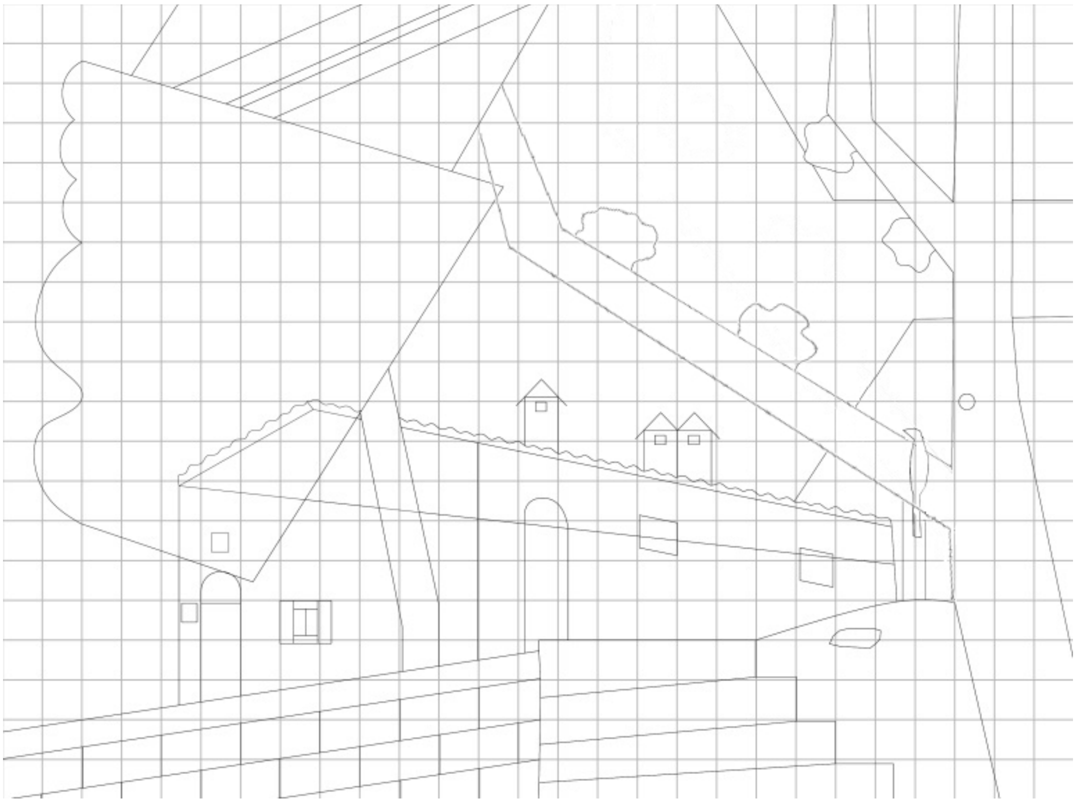

Figure 3.28. Sketch (supervised procedure) of picture in Figure 3.1.

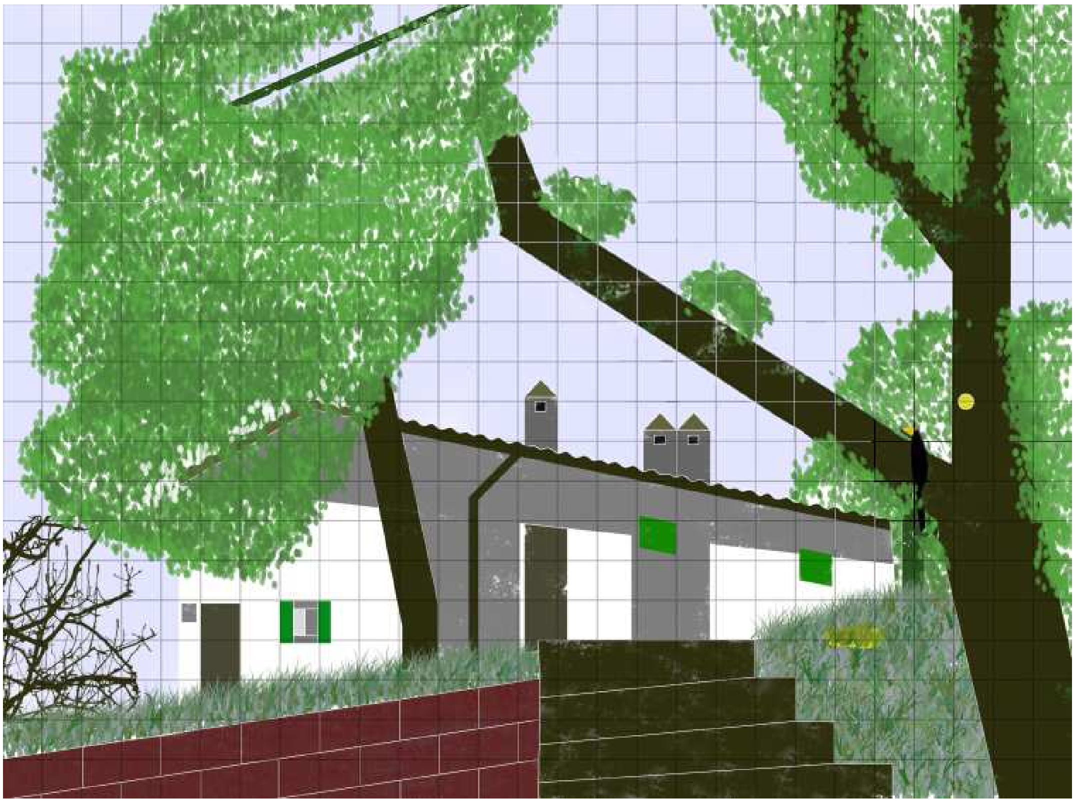

Figure 3.29. Colourful sketch (supervised procedure) of picture in Figure 3.1.

Figure 3.30. Reconstruction (supervised procedure) of picture in Figure 3.1.

Figure 3.31. Final reconstructed image (supervised procedure) for picture in Figure 3.1.

At this time, we performed the second and last evaluation of likelihood between the original image (shown in Figure 3.1) and the reconstructed one, shown in Figure 3.32.

Figure 3.32. Superposition of reconstruction (supervised procedure) and original picture.

Finally, the result for the likelihood assessment shown in Figure 3.33 values 79,20%.

Figure 3.33. Matrix relevant to likelihood assessment between the reconstruction drawing (supervised procedure) and the original picture.

3.4. Reconstructed image analysis.

The more visible difference between the reconstruction and the original image it's in the layout of images through the vertical axis: in particular, the reconstruction overestimates the dimensions of all components shown on the lowest half of original image. In fact, the textual description is the root cause of this overestimation, since it doesn't explicitly refer to heights of all visible artefacts (building and basement).

Due to this anomaly, we decided to perform another likelihood assessment between the original image and a new one, built with different parameters on all components shown on the reconstruction. This trick allows balancing the effect of overestimation:

- we'll consider 60% for the upper half of reconstructed image, shown in Figure 3.35
- and 40% for the lowest half

Figure 3.34. Reconstruction of image, proportioned 60/40.

Figure 3.35. Superposition of reconstruction (supervised procedure) and original picture.

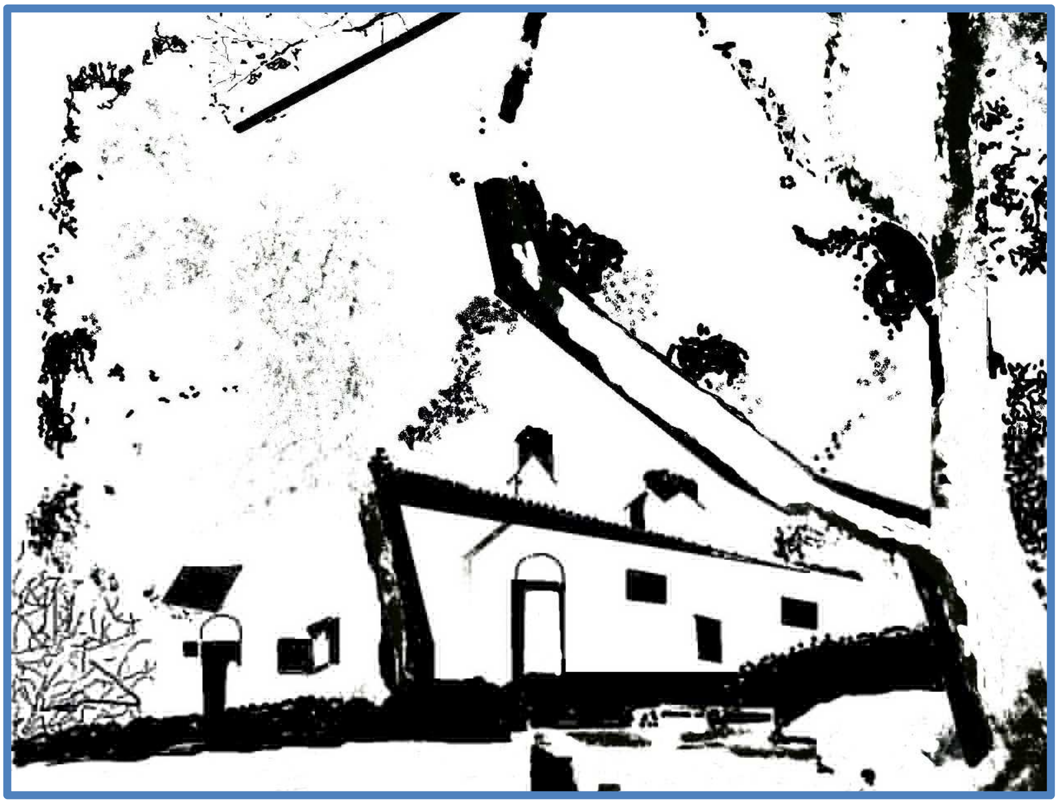

Figure 3.36. Matrix relevant to likelihood assessment between the proportioned drawing (supervised procedure) and the original picture.

Finally, the result for the likelihood assessment shown in Figure 3.36 values 85,92%.

Another note is relevant to the arbitrary limit assigned to the number of words to be used in the description: for sure, a low limit shows some problems related to scarce availability of contents. Let us recall that choosing to use 1.350 words corresponds to spend (averagely) 10 words for each square centimetre, while a description using 100 words for square centimetre would be really close to a raster image.
On the other hand, it's clear that raising the limit to 20 words for square centimetre would of course improve both quality and ease of interpretation on description; nevertheless, please be aware about problems related to costs for longer descriptions and more demanding interpretation skills.

Unfortunately, there are some details in the image that prove to be not enough defined, even over reconstruction and adjustment: maybe a more precise description would get a more punctual interpretation.

In the following paragraphs, we're numbering four examples, to allow further development and discussion. Naturally, that is not the full list of errors occurring in the reconstruction: there are some others details that could be studied in order to improve the overall image quality (for instance, the tree shape), but they're not in the scope of this work.

3.4.1. Improving opportunity for textual description: the dry bush.

The first example in the list refers to the shape of the dry bush, that’s visible in the low left corner of original image.

Figure 3.37. Dry bush detail on both original and reconstructed image.

Looking at the part of description dealing with this feature, it proves to be poorly detailed.

A third level consists of two grey/green elements, respectively in the lower left corner (above the "base”) […]:

- infrequent and tangled, the part on the left rises up to 1/3 of the image. The approximate form (though undulating in outline) is a kite with one side along the left edge, the next side horizontal and the major diagonal upward to the right by 60° to the horizontal, intersecting with the minor one in a ratio of 3:1) […]

Table 3.7. Dry bush detail on description.

In fact, the reconstruction correctly indicates the dry bush in the right position; nevertheless, dimensions and proportions are both imprecise.
In this case, the error should have been avoided by adding some information in the textual description. In particular, we must add another sentence to the text, as indicated in the following Table 3.8.

Its shape is an isosceles trapezoid (half a square), whose diagonal is lying on the image border.

Table 3.8. Image description, with object-based structure: detail for the dry bush.

3.4.2. Improving opportunity for textual description: the chimneys.

The second example comes from the chimneys position: they're exposed to sunlight in the original image (and they're actually overexposed in the picture).

Figure 3.38. Chimneys detail on both original and reconstructed image.

Looking at the part of description dealing with this feature, it proves to miss a detail.

[…] Overlying the brown profile, there are three vertical elements:

- the first block is over the elongated rectangular dark green element, on the isosceles trapezoid
- the second two, partially overlapping, are above the square dark green piece on the left (of isosceles trapezoid)

They all have a height equal to the homogeneous grey below and are topped with a thin rectangular profile, shiny brown, with slightly sloping gable. […]

Table 3.9. Chimneys detail on description.

In fact, the reconstruction correctly indicates all the three chimneys in the right positions; nevertheless, they're not exposed to sunlight and they're even shown as shady.
In this case too, the root cause of error lies in the textual description. To solve the problem and avoid difficulties of interpretation, we must add the plain indication of colour for these three elements, as indicated in the following Table 3.10.

At the top of the brown profile, there are three white vertical elements.

Table 3.10. Image description, with object-based structure: detail for the three chimneys.

3.4.3. Improving opportunity for textual description: shadows.

The third example refers to shadows on the front side of the building, as they're visible in the original image.

Figure 3.39. Shadows detail on original image.

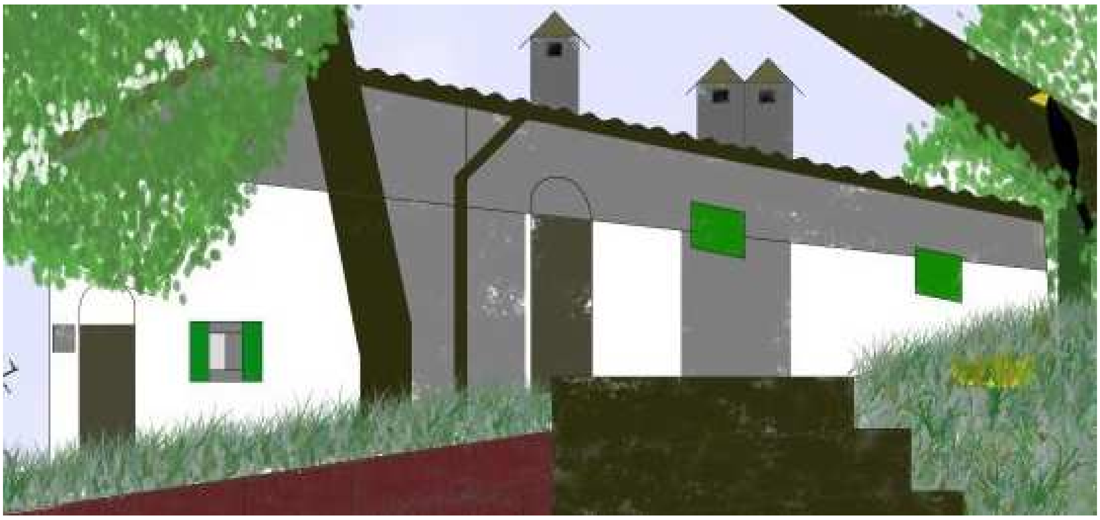

Figure 3.40. Shadows detail on reconstructed image.

Looking at the part of description dealing with this feature, it proves to be poorly detailed.

[…] The gross form of the spot (though undulating in outline) is: - an isosceles trapezoid (with an oblique vertical side equal to 6/10 of image height and the two bases 30° descendent on the right over the horizontal axis; one doubles the other and the longest is higher, as long as the oblique side) - […]

Table 3.11. Shadows detail on description.

In fact, the reconstruction correctly indicates the presence of a shadow on the front side of building; nevertheless, dimensions and shape are both imprecise and they don’t correspond to the ones shown in the original image.

Again, the root cause of error lies in the textual description. To solve the problem and avoid difficulties of interpretation, we must add another sentence, as indicated in the following Table 3.12.

The oblique side shows a light concavity in its inferior third.

Table 3.12. Image description, with object-based structure: detail relevant to the shadows on the building front side.

3.4.4. Improving opportunity for textual description: overlapping planes.

The last example shown refers to the overlapping visible planes on the original image.

Figure 3.41. Trees detail on both original and reconstructed image.

Looking at the part of description dealing with this feature, it seems suitably fitted.

[…] A third level consists of two grey/green elements, respectively […] and just behind the second linear element (in the foreground) […]

Table 3.13. Trees detail on description.

In fact, the reconstruction correctly indicates the presence of two trees; nevertheless, it doesn't show the correct positions nor consider the impossible overlapping of first and third level of original image.

In this case, however, the root cause of interpretation errors isn't actually in the textual description. No other sentence could be added to express additional information: it's apparently a mistake introduced by the "experience" of the system, a kind of licence that the interpreter uses from his own knowledge. Assuming that we found the correct origin of this error, it could be automatically avoided by an actual expert system on a computer platform.

Conclusions

The primary objective of this thesis is to develop strategies of elaboration, in an attempt to reduce both number and weight of computational steps, by increasing the effectiveness of the automation.
We applied some principles of Linguistics, finding new processing strategies for geomatic analysis (study of information contained in maps and images).
We demonstrated that the centring for databases comparison can be successfully improved by using the analysis of the connections and direct comparisons between the structures of the graphs.

The thesis project focuses on a possible new challenge of artificial intelligence, suggesting a theoretical proposal about image matching in the geomatic field: this study deals with the chance to integrate the concepts borrowed by liberal arts (humanistic field) in the context of automated data. The objective of this research is to measure the reliability of the rules of the Universal Grammar by Chomsky in the automatic reading of maps and images, through the syntactic recognition of comparison models among maps, images or 3D models, by using archetypes. Indeed, there is an analogy between the hierarchical structure (tree-shaped) of models and languages syntax: in fact, the combination rules-representations perfectly corresponds to the representation of syntactic structures in mind.

Grammars are defined as combination rules for objects primitives and they directly come from the language which is able and used to describe the model. We studied their expressive capacity through:

- analysis of the number of primitives of the chosen and adopted model. Naturally, a modest amount of atomic elements make easier both computation and management of an automated system;
- and analysis of rules, to be applied iteratively according to re-writing instructions.

Therefore, the best possible syntactic representation establishes a compromise between the number of existing relationships and the endless amount of primitives necessary for the model description, taking into account the technical constraints (the available computational power) and final result of the study.

Applications of the syntactic method for pattern recognition are several and refer to the use of different types of grammars: for example, this technique is used to analyse shapes of waves and boundaries (seismograms, electrocardiograms and electro-encephalograms). In the geomatic field, we could study the textures of digital images, or improve results in soil classification from satellite images: in fact, stochastic grammars can produce models for this kind of elaborations.

With reference to the matching between images and maps, the aim of this thesis is to prove the bijective correspondence between an image and a text relevant to its description, of course considering that quality of correspondence depends on both image definition and text complexity. The text shall perfectly correspond to visible structures in the image, without referring to single objects: the key point is that such a text isn't a map.

This thesis walked through a brief description of the state of the art in image processing, with a short recall of cultural heritage due to Computer Science, Logic and Psychology. Then we analysed the principles of computer vision, while verifying eventual hints for automatic pattern recognition.

The second chapter deals with Linguistics, through the history of the study of language, until the formulation of the Universal Grammar and the theory of formal languages, by Chomsky.

The third chapter finally reports a case study, relevant to my experience in an image representation. I deal with further representations of a single image, taken by a common digital camera, aiming to a textual description that could be eventually traced back to the original scene.
First of all, the picture is not widely recognizable, nor professional, in order to reduce the impact due to previous awareness of symbols and proportions.
Then I came to an image description, with object-based structure, targeting to build a conceptual map that streamlines the further development of (semi) automatic software: this should offer interpretation and classification of representations, as much as possible unambiguous and reliable.
A drawing-skilled guy performed the reconstruction of the image: I evaluated results for his two tentative (unsupervised procedure and supervised one), then a rescaling really improved final likelihood assessment values.

My work ends with an evaluation of the overall reliability of the proposed procedure, while suggesting some hints in order to improve details and correctness of description. The thesis focuses on the cultural reasons of a possible choice, therefore clues are just an indication for other researchers – but of course they shall be taken into due account when an automatic system is under construction.

I'd underline that an eventual refinement of description would improve the quality of reconstruction, by adding words to the text: personally, I expect that a text 25% longer (about

1.500 words) would significantly express the details the first version missed.
Therefore, the first and next development of this thesis is the expansion of the Italian text, including at least the identified clues, in order to check whether the quality of reconstruction will be eventually improved.

A computerized expert system, able to translate images into their textual descriptions, or to reconstruct an image representation starting with its text, is out of the scope of this thesis: the creation of software would have been too demanding for a single person, while it could have been completely useless in case the procedure cannot prove its reliability. Conversely, now that effectiveness of procedure seems to be proved, the automation could be an interesting field to explore, as an outcome of this thesis work.

Finally, there’s an addendum at the end of this book, explaining what kind of place Mount San Giorgio is: it’s a UNESCO World Heritage Site since 2003.

addendum

Mount San Giorgio

The thesis focuses on the original picture taken by Professor L. Mussio on the landscape at the summit of Mount San Giorgio (Canton Ticino, Switzerland), with his own non-professional digital camera. The place is shown in the pictures below (Figures 4.1 and 4.2), in other weather conditions and taken from different points of view.

Figure 4.1. Another view on the landscape at the summit of Mount San Giorgio.

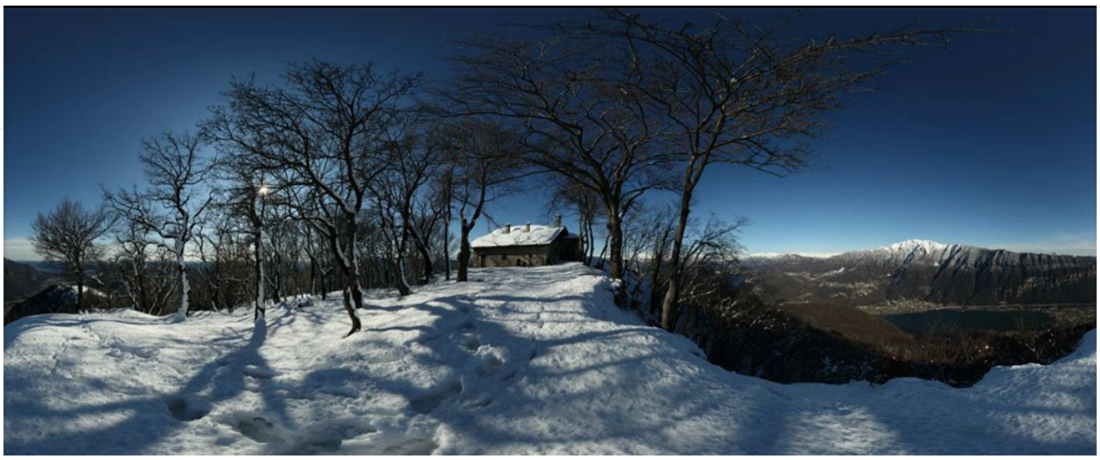

Figure 4.2. A snowy view on the landscape at the summit of Mount San Giorgio.

Monte San Giorgio is a wooded mountain (1.097m above sea level) located at the very South of Switzerland, prospecting the Lugano Lake between the south of canton Ticino in Switzerland and the region of Lombardy in Italy.

Figure 4.3. Location of Mount San Giorgio.

Monte San Giorgio became a UNESCO World Heritage Site in 2003, because it is “the best fossil record of marine life from the Triassic Period (245–230 million years ago). The sequence records life in a tropical lagoon environment, sheltered and partially separated from the open sea by an offshore reef. Diverse marine life flourished within this lagoon, including reptiles, fish, bivalves, ammonites, echinoderms and crustaceans. Because the lagoon was near land, the remains also include land-based fossils of reptiles, insects and plants, resulting in an extremely rich source of fossils."

Later, in 2010, also the Italian side of the mountain was added as an extension to the World Heritage Site.

Figure 4.4. Monte San Giorgio, shown on the left in the background of Lake Lugano.

Essential bibliography references

Abbagnano, N. (1971): *Dizionario di Filosofia*. UTET, Torino.

Aho, A.; Ullman, J. (1972): *The theory of Parsing Translation, and Compiling*. Prentice Hall, Englewood Cliffs.

Akmajian, A.; Demers, R. A.; Farmer, A. K.; Harnish, R.M. (1982): *Linguistica*. Il Mulino, Bologna.

Arnheim R. (1974): *Il pensiero visivo*. Einaudi, Torino.

Bara, B.G. (2003): *Il sogno della permanenza. L'evoluzione della scrittura e del numero*. Bollati Boringhieri, Torino.

Bayes, T. (1764): *Essays towards solving a problem in the doctrine of chances*. Philosophical Transactions of the Royal Society of London.

Bellone, T. (2002): *Note sul riconoscimento sintattico di modelli*. Rivista dell'Agenzia del Territorio, anno II, n. 2.

Bellone, T.; Comoglio, G. (2004): *Le Scienze cognitive e la Geomatica*. Rivista dell'Agenzia del Territorio, anno IV, n.3 anno 2004.

Bloomfield, L. (1962): *The Menomini language*. Yale University Press, New Haven.

Boetti, M. (2009): *Senza più il canone occidentale - La linguistica della vita.* Transfinito.eu

Boole, G. (1847): *The mathematical analysis of logic*. Macmillan, Cambridge.

Boole, G. (1854): *The laws of thought*. Bell, London.

Calzolari, N.; Lenci, A. (2004): *Linguistica computazionale. Strumenti e risorse per il trattamento automatico della lingua*. Mondo Digitale 2.

Canepari, L. (1979): *Introduzione alla fonetica.* Piccola Biblioteca Einaudi, Torino.

Castelfranchi, Y.; Stock, O. (2000): *Macchine come noi. La scommessa dell'intelligenza artificiale*. Laterza, Bari.

Chomsky, N. (1957). *Syntactic Structures*. Mouton, The Hague.

Chomsky, N. (1965): *Aspects of the theory of syntax*. MIT Press, Cambridge.

Chomsky, N. (1993): *Anno 501, la conquista continua. L'epopea dell'imperialismo dal genocidio coloniale ai giorni nostri.* Gamberetti Editore, Roma.

Chomsky, N. (2003): *Due ore di lucidità*. Baldini&Castoldi, Milano.

Chomsky, N.; Foucault, M. (2005): *Della natura umana. Invariante biologico e potere politico*. Derive Approdi, Roma.

Cohen, P.; Feigenbaum, E. (1981): *The handbook of artificial intelligence*. Kaufmann, Los Altos.

Cook, V. J. (1990): *La Grammatica Universale. Introduzione a Chomsky.* Il Mulino, Bologna.

de Saussure F. (1912): *Cours de linguistique générale*. Paris.

de Saussure, F. (1916): *Écrits de linguistique générale*. Paris.

Diamond, J. (1998): *Armi, acciaio e malattie. Breve storia del mondo negli ultimi tredicimila anni.* Einaudi, Torino.

Diamond, J. (2005): *Collasso. Come le società scelgono di vivere o morire*. Einaudi, Torino.

Fondelli, M. (1992): *Trattato di fotogrammetria urbana e architettonica*. Laterza, Bari.

Frege, G. (1973): *La struttura logica del linguaggio*. Bompiani, Milano.

Fu, K.S. (1974): *Syntactic methods in pattern recognition*. Academic Press, New York.

Fu, K. S. (1982): *Syntactic pattern recognition and applications*. Prentice Hall, Englewood Cliffs.

Giacalone Ramat, A.; Ramat, P. (1993): *Le lingue indoeuropee*. Il Mulino, Bologna.

Godart, L. (1992): *L'invenzione della scrittura. Dal Nilo alla Grecia*. Einaudi, Torino.

Hofstadter, D. R. (1984): *Gödel, Escher, Bach: un'eterna ghirlanda brillante*. Adelphi, Milano.

Hofstadter, D. R. (1996): *Concetti fluidi ed analogie creative*. Adelphi, Milano.

Lepschy, G. C. (1992): *La linguistica del Novecento*. Il Mulino, Bologna.

Lolli, G. (1992): *Incompletezza: saggio su Kurt Gödel*. Il Mulino, Bologna.

Marconi, D. (2001): *Filosofia e scienza cognitiva*. Laterza, Bari.

Markov, A. (1913): *An example of statistical investigation in the text of Eugene Onegin*. Proceedings of the Academy of Sciences of St. Petersburg.

Minsky, M. (1990): *La società della mente*. Adelphi, Milano.

Moro, A. (2006): *I confini di Babele*. Longanesi&C., Milano.

Quine, W. (1981):*Logica e grammatica*. Il Saggiatore, Milano.

Reed, S. K. (1989): *Psicologia cognitiva*. Il Mulino, Bologna.

Robins, R. H. (1971): *Storia della linguistica*. Il Mulino, Bologna.

Sansò, F. (1979): *Il trattamento statistico delle misure*. CLUP, Milano.

Sester, M. (1992): *Automatic model acquisition by learning*. IAPRS Vol. XXIX, Part B3.

Sprague, R.H.; Watson, H.J. (1996): *Decision support for management*. Prentice Hall, Upper Saddle River.

Tamburini, G. (2002): *I matematici e le macchine intelligenti*. Paravia Bruno Mondadori, Milano.

Togliatti, G. (1976): *Fondamenti di statistica*. Hoepli, Milano.

Turban, E. (1995): *Decision support and expert systems*. Prentice Hall, Upper Saddle River.

Von Humboldt, W. (1835): *La diversità delle lingue*. Laterza, Roma.

Wall, R. (1972): *Introduction to mathematical linguistics*. Prentice Hall, Englewood Cliffs.

Wittgenstein, L. (1939): *Lectures on the foundations of mathematics*. Cambridge.

Zamponi, A. (1976): *L'etimologia*. Zanichelli, Bologna.