# Elements of a Semantic Code

**Bernd-Olaf Küppers**

**Abstract** An important step towards a scientific understanding of complex systems will be the objectification and quantification of information that carries semantics, i.e., sense and meaning. In this paper a concept is developed according to which the general aspects of semantics—such as novelty, pragmatic relevance, selectivity, complexity and others—are understood as elements of a "semantic code". However, in contrast to its traditional usage, the term "code" does not refer to a set of rules for assignment or translation of symbols, but rather to a reservoir of value elements, from which the recipient configures a value scale for the evaluation of semantic information. A quantitative measure for the value scale is proposed.

## 1 Three Dimensions of Information

Information is based upon signs and symbols and sequences of these. Such sequences may be rich in content, that is, they can have a sense or a meaning.[1] The content, in turn, may become operative in various ways, because of the different kinds of reactions that it can induce in the recipient. In general, three dimensions of information can be distinguished. The "syntactic" dimension is understood as the ordered arrangement of symbols and the relationships between them. The "semantic" dimension includes the relationships between the symbols and also that for which they stand. Finally, the "pragmatic" dimension includes the relationships

[1] Here the two terms *sense* and *meaning* are used synonymously, and the distinction between them that is usual in linguistic philosophy is not retained.

B.-O. Küppers
Frege Centre for Structural Sciences, Friedrich Schiller University Jena, Jena, Germany
e-mail: bernd.kueppers@uni-jena.de

B.-O. Küppers et al. (eds.), *Evolution of Semantic Systems*,
DOI 10.1007/978-3-642-34997-3_4, © Springer-Verlag Berlin Heidelberg 2013

between the symbols, that for which they stand, and also the effect that they have upon the recipient.

However, resolving the concept of information into a syntactic, semantic and pragmatic dimension is only justified within the framework of a systematic analysis. Strictly speaking, the three dimensions make up an indissoluble unity, as one cannot allude to one of them on its own without automatically invoking the other two.

Syntax only takes on a meaning when both the symbols as such and their interrelationships have been stipulated. Semantics, in turn, presuppose a syntactic structure as carrier of the semantics, even though semantics go further than this. The same applies to pragmatics, since the sense and the meaning of information do not open up without reference of the symbols to the real world. Sense and meaning of information can only be objectified through the pragmatic effects that they have upon the recipient and which trigger his interactions with the surrounding world.

Because of the intimate intermeshing of semantics and pragmatics, they are frequently considered together as the so-called "semantico-pragmatic" aspect of information. However, the pragmatic aspect is accompanied by further facets of semantic information, expressed in its novelty, complexity, selectivity and other features. Nevertheless, it is purely a matter of definition whether one considers these as independent properties of semantic information or perhaps as various aspects of the pragmatic component that endow the latter with a fine structure. In any case the semantic aspect of information is something like a "hinge" between its syntactic and its pragmatic level.

The syntactic and the pragmatic aspects are ultimately equivalent to the structural and functional aspects of information. Even though function necessarily presupposes a structure as its carrier, this does not mean that functions can be traced back to their structures. Human language may provide a good example of this: here, the meaning and thus the pragmatic—i.e., functional—relevance of words and sentences are fixed through their syntax, but nonetheless cannot be derived from it. Rather, the semantics appear to stand in a "contingent" relationship to the syntax. In other words, the semantics of a syntactic structure are not necessarily the way they are. They are governed neither by chance nor by law.

Moreover, in human communication the exchange of semantic information is always an expression of human mental activities with all their individual and unique features. Can these be blanked out, allowing us to approach the general characteristics of semantic information? Above all: is there a semantic code that embodies the principles according to which meaningful information is constituted?

Questions of this kind appear to go well, and irrevocably, beyond the scope of the exact sciences, which are based upon objectification, idealisation, abstraction and generalisation. It is therefore understandable that the founding fathers of communication theory—Hartley, Shannon and others—were seeking a scientific approach to the concept of information that is independent of the *meaning* of information but that provides a measure of the *amount* of information.

In fact, communication technology is not concerned with the actual content of information; rather, it sees its primary task in passing the information in a reliable way from the sender to the receiver. The path of this passage, the so-called information channel, has to function such as to be nearly error-free as possible and to avoid corrupting the sequence of symbols on their way to the recipient. This presents the engineer with a technical challenge that is completely separate from the sense and the meaning the transmitted information may have for the recipient.

While it is true that communications engineers definitely speak of the "content" of information, they are referring not to its meaning, but rather to the sheer number of symbols that have to be transmitted. Thus, Shannon and Weaver state explicitly that: "In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent (...) as regards information" [19].

According to our intuitive understanding, information has the task of eliminating uncertainty. This thought also underlies Shannon's definition of information [17]. According to it, the probability $p_k$ with which the arrival of a particular message $x_k$ is to be expected out of a set of possible messages may be considered as a simple measure of its information content. On the basis of this idea the information content of a message $x_k$ is arranged inversely proportional to its expectation value $p_k$, or

$$I_k \sim 1/p_k. \tag{1}$$

If one demands from the information metric that, as a measure of a quantity, it should be additive, then one has to replace the above definition with

$$I_k = \mathrm{ld}(1/p_k) = -\mathrm{ld}\, p_k, \tag{2}$$

where we have introduced the binary (or dyadic) logarithm ld. This definition thus also takes account of the conventional use of binary coding in communications technology. If, in the simple case, a source of messages consists of $N$ messages, which all have the same expectation value for the recipient, then the recipient needs

$$I = \mathrm{ld}(N) \tag{3}$$

binary decisions to select a given message. For this reason the quantity $I$ is also termed the "decision content" of the message in question. In fact, the decision content had been proposed by Hartley [7] as a measure of the quantity of information some years before Shannon formulated the general expression above. Even though definition (2) does not consider the meaning of information, it points at least indirectly to a semantic aspect, namely that of the novelty of an item of information, insofar it is correlated with its inverse expectation probability.

The prior probabilities $p_k$ are fixed by the recipient's prior knowledge. They are, however, objectively determined quantities, i.e., they are the same for all recipients that share the same facilities for, and methods of, obtaining knowledge. One can also say that the information measure developed by Shannon is "objectively subject-related" [20]. In other words: information is not an absolute entity, but only a relative one—a fact which is also termed the "context-dependence" of information.

Shannon's measure of information refers to a message source $X$ of which the elements $\{x_1, \ldots, x_N\}$ make up a probability distribution $P = \{p_1, \ldots, p_N\}$. A characteristic feature of $X$ is its statistical weight

$$H = \sum_k p_k I_k = -\sum_k p_k \operatorname{ld} p_k, \tag{4}$$

which corresponds to the mean information content of a message source. $H$ is also termed the "entropy" of the source, because it has the same mathematical structure as the entropy function of statistical physics [see below, (11)].

An increase in information always takes place when through the arrival of a message the original distribution of expectation probabilities

$$P = \{p_1, \ldots, p_N\} \quad \text{with} \quad \sum_k p_k = 1 \quad \text{and} \quad p_k > 0 \tag{5}$$

becomes modified and converted into a new distribution

$$Q = \{q_1, \ldots, q_N\} \quad \text{with} \quad \sum_k q_k = 1 \quad \text{and} \quad q_k \geq 0. \tag{6}$$

If the gain of information is represented by the difference

$$H(P) - H(Q) = \sum_k p_k I_k - \sum_k q_k I_k \tag{7}$$

then this may lead to negative values, which would contradict our everyday picture of a gain, which we associate with a real increase in information. However, as Rényi [14] has shown, this deficiency can be resolved when one calculates the gain not by considering the difference between two mean values, but rather by first determining the gain in information for each single message and then calculating the mean value

$$H(Q|P) = \sum_k q_k [I_k(p_k) - I_k(q_k)] = \sum_k q_k \operatorname{ld}(q_k/p_k). \tag{8}$$

The Rényi entropy always fulfils the relation

$$H(Q|P) \geq 0. \tag{9}$$

Classical information theory, as developed on the basis of Hartley's and Shannon's work, always refers to a source of messages, which in turn is characterised by a probability distribution. Thus, the question automatically follows whether the information content of a given message can also be quantified without reference to its source. The basis for the measure of information would then be not a probability distribution, but rather a property of the message itself.

A measure of this kind has been proposed independently by Kolmogorov [8], Solomonoff [18] and Chaitin [4]. This approach has become known as algorithmic theory of complexity, since in it the concept of information is intimately linked to that of complexity. According to this idea, a sequence of binary digits is to be considered complex when the sequence cannot be compressed significantly, i.e., when there is no algorithm that is shorter than the sequence itself and from which the sequence can be derived. To express it in the terminology of computer science: the complexity $K$ of a binary sequence $S$ is given by the length $L$ of the shortest program $p$ of a computer $C$ from which $S$ can be generated:

$$K_C(S) = \min_{C(p)=S} L(p). \tag{10}$$

Thus, within this concept, the information content is defined through the complexity of a sequence of symbols. It provides, as it were, a measure of the "bulk" of the information contained in a message.

Like Shannon's definition of information, algorithmic information takes no account of the actual content of information. Rather, the algorithmic measure depends solely upon the complexity of an item of information, which in this case means the aperiodicity of its syntax. From a technological point of view, those restrictions are perfectly justified. However, in the natural sciences such a concept of information soon reaches the limits of its usefulness. This applies above all in the study of biological systems. This is because the results of modern biology, especially in the area of genomic analysis, show clearly that a profound understanding of biological phenomena requires access to the aspect of meaningful information. This applies not least to the problem of the origin of life itself, which is equivalent to that of the origin and evolution of information that has a meaning for the inception and sustenance of vital functions [9].

Insofar as syntax, semantics and pragmatics are intimately interrelated (albeit in ascending order of complexity), we are confronted in a fundamental way with the question of whether the semantico-pragmatic level of information is accessible to an exact, scientific analysis. Here—as we have already pointed out—we encounter major difficulties, because according to our traditional understanding of science the level of sense and meaning evades the kind of objectification and quantification that characterises the exact sciences.

The nature of these issues can be illustrated by the human language. As syntax, semantics and pragmatics represent increasingly complex dimensions of information, one's first thought might be to try and derive the semantics of a linguistic expression from the lower level of its syntax, i.e., from the sequence of phonemes, letters or other symbols which constitute this expression. However, this thought is short-lived, since there is no necessary connection between the syntax and the semantics of a message that would allow us to move logically from the one to the other. A garbled or incomplete text, for example, cannot in general be reconstructed by analysing the available fragments. This can at best be done against a background of additional information, which can serve as a semantic reference frame for the reconstruction.

In linguistics, the relationship between syntax and semantics is referred to as "arbitrary", because it does not allow any necessary connection to be discerned. According to this view, syntax is indeed a carrier of semantics, but semantics are not caused by the syntax. Rather, the irreducibility of semantics seems to be an example of "supervenience", i.e., the relationship of non-causal dependence between two properties $A$ and $B$.

It seems to be a plausible explanation that the arbitrariness of the relationship between syntax and semantics is a consequence of a convention that is determined by the community using the language in question. Nonetheless, the use of the terms "arbitrary" and "conventional" can be criticised, as they still imply a causation principle, even if this is only in the form of the fixation of the semantics by the decision of the language community. For this reason, it might be better to describe the relationship between syntax and semantics as "contingent", as this term is neutral in respect of causation; it merely expresses the fact that the relationship between syntax and semantics is not of necessity that which one observes it to be.

Syntax and semantics are thus independent dimensions of information insofar as their interrelationship is not law-like, even though they are indissolubly connected to one another. Furthermore, the term "semantics" is strongly associated with numerous connotations such as sense, meaning, usefulness, functionality, value, content and so forth. These in turn leave a relatively large scope for determining the semantic dimension of information.

There are two possible ways of coping with this problem. One is to regard precisely the multiplicity as a characteristic feature of semantics. In that case, one would retain the variety of possible definitions and try to build up a theory of semantics on the most general basis possible. The other is to attempt to narrow down the broad field of possible definitions, in order to sharpen the concept of semantics. In that case one would arrive at a definition that would be precise, but at the cost of considerable one-sidedness.

In the latter respect, the greatest progress has been made by the analytical philosophy of language. This already possesses an elaborate theory of semantics, at the centre of which is the logical construction of human language. For example, within this theory strict distinction is made between concepts such as "sense" and "meaning", which in non-scientific usage are largely synonymous.

According to Frege, the founder of logical semantics, the "meaning" of a symbol or an expression is the object that it denotes, while the "sense" is the manner in which it is presented. Thus, the terms "evening star" and "morning star" have the same meaning, as they refer to the same object, the planet Venus; however, as Frege explains, they have different senses, because Venus is presented in the expression "morning star" differently from the way in which it is presented in the expression "evening star". The fact that the morning star and the evening star are the same object requires no further justification. However, the statement that the "star" we see close to the rising sun is the same "star" that appears in the evening sky after sunset is not an obvious truth. Rather, it is a discovery that we owe to the Babylonian astronomers. Semantic precision of this kind was applied by Frege not only to proper nouns, but also to entire sentences. Ultimately, they led to profound insights into the truth value of linguistic statements.

## 2 The Language of Genes

Even though logical depth is a unique property of human language, we may consider the possibility that language is a general principle of natural organisation that exists independently of human beings. In fact, there is much evidence in support of this hypothesis. Thus, it is generally accepted that various forms of communication exist already in the animal kingdom, as revealed by comparative behavioural research. We also know that even plants use a sophisticated communication system, employing certain scent molecules, to inform each other and in this way to protect against the threat of danger from pests. And it has long been known that bacteria communicate with one another by releasing chemical signals.

However, more surprising is the fact that essential characteristics of human language are reflected even in the structure of the genetic information-carriers. This analogy does not just consist in a mere vague correspondence; rather, it embraces largely identical features that are shared by all living beings.

Let us consider some facts: the carriers of genetic information, the nucleic acids, are built up from four classes of nucleotide, which are arranged in the molecules like the symbols of a language. Moreover, genetic information is organised hierarchically: each group of three nucleotides forms a code-word, which can be compared to a word in human language. The code-words are joined up into functional units, the genes. These correspond to sentences in human language. They in turn are linked up into chromosomes, which are higher-order functional units, comparable to a long text passage. As in a written language, the "genetic text" includes punctuation marks, which label—for example—the beginning and end of a unit to be read. Last but not least, the chemical structure of the nucleic acids even imposes a uniform reading direction.

In addition to the parallels described here, there are further fundamental correspondences between the structure and function of genes and that of human language. These include, especially, the vast aperiodicity of nucleotide sequences and the context-dependence of genetic information. The context in turn is provided by the physical and chemical environment, which confers an unambiguous sense upon the (in itself) plurivalent genetic information [10].

Just as a printing error in a written text can distort the meaning, the replacement of a single nucleotide in the genome can lead to collapse of the functional order and thus to the death and decay of the organism. This shows that genetic information also has a semantic dimension; in other words, it possesses functional significance for the sustenance of life processes. As the dynamics of life are encoded in the genes, it would seem only consistent to speak of the existence of a molecular language.

Naturally, the analogy between the language of genes and that of humans would immediately break down if we were to make the entire richness of human language the measure of our comparison with the genetic language. The language of Nature is more to be understood as an all-pervading natural phenomenon, which has found its most elementary form of expression in the language of the genes and its highest in human language.

The use of terms such as "genetic information" and "the language of genes" is in no way an illegitimate transfer of linguistic concepts to the non-linguistic realm of molecules. On the contrary: the existence of a genetic, molecular language appears to be an indispensable prerequisite for the construction of living systems, as their complex functional organisation arises along the path of material instruction and communication. In fact, the unfolding and expression of the tremendous amount of information stored in the genome takes place stepwise in the form of an exceedingly intricate process of communication between gene and gene product. However, any communication requires certain rules and these can only be understood by an appropriate reference to the model of language.

In view of the broad-ranging parallels between the structures of human language and the language of genes, recent years have seen even the linguistic theory of Chomsky move into the centre of molecular-genetic research [16]. This is associated with the hope that the methods and formalisms developed by Chomsky will also prove suited to the task of elucidating the structures of the molecular language of genetics. How far this hope will carry, we must wait to see. However, in the most general sense it may be expected that the application of linguistic methods to biology will open up completely new theoretical perspectives.

Like the philosophy of language, biology uses a special concept of semantics. This stands here, as we have already stated, for the plan-like and purpose-directed self-construction of living matter which finally arises in its extreme functional order. Thus, biological semantics could just as well be termed "functional semantics". However, functions can as little be inferred from their underlying structures as the semantics of a linguistic expression can be determined from its syntax. Like the semantics of linguistics expressions, the semantics of the genetic information appear to present an irreducible property of living matter. Notwithstanding, any progress that linguistic research achieves concerning the problem of semantics will be an enormous help for a better understanding of the corresponding problem in biology.

This is also demonstrated by the following example. Even though the language of the genes lacks the logical depth of human language, we can draw a surprising parallel between the two concerning the distinction between "sense" and "meaning". Genetic information is passed down from parent to daughter organism primarily as an instruction, a mere "pre-scription", which is stored in the hereditary molecules. This form of existence can be compared to the meaning of a linguistic expression. However, the genetic information does not take effect until it is expressed in the fertilised ovum, and this expression is accompanied by a permanent re-assessment and re-evaluation of the information. In this process, the context-dependence of the genetic information comes into effect—precisely the property that imparts the "sense" to a word. In accordance with Frege, we can say that the terms "instruction" and "information" have the same meaning but different senses.

## 3 The General Structure of Language

If we wish to know what the language of Nature is, then we must first look for the general structure of language. However, we cannot solve this problem by simply making an abstraction based on human language, as this has unique features that are due to the unique position of humans as thinking beings. In human language we can pronounce judgements; formulate truths; and express opinions, convictions, wishes and so forth. Moreover, human language has various manifestations at the level of interpersonal communication: there are picture languages, gesture languages, spoken languages and written languages. Therefore, to uncover the general structure of language, we must develop our concept of language from the bottom up. Setting out from the most general considerations possible, we must endeavour to develop an abstract concept of language that reflects human language and at the same time is free from its complex and specific features.

This constraint immediately highlights an almost trivial aspect of language, one that is basic for any process of communication: successful communication is only possible when the partners in communication use a common pool of symbols. Further to that, sequences of such must also be structured according to rules and principles which are known to both the sender and the recipient. Such syntactic structures then form the framework for a symbol language shared by sender and recipient.

However, the syntactic structure must fulfil yet another condition: the sequences of symbols from which a language is built up must have an aperiodic structure. This is because only aperiodic sequences are able to encode sufficiently complex information. This can also be shown to be the case for human language.

Let us consider the language of symbols in more detail. Its character can be illustrated by reference to the communicative processes that take place in living matter. Here, the language is initially instructive in Nature, that is—setting out from genetic information—it sets up the posts that the innumerable elementary processes have to follow. However, the organism is anything but a piece of mechanical clockwork that follows a rigid set of symbols. It is better viewed as an extremely complex system of internal feedback loops, one that has continually to orientate and adjust its internal processes in order to preserve itself and its capacity for reproduction. To do this, it must continually take up new information from its environment (both inside and outside the organism) and evaluate and process this information.

The interactive procedures of expression and continual re-evaluation of information in living matter go beyond mere instruction, insofar as this level of signal-processing takes place with reference to the entire context of the signals. To take account of the full breadth of this aspect of communication, let us look at human language, which is context-dependent to a high degree. The context-dependence is even reflected in the inner structure of a language. And this is because sounds, words, etc. are always present in a complex network of interrelationships. Moreover, this is precisely what makes them elements of a language in the first place.

The "internal" structure of language has been investigated systematically within the framework of structuralism. According to this view, every language is a unique set of interrelated sounds and words and their meanings, and this cannot be reduced to its individual component parts, because each of these parts only acquires meaning within the context of the overall structure. Only through the overall structure can the elements of the language coalesce into a linguistic system; in this system, the elements are demarcated from one another and there is a set of rules that assign a linguistic value to each element.

In the light of such insights into the structure of human language one could toy with the idea of turning these insights on their head, regarding every ordered network of interrelationships as a linguistic structure. Such a radical reversal of the structural view of language is by no means new; it has long stood at the centre of a powerful current of philosophical thought, which is named "structuralism".

Its programme has been expressed by Deleuze in the following way: "In reality there are no structures outside that which is language—although it may be an esoteric or even a non-verbal language. The subconscious only possesses structure insofar as the sub/unconscious speaks and is language. Objects only possess structure insofar as they are considered to speak in a language that is the language of symptoms. Objects themselves only have structure insofar as they hold silent discourse, which is the language of symbols" [5].

Opinions may differ concerning the depth and breadth of the relevance of structuralism for our understanding of the world. Such a verdict will depend not least upon whether one regards structuralism as merely a methodological, if relatively exact, tool for the humanities and social sciences, or whether one accepts it as the ultimate authority for our understanding of the world. As long as structuralism recognises the limits that are placed upon any exact science, it will not claim to embrace the total wealth of reality. Rather, it will follow its own intention of investigating the structures of reality as such, i.e., independently of the forms in which these structures are manifested in reality.

## 4 The "Language" of Structures

The use of terms such as "information" and "language" in biology is often criticised as a naturalistic fallacy. Both terms, it is claimed, fail to characterise natural objects, being illegitimate projections from the non-scientific area of interpersonal communication and understanding into the area of natural sciences. However, this charge is not even justified in a scientific-historical sense. This is because the concept of information in biology is rooted not in the idea of communication but in that of material instruction. A clear example of this is provided by Schrödinger's elaborations on the structure of chromosomes in the 1940s [15].

The later information theory of Shannon likewise fails to provide any direct link to the forms of interpersonal communication. It is true that Shannon initially called his information-theoretical approach "communication theory"; however, he was interested only in the machines, the technology of message-transfer and the channels along which this proceeds (see above).

Consequently, the concept of information as used in the natural sciences and in engineering is not a "natural" concept, but rather a "structural" one. Let us look at this in more depth by considering the entropy of the source of a message (4). For this purpose, we must first lay open the structure of the idea of entropy. This leads us, as a first step, to the need to appreciate the distinction between two levels of description of a system: its microstate(s) and its macrostate.

In information theory, the macrostate describes the higher-order properties of a sequence of symbols, such as its length $n$ or the size $\lambda$ of the set of symbols it uses. In contrast, the microstate is one of the $\lambda^n$ possible ways of ordering the symbols. According to Shannon the quantity of information of such a sequence is given by the number of binary decisions that are needed to select a particular microstate out of the set of all possible microstates. The Shannon information thus represents the "potential" information contained in a macrostate, that is, everything that could possibly be known as a result of having complete knowledge of the entire microstate. In other words: Shannon's measure of information is determined by the accuracy with which the respective micro- and macrostates are defined.

Let us now consider the entropy function of statistical thermodynamics. Here, the macrostate of a material system is given by a number of state functions such as pressure, volume, temperature and the like, while the microstate is given by the exact specification of the position and momentum co-ordinates of all the particles that make up the system (or, put differently, a complete description of its approximately time-independent quantum state). As a rule, such a system contains an unimaginably large number of possible microstates. On the assumption that all the microstates have the same prior probability of being realised, the system will in most cases be in the macrostate that has greatest statistical weight, i.e., the greatest number of microstates. Such a distribution is only very rarely followed by another with a lower statistical weight if, as is usual with atomic or molecular systems, the particle numbers involved are large.

Boltzmann suggested that physical entropy should be understood as a strictly increasing function of the statistical weight of a distribution and the entropy of a macrostate should be set to be proportional to the number of its microstates. The mathematical formulation of this idea leads to the equation

$$S = -k \sum_i p_i \ln p_i \,. \tag{11}$$

Here $p_i$ is the probability of realisation of the $i$th microstate and $k$ is a proportionality factor known as Boltzmann's constant. For thermally isolated systems, all microstates possess the same prior probability $p_i = 1/p$, so that the above equation adopts the form

$$S = k \ln W, \tag{12}$$

where $W$ is the number of possible microstates that make up a macrostate. This number is also called the "thermodynamic probability"; it should however not be confused with the mathematical probability, which always lies in the range $0 \leq p \leq 1$.

Thus, apart from its sign, the Shannon entropy has the same mathematical structure as the Boltzmann entropy as given by (11). Strictly speaking, informational entropy and thermodynamic entropy are two structurally equivalent partition functions, of which one refers to the distribution of messages and the other the distribution of energy among the various quantum states.

In consequence of the structural aspects described above, it is therefore justifiable to equate information with negentropy. Boltzmann had already recognised this when he laid the statistical foundations of entropy, pointing out that the entropy function is at the same time related to the information that we have about the system [2]. Indeed, for the case of an ideal gas this is immediately obvious: the temperature of such a gas is directly proportional to the mean kinetic energy of the gas molecules and represents an *intensive* quantity—that is, it depends not upon the number of particles for which the mean value is calculated. It necessarily follows from this that there must be another quantity, complementary to temperature, that is proportional to the size or extent of the system. Were this not the case, then our knowledge of the total thermal energy of the system would be incomplete. This *extensive* quantity is the entropy.

The entropy function thus is a measure of the loss of information that arises in statistical physics when one dispenses with exact knowledge of the microstate. In this way, Boltzmann brought into focus—for the first time—the fundamental dependence of scientific statements upon the means and abstractions used to describe them.

The view of information and negentropy as being equivalent furnishes the concept of information not with a naturalistic, but rather with a structural interpretation. This leads to the collapse of the frequently expressed criticism that science wrongly regards information as a natural entity. It now becomes clear that this criticism rests upon a fundamental misunderstanding. In fact, the natural sciences rarely concern themselves directly with natural entities. Of much more importance for a causal understanding of Nature are the relationships between such entities. This becomes especially clear with the example of physical entropy. Although it is true that entropy is an object of scientific discourse that can be described by a physical quantity, it is not a natural entity in the same sense that trees, stones or other material objects are. Rather, the entropy of a system is a statistical partition function, the objective nature of which lies exclusively in the objective nature of the distribution of energy that it expresses.

Information is precisely *not* a "natural" entity in the narrow sense, but rather a "structural" entity that is structurally equivalent to the physical entropy. Consequently, information theory should be assigned neither to the natural sciences nor to the humanities. It belongs to the rapidly increasing branch of the so-called structural sciences. Their programme is to describe the comprehensive, abstract structures of reality, independently of where we encounter them and of whether they characterise living or non-living systems, natural or artificial ones. The structural sciences form a self-contained scientific view of reality, seeking the elements that combine the traditional natural sciences and the humanities.

In general, one obtains the laws of structural sciences by removing all the empirical constants from empirical laws and replacing them with logical constants. The structural laws then have the same syntactic structure as the empirical laws from which they were derived. In science, there are numerous examples of such structural equivalences. A very simple example from physics is the parity of structures of the law of gravity, which describes the attraction between two masses, and Coulomb's law, which describes the force between two charges. One may also mention the Fourier equation, which is applied in electrodynamics, hydrodynamics and thermodynamics. Another well-known example is that of the Lotka–Volterra equations, which are used for the description of dynamic systems not only in biology, but just as much in chemistry and economics.

Above all, the emergence of the structural sciences was due to the investigation of complex systems. As there are no additional natural laws that apply only to complex systems, their underlying laws must be sufficiently abstract to do justice to the vast diversity of complex phenomena, each with its individual characteristics. This explains the spectacular growth of the structural sciences in our day. Alongside information theory, the middle of the last century saw the establishment of cybernetics, game theory, systems theory and semiotics. The classical structural sciences have been enriched in recent decades by disciplines as important as complexity theory, network theory, synergetics and the theory of self-organisation. At the same time, the structural sciences form a mould for possible sciences, comparable with mathematics, which one may regard as a prototype of a structural science. Unlike mathematics, however, the structural sciences develop only through interaction with the experienceable structures of reality. And, most importantly: the structural sciences also contain the key that opens the door to the semantic dimension of information.

## 5 A Quantitative Approach to the Semantics of Information

Meaningful information in an absolute sense does not exist. Information acquires its meaning only in reference to a recipient. Thus, in order to specify the semantics of information one has to take into account the particular state of the recipient at the moment he receives and evaluates the information. In human communication the recipient's state is insofar a unique one as it is determined by his prior knowledge, prejudices, desires, expectations and so forth. But can those individual and historically determined circumstances ever become the subject of an exact science based upon generalised principles, rules and laws?

Against this background, one must give credit to the efforts of the cultural philosopher Cassirer to bridge over the apparent dichotomy between the particular and the general [3]. The particular, as Cassirer argues, does not become such by being a thing apart from general principles, but rather by entering into these in a relationship of ever-increasing multiplicity. The individual and particular—as we might also say—“crystallises” out of the network of its dependences upon general principles. This is a highly interesting figure of thought, and it points to a way in which one may comprehend the individual and particularly by reference to the general, even though this can never be performed exhaustively.

On the basis of this idea, we will now develop a novel approach to information that finally leads to a quantification of the value and thus the semantics of information. To do this, we must first see what the general rules are that constitute the semantics of information in the sense explained above. According to prevailing opinion, at least four general aspects are decisive: the novelty value, the pragmatic relevance, the selectivity and the complexity of information.

Already in the 1950s, Bar-Hillel and Carnap undertook the attempt to quantify the meaning of information on the basis of philosophical and linguistic considerations [1]. They allowed themselves to be guided by Shannon’s idea, according to which the essence of information consists in removing uncertainty. Setting out from this idea, they transferred Shannon’s information metric to linguistic statements and assessed the information content of such statements on the basis of how strongly the set of possible expected statements could be restricted by the arrival of a particular statement. However, this concept could only be realised within an artificial language, so that the range of its applicability necessarily remained extremely narrow.

In comparison with this, the idea of measuring semantic information by its pragmatic relevance has been much more fruitful. The pragmatic approach to the semantics of information has been favoured by many authors (see for example [13]). This approach takes account of the fact that meaningful information can be characterised by the reaction which it initiates on the part of its recipient. Such reactions can in turn express themselves in actions taken and thus provide a measurable indicator of the content of the information in question. The concept of the “pragmatic relevance” of a piece of information is by no means restricted to consciously acting recipients. Under certain circumstances the pragmatic relevance of a piece of information is already seen in the functional mosaic of the receiving system, as for example is the case in the living cell when genetic information is retrieved and processed.

This functional aspect of information has been investigated in detail by Eigen in his theory of the self-organisation and evolution of matter [6]. The theory shows that at the level of genetic information the evaluation of the functional content of information takes place through natural selection. Moreover, in this case the semantics of genetic information can even be objectified and quantified by its “selection value”. In general, the basic assertion holds: evaluated information is selective, whereby selectivity reveals itself as being a fundamental aspect of semantics.

Another theoretical approach which has been developed by the author is aimed at the prerequisites that have to be fulfilled by syntax if it is to be a carrier of semantic information [11]. One such prerequisite is the vast aperiodicity of the arrangement of the symbols, as only aperiodic sequences possess adequate scope for the encoding of meaningful information. From this point of view, the aperiodicity of the syntax can be considered as a realistic measure for the complexity of the semantics that can be built up on it.

In face of the different approaches to the semantic dimension of information one might obtain the impression that there is a basic contradiction between them. However, this is not the case. On the contrary: the breadth of this variation is a necessary consequence of the idea described above, according to which the semantics of information arise through the multiplicity of its general aspects.

Let us demonstrate this by referring to an example from human language. Consider the statement: "It will rain in Berlin tomorrow". Anyone receiving this message can evaluate it according to various criteria. It can, for example, be evaluated for its novelty. But if a recipient of the message has already heard the weather forecast, then its value for him will at best be that of confirming information already in his possession, and the novelty value will be low.

However, one may also ask what the pragmatic relevance of the message is. This is equally large for all its recipients who are likely to be exposed to the weather in Berlin of the following day. Conversely, it will have no (or only marginal) relevance for those recipients who in any case are planning to spend that day outside Berlin.

Comparable considerations apply to the selectivity of a message. Within the frame of all conceivable weather forecasts, our example is highly selective and, thus, considerably more meaningful than a message stating that tomorrow it will rain somewhere in the world. But even that information is dependent upon its recipient. For a recipient outside Berlin, both messages are presumably equally uninteresting. On the other hand, the selective value of the message will increase with the recipient's closeness to Berlin on the day in question.

Last but not least, the complexity of a piece of information clearly also plays a part in its evaluation. If one reduces the complexity of the message "It will rain in Berlin tomorrow" to the sentence "It will rain in Berlin", then the information content for most recipients will fall because of the message's poor selectivity in respect of the time point referred to.

This example demonstrates that general principles alone do not constitute a sufficient condition for determining the value and thus the semantics of a given item of information. It is rather the specific weighting of these principles by the recipient that gives the information its individual and particular meaning.

In general, the essential characteristics of semantics, such as the novelty value, pragmatic relevance, selectivity, complexity, confirmation and so forth, can be conceived of as elements $W_k$ of a semantic code $C_{sem}$:

$$C_{sem} = \{W_k\} \qquad (k = 1, 2, \ldots, n). \tag{13}$$

However, in contradistinction to the usual understanding of a code, the semantic code does not possess any rules for assignment or translation. Rather, the elements of the semantic code determine the value scale that a recipient applies to a piece of information received. Strictly speaking, the semantic code is a superposition principle that, by superimposition and specific weighting of its elements, restricts the value that the information has for the recipient and in this way becomes a measure for the meaning of the information.

If the elements $W_k$ have the weights $p_k$ (with $\sum_k p_k = 1$) for the evaluation of a piece of information $I_j$ by the recipient, then an adequate measure for the information value $W(I_j)$ would be a linear combination of the weighted elements $W_k$ :

$$W(I_j) = \sum_k p_{jk} W_k \quad \text{with} \quad \sum_k p_{jk} = 1 \qquad (j = 1, \ldots, m). \tag{14}$$

This measure, in turn, has the same mathematical structure as the entropy of a message source (4). However, in place of the weighted messages, relation (14) contains the weighted values $W_k$ of a chosen message $I_j$. At the same time, the number $k$ is a measure of the fine structure of the evaluation scale: the greater $k$ is, the sharper, i.e., the more differentiated is the evaluation by the recipient. In the limiting case, where the only value a recipient attaches to a message is its novelty, (14) reduces to the information measure (2) of classical information theory.

The information value $W(I_j)$ is a relative and subjective measure insofar as it depends upon the evaluation criteria of the recipient. However, for all recipients who use the same elements of the semantic code, and who for a given message $I_j$ assign the same weights to these elements, $W(I_j)$ is an objective quantity.

## 6 The Complexity of Semantic Information

The elements of the semantic code can be—at least in principle—quantified. For novelty value, Shannon's definition of information content is a good example. For complexity, the measure offered by algorithmic information theory seems appropriate. Let us look more closely at the latter, because it allows far-reaching conclusions to be drawn concerning the context-dependence of information. To investigate this, we take up the central thesis of philosophical hermeneutics, i.e., that one can only understand something if one has already understood something else (for this issue see [12]). This thesis we now put onto a quantitative basis. Expressed in the language of information theory, the question would be: how much information is needed in order to understand another piece of (meaningful) information?

An exact answer to this question seems at first glance impossible, as it still contains the problematical concept of "understanding". However, surprisingly, an answer is nonetheless possible, even though one must restrict the consideration to the minimum condition for any understanding. This minimum condition is the obvious one that information which is to be understood must first be registered by the receiver. On the other hand, there are good reasons to assume that a symbol sequence carrying information must have an aperiodic structure (see Sect. 5). For such sequences there is however no algorithm that would make it possible to deduce the rest of the sequence if only a part of it were given. Therefore, the recipient must be in possession of the entire information in question, i.e., of the entire sequence of symbols, before the actual process of understanding can commence. Thus, even the act of registration demands a quantity of information that has at least the same degree of complexity as the sequence of symbols that is to be understood.

Let us take an example. Consider again the information “It will rain in Berlin tomorrow”. We are not concerned with its truth content, but only with the complexity of the sequence of letters, which we can at the same time regard as a measure of the complexity of its semantics. In the above sentence, the complexity of the sequence of letters is at a maximum, since there is no algorithm that would be shorter than this sequence and with which at the same time the sequence could be extended or augmented. The structure, i.e., the syntax of this sentence is aperiodic and, in that sense, random. If, in contrast, the sequence were periodic, or largely periodic, then its inherent regularity, or law-like structure, would allow it to be compressed—or, if a part of it were already known, would allow the other part to be generated. This means that, from a syntactic point of view, meaningful letter sequences are always random sequences. However, this statement should not be inverted! Not every random arrangement of letters represents a meaningful sequence.

These conclusions are fundamental. They remain unaffected by the criticism that every language possesses syntactic rules according to which the words of the language are allowed to be assembled into correctly formed sentences. However, such rules only restrict the set of random sequences that can carry meaning at all. They do not allow any converse inference to be made about the meaning content itself. This is because the meaning of a sentence invariably depends upon the unique order of its letters.

One can express this in another way: meaningful information cannot be compressed without loss of some of its meaning. Of course, the content of a piece of information may sometimes be reduced to its bare essentials, as done in telegram style or in boulevard newspapers, but some information is always lost in this process.

Nevertheless, the fact that there is no compact algorithm for the letter sequence discussed above cannot be proved in a strict manner. This is because there might be some simple rule, hidden in the letter sequence, which we have not noticed until now. However, this hypothesis is arbitrarily improbable, as almost all binary sequences are aperiodic, i.e., random.

Even if the randomness of a given individual sequence cannot be proven, one can at least determine the proportion of sequences with—let us say—a complexity of $K = n - 10$ among all combinatorially possible binary sequences of length $n$. This is done by simply counting the sequences that can generate a sequence of complexity $K = n - 10$.

There are $2^1$ sequences of complexity $K = 1$ with this property, $2^2$ sequences of complexity $K = 2, \ldots\ldots$ and $2^{n-11}$ sequences of complexity $K = n - 11$.

The number of all algorithms of complexity $K < n - 10$ thus adds up to

$$\sum_{i=1}^{n-11} 2^i = 2^{n-10} - 2. \qquad (15)$$

As no algorithm with $K < n-10$ can generate more than one binary sequence, there are fewer than $2^{n-10}$ *ordered* binary sequences. These make up one $2^{10}$th of all $n$-membered binary sequences. This means that among all binary sequences of length $n$ only about every thousandth sequence is non-random and possesses a complexity $K < n - 10$.

To summarise the result: in order to understand a piece of information, one invariably needs background information that has at least the same degree of complexity as the information that is to be understood. This is the sought-for answer to the question of how much information is needed to understand some other piece of information. This finding gives the phenomenon of the context-dependence of information and language a highly precise form.

## 7 Concluding Remark

According to the model of the semantic code developed in this paper, the specific content of meaningful information is constituted by the superposition and the weighting of the general aspects of that information. It is quite conceivable that meaningful information may be assembled according to the same pattern in the brain. Thus, theoretical and experimental research suggests that brain cells, stimulated by sensory impressions, join up into a synchronous oscillating ensemble, whereby the individual information segments, scattered over different parts of the brain, are united into a self-consistent cognitive structure. In the light of the above considerations the synchronisation of oscillatory activity appears as the neuronal expression of a principle that we have identified as the generating principle of semantics. It may be that this principle will prove to be a useful model for the so-called neural code that neurobiology is seeking and which is assumed to be the guiding principle for the constitution of meaningful information in the brain.

## References

1. Bar-Hillel, Y., Carnap, R.: Semantic information. Br. J. Philos. Sci. **4**: 147–157 (1953)
2. Boltzmann, L.: Vorlesungen ber Gastheorie, 2 Bd. J. A. Barth, Leipzig (1896/1898)
3. Cassirer, E.: Substanzbegriff und Funktionsbegriff. Wissenschafliche Buchgesellschaft, Darmstadt (1980)
4. Chaitin, G.J.: On the length of programs for computing finite binary sequences. J. Assoc. Comput. Mach. **13**, 547 (1966)

5. Deleuze, G.: Woran erkennt man den Strukturalismus?, p. 8. Merve, Berlin (author's translation) (1992)
6. Eigen, M.: Selforganisation of matter and the evolution of biological macromolecules. Naturwissenschaften **58**, 465–523 (1971)
7. Hartley, R.V.L.: Transmission of information. Bell Syst. Tech. J. **7**, 535–563 (1928)
8. Kolmogorov, A.N.: Three approaches to the qunatitative definition of information. Problemy Peredachi Informatsii **1**, 3–11 (1965)
9. Küppers, B.-O.: Information and the Origin of Life. MIT, New York (1990)
10. Küppers, B.-O.: The context-dependence of biological information. In: Kornwachs, K., Jacoby, K. (eds.) Information. New Questions to a Multidisciplinary Concept, pp. 135–145. Akademie Verlag, Berlin (1995)
11. Küppers, B.-O.: Der semantische Aspekt von Information und seine evolutionsbiologische Bedeutung. Nova Acta Leopoldina **294**, 195–219 (1996)
12. Küppers, B.-O.: Information and communication in living matter. In: Davies, P., Gregersen, N.H. (eds.) Information and the Nature of Reality, pp. 170–184. Cambridge University Press, Cambridge (2010)
13. McKay, D.: Information, Mechanism and Meaning. MIT, New York (1969)
14. Rényi, A.: Probability Theory. Elsevier, Amsterdam (1970)
15. Schrödinger, E.: What is Life?. Cambridge University Press, Cambridge (1944)
16. Searls, D.B.: The language of genes. Nature **420**, 211–217 (2002)
17. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948)
18. Solomonoff, R.J.: A formal theory of inductive inference. Inform. Contrl. **7**, 1–22, 224–254 (1964)
19. Weaver, W., Shannon, C.E.: The Mathematical Theory of Communication, p. 8. University of Illinois Press, London (1963)
20. von Weizsäcker, C.F.: The Unity of Nature. Farrar Straus Giroux, New York (1981)

http://www.springer.com/978-3-642-34996-6

Evolution of Semantic Systems
Küppers, B.-O.; Hahn, U.; Artmann, S. (Eds.)
2013, XII, 232 p., Hardcover
ISBN: 978-3-642-34996-6