

André Włodarczyk

Ch. De Gaulle University (Lille)

Sorbonne University (Paris)

(Centre for Theoretical and Applied Linguistics)

e-mail: wlodarczyk.andre@gmail.com

ORCID: 0000-0001-6706-8702

CONCEPTS AND CATEGORIES: A DATA SCIENCE APPROACH TO SEMIOTICS*

Abstract. Compared to existing classical approaches to semiotics which are *dyadic* (signifier/signified, F. de Saussure) and *triadic* (symbol/concept/object, Ch. S. Peirce), this theory can be characterized as *tetradic* ([sign/semion]/[object/noema]) and is the result of either doubling the dyadic approach along the semiotic/ordinary dimension or splitting the ‘concept’ of the triadic one into two (semiotic/ordinary). Other important features of this approach are (a) the distinction made between *concepts* (only functional pairs of extent and intent) and *categories* (as representations of expressions) and (b) the indication of the need for providing the mathematical passage from the *duality between two sets (where one is a singleton) within systems of sets* to *category-theoretical monoids within systems of categories* while waiting for the solution of this problem in the field of logic.

Last but not least, human language expressions are the most representative physical instances of semiotic objects. Moreover, as computational experiments which are possible with linguistic objects present a high degree of systematicity (of oppositions), in general, it is relatively easy to elucidate their dependence on the concepts underlying signs. This new semiotic theory or rather this new research program emerged as the fruit of experimentation and reflection on the application of data science tools elaborated within the frameworks of Rough Set Theory (RST), Formal Context Analysis (FCA) and, though only theoretically, Distributed Information Logic (DIL).

The semiotic objects (s-objects) of this theory can be described in tabular datasets. Nevertheless, at this stage of formalisation of the theory, lattices (not trees) can be used as working representation structures for characterizing the components of concept systems and graphs for categories of each layer.

Keywords: Sign theory, human language, data science, formal concept, monoid category, theory of semiotics.

Since signs may also be regarded as objects, capable of being themselves represented in form of other signs, there is no need (at least at first) to make a distinction between objects and signs.

Maria Nowakowska (1980)

...a modern tradition of psychological research has developed in which concepts and categories are seen as central to theories of knowledge representation and long-term memory.

J. A. Hampton and D. Dubois (1993)

1. Introduction

The reason why expressions of human languages are specifically partial and ambiguous is, first of all, that they are built out of signs which have a double functionality *per se* (semantics) and *per alia* (ontology). In addition, language expressions are context-dependent (they rely heavily on all the cognitive sensory skills of human beings) which, in turn, roughly refer to objects and situations “out there”. Noteworthy, the classic theories of language either attributed the overall thinking capacities to language only or refused any participation of language in the process of thinking. In this essay, I adopt the hypothesis of mental code even though its existence is still to be proven by neuroscientists. Although both mental code and language and the relationship between them are basically culture-dependent, we can nevertheless expect that the nature of the primary structures of the code, language, and operations on them are universal to humans.

At this point, an important philosophical (but, as we will see, also methodological) problem arises: is it possible to construct theoretically a system of primitives (*a priori*) or can we only try to discover it experimentally (*a posteriori*)? To answer this question, it would be necessary to evaluate more than a half-century of research on transformational generative grammar, including all the attempts of its diverse technological implementations on computers and the results achieved in all the other frameworks of Natural Language Processing (NLP) technology as well. As a matter of fact, the search for primitive elementary concepts (in philosophy and language studies known as “universals”) has failed, at least as of now, and, in the domain of computer science, investigations into the deep “pivot language” for machine translation purposes have been discontinued and rather rapidly replaced by the elaboration of “transfer techniques” aimed at bypassing the “pivot language” representations during machine translation processes. Still, this first step which consisted of the theoretical construction of an overall ontology failed in implementation. Person-

ally, I believe that the next step on this way will need to pass through a period of rather long-lasting technological effort in research on human languages (and especially, on languages belonging to as divergent families as possible) using methods of **data science** (Knowledge Discovery in Databases – KDD), as an example see the research project “Computer-aided Acquisition of Semantic Knowledge – CASK (Włodarczyk 2007), including **machine learning** (neural networks). Usually, objects when perceived are considered to be denoted as cognitive objects. In the present theory, families of ordinary concepts (o-concepts) in the mind, which contribute to represent ordinary objects, are named *noemata*. Given that the vehicles (support) of signs have also a physical nature, families of semiotic concepts (s-concepts), which contribute to represent semiotic objects, are named *semions*. In addition, the term “object” in this paper is used in two different ways: (1) as the *single* (formal) *object* of one or more semiotic concepts (*s-concept*) and (2) as the *object* of a semiotic monoid category (*s-category*). In this theory, “traces” of signs are considered to have a double functionality: (1) as *objects per se* (objects) and (2) as *objects per alia* (para-objects).

Language signs, when used, are organised by degree of complexity in 5 layers of semiotic categories: *sounds*, *syllables*, *words*, *phrases* and *sentences* (*utterances*) which appear through emergence (synergetic interaction between objects and features) in a strict dependence of context (data processed by all the senses of humans and their knowledge). Therefore, signs are represented by categories and concepts which underlie these categories. In other words, semiotic categories appear in context-dependent emergence from concepts (or rather their families, or even aggregations of their families). Rules are used only in cases of necessity to create new schemata of signs (not yet in memory) but, even when schemata are ready to use, there is also a need to use recombination rules in order to extend the base utterances (or their schemata) into grounded language messages. In this paper, I am going to argue that today the adoption of computer technology will lead to further progress in order to create new frameworks for a tetradic semiotic theory based on the alignment of *semiotic* objects with *ordinary* objects, thus refining both the dyadic view of signs (issued from the Saussurean tradition) and the triadic one (issued from the Peircean tradition). In addition, this approach leads to the view of human language semantics as (a) the activation (*per se*) and (b) alignment of semiotic code units with ordinary code units replacing the logical interpretation of form by its sense or the transformation of “surface structures” into “deep structures”.

2. A short reminder of semiotic issues in linguistics and logics

Roughly speaking, there are two complementary classic theories of semiotic systems:

- (1) a **structural linguistic** theory in which the sign is defined as a *dyadic* relationship between *form* ('signifier') and *sense* ('signified') – Ferdinand de Saussure [1857–1913],
- (2) a **logical semiotic** theory in which the sign is an element of a *triangular* relationship between *representamen* (symbol), *interpretant* (comprehension device) and *referent* (object) – Charles Sanders Peirce [1839–1914].

In structural linguistics, the traditional couple of form and sense, considered to be like a coin which has two "inseparable faces", gave rise to what is often qualified today as the dyadic semiotic theory by de Saussure for whom form corresponds to the signifier (*signifiant*) and sense to the signified (*signifié*) of signs which are located in the memory of humans. In this traditional setting, language units are arranged hierarchically, units of higher layers (words, phrases and utterances) are said to be 'meaningful' while those of the lower ones (sounds, syllables) are, as it were, 'distinctive'. Note however that distinctivity is an inseparable characteristic of any structured collection of objects (cf. Lapis 2014). This dichotomy of notions applied to the hierarchical character of signs is supposed to determine two *big* layers of expression components and is known in structural linguistics as 'duality of patterning'. Furthermore, as the semiotic units of all the layers exhibit at the same time (a) *sequential* (syntagmatic, "actual") and (2) *parallel* (paradigmatic, "virtual") relationships, signs are considered to have a twofold nature: they *combine* and *commute*. In addition to this, signs form clusters following various opposition principles. Hence, the meaning of each of them is relative to that of the others. The term *semiosis* is often used to designate all the semantic changes that language signs undergo both in synchrony and diachrony.

According to the classic semiotics by Peirce, without taking into account, besides the *representamen* (symbol) and *referent* (object), the third component of semiosis, the *interpretant* (concept)¹, the distinction between *form* and *sense* is substantially inadequate. During its history, the Peircean model of semiosis underwent a few important interpretations (cf. Richards and Ogden 1923). The resulting model is widely known today under the name of "semiotic triangle" containing a modernised version of the triad of totally correlated notions: *sign*, *concept* and *object*. The notion of *concept* in the original Peircean version of the model of semiotics renders well the idea of conceptual representations of semantic situations. There is no

doubt that, at its basis, the notion of *interpretant* introduced a pragmatic element (the skill of the speech act participant, i.e. an agent endowed with human language ability) into the definition of sign. Indeed, the general theory of signs should account for the utilisation of signs by men (human beings equipped with cognitive *systems*²) in order to elucidate the interpretation (assignment function) of language expressions as transformation leading to conceptual representations and vice-versa.

Consider now that classic logic symbols are a sort of artificial signs:

1. their support is formal, hence presumed to be transparent,
2. nevertheless, they can be spelled or listened to, imitating human language signs,
3. their phonetic structure usually follows one of their host human languages,
4. to note variables and constants, logics needs to use *names*,
5. the only reason for using multi-character symbols for constants and variables is that the alphabets of their host human languages contain a limited number of letters (the theoretical “duality of patterning” of logic symbols is hardly conceivable),
6. they obey the fundamental laws of ‘distinctiveness’ and ‘combinatorics on words’, but these do not usually form any systematic opposition system (they are just lists of symbols),
7. they combine following very few syntactic rules,
8. they can be interpreted with respect to one or more formally defined closed (possible) worlds.

3. Data science tools for modelling linguistic signs

Due to the history of mankind and to the inevitable limitations of our experience in the world, human cognition can only be consistent locally. Human languages are highly complex systems with heterogeneous nature of “parts of speech” which need sophisticated scientific tools for their exploration. If language objects were homogeneous, the explanation of linguistic expressions would undoubtedly have a rather easy and completely recursive character to which rules of simple formal grammars could probably apply.³ Despite enormous efforts by many research teams, machine translation has been declared a hard problem of Artificial Intelligence. As a matter of fact, theoretical linguistic endeavours and human language processing techniques are not yet well equipped neither with appropriate meta-theories nor with advanced data scientific methods and tools.

Consequently, under no circumstances can we expect that representations using solely such simple structures as hierarchies (“trees”) suffice to represent and process human language expressions. Instead, linguists should revise their philosophical positions so that it becomes possible to apply the notions and tools of distributed computing in order to create/develop a truly scientific paradigm in the study of language. In other words, the investigations into language domain should be based on data validated with the help of computational tools for processing interactively produced descriptions, with **data mining** methods (using not just text-mining⁴). Bearing in mind the diversity of operations needed for processing data, I propose to use the following computational techniques (jointly or separately) among others:

1. *Rough Set Theory* (RST – Pawlak 1982) which provides discerning for concept *approximation*
2. *Formal Concept Analysis* (FCA – Ganter and Wille 1999; Wille 1982, 2005) which provides calibration for concept *scaling*
3. *Distributed Information Logic* (DIL – Barwise and Seligman 1997) which provides classification for concept *integration*

Tables with objects and features (datasets) are basic data structures used in computer science. The research on their nature gave rise to what has recently been launched as **data science** that tries to integrate not only the investigations into the deep learning mechanisms using the neuromimetic intelligence technology but also the techniques of Knowledge Discovery in Databases (KDD) using management techniques of semantic/ontological databases. In computer science, there are a few approaches⁵ to the problem of mappings between objects and their features as assembled in tabular datasets: (1) *information systems* (RST), (2) *contexts* (FCA) and (3) *classifications* (DIL).

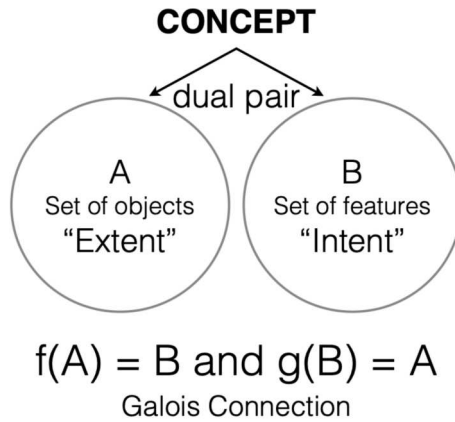
The way FCA looks at concepts is in line with the international standard ISO 704 that formulates the following definition: “A concept is considered to be a unit of thought constituted of two parts: extent and intent.” The extent consists of all objects belonging to the concept, whereas the intent comprises all attributes shared by those objects. Thus, the definition of a formal concept is based on the dual nature of the assignment function of features to objects but, from a strictly linguistic point of view, it is important to mention the fact that this duality cannot be found without the “opposition system” or, in FCA terms, without their “contexts” (tables of binary data).

The research of Karl Erich Wolff (2009) might serve as an interesting example of the use of several fields of data science with the aim of creating

a *relational semantic theory*. Let me quote the extended Formal Concept Analysis.

... we combine the structure of a many-valued context and the structure of a power context family in the notion of a Relational Data System such that each many-valued context, each power context family, and each concept graph can be represented by a Relational Data System. They will be used to define **Relational Semantic Systems** which offer a clear way of relational scaling (Wolff 2009: 67).

On the schema below, it is shown that formal concepts are dual pairs of sets of features as projected on objects. The duality is provided by the two functions $f(A) = B$ and $g(B) = A$ known as Galois Connection. This definition lies at the basis of the FCA theory by Rudolf Wille (1982, 2005) and by Bernhard Ganter and Rudolf Wille (1999).



Schema 1. The Definition of a Formal Concept as a Galois Connection

For the convenience of linguists, it may be handy to represent the model of systems with formal concepts as in the following triple:

- (1) $C = (A, \alpha, a)$ using the terminology as follows:
A is the *opposition system* to which a given semiotic unit belongs,
 α is a subset of features ($\{\text{intent}\}$) describing the *language units* of that system,
a – is a subset of *language units* ($\{\text{extent}\}$).

Most importantly, however, the concept defined this way, besides meeting the expectations of semioticians, has an important characteristic feature:

it is obtained automatically by the dual function which combines features with internalised traces of signs (s-objects). Let us compare by way of example two classic (logic and linguistic) analyses of such a simple expression like “white cat”. Following the syntax of the First Order Logic (FOL) language, we get the following formula:

$$(2) \exists x \text{ white}(x) \rightarrow \text{“there exists an } x \text{ such that } x \text{ is white”}$$

which after interpretation (i.e. after applying this formula to a set of cats) gives:

$$(2a) \exists \text{ cat white}(\text{cat}) \rightarrow \text{“there exists a cat such that this cat is white”}.$$

It goes without saying that only two layers are necessary for this analysis: a lower (elementary) layer of individuals and a higher (derivate) one of relations. Moreover, to make the interpretation happen, logics needs some application domain. In the case of FOL, logicians use set theoretical notions with respect to which their logical formulae take the truth values.

Now, in the proposed conceptual approach, signs being defined as pairs comprising “traces of expressions” and “features”, we get the following analysis (for the sake of brevity, only two out of 5 layers will be exemplified).

Layer 3 – *Words*: two semiotic units “white” and “cat” with their two underlying concepts:

- concept 1 = ({white}, {set of white objects}) and
- concept 2 = ({cat}, {pet}).

Layer 4 – *Phrases*: one semiotic unit “white cat” with its underlying concept: concept “cat” = ({cat}, {white}).

On every layer of the system, I propose to use the combination of the notion of **scaling** of *concept systems* (FCA) with that of **approximation** (RST) thanks to the fact that the *formal concepts* (FCA) correspond to *B-definable subsets of objects* (RST). Given that perception relies merely on partial views of objects, due to the synergy of (a) semiosis operating mostly on **sense** with (b) encapsulation operating mostly on **form**, semiotic representations help to improve (refine, deepen) these views through alignments not only with ordinary knowledge but also with data and information being acquired during the communication processes thanks to all the sensory devices of individuals. For this reason, one way to model these processes might be the utilisation of the *lower* and *upper* approximation techniques. Thus, combining the power of approximation (RST) with

that of scaling (FCA), it might become possible to process formal concepts associated with signs and, perhaps, even use them very successfully for modelling the underlying conceptual components of non-semiotic (ordinary) systems.⁶

The aim and meaning of Formal Concept Analysis as mathematical theory of concepts and concept hierarchies is to support the rational communication of humans by mathematically developing appropriate conceptual structures which can be logically activated (Wille 2005: 2).

The Formal Concept Analysis actually provides the mathematical capacity for defining concepts and scaling datasets. In FCA, all kinds of real-world entities (objects), once registered in datasets, are treated as *formal objects*. Indeed, in more general terms, **formal concepts** are very useful mathematical representations of both *real/fictitious* and *concrete/abstract objects*. The combination of FCA with RST is fruitful because the RST also presents a similar mathematical power in scrambling decisions while taking advantage of the ability to use approximation techniques.

Another reason for using data science technology in modern linguistics, is the fact that, being equipped with *autonomous* knowledge processing *systems*, humans use both *extensional* and *intentional* information which form their knowledge defined as an internal mental model of the external world. However, for clarity's sake concerning the feasibility, note that to make judgements and assertions, given that knowledge is inconsistent in face of the immense complexity of the world, man neither needs always to refer to the real world nor to be entirely rational.

4. Signs in the conceptual semiotic theory

It is namely Uta Priss (2002, 2020), a German mathematician, the author of the relational formal concept analysis (RFCA), who has formalised the traditional Peircean semiotic theory. She has also written many papers on the linguistic applications of FCA. Undoubtedly, one of the most interesting of them concerns language studies (especially contrastive linguistics) and treats human language expressions as *intents*. But in structural linguistics (the general theory of human language as a system), we need to account for the fact that (1) human language expressions *exist* in the real world in acoustic form, hence they also should be treated as objects, or perhaps more precisely, as “traces” of objects or *formal objects* (following the FCA framework terminology) and, consequently, that

(2) signs should be treated as one-element *extents* of formal concepts, not as their *intents*. For some linguists, this problem is similar to the one of *meta-language* for the sole reason that the objects of their investigations actually are languages. Some other linguists use the term *meta-language* to indicate special kinds of expression of their object language itself. However, even in the latter case, the definitions are often too large as they also cover problems which can be better captured using the term *meta-information* (cf. Włodarczyk 2013; Włodarczyk and Włodarczyk 2016a, 2016b, 2019a and 2019b).

As a matter of fact, applied to semiotic systems (and *a fortiori* to human languages), during the process of concept analyses (i.e. during communication), solely the formal concepts whose extents are composed of unique elements (singletons) are functionally efficient within the whole system.

Note however that, although the formal concepts which underlie the categorisation of semiotic objects are the result of their recognition in the real-world⁷ and memorisation as “traces” in the mind, it is clear that, using traditional terminology, one may say that signs are “vehicles” or “support” of information about not only real but also fictitious (abstract, mythical, symbolic etc.) objects. However, in the present approach to semiotics, the classical pair of notions form/sense is being replaced by the formal concept’s pair ({extent}, ({intent})), the latter being generalised as an outcome of its application to the expressions of all the layers of the information processing system. As will be shown below in section 6, processing signs consists in the awareness of the 5-layered emergence of *semiotic* units which are: (1) *phonetic*, (2) *syllabic*, (3) *lexico-grammatical*, (4) *syntactic* and (5) *sentential* (see Fig. 4). At this point again, I claim that this knowledge indeed needs to be, first of all and at least partly, treated in terms of ontologically grounded **formal concepts**, namely concepts whose extents have proto-objects as referents. Consequently, as is argued below, it is impossible to reduce either Form to *extent* or Sense to *intent*.

4.1. Semiotic and ordinary objects as formal objects

The distinction between semiotic formal objects (traces of signs) and ordinary formal objects (traces of things) is crucial for any theory of semiosis which supports the hypothesis of the existence of mental code, the structure of which is certainly more systematic (hence it is even more suitable for formal modelling) than the structure of human languages. Moreover, from the logic point of view, these two sorts of objects are linked together. Hypothetically, the correlation between both kinds of objects can be seen as an *alignment* rather than as any deductive system. In addition, both the com-

munication within human societies (**macro**-cognitive, social systems) and the common-sense knowledge of their members (**micro**-cognitive, individual systems) are better modelled using distributed systems than centralized ones. As languages undergo constant evolution, they form systems in which *semiotic* and *noemic* objects coexist in the distributed networks of knowledge. Intuitively, during the processes of conceptualisation (thinking and communicating) *s*-objects and *o*-objects both follow general laws of information transmission systems, hence, they can be assimilated to the merged clusters of smaller groups of concepts which can be represented by lattice diagrams.⁸ In functional linguistics, such collections of language units, where concepts are definable by the dual function as in FCA, are known as *opposition systems*.

4.2. Linguistic signs as objects and para-objects

In the proposed conceptual semiotic theory, formal objects are considered firstly as ‘traces’ of *semiotic* objects (objects *per se*) and, only in the second place, as aligned with *ordinary* objects or, more precisely, as para-objects (objects *per alia*) with the senses historically drawn from the perception of ordinary objects which are aligned with them (see Fig. 1). Linguistic signs are therefore objects and para-objects at the same time. Their peculiarity consists in that – in addition to conveying their proper (*per se*) semiotic information – they are *aligned* as objects *per alia* with the ordinary (non-semiotic) mental code constructs.⁹ Judging on the basis of analyses of human language signs seen as objects could perhaps make it possible to hypothesise that semiotic and ordinary objects, both being mental code units, are processed in the mind in a similar way (i.e. within hierarchies of layers) passing from lower (micro-) layer categories to higher (macro-) layer categories through multiple synergies between (1) features *per se* themselves, (2) features *per alia* themselves and between (3) features of encapsulation and those of semiosis (see Fig. 1, Fig. 2, and cf. section 6).

However, before this problem is properly solved, I will concentrate here on semiotic objects only, proposing to take advantage of the set-theoretical notion of *formal concepts* with one-element set of objects in extent and of the category-theoretical notion of *monoids*. Fig. 1 suggests that signs are processed “within” the layers of activated memory spaces¹⁰ where they are represented by (1) families of **semiotic formal concepts** (*semions*) and (2) **semiotic categories** (*semiotic units*). On the right side of fig. 1, ordinary real-world objects are represented by formal objects (“traces”) of families of ordinary formal concepts (*noemata*). In spite of all this, it is more than probable that *semions* and *noemata* are NOT processed in a similar way.¹¹

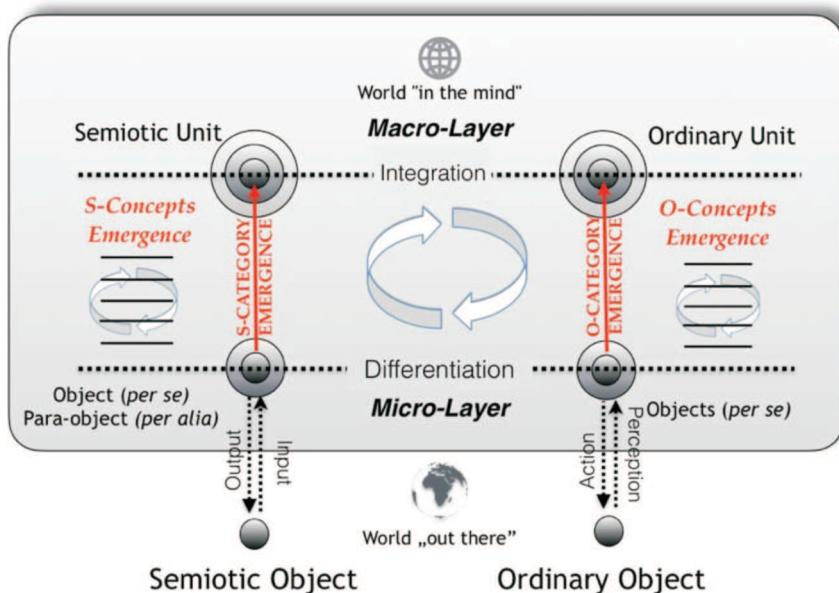


Figure 1. The Tetradic view of Emergence within Semiosis and Noesis and between them

Semioticians and linguists who traditionally consider that signs are objects which stand only *per alia* focus on the universal character of signs as substitutes of objects being themselves representations of elements of situations regardless of their belonging to the fragments of the world “out there” or their essentially abstract (only epistemic) nature. Contrariwise, linguists who restrict their views to signs as objects only *per se*, consider that signs determine specific mental representations (“the mind”) of external world objects; hence, departing from the common ontological ground of languages, thus paying attention mostly to the diversity of world languages. Both points of view appeared in the history of linguistic investigations as antagonistic, the first one is known as **universalistic**, and the second one as **relativistic**. In the present approach, introducing the distinction between *objects* (objects *per se*) and *para-objects* (objects *per alia*) makes it possible to “reconcile” the relativistic and universalistic points of view. It is important to mention that, just as semiotic categories sometimes differ very much in human societies speaking different languages, so ordinary (non-semiotic) categories are not the same for all societies and for all individuals because not only semiotic but also ordinary knowledge both depend on different cultures (different social and individual experiences). Hence, the ongoing cognitive investigations into ontology cannot have but a **normative** char-

acter. For this reason, we shall hopefully learn more about the problem of the alignment of semiotic categories with ordinary ones from neuroscience (of the brain), where active research is underway, than from cognitive psychology or cognitive linguistics.

Although linguistic signs should also be treated as physical objects, it is important to emphasise that concepts which represent signs (having their historically determined “literal” meanings) play an additional role, that of being aligned as semiotic para-objects with ordinary objects. This standpoint prevents me from treating as semiotic entities the *indices* which Peirce, in his semiotic theory, treated as a type of signs (e.g. smoke as an “index” referring to fire). As a matter of fact, due to human language ambiguity, there is a still prevailing misunderstanding of the important difference between *signs* (proper semiotic objects) which are aligned with mental representations (e.g. the English word “fire” referring to the mental representation of the object *fire*) and *indices* which should be classified as resulting from such cognitive capacities of human beings as learning and pattern matching (e.g. after having experienced a fire, one knows that – in the physical world – smoke (effect) is produced by fire (cause)). The issue here is that the nature of signs (semiotic units) lies in their reciprocal oppositions but that of objects (ordinary units) is still unclear. Nevertheless, *semiosis* should be seen as the dynamics of the “semantics (of signs)” in contrast to the dynamics of the “ontology (of things)” in *perception*.

5. Concepts and categories in human languages

As mentioned above, it is necessary to distinguish between the two following sorts of concept families:

1. ***Families of ordinary formal concepts (noemata)***, which are encoded in a sort of internal mental code and whose extents are “traces of things” (e.g. the category of an adult female person emerges amongst others from the underlying *ordinary formal concept*:
{image of a human being}, {sex: female}).
2. ***Families of semiotic formal concepts (semions)*** which are internally encoded but can be exteriorised (either in the form of human language expressions or, more generally, in the form of diverse signs) and whose extents are “traces of signs” (e.g. the category of the English word “woman” emerges amongst others from the underlying *semiotic formal concept*:
{trace of the word "woman"}, {gender: feminine}).

Indeed, I assume that (1) both semiotic objects (traces of signs) and ordinary objects (traces of things) are represented in the human mind by *s-categories* (semiotic units) and *o-categories* (ordinary units) respectively and that (2) these units are aligned with each other (see fig. 1). What is traditionally called *denotation* refers to the relationship between signs “in the mind” and objects in the world “out there”. What I propose is two kinds of relationships between signs (semiotic objects) and things (ordinary objects), both being internalised “in the mind” as (1) **singleton** extents of s-concepts and o-concepts, on the one hand, and as (2) **unique** objects of a special sort of categories (monoids); namely between semiotic categories (s-categories) and ordinary categories (o-categories), on the other hand. Hence, it is easy to predict the usefulness of notions such as *alignment* and/or *refinement* to characterise these two relationships between semiotic units and ordinary units.

Table 1
Elementary and Integrative Units of Language and Mental Codes

	Semiotic Units	Ordinary Units
Space of CATEGORIES	SIGNS	OBJECTS
Space of CONCEPTS	SEMIONS	NOEMATA

S-concepts and *o-concepts* constitute the building blocks of semions and noemata respectively and underlie the units that perhaps could be mathematically represented as *monoid* categories (i.e., categories having a unique object) that, in turn, integrate many of their features in the form of arrows pointing to themselves. Note that the proposed setting is an important paradigm shift in semiotic theory (table 1). As a matter of fact, once internalised, signs are characterised here as aggregates built as the result of the emergence of categories from a synergetic interaction between encapsulation and semiosis.

Classifications and their infomorphisms form a category in which classifications are objects and infomorphisms are morphisms (Burgin 2010: 473).

As will be shown below, all the conceptual “ingredients” of signs have singletons (one-element sets) as their extents within the dually defined pairs of ($\{\text{extent}\}$, $\{\text{intent}\}$).

Concerning *semiotic formal* concepts, their features are of two kinds:

- 1) **semiotic semantic** (traditionally considered in structural linguistics as displaying only ‘*distinctive*’ features of ‘form’) + features *per se* traditionally considered as displaying only ‘*significative*’ features of ‘sense’ and
- 2) **ordinary semantic** or **ontological semantic** (traditionally considered in structural linguistics as belonging only to ontology) with links *per alia* to noemata.

However, although the extents of semiotic concepts taken as objects *per se* have their proper semiotic features, the nature of their relationship as objects *per alia* with noemata is based on some sort of free unification that I will call provisionally **alignment**, not on interpretation. In other words, semiotic systems – judging from their most representative sort which are human languages – convey their own quasi-autonomous semantic ‘contents’ (information) and only **align** with ordinary knowledge.

5.1. Linguistic signs as categories (monoids)

The vertical arrows on Fig. 1 above indicate two kinds of inverse operations (recognition/production) which are at work between various layers of semiotic objects hierarchy. Hypothetically, operations of the same kind (referred to as perception/action) occur between layers of ordinary objects. The **semiotic formal objects** (“traces”, singleton extents of semions) should be treated as bearing their proper features, i.e. they form concepts which underlie the categories of the given layer in two ways: (1) as objects *per se* through synergies between encapsulation and semiosis giving rise to emergence, on the one hand, and (2) as objects *per alia* within a space of complex unifications (alignment, refinement and the like) with **ordinary formal objects** of noemata, on the other hand.

Operations such as the optimisation of formal concept systems underlie both the conceptualisation and categorisation of semiotic (and a fortiori perhaps also of ordinary) object systems. As represented on Fig. 2, concepts underlying signs of the lower (micro-)layer are subject to operations which result in the emergence of higher (macro-)layer signs. I propose to call this process of synergetic interaction between encapsulation and semiosis *conceptualisation*, while the process of building the higher layer units out of the lower ones I propose to call *categorisation*, which consists of free but complex unifications. The processes of categorisation are therefore due to the synergetic dynamics of concepts, making it possible to reduce signs and messages to solely relevant contents (see Fig. 2). Thus, categorisation concerns the construction of more complex units out of simpler ones on each

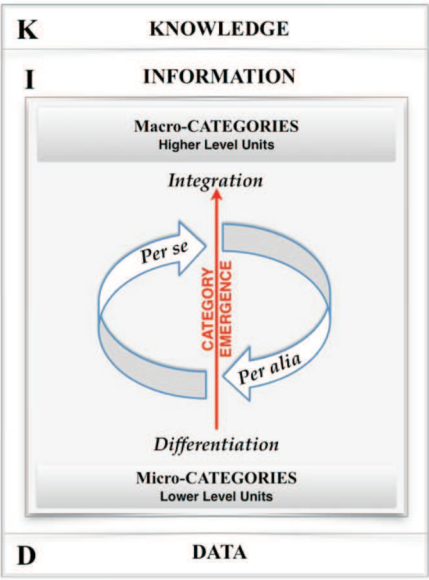


Figure 2. Information as the result of two kinds of processes (conceptualisation and categorisation) passing through multiple layers of its emergence from data to knowledge

layer (in section 6, see the discussion in more detail of the multi-layered structure of human language utterances).

The rotating arrows (inside box **I** on Fig. 2) depict the synergy (creative interaction) between (a) semions themselves and between (b) noemata themselves, in particular, and between (c) semions and noemata, in general, resulting in the formation of higher layer units (macro-categories) from the lower ones (micro-categories). Let us call *category emergence* the synergy (1) which occurs between all kinds of concepts (conceptualisation) and, depending on the context¹² in which new concepts are created, (2) gives rise to the essentially partial integration of the result into more or less complex units. Emergence is defined as a phenomenon that occurs between multiple layers of nature.

In philosophy, systems theory, science, and art, emergence occurs when an entity is observed to have properties its parts do not have on their own, properties or behaviours that emerge only when the parts interact in a wider whole. Emergence plays a central role in theories of integrative levels and of complex systems. For instance, the phenomenon of life as studied in biology is an emergent property of chemistry. (<https://en.wikipedia.org/wiki/Emergence>), last access: September 1st, 2022.

Applied to semiotics, and particularly to linguistics, this phenomenon sheds new light on the socio-natural ability of humans to communicate using very complex systems. Given that semions (defined as families of semiotic formal concepts which underlie the emergence of categories) have **single** element extents (language expression units), the language categories which result from emergence constitute a special kind of categories; they are **monoids** (i.e. categories known in mathematics as having a unique object with arrows pointing to itself only).

In fact, as described in section 6 below, the processes of transforming expressions seen as **data** into their meanings (information) and their implications into knowledge pass across a stratified system of components where the same operations (emergence and application of combination and recombination rules) are repeated at every layer. It is also noteworthy that besides Formal Concept Analysis (for **concepts** scaling) which inspired the foundations of the research program presented, two other data science approaches (a) Rough Set Theory (for **data** approximation), and (b) Distributed Information Logic (for **categories** distribution) are at work during the processes of information transmission/acquisition. And although the protagonists of these theories sometimes attempt to cover the whole domain of data science while extending their theories, in my view, all three mentioned approaches should be used altogether rather than as alternative competing ones. And in the same way as concepts and categories cannot be treated separately, so data science objects need to be processed together though on different stages of processing.

Table 2

The universal nature of Data-Information-Knowledge (DIK) Hierarchy
(every intelligent system acquires knowledge by the transformation
of data into information)

Domain	The DIK Pyramid		
Knowledge Engineering	Data	Information	Knowledge
Cybernetics	Support	Intermediary Transformations	Comprehension
Granular Computing	Hard_suit Granules	Link_suit Granulation	Soft_suit Complex granules
Neuroscience of Brain	Primary Sensory Area	Association Area	Higher Order Association Area
Data Science	Rough Set Theory	Formal Concept Analysis	Distributed Information Logic

In the proposed theoretical setting, **information** appears to be the result of an interface-like device (an invertible system) between boxes **D** and **K** (Fig. 2). Given that human languages can be seen as information transfer devices resulting of the conceptualisation and categorisation of signs, the new approach to language processing within the investigation methods should adopt the name of “human language processing” (HLP) defined as “encoding/decoding of information within the multi-layered semiotic data processing device”, instead of “natural language processing” (NLP) which has existed for at least half a century.

5.2. Encapsulation in human languages

The problem of encapsulation can be formulated as follows: while sign-objects are “public”, referent objects are “private” i.e. sign-objects encapsulate referent objects, which causes underspecification or partiality of meaning. Para-objects represent **partial views** (selected fragments) or **compressions** (e.g. some conjoined parts of a whole) as *opposed to* their counterparts (e.g. wholes) of object representations.¹³ This corresponds to the pragmatic principle in accordance with which – in order to make their statement effective and concise – speakers should not “say everything” about a given situation but only what is relevant for their current purpose.

As concerns the partiality of linguistic utterances, we showed in our previous research (cf. Włodarczyk and Włodarczyk 2013, 2019) that the information they convey does not consist exclusively of *ortho-information*, which would reflect the whole knowledge of speakers about the situation they talk about. What we called *meta-informative* devices¹⁴ make it possible to communicate only some chosen aspects of the situation and to highlight them among other parts of the same situation which are treated as secondary and can remain implicit (because “it goes without saying”), since they may be easily reconstructed on the basis of the common knowledge shared by users of a language and by the participants of a given conversation. Moreover, any chosen linguistic unit is “virtually” linked to other units belonging to the same subsystem. Let me call *para-information* such information that refers to that which is activated in memory though, using linguistic terminology, not “actualised”. As an example, let me evoke the meaning of the definite article “the” in English which results from the fact that it can be chosen among other articles (the indefinite “a” or even the absence of any article). Lexical units also belong to para-informative networks, e.g. saying that “someone divorced” entails that this person was previously married. Thus, both para-information and meta-information are

important components of linguistic information, making it possible to make linguistic messages as concise as possible. This property of human language utterances is seen as a defect by logicians whose formulae do not entail any implicit content. However, although it is difficult to model and formalise meta-informative and para-informative devices for language processing, it appears that native speakers are trained to use them unconsciously seemingly without special effort.

5.3. Semantics as alignment of semiotic categories with ordinary categories

Let me return now in more detail to the problem of alignment, in particular so far as it concerns human language semantics. Both semiotic and ordinary objects are analysed/produced by the cognitive systems of intelligent agents. Instead of the commonly accepted direct mapping between signs and objects of the world, I claim that semantics consists in the alignment of two different sorts of formal (mental) objects: semiotic units (semiotic para-objects) and cognitive units (non-semiotic objects). This point of view makes it possible to shed new light on the distinction between literal and proper senses not only from the point of view of the pragmatic use of language in human communication.¹⁵ As a matter of fact, the distinction between s-concepts and o-concepts can be seen as a step forward to the formalisation of the traditional opposition of literal and proper meaning. In Fig. 3, horizontal arrows represent the alignment of semiotic ontology with ordinary ontology, that of semiotic information with ordinary information and that of semiotic knowledge with ordinary knowledge. Thus, in the semiotic framework I proposed, alignment is one of the *proper* semantic functions of language.

In fact, as a result of this theoretical shift, linguistic expressions have two sorts of semantics: semantics *per se* and implications of semantics *per alia*. From the point of view of logics, propositions undergo both **interpretation** (satisfaction) and **derivability** (admissibility). But in the case of human language utterances, the **semantic** content of *semiotic units* (*semions* and *categorie*) draws from reciprocal **dependencies** (e.g. dependency of decisions from conditions) (a) between themselves and (b) between them and *ordinary units* (*noemata* and *o-categories*). Consequently, the semantic content of linguistic expressions is defined neither from a purely **realistic** viewpoint nor from a purely **nominalistic** one. Indeed, using formal concepts makes it possible to provide a quite new and more complete definition of semantics, namely, it enables us to treat jointly extent and intent from the very start.

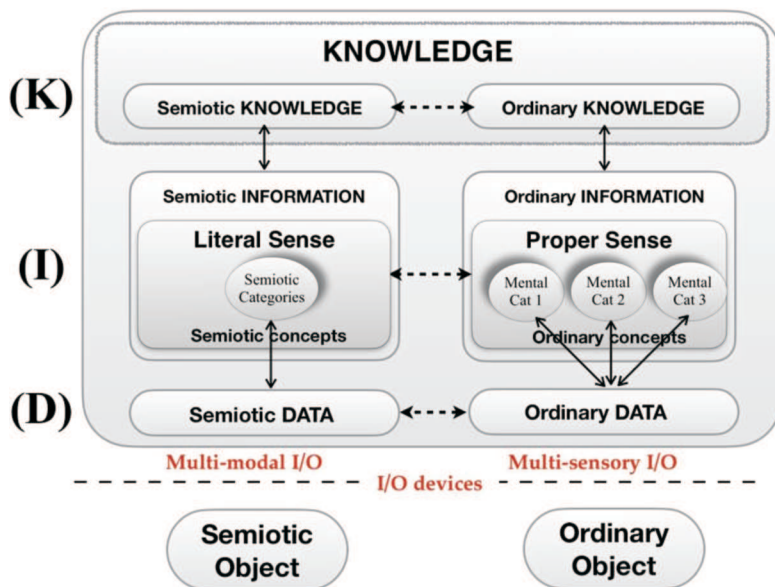


Figure 3. Relations between the multi-modal and multi-sensory Data (D), the Literal and Proper Senses (I) and the integrated semiotic and ordinary Knowledge (K)

Taking the above theoretical position makes it possible to replace the foundational issues of *possible world semantics* (ex. the problem of mythical beings like *Pegasus*) and even to cope with (apparently) deficient constructs. For example, the fact that the non-semiotic categories without alignments with the semiotic ones are felt as *unnamed* or *unspeakable*, and vice versa, that the semiotic categories without alignments with the non-semiotic ones are referred to as *incomprehensible* (ex. *abracadabra*). In formal terms, not all ordinary objects are paired with semiotic objects (i.e. are named) and, vice versa, not all semiotic objects designate ordinary objects (this is known as the coverage problem).

5.4. Ambiguity of category alignments

The alignment of semiotic categories with ordinary categories is (most probably always) problematic. For example, some signs are synonymous (designate the same referent) and some others are polysemous (designate more than one referent). As an example, let us show how we can resolve the problem of sense and meaning (Frege's *Sinn und Bedeutung*) caused by the use of two different signs to refer to the same object of the world out there. The planet Venus is named the “morning star” when observed at dawn and

the “evening star” when observed at dusk. Hence, the distinction “morning star” / “evening star” concerns the semiotic categorisation (with respect to the observer’s location on the scale of daytime) regardless to the fact that there are one or two ordinary objects. For this reason, it becomes possible to attach these attributes to the semiotic category and not to an ordinary category. Speakers using the expressions “morning star” and “evening star” may know or not that these expressions are signs aligned as para-objects with the same ordinary object, however when choosing one of them they select a different point of view to which different images, feelings and emotions are attached. “Morning” is used in human languages as a metaphor of beginning, renewal, optimism etc. while “evening” calls to mind opposite notions. To explain why metaphors are so important in human language (Lakoff and Johnson 1980) the distinction between *semiotic* concepts and *semiotic* categories, on the one hand, and *ordinary* concepts and *ordinary* categories, on the other hand, appears to be crucial. As another example, notice how poets and writers make use of the literal meaning of the expression “the sun is rising” although everyone knows since the times of Copernicus (and today, everyone can even see it on satellite pictures) what is the real ontology “hidden” behind this expression.

6. The multi-layered arrangement of semiotic categories in human languages

Conceptual semiotics constitutes the theoretical foundation of *conceptual linguistics*. It had its roots in multiple experimentations with language data within the computational framework of *interactive linguistics* (Włodarczyk and Włodarczyk 2019b) elaborated on the basis of data science methods. The results obtained in my research within this framework showed that a special kind of lattice is equivalent to a kind of neural network with regard to human language. This observation suggests therefore that the multi-layered arrangement of semiotic categories in human languages corresponds to a cascade of neural networks. More advanced experimental work is needed however in order to prove that it is possible to get two complementary models of the conceptual processing of data which underlies semiotic categories: one *conceptual* (qualitative, “symbolic”) and the other one *neuro-mimetic* (quantitative, “statistical”). Let me also assume that, here again, the qualitative and quantitative processes are most probably both active in parallel during the data-information-knowledge processing.

6.1. Multi-layered pairwise patterning of human language utterances

The processes of encapsulation and semiosis I sketched out above made it possible to reconsider the problem of double patterning as it is usually understood in classical structural linguistics. As a consequence, this led me to work out a new model of the layered patterning of information during the processes of communication in human language. My approach is close to some theses of stratificational linguistics (Lamb 1999) but, in general, it follows the philosophy of functional linguistics and should be seen as the attempt to systematise the points of view of structuralist theories of phonology, morphology and syntax, and to fill their gaps.

Indeed, the main characteristics of the structure of this new model of human language components is its **multi-layered pairwise patterning**. In structural linguistics, phonemes are considered to be distinctive units without meaning; consequently, the structural theory of double patterning (concerning the opposition of distinctive function vs. meaning function) draws a sharp limit between the first and second patterning, between phonemes (viewed as “distinctive units”) and all linguistic categories of superior layers beginning with morphemes up to whole utterances viewed as “meaning units” (see part A on the left of Fig. 4). Yet, considering that units on each layer can be treated as categories, it should also be possible to process the concepts which underlie them as having both extent and intent. I claim that it is possible to generalise the idea of “distinction/meaning” relationship at each structural layer of linguistic units (part B on the right of Fig. 4).

Thus, phonemes can be represented by families of formal concepts of the phonic layer making it possible to interpret voice sounds as arranged into language units (categories) of a higher layer: *syllables*. Similarly, categories of successive layers composing linguistic utterances (words, phrases and utterances) are conceptualised by families of formal concepts, in this case, forming the units (categories) of successive higher layers (morphemes, phrasemes and predicemes). On each layer of the hierarchy, new categories *emerge* due to synergy (i.e. multiple interaction processes between encapsulation (*per se*) and semiosis (*per alia*) operating on formal semiotic concepts). Therefore, (1) **syllables** and chains of syllables emerge out of voice sounds conceptualised as *phonemes*, (2) **words** emerge out of syllables and chains of syllables which undergo conceptualisation processes as *morphemes*, (3) **phrases** (chains of words) are conceptualised as *phrasemes* and, on top of all these layers, (4) **utterances** (chains of phrases with characteristic prosody) emerge from phrases conceptualised as the families of formal concepts I call *predicemes*.

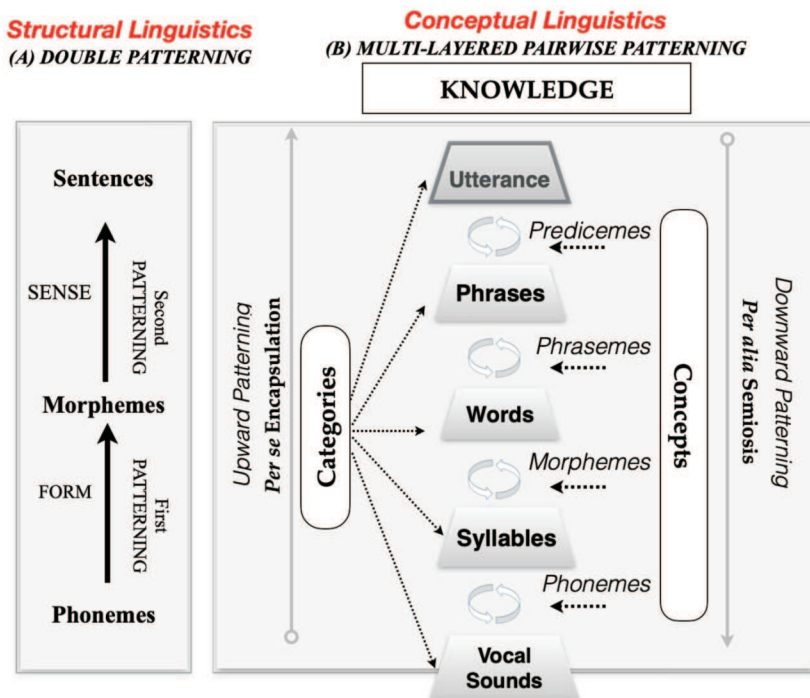


Figure 4. Essential Layers of Utterance Construction Architecture: (A) Double Patterning theory vs. (B) Multi-Layered Pairwise Patterning theory

As a matter of fact, in order to process a chain of sounds as a syllable, a chain of syllables as a word, a chain of words as a phrase and a chain of phrases as an utterance (information) speakers need to use their knowledge of the given language consisting in layers of *phonemes*, *morphemes*, *phrasemes* and *predicemes* (see Fig. 4). Indeed, only competent speakers of the language in which an expression is uttered have at their disposal concepts which are necessary for encoding/decoding units (expressions) of each successive layer of utterances in a given language. Persons who do not know the language in which an utterance is being uttered perceive only chains of physical voice sounds without any meaning.

From a more general perspective, let me just mention that besides the main functions of the human brain there are a large number of intellectual faculties (functions) which enable the speakers/hearers to encode/decode meaning (structured information) into/from linguistic units such as *sounds*, *syllables*, *words*, *phrase* and *utterances*. These units are due to three kinds of processing moods. They are: (a) products of grammar rules, (b) results of realisations of schemes (such as verb and/or noun va-

lences) which are *ready-made* (in the widest sense, ‘idiomatic’) expressions (as alternative to products of grammar rules) and (c) optionally, results of recombination rules as applied, first of all, to schemes but also to the products of grammar rules.

6.2. Concepts and categories of utterance layers

At this point, it seems necessary to add a few precisions about the need to distinguish thoroughly between observable linguistic categories and their underlying concepts. In the history of language science, the problems covered by the “category/concept” couple of notions was explicitly introduced first in the domain of phonetics, giving rise to the notion of *phoneme*¹⁶ as it is now used in linguistics. Phonemes are defined in the present framework as families of formal concepts, their intents consist of features which determine mostly the distinctivity of voice sounds within the given system. At first sight, this idea may seem misleading ... especially for linguists and other specialists working on language problems outside the data mining community for whom the wording *phonemes are not supposed to be concepts* would certainly be more than evident.¹⁷ The discrimination of phonemes is possible solely when one takes into consideration the system of “phonological oppositions” which can be represented by datasets to which the dual function (Galois connection, see the schema 1 above) can be applied in order to find out the formal concepts consisting of *extents* (objects i.e. sound traces) and *intents* (the latter being known in linguistics as “distinctive features” of phonemes). Nevertheless, from the viewpoint of data science, given that every object has properties, specialists assume that it is always possible to build datasets using features. Phonemes are concepts which cannot be directly observed, what speakers and hearers experience are different realisations of phonemes as **sounds** in different speech acts and different languages.

Let me address another important issue of structural linguistics: the definition of the category *word*. Surprisingly, the category of *word* has been definitively abandoned and replaced by that of *morpheme*, although recently some researchers consider that both concepts are necessary for the description of human languages. The main reason which was put forward for removing *word* from the inventory of linguistic notions was that it is difficult (if not quite impossible) to provide a formal (hence universal) definition of what speakers intuitively perceive as words in any language. In my approach, I keep both terms however within different ranges; *words* are treated as categories emerging from the layer of *syllables* thanks to the concept of morpheme, thus they are what is perceived and produced by language users

while morphemes are families of concepts underlying the emergence of words out of syllables.

It is also necessary to mention how the term *phraseme* is used in my model since it is understood differently in other linguistic theories.¹⁸ In this approach, phrasemes are sets of concepts determining the internal structure of phrases. Phrases emerge out of word chains. The features defining the category of noun phrases in a language make it possible to distinguish them from verb phrases whose features are different. The bundles of features characteristic of different types of phrases in a given language constitute their phrasemes. In human languages various phrasemes underlie different types of phrases.

In a similar way, *predicemes*¹⁹ are families of concepts underlying the emergence of utterances of a given language (regardless of whether these utterances are complete or not). Importantly, *predicemes* provide closure frames for predications, thus making it possible for utterances to emerge out of phrases (cf. Włodarczyk and Włodarczyk 2019a). The basic schema of predicemes is the subject-predicate structure which may be expressed in different ways in different languages, e.g. word order (SOV and SVO being the most frequent orders in all the languages of the world); *gender* and *number* agreement between subject and verb is a mark of the predicative structure in many European languages, etc. At this point, it is important to emphasise that one of the most innovative contributions of conceptual linguistics consists in highlighting the pragmatic nature of predication in human languages, understood as meta-information (cf. Włodarczyk and Włodarczyk 2013, 2016b, 2019a). In linguistic utterances, meta-information itself combines with ortho- and para-information. Indeed, all the complexity of linguistic utterances aims at making them the most efficient vehicles of information.

7. Conclusion

From a philosophical point of view, it is interesting to note that the computer tools for data mining created to advance research on all sorts of objects, not just languages, have been used here to represent the supposed functionality of linguistic systems. This is not a paradox, however, since the formalisms (e.g. logical notation conventions or Programming in Logic language Prolog) that in science underlie the symbolic representations are to a great extent inspired by the study of natural languages. In short, the functionality of tools is borrowed for elucidating the functionality of objects.

To conclude, let me outline the most salient theses of the proposed tetradic conceptual theory of semiotics:

- Mental systems are layered information processing devices forming a stratified configuration of spaces.
- Within mental systems, in the same way as ordinary objects, signs as semiotic objects are code units (I adopt the idea of ‘mental code’ in spite of the fact that its real existence is still to be proved by neuroscientists).
- Code units are components of conceptual systems; hence they have characteristics of morphisms.
- Within mental systems, objects within ordinary code units are traces of ordinary things, while objects within semiotic code units are traces of semiotic things.
- Code units representing ordinary objects preserve traces of objects of the world “out there”, however code units representing semiotic objects, in addition to features their traces also preserve, have their “proper semantic” features *per se*.
- Signs – like all other objects – are physical things of various kinds: articulatory, acoustic, graphical, material etc. As such, they are objects *per se*, but – in addition – unlike the ordinary (non-semiotic) objects, signs are also para-objects: they are objects *per alia*.
- Signs should be treated as categories within a layered system of information processing. In structural (Saussurean) linguistic terminology, objects of such categories are seen as ‘signifiers’ and all their features correspond to the ‘signified (face)’ of signs.
- Ultimately, the meaning of signs (or semantics) is obtained in the process of integration of information (most importantly assemblages of ortho-, meta- and para-information).

The above proposal opens a new perspective on the hierarchy of successive layers of semiotic objects which are building blocks of information as conveyed by language utterances. I propose to replace the explanation of the so called “double patterning” problem of classical linguistics, which draws the dividing line between “distinction” and “meaning” of signs, by that of *multi-layered pairwise patterning* reflecting the hypothesis according to which every linguistic unit at each successive layer building up an utterance can be at the same time **both** distinctive **and** meaningful, though to varying degrees.

Clearly, in addition to their quality as objects *per se* characterised by their own ‘material’ (physical) ontology augmented by their own semantic features, signs are objects *per alia* which are aligned with ordinary objects. For this reason, they draw their features from ordinary ontologies.

Due to their social nature, signs are more conservative than ordinary categories. Indeed, languages undergo fewer changes than cognition within the same period of time. From the point of view of their systemic nature, it is important to keep in mind that language ontologies are not the same in different languages, especially so when these languages belong to different families and cultural areas. For this reason, translating from one language to another is a hard task (theoretically not always feasible), especially as regards literary works of art which rely heavily on the literal sense of expressions.

Linguistic signs are used as *data* that need to be recognised and processed by cognitive systems as both concepts and categories. The undergoing operations should include, amongst others, **emergence** (due to the synergy of encapsulation and semiosis) of concepts themselves and **optimisation** (approximation and disambiguation) of underlying concept opposition systems. These kinds of operation should happen while passing through different layers of a *hierarchy of sign or sign combinations* in which the highest layer constitutes bunches of information giving rise, in turn, to the acquisition of knowledge. Human languages, as *semiotic objects* (i.e. seen as objects and para-objects), are proper to humans and have a social character, i.e. they rely on an inter-subjective consent between their bearers. Languages enable humans to fuse their reciprocal knowledge during instances of communication (Stacewicz and Włodarczyk 2020). This remark suggests the hypothesis about the existence of a sort of elementary presemiotic stage of communicative ability within the brains of certain species of animals but, as for now, more research needs to be done to acquire constructive knowledge on this problem.

What is specific about human languages is that language units or expressions (categories as defined here) are (1) *conceptual*, their underlying concepts being drawn from the ontological and epistemic references of noemata, (2) *self-consistent* though (3) *not self-sufficient* (they play their role exclusively when aligned with noemata). In our model, the structural linguistic definition of signs as pairs of ‘signifiers’ and ‘signified’ is enhanced by the adoption of a special kind of category (a monoid category). Although semiotic objects and ordinary objects are both represented by their traces within mental systems, traces of signs are not simply classic ‘signifiers’ (they are families of dual pairs of extent and intent) and the ‘signified’ face of signs cannot be simply elucidated by the classic notion of “concepts” (they are equally families of pairs of extent and intent, too).

The presented conceptual approach to semiotics fills the gap between the theory of signs and that of linguistic messages (sentences, utterances),

providing a new research paradigm for specialists of many disciplines (linguists, philosophers, logicians, neuroscientists, psychologists and computer scientists) about any semiotic system in an integrative manner.

NOTES

* I would like to thank Reviewers and Editors for all valuable comments and suggestions which helped me to improve the quality of the manuscript.

¹ Taking the *referent* for an ingredient of the theory, Saussure's approach is, in fact, rather triadic, too. The difference lies in that while the Saussurean pair "*signifier/signified*" is located in memory, the Peircean pair *representament/referent* seems to be situated in the real world, i.e. it is outside the memory.

² More generally, living beings or, as Henryk Greniewski [1903–1972] put it, "*organisms possessing central nervous systems endowed with the capacity to process information*" (Engl. trans. AW, Greniewski 1968: 30).

³ Computer scientists proposed multiple (about 20) formal approaches to grammars of natural languages. One of them, the Lexical-Functional Grammar (LFG), has – however without much success – laid at its basis some principles of the generative grammar elaborated by Noam Chomsky.

⁴ We called "interactive linguistics" (Włodarczyk 2007, 2015; Włodarczyk and Włodarczyk 2019b) the methodological computer-aided approach of the research area on human languages which is deeply rooted in the forthcoming *data science*.

⁵ Their nomenclature is sometimes rather strange, e.g., the not cited above "institutions" (Goguen and Burstall 1984; Goguen 1991).

⁶ Note however that this claim does not preclude the possibility of modelling concepts using quite different definitions which would be more suitable for the purposes of other applications.

⁷ Incidentally, it is worth mentioning that for reasoning, *agents* and *figures* have a double "embodiment", one is (a) *objective* (existing "out there"), the other one is (b) *subjective* (represented "in the mind"). In fact, an *inter-subjective* nature should also be envisaged because while, in the case of (b), representations located in the mind of individuals (with *traces* of objects), the inter-subjective nature of representations is imaginary and is the result of all kinds of ontology unification-like processes which occur in the mind of communicating agents.

⁸ As an example of such an opposition system, I built data tables of personal pronouns in English using FCA and decision logic and the result was a lattice whose upper nodes are attributes (*speaker, masculine, feminine, human*) and lower nodes objects (*I, you, he, she, it*). Note that the resulting latticework suggests a strong structural analogy with neural networks.

⁹ Note however that *ordinary* objects may perform two functions: they form *ordinary* categories (with objects on their own behalf, *per se*), but they may also be incidentally used as if they were objects of *semiotic* categories (as objects *per alia*), e.g., a fan on the door sometimes used to indicate the ladies' restroom in Japanese restaurants.

¹⁰ Note that this suggests that there are two kinds of operational ("working") memory ("short-term working memory" and "activated portion of long-term memory which is subject to undergo some alterations") at work during the processes of utterance analysis/synthesis.

¹¹ Neither the neural nature of *semions* nor that of *noemata* are known, as yet. However, to hypothesise about their supposed analogy might be arbitrary because of at least the two following reasons: (1) traces of signs have a double functionality (*per se* and *per alia*) and (2) they are combining in one dimension while traces of objects can develop their representations in the 3D space.

¹² Usually, two types of emergence are distinguished: weak and strong. Weak emergence occurs when objects of higher layers arise as a result of applying rules, while strong emergence arises due to additional factors coming from the environment (context) in which the system operates. Additional factors may include information from the environment coming through the senses (traditionally sight, hearing, taste, smell, touch).

¹³ By partial view, I mean any part of an object representation, not merely partially ordered sets.

¹⁴ In our previous research, we showed that meta-informative markers which point to *aboutness* (choice of the subject and direct object as global and local centres of attention, partition of the utterance into subject and predicate, selection of some temporal parts of situations as opposed to their whole development) are not only lexical (e.g. *to say*, *as concerns* etc.) but also grammatical (e.g. aspect of verbs etc.), cf. Włodarczyk and Włodarczyk 2013.

¹⁵ Marek Tokarz provided a modern pragmatic interpretation of this traditional linguistic problem (Tokarz, 2006, section “Znaczenie *dosłowne i niedosłowne*” [*Literal and Non-literal Meaning*]).

¹⁶ At the beginning of the 20. century, Nicolas Trubetzkoy (1939) introduced the notion of *phoneme*. This fact can be seen as a decisive step into the scientific approach in linguistics. Although phonetic sounds have an audio-articulatory nature (“language forms”), they are processed in the mind not only as “traces” of (physical) objects but also as ones of fully meaningful concepts.

¹⁷ To the best of my knowledge, it is Velina Slavova and her research team (at New Bulgarian University – Sofia) who provided the most reliable statistical results on the problems of sound symbolism and iconicity (2020).

¹⁸ Currently, most linguists understand phrasemes as collocations or idioms (as opposed to “free phrases”). In my approach all the combinations of the Word Layer which are NOT due to the weak emergence underlie the construction of coined phrases. The most constrained phrasemes are said to be coined, while some collocations allow for the free choice of at least one component.

¹⁹ As far as I know, the term *prediceme* was coined by Leon Zawadowski (1966).

REFERENCES

- Barwise, J. and Seligman, J. (1997). *Information Flow: The Logic of Distributed Systems*. Cambridge: Cambridge University Press.
- Burgin, M. (2010). *Theory of Information: Fundamentality, Diversity and Unification*. Singapore: World Scientific.
- Ganter, B. and Wille, R. (1999). *Formal Concept Analysis. Mathematical Foundations*. Berlin–Heidelberg–New York: Springer.

- Goguen, J. and Burstall, R. (1984). "Introducing institutions". In: Clarke, E. and Kozen, D. (eds), *Logics of Programs: Proceedings 1983. (Workshop Carnegie Mellon University Pittsburgh, June 6–8 1983. Lecture Notes in Computer Science..* vol. 164). Berlin/Heidelberg/New York/Tokyo: Springer, 221–256.
- Goguen, J. (1991). "A categorical manifesto". *Mathematical Structures in Computer Science*, 1(1), 49–67.
- Greniewski, H. (1969). "Język nauki" (The language of science). *Zagadnienia Naukoznawstwa (Problems of the Science of Science)*, 4, 1(13), 24–66.
- Hampton, J.A. and Dubois, D. (1993). "Psychological models of concepts". In: Van Mechelen, I. et al. (eds.), *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. London: Academic Press, 11–34.
- Kuznetsov, S.O. and Poelmans, J. (2013). "Knowledge representation and processing with formal concept analysis", *WIREs Data Mining Knowledge Discovery* 3, 200–215.
- Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.
- Lamb, S. M. (1999). *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam: John Benjamins.
- Lapis, W. (2014), *Semantyczne i syntaktyczne aspekty dystynktywności. (Semantic and Syntactic Aspects of Distinctivity)*. Dąbrowka: Wydawnictwo RYS.
- Nowakowska, M. (1980). "Semiotic systems and knowledge representation". *International Journal of Man-Machine Studies*, 13, 223–257.
- Pawlak, Z. (1982). "Rough sets". *International Journal of Computer & Information Sciences*, 11(5), 341–356.
- Pawlak, Z. (1987). "O analizie pojęć" ("About the analysis of concepts"). In: Bogusławski, A. and Bojar, B. (eds.), *Od kodu do kodu (From Code to Code)*, Warszawa: UW, 249–52.
- Priss, U. (2002). "Associative and formal concepts, A classification of associative and formal concepts". In: Priss, U., Corbett, D. and Angelova, G. (eds.), *Conceptual Structures, Integration and Interface. International Conference on Conceptual Structure 2002*. Berlin/Heidelberg: Springer, 354–368.
- Priss, U. (2020). "A preliminary semiotic-conceptual analysis of a learning management system". *Procedia Computer Science*, special issue. Cristani, M., Toro, C., Zanni-Merk, C., Howlett, R.J., and Lakhmi, C.J. (eds.), *Knowledge-Based and Intelligent Information & Engineering Systems, Proceedings of the 24th International Conference KES 2020*, vol. 176, Elsevier: 3702–3709.
- Richards, I.A. and Odgen Ch.K. (1923). *The Meaning of Meaning*. London: Kegan Paul.
- Slavova, V. (2020). "Emotional valence coded in the phonemic content – Statistical evidence based on corpus analysis". *Cybernetics and Information Technologies*, 20(2), 3–21. doi: 10.2478/cait-2020-0012

- Stacewicz, P. and Włodarczyk, A. (2010). “Modeling in the context of computer science – a methodological approach”, *Studies in Logic, Grammar and Rhetoric*. Special issue: Świączkowska, H. (ed.), *Philosophical Trends in the 17th Century from the Modern Perspective*, 20(33), 155–179.
- Stacewicz, P. and Włodarczyk, A. (2011). “O modelowaniu informatycznym ze szczególnym odniesieniem do badań nad sztuczną inteligencją”. *Zagadnienia Naukoznawstwa (Problems of the Science of Science)*, 4(190), 165–184.
- Stacewicz, P. and Włodarczyk, A. (2020). “To know we need to share – Information in the context of interactive acquisition of knowledge”. *Procedia Computer Science*, special issue. Cristani, M., Toro, C., Zanni-Merk, C., Howlett, R.J., and Lakhmi, C.J. (eds.), *Knowledge-Based and Intelligent Information & Engineering Systems, Proceedings of the 24th International Conference KES 2020*, vol. 176, 3810–3819.
- Tokarz, M. (2006). *Argumentacja. Perswazja. Manipulacja (Argumentation. Persuasion. Manipulation)*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Trubetzkoy, N.S. (1939). *Grundzüge der Phonologie*. Prague (*Travaux du Cercle Linguistique de Prague*, No 7).
- Wille, R. (1982). “Restructuring lattice theory: An approach based on hierarchies of concepts”. In: Rival, I. (ed.), *Ordered Sets*, Dordrecht/Boston: Reidel, 445–470.
- Wille, R. (2005). “Formal concept analysis as mathematical theory of concepts and concept hierarchies”. In: Ganter, B., Stamme, G., Wille, R. (eds.), *Formal Concept Analysis*. Berlin/ Heidelberg: Springer, 1–33.
- Włodarczyk, A. (2007). ハリ・ソルホンヌ大学理論・応用言語学研究所 (CELTA) — CASK (Computer-aided Acquisition of Semantic Knowledge) プロジェクト — (paper in Japanese). *Japanese Linguistics*, 21 (English version: <http://celta.paris-sorbonne.fr/anasem/papers/>).
- Włodarczyk, A. (2013). “Towards (Re)construction of the Theory of Linguistic Oppositions (within the Framework of Interactive Linguistics)”, *Workshop for FCA Tools and Applications* (at ICFCA’2013).
- Włodarczyk, A. (2015). “Informatyka szansą na rozwój naukowej lingwistyki” (Computer science as an opportunity for the development of scientific linguistics). In: Stacewicz, P. (ed.), *Od informatyki i jej zastosowań do światopoglądu informatycznego*. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 117–132.
- Włodarczyk, A. (2017). “Concepts and sign in the light of information systems” (A brief survey of an invited lecture). blog *Filozofia w informatyce*, 45. Spotkanie. Kraków. <https://filozofiainformatyki.wordpress.com>
- Włodarczyk, A. and Włodarczyk, H (2013). *Meta-Informative Centering in Utterances: Between Semantics and pragmatics*. Companion Series in Linguistics., Amsterdam: John Benjamins.

- Włodarczyk, A. and Włodarczyk, H. (2016a). “Trójwarstwowa struktura informacji w treści wypowiedzi (szkic o programie Gramatyki rozproszonej)”. *Investigationes Linguisticae*, 34, 73–112.
- Włodarczyk, A. and Włodarczyk, H. (2016b). “O pragmatycznej naturze predykcji (czyli o meta-informacji w orzekaniu językowym)”. *Poradnik Językowy*, 8, 7–21.
- Włodarczyk, A. and Włodarczyk, H. (2019a). “Qu’est-ce au juste que la prédication?” *Bulletin de la Société de Linguistique de Paris*, 64 (1), 1–54.
- Włodarczyk, A. and Włodarczyk, H. (2019b). “The interactive method for language science and some salient results”. *Zagadnienia Naukoznawstwa (Problems of the Science of Science)*, 3 (221), 73–92.
- Wolff, K.E. (2009). “Relational scaling in relational semantic systems”. In: Rudolph, S. and Dau, F. (eds.), *ICCS '09: Proceedings of the 17th International Conference on Conceptual Structures: Conceptual Structures: Leveraging Semantic Technologies*. Berlin/Heidelberg: Springer, 307–320.
- Zawadowski, L. (1966). *Lingwistyczna teoria języka*. Warszawa: Państwowe Wydawnictwo Naukowe.