Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 3810–3819

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

To Know we Need to Share – Information in the Context of Interactive Acquisition of Knowledge

Paweł Stacewicz[a], André Włodarczyk[b,*]

[a]*Warsaw University of Technology, Dept. of Administration and Social Sciences, Pl. Politechniki 1, 00-661 Warsaw, Poland*
[b]*Sorbonne University, Centre for Theoretical and Applied Linguistics - CELTA, 28 rue Serpente, 75006 Paris, France*

Abstract

Natural language interface requires much more complex processing than is currently assumed. We present a general review of foundational notions of data, information and knowledge, aiming at tentatively sketching out a set of subcomponents of an integrative linguistic theory of man machine interaction, bearing in mind that our model will favor further research on simulation of man man interaction. Our main concern is to show that in order to acquire knowledge humans need to extract it from multiple kinds of information as distributed in the content of natural language utterances.

Keywords: information; knowledge; ortho-information; meta-information; para-information; semantics; pragmatics; ontology

1. Introduction

Contemporary Natural Language Processing (NLP) techniques are mostly based on syntactic parsing which relies on the highly simplified view of interpretation as the interface between morphological form and semantic content. We propose to pay more attention to the fact that, though quite insufficient for representing meaning of language messages, linguistic knowledge is much more autonomous than is usually admitted in language sciences, psychology, logics and computer science (in both 1st and 2nd generation AI). Indeed, the semantic content of utterances is, in most approaches, directly engulfed in cognitive knowledge. We are going to show that producing and understanding information as conveyed by natural language messages requires more complex processing. Assuming that natural language utterances are always partial, we therefore reconsider radically the access to the semantic content of utterances by the distinction sometimes made between the **overt** (or explicit) content, being only a small part of integral information on the basis of which the **covert** (or implicit) content is supposed to be reconstructed. In the Distributed Grammar

* Corresponding author. Tel.: +48 508 968 431.
E-mail address: wlodarczyk.andre@gmail.com

1877-0509
10.1016/j.procs.2020.09.006

(DG) program, it is assumed that the interpretation of linguistic utterances requires taking into account at least three tiers of information (para-, ortho- and meta-information) and that these informational spaces are multiplied in such a way as to provide insight into cognitive **semiotics** as separate (though not independent) from **multi-modal** cognitive semantics.

In section 2, we discuss the most ambiguous today concept of information and describe the processes which lead from data expressed in some language to knowledge through the interpretation of data which transforms them into information that has to be verified in order to become a part of knowledge[1]. In section 3, we present the complexity of information expressed in natural languages. Natural language information consists of three tiers.

2. Between information and knowledge

2.1. Variety of meanings of the term "information"

At the current stage of civilizational development - more and more often referred to as the age of information and/or information society [7] - the term "information" is very common and also ambiguous. It is used primarily in communication contexts (when it comes to transmitting information by means of various types of media) and technical contexts (when it comes to processing information by means of computers and devices linked to them).

Despite the strong dominance of these two contexts, the attempt to define a maximally general concept of information[2] requires the consideration of at least three of its dimensions (in other words: semantic areas), which refer in turn to: (a) the world, (b) cognitive systems (including the human mind), (c) language. Let us explain this in more detail.

- **Ia.** In the physical dimension, information is understood as something in the world, that is, as a certain way of ordering or organizing, the material elements that form specific structures [18]; it is therefore a certain component of being, related to Aristotle's form.
- **Ib.** In the cognitive dimension, information is understood as something within the cognitive system [2], i.e., as cognitive content that can become identical to knowledge if the additional condition of a sufficiently good justification is met (given that the content is a statement of something).
- **Ic.** In the communication dimension, information is meant as something within a language [4], i.e. the content conveyed by means of a sequence of characters that are understandable to the recipient, such as gestures, verbal constructions, mathematical formulas, computer data, etc.

It is worth noting that today there is another way of understanding the term "information" as a specific type of code that can be stored in the memory of a computer, processed according to specific rules, transmitted, used to control external devices etc. [13]. This meaning can be found in phrases such as 'information processing systems' or simply information systems.

In this work, we do not treat the above-mentioned semantic area as an alternative to the ones listed above (Ia-Ic), because we believe that today computing systems already perform certain cognitive functions, such as object recognition, categorization, reasoning or learning [14], and in addition, they do so in a way that is increasingly close to cognitive autonomy (e.g. when they are based on self-organized artificial neural networks). Moreover, the content which is computationally coded, e.g. by means of artificial programming and communication languages, fits well within the dimension Ic; and as far as computational virtual objects are concerned, intertwining more and more closely with the physical world, they should be included within the dimension Ia [3].

[1] Some fragments of this article are based on two works published in Polish: [15] and [16].

[2] In this work we will not deal with the general concept of information, which can (perhaps) be abstracted from the above-mentioned semantic areas but will focus on information understood as cognitive content (Ib). It is worth noting that among the types of meanings which we pointed to, the most controversial is to consider information as something physical that exists in the physical world (Ia). As regards the ambiguity of the term "information" see also [5].

2.2. Information as cognitive content

Among the three aforementioned meanings of the term "information" we will focus here on Ib, i.e. information as cognitive content, which we will abbreviate to cc-information. By referring the above meaning to the cognitive-communicative activity of the human being - which for the time being is a blueprint and a desirable "point of destination" for artificial cognitive systems - we can see that the notion of cc-information combines all the above-mentioned meanings. This statement is justified by the following reconstruction of a characteristic human cognitive process that leads from structures and regularities observed in the world (see Ia) to information communicated, conveyed and shared by a certain community (see Ic). Here are the successive stages of that process [16]:

The human mind:

i) finds in the world (i.e: in the signals/stimuli that reach it) some *regularities* that appeal to it and which most probably are natural,
ii) reflects these regularities in a certain *internal code* (in the brain appearing as a neuronal record), combining with this code a certain meaning,
iii) can make of this code, once interpreted, a meaningful element and/or base of its own *knowledge*,
iv) can transcribe the interpreted and meaningful code (sometimes having the status of knowledge) into a certain intersubjectively accessible language and *communicate* it in this form to other cognitive subjects in order to *interact* with them effectively.

If we focus on the above-mentioned processes - especially points (ii) and (iii), which combine the concepts of cc-information with knowledge, we can refer to a scheme known from the theory of knowledge as the *DIK*[3] *Pyramid*. To begin with, we will briefly discuss a certain typical interpretation of it [1]).

The pyramid consists of three levels, the layout of which corresponds to the following observation: the information processed by the cognitive system is, on the one hand, encoded in a specific way and, on the other hand, can become the basis or element of knowledge of the system.

According to this observation, the lowest level of the pyramid is filled with data, i.e. certain uninterpreted chains of symbols of a specific code (e.g. language); above them is information, i.e. interpreted (meaningful) data, inter alia, in terms of the goals defined or achieved by the system; the level of knowledge, in turn, is built up on top of information - i.e. all the information verified by the system, and thus sufficiently well justified for it. The notion of justification is naturally associated with the activity of the human mind, which consciously refers to facts and/or inferences "embedding" new information in the system of accumulated knowledge. Nothing however stands in the way of using this concept - just like the concept of c-information - in isolation from the conscious activity of the subject, referring to purely formal relationships between elements of knowledge (and information aimed at being integrated into knowledge). With this approach, the automaton can also generate and communicate justifications. In addition, if the automaton is equipped with receptors and effectors that give it contact with physical reality, it can generate justifications based on observation of that reality.

Therefore, the pyramid comprises three levels - I. data, II. information and III. knowledge - which are grounded in hard facts, i.e. what is happening in the world around the pyramid[4]. To enter level I. the ability to recognize purely physical regularities in the world (which are reflected in data) is needed; to move from level I. to level II. the interpreting activity (of data) is necessary; and from level II. to level III. - verification or justification of the veracity of the information obtained.

In the sequel, we will sketch out the main points of the new model of the structure of linguistic utterances as elaborated within the framework of the Distributed Grammar program following which the explanation of utterances

[3] Data-Information-Knowledge.

[4] Interestingly, all the three key concepts of the DIK pyramid are called, in certain contexts, information. Data is often viewed as basic information, being the basis for further analyses and conclusions (e.g. data as a mathematical problem); interpreted data, called here simply information - is information in a semantic and pragmatic sense (having a certain meaning and somehow used by a processing system, including the human mind); and knowledge is understood as information in a strong sense (e.g.: scientific publications are both a record and a source of knowledge).

shows how the uninterpreted sign systems (standing for something else) differ, in the context of the communicating agents cognitive activity, from the meaningful and verified components of knowledge.

3. Information in natural language

The view of information in natural language that we present hereafter is based on our research within the Distributed Grammar program (DG)[5] which led us to conceive a meta-grammar of natural languages seen as interfaces rather than as operational codes, the productions of which are arbitrary ad hoc projections into (possibly multi-modal) representations of knowledge. Distributed Grammar [23, 25] is an integrative theoretical framework aiming at accounting for the relatively complex structure of simple utterances which is usually explained using terms such as topic, focus, subject, argument, predicate in separate theoretical proposals.

Information is often seen as a part of knowledge. As such it is always presented with respect to language: "*analyzing the concept of information, it is natural to come to the conclusion that information has three major aspects or issues: Syntax, Semantics and Pragmatics*" [1]. However, we point out that the above triad combines syntax with semantics and pragmatics in spite of the fact that semantics and pragmatics are information, but syntax is not. In our view, ontology is closer to semantics and pragmatics than syntax which, in fact, applies to information belonging to all of these domains: ontology, semantics and pragmatics. We consider that information selected from each of the above 3 cognitive domains characterises respectively the three tiers of the conceptual structure of utterance content we present here: **para-information, ortho-information and meta-information (POM)**.

In our linguistic theory, the material form (sounds) of even very simple natural language utterances (in addition to their referral to prosody, phonology, morphology and syntax) expresses content consisting of at least three kinds of information: (1) the ortho-informative ("properly" semantic) signification simultaneously with (2) the para-informative (ontological) identification and (3) the meta-informative (pragmatic) process of predication (in a strictly linguistic sense)[6]. Moreover, research on several natural languages, showed that what is immediately conceived and perceived during the synthesis and analysis of the most elementary language messages (utterances), is their meta-informative structure even before one gets some access to all other kinds of information. As a matter of fact, ortho- and para-information are less salient and need more processing power for producing and understanding linguistic utterances.

Hereafter is a schematic representation (fig.1) of the three kinds of information and the relations between them as expressed in a simple utterance (1):

(1) Brutus killed Caesar with a poniard.

Information components in the figure above are not all overtly expressed by a segment of the utterance, some of them remain covert. The morpho-syntactic form of the utterance gives the most direct access to the meta-informative content. On this tier, we find the nouns *Brutus* and *Caesar* which, in English, can be recognised – on the basis of their position before or after the verb – respectively as the subject and the (direct) object. The *subject* is that attention-driven phrase (ADP) which refers to the speakers global centre of attention (CA). The verb and its complements are what is being said about the subject, they are the constituent parts of what we call predicate. As a matter of fact, in the DG framework, the predicate reflects the meta-informative clustering of representation [25]. Consequently, predication is not a semantic notion but a pragmatic one, since meta-information depends on the speakers point of view, and more precisely, his own centres of attention presumably shared with the hearer.

In utterance (1), *Brutus* is the subject noun phrase (reduced to only one noun (proper name) without any determiner such as article or adjective), and it occupies the position of the global attention-driven phrase of this utterance pointing at the **main** centre of attention (CA). In other words, this utterance is first of all about *Brutus*. The predicate contains three constituents: a verb (whose content, in fact, as a rule, belongs primarily to ortho-information) and two noun phrases. The first one is the direct object *Caesar* which – from the point of view of the meta-informative structure of utterances – is the local attention-driven phrase pointing at the secondary CA; this means that in this utterance the main concern is *Brutus* and the secondary one is *Caesar*. Within the predicate, the second noun phrase *with a*

[5] This research was developed (2000-2014) at Centre de Linguistique Théorique et Appliquée (CELTA), Sorbonne University, Paris.
[6] N.B. The term predicate is used differently by logicians [25].

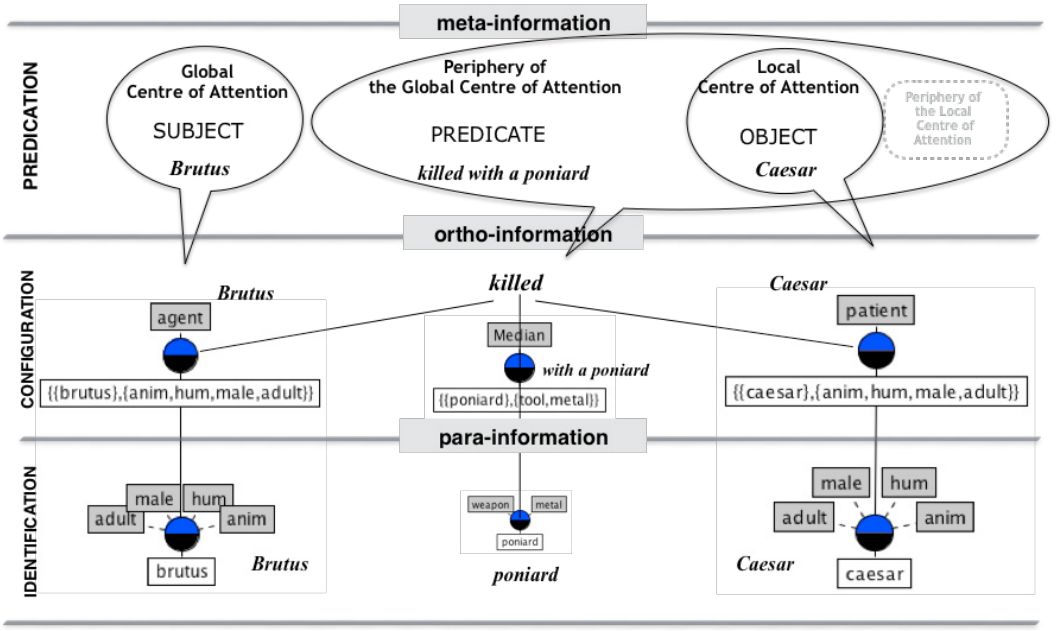

Fig. 1: Three information tiers and their mutual dependencies as expressed by utterance (1).

poniard, is the so called "indirect object". However, these kinds of phrase can hardly be considered as the "third" attention-driven phrase (ADP) for the following reasons: firstly, the English preposition *with* – as used in this context - expresses the means of an action (in our terminology the "median role"). Hence, it conveys ortho-informative content and, secondly, according to [10], a neurologist of attention, no more than two centres of attention (one global and one local) can be active at once, the third one seems to be impossible.

As concerns the two attention-driven phrases, i.e. the subject and direct object, the semantic role each of them plays in the situation expressed by the verb is not directly marked by its syntactic position. As a matter of fact, *Caesar* could be as well expressed not as the direct object of the verb in active voice, but as the subject of the verb in passive voice. The passive voice would make it possible on the level of meta-information to treat *Caesar* as the main CA and *Brutus* as the secondary one:

(2) Caesar was killed by Brutus.

In utterance (2), the semantic roles of *Caesar* and *Brutus* are the same as in utterance (1). But how do we know what their roles are? Knowledge about these roles is included in the verb valence[7], not in the morphological form of noun phrases nor in the syntactic structure of the utterance. The valence scheme of the verb *kill* contains three slots and English speakers know from the lexical meaning of this particular verb, that when it is in the active voice, its subject refers to the active role (*the killer*) and its object to the passive role (*the killed*), whereas the median role is partly expressed overtly by preposition *with* in the indirect object.

As shown on fig. 1, roles belong to utterance semantics, i.e. to its ortho-informative tier[8]. Native speakers learn the meaning of verbs together with the roles that participants play when taking part in the situation expressed by the verb. The semantic roles of the subject and object of a verb may however be different than that of action verbs like kill, depending on the semantic content of the verb itself and depending on its voice form, e.g. ; the subject of a verb with the meaning of feeling something refers to a passive participant, e.g. (3).

[7] Valence refers to the capacity of a verb to take a specific number and type of arguments (noun phrase positions). SIL Glossary of linguistic terms.

[8] Research on ortho-information was first introduced under the name of associative semantics in [20, 22].

Table 1: Intra-tier operations on information within utterance content.

Sort of information	Intra-tier operation	Result
meta-information	Centering	Distinguished part of information driven by the centres of attention
ortho-information	Configuration	Proper information content: situations & participants
para-information	Identification	Components of semantic situations: entities & relations

(3) Old people suffer from the heat.

The interpretation of the utterance content must also include the *identification* of entities spoken about, which takes place in the para-informative tier. On this level, we find individual concepts, for instance concerning entity classes entailing animate, human and non-human beings or inanimate things, and conceptual schemes expressing relations between entities, i.e. situations and the roles they contain. In utterance (1) *Brutus* and *Caesar* both belong to entities characterised as male human and adult agents, and the individual features of each is rooted in the readers knowledge about antique Rome where *Caesar* was an army general, political leader and writer; Brutus is known as Caesars adoptive son. *The poniard* belongs to inanimate things (figures) made of metal and used as weapons. Thus, in the para-informative tier each of the three noun phrases refers to a well identified entity.

In the para-informative tier, a selection procedure leads to the extraction of individual concepts and relational schemes out of the speakers knowledge. Then, they are embedded into the ortho-informative tier as participants playing roles in a chosen relational scheme, a frame which makes it possible to *configure* the situation spoken about. In each language, the situation frame is expressed by the valence schema of a verb. However, elements which have been *identified* in the para-informative tier and *configurated* in the ortho-informative one belong to a different sort than the components of the meta-informative structure which is closely related to the speech act and particularly to the speakers and hearers knowledge. On the meta-informative level, elements of para- and ortho-information are in turn selected and ordered depending (a) on the speakers centres of attention and also, as far as shared attention is concerned, on the hearers centres of attention, as well as (b) on the specific means the speakers have at their disposal in the language they speak. As will be shown further, these linguistic means consist both of valence schemas and grammatical recombination rules (verb voices, major constituents order etc.). Integration of information belonging to the three different tiers is made possible by the fact that participants and spatio-temporals anchors of the situation are identified as referring to the same entities. Table 1 summarizes the operations characteristic of each information tier.

In order to end successfully with comprehension, the interpretation of an utterance must access the ortho-informative tier of semantic representations. Selecting participants and/or spatio-temporal locations together with their respective views produces para-information. Ortho-information is generated when speakers make connections between entities in their mental representation space (configuration of situations). Meta-information (predication) takes place when speakers express their mental representations in linguistic utterances within which some fragments of ortho-information are distinguished thanks to attention centering. Recall however that information is only partially expressed as a result of application of the base schemes (valences) whose instantiations undergo various transformations which are subsequently mixed up and multiplexed in order to be exteriorised in the form of utterances.

3.1. Meta-information

The meta-information tier is the one of attentional partitioning of ortho-information into 1st-degree meta-information and 2nd-degree meta-information for the purpose of discriminating central and peripheral parts of it. The first level of meta-information concerns that of *base utterances* whose centres of attention are expressed by subject (global) and object (local). The second one concerns *extended utterances* in which the global ADP is the *topic* and the local one the *focus*. The most important property of topic and focus is their respectively *old* and *new* status of communicated information which is in contrast with the rest of the utterance. As a matter of fact, although this might seem at first glance contradictory with the definition of "information", communication in natural languages is based

on the alternation of *old* and *new*[9]. Moreover, the *old* or new status is not an inherent property of ortho-information but is based on the way in which it is communicated. The speaker is free to introduce some chunk of information either with a *new* or *old meta-informative status*: this is a major argumentative device in the strategy chosen to enrich or manipulate the hearers knowledge.

In the DG framework, base and extended utterances are defined as **pragmatic** units of discourse in contrast to simple and complex sentences understood as **syntactic** units [24]. As a pragmatic unit, each utterance contains at least one attention-driven phrase (ADP) referring to a centre of attention (CA). The ADP may have either the same or a different meta-informative status (*old* or *new*) than the rest of the utterance (the statement). In a base utterance there is no contrast between the status of the global ADP and that of the statement: it is either "all new" or "all old". On the other hand, the ADPs of extended utterances contrast with the statement. The topic bearing an old meta-informative status is in contrast with the new *comment*, the focus of new meta-informative status is in contrast with the old *background*. When a constituent of a base utterance (be it an ADP or another NP) undergoes topicalization or focalisation it is treated as an *extension* of the base utterance which, in many languages, is moved to the left or right periphery of the base utterance[10]. In so-called *free word order* languages (e.g. Polish, Russian), topic and focus are superimposed on other NPs mostly by means of prosody and word order.

Concepts coined for the description of meta-information have been elaborated using algorithms from Formal Concept Analysis – FCA [19] and Rough Set Theory – RST [11, 12]) as implemented in the Semana platform at CELTA[11]. It is worth noting that only two features (the old or new statuses, and the global or local property of attention-driven phrases) suffice to show the variety of attention-driven phrases in the form of a lattice (fig. 2).

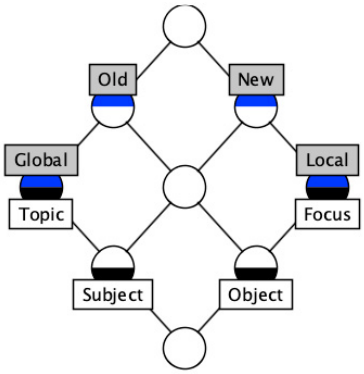

Fig. 2: Lattice revealing the interdependencies of attention-driven phrases as used in language utterances.

Base utterances are generated using valence schemes. In order to change the meta-informative structure of an utterance, languages make use of *recombination* rules some operating within base utterances and the others making it possible to build extended utterances. Recombination rules may be either lexical or grammatical. Lexical means of recombination are pairs of related verbal schemes, such as *buy/sell*, *give/receive*, thanks to which the same participants in a situation are introduced as global or local centres of attention. Changing the verb makes it possible to change the subject within a base utterance. Verb voice (active/passive) belongs to the meta-informative grammatical means used in recombination, cf. utterances (1) and (2). With the passive voice, the distinction of salience (global/local) is changed without changing the ortho-informative content of the utterance. The recombination of ADP by topicalization or focalisation makes it possible to build extended utterances: the most universal means are prosody and word order;

[9] "Connected speech unfolds as an unbroken sequence of "messages", in which the speaker is alternating between elements of given and elements of new; these map into the structures of the other grammatical units, most powerfully into those of the clause." ([6], p. 42).

[10] In German, the topic and focus are moved outside the limits of the syntactic structure of base utterances, to the first and to the last position in the utterance. In English, most often, topic is moved to the left periphery of the utterance whereas focus is mostly marked by intonation.

[11] See Interactive Linguistics at Centre for Theoretical and Applied Linguistics (Centre de Linguistique Théorique et Appliquée - CELTA), http://celta.paris-sorbonne.fr/anasem/indexIL.html

particles; syntactic constructions (such as cleft sentences for focalisation) are also commonly used [23]. Hereafter are two examples of the possible ways of topicalizing or focalizing one of the constituents of utterances (1) and (2).

(4) It is Brutus who killed Caesar with a poniard.
(*Brutus* is focalized by the cleft sentence form: *it is who*)

(5) As concerns Caesar, he was killed by Brutus with a poniard.
(*Caesar* is topicalized by the expression *as concerns*)

Moreover, in the DG framework, both the so called *rigid word order* (like in English) and *free word order* languages (like in Polish) can be dealt with in a consistent way: in both types of languages, the typical word order or, more precisely, major phrase order is a meta-informative marker used to introduce ADP and to distinguish between old and new information. In the first type of languages, word-order is concerned mainly with the first meta-informative level (subject and object of base utterances) and in the second type with the second meta-informative level (topic and focus of extended utterances). Yet, in no language does word-order serve only in base or extended utterances exclusively. Moreover, it must be emphasized that besides word-order, intonation is the most universal marker of meta-information.

3.2. Ortho-information

The ortho-information tier is determined by an encapsulation/activation function within a certain information frame (a configured network of information) concerning basically the roles of participants of the spatio-temporally anchored situation in question.

In the ortho-informative tier, information about roles played by participants and spatio-temporal anchors is configurated into situation frames. Conceptual schemes are mental representation frames through which it is possible to pick up and identify situations of the outer world. In the process of linguistic communication, conceptual schemes are matched with valence schemes, sort of ready-made linguistic frames making it possible to express oneself very briefly. In human languages, there is no direct access to ortho-information in an utterance: only literal meaning is explicit whereas cognitive content is barely alluded to and has to be reconstructed in the process of semantic interpretation. The semantic content of linguistic categories is distinct from their cognitive content; for instance, *gender* is not *sex*, nor *tense* in grammar is *time* in human experience, nor grammatical *number* matches directly the cognitive notion of *quantity*. Within situation frames, participants (expressed by noun phrases in the syntactic positions of subject, direct object and indirect object) are cast in active, passive and median **actancy**. But it has to be emphasized that in DG, the term *actancy* refers to a linguistic category which is kept apart from the non-semiotic cognitive notion of **agentivity**. Linguistic actants are characteristic of the language semantics but do not match directly cognitive roles of agents (animate entities) and figures (inanimate) in the conceptual representation of agentivity. Moreover, in DG, agentivity is not restricted to agents playing active roles as it is the case in contemporary linguistic theories[12], it concerns all sorts of roles played by participants in a situation, i.e. not only active but also passive and median roles. Actancy roles are explicit while agency roles have to be interpreted using world knowledge shared by the speaker and the hearer as shown in example (6).

(6) The door creaked.

In utterance (6), the subject *door* is an active pseudo-role; as a matter of fact, when we think it over, it is clear that the inanimate figure *door* is not an active participant of the situation ; we know from our experience that doors have to be moved by an active participant (be it an agent, e.g. a person, or a figure, e.g. the wind) in order to creak.

Actancy is a very abstract concept. For this reason, it generates a certain degree of ambiguity. On the contrary, **agentivity** is more concrete and precise. In human languages, actant roles are attributed to participants according to valence schemes. The core of syntax is made of valence schemes which speakers use “en bloc” and learn as fixed[13]

[12] Agent is the semantic role of a person or thing who is the doer of an event. [...] A prototypical agent is conscious, acts with volition (on purpose), and performs an action that has a physical, visible effect. SIL Glossary of linguistic terms, https://glossary.sil.org/term/agent-semantic-role (19.03.2020)

[13] In most cases, valence schemes have been generated by ancient grammar rules which are no longer in use.

syntactic structures. As a matter of fact, the knowledge of a language provides speakers with valence schemes (actancy) that make it possible to bring together information from the three layers (albeit partially) in a single expression. In 20th century research on natural language syntax, it has been shown that only a small part of linguistic utterances can be fully generated by inference rules while the biggest part are fixed non-compositional constructions. Linguistic utterances are built both with syntactic rules and valence schemes.

The main components of the ortho-informative tier are situation frames, actant roles and spatio-temporal anchors referring to the time and place in which a situation takes place (more on ortho-information in [20, 21, 22, 23]).

3.3. Para-information

The para-information tier is the space of overtly expressed concepts which are established with regard to other concepts such as alike (similar or opposite), but not expressed in the utterance.

Although identification is always present in the para-informative layer, it is rarely expressed otherwise than by naming an individual or a relationship. Morphemes such as "just", "also", "even" and "only, solely, merely" are used as *relative* identifiers which may be explicitly expressed in natural language utterances (Figure 3). Para-information concerns therefore the identification of concepts (among them those of entities and situations) when they emerge in the speakers mind not only in *reflexive* terms (in relation to themselves) but also in a *relative* way as a kind of 'alikeness' (as compared to other members of the same group of beings, class of figures, location in space and time etc.). Thus, at this level, in order to communicate information, speakers transmit what they know about both individual as well as relational cognitive and ontological concepts.

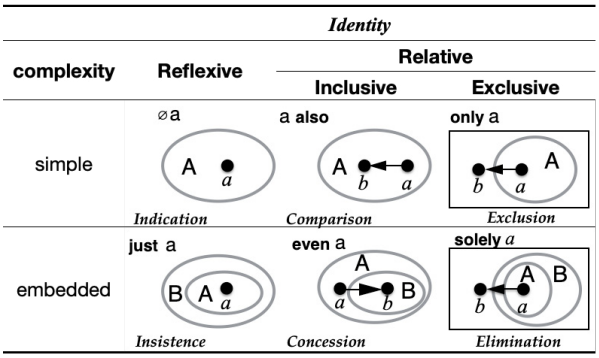

Fig. 3: Para-information concerns identity and alikeness

Here again, the selection of these concepts is guided by the verb valence patterns, but it is carried out first in mental representations before it is expressed in an utterance. It should be noted that adverbs, particles and deictic pronouns constitute the language expression of this identifying selection in the broad sense in which equivalence and complementarity also participate (as sufficient criteria of the distinctiveness of the semantic features). It seems interesting to note that logicians who choose to work on logical reasoning with identity give the identity relationship a separate status among all other relationships.

4. Conclusion

Modelling information as transmitted by natural languages, one needs to consider that: (a) knowledge components and communicated information are distributed in multi-stratified modules of the human mind, (b) information build up out of formal concepts (defined as duality between sets of objects and sets of features) represents the content of language utterances, (c) information contained in language utterances is stratified and (d) natural languages are interfaces but, due to the numerous changes during the incredibly long evolution, they have their "purely" linguistic semantics which is partial, and therefore needs developments in the "core" multi-modal knowledge.

For building artificial systems which would be supposed not only to cooperate efficiently in a certain environment but also to communicate with humans, it is necessary to take into account how humans actually obtain and convey

information (cognitive content) using linguistic expressions. In this paper, we have presented a semiotic conceptual grid covering three types of information as encoded in natural language utterances. Being developed in depth in more advanced studies, the proposed solutions may reveal to be an important contribution of non-IT researchers to purely technical investigations on information systems with natural language proficiency.

Taking into account any cognitive system, including artificial ones, various types of interactions can be considered, not limited to the exchange of information between human minds. Equally important is human interaction with the computing systems they, in fact, actually create. Without it, for example, it is difficult to imagine practicing modern science (we have in mind here improving and automating calculations, using digital models and simulations, and - at least partly - making and verifying hypotheses. It is also worth considering computing systems, e.g. multi-agent systems, which either operate on the basis of internal interaction between their modules or in interaction with other similar systems [8].

References

[1] Burgin, Mark (2010) "Theory of Information: Fundamentality, Diversity and Unification", World Scientific: Singapore, 672 p.

[2] Casagrande, David (1999) "Information as Verb: re conceptualizing Information for Cognitive and Ecological Models", *Georgia Journal of Ecological Anthropology* **3**:4-13.

[3] Chalmers D. J. (2017) "The Virtual and the Real", *Disputatio* **9**(46):309352.

[4] Cherry, C. (1978), "On Human Communication", 3rd ed, The MIT Press: Cambridge.

[5] Floridi L. (2010) "Information. A Very Short Introduction", Oxford University Press, Oxford.

[6] Halliday M.A.K. and Greaves W.S. (2008) "Intonation in the Grammar of English". London: Equinox.

[7] Ito Y. (1991) "Birth of joho shakai and johoka concepts and their diffusion outside Japan", *Keio Communication Review* **13**:3-12.

[8] Maudet N., Chaib-Draa B (2002) "Commitment-based and dialogue-game-based protocols: new trends in agent communication languages", *The Knowledge Engineering Review* **17**(02):157-179.

[9] Mitchell T.M., "Machine Learning", McGraw-Hill, Singapore 1997.

[10] Oberauer, Klaus (2003) "Selective attention to elements in working memory", *Experimental Psychology* **50**:257-269.

[11] Pawlak, Zdzisław (1981) "Information Systems – Theoretical foundations". *Information Systems* **6**(3):205218.

[12] Pawlak, Zdzisław (1991) "Rough Sets. Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishers, Dordrecht: 237

[13] Primiero, Giuseppe (2016) "Information in the Philosophy of Computer Science", Floridi (ed.) The Routledge Encyclopedia on the Philosophy of Information, p. 90-106.

[14] Russel S., Norvig P. (2020) "Artificial Intelligence: A Modern Approach", Pearson, London.

[15] Stacewicz, Paweł (2016) "O zwiazkach informacji ze sfera wartosci ze szczegolnym uwglednieniem idei racjonalnosci (On the relationship between information and the sphere of values with a particular regard to the idea of rationality)", Zubelewicz Jan (ed.) O niektorych wartosciach podstawowych (w kregu filozofii współczesnej)(On some basic values (in the circle of contemporary philosophy)), vol. 3, Wydawnictwo Naukowe Sub Lupa, p. 107-120.

[16] Stacewicz, Paweł (2017) "O redukcji informacji do danych (On reducing information to data)," Stacewicz P. (ed.) Rozne oblicza informacji (Different faces of information), p. 11-21.

[17] Stacewicz, Paweł (2019) "From Computer Science to the Informational Worldview. Philosophical Interpretations of Some Computer Science Concepts", *Foundations of Computing & Decision Sciences*, **44**:27-43.

[18] Stonier, Tom (1990) "Information and the Internal Structure of the Universe". Springer-Verlag, New York.

[19] Wille, Rudolf (1982) "Restructuring Lattice Theory: an Approach based on Hierarchies of Concepts", Rival I. (ed.), Ordered Sets, Reidel, Dordrecht-Boston: 445-470. Reprinted in: Ferré, S., Rudolph, S. (eds.): Formal Concept Analysis. ICFCA 2009. LNAI 5548. Springer, Heidelberg: 314339.

[20] Włodarczyk, André (2003) "Les Cadres des situations sémantiques", Études Cognitives / Studia Kognitywne 5, Warszawa:35-51, English translation: Frames of Semantic Situations, in Meta-Informative Centering in Utterances. John Benjamins Pub. Co., Amsterdam / Philadelphia 2013)

[21] Włodarczyk, André (2008) "Roles and Anchors of Semantic Situations". Études cognitives / Studia kognitywne VIII, SOW, Warszawa, 53-70.

[22] Włodarczyk, André & Włodarczyk, Hélène (2008) "Roles, Anchors and Other Things we Talk About: Associative Semantics and Meta-Informative Centering Theory". Intercultural Pragmatics, Vol. 5. No. 3., "Mouton Series in Pragmatics", Berlin/New York: Mouton de Gruyter, 345-366.

[23] Włodarczyk, André & Włodarczyk, Hélène (2013) "Meta-Informative Centering in Utterances – Between Semantics and Pragmatics", *Companion Series in Linguistics* N143, Amsterdam: John Benjamins, 306 p.

[24] Włodarczyk, André & Włodarczyk, Hélène (2017) "Subjecthood and Topicality are both Pragmatic Issues", Papers on and around the Linguistics of BA, ed. Harada Y., Shudo S., Takekuro M. Institute DECODE Waseda University, Tokyo: 1-10.

[25] Włodarczyk, André & Włodarczyk, Hélène (2019) "Qu'est-ce au juste que la prédication ? (What actually is predication)" Bulletin de la Société de Linguistique de Paris, t. CXIV (2019), fasc. 1, p. 1-54.