Some Notes on Extracting Linguistic Summaries Built with Epistemic Modalities and Natural Language Connectives of Equivalence

R. Katarzyniak, W. Lorkiewicz, D. Więcek

Wroclaw University of Science and Technology
Faculty of Computer Science and Management, Department of Informatics
Laboratory of Computational Semiotics and Interactive Systems

E-mail: {radoslaw.katarzyniak, wojciech.lorkiewicz, dominik.wiecek}@pwr.edu.pl

Abstract: In this paper we deal with an original technically oriented model for cognitive semantics. As the expected area of application we focus on the process of extraction of modal linguistic summaries from data managed by autonomous components of ambient systems and intelligent environments. As such, the cognitive semantics is defined for a particular case of modal natural language statements with epistemic modalities. The statements of interest are built with natural language operators, representing epistemic modalities (related to the main cognitive states of knowledge certainty: full certainty, strong belief and epistemic possibility), and natural language connectives of equivalence. Furthermore, an approach to their effective processing by autonomous computational systems is designed. An internal architecture of the autonomous computational component is designed with respect to modular model for natural language processing with separate modules for epistemic and semantic memory storage and processing. An original theoretical concept underlying the model of semantic memory is a holon defined as a collection of complementary linguistic protoforms. Finally, we provide several illustrative computational examples of linguistic summaries' extraction, based on artificial and real data.

Key words: linguistic description, linguistic summary, natural language connective, equivalence, epistemic modality, natural language engineering, autonomous system

I. INTRODUCTION

Over the last years linguistic summaries have become an important concept, intensively studied by data sciences. Such popularity results from the fact that linguistic summaries are an effective way of producing compact and human-friendly representations of data collections with which human end users can naturally cope in the context of modern smart environments and modern systems [8, 50]. Moreover, modern IT systems have become strongly intertwined with Big Data paradigm and related models [40], where ability to generate user-friendly responses is highly important.

Up to now multiple classes of linguistic summaries have been elaborated and utilized, each tailored to a particular data semantics and specific application domains. Fortunately, despite the fact that these applications are diverse, strongly case-related and highly dependent on a particular data meaning, it is still possible to design their clear categorization. A theoretical concept, which proves to be useful for an effective differentiation between distinct groups of linguistic summaries, is the concept of protoform [23]. It is worth mentioning, that the concept of protoform in the context of linguistic summaries has already been applied. In each practical case a protoform was understood as an

abstracted prototype of a more general structure, i.e. an underlying members of a group of natural language (summarizing) statements, mutually interrelated by their shape and content. Often practically interesting hierarchies of proto-forms were also defined and applied.

However what seems to be the main drawback of proto-forms application, is the fact that there exists no unified language of precise representation of protoforms in data intensive systems. Below, we indirectly illustrate this drawback by the fact that four different linguistic summaries approaches specify different schemes of protoforms.

Although it is usually not directly stated, it is still worth mentioning that the concept of protoform is in many ways related to linguistics. For instance, to some extent it corresponds to the well-known concept of derivation tree. However, for the sake of this study, it is significant and interesting that there exists a strong theoretical relation between the concept of protoforms and the paradigm of cognitive linguistics [10], as well as interactive linguistics. The latter, lesser known in the mainstream of linguistic research, is based on an original MIC theory developed by Włodarczyk and Włodarczyk in CELTA[1], and *provides the information technology field with valuable and methodologically effective approach to natural language modelling and analysis* [2].

The suggested relation between the protoform, linguistic summary, and the paradigms of cognitive and interactive linguistics seems to be especially promising and helpful for practical applications. Namely, particular models of natural language processing, developed with respect to both theoretical paradigms, provide valuable support for the analysis, design and implementation of these components of ambient systems and smart environments, in which autonomous production of natural (semi-natural) language-based linguistic summaries is assumed as a desirable functionality of the system.

In this paper we present several results from research and development efforts to design, implement and verify an original computational model for processing *modal* linguistic summaries complying with the assumptions of mutual influence between different technologies, supporting the analysis, design, and implementation of ambient systems and intelligent environments, and the paradigms of cognitive and interactive linguistics. In particular, we focus on specific models of cognitive semantics dedicated to particular classes of (semi)natural language statements built using natural language connectives of equivalence and (auto)epistemic operators of knowledge, believe and possibility[2]. Preliminary results for the analysis of this class of modal sentences, sometimes only partially in line with the paradigms of cognitive and interactive linguistics, were published elsewhere, e.g. [34, 43-45]. In this study we take into account additional elements, conceptual and theoretical, that increase the degree of consistency between the commonsense semantics of natural language equivalence-based modal statements and their technical counterparts of our interest.

The paper is organized into six major sections, followed by a short conclusion of the presented research.

Section II presents a brief overview and comparison of four groups of linguistic summarizations of knowledge. For each group a set of related protoforms is given and adequate source publications are cited. Moreover, the main conceptual difference between fuzzy-theory related and epistemic modality-based summaries is pointed out. In general, this section locates our original research in a broader context of similar projects, subject to the fact that the other research into linguistic summaries does not involve the concept of epistemic modality as am important aspect of communication.

Section III illustrates several examples of practical contexts, in which modal linguistic summaries are built from natural language operators of epistemic modality and sentence connectives of equivalence are used as natural language representations of summaries knowledge. Special attention is paid to explanation of specific conditions, in which the use of the latter linguistic means is natural and necessary to capture the actual cognitive state of an autonomous system, carrying out a linguistic summary.

Section IV introduces a formal model for autonomous systems capable of producing the aforementioned modal linguistic summaries. A dedicated architecture of the autonomous systems is described, including a detailed presentation of computational submodules for episodic and semantic memory management. At the general level of functionality, the episodic memory is treated as data corpus – to be linguistically summarized, while semantic memory plays the role of a specialized module for managing protoforms and internal representations of extracted linguistic summaries.

Section V provides a detailed presentation and discussion of the core theoretical concept underlying the production of modal linguistic summaries built with natural language connectives of equivalence and modal operators of knowledge, belief and possibility. In particular, focusing on the underlying concept called cognitive semantics of modal statement.

Section VI illustrates the behaviour of the presented approach through several computational examples and depicts

---

[1] Centre de Linguistique Théorique et Appliquée, Université Paris Sorbonne-Paris IV.

[2] In natural languages auto-epistemic operators are used in such statements as *I know that X. I believe that Y*, or *I find it possible that X.*. Importantly, extraction and production of the aforementioned statements is realised by autonomous, interactive and data intensive knowledge processing systems. It is worth underlying, that in all of the aforementioned linguistic cases the first person singular form is an apparent characteristics and necessity. However, it is quite obvious that communicating auto-epistemic modalities is a far more complex issue. Namely, it is necessary to remember that the auto-epistemic modality is indirectly assigned to multiple statements without a direct application of the first person singular, e.g. *It is possible that X.*. It is enough to note that the concept of auto-epistemic modality becomes inevitable if the statement is referred to (and interpreted within) a subjective perspective – internally developed by this statement's uttering subject.

how commonsense consistency of extracted linguistic summaries is realized in the proposed computational model.

II. BRIEF OVERVIEW AND DISCUSSION OF RELEVANT RESEARCH

For the purposes of our study, we find it sufficient to briefly overview only four groups of linguistic summaries, already developed (to various extent), studied and utilized in multiple contexts of Big Data, as well as ambient intelligent systems and smart environments. As aforementioned, each of the four groups can be naturally defined by the related protoform (set of protoforms) used.

II. 1. Fuzzy conditional summaries

The first class consists of statements being instances of the following two protoforms, having the form of generalized conditional structure:

$$\text{IF } x \text{ is } A \text{ THEN } x \text{ is } B. \tag{1}$$

$$\text{IF } x \text{ is } A_i \text{ THEN } y \text{ is } B_j \text{ with possibility } R_{i;j}. \tag{2}$$

Historically, these semi-natural language statements were usually referred to as *linguistic knowledge* rather than *linguistic summaries*. However, it is quite obvious that instances of both protoforms (1) and (2) are compact and human-friendly linguistic representations of relations established over data, spoken (or written) in a particular natural language. Usually, they are used as approximate linguistic descriptions for behavioural algorithms of systems, automatically identified from corpora (collections) of numerical data [26].

Examples of linguistic summaries being instances of protoform (1) are:
- IF x is Small THEN y is Small,
- IF x is Small THEN y is Big,

and protoform (2):
- IF x is Big THEN y is Small with possibility 0.1,
- IF x is Big THEN y is Big with possibility 0.9.

There exist multiple, already classic, works dealing with fundamental problems of linguistic summaries derived from protoforms of conditional-structured linguistic summaries (e.g. [42, 49, 51] or [41]). Modelling and analysis of commonsense interpretability of the summaries is included.

It is necessary to mention that in many ways all research on conditional-structured linguistic summaries refers to the well-known concept of linguistic variable [20], which seems especially relevant to our study.

II. 2. Yager's original approach and its developments

Strong theoretical research and practical developments also related to the second group of popular linguistic summaries that was originally introduced by Yager [47] and substantially extended in many subsequent works, e.g. [16-18]. The main protoforms underlying the second group of linguistic summaries are as follows:

$$QR's \text{ are } S. \tag{3}$$

$$QBR's \text{ are } S. \tag{4}$$

where Q denotes a linguistic quantifier, R points at a set of objects, in a broad sense B defines a subset of R, and F is (a name of) a property attributed to objects from R. In consequence, summaries derived from these protoforms are linguistically quantified propositions [22].

Examples of linguistic summaries being instances of protoform (3) are:
- Most trees are high.
- Some trees are very old.

and protoform (4):
- Some active sensors are highly overheated.
- All young employees are very well paid.

Again, this line of research is apparently rooted in the original concept of linguistic variable [20], too, and in many ways is related to Zadeh's paradigm of computing with words [13, 24, 25].

II. 3. Linguistic summaries of periodical data

The third group of relatively well studied linguistic summaries can be in a way treated as a generalization of the previous group. However, the new element is the included aspect of time. In particular, the semantics of summarized data in this case is in various ways temporally labelled.

An example of related protoform of our interests is:

$$Prec_1 Time \text{ x is } Pdt \text{ with a period of } Prec_2 Unit. \tag{5}$$

where $Prec_1$ and $Prec_2$ are precision adverbs, $Time$ is a time contextualisation expression, Pdt is a linguistic term for particular periodicity assessment, and $Unit$ represents a particular time period. Variable x refers to a particular event.

Examples of instances of linguistic summaries, derived from protoform (5), are:
- *During the first decade*, the phenomenon is *periodic* with a period of *about 1 week.*
- *From 10 a.m. to 12 a.m.*, the signal is *highly periodic* with a period of *more or less 10 seconds.*

Over the last years, the group of temporally-related linguistic summaries has been an object of intensive theoretical study and application-oriented development. Although the body of results is still rather modest, as compared to previously mentioned groups of linguistic summaries, however, it consists of strong computational models (see [11, 14, 15] for interesting and valuable examples). Again, all of them are in many ways related to the original concept of linguistic variable [20] and Zadeh's paradigm of computing with

words [13, 24, 25].

II. 4. Modal linguistic summaries

The fourth group of linguistic summaries relevant to our study to be briefly communicated in this discussion we choose the set of (semi-)natural language statements derived from the following protoforms, already considered in our previous research, e.g. [37-39]:

$$\Pi(o \text{ is } S).$$
$$\Pi(o \text{ exhibits } S). \tag{6}$$

$$\Pi(o \text{ is not } P).$$
$$\Pi(o \text{ does not exhibit } P). \tag{7}$$

$$\Pi(o \text{ is } P \text{ and is } Q).$$
$$\Pi(o \text{ exhibits } P \text{ and exhibits } Q). \tag{8}$$

$$\Pi(o \text{ is } P \text{ and is not } Q).$$
$$\Pi(o \text{ exhibits } P \text{ and does not exhibit } Q). \tag{9}$$

$$\Pi(o \text{ is not } P \text{ and is } Q).$$
$$\Pi(o \text{ does not exhibit } P \text{ and exhibits } Q). \tag{10}$$

$$\Pi(o \text{ is not } P \text{ and is not } Q).$$
$$\Pi(o \text{ does not exhibit } P \text{ and does not exhibit } Q). \tag{11}$$

where Π is instantiated by an (auto)epistemic operator of knowledge certainty (uncertainty), and P and Q are properties attributed to object o.

A collection of (auto)epistemic modalities[3] considered in [37-39] consisted of the following three linguistic operators to communicate epistemic modality:

- the language phrase *I know that...* for verbal (or written) presentation of epistemic certainty (further formalized by *Know*(.)).
- the language phrase *I believe that...* for verbal (or written) presentation of slightly reduced certainty (further formalized by *Bel*(.)).
- the language phrase *It is possible that...* for verbal (or written) presentation of distinctly reduced certainty (further formalized by *Pos*(.)).

As aforementioned, the indirect usage of the first person singular is not a coincidence, i.e. in this way potential receivers are informed that the communicated content is extracted from a strictly subjective cognitive perspective of an autonomous system.

Possible illustrative examples of linguistic summaries, derivable from protoforms (8)-(11) are:

- I know that the object o is luminous if and only if it has a high temperature.
- I believe that the object o is not luminous if and only

if it does not have a high temperature.
- It is possible that the object o is luminous if and only if it does not have a high temperature.

The fact why instances of natural language statements derivable from protoforms (8)-(11) can (and should) be treated as linguistic summaries of some data is discussed in the following sections.[4]

It is worth mentioning that with a high degree of certainty it can be assumed that the three types of epistemic modality are captured by (and communicable with) the majority (if not all) contemporary natural languages.

Finally, it should be underlined that a regular theory regarding modal linguistic summaries has already been developed. It supports the way in which this class of summarizing statements can be evaluated against their pragmatic adequacy and commonsense consistency, and further applied to practical contexts. Such support is ensured by an original theory of modal language grounding (the grounding theory for short) which was presented in parts elsewhere: detailed sub-theories of natural language grounding relevant to protoforms (6)-(7) in [30, 36] and sub-theories related to protoforms (8)-(11) in [29, 31, 32, 35].

II. 5. Comparing of linguistic summaries – commonsense meaning and unconventional applications

At the general level the pragmatic function of the four classes of linguistic summaries (derivable from the presented cases of protoforms) seems to be similar, i.e. to communicate summarized knowledge about an aspect of a domain by means of natural language statements. However, there exists an important underlying difference between the first three classes of linguistic summaries and the latter one (derived from protoforms (6)-(11)).

Similarly as in [1], we assume the following list of elements involved in creation of linguistic descriptions and derivable from the first three classes of protoforms:

Input data

usually collections of numeric data, possibly with an associated temporal and/or spatial component, defined as input variables.

Linguistic variables

defined on the input variable domain as a set of fuzzy sets which label and categorize that domain. Each label in a linguistic variable is associated to a mathematical fuzzy definition in the form of a membership function.

Fuzzy quantifiers

These are also defined via fuzzy membership functions.

---

[3] See [19] for a comprehensive presentation and discussion of the concepts.

[4] See [39] to find out examples of complex, more extended and particular domain-related statements, formulated in a natural language, with a modal linguistic summary as their core element.

Evaluation criteria

> used to discriminate the most adequate linguistic descriptions (linguistic summarization), e.g. the data coverage degree, the sentence fulfillment degree, the relevance and the description length.

Protoform

> an abstract prototype for particular linguistic summaries of interest, used as a base for natural language statement derivation.

A similar, but not the same, list of elements can be proposed for the class of modal linguistic statements, based on sub-theories discussed in papers [29-32, 35, 36]:

Input data

> collections of data (not necessarily numeric), usually with an associated temporal and/or spatial component.

Linguistic variables

> defined on the input variable domain as a set of *crisp* sets which label and categorize that domain. Again, each label in a linguistic variable is associated to a crisp definition in the form of a membership function. In general, it is a set of properties' names, where each name of a property is associated to a particular value from the input variable domain.

Linguistic operators of (auto-)epistemic modalities

> These are defined via an original concept of cognitive semantics.

Protoform

> an abstract prototype for particular linguistic summaries of interest, used as a base for natural language statement derivation.

Evaluation criteria

> used to derive (and choose) adequate modal descriptions (modal linguistic summarization) that are realized by the above mentioned concept of cognitive semantics.

A deeper study into the nature of aforementioned groups of linguistic summaries forces the conclusion that main differences between both groups of summaries result from the fact that they are used to communicate two fundamentally different types of knowledge vagueness. Namely, the linguistic summaries derivable from protoforms (1)-(7) cover a case of fuzziness understood as lack of crisp borders of particular concepts [22], while the remaining summaries are used to

communicate epistemic uncertainty [21], resulting from subjectively experienced knowledge incompleteness (see [27] for an interesting classification of knowledge incompleteness types). In conclusion, it is the concept of a linguistic variable that is naturally related (and applicable) to all models of semantics, supporting linguistic summaries derivable from protoforms (1)-(7), while it is still not needed in our detailed sub-theories covering the semantics of modal linguistic summaries [29-32].[5]

It is quite obvious that the main body of research and development results on linguistic summaries covers mainly the case of fuzzy-related statements. However, for at least one reason the other results, related to epistemic uncertainty, can and should be evaluated as practically important and relevant: although it is rarely stressed in the literature on cognitive linguistics, it is still true that the concept of epistemic uncertainty is originally derived from the concept of system's autonomy. Therefore, if *autonomous* computational artificial components with abilities of natural language production are taken into account as a functionally required element of a broader system, the problem of computationally realized autonomous production of modal linguistic summaries becomes substantial and practically non-negligible. Therefore, modern interactive systems, especially those based on such paradigms as Big Data, ambient intelligence and smart environments, are the main application area targeted by our proposals presented in further subsections. The following motivating scenarios show, in a more detailed manner, particular circumstances in which autonomous production of modal linguistic summaries becomes unavoidable.

III. MOTIVATING SCENARIOS FOR EXTRACTION OF LINGUISTIC SUMMARIES

The fact that the class of studied linguistic summaries consists of natural language statements built with three types of operators, representing epistemic modalities, significantly changes practical contexts in which our summaries can be utilized. It is crucial that a final model for managing linguistic summaries' extraction (to be developed in the forthcoming sections) has to take into account *a necessary encapsulation of summarizing processes by a related autonomous system*. The related system has to be treated as one and the only one owner of such processes. A necessity to explicitly distinguish between the model components, which represent purely private, internal (cognitive) structures and internal knowledge bases of autonomous summarizing system, and the model components, which represent all entities of a world (a domain) external to the system, is a fundamental and inevitable theoretical assumption of the approach.

---

[5] For the sake of completeness of presentation, it is worth mentioning that some introductory steps have already been taken towards an effective integration of simple fuzzy linguistic summaries and modal linguistic summaries [46]. The results bring us closer to effective integration of processing of epistemic uncertainties and fuzzy imprecision in the same autonomous computational component. The related research into this issue is underway.

In other words: the target model of linguistic summaries' extraction from internal and encapsulated knowledge base must include a dedicated model for cognitive processes carrying out data summarization. The latter is contrary to models of linguistic summaries' extraction in which only a fuzzy semantics of data summarising statements is required, while all references to cognitive states of autonomous systems carrying out the summarization are negligible as irrelevant.

A clear distinction of the two cognitive perspectives, from the overall body of cognitive processes carrying out computational actions contributing to linguistic summaries' extraction, is not an easy task. Fortunately, the following two scenarios of data collection and summarization, based on two apparently distinct concepts of external and embedded (internal) data bases, allow for better explanation of the very specific nature of extracting linguistic summaries in which natural language connective of equivalence are combined with dedicated labels for epistemic modalities. Moreover, presented scenarios are intended to ground our technical research into linguistic summaries' engineering in relevant but non-technical theories, dealing with natural language processing.
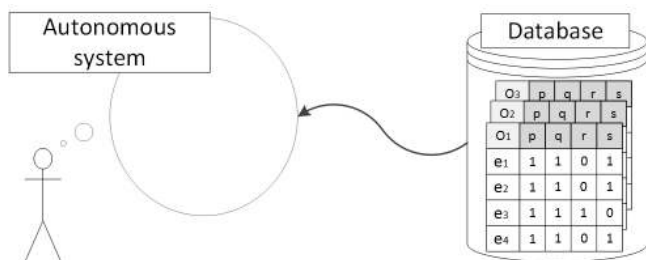

Fig. 1. Internal vs. external knowledge base – a generalized initial state

Both illustrative scenarios discussed below refer to the same set of conceptually simple, but still realistic situations shown in the forthcoming figures.

Firstly, the figures include a graphic representation for an autonomous computational system, i.e. a small figure of a man. The autonomous system is assigned an intentional stance to analyze available data in order to find out if and how each pair of events can be treated as equivalent, in the sense of their co-appearance. In general, our definition of an event can be eventually very broad, but for the sake of this research it is sufficient to consider a class of simplified events, related to a simple discrete fact of exhibiting a property by an object. Sample events are (a) exhibiting property P by an object o and (b) exhibiting property Q by an object o)[6].

Secondly, in all figures a graphic representation for an external data base is given. The external data base is intended to store representations of the events that occurred

in the past. Fig. 1 shows a complete content of the external database consisting of information on events occurred in a multi-object world. For the sake of simplicity, in all other figures only a portion of the external database is illustrated, i.e. associated with a single object.

Thirdly, in all figures a graphic representation of an internal database, encapsulated by the autonomous system, is proposed (see a circle associated with a small figure of a man). Similarly to the external database, the internal database is also intended to store data on events that occurred in an external world. However, for individual figures contents of the internal database will vary, depending on the purpose of a particular example of interest.

Fourthly, linguistic summaries, developed by the autonomous computational system, are represented as a collection of modal formulas. The formulas represent instances of particular protoforms, all based on sentence connective of equivalence and operators of epistemic modalities. Obviously, the formulas can always be easily reformulated into (stated as) adequate natural language statements.

From the autonomous system's perspective, the external data base plays a basic role of the primary source of knowledge about some events. Therefore, in order to acquire information about an event primary represented in the external data base, the autonomous system needs to retrieve adequate and event-related items of data from the external data base, and then save them in the internal (embedded) data base. An item retrieval operation is completed if and only if related data items have been successfully stored in the internal data base. In fact, a data item can be treated as a portion of knowledge cognitively acquired by an autonomous system if and only if a dedicated representation of the item (perhaps a direct copy of data from the external database) has been stored in the internal database (encapsulated / incorporated by this system). Otherwise the retrieved item would be inaccessible to all cognitive processes carried out by the autonomous system. The content of the internal data base defines the actual scope of the autonomous system's knowledge about recognizable external events. In other words, only part of the encapsulated data can contribute to what the autonomous system knows about external events)[7].

Fig. 1 represents an initial state treated as a reference point for both of the knowledge processing scenarios (and linguistic summary extraction tasks) considered below. Let us note that all primary knowledge about external events' appearances is stored in the dedicated external data base. In both scenarios the internal database will mediate between a linguistic summary and the primary knowledge source, i.e. the external database. However, the way the internal database is taken into account in a particular scenario and influence the autonomous knowledge state will vary significantly.

---

[6] Further, we refer to these and similar events by symbols $P(o)$ and $Q(o)$, respectively.
[7] Conversely, removing a data item from the internal data base is functionally equivalent to forgetting this item.

III. 1. Externally grounded linguistic summarization of data

Let us now consider the first aforementioned scenario in which summarization of knowledge is, in a particular sense, carried out in direct relation with events' representations originally stored in the external database. For the sake of presentation simplicity let us further assume that the target range of data for summarizing processes is restricted to all and only to these external data which refer to a single object o_3. Sample consecutive steps of the scenario are presented in Fig. 2-6.
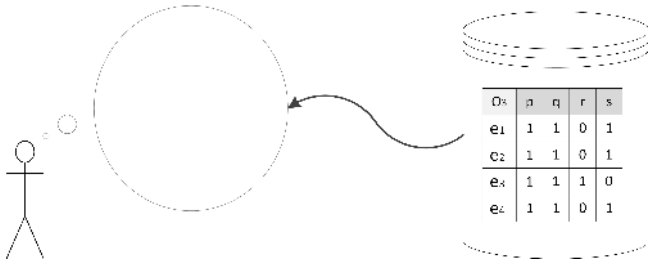

Fig. 2. Externally grounded summarization of data: state 1 = initial state

Under the first scenario the linguistic summary is established on a regular basis, i.e. to-date each new data item from an external database is retrieved and saved in the internal database. After a new data item has been saved, relevant linguistic summaries built from sentence with connectives of equivalence and summarizing retrieved data are computed.
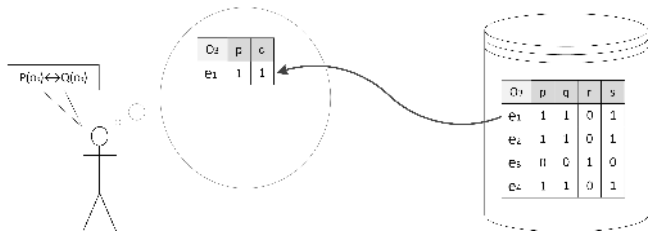

Fig. 3. Externally grounded summarization of data: state 2 = one item acquired

Fig. 2 shows an initial step at which the external database is restricted to all event-related data on an object o_3.
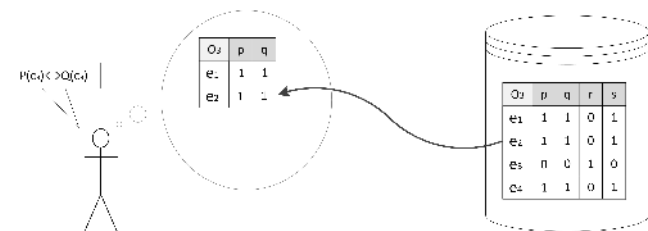

Fig. 4. Externally grounded summarization of data: state 3 = two items acquired

Fig. 3 shows the state which the autonomous system adopted after having acquired a data item on an observed state of properties P and Q (a simplified event). Obviously, the data item originated from the external data base. The actual subjectively experienced knowledge state of the system has changed as compared with the previous state (illustrated in Fig. 2). The entire collection of data stored internally by the autonomous system is checked against the appearance of equivalent events. In the situation depicted on Fig. 3 the entire internal data base developed by the system supports the equivalence $P(o_3) \Leftrightarrow Q(o_3)$. It is quite obvious that such equivalence is strengthened whenever the co-appearance of both properties in object o_3 is experienced.

The same transformation occurs between state 2 (see Fig. 3) and state 3 (see Fig. 2). The volume of a new internal database in state 3 is increased by 1, compared to state 2. However, the entire content of the internal database continuous to support the equivalence $P(o_3) \Leftrightarrow Q(o_3)$, and only this equivalence.
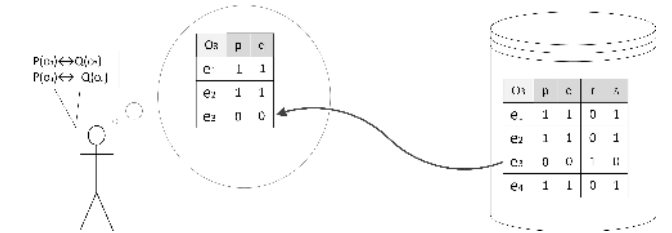

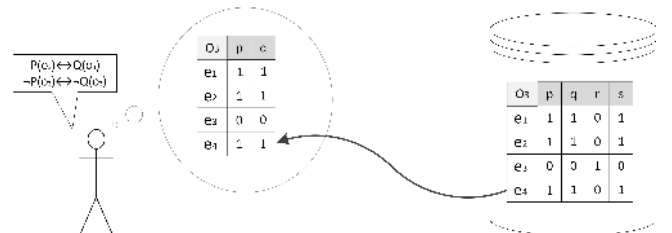Fig. 5. Externally grounded summarization of data: state 4 = three items acquired

Fig. 6. Externally grounded summarization of data: state 5 = four items acquired

A new situation depicted in Fig. 5, occurs after a new data item is acquired by the summarizing system. Namely, the internal database is supplemented by an event representation that supports the equivalence $\neg P(o_3) \Leftrightarrow \neg Q(o_3)$. The newly acquired data item does not influence in any way, i.e. does not support or contradict the equivalence $P(o_3) \Leftrightarrow Q(o_3)$. As such, it seems reasonable and adequate to extract two linguistic summaries based on both formulas shown in Fig. 5.

State 5 depicted in Fig. 6 emerges as a result of extension of the contents of the internal database by a knowledge piece supporting the equivalence $P(o_3) \Leftrightarrow Q(o_3)$. Therefore, even if the volume of the internally collected contents increases,

the form of linguistic summaries remains unchanged. Natural language statements communicating the meaning of two equivalences $P(o_3) \Leftrightarrow Q(o_3)$ and $\neg P(o_3) \Leftrightarrow \neg Q(o_3)$ are obtained by the summarizing autonomous system (a human end-user) as linguistic summaries of the primary external database, although the internal database substantially supported the realization of knowledge processing tasks.

Let us note that at each stage of the first scenario the summarizing autonomous system is assigned a state of certain knowledge regarding the equivalence relation between events under consideration. From the system's point of view, the current resulting knowledge is based *on and only on* all data contents which has ever been incorporated into the internal database. No other conceptual structure for knowledge representation, e.g. an empty slot for possible new knowledge items, is involved and considered as an important factor affecting the current result of summarization. For everything the autonomous system has collected up to a point, she is sure for her current resulting model of equivalence and acts as if a current task of summary extraction has been completed.

In conclusion, it is worth noting that the first scenario of linguistic summaries' extraction converges with natural behaviour of human end-user determining forms of possible and adequate linguistic summaries whenever a new data item comes from his or her external environment and with a manner of not constraining expectations regarding size and contents of primary source of data to be summarized.
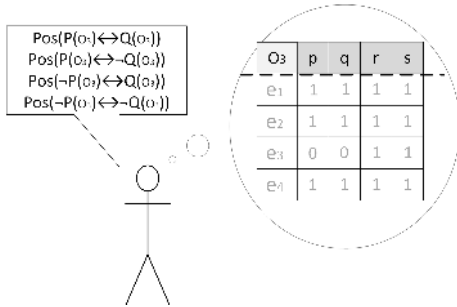

Fig. 7. Internally grounded summarization of data: state 1 = no item retrieved

III. 2. Internally grounded linguistic summarization of data

Let us now consider a second scenario for linguistic summaries production in which summarization processes are launched after the internal database is completed regarding its content. In other words, we will not deal with a scenario in which the linguistic summary is in a way updated after a new piece of data is acquired, but we concentrate on a sequence of knowledge states emerging step by step during the suc-

cessive transfer of data from the internal database to a working memory controlled by the autonomous system. Although the difference seems relatively slight, a deeper analysis leads to a conclusion that possible consequences are fundamental and significant, e.g. for the form of final linguistic summaries. The forthcoming presentation clarifies the issue utilising figures Fig. 7-11.

Fig. 7 depicts an initial state of summarization. A key element of the following scenario lays in the fact that data to be summarized is in entirety located in the internal database and the volume of collected data is computationally available for the autonomous system. As such, the system "knows" that four data items are stored in the internal database, but "she is not aware" of their content[8]. Moreover, it is the main reason why the knowledge state about the entire collected content is described by four complementary equivalences related to states of properties P and Q in the object o_3. Possibility operator is a very natural choice to mark the certainty level of all alternative equivalences to communicate their certainty as linguistic summarizations.
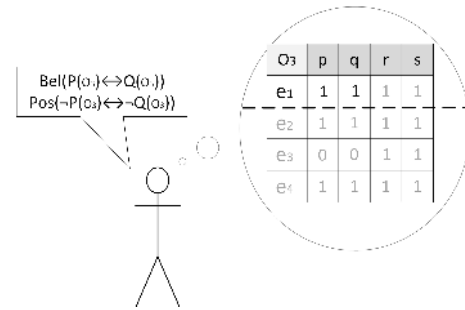

Fig. 8. Internally grounded summarization of data: state 2 = one item retrieved

Fig. 8 illustrates another state of the process. The first data item retrieved from the internal data base and transferred into the working area (in this sense explicitly included in a summary being extracted), substantially changes the current state of knowledge about the entire content of collected database to be summarized. Namely, a pair of the previous four alternative summaries are cognitively rejected as inconsistent with the just acquired data item. At the same time one of the remaining equivalences becomes stronger regarding the assigned level of certainty: the second modal equivalence has just been strengthened by a particular experience, while the other modal equivalence remains unsupported by yet retrieved data. The next step of the process contributes to the final result in a similar manner. Related results are depicted in Fig. 8.

---

[8] In our approach we often refer to *the so-called intentional approach to modelling systems*. It is worth mentioning that there exists a rich and comprehensive literature about this issue (see such excellent works as [5] or [6]).
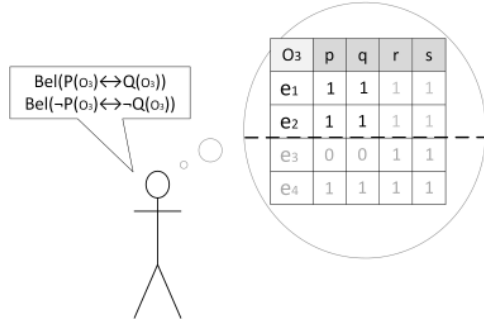
Fig. 9. Internally grounded summarization of data: state 3 = two items retrieved
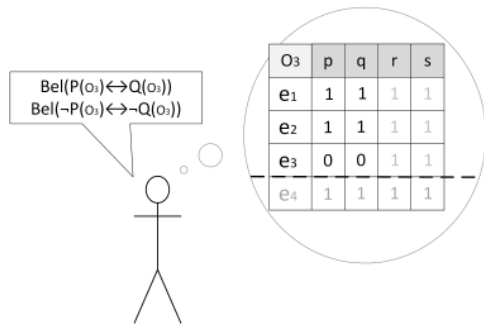

Fig. 10. Internally grounded summarization of data: state 4 = three items retrieved
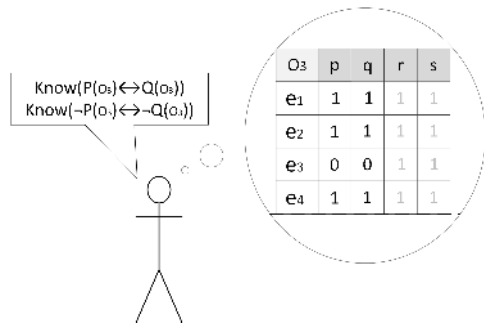

Fig. 11. Internally grounded summarization of data: state 5 = four items retrieved

Fig. 9 illustrates the next intermediate state of the final result computation. However, this time the third data item retrieved makes both equivalences equally certain, provided that the certainty is treated as experienced by the autonomous system herself. The remaining unprocessed content consists of only one data item. After the remaining item is retrieved and transferred into the working memory, the situation becomes certain, as regards to the final form of equivalence relation between events under consideration. For everything that the autonomous system has collected, she becomes certain about the actual form of events-related equivalence. The process of linguistic summarization completes and the final result is depicted in Fig. 11.

In this paper we omit a *detailed* explanation of psychological rationality of the second scenario by comparing this scenario with psychological models of semantics of modal statements. However, at least the following comments need to be made.

Firstly, the process of natural language generation by human systems is carried out including various types of collected knowledge. In particular, it is often inevitable that the process refers simultaneously to both episodic and semantic knowledge, i.e. both developed internally by a living system and stored in her internal databases.

Secondly, the overall process of natural language generation, including a case of generation of modal linguistic summaries, consists of multiple cognitive sub-processes which are carried out on various levels of living system's *awareness*. In fact, advanced and well supported non-technical models of natural language processing exist which manifestly refer to such property of language production and understanding. For instance, there exists a rich legacy of research into such concepts as deep cognition, implicit cognition, shallow level of cognition or deep level of cognition, i.e. worth mentioning regarding this issue is [12]). However, although the models contribute substantially to better understanding of actual processes of language production by humans, they are usually incompatible with the actual nature of artificial autonomous systems to which consciousness-related concepts, e.g. implicit cognition, are usually redundant.

The latter remark does not mean that there do not exist concepts that may be treated as artificial counterparts of the above mentioned human oriented concepts of deep and implicit cognition, which are particularly interesting to our study. Fig. 11 shows a situation in which a reference to the concept of deep cognition and related psychological phenomena seem to be natural and necessary. It is assumed that in Fig. 11 an autonomous computational system intends to determine autonomously the best possible linguistic summary of a distributed database (regarding two events related to properties *P*, *Q* and an object o_3). If cardinality of database is inconveniently high, as well as time and/or computing capabilities are substantially limited, the situation depicted in the figure can be effectively redefined by means of such concepts – as the internal database or implicit cognition, provided that the size of the entire distributed database is treated as known to the autonomous system and the entire distributed database is identified with the internal database. It is our claim that such conceptualization of computational processes carried out in the entire (Big Data based) distributed information system, depicted in Fig. 11, makes it possible to effectively and rationally produce instances of linguistic summaries, built with natural language connectives of equivalence and natural language operators of epistemic modalities.
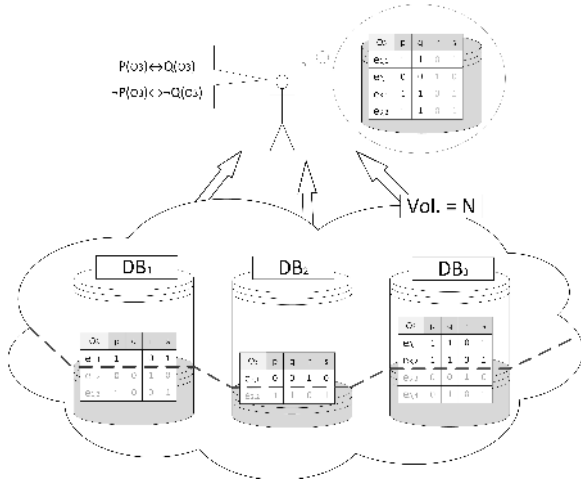
Fig. 12. Internally related summarization of a distributed database: a relevant context of application

In the following sections a technically oriented approach to modelling the above introduced ideas and issues is presented, discussed and partially verified.

IV. A MODEL FOR TECHNICALLY ORIENTED GROUNDING OF LINGUISTIC SUMMARIES

In this paper we strictly focus on presentation of the process of extracting linguistic summaries with epistemic modalities and natural language connectives of equivalence. Let us now introduce the formal framework for managing and grounding the aforementioned linguistic descriptions. It is worth underlying, that the presented model follows directly from the original theory of epistemic modalities' grounding [30, 31, 33] adopted and extended to a case of natural language connectives of equivalence.

The overall architecture (see Fig. 13) of the presented approach falls into the category of ubiquitous computing and ambient intelligent systems. Yet, the underlying structures are rather simple and consist of several different data storages, interconnected modules, and processes that govern interconnections between modules. In particular, different data storages involve storage of episodic and semantic knowledge, as well as storage of linguistic summaries. Further, different modules manage aforementioned data storages, whereas interconnections of the modules define how these storages are linked with each other and involve three major processes: perception, semantic memory transition, and grounding. For the sake of simplicity we further limit our presentation to a case of a single agent system. However, please note that the presented approach can be extended to a case of multi-agent system – population of individual agents.
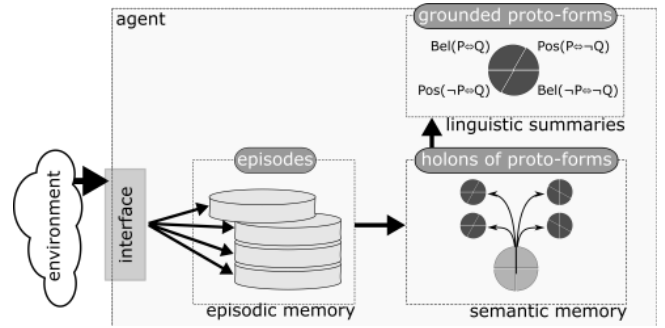

Fig. 13. Overall architecture

IV. 1. Representation of episodic perceptions

At the very core of the proposed approach is an individual autonomous agent. As aforementioned, the original theory of epistemic modalities' grounding assumes that the subject of knowledge (autonomous agent) is situated in a world and collects empirical knowledge about the world states. Each observation of the world is given as a basic knowledge granule and contributes to the content of an episode (functionally related to episodes in neurolinguistic models [9]).

We further assume that the external environment *World* consists of a set of atomic objects $o \in O$. Each object from the external environment can either exhibit or not exhibit a property $P \in \Delta$ from a finite set of all observable properties $\Delta P = \{P_1, \ldots, P_K\}$. Due to the technical nature of the proposed approach individual world states are relativized to individual discrete time points $t \in T$. For the sake of notation brevity, we further assume that property $P_i(t)$ represents a set of all objects that exhibit property P_i at time point t. Further, $O/P_i(t)$ represents a set of all objects that do not exhibit property P_i at time point t. Consequently, the following world profile tuple formally represents actual world's states:

$$WP(t) = <O, P_1(t), \ldots, P_K(t)> . \quad (12)$$

An individual embedded agent is capable of perceiving the external environment through a set of dedicated sensors. As such, these sensors are a form of a direct interface between the agent and the external world, and results of individual observations are encapsulated in agent's internal knowledge base (see Fig. 13). Such a knowledge base is a strictly private body of knowledge about the state of the external environment. It consists of information regarding the state of different objects, available in the environment at current time point t, in regards of the perceivable[9] properties $P_1, \ldots, P_K$. As such we further assume that a set $P_k^+(t)$ consists of all objects $o_i \in O$ that at time point t where registered by the agent to exhibit property P_k, whereas set $P_k^-(t)$ consists of all objects $o_i \in O$ that in time point t where registered to not to exhibit property P_k.

---

[9] Available to the agent's perceptual system.

The following base profile tuple is introduced to formally represent internally stored (embodied) observation – episode:

$$BP(t) = < O, P_1^+(t), P_1^-(t), \dots, P_K^+(t), P_K^-(t) > \quad (13)$$

The introduced tuples are sometimes called *t*-related world profile (Eq. 12) and *t*-related base-profile (*t*-related episodes) (Eq. 13), respectively [30, 36].

In particular, for each $k = 1, \dots, K$, $o \in P_k$ holds *if and only if* at time point t object o exhibited property P_k, otherwise the property was not exhibited; for each $k = 1, \dots, K, o \in P_K^+(t)$ *if and only if* at time point t object o is represented in *t*-related episode as exhibiting the property P_k; for each $k = 1, \dots, K$, $o \in P_k^-$ *if and only if* at time point t the object o is represented as non-exhibiting property P_k in *t*-related episode.

We further assume that the agent's perception system is correct[10]. In short, the agent cannot perceive anything that is outside of (contradictory to) the current state of the world. Formally the correctness is defined in strict reference to the world profile (see Eq. 14). In particular, at a given time point t the agent cannot perceive an object o exhibiting certain property $o \in P_k^+(t)$ if the world profile contradicts such a state $o \notin P_k(t)$. Moreover, at a given time point t the agent cannot perceive an object o not exhibiting certain property $o \in P_k^-(t)$ if the world profile contradicts such a state $o \in P_k(t)$.

$$\begin{aligned} o \in P_k^+(t) \Rightarrow o \in P_k(t) \\ o \in P_k^-(t) \Rightarrow o \notin P_k(t). \end{aligned} \quad (14)$$

Internal representation of observations conducted by the autonomous agent up to time point t can be represented as a set of established base profiles that represent individual episodes experienced by the agent. The overall content of the episodic knowledge base is represented by the following collection of base-profiles:

$$Episodes(t) = \{BP(t_n) : t_n \in T \quad \text{and} \quad t_n \leq t\}. \quad (15)$$

As aforementioned, the role of *Episodes(t)* is crucial due to its ultimate function in the induction of subjective references to the external world given in a form of linguistically oriented models stored in the semantic memory, especially in situations when direct world's observation is restricted or impossible for some reasons.

In each state of the system (defined by the time point $t \in T$) agent's empirical knowledge is divided into two parts. The first, called *shallow*, represents a conscious part of the knowledge processing processes and includes a portion of already processed *episodes*. The second, called *deep*, represents the rest of the empirical material, which is beyond the reach of conscious cognition. It includes still unprocessed – raw – *episodes*. As it has already been shown, such a division has its justification and is widely described in the cognitive science literature. The first part represents the focal point (point of attention) of the entity, whilst the other represents the unconscious (to the entity) empirical knowledge. It seems an obvious analogy to that of the distribution of processed and unprocessed part of knowledge. In particular, the existence of such a knowledge division implies the need for markers of modal uncertainties, taking into account the impact of unconscious empirical knowledge on the current context expression.

The set of *Episodes* is naturally divided into two levels: deep and shallow. At each particular time point $t \in T$, the related *Episodes* is given as a partition of *Episodes(t)* into the shallow and deep level of knowledge, represented by the following tuple:

$$Episodes(t) = (\overline{Episodes(t)}, \underline{Episodes(t)}), \quad (16)$$

where $\overline{Episodes(t)}$ states for the experience which is located at the shallow cognitive level (at the time point *t*), and $\underline{Episodes(t)}$ states for the remaining experience (located at the deep cognitive level and at the same *t*).

IV. 2. Representation of semantic knowledge

Typically there is a gradual transition from episodic to semantic memory, in which episodic memory reduces its sensitivity and association to particular events, so that the information can be stored as more general knowledge. Similarly, in the proposed approach the episodic knowledge fills in adequate semantic structures eventually reducing the original association to a particular past event.

The main assumption of the grounding theory is that linguistic statements are inseparably connected to so-called mental language holons. Language holons represent embedded summarization of empirical episodic experiences gathered by an individual. In particular, each holon relates to experiences, though reduced their sensitivity and association to particular events, strictly related to certain subject(s). As such, the holon represents individually established semantic structure that represents, i.e. in the sense it summarizes past experiences of an individual with a particular subject(s).

It should be underlined that the structure of the holon is strictly related to a given protoform at hand. Moreover, a holon has the noteworthy property of playing the role of a whole and a part at the same time. As such a holon fills in a certain protoform with the content (whole), thus represents its particular instantiation, and at the same time is composed

---

[10] In the presented paper it is assumed that the perception system of an individual agent is additionally complete. In particular, it is not possible that at a given time point t agent is unable to determine the state of a given property P_k (or properties) in a given object o. Nevertheless, it is worth mentioning that an incomplete model is also considered and is further researched by the authors elsewhere.

of internal structures (parts) that result from transition from episodic memory. For example, in the case of simple modalities [30] a holon relates to a protoform of type $\Omega(p)$ (where Ω is a modal operator, and p refers to object exhibiting property P). Its internal structure is then comprised of two interconnected parts (see Fig. 14) – summaries of individual experiences that involved this object <u>to exhibit</u> property P, and summaries of individual experiences that involved this object <u>not to exhibit</u> property P.
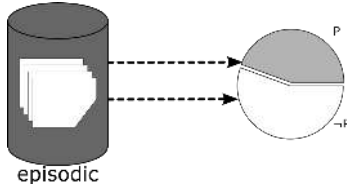

Fig. 14. Model of a simple holon

We can note that each part of the holon relates to a different cognitive state the individual can experience, whereas the whole holon refers to the relation between these cognitive states. As such, we can state that in many ways language holons are similar to mental models known from the cognitive linguistics and psychology [28].

Consequently, we can note that established linguistic summary related to the protoform is grounded in agent's individual experiences in a sense that it is grounded in the so-called mental language holons.

From the pragmatic point of view, mental language holons are higher level summarizations (semantic generalizations) of relative share of complementary bodies of experiences related to particular subjects (or their conjunctions). For the sake of completeness it is worth mentioning that at the technical level mental language holons can be treated as complexes of complementary classification rules (see Fig. 15).
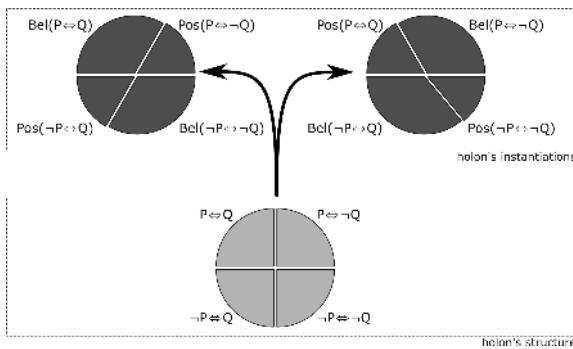

Fig. 15. Model of holon – case of equivalence

Let us now focus on the formal tool to represent different classes of cognitive states involved in the process of grounding linguistic summaries with epistemic modalities and natural language connectives of equivalence. In particular, we

focus on the presentation of the holon, i.e. construction of its parts and construction of the whole. As such, the focal point of the following presentation is the case of $\Omega(p \Leftrightarrow q)$, $\Omega(p \Leftrightarrow \neg q)$, $\Omega(\neg p \Leftrightarrow q)$, and $\Omega(\neg p \Leftrightarrow \neg q)$ protoforms.

For the sake of brevity of definitions and ease of presentation let us further focus on a case of a single object $o \in O$. Bear in mind that this simplifying assumption does not limit the generality of the presented model. In particular, all the presented definitions and theorems can be easily extended to a case, where multiple objects are taken into consideration.

The first tool is the notion of a grounding set. Grounding set $C^{P,Q}$ contains all base profiles that are available in agent's current knowledge state $Episodes(t)$, i.e. gathered agent's observations of the external world up to time point t, and which represent experiences relating to information about properties $P \in \Omega$ and $Q \in \Omega$.

Such a grounding set is naturally partitioned into four basic sub-parts, each representing different combination of available information regarding the properties $P \in \Omega$ and $Q \in \Omega$. For instance, one such part focuses on the cognitive state relating to the case where both properties $P \in \Omega$ and $Q \in \Omega$ were registered to exhibit in object o. Another part focuses on the cognitive state relating to the case where property $P \in \Omega$ was registered to exhibit in object o and property $Q \in \Omega$ was registered not to exhibit in object o. The third part focuses on the cognitive state relating to the case where property $P \in \Omega$ was registered not to exhibit in object o and property $Q \in \Omega$ was registered to exhibit in object o. The final part focuses on the cognitive state relating to the case where both properties $P \in \Omega$ and $Q \in \Omega$, where registered not to exhibit in object o.

Formally, all empirical support which is functionally accessible to the episodic memory and related to semantically complementary protoforms is given by the following interrelated subsets of *Episodes(t)*:

$$C^{p \wedge q}(t) = \{E(t_n) : t_n \leq t, o \in P^+(t_n) \text{ and } o \in Q^+(t_n)\}, \tag{17}$$

$$C^{p \wedge \neg q}(t) = \{E(t_n) : t_n \leq t, o \in P^+(t_n) \text{ and } o \in Q^-(t_n)\}, \tag{18}$$

$$C^{\neg p \wedge q}(t) = \{E(t_n) : t_n \leq t, o \in P^-(t_n) \text{ and } o \in Q^+(t_n)\}, \tag{19}$$

$$C^{\neg p \wedge \neg q}(t) = \{E(t_n) : t_n \leq t, o \in P^-(t_n) \text{ and } o \in Q^-(t_n)\}. \tag{20}$$

We can note that the grounding set C is a particular technical realisation of the linguistic holon. Such a holon as a whole represents an instantiation of a certain protoform,

and as a part is naturally partitioned into four interconnected components $C^{p \wedge q}$, $C^{p \wedge \neg q}$, $C^{\neg p \wedge q}$, and $C^{\neg p \wedge \neg q}$.

It is worth mentioning a basic comment why we utilise holons comprised of basic conjunctions between two properties P and Q in order to infer linguistic summaries with natural language connectives of equivalence. In particular, only by experiencing real co-occurrence of properties P and Q (conjunction) at certain time point t in the empirical database of an individual, it is possible to construe particular equivalence relation. If we can recollect in our empirical memory a situation in which both properties P and Q co-existed (are exhibited at the same time), and no contradictory situations were registered, it is only then that we can state P *if and only if Q*. Naturally, in order to comply with certain commonsense restrictions, i.e. strictly related to natural language understanding, the co-occurrence of the properties needs to fulfil certain conditions and dependencies. In essence, the proposed model thrives at an approach to formally represent and implement the aforementioned restrictions.

Analogous to the case of an *Episodes(t)*, we further consider a dichotomous partition of the episodic content of the linguistic holon. In particular, the *shallow* portion of the grounding set represents a conscious part of the knowledge processing processes and includes a summary of the portion of already processed *episodes*, whereas *deep* portion of the grounding set represents a summary of the rest of the empirical material, which is beyond the reach of conscious cognition. For instance, it includes some general indication of still unprocessed – raw – *episodes*.

Furthermore, each grounding set is divided into shallow and deep level:

$$C^{p \wedge q}(t) = (\overline{C^{p \wedge q}(t)}, \underline{C^{p \wedge q}(t)}), \qquad (21)$$

$$C^{p \wedge \neg q}(t) = (\overline{C^{p \wedge \neg q}(t)}, \underline{C^{p \wedge \neg q}(t)}), \qquad (22)$$

$$C^{\neg p \wedge q}(t) = (\overline{C^{\neg p \wedge q}(t)}, \underline{C^{\neg p \wedge q}(t)}), \qquad (23)$$

$$C^{\neg p \wedge \neg q}(t) = (\overline{C^{\neg p \wedge \neg q}(t)}, \underline{C^{\neg p \wedge \neg q}(t)}). \qquad (24)$$

Let us further provide some comments on the division between the shallow and the deep part of the grounding set. Such a distinction is crucial, as during the processing stage we do not possess the complete information about $C^i(t)$. In particular, every incoming episode at time point t is classified (or not) into one of the four instances of the grounding set (Eq. 17, Eq. 18, Eq. 19, Eq. 20). When successfully classified, it falls into the processed (shallow) part of the grounding set $\overline{C^i(t)}$. Furthermore, such a distinction is crucial to introduce subtle dependencies between *deep* and *shallow* levels in the context of conditional statement and its formalisation. Such a relation is further referred to as *relative grounding strenght* (Eq. 25, Eq. 26, Eq. 27, Eq. 28) and is further described in the following section.

In conclusion, the aforementioned gradual transition from episodic to semantic memory is realised by the process of establishing linguistic holons represented in the form of grounding sets. This particular process defined in the grounding theory [30, 31] is certainly a more general process and does not only relate to linguistic summaries with language connectives of equivalence. In particular, similar definitions are proposed for other language connectives, i.e. conjunctions, disjunctions and conditional statements.

It should be underlined that this transitioning process imposes a natural division of the episodic knowledge gathered by an individual into complementary, yet competing, mental models, each supporting different instantiation of a certain linguistic protoform. As a whole, such instantiation can be regarded as a form of representation of a cognitive state of an individual. In particular, it seems a really good indication of the state of an individual's thought processes and current state of mind.

Finally, the transition from episodic memory, here represented as a collection of episodes, leads to establishment of semantic memory, here represented as linguistic holons – grounding sets.

V. COGNITIVE SEMANTICS FOR MODAL EQUIVALENCES

Once the semantic module is in place, we can focus on the final process of information processing. In particular, the process of establishing linguistic representation (linguistic summaries) adequate to the current cognitive state (semantic knowledge) of an individual is fundamental to cognitive semantics. This process is strictly related to the process of grounding of linguistic statements [48] and a general concept of semiotic triangle [3], and is the central stage of the grounding theory model. The assumed grounding theory is based on a technical model of so-called cognitive semantics for natural language statements, limited to some scope of quasi-natural language statements, modal in the epistemic sense.

The fundamental concept of the cognitive semantics is the notion of semiotic symbol [4]. With respect to the so-called semiotic definition of symbols, each such symbol should be considered in relation to the cognitive agent that is the carrier of their sense. The role of semiotic symbols in semantic communication is defined by three elements of the semiotic triangle: the material form of the symbol, e.g. a linguistic representation available to the entire population, the sense assigned to the material form by a certain cognitive agent, and an object usually located in the external world to which the semiotic symbol is referred by a certain cognitive agent.

Semiotic symbols are understood as an integrated form-meaning package – a common form of the external repre-

sentation (character), and significance in the context of an individual agent's cognitive system (sense). Such a semiotic symbol relates directly to a particular proto-form), e.g. $Know(p \Leftrightarrow q)$. In essence, it is associated with a certain external representation (that can be expressed/uttered by an individual), e.g. *I know that o exhibits P is equivalent to o exhibits Q*, and is instantiated by a certain cognitive state (the the individual is in) (see Sec. V. 1.), e.g. grounding set $C = \langle C^{p \wedge q}, C^{p \wedge \neg q}, C^{\neg p \wedge q}, C^{\neg p \wedge \neg q} \rangle$.

In the presented work four basic instantiations of general protoforms, considered in the original theory of epistemic modalities grounding, can be derived. Each instantiation realises known cases of expressing realizable knowledge states about the current state of property P and Q in object o:

- "*o* exhibits *P* is equivalent to *o* exhibits *Q*",
- "I know that *o* exhibits *P* is equivalent to *o* exhibits *Q*",
- "I believe that *o* exhibits *P* is equivalent to *o* exhibits *Q*",
- "(For all I know) it is possible that *o* exhibits *P* is equivalent to *o* exhibits *Q*".

Each of the aforementioned general protoforms can be later instantiated as four basic protoforms, i.e. each for different case "*o* exhibits *P* is equivalent to *o* exhibits *Q*", "*o* exhibits *P* is equivalent to *o* not exhibiting *Q*", "*o* not exhibiting *P* is equivalent to *o* exhibits *Q*", and "*o* not exhibiting *P* is equivalent to *o* not exhibiting *Q*".

In particular, we focus on a class of linguistic summaries that require utilising linguistic means used in natural language to communicate different levels of trust in the appropriateness of generated utterances – auto-epistemic belief (see [19] for adequate concept discussion). In particular, it is worth noting that the concept of auto-epistemic modality is directly related to a subjective perspective of an individual and internally developed by the subject utilising such a statement. In natural language modal epistemic operators serve such a function, for instance operators of certainty (I know that...), belief (I believe that...), and possibility (I find it possible that...). As such, in the presented paper we focus on these three classes of modal operators representing different layers of epistemic modality (see Sec. II. 4.). With a high degree of certainty, it can be assumed that the aforementioned three types of epistemic modality are captured by (and communicable with) the majority (if not all) contemporary natural languages. Moreover, it is worth mentioning that both statements "*o* exhibits *P* is equivalent to *o* exhibits *Q*" and "(For all I know) it is possible that *o* exhibits *P* is equivalent to *o* exhibits *Q*" are similar. However, both results from different internal processes. The former is a statement of fact, whilst the latter implies an internal belief of an individual unit.

V. 1. Cognitive state's importance

In the proposed approach we follow the stance by Dennett [6] "Exposure to x, that is sensory confrontation with x over some suitable period of time, is the normally sufficient condition for knowing (or having true beliefs) about x". As such, relative intensities (strengths) of internal components of a holon, i.e. sub-part of the grounding sets C, are a good condition for establishing beliefs. Furthermore, calculating relative strengths of each parts of the holon are a sufficient form of indication of how important a particular cognitive state is within the individual.

Moreover, it is quite natural and practically rational that whenever knowledge about the actual state of P or Q in object o is missing, the *t*-related content of $C^{p \wedge q}$, $C^{p \wedge \neg q}$, $C^{\neg p \wedge q}$ and $C^{\neg p \wedge \neg q}$ may be used to fill the existing knowledge gap.

It is assumed in the theory that each epistemic operator of modality (*Possibility*, *Believe*, *Knowledge*) is clearly related to a certain scope of previously mentioned relevant intensity of summarised empirical evidence, related to the sense of proto-forms considered. In consequence, to each linguistic protoform, always related to one and only one part of a relevant mental language holon, certain intensity of summarized (embodied) experience of a subject (or binary conjunctive subject) is assigned. In the grounding theory this intensity is numerically represented by the relative grounding strength, i.e. $\lambda^{p \Leftrightarrow q}(t)$, $\lambda^{p \Leftrightarrow \neg q}(t)$, $\lambda^{\neg p \Leftrightarrow q}(t)$ and $\lambda^{\neg p \Leftrightarrow \neg q}(t)$.

Assigned to all linguistic protoforms under consideration, the relative grounding strengths can be expressed as follows:

$$\lambda^{p \Leftrightarrow q}(t) = \frac{\left| \overline{C^{p \wedge q}(t)} \right| + \left| \overline{C^{\neg p \wedge \neg q}(t)} \right|}{|Episodes(t)|}, \qquad (25)$$

$$\lambda^{p \Leftrightarrow \neg q}(t) = \frac{\left| \overline{C^{p \wedge \neg q}(t)} \right| + \left| \overline{C^{\neg p \wedge q}(t)} \right|}{|Episodes(t)|}, \qquad (26)$$

$$\lambda^{\neg p \Leftrightarrow q}(t) = \frac{\left| \overline{C^{\neg p \wedge q}(t)} \right| + \left| \overline{C^{p \wedge \neg q}(t)} \right|}{|Episodes(t)|}, \qquad (27)$$

$$\lambda^{\neg p \Leftrightarrow \neg q}(t) = \frac{\left| \overline{C^{\neg p \wedge \neg q}(t)} \right| + \left| \overline{C^{p \wedge q}(t)} \right|}{|Episodes(t)|}, \qquad (28)$$

where $|X|$ denotes the cardinality of set X. It is worth mentioning that due to complete observations of an agent $|Episodes(t)| = |C^{p \wedge q}(t)| + |C^{p \wedge \neg q}(t)| + |C^{\neg p \wedge q}(t)| + |C^{\neg p \wedge \neg q}(t)|$.

We can note that in case of natural language connectives of equivalence there is internal dependency between grounding strengths of different parts of holon, i.e. $\lambda^{p \Leftrightarrow q} = \lambda^{\neg p \Leftrightarrow \neg q}$ and $\lambda^{p \Leftrightarrow \neg q} = \lambda^{\neg p \Leftrightarrow q}$. In particular, in natural language both protoform pairs $(\Omega(p \Leftrightarrow q), \Omega(\neg p \Leftrightarrow \neg q))$ and $(\Omega(p \Leftrightarrow \neg q), \Omega(\neg p \Leftrightarrow q))$ are internally intertwined. Occurrence of one experience supporting a certain element of such

a pair is maybe not reinforcing but rather is not weakening the relative grounding strength of the other element of such a pair.

Finally, in the proposed model cognitive state's importance is strictly related to the intensity of summarised empirical evidence and is expressed using the notion of relative grounding strength of a particular cognitive state. In particular, each part of the holon can be associated with a particular score – relative grounding strength. These scores are further used to estimate which protoforms can be grounded in the agent's individual experience.

V. 2. Expressing auto-epistemic beliefs

As aforementioned, it the presented approach we investigate linguistic summaries with epistemic modalities, for natural language connectives of equivalence. Epistemic modalities are represented using three modal operators of belief.

In particular, we assume that all linguistic summaries studied in this paper are instantiations of the following protoform $\Omega(p \leftrightarrow q)$, where $\Omega \in \{Know; Pos; Bel\}$. The related linguistic representations of the protoforms are as follows:

1. Protoform $Know(p \Leftrightarrow q)$ can be communicated as *It follows that p holds if and only if q holds.*
2. Protoform $Bel(p \Leftrightarrow q)$ can be communicated as *I believe that p holds if and only if q holds.*
3. Protoform $Pos(p \Leftrightarrow q)$ can be communicated as *It is possible that p holds if and only if q holds.*

Each of the aforementioned general protoforms can be later instantiated as four basic protoforms, with different $p \Leftrightarrow q$, $p \Leftrightarrow \neg q$, $\neg p \Leftrightarrow q$ and $\neg p \Leftrightarrow \neg q$.

In order to determine which of these instantiations should be derived from the basic protoform, as adequate to a certain knowledge state, original definitions for cognitive semantics of all possible instantiations were proposed (in a form of the so-called epistemic satisfaction relations). Their consistency with the predefined set of common-sense postulates can be analytically proved following an analogous reasoning of that presented in [30, 31].

The most important conclusion from the theory of epistemic modality grounding is important and interesting for purely pragmatic reasons. Namely, one can prove in a strict analytic way the following general theorem: *the relatively simple and intuitively acceptable concepts of grounding sets, modality thresholds and epistemic satisfaction relations are fully sufficient for designing an artificial agent capable of producing consistent and intuitively acceptable sets of linguistic summaries, stated in quasi-natural epistemic modal languages.* As aforementioned, proofs and a more advanced discussion of this result are given elsewhere. In particular, the choice criteria are given as a dedicated system of so-called modality thresholds – each threshold defining an upper or lower level of range of this relative grounding strength which makes the use of a certain epistemic operator of modality possible.

In the proposed approach, the fundamental role appears to be played by the collection of modality thresholds $\lambda^{\Leftrightarrow}_{minPos}$, $\lambda^{\Leftrightarrow}_{maxPos}$, $\lambda^{\Leftrightarrow}_{minBel}$, $\lambda^{\Leftrightarrow}_{maxBel}$ and $\lambda^{\Leftrightarrow}_{Know}=1$. In particular, the collection of modality thresholds mediates between the overall content of episodic knowledge base and adequate utilization of summarized episodes' content in a creation of linguistic summaries of current knowledge states, especially when some data on the current state is not accessible. Moreover, the collection becomes the main cognitive mechanism responsible for ensuring proper production of acceptable and intuitively consistent sets of modal summaries of collected data.

An interesting result from the theory of grounding, for the practice perhaps the most important one, is that the system of modality thresholds cannot be freely chosen. Namely, in order to guarantee common sense consistency of (written and verbal) language behaviour the system of modality thresholds has to fulfil some predefined set of requirements, accepted in the theory of grounding, as a reflection of common sense pragmatics applied in actual contexts to natural language operators of knowledge, belief, and possibility. The fact that written and/or verbal behaviour, produced by a technical system based on the theory of grounding, is actually consistent, from the semiotic and pragmatic point of view, can be analytically proved and verified.
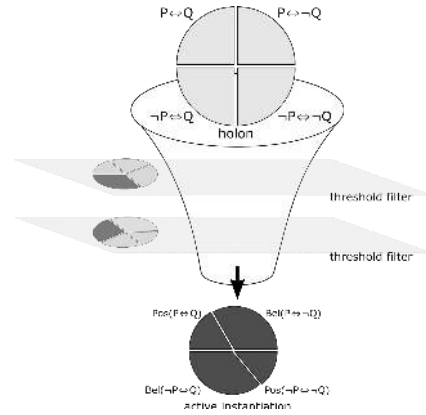
Fig. 16. Model of threshold activation step

For the sake of completeness we further present all the main definitions for proper grounding of instantiations of all considered protoforms.

Protoform $\Omega(p \Leftrightarrow q)$:

$Episodes(t) \models_G$ "It is possible that o exhibits P is equivalent to o exhibits Q"

$PS(t) \models_G Pos(p \Leftrightarrow q)$ holds iff:

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge q}(t)}\right| = 0 \qquad (29)$$

$$0 \leq \lambda^{p \Leftrightarrow q}(t) \leq \lambda^{\Leftrightarrow}_{maxPos} \qquad (30)$$

$Episodes(t) \models_G$ "I believe that o exhibits P is equivalent to o exhibits Q"

$PS(t) \models_G Bel(p \Leftrightarrow q)$ holds iff:

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge q}(t)}\right| = 0 \qquad (31)$$

$$\lambda^{\Leftrightarrow}_{minBel} \leq \lambda^{p \Leftrightarrow q}(t) < 1 \qquad (32)$$

$$\left|\overline{C^{p \wedge q}(t)}\right| > 0 \qquad (33)$$

$Episodes(t) \models_G$ "I know that o exhibits P is equivalent to o exhibits Q"

$PS(t) \models_G Know(p \Leftrightarrow q)$ holds iff:

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge q}(t)}\right| = 0 \qquad (34)$$

$$\lambda^{p \Leftrightarrow q}(t) \cong 1 \qquad (35)$$

$$\left|\overline{C^{p \wedge q}(t)}\right| > 0 \qquad (36)$$

Protoform $\Omega(p \Leftrightarrow \neg q)$:

$Episodes(t) \models_G$ "It is possible that o exhibits P is equivalent to o not exhibiting Q"

$PS(t) \models_G Pos(p \Leftrightarrow \neg q)$ holds iff:

$$\left|\overline{C^{p \wedge q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge \neg q}(t)}\right| = 0 \qquad (37)$$

$$0 \leq \lambda^{p \Leftrightarrow \neg q}(t) \leq \lambda^{\Leftrightarrow}_{maxPos} \qquad (38)$$

$Episodes(t) \models_G$ "I believe that o exhibits P is equivalent to o not exhibiting Q"

$PS(t) \models_G Bel(p \Leftrightarrow \neg q)$ holds iff:

$$\left|\overline{C^{p \wedge q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge \neg q}(t)}\right| = 0 \qquad (39)$$

$$\lambda^{\Leftrightarrow}_{minBel} \leq \lambda^{p \Leftrightarrow \neg q}(t) < 1 \qquad (40)$$

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| > 0 \qquad (41)$$

$Episodes(t) \models_G$ "I know that o exhibits P is equivalent to o not exhibiting Q"

$PS(t) \models_G Know(p \Leftrightarrow \neg q)$ holds iff:

$$\left|\overline{C^{p \wedge q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge \neg q}(t)}\right| = 0 \qquad (42)$$

$$\lambda^{p \Leftrightarrow \neg q}(t) \cong 1 \qquad (43)$$

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| > 0 \qquad (44)$$

Protoform $\Omega(\neg p \Leftrightarrow q)$:

$Episodes(t) \models_G$ "It is possible that o not exhibiting P is equivalent to o exhibiting Q"

$PS(t) \models_G Pos(\neg p \Leftrightarrow q)$ holds iff:

$$\left|\overline{C^{p \wedge q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge \neg q}(t)}\right| = 0 \qquad (45)$$

$$0 \leq \lambda^{\neg p \Leftrightarrow q}(t) \leq \lambda^{\Leftrightarrow}_{maxPos} \qquad (46)$$

$Episodes(t) \models_G$ "I believe that o not exhibiting P is equivalent to o exhibiting Q"

$PS(t) \models_G Bel(\neg p \Leftrightarrow q)$ holds iff:

$$\left|\overline{C^{p \wedge q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge \neg q}(t)}\right| = 0 \qquad (47)$$

$$\lambda^{\Leftrightarrow}_{minBel} \leq \lambda^{\neg p \Leftrightarrow q}(t) < 1 \qquad (48)$$

$$\left|\overline{C^{\neg p \wedge q}(t)}\right| > 0 \qquad (49)$$

$Episodes(t) \models_G$ "I know that o not exhibiting P is equivalent to o exhibiting Q"

$PS(t) \models_G Know(\neg p \Leftrightarrow \neg q)$ holds iff:

$$\left|\overline{C^{p \wedge q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge \neg q}(t)}\right| = 0 \qquad (50)$$

$$\lambda^{\neg p \Leftrightarrow q}(t) \cong 1 \qquad (51)$$

$$\left|\overline{C^{\neg p \wedge q}(t)}\right| > 0 \qquad (52)$$

Protoform $\Omega(\neg p \Leftrightarrow \neg q)$:

$Episodes(t) \models_G$ "It is possible that o not exhibiting P is equivalent to o not exhibiting Q"

$PS(t) \models_G Pos(\neg p \Leftrightarrow \neg q)$ holds iff:

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge q}(t)}\right| = 0 \qquad (53)$$

$$0 \leq \lambda^{\neg p \Leftrightarrow \neg q}(t) \leq \lambda^{\Leftrightarrow}_{maxPos} \qquad (54)$$

$Episodes(t) \models_G$ "I believe that o not exhibiting P is equivalent to o not exhibiting Q"

$PS(t) \models_G Bel(\neg p \Leftrightarrow \neg q)$ holds iff:

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge q}(t)}\right| = 0 \qquad (55)$$

$$\lambda^{\Leftrightarrow}_{minBel} \leq \lambda^{\neg p \Leftrightarrow \neg q}(t) < 1 \qquad (56)$$

$$\left|\overline{C^{\neg p \wedge \neg q}(t)}\right| > 0 \qquad (57)$$

$Episodes(t) \models_G$ "I know that o not exhibiting P is equivalent to o not exhibiting Q"

$PS(t) \models_G Know(\neg p \Leftrightarrow \neg q)$ holds iff:

$$\left|\overline{C^{p \wedge \neg q}(t)}\right| = 0 \text{ and } \left|\overline{C^{\neg p \wedge q}(t)}\right| = 0 \qquad (58)$$

Tab. 1. Exemplar database of Episodes - Scenario I

Episodes	P^+	P^-	Q^+	Q^-	R^+	R^-	S^+	S^-
e_1	o	-	-	-	-	o	-	o
e_2	-	-	o	-	-	o	-	o
e_3	o	-	o	-	-	-	o	-
e_4	-	o	-	o	o	-	o	-
e_5	o	-	o	-	o	-	o	-

$$\lambda^{\neg p \Leftrightarrow \neg q}(t) \cong 1 \tag{59}$$

$$\left| \overline{C^{\neg p \wedge \neg q}(t)} \right| > 0 \tag{60}$$

Establishing the resulting set of grounded protoforms directly translates to a set of linguistic summaries with epistemic modalities and natural language connectives of equivalence, i.e. related linguistic representations of the protoforms. As such, it is the final step of the introduced approach. Moreover, the established set of linguistic summaries forms the final data storage, in the proposed architecture, of linguistic summaries (see Fig. 13).

VI. COMPUTATIONAL EXAMPLES

In this section we presented two hypothetical computational scenarios, which help illustrating the presented model. Then we presented an example showing the attempt to apply this model for the purpose of data mining on a popular open dataset *Housevoters* [7]. All these examples are intended to illustrate the process of linguistic summarization to modal equivalences form, divided into subsequent processing steps.

VI. 1. Scenario I

Let the episodic knowledge consist of a set of objects $O = \{o\}$ (single object case), with properties $\Delta = \{P, Q, R, S\}$, and database of episodes $Episodes = \{e_1, e_2, e_3, e_4, e_5\}$ given in Tab. 1 and Tab. 2. Further, let commonsense (natural language) interpretation of symbols p, q, r and s be given as '*Object o exhibits property P*', '*Object o exhibits property Q*', '*Object o exhibits property R*', and '*Object o exhibits property S*', respectively.

For the sake of simplicity let us further assume that the database $Episodes$ remains unchanged throughout all time points $t_0 < t_1 < t_2 < t_3 < t_4 < t_5$, at which linguistic summaries are to be uttered.

Further, we assume the following shallow and deep level division in processing states of available $Episodes$:
$\overline{Episodes}(t_0) = \emptyset, \underline{Episodes}(t_0) = \{e_1, e_2, e_3, e_4, e_5\},$
$\overline{Episodes}(t_1) = \{e_1\}, \underline{Episodes}(t_1) = \{e_2, e_3, e_4, e_5\},$

$\overline{Episodes}(t_2) = \{e_1, e_2\}, \underline{Episodes}(t_2) = \{e_3, e_4, e_5\},$
$\overline{Episodes}(t_3) = \{e_1, e_2, e_3\}, \underline{Episodes}(t_3) = \{e_4, e_5\},$
$\overline{Episodes}(t_4) = \{e_1, e_2, e_3\}, \underline{Episodes}(t_4) = \{e_5\},$
$\overline{Episodes}(t_5) = \{e_1, e_2, e_3, e_4, e_5\}, \underline{Episodes}(t_5) = \emptyset.$
Let us notice that: $\emptyset = \overline{Episodes}(t_0) \subset \overline{Episodes}(t_1) \subset \overline{Episodes}(t_2) \subset \overline{Episodes}(t_3) \subset \overline{Episodes}(t_4) \subset \overline{Episodes}(t_5) = Episodes(t).$

Finally, we assume the following modality thresholds values: $\lambda^{\Leftrightarrow}_{minPos} = 0.0$, $\lambda^{\Leftrightarrow}_{maxPos} = \lambda^{\Leftrightarrow}_{minBel} = 0.5$, and $\lambda^{\Leftrightarrow}_{maxBel} = \lambda^{\Leftrightarrow}_{Know} = 1$. It should be stressed that the assumed modality thresholds fulfil commonsense consistency requirements imposed on linguistic summaries.

An approach to a formal prove of the property can be found elsewhere (e.g. [31]).

Tab. 1 shows a simple knowledge base consisting of 5 episodes. Let us examine the presented model behaviour in given points of time. In particular, let us consider the case of equivalence between the properties P and Q.

In the first step knowledge base is empty $\overline{Episodes}(t_0) = \emptyset$, therefore we can neither confirm nor exclude whether any of the summaries is true. According to the terms outlined in the previous section (Eq. 29, Eq. 30, Eq. 37, Eq. 38, Eq. 45, Eq. 46, Eq. 53, Eq. 54) the summary set is as follows:
$PS(t_0) \models_G \{Pos(p \Leftrightarrow q), Pos(p \Leftrightarrow \neg q), Pos(\neg p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

In the next step t_1 episodic memory getting e_1 ($\overline{Episodes}(t_1) = \{e_1\}$), but it does not contain complete information about properties P and Q, therefore each of the sets $\overline{C^{p \wedge q}(t_1)}, \overline{C^{p \wedge \neg q}(t_1)}, \overline{C^{\neg p \wedge q}(t_1)}, \overline{C^{\neg p \wedge \neg q}(t_1)}$ remains unchanged (so also $\overline{Episodes}(t_1) = \emptyset$) and the result set remains unchanged:
$PS(t_1) \models_G \{Pos(p \Leftrightarrow q), Pos(p \Leftrightarrow \neg q), Pos(\neg p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

This situation is analogous for the t_2:
$PS(t_2) \models_G \{Pos(p \Leftrightarrow q), Pos(p \Leftrightarrow \neg q), Pos(\neg p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

The next episode (e_3) taken into account in subject's memory is classified to $\overline{C^{p \wedge q}(t_3)} = \{e_3\}$. Thus, according to conditions (Eq. 45, Eq. 47, Eq. 50) protoform $\Omega(p \Leftrightarrow \neg q)$ is excluded and according to conditions (Eq. 37, Eq. 39, Eq. 42) protoform $\Omega(\neg p \Leftrightarrow q)$ is also excluded. However, the appearance of this episode reinforces the relative ground-

ing strength:

$$\lambda^{p\Leftrightarrow q}(t_3) = \frac{\left|\overline{C^{p\wedge q}(t_3)}\right| = 1 + \left|\overline{C^{\neg p\wedge\neg q}(t_3)}\right| = 0}{|Episodes(t_3)| = 3} = 0.33$$

(61)

but not enough to fulfilled conditions (Eq. 32, Eq. 33) for protoform $\Omega(p \Leftrightarrow q)$, and conditions (Eq. 56, Eq. 57) for protoform $\Omega(\neg p \Leftrightarrow \neg q)$. Thus the summary is as follows: $PS(t_3) \models_G \{Pos(p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

The situation is slightly different in the next step, this time the episode e_4 goes to $\overline{C^{\neg p\wedge\neg q}(t_4)} = \{e_4\}$ and the relative grounding strengths are computed as follows:

$$\lambda^{p\Leftrightarrow q}(t_4) = \frac{\left|\overline{C^{p\wedge q}(t_4)}\right| = 1 + \left|\overline{C^{\neg p\wedge\neg q}(t_4)}\right| = 1}{|Episodes(t_4)| = 3} = 0.66$$

(62)

$$\lambda^{\neg p\Leftrightarrow\neg q}(t_4) = \frac{\left|\overline{C^{p\wedge q}(t_4)}\right| = 1 + \left|\overline{C^{\neg p\wedge\neg q}(t_4)}\right| = 1}{|Episodes(t_4)| = 3} = 0.66$$

(63)

Under the terms of Eq. 32 and Eq. 56, for both mentioned protoforms, this strength is more than $\lambda^{\Leftrightarrow}_{minBel} = 0.5$, so conditions Eq. 33, Eq. 57 are fulfilled, and grounded set of modal equivalences has the form: $PS(t_4) \models_G \{Bel(p \Leftrightarrow q), Bel(\neg p \Leftrightarrow \neg q)\}$

The appearance of last episode e_5 increase cardinality of holon in the subset $\overline{C^{p\wedge q}(t_5)} = \{e_3, e_5\}$. In this step we processed the whole database of episodes and:

$$\lambda^{p\Leftrightarrow q}(t_5) = \lambda^{\neg p\Leftrightarrow\neg q}(t_5) = 1.0$$

(64)

Thus, according to conditions (Eq. 35, Eq. 59, Eq. 33, Eq. 57), the final set is as follows: $PS(t_5) \models_G \{Know(p \Leftrightarrow q), Know(\neg p \Leftrightarrow \neg q)\}$

This short example illustrates the process of generation linguistic summarisation as modal equivalences. This process presents the range from the possibility level of all protoforms, through excluding two of them to knowing that $p \Leftrightarrow q$ and $\neg p \Leftrightarrow \neg q$ occurs.

VI. 2. Scenario II

Let us assume identical circumstances as in the first scenario and analyze the knowledge base given in Tab. 2.

In the first step t_0, as in the previous case, we assume the possibility of all four protoforms: $PS(t_0) \models_G \{Pos(p \Leftrightarrow q), Pos(p \Leftrightarrow \neg q), Pos(\neg p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

In the next step, t_1, including episode e_1, we increase the cardinality of holon subset $\overline{C^{p\wedge q}(t_1)} = \{e_1\}$, and the relative grounding strengths are as follows:

$$\lambda^{p\Leftrightarrow q}(t_1) = \lambda^{\neg p\Leftrightarrow\neg q}(t_1) = 0.2$$

(65)

$$\lambda^{p\Leftrightarrow\neg q}(t_1) = \lambda^{\neg p\Leftrightarrow q}(t_1) = 0$$

(66)

Simultaneously, according to the conditions (Eq. 45, Eq. 47, Eq. 50) protoform $\Omega(p \Leftrightarrow \neg q)$ is excluded and according with the conditions (Eq. 37, Eq. 39, Eq. 42) protoform $\Omega(\neg p \Leftrightarrow q)$ is also excluded. Accordingly, the summary at this point contains: $PS(t_1) \models_G \{Pos(p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

The next episode (e_2) recalled by the subject's memory increases the cardinality of set $\left|\overline{C^{\neg p\wedge\neg q}(t_2)}\right|$ to 2 and increases the relative grounding strengths $\lambda^{p\Leftrightarrow q}(t_2)$ and $\lambda^{p\Leftrightarrow\neg q}(t_2)$ to 0.4. However, according to the conditions (Eq. 32, Eq. 56), it is still lower than $\lambda^{\Leftrightarrow}_{minBel} = 0.5$. Thus, the resultant set remains unchanged: $PS(t_2) \models_G \{Pos(p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

Subsequently, by adding e_3, the belief in equivalence $p \Leftrightarrow q$ is strengthened. Therefore, the relative grounding strengths are $\lambda^{p\Leftrightarrow q}(t_3) = \lambda^{\neg p\Leftrightarrow\neg q}(t_3) = 0.6 > \lambda^{\Leftrightarrow}_{minBel}$. The conditions Eq. 31 and Eq. 33 are fulfilled, so in the case of protoform $\Omega(p \Leftrightarrow q)$ we can use linguistic label Bel. But in the case of protoform $\Omega(\neg p \Leftrightarrow \neg q)$ the condition Eq. 57 is still not fulfilled. As such, the label will remain at the same level of certainty – possibility Pos: $PS(t_3) \models_G \{Bel(p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)\}$

The subsequent episode (e_4) is classified into a holon subset $\overline{C^{p\wedge\neg q}(t_4)} = \{e_4\}$. This slightly complicates matters. Grounding of the correct, non-contradictory (from the classical logic stance) summary set requires additional conditions to be fulfilled (Eq. 29, Eq. 37, Eq. 45, Eq. 53). Unfortunately, none of these conditions is satisfied, therefore the resulting set is empty: $PS(t_4) \models_G \emptyset$

Tab. 2. Exemplar database of Episodes - Scenario II

Episodes	P^+	P^-	Q^+	Q^-	R^+	R^-	S^+	S^-
e_1	o	-	o	-	-	o	-	o
e_2	o	-	o	-	-	o	-	o
e_3	o	-	o	-	-	-	o	-
e_4	o	-	-	o	o	-	o	-
e_5	o	-	o	-	o	-	o	-

In the last step, despite the strengthening set $\overline{C^{p \wedge q}}(t_5)$, the resulting set is still empty:

$$PS(t_5) \models_G \emptyset$$

This is due to the fact that none of the previously mentioned conditions cannot be met.

This scenario shows that the presented model may not always generate some result. Sometimes a set of data do not allow to ground the statement in a form of equivalence, because these data not contain equivalence relationships, or these equivalences are mutually in contradiction.

VI. 3. Real data computation

This section presented a scenario applying the modal equivalences grounding model on dataset *Housevoters* [7], which represent 1984 United Stated Congressional Voting Records, classified as Republican or Democrat.

This data set includes votes for each of the U.S. House of Representatives Congressmen on the 16 key votes identified by the CQA. The CQA lists nine different types of votes: voted for, paired for, and announced for (these three simplified to yea), voted against, paired against, and announced against (these three simplified to nay), voted present, voted present to avoid conflict of interest, and did not vote or otherwise make a position known (these three simplified to an unknown disposition).

This dataset contains the following attributes:
1. Class Name: 2 (democrat, republican)
2. handicapped-infants: 2 (y,n)
3. water-project-cost-sharing: 2 (y,n)
4. adoption-of-the-budget-resolution: 2 (y,n)
5. physician-fee-freeze: 2 (y,n)
6. el-salvador-aid: 2 (y,n)
7. religious-groups-in-schools: 2 (y,n)
8. anti-satellite-test-ban: 2 (y,n)
9. aid-to-nicaraguan-contras: 2 (y,n)
10. mx-missile: 2 (y,n)
11. immigration: 2 (y,n)
12. synfuels-corporation-cutback: 2 (y,n)
13. education-spending: 2 (y,n)
14. superfund-right-to-sue: 2 (y,n)
15. crime: 2 (y,n)
16. duty-free-exports: 2 (y,n)
17. export-administration-act-south-africa: 2 (y,n)

The Tab. 3 presents the results of software computation for proforms $\Omega(Physician\_fee\_freeze \Leftrightarrow Adoption\_budget\_resolution)$ divided into various stages of data processing.

As in the previous scenarios, the model starts allowing possibility of occurrence all protoforms $\Omega(pff \Leftrightarrow abr)$, $\Omega(pff \Leftrightarrow \neg abr)$, $\Omega(\neg pff \Leftrightarrow abr)$, $\Omega(\neg pff \Leftrightarrow \neg abr)$.

Over time (consequently recalling next episodes to the holon) protoforms $\Omega(pff \Leftrightarrow \neg abr)$ and $\Omega(\neg pff \Leftrightarrow abr)$ are excluded. Then strengthened are $\Omega(pff \Leftrightarrow abr)$ and $\Omega(\neg pff \Leftrightarrow \neg abr)$, but not so that the relative grounding strengths exceed the $\lambda^{\Leftrightarrow}_{minBel} = 0.5$.

Tab. 3. Results for *Physician_fee_freeze(pff)* and *Adoption_budget_resolution(abr)*

Process	Result set
0%	$\{Pos(pff \Leftrightarrow abr), Pos(pff \Leftrightarrow \neg abr),$ $Pos(\neg pff \Leftrightarrow abr), Pos(\neg pff \Leftrightarrow \neg abr)\}$
10%	$\{Pos(pff \Leftrightarrow abr), Pos(\neg pff \Leftrightarrow \neg abr)\}$
20%	$\{Pos(pff \Leftrightarrow abr), Pos(\neg pff \Leftrightarrow \neg abr)\}$
30%	$\{Pos(pff \Leftrightarrow abr), Pos(\neg pff \Leftrightarrow \neg abr)\}$
40%	$\{Pos(pff \Leftrightarrow abr), Pos(\neg pff \Leftrightarrow \neg abr)\}$
50%	$\{Pos(pff \Leftrightarrow abr), Pos(\neg pff \Leftrightarrow \neg abr)\}$
60%	$\{Bel(pff \Leftrightarrow abr), Bel(\neg pff \Leftrightarrow \neg abr)\}$
70%	$\{Bel(pff \Leftrightarrow abr), Bel(\neg pff \Leftrightarrow \neg abr)\}$
80%	$\emptyset$
90%	$\emptyset$
100%	$\emptyset$

Such a situation takes place after processing 60% of dataset, so then the grounded set is as follows: $\{Bel(pff \Leftrightarrow abr), Bel(\neg pff \Leftrightarrow \neg abr)\}$.

During further processing ($> 70\%$) we can notice that certain contradictions appear (from the logical consistency point of view), therefore the resulting set is empty.

Tab. 4 represents a case where the data is divided into particular stages of processing. In particular, it contains the results of performed software-based computation for protoform $\Omega(El\_salvador\_aid \Leftrightarrow Aid\_to\_nicaraguan\_contras)$ limited only to *Democrat* class.

In the beginning the model cannot exclude any of the possibilities: $\Omega(esa \Leftrightarrow anc)$, $\Omega(esa \Leftrightarrow \neg anc)$, $\Omega(\neg esa \Leftrightarrow anc)$, $\Omega(\neg esa \Leftrightarrow \neg anc)$. But consequently recalling next episodes to the holon, protoforms $\Omega(esa \Leftrightarrow \neg anc)$ and $\Omega(\neg esa \Leftrightarrow anc)$ are excluded.

Then sets $\Omega(esa \Leftrightarrow anc)$ and $\Omega(\neg esa \Leftrightarrow \neg anc)$ are enhanced, so the relative grounding strengths also, but not enough to exceed required threshold $\lambda^{\Leftrightarrow}_{minBel} = 0.5$.

This situation occurs until processing reaches 40% of episodes. Up to that point the result set remains unchanged: $\{Pos(esa \Leftrightarrow anc), Pos(\neg esa \Leftrightarrow \neg anc)\}$.

Unfortunately, all further processing ($> 40\%$) cannot avoid contradicting protoforms (Eq. 29, Eq. 37, Eq. 45, Eq. 53). As such, the model cannot ground any of the linguistic statements using natural language connectives of equivalence, i.e. the linguistic data result is an empty set.

VI. 4. Discussion

In conclusion, all presented computational examples depict the behaviour of the process of establishing a linguistic summary in the form of modal conditionals of equivalence at various stages of processing. Using the above examples

we illustrate the behaviour of the grounding process for two classes of hypothetical scenarios, i.e. case of establishing the highest epistemic modality and case of inability to establish particular epistemic modality.

Tab. 4. Results for *El_salvador_aid(esa)* and *Aid_to_nicaraguan_-contras(anc)* for Democrats

Process	Result set
0%	$\{Pos(esa \Leftrightarrow anc), Pos(esa \Leftrightarrow \neg anc),$ $Pos(\neg esa \Leftrightarrow anc), Pos(\neg esa \Leftrightarrow \neg anc)\}$
10%	$\{Pos(esa \Leftrightarrow anc), Pos(\neg esa \Leftrightarrow \neg anc)\}$
20%	$\{Pos(esa \Leftrightarrow anc), Pos(\neg esa \Leftrightarrow \neg anc)\}$
30%	$\{Pos(esa \Leftrightarrow anc), Pos(\neg esa \Leftrightarrow \neg anc)\}$
30%	$\{Pos(esa \Leftrightarrow anc), Pos(\neg esa \Leftrightarrow \neg anc)\}$
40%	$\{Pos(esa \Leftrightarrow anc), Pos(\neg esa \Leftrightarrow \neg anc)\}$
50%	$\emptyset$
60%	$\emptyset$
70%	$\emptyset$
80%	$\emptyset$
90%	$\emptyset$
100%	$\emptyset$

Analysing the results in the case of establishing the highest epistemic modality (Scenario I) we can notice that the underlying initial uncertainty experienced by an agent is reflected in the grounded statements: $Pos(p \Leftrightarrow q), Pos(p \Leftrightarrow \neg q), Pos(\neg p \Leftrightarrow q), Pos(\neg p \Leftrightarrow \neg q)$. In particular, the inherent initial strength of a deep level of cognition and the early inability to establish direct correlation between the properties P and Q affects the agent's ability to reason about equivalence $P \Leftrightarrow Q$. Only after perceiving a particular state, where both properties are exhibited, is the agent allowed to establish higher levels of certainty expressed by the higher epistemic modality.

In the case of inability to establish particular epistemic modality (Scenario II) we can notice that the early underlying certainty experienced by an agent is similar to the aforementioned case - mostly due to the inherent initial strength of a deep level of cognition. Furthermore, in this scenario the situation involves the agent to process, in the shallow level of cognition, an opposing episode, where the property P is exhibited and the property is not exhibited Q. As such, the agent has actively registered a counterexample of the equivalence $P \Leftrightarrow Q$.

Furthermore, the behaviour of the grounding process applied to a classical data mining dataset Housevoters uniquely pinpoints the influence of shallow and deep levels of cognition on the ability to ground and establish particular equivalence statements. In particular, we can notice that with the increasing number of processed episodes (here votes) the system is limiting the effects of the deep level and thus is able to establish higher certainty levels (from Pos to Bel). However, whenever a contradictory episode is analysed in the shallow level then the system is certain that particular equivalence is not suitable to represent the gathered information.

As such, we can notice the underlying nature of the proposed class of linguistic summaries, particularly the fact that they are used to communicate a specific type of knowledge vagueness. Namely, the summaries are used to communicate epistemic uncertainty that results from subjectively experienced knowledge incompleteness. In particular, it directly results from the scenarios that within the presented approach the individual system has to be treated as the sole owner of the summarizing processes. Therefore, there is an obvious necessity to explicitly distinguish between purely private and internal (cognitive) structures of autonomous system, and all entities of an external world (a domain). Only in such settings is it possible to introduce proper mechanisms of extracting linguistic summaries in which natural language connectives of equivalence are combined with a dedicated label of epistemic modality.

VII. CONCLUSIONS

In this paper we have discussed selected issues on an original approach to extraction of linguistic summaries by autonomous computational agents. In particular, a case of (semi-)natural language statements built with epistemic modalities and connectives of equivalence was considered as a special case of a broader (semi-)natural language of epistemic modalities. As compared to the well-known classical interpretation of linguistic summaries, based on the fuzzy set theory, the underlying nature of the proposed class of linguistic summaries is used to communicate a fundamentally different type of knowledge vagueness. Namely, the classical approach covers a case of fuzziness understood as a lack of crisp borders of concepts, whilst here the summaries are used to communicate epistemic uncertainty, i.e. uncertainty resulting from subjectively experienced knowledge incompleteness. In conclusion, the proposed model deals with a different class of linguistic concepts, namely modal operators of *knowledge*, *belief* and *possibility*.

It was reasoned out why combining epistemic modalities and sentence connectives of equivalence are rather an unusual case of natural language statement processing. Namely, it was argued (illustrated in a realistic knowledge management motivating scenario in Sec. III.) that the commonsense meaning of the class of statements inherits phenomena closely related to the concept of *implicit* (so-called *deep*) cognition, the latter being rather incompatible with the nature of artificial computational agents. However, it was also shown (e.g. by the aforementioned motivating scenario) that in certain practical contexts characterized with

constrained access to collected empirical data, an interactive computational agent, capable of (semi-)natural language processing, should be equipped with adequate technical counterparts.

It should be noted that the reported research belongs to a broader long-term R&D project within which modal linguistic summaries are studied and utilized, along with the concept of an artificial cognitive system.

Acknowledgement

References

[1] A. Ramos-Soto, A. Bugarín, S. Barro, *On the role of linguistic descriptions of data in the building of natural language generation systems*, Fuzzy Sets and Systems **285**, pp. 31–51 (2016).

[2] A. Włodarczyk, *Grounding of the meta-informative status of utterances*, in Meta-informative Centering in Utterances: Between Semantics and Pragmatics, A. Włodarczyk and H. Włodarczyk, Eds. Amsterdam: John Benjamins Publishing Co., pp. 41–58, (2013).

[3] C. Ogden, I. Richards, *The meaning of meaning*, Harcourt, Brace, 1923.

[4] C. Peirce, *Collected Papers of C. S. Peirce*, Harvard University Press, Cambridge, 1932.

[5] D. Dennett, *The Intentional Stance*, MIT Press, 1989.

[6] D. Dennett, *True believers: The intentional strategy and why it works*, In Mental Representation: A Reader. Blackwell, 1994.

[7] Dataset Housevoters, *archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records*, Congressional Quarterly Almanac, 98th Congress, 2nd session 1984, Volume XL: Congressional Quarterly Inc. Washington, D.C., 1985.

[8] E. Aarts, B. de Ruyter, *New research perspectives on Ambient Intelligence.* Journal of Ambient and Smart Env. **1**, 50-14 (2009).

[9] F.J. Stachowiak, *Tracing the role of memory and attention for the meta-informative validation of utterances*, in Meta-informative Centering in Utterances: Between Semantics and Pragmatics, A. Włodarczyk and H. Włodarczyk, Eds. Amsterdam: John Benjamins Publishing Co., pp. 121-142 (2013).

[10] G. Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, Chicago: University of Chicago, 1987.

[11] G. Moysea, M.-J. Lesota, *Linguistic summaries of locally periodic time series*, Fuzzy Sets and Systems **285**, 94–117 (2016).

[12] G. Underwood, *Implicit Cognition*, Oxford University Press, 1996.

[13] J. Kacprzyk, *Computing with words is an implementable paradigm: fuzzy queries, linguistic data summaries, and natural-language generation*, IEEE Trans. Fuzzy Syst. **18**(3), 451–472 (2010).

[14] J. Kacprzyk, A. Wilbik, S. Zadrożny, *An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation*, Int. J. Intell. Syst. **25**(5), 411–439 (2010).

[15] J. Kacprzyk, A. Wilbik, S. Zadrożny, *Linguistic summarization of time series using a fuzzy quantifier driven aggregation*, Fuzzy Sets Syst. **159**(12), 1485–1499 (2008).

[16] J. Kacprzyk, R.R. Yager, *Linguistic summaries of data using fuzzy logic*, International Journal of General Systems **30**, 133–154 (2001).

[17] J. Kacprzyk, S. Zadrożny, *Computing with words in intelligent database querying: standalone and Internet-based applications*, Information Sciences **34**, 71–109 (2001).

[18] J. Kacprzyk, S. Zadrożny, *Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools*, Information Sciences **173**, 281–304 (2005).

[19] J. Nuyts, *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective*, Amsterdam: John Benjamins Publishing Co., 2001.

[20] L.A. Zadeh, *The concept of linguistic variable and its applications to approximate reasoning - I (II, III)*, Information Sciences **8**(8,9), 199–249 (301-357, 43-80) (1975).

[21] L. Talmy, *Toward a Cognitive Semantics*, Cambridge, MA: MIT Press, 2000.

[22] L.A. Zadeh, *A computational approach to fuzzy quantifiers in natural languages*, Computers and Mathematics with Applications **9**,149–184 (1983).

[23] L.A. Zadeh, *A prototype-centered approach to adding deduction capabilities to search engines – the concept of a protoform*, BISC Seminar, University of California, Berkeley, 2002

[24] L.A. Zadeh, *From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions*, Int. J. Applied Mathematics and Computer Science **12**(3), 307–324 (2002).

[25] L.A. Zadeh, *Fuzzy logic = computing with words*, IEEE Trans. Fuzzy Syst. **4** (2), pp. 103–111, 1996.

[26] L.A. Zadeh, *Outline of a new approach to the analysis of complex systems and decision processes*, IEEE Trans. Systems Man Cybernet. SMC-3, pp. 28–44, 1973.

[27] P. Smets, *Varieties of ignorance*, Information Sciences **57-58**, 135-144 (1991).

[28] P.N. Johnson-Laird, *Mental models. Towards a cognitive science of language, inference, and consciousness*, Cambridge Uni. Press, 1983.

[29] R. Katarzyniak, *On some properties of grounding nonuniform sets of modal conjunctions.* International Journal of Applied Mathematics and Computer Science **16**(3), 399–412 (2006).

[30] R. Katarzyniak, *On some properties of grounding simple modalities*, Systems Science **31** (3), 59–86 (2005).

[31] R. Katarzyniak, *On some properties of grounding uniform sets of modal conjunctions.* Journal of Intelligent Fuzzy Systems **17**(3), 209–218 (2006).

[32] R. Katarzyniak, *Some notes on grounding singletons of modal conjunctions*, Systems Science **32**(1), 45–55 (2005).

[33] R. Katarzyniak, D. Więcek, *An approach to extraction of linguistic recommendation rules – application of modal conditionals grounding*, Lecture Notes in Computer Science Lecture Notes in Artificial Intelligence, vol. 7653, pp. 249-258, 2012.

[34] R. Katarzyniak, D. Więcek, *Modal equivalences as linguistic summarisation of data resources*, [In:] *Information Systems Architecture and Technology: Intelligent Information Systems, Knowledge Discovery, Big Data and High Performance Computing*, Wrocław University of Technology, Wrocław (Poland), pp. 23-32, 2013.

[35] R. Katarzyniak, N.T. Nguyen and J.C. Jain, *A model for fuzzy grounding of modal conjunctions in artificial agents*, Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 4953, pp. 341-350, 2008.

[36] R. Katarzyniak, N.T. Nguyen and J.C. Jain, *Soft computing approach to contextual determinantion of grounding sets for simple modalities*, Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 4692, pp. 230-237, 2007.

[37] R. Katarzyniak, W. Lorkiewicz, J. Sobecki *An architecture for verbal summaries in smart mobile devices*, 2015 IEEE 15th International Conference on Data Science and Data Intensive Systems, Sydney, Australia. Piscataway, NJ: IEEE, pp. 540-547, 2015.

[38] R. Katarzyniak, W. Lorkiewicz, J. Sobecki *A model for linguistic summaries of results from usability studies*, 2015 IEEE 15th International Conference on Data Science and Data Intensive Systems, Sydney, Australia. Piscataway, pp. 103-110, NJ: IEEE, 2015.

[39] R. Katarzyniak, W. Lorkiewicz, O. Krejcar, *Linguistic alerts in information filtering systems: towards technical implementations of cognitive semantics*, 2016 ICEIS 18th International Conference on Enterprise Information Systems, 25-28 April 2016 Rome, Italy, pp. 512-519, SCITEPRESS, 2016.

[40] R. Kitchin, *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences*, Sage Pub. Ltd., 2014.

[41] R.J.G.B. Campello, W.C. Amaral, *Modeling and linguistic knowledge extraction from systems using fuzzy relational models*, Fuzzy Sets and Systems **121**, 113–126 (2001).

[42] R.M. Tong, *Synthesis of fuzzy models for industrial processes – some recent results*, Internat. J. General Systems **4**, 143–162 (1978).

[43] R.P. Katarzyniak, *Extracting Modal Implications and Equivalences from Cognitive Minds*, Proc. of the 7th Int. Conf. Discovery Science – DS 2004, 2-5 October 2004, Padova (Italy), LNCS, vol. 3245, pp. 420-428, 2004.

[44] R.P. Katarzyniak, *The Language Grounding Problem and its Relation to the Internal Structure of Cognitive Agents*, Journal of Universal Computer Science **11**(2), 357–374 (2005).

[45] R.P. Katarzyniak, W.A. Lorkiewicz, D. Więcek, *Modal linguistic summaries based on natural language equivalence with cognitive semantics*, Proc. 12th Int. Conf. on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE, pp. 1904-1909, 2016.

[46] R.P.Katarzyniak, G. Popek, *Integration of modal and fuzzy methods of knowledge representation in artificial agents*, International Journal of Software Engineering and Knowledge Engineering **23**(1), 13–29 (2012).

[47] R.R. Yager, *A new approach to summarization of data*, Information Sciences **28**, 69–86 (1982).

[48] S. Harnad, *The symbol grounding problem*, Physica D: Nonlinear Phenomena **42**(1-3), 335–346 (1990).

[49] T. Takagi, M. Sugeno, *Fuzzy identification of systems and its applications to modeling and control*, IEEE Trans. Systems Man Cybernet. SMC-15, pp. 116–132, 1985.

[50] W. Minker, R. López-Cózar, M. McTear, *The role of spoken language dialogue interaction in intelligent environments.* Journal of Ambient and Smart Environments **1**, 31–36 (2009).

[51] W. Pedrycz, *An identification algorithm in fuzzy relational systems*, Fuzzy Sets and Systems **13**, 153–167 (1984).

Radosław Katarzyniak Personal Data: born 01 June 1966, Kalisz, Poland; Current Position: Professor of Computer Science, Wrocław University of Science and Technology (WrUST), Faculty of Computer Science and Management, Wrocław, Poland; Head of Computational Semiotics and Interactive Systems Group; Academic Education: 1991: M.Sc. in Computer Science from WrUST; 1999: PhD in Computer Science from WrUST; 2009: D.Sc. (Habilitation) in Computer Science from AGH University of Science and Technology, Poland; Professional Activities: organizing chair of multiple conferences (IEA-AIE'08; International Conference on Collective Computational Intelligence – ICCCI'09, ICCCI'10, ICCCI'11; Asian Conference on Intelligent Information and Database Systems – ACIIDS'09, ACCIDS'10); editor and co-editor of monographs on application of computational intelligence methods to knowledge management; guest editor of special issues in international journals on multiagent systems and computational intelligence applications; author of about 150 research journal and conference papers; Current Research Interests: autonomous systems, cognitive linguistics, natural language engineering, knowledge management, computational intelligence.
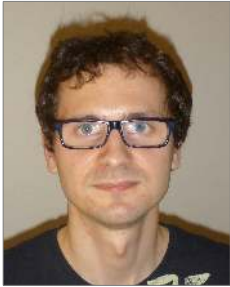
Wojciech Lorkiewicz Personal Data: born 02 March 1983, Wrocław, Poland; Current Position: Assistant Professor, Wrocław University of Science and Technology (WrUST), Department of Informatics, Wrocław, Poland; Member of Computational Semiotics and Interactive Systems Group, WrUST; Academic Education: 2007: M.Sc. in Computer Science from WrUST, 2013: PhD in Computer Science from WrUST, and Swinburne University of Technology, FICT, Melbourne, Australia; Professional Activities: Organising Committee member of several international conferences (IEA-AIE'08, ICCCI'09, ACIIDS'09, ICCCI'11, ACIIDS'11, ICCCI'12), Member of KES Focus Group on Agent and Multi-agent Systems; Author and co-author of about 25 research journal and conference papers; Current Research Interests: artificial and ambient intelligence – agent and multi-agent systems modelling; artificial cognition – semantic communication, language processing and evolution, computational cognition; knowledge management systems – computational methods in knowledge processing and information retrieval.

Dominik Więcek Personal Data: born 31 January 1987, Syców, Poland; Current Position: PhD Student, Wrocław University of Science and Technology (WrUST), Department of Informatics, Wrocław, Poland; Senior Software Developer at Applications and Systems Department, Senior IT Assistant at Wrocław Centre for Networking and Supercomputing, WrUST; Member of Computational Semiotics and Interactive Systems Group, WrUST; Academic Education: 2011: M.Sc. in Computer Science from WrUST; Current Research Interests: artificial and ambient intelligence – agent and multi-agent systems' research and development, computational cognition – semantic communication, natural language engineering; knowledge management.