Are Objects Ontologically Dependent on Processes?

(Extended Abstract)

Antony Galton[1]

1 Substance Ontology *vs* Process Ontology

The dominant ontological paradigm throughout Western philosophy has been the Aristotelian *substance ontology*, according to which substances — including what we would nowadays more generally refer to as *objects* or *things* — are independent entities ("first-class citizens" of the ontology), whereas processes and events are "accidents", that is, "second-class" entities that are dependent for their existence on the objects that enact or participate in them.

Despite its dominance, the substance ontology has not gone unquestioned. Repeatedly in the history of philosophy there have been thinkers who have sought to elevate *processes* to the rank of first-class independently existing ontological elements, sometimes even reversing the dependence to make objects ontologically dependent on processes. Indeed, this line of thinking had already appeared before Aristotle, being traceable at least as far as Heraclitus, more than a century earlier.

In more modern times, some form of process ontology has been advocated by thinkers such as Whitehead [17], Heidegger [8], Rescher [13], and Dupré [2], amongst others, and has particularly found favour amongst philosophers for whom the primary focus of interest is living organisms, as opposed to, for example, human artefacts. Outside Western philosophy, Buddhist philosophy in particular has had a long history of treating processes as more fundamental than substances [12].

My purpose in this talk is to examine a number of considerations that I consider to be of key relevance to the debate between substance ontologists and process ontologists. My general orientation will be towards the latter, but before presenting a positive case I begin by drawing attention to what I believe is a serious problem with substance ontology.

2 The Cinematographic Model of Reality

If the fundamental elements of reality are substances, how are processes and change to be accounted for? In mainstream thinking, the standard account of change is somewhat as follows. At each instant of its existence, an object has certain properties; but it need not have the same properties at every instant. Change is what happens when an object x has some property P at some time t_1, but lacks that property at some distinct time t_2:

$$\exists t_1 \exists t_2 \; P(x, t_1) \wedge \neg P(x, t_2).$$

Suppose at a given instant t we take together all the properties of all the objects existing at that instant, and call that ensemble the state of the world, or *world-state* that holds at t: call this $ws(t)$. On this view, the history of the world essentially consists of the mapping ws from instants to world-states. Each world-state is itself unchanging, and the existence of change in the world reduces to differences between successive world-states.

I call this the *cinematographic model of reality*. In the cinema we think we see moving images on the screen. But we know that in fact nothing moves: what we are seeing is a succession of still images in each of which there is no change or movement. The view of change described above, which is naturally associated with the substance ontology, works in the same way. Motion and change are reduced to some kind of "illusion" produced by stringing together world-states from which motion and change are absent.

This a problem for us if we believe that changes in the world are continuous, as expressed by Leibniz in the catchphrase *natura non facit saltus* ("nature does not make jumps"). Continuity is important for substance ontology because it provides one of the major underpinnings of the concept of *identity*. If asked what it means to say that the person you see in front of you now is the *same* person as the one you saw entering the room at the beginning of the talk, you may reply in various ways, but most probably you will invoke some form of continuity, be it spatio-temporal, material, or psychological.

On the cinematographic view, we require the functions ascribing quantitative property values such as position to substances to be continuous. What this means is that by focussing on sufficiently small time periods, we can make the changes that we see as small as we like. This is what is expressed in the standard mathematical definition of continuity for a numerical function:

$$\forall t \, \forall \epsilon > 0 \, \exists \delta > 0 \, \forall t' \, (|t - t'| < \delta \rightarrow |f(t) - f(t')| < \epsilon).$$

Now the problem is that this formula — which expresses a *mathematical* notion of continuity — only comes close to capturing something that resembles the physical continuity we believe we see in the world if the set T of times over which the variables t and t' range satisfies the Dedekind completeness condition: any set of numbers with an upper bound has a least upper bound. This in turn means that T must be modelled by something like the set of real numbers $\mathbb{R}$.[2] You can easily see what goes wrong with, say, the rational numbers ($\mathbb{Q}$) by constructing a function which maps x to 1 if $x^2 < 2$ and to 0 otherwise; this function is continuous on $\mathbb{Q}$ but gives a very bad model of any physically continuous variation.

There are two reasons why we should feel uncomfortable about having to use the real numbers to model space and time. The first is

[1] University of Exeter, UK, email: apgalton@ex.ac.uk

[2] Perhaps it does not have to be the full non-denumeratble set of real numbers: maybe a denumerable set such as the computable numbers, or the real algebraic numbers, which satisfy the first-order form of Dedekind completeness (quantifying over first-order expressible sets only) but not the full second-order form, would do.

that they require us to admit the existence of *actual infinities*. Within any time interval, however short, there must be infinitely many instants, one for each real number falling between the numbers assigned to the endpoints of the interval — and likewise with spatial regions. The second problem is how we are to account for the facts of *duration* (in time) and *extent* (in space). If a time interval is nothing but an agglomeration of durationless instants, where does its duration come from? However many zeros you add together, you will never get anything other than zero; and yet we are invited to suppose that the positive duration of a time interval somehow emerges from the combination of all the instants it contains.

None of this means that we should give up using real number mathematics for the purposes of science. The use of $\mathbb{R}$ to model physical continua may be a metaphysical fiction — indeed a metaphysical absurdity — but it is the only way we know of providing a formal underpinning for the kinds of calculations needed to obtain quantitatively meaningful results about the real world. It supports the "unreasonable effectiveness of mathematics in the natural sciences" [18]. Despite that, I believe the foregoing considerations render the traditional substance ontology untenable.

3 Taking Time and Change Seriously

In the argument presented above, the notion of *duration* played a key role; this is as it should be, for surely duration is the very essence of time. That is why it is so absurd to try to build time out of durationless instants. Furthermore, the essence of duration is surely *change*: any duration has parts, and without change, how can one part of a duration be distinguished from another?

Although I do not postulate any such thing as a minimum duration in actuality, subjectively the present moment seems to be a kind of minimum duration, reflecting the limit of the temporal resolution of our immediate experience. By technological means, we are able to measure temporal separations much shorter than we can discriminate ourselves, but no technology can identify for us an absolute minimal duration. What is crucial here is that time always comes endowed with duration; durationless time seems utterly contradictory. And therefore time depends on change: a temporal duration completely lacking in change also seems contradictory.

It follows from this that the present, which is part of time given in experience, encompasses change. This is radically opposed to the standard picture of real-number time, in which the present is portrayed as a durationless instant and therefore completely lacking in change. I propose therefore that we should embrace the notion of the *dynamic present*. A 'snapshot' of the world, if such can be conceived, so far from being the static 'time slice' postulated by conventional ontologies, itself contains motion and change: the processes of the world are active at all times of their existence. Elsewhere I have advocated a view of processes as being, if not actual continuants, at least in important respects *continuant-like* [5, 6, 7]. Processes are wholly present at each time of their existence, and have attributes which themselves can undergo change, resulting in *metaprocesses*, such as the fluctuation in loudness of a noise, or the acceleration of a motion. This is fully in accord with the picture I am painting here.

In order to allay possible misunderstanding, I must clarify that I am here using the word 'process' to denote something that is ongoing, open-ended, a pattern of activity rather than a completed episode with a beginning and an ending. It is the sense of process that is aligned with *activity* in Vendler's well-known classification of verb meanings into achievements, accomplishments, states and activities [16]; in [11], these are called processes. It also corresponds to the category 'process' in the DOLCE ontology. This is *not* the same as the 'processes' of BFO [1]. A BFO process is a four-dimensional entity, similar to the completed episodes of the kind I mentioned earlier; in more standard terminology this could be described as an *event* rather than a process.

It cannot be stressed too strongly that most of what I say about processes in this talk would be false or unintelligible if 'process' were to be understood as meaning a BFO process. A process ontology, BFO-style, would be *perdurantism*: like the substance ontology (or *endurantism*), it too is unable really to come to terms with the notion of real change. Whereas endurantism must simulate change cinematographically by compounding together infinitely many changeless instants, perdurantism reduces change to difference amongst the temporal parts of a four-dimensional whole which represents a substance by means of its entire life-history. In the kind of process ontology I am advocating, change is fundamental: it is a primitive given, not to be explained away in terms of changelessness.[3]

4 Object, Process and Event as Ontological Aspects

In [6] I defined two contrasting, and complementary, perspectives we use in describing the world. On the one hand, we may describe reality in terms of how it presents itself to us in experience, taking, in other words, a present-tense point of view. I call this the *experiential perspective*; forms of description appropriate to this perspective may be said to present elements of reality under an experiential *aspect*. In this aspect we see objects and ongoing processes. On the other hand, we can as it were stand back from the immediately experienced world and present a synoptic view, integrating over a period of time to form temporally extended episodes that we call events. This is called the *historical aspect*. The use of the term 'aspect' here reflects its use in linguistics to denote distinctions such as that between perfective and imperfective forms of the verb, which typically correlate with the historical and experiential aspects respectively.

Within the experiential perspective we find both objects and processes, but contrary to the impression given by the substance ontology, it is not always easy to distinguish between these two categories. In previous work [3, 4] I drew attention to what I called *dual-aspect* or *multi-aspect* phenomena, such as rivers, waterfalls, ocean currents, waterspouts, hurricanes, and tornadoes; depending on how one views these, they may equally well be described as objects, processes, or events. A hurricane, for example, as captured in a series of satellite images, has size, shape, position, and motion along a trajectory, all features typically associated with objects; but to the unfortunate villager in the path of the hurricane, it is not an object but a process, a ferocious rushing of the wind, wreaking destruction as it passes by; or else, taking the duration of its passing as a whole, it is an event.

There are, indeed, complex interdependencies between the objects and processes we see in the world. In the spirit of [7], we can readily recognise that a human being, for example, is dependent on its metabolic processes by which the living fabric of the body is maintained in working order; but these processes are themselves dependent on the molecules which participate in them. They in turn are dependent on the atomic and subatomic processes which hold them together. For a process philosopher the nature of the dependence of an object on processes might best be described in terms of constitution: the processes are *constitutive* of the object. This does not mean that the object 'does not exist' but rather tells us something important

[3] I am indebted to Anne Sophie Meincke for helping me to clarify my understanding of the deficiencies of both endurantism and perdurantism.

about what it means *for* an object to exist: in effect, it suggests that existence is a process rather than a state.

5 An object image schema

If 'object' and 'process' are to be understood as aspects of the same experiential reality then an appropriate approach to explaining their relationship might be to examine the cognitive apparatus through which we apprehend and experience them. Here I explore the potential of the *image schema* as a tool to aid us in this examination. An image schema [9, 10] is a recurring "semi-abstract" pattern by which we can mould our raw experiences into a structured understanding of the world. As Johnson [9] puts it, "image schemata operate at a level of mental organization that falls between abstract propositional structures, on the one side, and particular concrete images, on the other". Typical examples of image schemas include PATH, CONTAINER, FORCE, LINK and COLLECTION.

Johnson's list of image-schemas includes both PROCESS and OBJECT, but he says frustratingly little about these. Subsequent authors have elaborated on them somewhat, but I have the impression that these schemas have not been subject to the same scrutiny as many others. Santibáñez [14] attempts to characterise the main features of an OBJECT image-schema, but seems to be limited to rather a narrow conception of 'object' as prototypically middle-sized, solid, inert, and manipulable: cars, books, keys, tables, and apples rather than hurricanes, animals and trees. Szwedek [15] distinguishes static and dynamic object schemas, but the dynamic schemas exclusively relate to the external processes of objects and in no way support any notion that objects may be inherently processual in nature. Can we do better than this in support of the process-ontological theory?

What features do we expect to find in an object in the most general sense? Most fundamentally there is *persistence*: an object manifests a certain stability, either of matter or form. This is achieved through a balance of internal processes (a living organism undergoes a constant exchange of matter with its environment, rebuilding itself throughout its life; an inert chunk of matter maintains a balance between attractive and repulsive forces at atomic and subatomic scales). An object owes its identity to this stability. But objects also enact changes, manifested as their external processes or *activities*. Changes can be intrinsic, affecting the matter and form of the object itself, or extrinsic, affecting its relation to its environment — the most obvious extrinsic change is change of location, or motion. These changes or motions can come about either autonomously, as a result of the object's inner processes, or in response to *impingements* by environmental processes: thus an object's external processes comprise both *action* and *reaction*. But because of stability, an object is able also to resist impingements, giving us *resilience*. Finally, if the balance of the internal processes cannot be maintained, perhaps as a result of external forces too strong to respond to while maintaining stability, the object may undergo *destruction*, involving the separation or dissolution of its parts. All these features are illustrated schematically in Figure 1, where the arrows represent various kinds of causal relation.

The key point to observe here is that the object schema is specified in terms of processes. If this schema comes anywhere close to capturing what it is to be an object, then objects must be dependent on processes. For a thoroughgoing process ontology one must go further than this and maintain that objects are dependent on *nothing but* processes; only thus could we claim that objects are constituted by processes, that the nature of objects is essentially processual. All the considerations I have put forward in this talk tend in this direction, but I readily admit that the case is not proven.

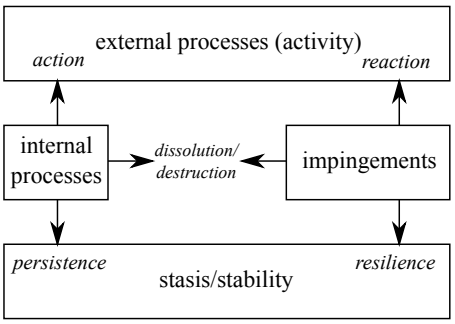

Figure 1. A sketch for a process-oriented OBJECT schema.

REFERENCES

[1] Robert Arp, Barry Smith, and Andrew D. Spear, *Building Ontologies with Basic Formal Ontology*, The MIT Press, Cambridge MA and London UK, 2015.

[2] John Dupré, *Processes of life: Essays in the philosophy of biology*, Oxford University Press, 2012.

[3] A. P. Galton, 'Desiderata for a spatio-temporal geo-ontology', in *Spatial Information Theory: Foundations of Geographical Information Science*, eds., W. Kuhn, M. F. Worboys, and S. Timpf, pp. 1–12, Berlin, (2003). Springer.

[4] A. P. Galton, 'Fields and objects in space, time, and space-time', *Spatial Cognition and Computation*, **4**(1), 39–67, (2004).

[5] A. P. Galton, 'On what goes on: The ontology of processes and events', in *Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)*, eds., B. Bennett and C. Fellbaum, pp. 4–11. IOS Press, (2006).

[6] A. P. Galton, 'Experience and history: Processes and their relation to events', *Journal of Logic and Computation*, **18**(3), 323–340, (2008).

[7] A. P. Galton and R. Mizoguchi, 'The water falls but the waterfall does not fall: New perspectives on objects, processes and events', *Applied Ontology*, **4**(2), 71–107, (2009).

[8] Martin Heidegger, *Sein und Zeit*, Max Niemeyer, Tübingen, 1927.

[9] Mark Johnson, *The Body in the Mind: The bodily basis of meaning, reason and imagination*, Chicago University Press, Chicago, 1987.

[10] George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, Chicago University Press, Chicago, 1987.

[11] Marc Moens and Mark Steedman, 'Temporal ontology and temporal reference', *Computational Linguistics*, **14**, 15–28, (1988). Reprinted in Mani, I., Pustejovsky J, and Gaizauskas, R, *The Language of Time: A Reader*, Oxford, 2005.

[12] Process philosophy and Buddhism: Process metaphysics versus substance metaphysics. http://rational-buddhism.blogspot.co.uk/2013/10/process-philosophy-and-buddhism-process.html. Rational Buddhism blog post, Wednesday 23 October 2013.

[13] Nicholas Rescher, *Process Philosophy*, University of Pittsburgh Press, 2000.

[14] Francisco Santibáñez, 'The object image-schema and other dependent schemas', *Atlantis*, **24**(2), 183–201, (2002).

[15] Aleksander Szwedek. The OBJECT image schema. https://www.academia.edu/31630572/The_OBJECT_image_schema. Accessed 17/3/2017.

[16] Zeno Vendler, 'Verbs and times', in *Linguistics and Philosophy*, chapter 4, Cornell University Press, Ithaca, (1967). Reprinted in Mani, I., Pustejovsky J, and Gaizauskas, R, *The Language of Time: A Reader*, Oxford, 2005.

[17] A. N. Whitehead, *Process and Reality: An Essay in Cosmology*, MacMillan, New York, 1929.

[18] Eugene Wigner, 'The unreasonable effectiveness of mathematics in the natural sciences', *Communications in Pure and Applied Mathematics*, **13**(1), (1960).

Towards Building Ontologies from Crowdsourced Data

Paula Chocron[1] and **Dagmar Gromann**[2]

Abstract. Building computational conceptual models that are flexible and context-independent is an important challenge given the growing interest in cross-domain applications that need to deal with real-world objects. While crowdsourcing methods have been used extensively in ontology engineering and evaluation, few approaches apply these methods to retrieve data. However, retrieving data from a large crowd promises to elicit the whole spectrum of associative knowledge humans use to cognitively represent concepts. In this paper, we propose to use two crowdsourcing techniques - a mechanized labour-based and a game-based approach - as a data acquisition method and a semi-automated approach for building a domain-specific ontology from this data for the concept *city*. Our approaches are indirect, asking participants to describe instances of cities. In a second phase, we implement techniques to extract semantic classes from the obtained data. We evaluate our results against another collaboratively built ontology for *city*, extracted from Wikipedia. We compare the techniques used, analysing benefits and drawbacks for each one.

1 Introduction

The past decade has witnessed an increasing interest in integrated visions that propose a tight and optimised relation between Web technologies and everyday real-life domains, such as the Internet of Things [1] or Smart Cities [7]. These efforts require sound computational models that represent concepts ubiquitous in our everyday life. Moreover, since different domains interact in these approaches, the models need to be as context-independent as possible, very flexible, and easy to adapt to diverse applications or to change and evolution.

The computational modeling of these concepts is a difficult problem that has been largely discussed from both the philosophical and the engineering perspective. Concepts that people can easily understand in informal conversations can be challenging to define and model formally, something that seems to be particularly true for spatial and geographic concepts, as it can be seen in the well-known discussion about the definition of *forest* (see, for example, [4]). An alternative solution would be relying on a group of *experts* to build the concept descriptions. This is, on the one hand, a time- and cost-intensive task [31]. On the other hand, it is not free of arbitrariness, since the resulting conceptualisations can be biased by the view of the person(s) in charge of building them. We believe that relying on a large population for acquiring associative knowledge drastically reduces this potential bias.

In the past years, *crowdsourcing* techniques received increasing attention as methods that can overcome these two problems. Crowdsourcing consists of a collective of non-experts (a *crowd*) performing short and accessible tasks that are then combined to tackle a larger problem. Crowdsourcing methods are particularly well suited for tasks that are difficult to automate completely, but are at the same time too large to be completed by just one person, or that benefit from the diversity of the participants.

In this paper we propose a method that uses two crowdsourcing techniques – a mechanized labour-based and a game-based approach – to build concepts for ontologies, identifying different types of knowledge that humans use to describe instances. In line with results of Parasca et al. [22] we observe that people extensively rely on prior knowledge as well as synonymy, antonymy, and hypernymy to describe concepts. This kind of task can benefit from collaboration, as it is discussed in [15], where the authors compare conceptual maps created individually and collaboratively, concluding that the latter ones are of higher quality.

Using crowdsourcing can provide a description that reflects the collective perception of a concept, identifying categories that are socially relevant, but not immediate when explicitly trying to define them. For instance, five participants in the mechanized labour-based approach used *café* to describe *Paris*. In the approach using a word guessing game, four participants used *baguettes* to describe *Paris*, which was guessed correctly upon this first hint in four different games. When asked directly to describe *city* it is unlikely that participants would provide similarly specific associations. Our approach is still preliminary, but the data obtained[3] can be used as a pattern to describe concepts in applications where the social aspect is relevant or as a human-created standard to evaluate automated techniques. We also believe that the game-based approach can provide a valuable method for cognitive applications, since it can specify the type of knowledge to be elicited in a way that is entertaining to participants.

We chose to perform our experiments using the concept of *city*. The concept of *cityness* has been extensively discussed in the urbanistics literature, remarking its social and dynamic

[1] Artificial Intelligence Research Institute (IIIA-CSIC) and Universitat Autònoma de Barcelona, Spain, email: pchocron@iiia.csic.es
[2] Artificial Intelligence Research Institute (IIIA-CSIC), Spain, email: dgromann@iiia.csic.es

[3] Data obtained from and source code used in this approach can be found at `https://github.com/paulachocron/CrowsourcedKnowledgeAcquisition`

aspects. In particular, we consider *city* to be a good concept to perform this experiment, since it has clear instances which are in general well-known by a random crowd. In addition, although a city can be uniquely identified by means of its coordinates, these are in general not the most immediate characteristics that come to mind, and the resources used when describing an instance are very varied.

After introducing related work, Section 3 explains the design and implementation of two crowdsourcing methods that retrieve data by asking participants to describe instances of a city. The mechanized labour-based technique asks participants directly, while the second one presents the task in the form of a game, making it more attractive to participants. This game-based method can be seen as an extension of the first one, that could potentially complete the descriptions obtained. In Section 4 we explain a post-(crowdsourcing)-processing phase, in which we implement two methods to automatically extract categories related to cities from the crowdsourced data. We evaluated our approach by comparing its results to another crowdsourced description of cities that we extract from the Wikipedia Tables of Contents of city pages.

2 Related Work

Crafting ontologies manually is a costly task, and the obtained results are not free of arbitrariness. For these reasons, the field of ontology learning has been extensively studied in the past years [18]. Many of these approaches, particularly those from the first years of the area's development, rely on predefined patterns and rules or static resources, such as WordNet [31]. However, these static approaches have two drawbacks, namely they are neither scalable nor easily portable between domains. Recent approaches seek to be more dynamic, for example by using machine learning to extract relations from an existing seed ontology [23] or to develop axioms extracted from text [27].

Using static resources in ontology learning is not straightforward due to the multiplicity of senses associated with each word. To address this problem, Bentivogli et al. [5] associate senses with a WordNet domain ontology they create and which we also use herein to classify words. A similar idea is presented by Izquierdo et al. [14] who associate the Kyoto ontology of the project with WordNet senses and also a number of upper level ontologies. Those associations are then used to present a class-based word sense disambiguation method. Alternatively, distributional semantic approaches have been investigated for word sense disambiguation with context-poor data sets. For instance, Basile et al. [3] extract DBpedia glosses for each word in tweets and then compute the cosine similarity between the context of the word in the tweet and each gloss to find the most related one(s), a second approach we adapt in this paper. Similarity between sets of words can be computed by composing their vectors in different ways; in [3] the authors use addition.

The use of crowdsourcing techniques in ontology learning has received considerable attention in the past few years. One prominent example can be found in the approach of Hanika et al. [13], who present approaches that solve some specific tasks in ontology engineering via crowdsourcing, mainly related to specifying relations between terms and verification. Eckert et al. [11] propose a method to crowdsource concept hierarchies, asking users questions about how concepts are related. In general, these methods start from an existing set of concepts and attempt to add relations collaboratively, without allowing the discovery of new concepts. Crowdsourcing has also been applied to ontology evaluation (both for the subclass-superclass hierarchy [21] and for entire ontology statements [32]) and alignment [24].

In an effort to make crowdsourcing tasks more appealing to participants, the idea of *gamification* was introduced that presents the problems to be solved in form of an interactive game, which is thought to foster motivation to participate [25, 29]. Some well-known examples are Duolingo [28], an approach to crowdsource the translation of the Web, and reCAPTCHA [30], a method for digitizing paper copies of documents. Parasca et al. [22] utilize a guessing game to elicit associative knowledge to define words provided to the players. They analyse the type of associative knowledge obtained this way and suggest their data set to be used to evaluate distributional approaches, which makes their method more focused on linguistics. Verbosity [29], a game to elicit commonsense facts in a structured way, restricts the elicitation of knowledge to specifically related items in order to obtain already related expressions. We decided against this design to have a non-restrictive elicitation of associative knowledge.

Individual ontology engineering tasks have been crowdsourced as games as well, such as for classification and population [26]. In [20] a game is proposed to obtain attributes for concept descriptions. Their approach is explicit in that it asks players to name properties directly. In combination with ontologies, a specific part of the ontology building task is usually crowdsourced but not the knowledge acquisition step that precedes the ontology building as in our approach.

3 Crowdsourcing Concept Descriptions

We propose two different crowdsourcing mechanisms to collect descriptions of city instances that will later be used to build the concepts related to city. The approach, from the data collection to the evaluation, is illustrated in Figure 1, where rectangles are steps and circles are the different techniques that we explore.

Crowdsourcing methods can be divided into two major categories [10]. *Explicit* methods are those in which participants are asked directly to perform some task, and then their work is combined to solve a larger problem. In the *implicit* approach users are given a task that is only indirectly related to the problem to solve; then some kind of post-processing is applied to extract the desired information from the retrieved data. The implicit approach allows designers to make the task more attractive for participants, for example by presenting it as a game. For some tasks, implicit approaches can result in more interesting and fine-grained results, and for some experiments it is crucial that participants do not know the final objective to avoid cognitive biases.

We used two different implicit crowdsourcing methods, in which we ask participants to describe specific instances of cities either as a direct question or as part of a game to obtain a general characterization of *city* as a concept. We believe that using an implicit approach for our problem can lead to richer and more fine-grained ontologies than the explicit one. However, the direct comparison to this kind of task is yet

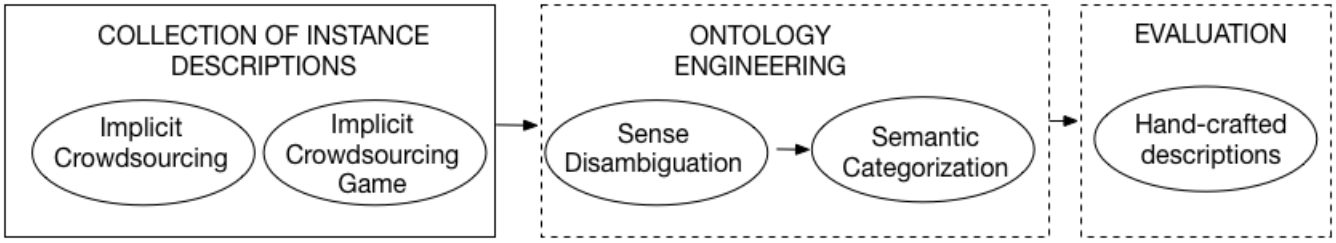

Figure 1. Overview of Concept Formation Method

to follow. The kind of technique we propose here is particularly applicable when describing abstract concepts that do not have a clear physical correspondence, where the properties are less evident. To form concepts, we group the descriptions we obtain from the crowdsourced data by their sense that we disambiguate and by their semantic category that we extract.

A first approach, which we call *mechanized labour-based knowledge acquisition*, uses a popular crowdsourcing online platform (*CrowdFlower*) in which users can register and complete mechanical tasks in exchange for payment. The task we proposed consisted in directly asking participants to provide the first ten words they associate with a city name displayed to them. In the second approach we adapted the well-known Taboo game, in which one player describes a city instance to the other player without using any of the words on a list that is provided to them. This second technique was developed as an extension to the first one, and it therefore uses the results of the first one as words that cannot be used in the game in an effort to obtain more fine-grained, original descriptions of *city*.

For both methods we used as input a list of city names that we built with input from online listings of popular cities retrieved by a search engine query. Participants had to be first language English speakers since the data we wished to obtain are in English. They were always given the possibility to skip a city if they were not familiar enough with it. They were explicitly instructed to only use common nouns with adjectival/verb modifiers. There are two main reasons for this restriction: (1) proper nouns trivialize the identification of cities as they uniquely identify them, e.g. Eiffel for Paris, and (2) we were interested in ontology building from common language and not based on instances or named entities. Participants were explicitly instructed to comply with this input restriction, in the mechanized labor-based approach even tested on them, and non-conforming characterizations were omitted from the final data set. In the following subsections we explain each of the two approaches in detail.

3.1 Mechanized Labour-Based Knowledge Acquisition

Description

In the first of our approaches we ask participants directly to provide descriptions of a given city. This was implemented using the online crowdsourcing platform *CrowdFlower*. In a crowdsourcing task, participants were provided with the name of a city (for example, Paris), its country name (France), and its latitude and longitude (48.85° N, 2.35° E), plus ten input fields for city descriptions as illustrated on the right hand side of Figure 2. The instructions clearly described the task and specified the input restrictions to describing cities by using common nouns with possible adjectival/verbal modifier as description, no words other than English including loan words, and a categorical omission of personal opinions. For instance, *stinky cheese* to describe *Paris* by means of one of its most prominent food exports would be permissible in the sense of providing a common noun in English but clearly violates the instruction of not providing a personal opinion. Each page in the task displayed a total of five city instances to each participant, asking for their descriptions following the specified input restrictions.

The left hand side of Figure 2 shows the first part of a page displayed to the participants, where they could indicate whether they had heard of the city, had been there, or are not familiar with it. The right hand side of Figure 2 shows the view when clicking one of the first two options, where the user is prompted to provide a maximum of ten words they associate with that city. We uploaded a list of 300 cities that we collected from online resources and there was no limit to how many descriptions participants could provide.

We took several measures to ensure the quality of the results for this crowdsourcing technique. First, we conducted a pilot study to evaluate the kind of results we were to expect and to improve our test setup. Second, we asked each participant 20 test questions to ensure their ability to comply with the instructions regarding the input restrictions. For example, we asked participants whether Breaking Bad is an adequate description of Albuquerque, USA, to test the ability to differentiate common from proper nouns. Only participants with an accuracy above 70% on the test questions were permitted to complete city descriptions. Finally, answers on which participants spent less than ten seconds on the question were not considered, a setting enabled by the platform.

This measure was taken because it can be reasonably assumed that a participant who spends less than on average ten seconds on the description of five city instances has not taken the time to provide enough descriptions or delivers poor quality. For instance, if a participant provides *cathedral* for all five cities and no other input it might be possible to do so within ten seconds, but this does not mean that *cathedral* is an adequate description for those cities. It is highly unlikely that a human being is able to process five city names and type at least one association for them within ten seconds. This is a common quality measure in crowdsourcing approaches and in combination with the test question turned out to be a very effective way to obtain high quality contributions. The pilot study was particularly important since the improved actual study received a better evaluation by the participants, which increased participation. The more participants, the faster the

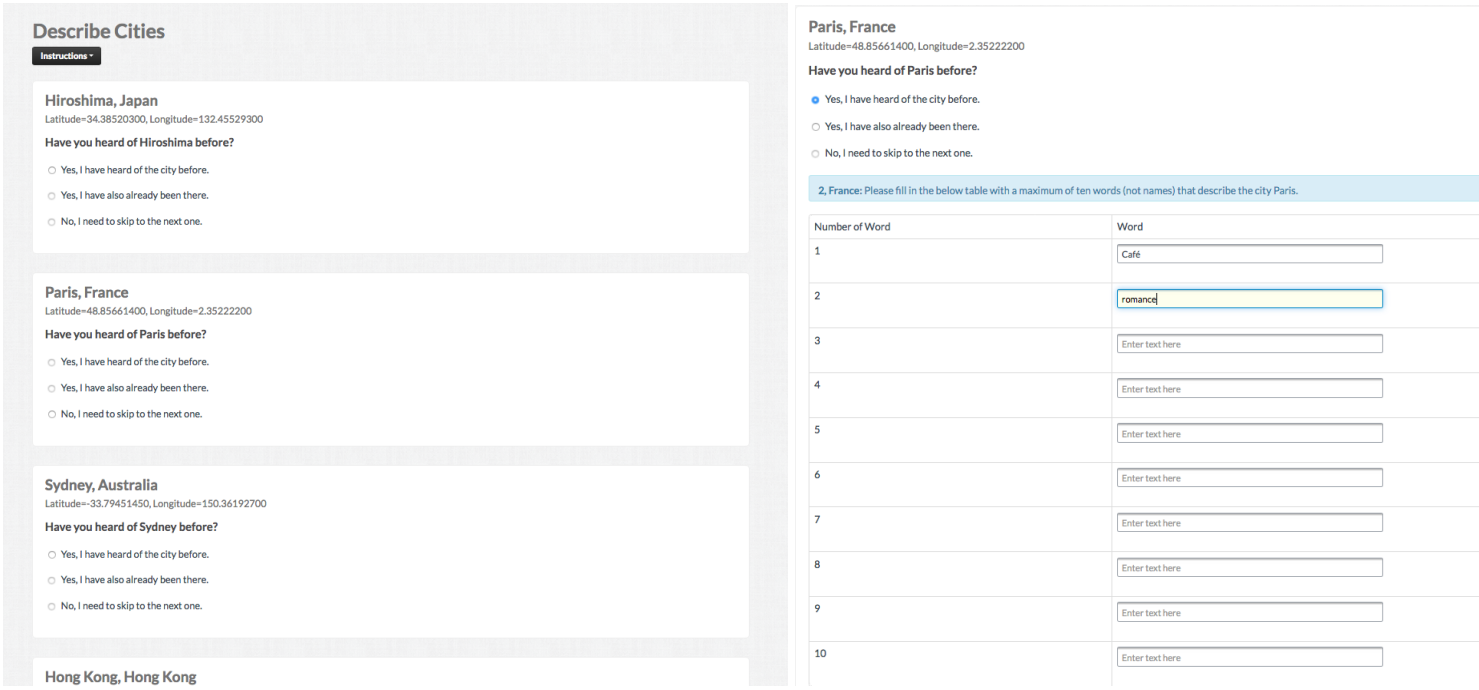

Figure 2. Example Question on *CrowdFlower*

specified number of targeted descriptions can be achieved.

Results

The task was online on *CrowdFlower* for six days asking participants to describe a total of 300 cities. At the end of the task, 82 participants mainly located in the United States (65%) and the United Kingdom (30%) curated a total of 3,616 descriptions for 275 of the 300 cities. For the remaining 25 cities, no descriptions were provided since the participants had the option to indicate that they did not know a city displayed to them in the task. We did not limit the number of descriptions that could be provided by an individual participant in the overall task and some top contributors provided up to 85 descriptions. The proportion of participants that passed all the trust tests (70% accuracy on the test questions and answer time not too short) was 91%, which shows that the task was well designed and accessible for participants. This argument is further supported by a final contributor satisfaction rating of 4.2 out of 5 potential points for our task.

The 3,616 descriptions we obtained for the 275 cities contained duplicates. In a first processing phase, we kept duplicate descriptions across the data set but de-duplicated the ones for each city. For instance, we would de-duplicate the five mentions of *café* relating to *Paris, France* to one while keeping *café* as a description of *Tangier, Morocco*. This was done automatically by applying similarity measures from the WordNet Similarity for Java (WS4J) library[4] combined with the Levenshtein distance [17]. This approach was also used to identify the descriptions that were provided most frequently, for example, there are in total 576 descriptions that were provided by more than one user. For instance, for *Paris* the description *café* was identified as one of the most frequent ones as it was provided by five distinct participants. To reduce the bias of individual associations regarding specific city instances, we decided to choose only those frequent descriptions for the concept formation task.

3.2 Game-Based Knowledge Acquisition

Our second crowdsourcing technique presented the city description task as a game that was designed by adapting the popular board game Taboo™[5], which we call GUESSence. First, we will briefly describe the original Taboo™ game, then our version, and finally the results we obtained from the game.

Description of the Taboo Game

The popular board game Taboo is a word guessing game, where an even number of players is grouped into competing teams. One player of one team draws a card and describes a word it provides, with the objective of having their team colleagues guess the word without using the word itself or five related words indicated on the same card. The five related words are called taboo words and neither the words themselves, their components, nor their inflectional variants can be used. For instance, if the word is *basketball*, it is not permitted to use *ball*, *baskets*, or *basketball* to describe it, and neither is it permitted to use any of the given taboo words, for instance *court, player, bounce, baseline, game*. The goal is to have your team guess as many words as possible within the alloted times. Players take turns at describing words and after the alloted time is over, it is the turn of the next team. Each team receives one point for a correct guess and the team with most points at the end of the game wins.

[4] `https://code.google.com/archive/p/ws4j/`

[5] `http://www.hasbro.com/common/instruct/Taboo(2000).PDF`

Description of GUESSence

In our two-player version of Taboo™, there are two roles a player might assume: describer and guesser. The describer provides hints to the guesser that describe a given city and the guesser responds with a city name that is believed to be the correct result. As a further restriction, the describer is not allowed to use any of the phrases that are provided as *taboo words* along with the city name and the country it is located in. The objective of the game is that the guesser finds out which city is being described. The game is collaborative, since when the guesser names the correct city both players win. As an example depicted in Figure 3, if the city to describe was *Paris, France*, the describer would get the list of taboo words *capital, café, romance, croissant, art, tower, fashion, museum, palace, terrorist, cathedral*. Then a game could be developed as exemplified in the following, where the describer starts the game by saying *river* and the guesser provides the first guess. When the guesser types the correct city name the game was successful and both players are automatically assigned to a new game.

describer:	River
guesser:	London
describer:	Famous pastries
guesser:	Vienna
describer:	Hunchback
guesser:	Paris

The taboo words for this game were the descriptions obtained from the first data collection method, ensuring that there was no overlap between the data set gathered with the first knowledge production method and this one. Additionally, we intended to analyse if this would trigger the use of descriptions belonging to different categories in an effort to find original descriptions. We only used cities for which we had enough taboo words, omitting all cities with less than five descriptions in the first task. This reduced the number of 275 cities from the first task to 244 in the game. In an effort to make each game equally challenging, we decided to include some hand-picked additional salient descriptions as taboo words if they were missing, such as *canal* for *Venice*. By hand-picked we mean we complemented the most frequent descriptions of the first task by associations that only one or two *CrowdFlower* participants had who indicated they had visited the city before.

We determined the number of taboo words on the basis of the number of descriptions provided for each city by the crowd of the mechanized labour-based approach. This was based on the assumption that the more associations participants of the first crowdsourcing method had with a specific instance of a city, the easier it was to retrieve those associations and the more well-known specific associations for that city might be. To make the description of each city instance equally challenging in the game-based crowdsourcing approach, we kept more taboo words for cities with a large number of associations while keeping fewer taboo words for cities with a lower number of associations from the first crowd. Concretely, cities with more than 25 descriptions were equipped with 12 taboo words in the game, cities with 20 to 25 descriptions were assigned 10 taboo words, and 8 taboo words were provided for cities with less than 20 descriptions from the first crowd on *CrowdFlower*. Cities with less than five descriptions from less than two participants were omitted in the game-based approach since this can be seen as an indication that not enough associations could be elicited from the crowd and thus the city might be more challenging to describe in the game-based approach than the other city instances.

Crowdsourcing

Since the game requires collaboration, participants had to be online simultaneously, which made the use of crowdsourcing platforms challenging. As we observed in the first data collection method, the number of participants that simultaneously accessed the crowdsourcing task rarely exceeded two at a time. Thus, we had to devise an alternative method to recruit participants. We contacted colleagues within the ESSENCE project[6] at the School of Informatics at the University of Edinburgh, and asked them to send out an invitation to participate on internal mailing lists and invite personal contacts. This local recruitment should ensure that participants in the game were first language English speakers in line with the first crowdsourcing method and this was the only partner institution located in an English speaking country. As an incentive to participate, we offered a small shopping voucher. To avoid personalized hints when participants knew each other, the method for joining two players in a game was automated and both players had no means of identifying who the other player was.

We developed an online platform[7] and pre-scheduled game sessions with up to nine players at a time, who were assigned automatically and anonymously to two-player games. The first player to log onto a game would be assigned the guesser role. The second player to join a game session would be the describer, who in contrast to the guesser would see the city name, country, and taboo words, the view that is illustrated in Figure 3. The game starts with the describer providing a hint and ends with the correct guess from the guesser. Players were newly assigned automatically and anonymously to each game. The same input restrictions as in the first crowdsourcing method were applied for the hints of the describer, which was additionally enforced by automated warning messages. For those warnings, we implemented several named entity recognition processes using NLTK [6] and the Stanford Named Entity Recognizer (NER) [12]. Upon a correct guess, the interface automatically redirects both players to a success page. Should the automatic detection of a correct guess fail because of spelling errors or other reasons, we provide a "Guess Correct" button on the interface. If a player decided to leave the game and start a new game, the other player was automatically informed by the system that the game was over.

A total of 30 participants played the game in five online sessions. This resulted in 316 games, of which 174 were successful, i.e., the city was guessed correctly. We decided to limit our data set for the concept formation process to successful games only. This ensures the quality of the hints, i.e., they are indeed associated with the city being described to a degree that allows a human player to identify the city. The lengths of the games varied substantially. At times players were able to

[6] `https://www.essence-network.com/`
[7] `http://taboo.iiia.csic.es`

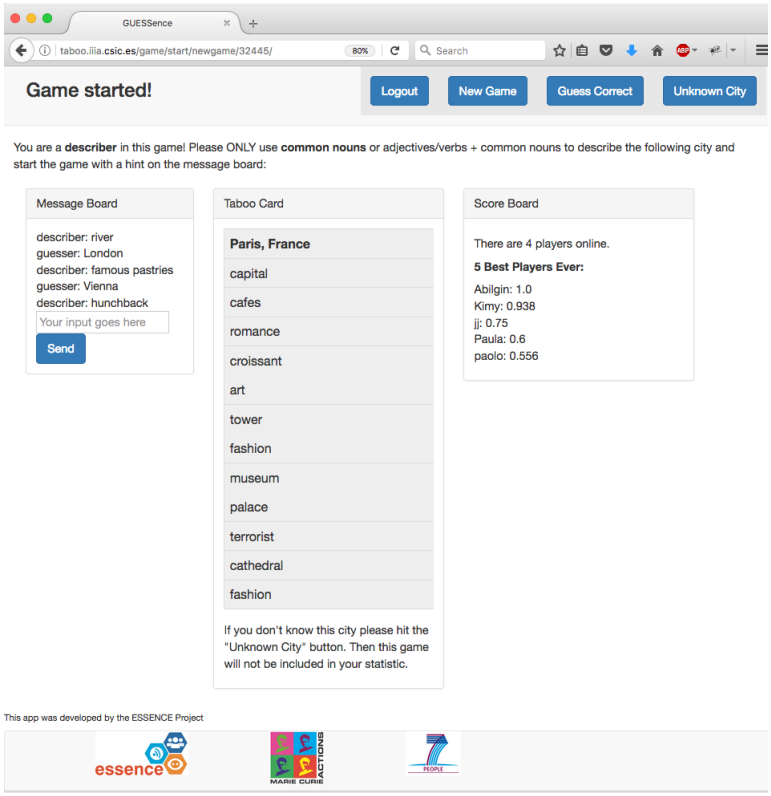

Figure 3. Example GUESSence Game: Describer View

guess the correct city name upon receiving the first hint, such as *baguettes* immediately triggered the response *Paris* when indeed *Paris* was the target city. In other games players required up to ten hints to correctly guess the target city. It can be presumed that first hints that immediately and in several games trigger the correct target city are more associated with the instance then hints that require further descriptions to lead to a correct result. For instance, *baguettes* as a first hint for *Paris* lead to four successful games with different players, which means that *baguettes* is closely related to the city. We could observe a strategy of guessers to first provide the capital of a country they presumed the target city to be in, so it could be argued that *baguettes* is more related to the country itself then the city. Indeed, the types of knowledge used by the participants to describe a city came from different categories, such as regions, continents, climate, food, fauna, flora, among many others.

Those successful games were manually evaluated by 12 ontology engineers and researchers regarding their compliance to the restriction to common nouns and the rules of the game, e.g. not containing a Taboo word, which reduced our data set of successful games to 73 games of 62 cities and a total number of 202 descriptions of cities. This set of 202 descriptions of the second technique and the 322 descriptions from the first crowdsourcing technique provided the input to our concept formation method.

4 Building Ontologies

Building ontologies from natural language descriptions is a four step process that requires *concept formation* (e.g. [8]), extraction of hierarchical relations (e.g. [9]), learning non-hierarchical relations (e.g. [23], and finally extracting axioms (e.g. [27]). In this paper we focus on the first step of forming concepts from the instance descriptions we obtained in the two crowdsourcing processes. Concept formation is usually understood as the process of grouping terms into classes based on their shared semantic properties [31]. This involves detecting term variants, which is commonly done by using predefined background knowledge [31].

In our data, the available semantic property is the relation of each word to a city instance. Thus, we need to disambiguate each word and then group them by means of their related senses that we retrieved form existing lexical resources. We also retrieved predefined semantic categories into which data from both crowdsourcing approaches are classified. In this section we will differentiate between the data from the two approaches by referring to data from the mechanized labour-based approach as *taboos* and descriptions from the game-based approach as *hints*. To continue with our example, the hint *baguette* is related to *Paris* and the sense we want to obtain is that of "a long, narrow loaf of French bread" with the semantic category *Food*, where the category provides the general characteristic people use to describe a city. This is the ultimate objective of our approach: find the types of concepts that people associate with the general concept *city*. To evaluate our obtained concepts, we use an ontology that we created manually based on Wikipeda Tables of Contents (TOCs) of specific instances of cities. We will first explain how we obtained this evaluation ontology and then detail the approach to disambiguate and classify the crowdsourced data.

4.1 Building an Evaluation Ontology

When searching for characteristics of specific cities, we found that the Table of Contents (TOCs)[8] of Wikipedia pages for city instances shares some interesting characteristics with the kind of ontology that we attempt to build. ToCs are organised hierarchically, including subsections and even lower levels. As a collaborative encyclopedic resource where each page is quality assured by several community members, Wikipedia could be considered a long-term crowdsourcing approach and due to its quality assurance can be considered an excellent testbed for the concepts we form. Thus, we decided to manually build an ontology from TOCs of Wikipedia pages as a gold standard ontology for the characterization of city.

To build this ontology, four ontology engineers extracted Wikipedia TOCs of 20 randomly selected cities from our list of 300 cities. We merged the TOCs of those city pages, keeping the most general ones. In this way, we removed categories that were very specific to one city or region (such as "2.1.1 Legend of the founding of Rome"). In general it was easy to achieve an agreement, which suggests a high degree of consistency in Wikipedia's ToCs. Since many participants in the crowdsourcing processes used descriptions related to the country or region the city is in to describe the city, we repeated the ontology construction process from TOCs for countries, regions, and continents, which provided us with a four-layer gold standard ontology we used to evaluate our ontology built from crowdsourced data[9].

4.2 Concept Formation

To group our data into semantic categories, we retrieved the available senses and classifications for each noun and noun phrase from WordNet and an online dictionary, called Word Reference[10]. We opted for this combination since we found that semantic categories provided by Word Reference frequently complement WordNet domains [5]. An unexpected level of complexity was found at this point, triggered by the multiplicity of senses that may exist for each hint. In many cases, different senses corresponded to different categories; to continue with our example, baguette can be interpreted as "a small convex molding especially one of semicircular", which is classified as Architecture instead of the desired category Food for this hint. To overcome this problem, we implemented and compared several word sense disambiguation techniques. Since the focus of this paper is on crowdsourcing and ontology building, we will only briefly report on the most successful technique we applied.

By means of a word sense disambiguation approach based on distributional semantics adapted from [2], we identified the meaning that is semantically closest to the the general concept of city from WordNet and Word Reference definitions. We used each word of our input data in a vector representation and composed it with the vector of the city name. In a second step, we created a vector representation of each extracted definition and compared it with the first vector. The one combination of a definition vector with the $(word, city)$ that returned the highest similarity measure was the sense we chose automatically to be closest to the general concept of city. For instance, in the case of the two *baguette* senses related to *Food* and *Architecture*, the disambiguation process compares the vector for $(baguette, Paris)$ with the vector for each sense that is composed from the words in their glosses and returns a higher similarity for *Food*. Our composition method for vectors is that of averaging as implemented in the word2vec library [19]. All obtained senses were manually evaluated by two raters and only senses that both raters agreed upon were kept. We determined an F-Measure of 82% (164 taboos from 199) for the first crowdsourcing platform data set and 80% (89 hints from 112) for the second game-based data set. Given the highly ambiguous nature of the input data and the lack of context, we consider this a good result. For instance, boot has 7 different senses and no obvious connection to Wellington. Our algorithm identifies the correct sense of 'footwear', since the description most likely hinted at the famous 'Wellington rubber boot'.

We implemented two different approaches to extract predefined categories from descriptions of particular cities: categories from an online dictionary and ontology classes mapped to WordNet senses. In the first technique, we exploited the fact that Word Reference associates words with general labels that can be generally seen as its superordinate class. For example, Sushi is labeled as Food as is baguette which groups them together. To use this information, we first extracted all nouns in the crowdsourced city descriptions and then retrieved all existing glosses and categories from Word Reference where available. We used our distributional semantic disambiguation approach on the glosses to which labels were assigned to find the best gloss and extract its associated category. Our second approach consists of using WordNet to obtain the semantic classes mapped to its senses. Due to the fine-granular nature of WordNet senses, it was necessary to use ontologies associated with WordNet synsets to obtain general categories, as it is described in detail below.

In general, both Word Reference (WR) and WordNet (WN) contained definitions for the words in the descriptions, although there were some exceptions, most of which corresponded to words that can be considered from foreign languages (for example, names of foods or sports). For the majority of the words that were defined, the resources included the sense that was intended by the describer as one of the possible definitions, as it is shown in Table 1. We use the name *hints* to refer to data obtained in the game from the describer and *taboos* for the results of the first mechanized labour-based task. By *not available* we mean that the word is in the resource but the required sense is not. In some other cases, the describer used the word in a very complex or informal way, which was not included in the utilized resources. This is the case of, for example, using sack to describe Sacramento. The evaluation was done manually by three researchers.

We also measured the number of available correct categories in Word Reference as depicted in Table 2. In this case the values are lower, because many glosses in Word Reference are not classified into a category. At times, the categorization in the resource is not entirely accurate, as for instance beer is classified as wine instead of alcoholic beverage. Nevertheless, in most cases the quality of the categories is surprisingly high

[8] https://en.wikipedia.org/wiki/Help:Section#Table_of_contents_.28TOC.29

[9] The ontology is available at `https://github.com/paulachocron/CrowsourcedKnowledgeAcquisition`

[10] http://www.wordreference.com

	Available	Not Available
WN (taboos)	199	1
WN (hints)	112	6
WR (taboos)	194	6
WR (hints)	109	9

Table 1. Availability of correct senses for WordNet and Word Reference

and the accuracy we obtained exceeds 80% for both data sets. This means that 80% of the extracted semantic categories were correctly assigned, which is a good result given the lack of context.

	Available	Not Available	Correct	Accuracy
taboos	132	64	119	0.90
hints	77	42	65	0.84

Table 2. Availability of categories in Word Reference

We also evaluated the Word Reference categories by comparing them with the city ontology from TOCs. We performed this evaluation only for the categories that were disambiguated correctly with the distributional semantics approach, since we are interested in how far the data collected by crowdsourcing reflect a proper description of the general concept city. These categories are not organized in a taxonomy, and thus only the number of semantically equivalent categories with the Wikipedia resource was analysed.

When removing duplicates in the categories from Word Reference, we obtained a total of 29 categories for the hints and 31 for the taboos. In Table 3, we show the matching proportion with the Wikipedia taxonomy of cities, modulo obvious term alignment (such as Food ≡ Cuisine). We count in N the categories that did not match, and then present the ones that were present directly, the ones that were subconcepts of a present one (for example Mammal) and the ones that were present in the Wikipedia taxonomies for region or country. In both cases, 9 of the 12 first-level categories in the Wikipedia taxonomy were represented, either by themselves (in 6 cases) or by one of their subcategories.

	N	N (%)	Present	Subconcepts	Other tables
taboos	5	16%	15	5	6
hints	2	6%	18	8	1

Table 3. Evaluation with Gold Standard Ontology for the WR categories

To classify the WordNet definitions, we used existing mappings to ontology concepts in YAGO[11] and the Kyoto mapping to WordNet [16] to extract classes associated with descriptions. Using the disambiguated WordNet senses we first extracted all WordNet Domains (WNDs) [5] associated with them from YAGO. This resulted in 17 for the hint data set and 6 for the taboo data set, of which one for each data set was incorrect. Those numbers already refer to de-duplicated senses used as input, that is, each WND was only counted once. Where there was no WordNet domain, we queried associated Kyoto classes and used the Word Reference categories to automatically select the best class. This was then again evaluated manually. We extracted a total of 78 classes for senses for the hint and taboo data set with an accuracy of 80% for the first and 91% for the second.

In a nutshell, our approach to concept formation consists of disambiguating individual specifications of city instances and grouping them by semantic categories. We obtain the semantic categories from WordNet domains and Word Reference categories. Those are then compared to concepts of the Wikipedia ontology we created as a means to evaluate our approach. Each concept created by this method comprises numerous characterizations of city instances. The set of concepts we obtain characterize the general concept *city*.

4.3 Comparison of Acquisition Approaches

Since the data of the mechanized labor-based acquisition method served as input to the game-based approach, the data sets we obtained are mutually exclusive. On a higher level of abstraction, that is, on the level of the semantic categories we obtained from Word Reference, the overlap between the obtained categories was 60%. Many of the remaining 40% that did not correspond across the two data sets were more specific concepts. For instance, the hints resulted in *Religion* while the taboos also returned *Eastern Religion*. Another example of the reverse phenomenon is that the hints only lead to *Clothing* while the taboos also lead to *Textiles*. Based on some concepts returned, such as *Drugs*, and an examination of the crowdsourced data we found that the game-based approach tends to provide more informal descriptions than the mechanized labour-based approach. This could be explained by the fact that the latter task itself has a higher degree of formality where participants are formally paid, while a word guessing game can be considered an informal situation. However, in some cases the use of informal words to avoid using the forbidden taboo words has been observed as a strategy. For instance, a player used *haggling* to describe *Bangkok* since the crowd on the online platform had provided *bargaining* as a taboo word.

The paid crowdsourcing task was conducted mainly by participants in the United States and also several participants from the United Kingdom (UK). In contrast, the game-based method was exclusively based on participants from Edingburg, UK. It could be expected that this difference in country of residence of the participants would affect the nature of the data in the sense of introducing a cultural bias. However, with very few exceptions no local expressions or dialect could be observed in the data. Interestingly the difference between American and British English was used to describe specific cities, such as an American participant using *soccer* to describe *Newcastle* in the UK. The same phenomenon occurred with other languages, such as a British participant in the labor-based approach who used the German *s-bahn* (rapid transit railway) to characterize Berlin in Germany or a player in the game describing Osaka with *okonomiyaki* (a type of Japanese pancakes). We kept the British expression in the previous example, but did not keep the German or Japanese in the data sets. In the paid crowdsourcing process we believe that the formality of the task as well as the nature of the task might have led people to opt for a more standard version of English. In the game-based approach people were

[11] www.yago-knowledge.org/

neither informed nor aware of the fact that all other participants were equally based in Edinburgh. Since participants met online without any means of identifying the other player in each game, virtually no local expressions could be found in the final dataset. However, participants presumed to be in the same region since they repeatedly used cardinal directions to describe cities, such as *West*, which only helps in the game if this direction points to the same locations for both players.

5 Discussion

The two crowdsourcing techniques that we used in this approach returned useful input for the ontology building process. We found that the time needed to obtain data from the mechanized labour-based approach exceeded the time of the game-based approach to return the same amount of data. The former was running for more than a working week, while the latter achieved the same in just five sessions each a bit more than an hour. This is because the incentive to participate in an interactive online game seemed much higher. In fact, participants asked for the permission to play again after the first session, and four of the 30 participants joined a second session. One crucial point in accelerating the labour-based approach was the overall contributor satisfaction. In the pilot study, which also served to evaluate the accuracy of our test questions, we obtained a lower ranking of only 3.5 out of 5 points and several comments from contributors regarding suggestions to improve the quality of the test questions. In the actual study, we obtained a rating of 4.2 out of 5 points, which considerably increased the frequency with which new participants joined the task. Thus, we could observe that a high ranking can considerably accelerate the labour-based approach, however, it is unlikely that it even a the highest possible ranking would come close in performance to the game-based approach.

The low number of and locally restricted recruitment method of participants for the game-based approach might have a biasing effect on the data set obtained from this method. While the results from the concept formation stage seem to be comparable to the data obtained from the labour-based crowdsourcing step, we still believe that it would be more interesting to compare both methods with a larger number of participants. To this end, we have already developed a mobile app version of GUESSence that will allow an asynchronous access to the game and allow for an open access without any restriction to a specific region.

When applying the ontology building process to the data sets obtained with both crowdsourcing techniques, it can be observed that the mechanized labour-based technique returns more specific categorizations. This implies that the game is not useful as an extension of the first method, as we initially suggested. However, there are only minor differences in the results obtained with both approaches, and they both perform well when compared to our gold standard ontology. This means that the game returned results that were as robust as the direct technique, and can therefore be used on its own to retrieve descriptions. These kind of describing and guessing games are very popular, easy to play, and can be extended to different domains, which turns them into a good candidate to obtain this kind of data. They can even be included in an online gaming platform, which would provide very large sets of data.

This extensibility of the game-based acquisition method to other domains or other target concepts than *city* is also true for the labour-based approach. Any instances of a specific concept could be used to trigger responses from participants as long as they are general enough to be known to a crowd of non-experts, such as *Food* as a general category or *Movies*. The results of the first approach could again be used as taboo words for the game. Alternatively, the taboo words could be generated manually. However, the design of the game relies on those input restrictive characterizations we call taboo words. The main restriction in terms of portability of the approach to other domains are: a) the domain and its instances have to be known by the participants to a degree that allows them to describe them, b) the domain has to be covered by some predefined knowledge or lexical resource to allow for the disambiguation and classification approach as it is proposed here to work. This means that highly domain-specific contents, such as *fishing rod* and the specific types of rods as instances, would not work well for this approach since a low number of participants would be able to describe their characteristics and few resources would contain the words, however, the more general category *sport* or *sport equipment* could be used.

Regarding the concept formation, the two resources that we used for the extraction of senses (Wordnet and Word Reference) were accurate in that they contain the correct sense for most of the words in the city descriptions. Word Reference is convenient because it already provides a classification of the senses in the form of semantic category, however, there are many senses for which that classification is missing, which results in a great loss of useful data. A resource like this one but with a complete classification would be ideal for our purposes. For WordNet, the labelling feature is not immediately available, so more complex techniques need to be implemented to retrieve a classification of our data. In both cases there were many other senses available, so some kind of sense disambiguation is necessary.

The comparison of the categories retrieved from Word Reference with the ones in the gold standard shows that in most cases the labels match. Some of the ones that do not match directly are in turn subcategories of Wikipedia labels, which seems to show that creating an organized taxonomy using the Word Reference taxonomies as seeds would be a promising direction. In other cases, the categories match with others in the Wikipedia taxonomies for *country* or for *region*. This should be taken into account when using this kind of approach, since players tend to describe instances not only with their properties but also with properties from their upper categories.

6 Conclusions

We presented and discussed methods to automatically build concept descriptions from crowdsourced data. The two crowdsourcing techniques that were proposed gave good results in terms of quantity and reflecting associative knowledge. The results obtained with the game-based approach are as robust as the ones obtained with the mechanised one, although slightly less fine-grained. The final results we obtained could be very valuable as a seed for ontology learning to be extended with hierarchical, non-hierarchical relations, and axioms.

Multiple directions of research are derived naturally from

this work. Regarding the crowdsourcing methods presented, it would be interesting to compare, for this particular problem, the use of implicit techniques, like the ones we propose, with explicit ones. The implicit techniques already have as an advantage that they can be easily presented as a game, making the task more attractive. However, it would be interesting to compare the differences in the results. To this end, a third experiment in which users are asked directly to name properties of cities should be performed.

At the moment we focus on concept formation. In order to have more interesting and useful ontologies, this part should be developed to extract hierarchical and more informative non-hierarchical relations. There is a variety of approaches that tackle the relation extraction problem, both with automated and crowdsourcing techniques. However, their adequacy to our problem should be analysed, since they are not particularly designed to identify relations between a concept and its attributes.

ACKNOWLEDGEMENTS

This research has been funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 607062 /ESSENCE: Evolution of Shared Semantics in Computational Environments/.

References

[1] Luigi Atzori, Antonio Iera, and Giacomo Morabito, 'The internet of things: A survey', *Computer networks*, **54**(15), 2787–2805, (2010).

[2] Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro, 'An enhanced lesk word sense disambiguation algorithm through a distributional semantic model', in *COLING*, pp. 1591–1600, (2014).

[3] Pierpaolo Basile, Annalina Caputo, Giovanni Semeraro, and Fedelucio Narducci, 'Uniba: Exploiting a distributional semantic model for disambiguating and linking entities in tweets', in *Proceedings of the the 5th Workshop on Making Sense of Microposts co-located with the 24th International World Wide Web Conference (WWW 2015), CEUR Workshop Proceedings 1395, CEUR-WS.org*, eds., Matthew Rowe, Milan Stankovic, and Aba-Sah Dadzie, p. 62, (2015).

[4] Brandon Bennett, 'What is a forest? on the vagueness of certain geographic concepts', *Topoi*, **20**(2), 189–201, (2001).

[5] Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta, 'Revising the WordNet domains hierarchy: semantics, coverage and balancing', in *Proceedings of the Workshop on Multilingual Linguistic Resources*, pp. 101–108, (2004).

[6] Steven Bird, 'NLTK: the natural language toolkit', in *Proceedings of the COLING/ACL on Interactive presentation sessions*, pp. 69–72, (2006).

[7] Hafedh Chourabi, Taewoo Nam, Shawn Walker, J Ramon Gil-Garcia, Sehl Mellouli, Karine Nahon, Theresa A Pardo, and Hans Jochen Scholl, 'Understanding smart cities: An integrative framework', in *Proceedings of the 45th Hawaii International Conference on System Sciences (HICCSS)*, pp. 2289–2297, (2012).

[8] Philipp Cimiano, 'Ontology learning from text', in *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*, ed., Philipp Cimiano, 19–34, Springer, (2006).

[9] Philipp Cimiano, Andreas Hotho, and Steffen Staab, 'Learning concept hierarchies from text corpora using formal concept analysis', *Journal of Artificial Intelligence Research (JAIR)*, **24**(1), 305–339, (2005).

[10] Anhai Doan, Raghu Ramakrishnan, and Alon Y Halevy, 'Crowdsourcing systems on the world-wide web', *Communications of the ACM*, **54**(4), 86–96, (2011).

[11] Kai Eckert, Mathias Niepert, Christof Niemann, Cameron Buckner, Colin Allen, and Heiner Stuckenschmidt, 'Crowdsourcing the assembly of concept hierarchies', in *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, pp. 139–148, (2010).

[12] Jenny Rose Finkel, Trond Grenager, and Christopher Manning, 'Incorporating non-local information into information extraction systems by gibbs sampling', in *Proceedings of the 43rd annual meeting on association for computational linguistics*, pp. 363–370, (2005).

[13] Florian Hanika, Gerhard Wohlgenannt, and Marta Sabou, *The uComp Protégé Plugin: Crowdsourcing Enabled Ontology Engineering*, 181–196, Springer International Publishing, 2014.

[14] Rubén Izquierdo Beviá, Armando Suárez Cueto, German Rigau Claramunt, et al., 'Word vs. class-based word sense disambiguation', *Journal of Artificial Intelligence Research*, (2015).

[15] So Young Kwon and Lauren Cifuentes, 'The comparative effect of individually-constructed vs. collaboratively-constructed computer-based concept maps', *Computers & Education*, **52**(2), 365 – 375, (2009).

[16] Egoitz Laparra, German Rigau, Piek Vossen, et al., 'Mapping wordnet to the kyoto ontology.', in *LREC*, pp. 2584–2589, (2012).

[17] Vladimir I Levenshtein, 'Binary codes capable of correcting deletions, insertions, and reversals', in *Soviet physics doklady*, volume 10, pp. 707–710, (1966).

[18] Alexander Maedche, *Ontology learning for the semantic web*, volume 665, Springer Science & Business Media, 2012.

[19] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, 'Efficient estimation of word representations in vector space', *ICLR 2013*, (2013).

[20] Peyman Nasirifard, Slawomir Grzonkowski, and Vassilios Peristeras, 'Ontopair: Towards a collaborative game for building owl-based ontologies', in *Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb), Workshop at the 5th European Semantic Web Conference (ESWC08)*, (2008).

[21] Natalya F Noy, Jonathan Mortensen, Mark A Musen, and Paul R Alexander, 'Mechanical turk as an ontology engineer?: using microtasks as a component of an ontology-engineering workflow', in *Proceedings of the 5th Annual ACM Web Science Conference*, pp. 262–271, (2013).

[22] Iuliana-Elena Parasca, Andreas Lukas Rauter, Jack Roper, Aleksandar Rusinov, and Guillaume Bouchard Sebastian Riedel Pontus Stenetorp, 'Defining words with words: Beyond the distributional hypothesis', *ACL 2016*, 122, (2016).

[23] Alina Petrova, Yue Ma, George Tsatsaronis, Maria Kissa, Felix Distel, Franz Baader, and Michael Schroeder, 'Formalizing biomedical concepts from textual definitions', *Journal of biomedical semantics*, **6**(1).

[24] Cristina Sarasua, Elena Simperl, and Natalya F. Noy, *CrowdMap: Crowdsourcing Ontology Alignment with Microtasks*, 525–541, Springer Berlin Heidelberg, 2012.

[25] Neil Savage, 'Gaining Wisdom from Crowds', *Communications of the Acm*, **55**(3), 13–15, (2012).

[26] Katharina Siorpaes and Martin Hepp, 'Ontogame: Towards overcoming the incentive bottleneck in ontology building', in *Proceedings of the 2007 OTM Confederated International Conference on On the Move to Meaningful Internet Systems - Volume Part II*, OTM'07, pp. 1222–1232, Berlin, Heidelberg, (2007). Springer.

[27] Johanna Völker, Daniel Fleischhacker, and Heiner Stuckenschmidt, 'Automatic acquisition of class disjointness', *Web Semantics: Science, Services and Agents on the World Wide Web*, **35**, 124–139, (2015).

[28] Luis von Ahn, 'Duolingo: learn a language for free while helping to translate the web', in *Proceedings of the 2013 international conference on Intelligent user interfaces*, pp. 1–2, (2013).

[29] Luis Von Ahn, Mihir Kedia, and Manuel Blum, 'Verbosity: a game for collecting common-sense facts', in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 75–78. ACM, (2006).

[30] Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum, 'recaptcha: Human-based character recognition via web security measures', *Science*, **321**(5895), 1465–1468, (2008).

[31] Wilson Wong, Wei Liu, and Mohammed Bennamoun, 'Ontology learning from text: A look back and into the future', *ACM Computing Surveys (CSUR)*, **44**(4), 20, (2012).

[32] Maayan Zhitomirsky-Geffet, Eden S Erez, and Bar-Ilan Judit, 'Toward multiviewpoint ontology construction by collaboration of non-experts and crowdsourcing: The case of the effect of diet on health', *Journal of the Association for Information Science and Technology*, (2016).

A Cognitive Ontology of Rhetorical Figures

Randy Allen Harris,[1] Chrysanne Di Marco,[2] Ashley Rose Mehlenbacher,[3] Robert Clapperton,[4] Insun Choi,[5] Isabel Li,[6] Sebastian Ruan,[7] and Cliff O'Reilly[8]

Thought is metaphoric ... and the metaphors of language derive therefrom.
—I. A. Richards, *The Philosophy of Rhetoric*

Abstract. We report on an ongoing research project that uses OWL and a corpus of figurative instances to explore the cognitive dimensions of rhetorical figures. We argue that figures beyond familiar tropes like metaphor and metonymy are deeply cognitive. We adopt the position that figures are form/function pairings which make them especially promising for Advanced NLP, taking one figure as exemplary, the reverse-repetition scheme, antimetabole (e.g., "All for one and one for all"). We illustrate three different methodologies for building ontologies—top down, middle out, and bottom up. And we demonstrate the fruitfulness of cognitive ontological modelling by the way it reveals new categorical arrangements, new conceptual relationships, and important but hitherto unsuspected properties of figures, their multiplicity (figures working together) and their multiplexity (figures having multiple types of instantiations).

1. INTRODUCTION

Rhetoric is the ancient study of persuasion, with particular attention to the effects of expressive style on belief, action, and knowledge. Rhetorical figures are units of style, and their effectiveness is based on their cognitive appeal. The burgeoning work on metaphor, metonymy, and a very small handful of other tropes has established that fact for conceptual figures. But the same is true of schemes, figures based not in semantics ("meaning") but in phonology, morphology, lexis, and syntax ("form"). Many figures, for instance, like ones we take up in this paper, are based on repetition, and we all know how effective repetition is in aiding memory, shaping belief, and supporting reason. A syllogism is impossible without repetition.

Remarkably, however, rhetorical theories of figuration have rarely been mapped against the results of cognitive science, a metadiscipline that is to the current intellectual environment what evolutionary theory was at the turn of the twentieth century. In an earlier paper ([16]), we made the case that ontological representation was a natural way to chart the cognitive dimensions of figures, with the promise of considerable Advanced Natural Language Processing (ANLP) payoffs. While this project has connections with other work in the field, most notably suggestions by James Crosswhite, a rhetorician ([6]), and the explicit proposal by Floriana Grasso, a computer scientist ([12], [14]), of a Computational Rhetoric subfield, it is surprisingly novel. Crosswhite briefly suggests the importance of rhetorical figures for argumentation studies, a suggestion Grasso endorses, both of them in a computational context. But neither of them make any suggestions of how figures might be made computationally amenable, what functions they might serve in argument mining or other ANLP domains.

Our project is innovative in the field of rhetoric, where the cognitive dimensions of figures have seen very little attention (though, see [10], [4]), computational methods have not been used to investigate figures, and ontological research is very rare; and our project is unique, to the best of our knowledge, in computer science, artificial intelligence, and NLP, where rhetorical figures are largely unknown and rhetoric is poorly understood.[9] An excellent recent survey of argument mining, which conspicuously mentions rhetoric many times, nevertheless does not mention rhetorical figures ([20]).

We say "*surprisingly* novel," for two main reasons. Firstly, Jeanne Fahnestock's work shows very clearly how relevant figures are to argumentation ([8], [9], [15], [28]). They epitomize lines of argument. Secondly, many figures are highly amenable to computational modelling and detection. Rhetorical figures are virtually tailor-made for ANLP.

We have been heartened to see a few others adapt our approach or pursue our goals, work we have profited considerably from. Dubremetz and Nivre ([7]) and Hromada ([18]) have had promising success detecting figures. Mladenović and Mitrović

[1] Corresponding author and Director of RhetFig Project; English Language and Literature, University of Waterloo, Waterloo ON Canada N2L 3G1; Email: {raha}@uwaterloo.ca
[2] Cheriton School of Computer Science, University of Waterloo, Waterloo ON Canada N2L 3G1
[3] English Language and Literature, University of Waterloo, Waterloo ON Canada N2L 3G1
[4] School of Professional Communication, Ryerson University, Toronto ON Canada M5B 2K3
[5] Mathematics, University of Waterloo, Waterloo ON Canada N2L 3G1
[6] Biology, University of Waterloo, Waterloo ON Canada N2L 3G1
[7] Cheriton School of Computer Science, University of Waterloo, Waterloo ON Canada N2L 3G1
[8] Independent scholar, London.

[9] Mann and Thompson's Rhetorical Structure Theory (RST; [21]) is symptomatic of this (understandable) ignorance. While it has made some valuable insights into text linguistics, it is simply incorrectly named, by scholars who appear to know little or nothing about rhetoric. RST has really to do with text *coherence* rather than with *rhetoric* as traditionally understood, as the study of suasive language.

([22]) have built an impressive ontology of rhetorical figures in Serbian.

Our RhetFig Project has continued to develop ([19]), and we report on it further in this communication. We are building a comprehensive cognitive ontology of rhetorical figures in OWL—that is, an ontology organized along known cognitive affinities (like repetition, similarity, and contrast)—for Linguistic Linked Open Data (LLOD). We envisage a wide range of potential applications, including genre detection, sentiment analysis, and argument mining. After working with both a top-down and a middle-out methodology, both having their value, and both discussed below, we have shifted to a data-centric, bottom-up methodology, developing the ontology to accommodate specific instances, rather than the abstract classes.

2. RHETORICAL FIGURES ARE COGNITIVE

The machinery of argumentation specifically, persuasion generally, is guided by the organizational principles and natural affinities of the human mind. No one disputes this for analogy and metaphor, research into which goes back decades in cognitive science, but it is demonstrably true for a wide range of other rhetorical patterns. We know, for instance, that words aid in the recall of certain other words, those which share some of their phonological, morphological, or semantic characteristics (if you hear "key" you more readily call up words like "cue" "keystone" or "door"). *Lexical priming*, the phenomenon is called, and it is often described as a "spreading activation" among related words in neural net models. We know, too, that the brain is particularly attuned to sounds of similar duration and frequency. We know that repetition is critically important for fixing sequences in memory (one repeats a phone number or an address to remember it). All of these brain-function facts are reflected in the schemes and tropes of rhetorical theory.

Consider, for instance, the implications of Kenneth Burke's remarks about antithesis (a trope of juxtaposed oppositional meaning):

> [I]magine a passage built about a set of oppositions ("we do this, but they on the other hand do that; we stay here; but they go there; we look up, but they look down," etc.) Once you grasp the trend of the form, it invites participation regardless of the subject matter. Formally, you will find yourself swinging along with the succession of antitheses, even though you may not agree with the proposition that is being presented in this form. ... Thus, you are drawn to the form, not in your capacity as a partisan, but because of some "universal" appeal in it. [3]

Burke shows very clearly here how cognitive inclination (swinging along) works linguistically. In doing so, he opens the question of how it is that minds function when they are exposed to the traditional furniture of rhetoric: broad patterns of reasoning and figuration.

What, we want to know, are the sources of the "universal" appeal that Burke identifies? Or, put in the terms of a highly rhetorical, ubiquitous contemporary praxis, why does that chocolate-bar jingle stay with you endlessly, despite your lack of interest in the chocolate bar, despite even your active dislike of the jingle? The intuitive answer is that there is something in the way your brain operates that allows the jingle to exploit it, colonize it, set up shop. And the intuitive answer is surely right.

Your brain functions with affinities for rhythm, repetition, similarity, and difference (among others). The jingle is rhythmic, repetitious, and contains patterns based on similarity and difference (rhyme, assonance, alliteration, ...); indeed, rhythm itself is a function of repetition, similarity and difference. The jingle, in short, insinuates itself into your consciousness by exploiting the submerged processes from which your consciousness emanates.

Look again at the three antitheses Burke offers:

- we do this, but they on the other hand do that;
- we stay here; but they go there;
- we look up, but they look down;

Antithesis is a trope (conceptual opposition), but it is quickly clear that antithesis is not the only figure at work here (figures very frequently work in concert). Other formal patterns suffuse the passage. I will give their rhetorical names, but what is important is the patterns those names identify. The passage includes parison (parallel syntax), isocolon (parallel prosody), epanaphora (clause-initial lexical repetition), and mesodiplosis (clause-medial lexical repetition), all of which contribute to the "swinging along" Burke's example illustrates, and all of which partake of the same universal character—universal for two reasons.

First, figuration is linguistically inescapable. Figures are omnipresent in language, communicating intentions and desires, coding information and attitudes, propagating belief and knowledge. There is no degree-zero, purely literal language. Language cannot but be figured; it flows in what Edward Sapir figured as "well-worn grooves of expression" ([26]). These grooves can be used or abused, optimized or overblown—which is where formal theories of figuration come in—but there are no other grooves. So, when we look to figuration, we look to primal organizing patterns of language use.

Second, and causally related to the first, figuration reflects the way our brains percolate and process. This reflection is perhaps most evident in the overtly purposive use of figures. The traditional literary purpose, generating aesthetic pleasure, is best known. But mnemonic formulas ("i before e except after c"), proverbs ("a bird in the hand is worth two in the bush"), oral traditions ("rosy-fingered dawn"), children's literature ("I meant what I said and I said what I meant / an elephant's faithful, one hundred percent")—in short, all linguistic configurations serving purposes in which cognitive functions like attention, learnability, and recall are at a premium—take a form that rhetorical theorists in the classical and early-modern periods identified as a figure. It is this insight, coupled with the high degree of overlap between the organizing principles of cognition and the organizing principles of figuration (similarity, contrast, balance, repetition, and the like), that motivates cognitive rhetoric and argues for ontological representation.

Let's turn to an extended example, the seemingly esoteric scheme of reverse lexical repetition, antimetabole.

3. ANTIMETABOLE

This figure (from Greek *anti* "in opposite direction" + *metabole* "turning about") is a poorly researched and largely unknown but remarkably widespread rhetorical device characterized by symmetrical lexical inversion. Its most famous modern exemplum may be John F. Kennedy's summarizing antimetabole in his inaugural address:

(1) Ask not what your country can do for you. Ask what you can do for your country

But it appears in almost every imaginable discourse:

Science

(2) If you press a stone with your finger, the finger is also pressed by the stone. (Newton)

(3) La vie, c'est le germe et le germe, c'est la vie. (Pasteur)

(4) [T]he wire moves in opposite circles round each pole and/or the poles move in opposite circles round the wire. Faraday).

Political Science

(5) It is not the consciousness of men that determines their being, but, on the contrary, their social being that determines their consciousness. (Marx)

(6) When the people fear their government, there is tyranny; when the government fears the people, there is liberty (Jefferson)

Algebra

(7) $m + n = n + m$; $m/n \neq n/m$

Logic

(8) $(p \,\&\, q) \leftrightarrow (q \,\&\, p)$

Advertising

(9) I am stuck on Band-Aid, and Band-Aid's stuck on me

(10) StarKist doesn't want tuna with good taste, StarKist wants tuna that tastes good. (Starkist Tuna)

(11) Friendly Americans with America friends. (United States Travel Service)

Ordinary Language

(12) When the going gets tough, the tough get going.

(13) Winners never cheat, and cheaters never win

(14) A place for everything, and everything in its place

Antimetaboles also show up in all the places one would normally expect to find a 'fancy' figure like this—fiction, poetry, philosophy, oratory, popular music (hip hop is full of antimetaboles). It is present in the ancient texts of China, India, Egypt, and Mesopotamia, and in the daily twitter feeds and status updates of the 21st century. They course through social media; Mark Zuckerberg's recent claim about Facebook, for instance:

(15) We don't build services to make money; we make money to build better services

It was tweeted and retweeted, posted and reposted, far and wide, despite its obvious questionability and clear promotional purpose.

Antimetaboles, in short, are rampant, which gives us some very significant facts about language and the mind—chiefly, the profound importance of cognitive affinities for all facets of communication. Cognitive linguists have solidly established the fundamental importance of what they call "conceptual metaphor" (we call them "analogic frames") and "conceptual metonymy" ("correlational frames"). But these affinities (that is, of the mind for similarities and correlations) are only two of the several mental dispositions that shape our perception, reasoning, memory, and communication.

Antimetabole leverages three cognitive affinities: symmetry, opposition, and repetition. One might note, to start, that all three of these affinities are important to the gestalt-theory perceptual tool-chest, which has been empirically ratified very thoroughly. But more specifically, we respond more favourably and recall more easily symmetrical patterns (symmetrical faces and bodies are judged more attractive than asymmetrical faces and bodies; abstract symmetrical graphics, such as the yin/yang, are recalled more quickly and robustly than asymmetrical graphics). The two cola of the antimetabole are symmetrical with each other. We categorize by similarities, of course, which is the affinity underlying metaphor, but also by opposition, and many base-level concepts are organized in oppositional dyads (up/down, in/out, adult/child), as the Burke passage illustrates. The lexical sequencing in antimetabole is formally opposite. At the deepest operational levels of the brain, there are repetitions of neuronal-population firing patterns, reflected at the level of cognition by the importance of repeated stimuli. Antimetaboles have a double repetition.

Antimetaboles are aesthetically pleasing, memorable, and culturally pervasive because they stack up three cognitive affinities. Other rhetorical figures leverage these affinities (and others) in a range of similar, different, and overlapping ways. Figure 1 illustrates some of these ways, which our RhetFig Ontology systematizes.

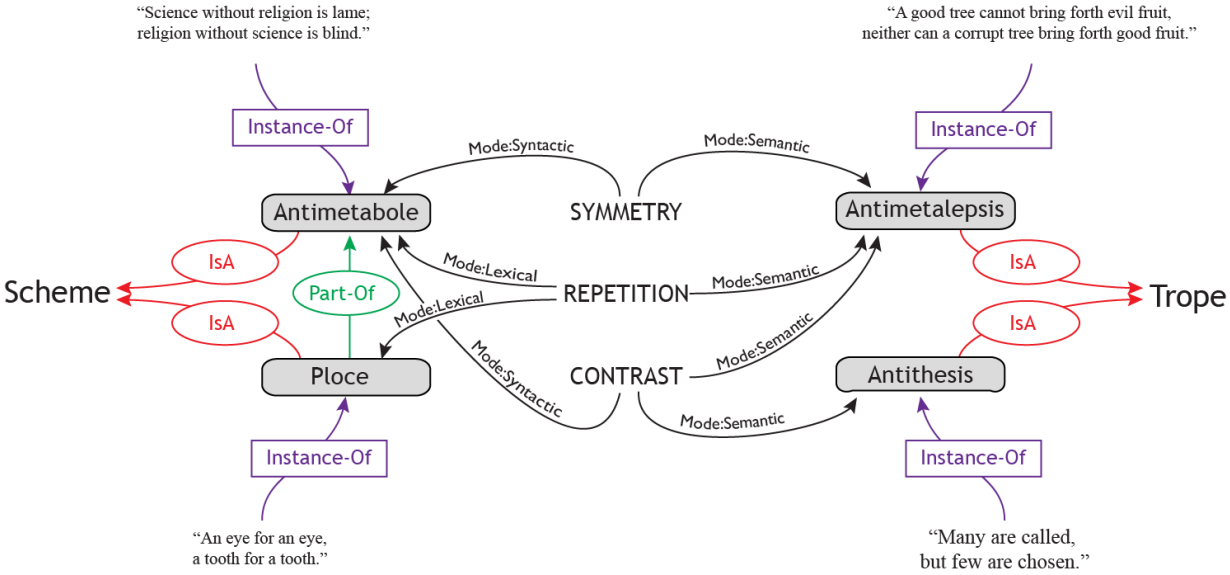

Figure 1. Ontological relations among figures, including antimetabole

4. FORM/FUNCTION PAIRINGS

If our work only went this far, one might fairly criticize it with a familiar line of attack against rhetoric, that it is 'merely stylistic,' that it tells us only about the peripheries of language, nothing about its conceptual depths. But most humanists, and all rhetoricians, realize that there is nothing *mere* about style, that style in fact offers different routes to those depths. Antimetabole charts three specific routes to a highly constrained, inter-related set of conceptual alignments ([10], [11], [25], [28]). Newton's and Pasteur's examples above are antimetaboles of reciprocality; the law of commutation, irrelevance-of-order; the "winners" and "place" ordinary-language examples, comprehensiveness.

What is compelling about this tight form/function coupling from a computational perspective is that once you find the form, the function is highly predictable. And the form of antimetabole, and of all schemes (tropes are a trickier matter), is computationally very tractable, in the sense that they can be modelled formally. Here is how Harris and Di Marco [16] render it as a Regular Expression (using the Waterloo Figure Representation formalism):

$$W^x \ldots W^y \ldots W^y \ldots W^x$$

In our Newton instance, for instance, W^x = finger, W^y = stone; the triple dots loosely express intervening elements (possibly null). This sort of representation is flexible and resilient—extensible to other constituents (phonemes, syllables, and phrases, as well as to words) and to all of the basic figurative operations (omission, addition, and substitution, in addition to iteration and permutation, as above), and it has proven value in computational research.

This representation is amenable to coding, and has been rendered in Perl, for instance ([17]). It has now been used for such purposes as text summarization ([1], [2]) and figure detection ([7], **Error! Reference source not found.**, [11], [17], [19], [24] [27]). But it was never meant as anything more than a sketch of possibilities; our project is regularizing the formalism, axiomatizing it with definitions of necessary concepts like identity, proximity, and sequence; and enforcing an overall rigour on the computational representations. Regular Expressions, too, only do half the job, by representing the pattern of figures; we need a parallel formalism for the functions they serve.

We are exploring how the necessary formalism may be found in Construction Grammar (CxG), a new morphosyntactic architecture developed over the last decade in close conjunction with Frame Semantics ([18]). In the Newton instance above (2), the function is reciprocality; Newton utilizes the antimetabole to express the utter balance of reciprocal forces. To the precise extent that the finger presses on the stone, the stone presses on the finger. Figure 2 captures this conceptual structure in CxG terms.

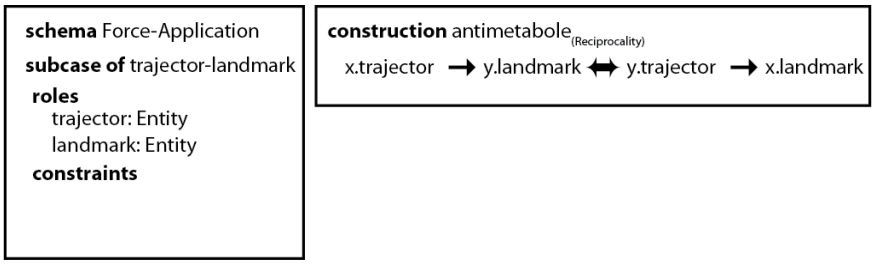

Figure 2. CxG representation of the reciprocal function of antimetabole

Figure 2 is only partially applicable, in that it captures antimetabolic instances like the Newton example (2), but fails with others, like the Pasteur instance (3) and the algebraic instance (7), in neither of which are their trajectors. But it provides a template for the sorts of CxG expressions needed for figural representation.

The abstract sequence of the antimetabole, represented in the Waterloo Figure Representation Notation above ([16], [25]), is precisely the kind of pattern a computational text analyser can find very easily; the abstract conceptual structure represented in Figure 2, and the combination of computational tractability with narrow functional range means that antimetaboles can be used to digitally diagnose major discursive themes and epitomize arguments. Here is where our work falls into the closest alliance with schools of Computational Argumentation.

Despite the hundreds of rhetorical figures that have been catalogued over two millennia, they fall into a relatively few, partially overlapping classes. While we have not worked out an exhaustive set of classes and relations, we are especially intrigued by the way in which the natural organizing principles of figures manifest well known cognitive affinities, like comparison, contrast, and symmetry, and by the interplay of well-known linguistic operations in the patterning of figures, like addition, deletion, and permutation. Our work applies these insights to the construction of a cognitive ontology of rhetorical figures.

5. ONTOLOGY DEVELOPMENT METHODOLOGIES

Rhetorical figures are prime candidates for a cognitive ontology. They are both richly, demonstrably cognitive, and richly, demonstrably valuable for NLP tasks like argument mining. The rhetorical tradition also suggests they hold considerable promise for a range of other automated text processing tasks, such as authorship attribution, genre detection, sentiment analysis, credibility assessment, pathology diagnosis and monitoring, voice interaction systems … virtually any activity that implicates pragmatics or style. Ontological representation seemed so obvious to us, so natural, so inevitable.

We started with a top-down approach. We used the categories: Cognitive Affinities (such as CONTRAST, SIMILARITY, SEQUENCE, REPETITION, and POSITION); Linguistic Domains (such as PHONEMES, SYLLABLES, MORPHEMES, WORDS, and PHRASES); and Figure Kind (such as TROPES and SCHEMES, as well as two somewhat novel categories, CHROMA and MOVES—see [4], [11]). Antimetabole, for instance, is a scheme of words utilizing affinities for repetition, sequence, and contrast, or <SCHEME; WORD; REPETITION, SEQUENCE, CONTRAST>.[10]

We followed this approach to considerable profit. We quickly realized that we would do better with multiple linked ontologies than trying to incorporate everything into one ontology, so we built a Cognitive Affinities ontology and an abbreviated Linguistic Domains ontology (we are hopeful that we will find existing ontologies that we can link to our RhetFig Ontology, and are aware in particular of multiple existing linguistic ontologies, but we have not had the chance to pursue this very far; we are still in the sandbox phase). Since OWL uses subsumption as its guiding organizational principle, we capitalized on the features of inheritance, which led us to reanalyse the conventional taxonomy of figures. So, for instance, there is figure known simply as *ploce* in the rhetorical tradition, a figure of unconstrained lexical repetition (that is, <SCHEME; WORD; REPETITION>), as in these instances:

(16) The best surprise is no surprise at all. (Holiday Inn marketing slogan)

(17) Villain, villain, smiling, damned villain! (Shakespeare, *Hamlet*)

But antimetabole, as well as a host of other lexical repetition figures, all of which are constrained in various ways, also satisfy the description <SCHEME; WORD; REPETITION>, except that they are constrained by other affinities, such as sequence or position. Antimetabole, for instance, is constrained by relative sequence (the repetitions must be in reverse order). Epanaphora is constrained by position. We developed a hierarchy corresponding to this realization, partially illustrated in Figure 3.

This new hierarchy provides us with a different understanding of figures in terms of their mutual relations based on cognitively principled categorization.

But the welter of rhetorical figures that have been catalogued over the millennia since their discovery soon made the top-down approach rather unwieldy and we opted for a middle-out development methodology, focussing only on figures of repetition. These include lexical repetitions (such as epanaphora and antimetabole), but also repetitions of phonemes (as in alliteration, repeating word-initial consonants), syllables (as in rhyme, repeating word-final syllables), morphemes (as in homoioptoton, repeating affixes), and so on. It also includes tropes such as synonymia (repeating concepts in different words), but we have restricted ourselves to schemes.

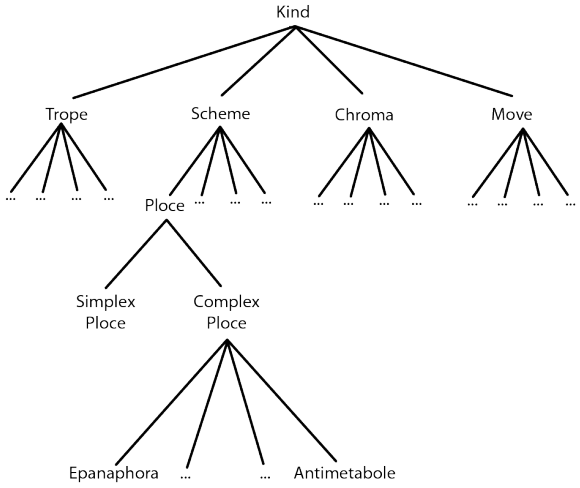

Figure 3. A new understanding of rhetorical figures of lexical repetition, based on the inheritance characteristics of ontologies

Again the approach was profitable, but again we ran into complications. On the profit side of the ledger, we found efficient ways to characterize the similarities and differences of figures. For instance, polyptoton is a figure in which the lemma is repeated but with different morphology, while in homoioptoton morphemes are repeated but with different lemmata, and (as you recall) ploce is a figure in which lexemes are repeated (i.e., both the lemma and the morpheme are repeated). Our use of linguistic domains in our categorization also allowed us to make further distinctions that had not been observed in the rhetorical tradition. Consider the following instances (all from James Baldwin's *Go Tell it on the Mountain*):

(18) It was later to become his proud testimony that he hated his sins—even as he ran towards sin. (Ploce)

(19) He wanted to be with these boys in the street, heedless and thoughtless. (Homoioptoton)

(20) You know honey, the Word tell us to hate the sin but not the sinner. (Polyptoton)

The formal structure of instances (18) and (20) superficially look the same. Both have the repeated lemmata with different morphemes. But (18) we categorize as ploce, because the morphology is inflectional, with negligible syntactic or semantic effects (*sins*, *sin*), while (20) is polyptoton because the morphology is derivational, with marked syntactic and/or semantic effects (*sin*, *sinner*). By categorizing these figures in the directions that OWL led us, we end up with a very efficient system that reminds one of the Principles and Parameters programme of Chomskyan linguistics:

Ploce <SCHEME; WORD-LEXEME; REPETITION>

Homoioptoton <SCHEME; WORD-LEMMA$_X$, MORPHEME$_Y$; REPETITION$_Y$, CONTRAST$_X$>

Polyptoton <SCHEME; WORD-LEMMA$_X$, MORPHEME$_Y$; REPETITION$_X$, CONTRAST$_X$>

[10] The formalism is as simple as it looks. The expressions are delimited with angled brackets. The main classes of information (figure type(s), linguistic domain(s), and cognitive affinity(/ies)), all mandantory, are separated by semi-colons. The specific type(s), domain(s), and affinity(/ies), when there are more than one, are separated by commas.

(Subscripts are markers of affiliation: for homoioptoton, MORPHEME goes with REPETITION; WORD-LEMMA goes with CONTRAST; vice versa for polyptoton.) Homoioptoton is a figure in which lemmata contrast, but a morpheme repeats. Polyptoton is a figure in which a lemma repeats, but morphemes contrast.

Ploce is, in these terms, a figure of simple lexical repetition, in which morphology goes unmentioned because it is inflectional. Homoioptoton and polyptoton both share the affinities of repetition and contrast, and both share the linguistic elements of word-lemma and morpheme. They differ only in terms of affinity-element assignments. The similarity between this treatment and the "switchbox" of Principles and Parameters ([5]) is suggestive on this level. Set the switches one way, you get polyptoton, set them another, homoioptoton. When the cognitive charting of figures advances to become fully cross-linguistic—as any cognitive attention to data must—we feel that there may be value in a framework of this type (though, of course, we can't know until the data is in *what* approach will be of value, and there is not even an initiative to collect such data). Free word order languages and fixed word order languages, for instance, will have different propensities toward figures implicating lexical sequence and figures implicating morphology.

The middle-out method soon hit a wall of diminishing returns as well—not, in this case, because of the great variety of figures but because of the limitations of OWL, and of description logics more generally, which present difficulties with capturing important notions like sequencing and indexing (since we need to represent relative and absolute order). We are familiar with some work that addresses some of these limitations through a tight and customized linking of ontologies. O'Reilly's work in particular ([23], [24]) modelled concepts and relationships such as hasFirstWord, hasLastWord, hasNextWord, and so on, in one ontology. Rhetorical Figures like epanaphora and epistrophe (clause-initial and clause-final lexical repetitions, respectively; they are types of complex ploce in our terms) went into another ontology, with Semantic Web Rule Language (SWRL) logic rules defining figures. While we feel these methods are productive, and we encourage O'Reilly to continue this work, others to emulate it, we are hopeful of more direct ways to express cognitive affinities like sequence and position. More importantly, we increasingly found that our instances were not falling into neat categories, and we changed to a data-centric approach, building the ontology from the bottom up.

We began by somewhat arbitrarily choosing a single instance of antimetabole, an exchange from Hemingway's *For whom the Bell Tolls*:

(21) "Coward," Pablo said bitterly. "You treat a man as a coward because he has a tactical sense. Because he can see the results of an idiocy in advance. It is not cowardly to know what is foolish."
"Neither is it foolish to know what is cowardly," said Anselmo, unable to resist making the phrase.

This seemed like an unassuming little antimetabole in which one character uses the reverse order of the figure in a kind of refutation of another character's utterance. We like it especially for the phrase that Hemingway uses to summarize Anselmo's motivation, "unable to resist making the phrase." Just as the first-one-order-then-the-other-order is the most natural and iconic way to represent the commutative principle in math and logic—i.e., (7) and (8), which exemplify the irrelevance-of-order antimetabolic function—a reversal of someone's utterance seems the most natural and iconic way to represent its repudiation. Hemingway, one of the most resolutely anti-rhetorical authors (an anti-style stylist, if you will), found the venerable antimetabole irresistible here to express the antipathy between his two characters, ascribing the compulsion to one of those characters.

Approaching categorization from the bottom had two unexpected benefits, discovering the multiplicity of figures and the multiplexity of figures.

The Hemingway instance (21), we soon realized, contains a multiplicity of figures, so that it was an instance not just of antimetabole, but of other figures as well, and if we were going to use our data to learn about the functional role of figures in language, we would have to be more thorough in our classifications. Fortunately, OWL intrinsically allows assigning one instance to multiple classes. Some of these other figures seemed somewhat incidental to the refutational function of the antimetabole—polyptoton, for instance (*coward*, *cowardly*), and epanaphora (*because he …*, *because he …*). Others are more integral—in particular, mesodiplosis (clause-medial lexical repetition; here *… to know what is …*, *… to know what is …*). Indeed, as we looked through our data, we came to realize that most of the most clearly functional antimetabolic instances were not just reverse lexical repetitions but reverse lexical repetitions around another stable lexical repetition. That is, they are more fully represented in the following terms (where the W^x and W^z constituents represent the antimetabole, the W^y constituents represent the mesodiplosis):

$$W^x \ldots W^y \ldots W^z \ldots W^z \ldots \ W^y \ldots W^x$$

If you look back at the antimetabolic instances in this paper, armed with this insight, the structure is impossible to miss. For (2) the mesodiplosis constituent is *press*, for (4) it is *moves in opposite circles*, for (5) it is *that determines their*, for (6) it is *fear*, for (7) it is the addition and division signs, and so on.

Moreover, some figures have particularly strong bonds with one another. Antimetabole manifests a syntactic (sequential, formal) opposition. Antithesis manifests semantic (conceptual, ideational) opposition. So, the two frequently co-occur, reinforcing each other. A huge proportion of antimetaboles are also antitheses, like the famous Kennedy aphorism (1), as well as the Marx (5) and Zuckerberg (15) instances. More compellingly, the combination of figures usually doubles down on the functional constraints. All of the antimetabole-antithesis expressions we have found combine the comprehensiveness function of the first figure with the rejection function of the second figure. Adding antithesis to antimetabole, that is, always produces an utter rejection of one proposition, concomitant with its utter replacement by the other, either assertively (Kennedy, Marx) or implicationally (Zuckerberg).

Rhetorical functions, in short, are very often the result of specific compounding effects. Figures work in multiplicities.

The data, we also discovered through the careful attention that ontological considerations forced on us, is multiplex. In particular, we found as we worked through the instances, especially in our focus on linguistic domains, that what we thought was one figure, antimetabole (and what the rhetorical tradition often treated as one figure, sometimes two) was in fact at least seven distinct patterns, which we identify as follows (some of the instances are repeated from above, under this new categorization; some of them are original):

Antimetathesis

(22) She sells sea shells by the seashore.

Antimetabole

(23) It's Trump calling real journalism "fake news" and holding up "fake news" as real journalism. (tweeted by (@ExportedFromMI)

Morphological Chiasmus

(24) Friendly Americans with America friends.

Syntactic Chiasmus

(25) Despised, if ugly; if she's fair, betrayed. (Leapor. "Essay on Woman")

Antimetalepsis

(26) [T]he new metaphor — the brain as a calculating machine — is ... fraught with the danger of anthropomorphizing the machine and mechanizing the man.

Pseudo-Antimetabole

(27) StarKist doesn't want a tuna with good taste. StarKist wants a tuna that tastes good.

Implied Antimetabole

(28) Britannia waives the rules [evoking the familiar inverse phrase, "Britannia rules the waves."].

We call this grouping of figures *The Chiastic Suite*, after one of the related terms for *antimetabole* from the rhetorical tradition, *chiasmus*, itself named after the Greek letter chi (χ, the repeating elements are said to 'cross' each other).

There's more going on here than just a new data story. Identifying this cluster of figures from the instances (each and every one of these configurations has been held to exemplify antimetabole, or one of its synonyms, in the tradition) allows us to 'purify' antimetabole for analytic terms. This is most obvious with pseudo-antimetabole. It is a *pseudo*-antimetabole because the two occurrences of *taste* in our example evoke different meanings (therefore, different words). The first word is *taste* in cultured appreciation of the arts sense, the second is *taste* as in the gustatory sense. There is, in other words, no repetition of *taste*; rather, there is a pun (formally, a paronomasia), a 'repetition' of the semiotic vehicle without the semiotic interpretant, along with a repetition of *good*. If one builds an automatic detector for antimetabole in order to do argument mining and comes across an instance like (27), the algorithm needs to be able to filter it out.

On the other hand, antimetalepses pattern conceptually very much like antimetaboles, which suggests one should widen the search pattern to include instances of 'crossed' synonymia, for argument mining. Instance (26), for example, serves a comprehensiveness function; the claim of the sentence is that the mind-as-machine metaphor that was so prevalent in early cognitive science is completely dangerous, because it misconceives minds and it misconceives machines. We don't have the space to map out all of the chiastic suite—and, to be frank, we haven't fully done so; the interactions are tricky, calling for considerably more investigation—but they demonstrate the productivity of cognitive ontologies for research into linguistic patterns.

6. CONCLUSION

Cognitive ontologies are highly productive, and vastly under-utilized, research tools for ANLP, especially in matters of style, pragmatics, and argumentation. Our research project demonstrates their virtues in multiple ways. Figure ontologies provide mechanisms for plumbing texts. And they reveal unsuspected features of linguistic configurations. But this research is in very early days. Our future plans for the project include (i) exploring other ontological representations, (ii) developing our annotation scheme, (iii) linking other ontologies, (iv) applying our bottom-up methodology in other figurative domains, (v) expanding our research to include other categories of rhetorical figures, beyond schemes, and (vi) expanding and refining our detection engine.

i. OWL has been particularly restrictive in its limitation to the IsA relation, which is insufficient both for the relations among many figures (including, for instance, various kinds of meronymy) and for the kind of information we foresee necessary for applications. So, we are looking at ways to incorporate other relations in OWL, and to augment OWL's expressiveness in other ways. But we are not committed to OWL. So we are also looking for other representational languages.

ii. The annotation scheme has multiple challenges. We only annotate the *form* of the figures at this stage, but annotating their functions is equally important, perhaps more so for activities like argument mining and text summarization. The tendency for figures to cluster in the same stretches of also makes for very messy annotation, so we are exploring stand-off markup. In both instances, we will ensure standardization as best we can, to increase the usefulness of our work, by following Apache UIMA (Unstructured Information Management) guidelines.

iii. We have been unsuccessful at finding other ontologies we can link to ours, and plan to expand our efforts at finding them, especially linguistic and cognitive ontologies.

iv. The bottom-up strategy has its limitations, but has been the most productive methodology we have adopted. We began with a single instance, but it was not arbitrarily chosen. It was an antimetabole, which we knew had various interesting implications and it has led to the discovery of the chiastic suite. We plan to begin another 'mini-ontology,' bottom-up with instances of the figures gradatio, because of its relation to the scheme anadiplosis (anadiplosis is lexical repetition at the end of one phrase and the beginning of the following phrase; gradatio is two or more successive anadiploses, so anadiplosis is a PART-OF gradatio), such tropes as incrementum and decrementum (successive words with semantic inclines or declines), and the compound figure (part scheme, part trope), climax (gradatio + incrementum). We apologize for the necessarily enigmatic description of these plans Adequately defining and exemplifying the figures here would require too much space. But we invite interested readers to consult Fahnestock's ([8]) chapter on incrementum and gradatio for more detail about some of these figures, many examples, and an account of how they function in argumentation.

v. Work on a gradatio mini-ontology, as we've just suggested, will involve the inclusion of tropes and compound figures, but there is room for expansion into other tropes, into chroma as well, and other figures as well. There has been considerable work on Argument Schemes in argument mining, for instance, a nice overview of which can be found in a recent paper by Argument Scheme guru, Douglas Walton, with Fabrizio Macagno ([29]). There are many figures, chiefly the ones we call Moves, that are strikingly similar to Argument Schemes (for instance, prolepsis, which is the projection and refutation of a counter-argument, and paralipsis, the insinuation of premises

while decrying their inclusion in an argument), and we envision a convergence of our research with Walton's line, starting with the incorporation of moves into our ontology.

vi. Jakub Gawryjolek developed JANTOR (Java ANnotation Tool Of Rhetoric) in 2009 ([11], [12]) at the outset of our project, an ambitious and accomplished but preliminary tool for the detection and annotation of rhetorical figures in HTML files. We plan to update and enhance JANTOR and make it available for other research groups

REFERENCES

[1] M. Alliheedi and C. Di Marco. Rhetorical Figuration as a Metric in Text Summarization. *Advances in Artificial Intelligence, Lecture Notes in Computer Science* 8436: 13-22 (2014).

[2] M. Alliheedi. *Multi-document Summarization System using Rhetorical Information.* M.S. thesis David Cheriton School of Computing, University of Waterloo, Waterloo ON Canada. (2012).

[3] K. Burke. *Rhetoric of Motives.* University of California Press, Berkeley, CA (1969).

[4] L. Chien and R. A. Harris. Scheme Trope Chroma Chengyu: Figuration in Chinese Four-Character Idioms. *Cognitive Semiotics* 10(6):155-178. (2011).

[5] A. N. Chomsky. *New Horizons in the Study of Language and Mind.* Cambridge University Press, Cambridge, UK. (2000).

[6] J. Crosswhite. Rhetoric and computation, Symposium on Argument and Computation. Bonskeid House, Perthshire, Scotland. June 27, 2000.

[7] M. Dubremetz and J. Nivre. Rhetorical Figure Detection: The Case of Chiasmus. *Proc. of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature, Denver, CO, June 4*. (2015).

[8] J. Fahnestock. *Rhetorical Figures in Science*. Oxford University Press, New York (1999).

[9] J. Fahnestock. Figures of Argument. *Informal Logic* 24.2:115-135. (2004).

[10] J. Fahnestock. Rhetoric in the Age of Cognitive Science. In: *The Viability of Rhetoric*. Richard Graff (Ed). State University of New York Press, New York, 159-180 (2005).

[11] J. Gawryjolek. *Automated Annotation and Visualization of Rhetorical Figures.* MSc thesis. David Cheriton School of Computing, University of Waterloo, Waterloo ON Canada. (2009).

[12] J. Gawryjolek, C. Di Marco, and R. A. Harris. An Annotation Tool for Automatically Detecting Rhetorical Figures. *CMNA (Computational Models of Natural Argument),* 13 July 09, Pasadena, CA. (2009).

[13] F. Grasso. Towards Computational Rhetoric. *Informal Logic* 22: 195-229. (2002).

[14] F. Grasso. Towards a Framework for Rhetorical Argumentation. EDILOG 2002 - *Proceedings of the 6th workshop on the semantics and pragmatics of dialogue*, Edinburgh, UK, 4-6 September 2002, p. 53-60.

[15] R. A. Harris. Figural Logic in Mendel's Experiments on Plant Hybrids. *Philosophy and Rhetoric* 46.4: 570-602. (2013).

[16] R. A. Harris and C. Di Marco. Constructing a Rhetorical Figuration Ontology. *AISB (Artificial Intelligence and Simulation of Behaviour)*, 8 April 09, Edinburgh, Scotland. (2009).

[17] D. H. Hromada. Initial Experiments with Multilingual Extraction of Rhetoric Figures by Means of PERL-Compatible Regular Expressions. *Proc. of the Second Student Research Workshop associated with RANLP*, Hissar, Bulgaria. (2011).

[18] T. Hoffmann and G. Trousdale (Eds). *The Oxford Handbook of Construction Grammar*. Oxford University Press, New York (2013).

[19] A. R. Kelly [now Mehlenbacher], N. Abbott, R. A. Harris and C. Di Marco. Toward an Ontology of Rhetorical Figures. *Proc. of the 28th ACM International Conference on Design of Communication*, 123-130. (2010).

[20] M. Lippi and P. Torroni. Argumentation Mining: State of the Art and Emerging Trends. ACM Transactions on Internet Technology 16.2:1-25

[21] W. C. Mann and S. A. Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8.3:243-281. (1988).

[22] M. Mladenović and J. Mitrović. Ontology of Rhetorical Figures for Serbian. *Text, Speech, and Dialogue*. Volume 8082 of the series *Lecture Notes in Computer Science*, 386-393. (2013).

[23] C. O'Reilly. *Lassoing Rhetoric with OWL and SWRL*. MSc thesis, Computer Science, University of Westminster, London. (2010).

[24] C. O'Reilly and S. Paurobally. Lassoing Rhetoric with OWL and SWRL. Unpublished. Available http://www.academia.edu/2095469/Lassoing_Rhetoric_with_OWL_and_SWRL (2010).

[25] S. Ruan, C. Di Marco, and R. A. Harris. *CMNA XVI - Computational Models of Natural Argument.* New York, NY, 9 July. (2016).

[26] E. Sapir. *Language: An Introduction to the Study of Speech*. Harcourt, Brace, New York (1921).

[27] C. Strommer. *Using Rhetorical Figures and Shallow Attributes as a Metric of Intent in Text*. PhD dissertation, David Cheriton School of Computing University of Waterloo, Waterloo ON Canada. (2011).

[28] C. W. Tindale. *Acts of Arguing: A Rhetorical Model of Argument,* State University of New York Press, Albany, NY (2000).

[29] D. Walton and F. Macagnob. A Classification System for Argumentation Schemes. *Argument and Computation* 6.3:219-245.

Representing Types through Image Schemas and Patterns

Claudio Masolo and **Emilio Sanfilippo**[1]

Abstract. We propose a formal framework to represent types intended as (intentional) complex properties characterised in terms of (i) simple *qualities*, e.g., colors or weights; (ii) *structural properties*, i.e., properties concerning the spatial distribution of qualities, e.g., being polka-dotted or being uniformly dense; and (iii) *structured properties*, i.e., properties concerning the topological arrangement of the components of assemblies. Our framework is inspired by work done in cognitive science, in particular by Gärdenfors' theory of conceptual spaces, as well as by the notion of image schema, the latter being a theoretical construct to represent general patterns and the way they apply to heterogeneous cases. Hopefully, our approach can also contribute to the understanding of the notions of image schema and conceptual space.

1 Introduction

The formal representation of *types* is of fundamental relevance for conceptual modelling and knowledge representation purposes [17]. In the context of engineering design, for example, it is common to distinguish between *artefacts* as individual spatio-temporal objects, and *artefact types* as complex properties that classify individual artefacts. Take, e.g., John's and Mary's Fiat500 cars, namely two distinct objects. Classifying them as artefacts of the type Fiat500 means that they have a certain shape, weight, height and spatial arrangement of components that are characteristic of the Fiat500 type. Generalising from design, the distinction between types and their corresponding instances apply to disparate categories commonly used in knowledge representation, e.g., events, among others.

The modelling of types is challenging. Philosophers have been discussing about the nature of properties since the early days of philosophical disputes, e.g., whether they are mind-independent universals, or whether they reduce to human ways of categorising phenomena.[2] The latter view is at the core of cognitive studies according to which knowledge acquisition and perception are guided from (or give rise to) some sort of *conceptual structures* [4, 20].

Departing from a metaphysical stance, we propose an approach for the representation of types grounded on cognitive theories. More specifically, we shall focus on the *compound* nature of types, namely, the fact that they are characterised in terms of properties concerning (i) qualitative aspects (*qualities*), (ii) the distribution of such qualities across space and/or time (*structural properties*), or (iii) other types each one classifying the component of a topologically structured assembly (*structured properties*).

The proposed framework is based on a modified version of Gärdenfors' theory of conceptual spaces [4], the latter being well-suited to represent types as multi-dimensional properties, therefore as being composed of other properties. The framework is also inspired by studies on *image schemas* (e.g., [9, 11]). Even though the latter are challenging to be captured in detail, there is a relative consensus in understanding them as sorts of (very general) patterns. Johnson claims that an image schema is "a dynamic pattern that functions somewhat like the abstract structure of an image, and thereby connects up a vast range of different experiences that manifest this same recurring structure" [9, p.2]. According to Langacker, image schemas are "schematized patterns of activity abstracted from everyday bodily experience, especially pertaining to vision, space, motion, and force" [11, p.42]. Recent approaches devoted to formal aspects describe image schemas as "patterns abstracting from spatio-temporal experiences" [10, p.155], or "mental patterns [that] may be combined with each other to generate more complex structures" [7, p.21]. We find image schemas useful theoretical constructs to represent recurrent configurations and the way they apply, in different contexts and with different modalities, to heterogeneous cases. The paper is not a contribution to clarifying or formalising conceptual spaces or image schemas; it rather proposes a way to re-elaborate some approaches in cognitive studies to deal with some knowledge representation issues, as done, e.g., in [10]. Hopefully, however, we can also shed some light on the very notions (and limitations) of conceptual space and image schema.

The paper is structured as follows. In Sect. 2 we introduce the main motivations behind our work. Sect. 3 presents the overall framework for the representation of qualities. Sect. 4 introduces the machinery to represent image schemas, which are used in Sect. 5 to represent patterns of qualities. Sect. 6 presents how patterns may be used to classify the structural properties of objects according to either time or space. Finally, Sect. 7 shows how structured properties and complex types can be represented in the proposed framework.

2 Properties: Types and qualities

In conceptual modelling and knowledge engineering, one commonly distinguishes between *individuals* and *properties*, including a predication or classification relationship holding between them. Recalling the example mentioned in the introduction, John's and Mary's cars are two individuals of the same type Fiat500, namely, they are both classified by the same (complex) property.

Properties bear an *intensional* nature, i.e., differently from sets, they do not reduce to their members (extensions). As a consequence, different properties can classify the same individuals, e.g., *being triangular* and *being trilateral*.[3] We distinguish two main kinds of

[1] Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy, email: {masolo,sanfilippo}@loa.istc.cnr.it

[2] Roughly, this is the distinction between realists and nominalists concerning the metaphysics of properties.

[3] The intensional characterisation of properties is relevant for different domain-specific scenarios. For example, in the context of design, artefact types (properties) are not usually specified by listing their instances but by sets of constraints on relevant characteristics.

properties, namely *types* and *qualities*. Types are *compounds* of simpler properties. Among these properties there are the qualities that are the simplest properties, i.e., they are not further decomposable.[4] For example, the type Fiat500 is composed by qualities like colour, shape and weight, among others, which constraint the holistic properties of Fiat500 cars.

Some types classify assemblies with a specific spatial structure. Fiat500 cars, for instance, consist of disparate components properly arranged, e.g., bodywork, engine and wheels. From this perspective, types do not reduce to compounds of qualities; they are characterised also in terms of what Armstrong calls *structured properties*. Similarly, Fiorini and colleagues [21, 22] claim that objects recognition and categorisation is not only based on the global (aka holistic) features that objects manifest, but it is also achieved by identifying objects' structures in terms of their parts and the way in which parts are arranged. A theory of cognition has, therefore, to deal with conceptual structures about both holistic and structural information.

From a representational perspective, we propose to model the type/quality distinction following the *similarity space theory of concepts* [20] and, in particular, the theory of *conceptual spaces* introduced by Gärdenfors [4]. The latter theory provides the basic framework to convey the intensional nature of properties, to characterise the complex nature of types in opposition to qualities, as well as to compare different types on the basis of their degree of similarity.

Gärdenfors [4] represents *concepts*—e.g., car, apple, etc.—as regions in a *space* obtained by composing a given number of *domains*—e.g., colour, taste, etc. The main reason to decompose a space into domains is the assumption that the properties in the domains may be attributed to individuals *independently* of the properties in other domains; e.g., the weight of an individual is independent of its temperature or colour.[5] Additionally, conceptual spaces have the peculiarity to be endowed with a distance relation representing degrees of similarity: the closer are the properties (within a domain), the more similar are the individuals that exhibit such properties.

The representation of structured properties has proven challenging in conceptual spaces. In some studies, Fiorini and colleagues [21, 22] propose a theory of part-whole relations with the purpose of grounding structural relations in cognition. The authors extend Gärdenfors' theory with *structural spaces*, which allow for the representation of the arrangements of the parts within a whole according to certain configurations.[6] However, the proposed framework has a mathematical and combinatorial nature, whereas the modality according to which the general patterns apply to specific cases is missing. Furthermore, the spatial structure of objects is not only relevant to represent how types (and the entities they classify) relate in a part-whole manner. Indeed, it also plays a fundamental role in cases in which types are characterised by *distributional* properties [18]. Consider, e.g., a car-type whose bodywork-type is characterised by a colour lozenge pattern distributed across the overall instances of the bodywork-type. In a design perspective, there are no components of the bodywork-type, each being coloured in a certain way. We consider these properties, called *structural* properties, in Sect. 6. It remains unclear how Fiorini and colleagues would approach this case.

[4] The distinction between types and qualities is reminiscent of the one, between *classes* and *attributes* assumed in the field of conceptual modelling, as, e.g., in the UML class-diagram [17]. Furthermore, some qualities, e.g., colours, may be compound, too [4]; we do not consider this aspect here.

[5] The domains are not totally independent, i.e., they may be *correlated*.

[6] A point in a structural space represents also the particular configuration of the parts of an object [21].

3 Qualities

We start by introducing two disjoint kinds of properties: $\mathrm{QT}(x)$ stands for "x is a *quality*", and $\mathrm{TY}(x)$ stands for "x is a *type*". This distinction roughly reflects the one between simple and compound properties, i.e., in terms of conceptual spaces, the distinction between properties (in a domain) and concepts. Qualities are partitioned into a finite number η of *domains* D_i, e.g., the domains of colour, shape and weight. Axiom (Ax1) guarantees that all the qualities belong to at least one domain, while (Ax2) assures that the domains are disjoint.

Ax1 $\mathrm{QT}(x) \leftrightarrow \bigvee_{i=1}^{\eta} \mathrm{D}_i(x)$

Ax2 $\bigwedge_{i\neq j=1}^{\eta} (\mathrm{D}_i(x) \rightarrow \neg \mathrm{D}_j(x))$

For instance, $\mathrm{D}_{\mathrm{colour}}(\texttt{red})$ states that the quality *being red* belongs to the colour-domain, while $\mathrm{D}_{\mathrm{weight}}(\texttt{1506kg})$ that the quality *being 1506kg heavy* belongs to the weight-domain.[7]

Every quality domain has a *top-quality*, i.e., a quality that subsumes all the qualities in the same domain (Ax3). The top-quality of a domain is noted with the name of the domain, e.g., the top-quality of $\mathrm{D}_{\mathrm{weight}}$ is noted with `weight`.

Ax3 $\bigwedge_{i=1}^{\eta} \exists d \forall q (\mathrm{D}_i(q) \rightarrow q \sqsubseteq d)$

We include *time* and *space* among the quality domains represented, respectively, with TM and SP (rather than $\mathrm{D}_{\mathrm{time}}$ and $\mathrm{D}_{\mathrm{space}}$). This move is motivated by the fact that time and space are structured similarly to other domains. However, we attribute them a peculiar role in the classification relation, since they *qualify* the classification (see Sect. 3.1). Recall that some philosophers advocate the ontological primacy of time and space. In particular, spatial relations are claimed to not be reducible to property-grounded relations [8]. We do not however enter into this discussion; our choice is just aimed at simplifying the formalisation. Proper qualities, noted as PQT and simply called qualities, are the qualities that are neither times nor spaces.

Domains contain comparable qualities. More generally, one can think of domains as sets of intensionally interlinked qualities. We introduce a (intensional) *subsumption* relation between qualities belonging to the same domain; see (Ax4), where $x \sqsubseteq y$ stands for "the quality x *is intensionally subsumed* by the quality y". D_i-qualities cannot be subsumed by D_j-qualities (with $i \neq j$). For example, `scarlet` cannot be subsumed by any quality in $\mathrm{D}_{\mathrm{weight}}$; it could be rather subsumed by `red`, i.e., $\texttt{scarlet} \sqsubseteq \texttt{red}$.

Ax4 $x \sqsubseteq y \rightarrow \bigvee_{i=1}^{\eta} (\mathrm{D}_i(x) \wedge \mathrm{D}_i(y))$

Proper subsumption $\sqsubset$ is defined in (Df1). Formally, $\sqsubseteq$ is a discrete and atomic partial order: it is reflexive, antisymmetric, transitive, discrete, and atomic, i.e., every quality subsumes an *atomic* quality, a quality that does not properly subsume any other quality (Df2). Ontologically, atomic qualities represent how the world is; epistemologically, they represent the maximal resolving power one disposes of, e.g., the finest resolution of measurement devices.

Df1 $x \sqsubset y \triangleq x \sqsubseteq y \wedge \neg y \sqsubseteq x$

Df2 $\mathtt{AT}_{\mathrm{QT}}(x) \triangleq \mathrm{QT}(x) \wedge \neg \exists y (y \sqsubset x)$

Quality domains may be structured according to further relations. Images schemas often refer to topological or order structures. We avoid to introduce a full metrics, as usually done for conceptual spaces, by assuming all domains to be endowed with at least the *con-*

[7] We note individual constants using the `typewriter` font.

nection relation $\bowtie$.[8] *Ordered* domains, e.g., TM, $\mathrm{D}_{\mathrm{weight}}$, or $\mathrm{D}_{\mathrm{length}}$, are also (linearly) ordered by the *precedence* relation $\prec$. Both connection and precedence hold only among qualities in the same domain (Ax5)-(Ax6).

Ax5 $x \bowtie y \rightarrow \bigvee_{i=1}^{\eta}(\mathrm{D}_i(x) \wedge \mathrm{D}_i(y))$

Ax6 $x \prec y \rightarrow \bigvee_{i=1}^{\eta}(\mathrm{D}_i(x) \wedge \mathrm{D}_i(y))$

Notice that qualities may subsume different (atomic) qualities, therefore, times, weights, colors, shapes, etc. are more similar to regions rather than to points, i.e., $\sqsubseteq$ could be interpreted mereologically. In this perspective, precedence ($\prec$) could be seen as the disjunction of the relations of *meet* and *before* introduced in [1], while the RCC calculus [23] may be considered for connection ($\bowtie$). The detailed axiomatic treatment of $\prec$ and $\bowtie$ is not relevant for our purposes; we want just to highlight the possibility to have structural relations (in addition to subsumption) that are defined in several (or all) quality domains. As we will see, this aspect is fundamental to characterise image schemas and patterns.

3.1 Classification

We analyse hereby how qualities can be attributed to individuals. Since qualities are in the domain of quantification, the standard predication mechanism of FOL cannot be adopted. It is thus necessary to introduce a new primitive relation, named *classification*, and to establish under what conditions individuals are classified by qualities. Here we focus on the classification of *physical objects* (noted OB), aka *continuants* or *endurants*, that are in, and can change through, space and time, e.g., my car, Barack Obama, the earth, etc. The framework can be however applied, with minimal tuning, also to *events*, aka *occurrents*, *perdurants*.

We start from a *local* notion of *direct classification* under qualities: $\mathtt{dCF}_{\mathrm{QT}}(q, x, s, t)$ stands for "the (proper) quality q directly classifies the object x *as it is at space* s *and time* t".[9] Given the possibility for objects to change through both time and space, the classification is spatio-temporally *qualified*.[10] For instance, John can decide to paint his red volleyball half in orange and half in blue.

Ax7 $\mathtt{dCF}_{\mathrm{QT}}(q, x, s, t) \rightarrow \mathrm{PQT}(q) \wedge \mathrm{OB}(x) \wedge \mathrm{SP}(s) \wedge \mathrm{TM}(t)$

One could interpret $\mathtt{dCF}_{\mathrm{QT}}$ taking a *mereological* stance: when $\mathtt{CF}_{\mathrm{QT}}(q, x, s, t)$ holds, the entity that has the quality q is not x but a different object, namely the part of x that, at t, is (exactly) located at space s. We assume a weaker position that does not equate the resolution of space with the one of objects, i.e., atomic objects can be located in non-atomic spaces. We come back to this point in Sect. 7.

By relying on subsumption, (Df3) defines the relation of *local indirect classification*: q indirectly classifies x as it is at space s and time t when x is locally classified under a quality that is (properly) subsumed by q. For example, if $\mathtt{crimson} \sqsubset \mathtt{red}$ and $\mathtt{dCF}_{\mathrm{QT}}(\mathtt{crimson}, \mathtt{jball}, \mathtt{s}, \mathtt{t})$, then $\mathtt{iCF}_{\mathrm{QT}}(\mathtt{red}, \mathtt{jball}, \mathtt{s}, \mathtt{t})$ ($\mathtt{jball}$ stands for John's volleyball).

Df3 $\mathtt{iCF}_{\mathrm{QT}}(q, x, s, t) \triangleq \exists z(\mathtt{dCF}_{\mathrm{QT}}(z, x, s, t) \wedge z \sqsubset q)$

(Df3) captures a disjunctive reading of classification: it is enough to be directly classified under *one* of the qualities that are subsumed by q to be indirectly classified under q. For instance, suppose $\mathtt{crimson} \sqsubset \mathtt{red}$ and $\mathtt{scarlet} \sqsubset \mathtt{red}$. Both scarlet and crimson entities are indirectly classified under $\mathtt{red}$. The difference between direct and indirect classification is *cognitively* and *empirically* relevant. Direct classification is the result of a direct observation or sensation while indirect classification relies on an abstraction process. Objects can be thus directly classified under non-atomic qualities, i.e., it is possible to have partial or general information due to the resolution at which the world is accessed. It is possible, for instance, to know that John's volleyball is red without knowing its exact shade. This means that direct and indirect classification are not mutually exclusive. For instance, in an empirical context, it is possible to have measurements about the same object taken with devices that have different resolutions. This is a departure from the Gärdenfors' theory of conceptual spaces where an object is always represented by one single point in a space.[11] The general relation of local classification abstracts from the direct vs. indirect distinction (Df4).

Df4 $\mathtt{CF}_{\mathrm{QT}}(q, x, s, t) \triangleq \mathtt{dCF}_{\mathrm{QT}}(q, x, s, t) \vee \mathtt{iCF}_{\mathrm{QT}}(q, x, s, t)$

To clarify what the qualification "at space s and time t" means we introduce the primitive relation of *being present*: $\mathtt{PRE}(x, s, t)$ stands for "the object x *is present* at the spatio-temporal region identified by space s and time t", i.e., "at t, x occupies s" (Ax8).[12] As said, we focus on spatio-temporal objects, i.e., objects that are present both in time and space (Ax9). The local classification (at s and t) of an object requires its presence (at s and t), see (Ax10).

Ax8 $\mathtt{PRE}(x, s, t) \rightarrow \mathrm{OB}(x) \wedge \mathrm{SP}(s) \wedge \mathrm{TM}(t)$

Ax9 $\mathrm{OB}(x) \rightarrow \exists st(\mathtt{PRE}(x, s, t))$

Ax10 $\mathtt{CF}_{\mathrm{QT}}(q, x, s, t) \rightarrow \mathtt{PRE}(x, s, t)$

First note that being present at s and t is not equivalent to being present at s *and* being present at t. Second, and more importantly, being present at s and t is not a disjunctive abstraction from being present at some atomic subregions of s and t.[13] For PRE we have a sort of conjunctive reading: being present at s and t requires to be present at every subregion (Ax11).[14]

Ax11 $\mathtt{PRE}(x, s, t) \wedge r \sqsubset s \wedge u \sqsubset t \rightarrow \mathtt{PRE}(x, r, u)$

Spatial location (at t) is defined in (Df5).[15] (Df6) introduces the classification of an object as it is at its whole spatial location, i.e., it determines the *global* (aka *holistic*) qualities of an object.

Df5 $\mathrm{L}_{\mathrm{SP}}(x, s, t) \triangleq \mathtt{PRE}(x, s, t) \wedge \neg\exists r(\mathtt{PRE}(x, r, t) \wedge s \sqsubset r)$

Df6 $\mathtt{CF}_{\mathrm{QT}}(q, x, t) \triangleq \exists s(\mathrm{L}_{\mathrm{SP}}(x, s, t) \wedge \mathtt{CF}_{\mathrm{QT}}(q, x, s, t))$

Extensional relations between qualities can be introduced by relying on the global $\mathtt{CF}_{\mathrm{QT}}$. By (Df7), $x \subseteq y$ (read as "x is (extensionally) *included* in y") holds when every entity globally classified at

[8] Some domains could have alternative topological organisations, e.g., the colour spindle or the RGB colour wheel, see [4]. In the following we assume domains to have a unique topological organisation.

[9] Local classification is much more informative than holistic classification.

[10] Classification could also be qualified with proper qualities. For instance, $\mathtt{dCF}_{\mathrm{QT}}(\mathtt{10kg}, x, \mathtt{red}, t)$ would be read as "x weights 10kg as it is at colour red and time t". We leave this interesting extension to future work.

[11] See [5, 14] for some criticisms about this assumption.

[12] Here we assume a sort of container-like notion of space. One could substitute spaces with *places*, spatial entities that are more abstract than spaces and that can be defined relatively to physical objects, e.g., the top, the interior, etc., see [2]. Similarly for times.

[13] As for the characterisation of types (see Sect. 7), approximate localisation in space-time can be represented by means of rough sets (see [14]).

[14] One could think to introduce a *conjunctive* classification primitive cCF that behaves similarly to PRE. For instance, $\mathtt{cCF}(\mathtt{red}, x, s, t)$ would imply a classification by all the color-qualities subsumed by red, e.g., $\mathtt{cCF}(\mathtt{scarlet}, x, s, t)$, $\mathtt{cCF}(\mathtt{crimson}, x, s, t)$, etc. In Sect. 6 we introduce structural properties, that, in our opinion, are stronger tools to characterise properties that require an object to have different qualities of the same kind.

[15] The location is unique only when specific constraints on SP hold.

any time by the quality x is also globally classified, at the same time, by y. For example, assume that all the red entities in the domain of quantification also weigh 1kg, then we have $\texttt{red} \subseteq \texttt{1kg}$. (Th1) can be trivially proved. However, the vice versa does not hold. This makes explicit the intensional nature of $\sqsubseteq$. Furthermore, different qualities may have the same extensions, that is, $x \subseteq y \wedge y \subseteq x$ does not imply $x = y$. Finally, note that empty-qualities, i.e., qualities that classify no object, are included in, but not subsumed by, all the qualities.

Df7 $x \subseteq y \triangleq \text{PQT}(x) \wedge \forall zt(\texttt{CF}_{\text{QT}}(x,z,t) \rightarrow \texttt{CF}_{\text{QT}}(y,z,t))$

Th1 $x \sqsubseteq y \rightarrow x \subseteq y$

4 Image schemas

As outlined in [16], images schemas are very abstract spatial structures with qualitative (topological) characteristics, i.e., they are not precisely characterised in terms of (geometric) magnitude or shape. This makes image schemas "highly flexible preconceptual and primitive patterns" [16, p.217] that can be instantiated in different contexts. They are 'malleable' enough "to fit many similar, but different, situations that manifest a recurring underlying structure" [9, p.30]. Our idea is to capture these abstract structures by relying on the qualitative relations of connection ($\bowtie$) and order ($\prec$) defined in the quality domains, i.e., intuitively, an image schema classifies tuples of qualities satisfying structural constraints expressible via $\bowtie$ and $\prec$.

Technically, image schemas are represented by *higher-level qualities*, qualities of tuples of homogeneous qualities. As we will see, this approach generalises the one in [4] where 'patterns' of locations of objects along a quality domain are represented as higher-level properties (sets of tuples in *product spaces*).

Tuples, represented by TU, are introduced into the domain of quantification. They are disjoint from objects, as well as from qualities and types. Tuple variables and tuple constants are noted with $\bar{u}, \bar{v}$, etc. Tuples are formally characterised by following the strategy usually exploited to reify relationships in FOL [6]. To cope with a first-order formalisation, tuples are assumed to have a maximal length ℓ. We thus introduce ℓ primitive binary relations $\multimap_{\mathtt{i}}$, where $x \multimap_{\mathtt{i}} \bar{u}$ refers to the ith element x of the tuple $\bar{u}$, see (Ax12) and (Ax13). Axioms (Ax14) and (Ax15) assure that tuples have at least two elements, and when tuples have the ith element, they also have all the previous elements. (Ax16) establishes the identity criterion for tuples, i.e., two tuples are identical if they have the same elements in the same order. It is easy to prove the unicity of the tuple $\bar{u}$ such that $x_1 \ldots x_n \multimap \bar{u}$ (with $n \leq \ell$), see (Df8), that is noted $\langle x_1, \ldots, x_n \rangle$. As expected, the same element can appear in different positions of the same tuple, e.g., $\langle \mathtt{a}, \mathtt{b} \rangle$, $\langle \mathtt{a}, \mathtt{b}, \mathtt{a}, \mathtt{a} \rangle$ or $\langle \mathtt{a}, \mathtt{a} \rangle$ are all valid and different tuples. (Df9) establishes when two tuples have the same length, while $\text{TU}_{\neq}$ identifies the tuples with all different elements (Df10). The subsumption relation between tuples is defined in (Df11).

Df8 $x_1 \ldots x_n \multimap \bar{u} \triangleq \bigwedge_{\mathtt{i}=1}^{n}(x_{\mathtt{i}} \multimap_{\mathtt{i}} \bar{u}) \wedge \bigwedge_{\mathtt{i}=n+1}^{\ell} \neg\exists x(x \multimap_{\mathtt{i}} \bar{u})$

Df9 $\bar{u} \equiv_{\text{L}} \bar{v} \triangleq \text{TU}(\bar{u}) \wedge \bigwedge_{\mathtt{i}=1}^{\ell}(\exists x(x \multimap_{\mathtt{i}} \bar{u}) \leftrightarrow \exists y(y \multimap_{\mathtt{i}} \bar{v}))$

Df10 $\text{TU}_{\neq}(\bar{u}) \triangleq \text{TU}(\bar{u}) \wedge \bigwedge_{\mathtt{i} \neq \mathtt{j}=1}^{\ell} \forall xy(x \multimap_{\mathtt{i}} \bar{u} \wedge y \multimap_{\mathtt{j}} \bar{u} \rightarrow x \neq y)$

Df11 $\bar{u} \sqsubseteq_{\text{TU}} \bar{v} \triangleq \bar{u} \equiv_{\text{L}} \bar{v} \wedge \bigwedge_{\mathtt{i}=1}^{\ell} \forall xy(x \multimap_{\mathtt{i}} \bar{u} \wedge y \multimap_{\mathtt{i}} \bar{v} \rightarrow x \sqsubseteq y)$

Ax12 $\bigwedge_{\mathtt{i}=1}^{\ell} \forall x\bar{u}(x \multimap_{\mathtt{i}} \bar{u} \rightarrow \text{TU}(\bar{u}))$

Ax13 $\bigwedge_{\mathtt{i}=1}^{\ell} \forall xy\bar{u}(x \multimap_{\mathtt{i}} \bar{u} \wedge y \multimap_{\mathtt{i}} \bar{u} \rightarrow x = y)$

Ax14 $\text{TU}(\bar{u}) \rightarrow \bigvee_{\mathtt{i} \neq \mathtt{j}=1}^{\ell} \exists xy(x \multimap_{\mathtt{i}} \bar{u} \wedge y \multimap_{\mathtt{j}} \bar{u})$

Ax15 $\text{TU}(\bar{u}) \rightarrow \bigwedge_{\mathtt{i}=2}^{\ell}(\exists x(x \multimap_{\mathtt{i}} \bar{u}) \rightarrow \exists y(y \multimap_{\mathtt{i}-1} \bar{u}))$

Ax16 $\bar{u} = \bar{v} \leftrightarrow \text{TU}(\bar{u}) \wedge \bigwedge_{\mathtt{i}=1}^{\ell} \forall x(x \multimap_{\mathtt{i}} \bar{u} \leftrightarrow x \multimap_{\mathtt{i}} \bar{v})$

For our purposes, it is important to individuate tuples of qualities belonging to the same domain, i.e., $[\text{D}_i]$TU-tuples (see (Df13) where $\multimap$ is defined in (Df12)), e.g., [TM]TU-, [SP]TU-, or $[\text{D}_{\text{color}}]$TU-tuples. DTU abstracts from the specific quality domain (Df14). Hedblom and colleagues [7] define *paths* as "collection[s] of two or more sites, which are connected by successor relationships" [7, p.27]. DTU-tuples can be seen as paths where sites are qualities (locations of objects) in a domain and the successor relation is captured by the order of the elements in the tuple. For this reason, we will use the terms path and DTU-tuple interchangeably.

Df12 $x \multimap \bar{u} \triangleq \bigvee_{\mathtt{i}=1}^{\ell} x \multimap_{\mathtt{i}} \bar{u}$

Df13 $[\text{D}_i]\text{TU}(\bar{u}) \triangleq \text{TU}(\bar{u}) \wedge \forall x(x \multimap \bar{u} \rightarrow \text{D}_i(x))$

Df14 $\text{DTU}(\bar{u}) \triangleq \bigvee_{i=1}^{\eta} [\text{D}_i]\text{TU}(\bar{u})$

Image schemas, represented by IM, are disjoint from all the other kinds of entities, QT included, i.e., QT collects only the qualities that directly classify objects. Image schema variables (constants) are noted with ι, σ, etc. (using small caps). We assume (tuples of) qualities to be static, i.e., the direct classification $\texttt{dCF}_{\text{IM}}$ between image schemas and tuples of qualities does not need to be spatio-temporally qualified. In addition, we assume image schemas to classify only homogeneous qualities (Ax17).

Ax17 $\texttt{dCF}_{\text{IM}}(\sigma, \bar{u}) \rightarrow \text{IM}(\sigma) \wedge \text{DTU}(\bar{u})$

Intuitively, image schemas capture general structural constraints. This idea can be formalised by means of constraints like (Ax18)-(Ax20), which characterise, respectively, the image schema 3MC of triples of mutually connected qualities, the image schema PWC of pairwise connected qualities, and the image schema INC of increasing qualities. These constraints do not refer to specific quality domains, they involve only shared structural relations. Image schemas are thus cross-domains. For example, $\langle \texttt{1kg}, \texttt{2kg}, \texttt{4kg} \rangle$ and $\langle \texttt{1m}, \texttt{7m} \rangle$ are both instances, with different length, of INC.

Ax18 $\texttt{dCF}_{\text{IM}}(\text{3MC}, \bar{u}) \leftrightarrow \exists xyz(xyz \multimap \bar{u} \wedge x \bowtie y \wedge y \bowtie z \wedge z \bowtie x)$

Ax19 $\texttt{dCF}_{\text{IM}}(\text{PWC}, \bar{u}) \leftrightarrow \bigwedge_{\mathtt{i}=1}^{\ell-1} \forall xy(x \multimap_{\mathtt{i}} \bar{u} \wedge y \multimap_{\mathtt{i}+1} \bar{u} \rightarrow x \bowtie y)$

Ax20 $\texttt{dCF}_{\text{IM}}(\text{INC}, \bar{u}) \leftrightarrow \bigwedge_{\mathtt{i}=1}^{\ell-1} \forall xy(x \multimap_{\mathtt{i}} \bar{u} \wedge y \multimap_{\mathtt{i}+1} \bar{u} \rightarrow x \prec y)$

Note that the form of these constraints—for instance, the fact that they involve only the $\prec$ and $\bowtie$ primitives—can be regulated only at the meta-level. The structural nature of image schemas is thus only poorly captured by our FOL framework. A second problem concerns the intensional nature of image schemas. Gärdenfors builds patterns in a purely mathematical manner as subsets of product spaces. Accordingly, Gärdenfors' patterns have a purely *extensional* nature. Vice versa one could think, for instance, that the 3-long patterns *being equidistant* and *forming the same angles* defined on triples of (punctual) spaces are co-extensional but different. In principle, our framework is compatible with this intensional stance. However, the formulas that characterise the *being equidistant* and *forming the same angles* patterns are (in Euclidean geometry) logically equivalent. One should then assume the *way* patterns are characterised through axioms to impact the identity of patterns. The proper characterisation of the intension of image schemas is left for future work.

Our image schemas are similar to predicates that apply to tuples of qualities. In this sense they are close to the patterns of Gärdenfors [4]. There are however some important differences that show that our image schemas are more abstract and more flexible than patterns. First, in [4], all the instances of a pattern are tuples of a product

space obtained from a given domain. For instance, following [4], `lighter` and `shorter` are two different patterns. Vice versa, the image schema INC is flexible enough to apply both to weights and lengths. Sect. 5 shows how the relations *being lighter than* and *being shorter than* are obtainable by applying the image schema INC to weights and lengths, respectively. Second, according to [4], all tuples instantiating a pattern have the same length. We have already shown that this does not hold for image schemas, which may be indeed *cyclic* (see, e.g., [7]) or—using the terminology of Galton [3]—*open*, i.e., without a pre-established length.

5 Patterns

We introduce hereby *patterns* that result from the instantiation of image schemas in given contexts. Intuitively, a context localises an image schema in the sense that it constrains the qualities the schema can apply to. We represent contexts by DTU-tuples and patterns (noted with PT) by couples $\langle \sigma, \bar{c} \rangle$ where $\text{IM}(\sigma)$ and $\text{DTU}(\bar{c})$, i.e., they are image schemas together with contexts of application. Pattern variables and constants are noted $\hat{p}, \hat{q}$, etc. The context $\bar{c}$ constrains the schema σ in the sense that it filters out all the tuples classified by σ that are not subsumed by $\bar{c}$, see (Df15).[16] Furthermore, $\bar{c}$ sets the length of the pattern, i.e., by (Df11), all the instances of $\langle \sigma, \bar{c} \rangle$ have the length of $\bar{c}$. Patterns have then a fixed lenght.

Df15 $\mathtt{dCF}_{\text{PT}}(\langle \sigma, \bar{c} \rangle, \bar{u}) \triangleq \text{PT}(\langle \sigma, \bar{c} \rangle) \wedge \mathtt{dCF}_{\text{IM}}(\sigma, \bar{u}) \wedge \bar{u} \sqsubseteq_{\text{TU}} \bar{c}$

Following Gärdenfors, our patterns can be used to classify tuples of *objects*, i.e., to represent *internal relations* among objects (Df16).[17] For instance, $\langle \text{INC}, \langle \mathtt{weight}, \mathtt{weight} \rangle \rangle$ represents *being lighter than*, while $\langle \text{INC}, \langle \mathtt{lenght}, \mathtt{lenght} \rangle \rangle$ represents *being shorter than*.

Df16 $\mathtt{dCF}(\langle \sigma, \bar{c} \rangle, \bar{x}, t) \triangleq \text{PT}(\langle \sigma, \bar{c} \rangle) \wedge \bar{c} \equiv_{\text{L}} \bar{x} \wedge \exists \bar{u} (\mathtt{dCF}_{\text{PT}}(\langle \sigma, \bar{c} \rangle, \bar{u}) \wedge \bigwedge_{\mathtt{i}=1}^{\ell} \forall u x (u \multimap_{\mathtt{i}} \bar{u} \wedge x \multimap_{\mathtt{i}} \bar{x} \rightarrow \mathtt{CF}_{\text{QT}}(u, x, t))$

The subsumption relation between patterns is defined in (Df17). Note that (Df17) requires the identity of image schemas, i.e., only instances of the same schema are comparable. For instance, by requiring tuples to have specific color-qualities, $\langle \text{PWC}, \langle \mathtt{red}, \mathtt{blue}, \mathtt{red} \rangle \rangle$ is subsumed by $\langle \text{PWC}, \langle \mathtt{color}, \mathtt{color}, \mathtt{color} \rangle \rangle$.

Df17 $\langle \sigma_1, \bar{c}_1 \rangle \sqsubseteq_{\text{PT}} \langle \sigma_2, \bar{c}_2 \rangle \triangleq \sigma_1 = \sigma_2 \wedge \bar{c}_1 \sqsubseteq_{\text{TU}} \bar{c}_2$

To sum up, a pattern $\langle \sigma, \bar{c} \rangle$ captures the same structure of an image schema σ but it is less abstract than σ because (i) it has a fixed-length (set by $\bar{c}$); and (ii) it focuses on the (DTU-)tuples subsumed by $\bar{c}$.

6 Structural properties

With the technical and conceptual machinery previously introduced, we now approach the representation of *structural properties* of objects, i.e., properties that take into account the *way* a pattern of qualities is exhibited by an object. Here we consider temporal and spatial ways (aka modalities) of exhibiting patterns. Roughly, our idea is that an object has a structural quality when it exhibits a pattern of qualities temporally or spatially arranged according to a second pattern. Structural properties, noted with SPR, are represented by couples of patterns $\langle \hat{q}, \hat{a} \rangle$, where $\hat{a}$ constrains the (temporal or spatial) arrangement of the qualities constrained by $\hat{q}$. We assume the context of $\hat{a}$ to have the same length of the context of $\hat{q}$ and to be (i) a $[\text{TM}]\text{TU}$-tuple or (ii) a $[\text{SP}]\text{TU}$-tuple. Structural properties with a temporal arrangement (case (i)) are called *historical properties* (see [15]), while the ones with a spatial arrangement (case (ii)) are called *distributional properties* (see [18]). (Df18) defines subsumption between structural properties.

Df18 $\langle \hat{q}_1, \hat{a}_1 \rangle \sqsubseteq_{\text{SPR}} \langle \hat{q}_2, \hat{a}_2 \rangle \triangleq \text{SPR}(\langle \hat{q}_1, \hat{a}_1 \rangle) \wedge \text{SPR}(\langle \hat{q}_2, \hat{a}_2 \rangle) \wedge \hat{q}_1 \sqsubseteq_{\text{PT}} \hat{q}_2 \wedge \hat{a}_1 \sqsubseteq_{\text{PT}} \hat{a}_2$

The classification of an object under an historical property is defined in (Df19): at t, the object x is classified by $\langle \hat{q}, \hat{a} \rangle$ (where $\hat{a}$ is a temporal arrangement) when, *before* t, following the *time* $\hat{a}$-path $\bar{a}$, the object x *holistically* follows the *quality* $\hat{q}$-path $\bar{q}$.[18] According to (Df19), *historical* properties hold in virtue of the temporal distribution of the qualities the classified object had in its past history.[19]

Df19 $\mathtt{tCF}(\langle \hat{q}, \hat{a} \rangle, x, t) \triangleq \text{SPR}(\langle \hat{q}, \hat{a} \rangle) \wedge [\text{TM}]\text{PT}(\hat{a}) \wedge \exists \bar{q} \bar{a} (\mathtt{dCF}_{\text{PT}}(\hat{q}, \bar{q}) \wedge \mathtt{dCF}_{\text{PT}}(\hat{a}, \bar{a}) \wedge \text{TU}_{\neq}(\bar{a}) \wedge \bigwedge_{\mathtt{i}=1}^{\ell} \forall q a (q \multimap_{\mathtt{i}} \bar{q} \wedge a \multimap_{\mathtt{i}} \bar{a} \rightarrow (\mathtt{CF}_{\text{QT}}(q, x, a) \wedge a \prec t)))$

The classification of an object under a distributional property is defined in (Df20): at t, the object x is classified under $\langle \hat{q}, \hat{a} \rangle$ (where $\hat{a}$ is a spatial arrangement) when, following the *space* $\hat{a}$-path $\bar{a}$, the object x *locally* follows the *quality* $\hat{q}$-path $\bar{q}$. According to (Df20), distributional properties hold in virtue of the spatial distribution of the qualities of the classified object.

Df20 $\mathtt{sCF}(\langle \hat{q}, \hat{a} \rangle, x, t) \triangleq \text{SPR}(\langle \hat{q}, \hat{a} \rangle) \wedge [\text{SP}]\text{PT}(\hat{a}) \wedge \exists \bar{q} \bar{a} (\mathtt{dCF}_{\text{PT}}(\hat{q}, \bar{q}) \wedge \mathtt{dCF}_{\text{PT}}(\hat{s}, \bar{a}) \wedge \text{TU}_{\neq}(\bar{a}) \wedge \bigwedge_{\mathtt{i}=1}^{\ell} \forall q a (q \multimap_{\mathtt{i}} \bar{q} \wedge a \multimap_{\mathtt{i}} \bar{a} \rightarrow \mathtt{CF}_{\text{QT}}(q, x, a, t)))$

Intuitively, the same object can be classified under a structural property by following different (cognitive) procedures, i.e., by following different paths. For instance, consider the distributional property $\langle \langle \text{PWC}, \langle \mathtt{color}, \mathtt{color} \rangle \rangle, \langle \text{PWC}, \langle \mathtt{space}, \mathtt{space} \rangle \rangle \rangle$, i.e., *having two connected colors at connected spaces*. Consider $\langle \mathtt{red}, \mathtt{orange} \rangle$, $\langle \mathtt{orange}, \mathtt{red} \rangle$, $\langle \mathtt{s}_1, \mathtt{s}_2 \rangle$ and $\langle \mathtt{s}_2, \mathtt{s}_1 \rangle$ (where these colors and these spaces are assumed to be connected). If an object is (at t) `red` at $\mathtt{s}_1$ and `orange` at $\mathtt{s}_2$ then, according to (Df20), we can follow $\langle \mathtt{red}, \mathtt{orange} \rangle$ according to $\langle \mathtt{s}_1, \mathtt{s}_2 \rangle$, or $\langle \mathtt{orange}, \mathtt{red} \rangle$ according to $\langle \mathtt{s}_2, \mathtt{s}_1 \rangle$. Thus, `sCF` concerns only how the object is at given spaces, not how it is (cognitively) explored. However, assume now the object is (at t) also `yellow` at space $\mathtt{s}_3$ (with $\mathtt{s}_2 \bowtie \mathtt{s}_3$). This object is classified by the above distributional property by following the paths $\langle \mathtt{red}, \mathtt{orange} \rangle$ according to $\langle \mathtt{s}_1, \mathtt{s}_2 \rangle$ or, alternatively, by following the paths $\langle \mathtt{orange}, \mathtt{yellow} \rangle$ according to $\langle \mathtt{s}_2, \mathtt{s}_3 \rangle$, i.e., the object exhibits the same structural property at different spaces. To avoid that, one could assure that the spatial location of the object (or its visible surface) is wholly covered by the spatial path. Analogous considerations hold for `tCF`.

7 Types

Gärdenfors's concepts are compounds of properties (belonging to different domains), i.e., they are regions in the multi-dimensional space composed by the domains. We extend this notion of concept (and conceptual space) by characterising our types in terms of both

[16] Note that $\mathtt{dCF}_{\text{PT}}(\langle \sigma, \bar{c} \rangle, \bar{c})$ does not always hold.

[17] An internal relation R is a relation such that the truth-value of $\mathtt{R}(a, b)$ depends exclusively on similarity judgments along quality domains concerning the objects a and b only. See [13] for more details. Note that, our image schemas, through contexts and patterns, can apply to objects. This seems to contrast with the position in [11, 7], where images schemas, in particular the PATH schema, only apply to events or activities.

[18] Where $[\text{D}_i]\text{PT}(\langle \sigma, \bar{c} \rangle) \triangleq [\text{D}_i]\text{TU}(\bar{c})$.

[19] Clearly these properties are relevant also to characterize events (even though in this case the classification is not temporally qualified).

qualities (QT-instances) and structural properties (SPR-instances). The properties in the union of QT and SPR are called *features*. Notice that, even though distributional properties depend on how a given quality pattern is spatially arranged, they are still holistic, i.e., they apply to the whole objects. At the end of the section, we add (i) *structured* properties (to be distinct from *structural* ones), i.e., following Armstrong, properties that hold in virtue of the way the classified object is mereologically structured into components of given types; and (ii) *relational* properties, i.e., properties that hold in virtue of the way the classified object connects with the environment.

Characterisation—$\mathtt{CH}(x, y)$ reads as "the type x is *characterised* by the feature (quality or structural property) y"—represents the link between types and features. Types have a multi-dimensional nature, they are characterised in terms of at least two, but usually several, domains (Ax21), where $[\mathrm{D}_i]\mathrm{SPR}(\langle \hat{q}, \hat{a}\rangle) \triangleq [\mathrm{D}_i]\mathrm{PT}(\hat{q})$.[20]

Ax21 $\mathrm{TY}(x) \rightarrow \exists yz(\mathtt{CH}(x, y) \wedge \mathtt{CH}(x, z) \wedge \bigvee_{i=1}^{\eta}((\mathrm{D}_i(y) \vee [\mathrm{D}_i]\mathrm{SPR}(y)) \wedge \neg\mathrm{D}_i(z) \wedge \neg[\mathrm{D}_i]\mathrm{SPR}(z)))$

The features of types are not ordered, types just *cluster* features (belonging to different domains). Indeed, these features are not necessarily atomic (see (Df2)), e.g., the Fiat500 type can be characterised by the red colour even though the colour-domain could contain more specific shades of red, like crimson, scarlet, etc.

Following the classical theory of types, one can assume that the features of a type express necessary and sufficient conditions to be classified by the type. In this classical perspective, the classification of an object by a type reduces to its classification by *all* its features. Structural qualities slightly complicate the scenario because depending on the kind of arrangement (temporal vs. spatial) one needs to consider $\mathtt{tCF}$ or $\mathtt{sCF}$. The definition of the classification of an object under a type ($\mathtt{CF}_{\mathrm{TY}}$) is given in (Df21).

Df21 $\mathtt{CF}_{\mathrm{TY}}(x, y, t) \triangleq \mathrm{TY}(x) \wedge \forall z(\mathtt{CH}(x, z) \rightarrow ((\mathrm{QT}(z) \rightarrow \mathtt{CF}_{\mathrm{QT}}(z, y, t)) \vee ([\mathrm{TM}]\mathrm{SPR}(z) \rightarrow \mathtt{tCF}(z, y, t)) \vee ([\mathrm{SP}]\mathrm{SPR}(z) \rightarrow \mathtt{sCF}(z, y, t))))$

There exist several possibilities to weaken the classical view on classification under types. For instance, one may introduce two characterisation-like relations to grasp the distinction between *necessary* and *optional* features of types, or introduce a 'weight' for each feature (as done for concepts in conceptual spaces). One may also distinguish the set of features sufficient to be classified by a type from the set of features sufficient to *not* be classified by a type.[21] Classification could also rely on the metrics of the domains (and on the metric of the overall space), i.e., one could see our notion of type as an extension of the one of *prototype* (see [24]) and capture classification (categorisation) on the basis of the distance between the qualities of the object and the features of the type.

Types can be organised by means of subsumption $\sqsubseteq_{\mathrm{TY}}$ (Df22). Given the definition of $\sqsubseteq_{\mathrm{SPR}}$ (Df18), $\sqsubseteq_{\mathrm{TY}}$ is grounded in the subsumption between qualities, i.e., following Gärdenfors, the way types are organised mainly depends on the organisation of qualities. Note that a type characterised by a (structural) quality can subsume only types characterised by (structural) qualities (in the same domain). It is trivial to prove that $\mathtt{CF}_{\mathrm{TY}}$ is closed under $\sqsubseteq_{\mathrm{TY}}$ (Th2), i.e., indirect classification under types is encapsulated into $\mathtt{CF}_{\mathrm{TY}}$.

Df22 $x \sqsubseteq_{\mathrm{TY}} y \triangleq \mathrm{TY}(y) \wedge \forall z(\mathtt{CH}(y, z) \rightarrow \exists w(\mathtt{CH}(x, w) \wedge (w \sqsubseteq z \vee w \sqsubseteq_{\mathrm{SPR}} z))$

Th2 $\mathtt{CF}_{\mathrm{TY}}(x, y, t) \wedge x \sqsubseteq_{\mathrm{TY}} z \rightarrow \mathtt{CF}_{\mathrm{TY}}(z, y, t)$

As observed by Fiorini and colleagues [21], object recognition and categorisation are grounded not only in holistic properties but also in the identification of parts of the objects and the way these parts are structured. The number, the types, and the arrangement of parts are all essential aspects to found structural similarities between objects, an idea very close to the one of structured properties of Armstrong.

We start to sketch the formalisation of structured properties by extending the notion of pattern's context: patterns have the form $\langle \sigma, \bar{c}\rangle$ where $\bar{c}$ is now a DTU-tuple *or a tuple of types*. *Structured properties*, noted with TPR, have the same form of structural properties, i.e., $\langle \hat{p}, \hat{a}\rangle$, but now $\hat{p}$ is a pattern with a type-context.[22] We then consider a temporary parthood relation defined on objects: $\mathtt{PART}(x, y, t)$ stands for "the object x is part of the object y at time t" (see [12] for a FOL axiomatisation). Notice that, by excluding the possibility to have spatially co-localised objects, PART can be reduced to spatial inclusion.[23] The *local* classification by the type p of the object x as it is at space s and time t is the classification under p of the part of x that, at t, is exactly located at s (Df23).[24]

Df23 $\mathtt{CF}_{\mathrm{TY}}(p, x, s, t) \triangleq \exists y(\mathtt{PART}(y, x, t) \wedge \mathrm{L}_{\mathrm{SP}}(y, s, t) \wedge \mathtt{CF}_{\mathrm{TY}}(p, y, t))$

The classification of an object under a structured property is defined in (Df24): at t, the object x is classified under $\langle \hat{q}, \hat{a}\rangle$ (where $\hat{a}$ is a spatial arrangement and $\hat{q}$ a type-pattern) when, following the *space* $\hat{a}$-path $\bar{a}$, the object x *mereologically* follows the *type* $\hat{q}$-path $\bar{q}$, i.e., its components selected by the spaces in $\bar{a}$ are instances of the correspondent types in $\bar{q}$.[25]

Df24 $\mathtt{mCF}(\langle \hat{q}, \hat{a}\rangle, x, t) \triangleq \mathrm{TPR}(\langle \hat{q}, \hat{a}\rangle) \wedge [\mathrm{SP}]\mathrm{PT}(\hat{a}) \wedge \exists \bar{q}\bar{a}(\mathtt{dCF}_{\mathrm{PT}}(\hat{q}, \bar{q}) \wedge \mathtt{dCF}_{\mathrm{PT}}(\hat{a}, \bar{a}) \wedge \mathrm{TU}_{\neq}(\bar{a}) \wedge \bigwedge_{\mathtt{i}=1}^{\ell} \forall qa(q \multimap_{\mathtt{i}} \bar{q} \wedge a \multimap_{\mathtt{i}} \bar{a} \rightarrow \mathtt{CF}_{\mathrm{TY}}(q, x, a, t)))$

Our approach differs from the one of Fiorini and colleagues [21, 22] for two main reasons. First, in [21, 22] objects are always though as completely specified, i.e., as already noticed, an object is represented by a point in a conceptual space. By extending conceptual spaces with structural spaces an object comes already with all its parts (of given types) and the position of these parts with respect to the whole. Vice versa our framework can represent *partial information*, including mereological one, about objects. Second, and more importantly, [21, 22] focus on the representation of structural information in the framework of conceptual spaces where the similar-

[20] (Ax21) excludes the possibility to characterise a type in terms of both a D_i-quality and a $[\mathrm{D}_i]$SPR-property. This could be criticized. For instance, one could assume that an object with a given colour pattern could also have an holistic colour, i.e., the (conventional) colour of the whole object emerges from the colour pattern. This situation could be represented by providing a link between colour distributional properties and colour qualities, i.e., the classification under a colour quality would be inferred by the one under the corresponding strctural property. This is another interesting point that we do not address here.

[21] This approach can be represented by means of *rough sets* [19].

[22] Whether and how $\prec$ or $\bowtie$ can be defined on types is not taken into account in this paper. The only image schema that applies to tuples of types is the one that allows to built the paths.

[23] E.g., $\mathtt{PART}(x, y, t) \triangleq \exists sr(\mathrm{L}_{\mathrm{SP}}(x, s, t) \wedge \mathrm{L}_{\mathrm{SP}}(y, r, t) \wedge s \sqsubseteq r)$.

[24] $\mathtt{CF}_{\mathrm{TY}}(p, x, s, t)$ may also be introduced without reference to parts. The case of types characterised only in terms of qualities is a straightfoward generalisation of (Df20). The case of types characterised in terms of structural (and structured) properties requires to change (Df20) to assure the spaces considered in the tuple $\bar{a}$ to be included in s.

[25] Similarly to what noticed about (Df20) at the end of Sect. 6, (Df24) does not guarantee x to be an assembly of only the components selected by $\bar{a}$, i.e., x may have components not considered by the structured property $\langle \hat{q}, \hat{a}\rangle$.

ity relation is central. The construction mechanism behind structural spaces is purely mathematical, i.e., structural spaces are the mathematical product of the spaces that represent the parts and the ones that represent the whole-centered position of the parts. Our approach is more explicative, since it makes explicit the dependence of structural and structured properties from the instantiation of images schemas, namely, it offers a basis to explain structural similarities in terms of image schemas.

A *relational* property of an object x can be seen as a structured property of an object (a system) y that has x among its components, i.e., it represents the way x is linked to the other components of y. The *relational* classification of objects under structured properties is defined in (Df25).[26]

Df25 $\mathtt{rCF}(\langle \hat{q}, \hat{a} \rangle, x, t) \triangleq \mathrm{TPR}(\langle \hat{q}, \hat{a} \rangle) \wedge [\mathrm{SP}]\mathrm{PT}(\hat{a}) \wedge$
$\exists y r \bar{q} \bar{a} (\mathtt{dCF}_{\mathrm{PT}}(\hat{q},\bar{q}) \wedge \mathtt{dCF}_{\mathrm{PT}}(\hat{a},\bar{a}) \wedge \mathrm{TU}_{\neq}(\bar{a}) \wedge$
$\mathrm{L}_{\mathrm{SP}}(x,r,t) \wedge \mathtt{PART}(x,y,t) \wedge r \multimap \bar{a} \wedge$
$\bigwedge_{\mathtt{i}=1}^{\ell} \forall q a (q \multimap_{\mathtt{i}} \bar{q} \wedge a \multimap_{\mathtt{i}} \bar{a} \rightarrow \mathtt{CF}_{\mathrm{TY}}(q,y,a,t)))$

(Df21) and (Df22) can then be extended to include structured and relational properties among the features that characterise types. These properties are fundamental to model assembled artefacts.

8 Conclusions

We presented a formal approach for the representation of compound types by modifying and extending Gärdenfors' theory of conceptual spaces. This is done by considering the notion of image schema and by generalising Gärdenfors' approach to patterns. In our framework, image schemas represent abstract (topological) structures, whereas patterns apply image schemas to specific quality-domains. Patterns are then used to specify either the historical properties of objects, namely constraints over their evolution in time, or objects' distributional or structured properties.

REFERENCES

[1] James F. Allen and Patrick J. Hayes, 'A common-sense theory of time', in *International Joint Conference on Artificial Intelligence (IJCAI85)*, pp. 528–538, Los Angeles, (1985).

[2] Maureen Donnelly, 'Relative places', *Applied Ontology*, **1**(1), 55–75, (2005).

[3] Antony Galton, 'Processes as patterns of occurrence', in *Process, Action, and Experience*, ed., R. Stout, Oxford Univeristy Press, (to appear).

[4] Peter Gärdenfors, *Conceptual spaces: The geometry of thought*, MIT press, 2000.

[5] Christopher Gauker, 'A critique of the similarity space theory of concepts', *Mind & Language*, **22**(4), 317–345, (2007).

[6] Terry Halpin and Tony Morgan, *Information Modeling and Relational Databases*, Morgan Kaufmann, second edition edn., 2008.

[7] Maria M Hedblom, Oliver Kutz, and Fabian Neuhaus, 'Image schemas as families of theories', in *Proceedings of the 4th International Workshop on Computational Creativity, Concept Invention, and General Intelligence (C3GI 2015)*, eds., TR Besold, K.-U. Kühnberger, M Schorlemmer, and Smaill A., pp. 19–33, (2015).

[8] Ingvar Johansson, 'Pattern as ontological category', in *Formal Ontology in Information Systems*, ed., N. Guarino, 86–94, IOS Press, Amsterdam, (1998).

[9] Mark Johnson, *The body in the mind: The bodily basis of meaning, imagination, and reason*, University of Chicago Press, 1987.

[10] Werner Kuhn, 'An image-schematic account of spatial categories', in *Proceeding of the International Conference on Spatial Information Theory (COSIT 2007)*, eds., S. Winter and B. Kuipers, volume LNCS 4736, pp. 152–168. Springer, (2007).

[11] Ronald W. Langacker, *Cognitive grammar: A basic introduction*, Oxford University Press, 2008.

[12] Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, and Alessandro Oltramari, 'Wonderweb deliverable d18', Technical report, CNR, (2003).

[13] Claudio Masolo and Daniele Porello, 'A note on the representation of relations in conceptual spaces', in *Atti della XII Conferenza Annuale della Associazione Italiana di Scienze Cognitive*, ed., M. Cruciani, volume 9, pp. 117–121, (2015).

[14] Claudio Masolo and Daniele Porello, 'Understanding predication in conceptual spaces', in *Formal Ontology in Information Systems: Proceedings of the 9th International Conference (FOIS 2016)*, eds., R. Ferrario and W. Kuhn, pp. 139–152. IOS Press, (2016).

[15] Claudio Masolo, Laure Vieu, Yoshinobu Kitamura, Kouji Kozaki, and Riichiro Mizoguchi, 'The counting problem in the light of role kinds', in *Proceedings of the AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, eds., E. Davis, P. Doherty, and E. Erdem. AAAI Press, (2011).

[16] Todd Oakley, 'Image schemas', in *The Oxford handbook of cognitive linguistics*, eds., D. Geeraerts and H. Cuyckens, chapter 9, 214–235, Oxford University Press, (2007).

[17] Antoni Olivé, *Conceptual modeling of information systems*, Springer Science & Business Media, 2007.

[18] Josh Parsons, 'Distributional properties', *Lewisian themes*, 173–180, (2004).

[19] Zdzisław Pawlak, *Rough sets: Theoretical aspects of reasoning about data*, volume 9, Springer Science & Business Media, 1991.

[20] Diana Raffman, 'Similarity spaces', in *The Oxford Handbook of Philosophy of Perception*, ed., M. Matthen, 679–693, Oxford Univeristy Press, (2015).

[21] Sandro Rama Fiorini and Mara Abel, 'Part-whole relations as products of metric spaces', in *Proceedings of the IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI 2013)*, pp. 55–62, (2013).

[22] Sandro Rama Fiorini, Peter Gärdenfors, and Mara Abel, 'Representing part-whole relations in conceptual spaces', *Cognitive Processing*, **15**(2), 127–142, (2014).

[23] David A. Randell and Antony G. Cohn, 'A spatial logic based on regions and connections', in *International Conference on Principles of Knowledge representation and Reasoning (KR 92)*, eds., B. Nebel, C. Rich, and W. Swartout, pp. 165–76. Morgan Kaufmann, (1992).

[24] Eleanor Rosch, 'Principles of categorization', in *Cognition and categorization*, eds., E. Rosch and B. B. Lloyd, 27–48, Lawrence Erlbaum, (1978).

[26] (Df25) must be modified in case spatially coincident objects are admitted.

Using the enaction paradigm as a basis for database exploration to favor users sensemaking

Landy Rajaonarivo, Pierre De Loor, Eric Maisel, Matthieu Courgeon [1]

Abstract.
This paper presents an interface between a user and a database inspired by the enactive field of cognitive science. This interface aims to favor the knowledge understanding for the user during his exploration of data. It proposes data progressively by estimating, in real time, the user centers of interest. Thus, a co-evolution between the user representations of the links between data and the navigation environment appears. To address the embodiment side of the cognition promoted by the enactive field, the user is immersed in a virtual museum that grows progressively, in real time. The data are embodied in 3D objects endowed with an interactive behavior. The structure of the data such as ontologies are also embodied in interactive graphical objects. Thanks to the co-evolution between the user and the environment, the resulting museum is specific for each user. The paper presents some arguments to raise this approach and provides some explanations about the links between the behavior of the user relative to the embodied data and the realtime evaluation of their respective relevances.

Keywords: data exploration, intelligent user interface, enaction, co-evolution, ontologies, embodiment, metaphor, sensemaking.

1 Positioning

In cognitive science, one paradigm that addresses the notion of human sensemaking is enaction which claims that cognition relies on the history of the interactions between an individual and his environment [5, 11]. Living beings modify their environment which, in return, impacts on the possibility of actions for them. After a while, this co-construction of the *agent* and his environment implies a strong interdependence between them. This coupling is inseparable to the explanation of the cognition and of the representations of individuals. The coupling lets emerge a specific understanding, individual to each person. Concept like affordances [2] is in line with these principles because it characterizes the influence of the environment in humans' actions. Enaction is also inspired by some phenomenologists, which considers that Humans find a stable representation of their own world from sensorimotor invariant [6, 7]. Eventually, enaction is also close to constructivism and embodied cognition because it highlights the role of the action in cognition and then, in sensemaking.

Our positioning is to take enaction as a basis of inspiration to develop awareness of users that should understand a huge amount of data. The challenge here, is to propose a user interface which favors the emergence of understanding. Currently, research engines only propose a list of suggestion to users according to some keywords but rarely address the embodied part of the cognition nor the real time behavior of the user relative to the propositions of the research engine. In our approach, we embody data in 3D objects which are introduced in a virtual environment. This last is used by the user to favor his embodiment relative to the data. We use also an on-line self-organization of the graphical representation of the data. Our claim is that it is more important to let the user actively discover links between information, making choice among these links, than giving to him a list of propositions without some cues.
In the literature, some works claim their links with enaction. For example, Kaipainen et al propose in [4] an enactive system allowing to study the co-influence between the human experience and the technologies. This system is based on the psycho-physiology of facial expressions and avatars that mirror them. Another enactive system is proposed in [8] which is implemented to design the bodily interaction with a virtual agent.

Our study can be considered as less embodied (even if it is one of our objective to immerse the user) than these previous works but more focused on the coevolution between humans and systems and moreover, more focused on the human sensemaking through his interaction and the subsequent evolution of his environment.

Indeed, our work aims to mix adaptive methods of information visualization and self-organization of data while introducing the user in the interactive loop. We also define an interactive intelligent system designed to guide the user to construct a representation of a problem. By this way, we transform the idea of artificial intelligence based of predefined knowledge (classical artificial intelligence) toward an evolving, not predefined and human coupled system that we design as an enaction-based artificial intelligence system [1]. This paper is structured as following, section 2 presents the theoretical principles of our proposition. This proposition implies a realtime evaluation of the user's center of interest according to some concepts. Section 3 is focused on the formalization of the concepts through ontologies as well as on the formalization of the evaluation of user's center of interest. Section 4 provides an illustration of the results obtained with this proposition.

2 Principles

We propose an interface for database exploration favoring the co-construction of the user sensemaking and his environment that we call *"enactive interface"* in reference to the inspiration that this research field gives to us for this study. The user is considered as the *agent* of the theoretical stance of enaction and his environment is a virtual *living* museum. It represents the database as well as the se-

[1] ENIB/CERV Lab-STICC, Brest, France, email: {rajaonarivo, deloor, maisel, courgeon}@enib.fr

mantical relations between the data. Our virtual museum is considered as 'living' because it evolves in realtime according to the user's behavior.

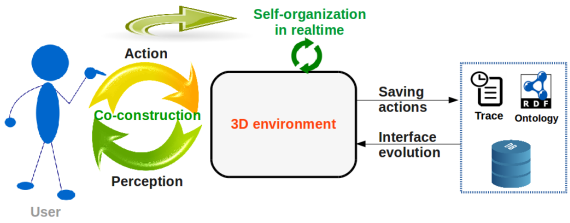

Figure 1. Enactive interface for database exploration

The figure 1 shows the functional principles of our proposition: the user interacts with a virtual environment and all his interactions are recorded. These are used to estimate the centers of interest of the user. The museum evolves and grows according to the estimation of the users interests. By *evolving*, we mean that the objects presented to the user are dynamically organized and by *growing* we mean that the topics of the next rooms are also defined dynamically. The evaluation of the user is provided from his actions in the virtual museum. For that, the user can use different interactive tools that are linked with some semantic concepts of the data (topic, space and time). This semantic concepts are defined by our research partners in cultural heritage and linguistics domains and formalized by three ontologies. For example, cultural topics are in an ontology that contains concepts like society, economy, mentality, natural cultural heritage, industrial cultural heritage, etc (dozens of concepts). The figure 2 illustrates the structure of these abstract concepts at a macro level and the figure 3 details the micro-level of one of the abstract concept *"Patrimoine industriel et artisanal"* (which means in English Industrial and artisanal heritage).

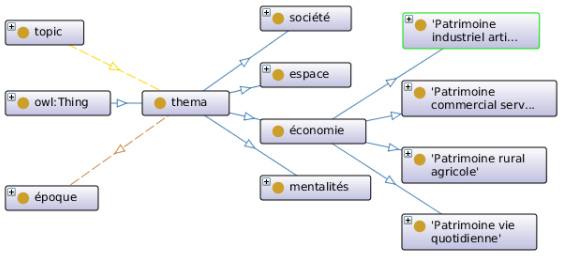

Figure 2. Structure of the concepts at macro level

The topos ontology can be described in the same manner. It defines places like towns, regions, geographical coordinates. Chronos ontology is divided in dates, years, century, ages, etc.. The cultural heritage objects of the database are connected to the concepts located at the leaf of the ontology.

Interactive tools are adapted to the meaning of these ontologies. For example, a map tool allows for the designation of places that represent the ontology structure (see Figure 4 on the bottom right). The use of this interactive tools as well as the behavior of the user

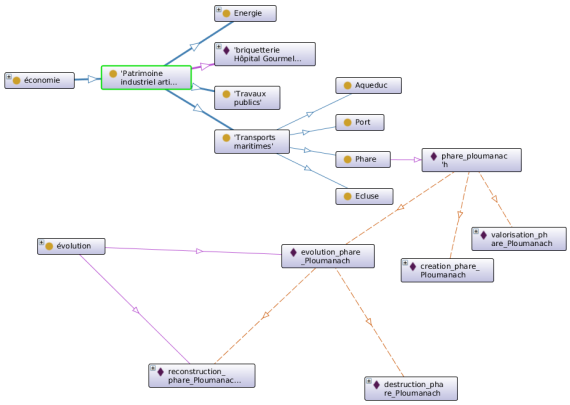

Figure 3. Structure of the concepts at micro level

can be interpreted in terms of concepts that interest him. Then, as the data are tagged with concepts, and as these concepts are linked together through ontologies, it is possible to choose data semantically close to the estimated user's center of interest. Moreover, as the concept are hierarchically organized, it is possible to choose more or less abstract concepts for the topics of the next rooms. Thanks to virtual environment, the user doesn't manipulate directly the concepts nor the ontologies. For example, if he stays a certain amount of time in front of an object, the system deduces that the concepts associated to this object are important. Some interactive tools are more explicitly linked with the ontology: for example, each object are linked with a place in the world (and are tagged with this place which is in the ontology). By this way, concepts that are easy to be represented *naturally* (as place in the world) and are explicitly represented with the tools while concepts that are more abstract, are just deduced from the behavior of the user. In short, the environment influences the user by suggesting objects that are linked semantically but not explicitly. It is their apparition during the visit and the distance that separate them, that express a semantical proximity. In the same time, the understanding of the user as well as its corresponding behavior evolves. It lets appear the co-evolutive loop between the agent and its environment.

3 Formalization

These section presents how the system estimates the center of interest of the user and how it will update the environment. For that, we specified a list of actions that the system can infer from the user's behavior his visit. The actions which are considered are the following:

- Enter in a room
- Display the information about an object
- Stay in front of an object
- Bypass an object
- Use a navigation tool
- Take an object

As each element of the museum is connected to at least one concept in the ontologies, we attribute a value to each concept that we call relevance and this last will be updated through the actions of the user. All user actions have a degree of interest, so, the updating of the relevance of one concept is different from an action to another. We suppose that:

- We have n types of possible actions and each type $type_i$ has a weight w_i which indicates the degree of interest of the user according to his actions. For example, when the user stays in front of an object, his center of interest on this object is lower compared to his center of interest when he takes this object with him. The more w is high, the more the user is supposed interested.
- An action a is defined by $a = \{type_a, \Gamma_a, t_a\}$ where Γ_a:a set of concepts involved in the action a, $type_a$: the type of the action a having a weight w_a, t_a: the time that the action a has been realized.

We consider now that the user do an action a of type $type_a$ at the moment t_a. So,

$\forall c \in \Gamma_a, \tau_c(t) = \tau_c(t - dt) + [w_a * (\tau_{max} - \tau_c(t - dt))] * dt$
where τ_c is the relevance of the concept c and τ_{max} the maximal relevance of all concepts.

As our objective is to favor the data discovery for helping understanding on a domain but not the research for data, we propose to the user not only the relevant concepts but also the concepts semantically close from the relevant concepts. For that, we introduce the measure of similarity distance between two concepts in a same ontology through the measure of distance proposed in [12].

It is based on the fact that concepts are hierarchically organized in trees of nodes. Each nodes represent one concept of the ontology. Thus, a concept has an upstream concept and some potential downstream node and there is a unique root concept which has no upstream concept.

The conceptual similarity between c_1 and c_2 is dependent on the least common superconcept of c_1 and c_2 designed as c_3 as well as on the root concept of the ontology :

$ConSim(c_1, c_2) = \frac{2*N_3}{N_1+N_2+2*N_3}$ where, if we consider that each concept is a node in the hierarchical tree of the ontology :

N_1 is the number of nodes on the path from c_1 to c_3. N_2 is the number of nodes on the path from c_2 to c_3 . N_3 is the number of nodes on the path from c_3 to $root$.

4 Example

We show through the following example how and when our system evolves according to the interactions of the user during his exploration.

Figure 4. Data exploration system using virtual museum as navigation metaphor

Figure 4 Shows the virtual museum. The data are embodied in photo frames and the interactive tools are accessible through the three icons in the top left.

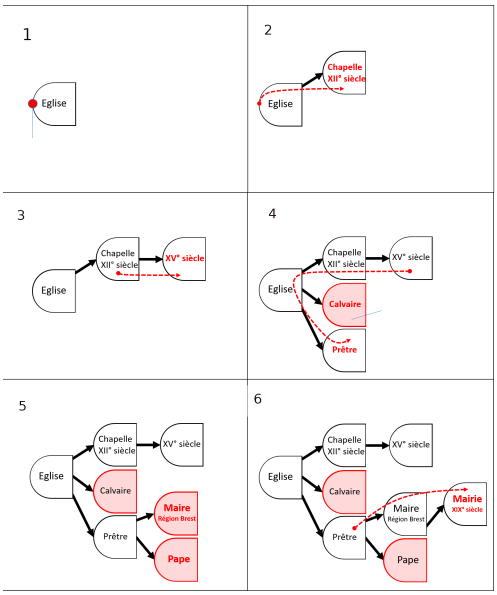

Figure 5. example of user's exploration path with dynamic creation of the museum's rooms: the red color rooms are the rooms which are not visited by the user. Legends: église: church, chapelle: chapel, prêtre: clergyman, calvaire: calvary, mairie: town hall, maire: mayor

Figure 5 illustrates the growing of the virtual museum during the user's navigation.

Each room is tagged by one or two concepts providing from the ontologies. Each object of each room is linked to concepts that are considered close to the room concepts. Moreover, the distance between the object is relative to the semantical distance defined in section 3. This grouping system is supposed to help the user to explore data that interest him. In the beginning of the interaction, the museum has just one room and a number of data is limited by the size of the room. For the example in the Figure 5, the system presents data related to the topic:*Eglise* (which means Church in English). We enumerate below the process of the dynamical creation of the other rooms and object selection :

1. The system proposes to the user a data extract about a given topic.
2. Referring with the user's interaction with the data proposed, the system creates another room close to the user. The topic of this new room is more or less close semantically to the room's topic where the user is. The topic is displayed on top of the room's door and there is a sound of door opening. It lets the system to influence the user to explore another data which may interest him (affordances). For this simulation, the user decided to enter in this new room (the course of the user is in red dotted lines).
3. At this step, the system creates a third room, the user entered in this room, but he was not interested by the data inside this room, so he went back to the second created room where he was.
4. The user continued his path in order to return to his initial position.
5. The system takes into account the displacement of the user in the environment and the time during which he has stayed in each room. After that, it creates new rooms with topics estimated close

to user's interest. For this step, the user ignores the room with a topic *Calvaire* but he visited the room with the topic *Prêtre*. When the user entered in one room, the system makes him aware of the impact of his interactions on his exploration by creating new rooms.

6. This last step shows that the user token his last path to return back.

This example shows that our system draws attention of the user during his exploration by proposing new information, and by indicating paths of navigation. It takes into account the user's interaction to propose the new information. So, the system could always propose new information even if the user did nothing.

5 Next stages

In the future, we plan to exploit the notion of affordances to improve the understanding knowledge for the user. For example, in the actual version, when new rooms are defined, a door opens and a sound that draw user's attention is emitted. We imagine also to improve the relationship between the user and the environment by inspiring of the models presented in [3]. In the future, we will work on the shapes of these doors and of other graphical representation. We will also work on the position of the objects relative to the user (below, above, small, big ...). Then, we plan to test immersive technologies like VR headset to address the effect of the embodiment on human understanding during data exploration. For this, we want to evaluate if the fact that the user walks physically into the data improve his ability to remember some information or links between these information. Finally, we also work on the mean to favor serendipity by the use of the structure of the ontology in parallel with some exploration policies to favor novelty during the exploration. The reader who needs more technical information about this study can refer to [10, 9].

REFERENCES

[1] Pierre De Loor, Kristen Manac'h, and Jacques Tisseau, 'Enaction-Based Artificial Intelligence: Toward Co-evolution with Humans in the Loop', *Minds and Machines*, **19**(3), 319–343, (oct 2009).

[2] James J. Gibson, 'The theory of affordances', *Perceiving, Acting, and Knowing*, (1977).

[3] Fahim T. Imam and Thomas R. Dean, 'Affordances in representing the behaviour of event-based systems', in *JOWO@FOIS*, (2016).

[4] Mauri Kaipainen, Niklas Ravaja, Pia Tikka, Rasmus Vuori, Roberto Pugliese, Marco Rapino, and Tapio Takala, 'Enactive systems and enactive media: embodied human-machine coupling beyond interfaces', *Leonardo*, **44**(5), 433–438, (2011).

[5] H R Maturana and F J Varela, *Autopoiesis and Cognition: The realization of the living*, Reidel: Boston, 1980.

[6] M Merleau-Ponty, *Phénoménologie de la Perception*, Collection " Tel ", 1990, éditions g edn., 1945.

[7] A Noë, *Action in Perception*, Cambridge, MA: MIT Press, 2004.

[8] Roberto Pugliese and Klaus Lehtonen, 'A framework for motion based bodily enaction with virtual characters', in *International Workshop on Intelligent Virtual Agents*, pp. 162–168. Springer, (2011).

[9] Landy Rajaonarivo, Eric Maisel, Matthieu Courgeon, and Pierre De Loor, 'Inline co-evolution between users and information presentation for cultural heritage data exploration', *22nd annual meeting of the intelligent user interfaces, IUI*, (2017).

[10] Landy Rajaonarivo, Eric Maisel, and Pierre De Loor, 'An enactive based realtime 3d self-organization system for the exploration of a cultural heritage data base', *20th International Conference Information Visualisation*, (2016).

[11] F J Varela, E Thompson, and E Rosch, *The embodied mind: Cognitive science and human experience*, The MIT Press, 1992.

[12] Zhibiao Wu and Martha Palmer, 'Verbs semantics and lexical selection', in *Proceedings of the 32Nd Annual Meeting on Association for Computational Linguistics*, ACL '94, pp. 133–138, Stroudsburg, PA, USA, (1994). Association for Computational Linguistics.

Ontologies as Backbone of Cognitive Systems Engineering

Ricardo Sanz[1] **and Julita Bermejo**[1] **, Juan Morago**[1] and **Carlos Hernández**[2]

Abstract. Cognitive systems are starting to be deployed as appliances across the technological landscape of modern societies. The increasing availability of high performance computing platforms has opened an opportunity for statistics-based cognitive systems that perform quite as humans in certain tasks that resisted the symbolic methods of classic artificial intelligence. Cognitive artefacts appear every day in the media, raising a wave of mild fear concerning artificial intelligence and its impact on society. These systems, performance notwithstanding, are quite brittle and their reduced dependability limips their potential for massive deployment in mission-critical applications —e.g. in autonomous driving or medical diagnosis. In this paper we explore the actual possibility of building cognitive systems using engineering-grade methods that can assure the satisfaction of strict requirements for their operation. The final conclusion will be that, besides the potential improvement provided by a rigorous engineering process, we are still in need of a solid theory —possibly the main outcome of cognitive science— that could sustain such endeavour. In this sense, we propose the use of formal ontologies as backbones of cognitive systems engineering processes and workflows.

Keywords: Cognitive systems; trust; dependability; engineering processes; autonomy; life-cycle; ontology.

1 INTRODUCTION

These days we are seeing in the media a continuous flow of reports about self-driving cars, mobile phone natural language assistants or machines that win at games traditionally considered reserved to humans (e.g. Go or Texas hold 'em poker). Artificial intelligence (AI) seems to be re-flourishing and this is raising a global awareness of its potential and a global concern of its risks. The recently created *Partnership for AI*[3] has been "Established to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society."

The current flourishing of AI is characterised by an availability of high performance computing platforms that has opened an new opportunity for statistics-based cognitive systems. These systems can perform quite as humans in certain tasks that resisted approaches based on the symbolic methods of classic artificial intelligence and the formal representation approaches of cognitive science. More or less cognitive artefacts are appearing every day in the media, raising a wave of mild fear concerning artificial intelligence and its potential impact in society —both positive and negative. However, from a real-world systems usability perspective —the perspective of systems engineers— these systems are still quite brittle and their reduced dependability is seriously limiting their potential for massive deployment in mission-critical applications. A clear example are autonomous cars, where this intrinsic brittleness when deployed in real roads is limiting their widespread deployment.

In this paper we will analyse the issue of the *engineering of dependable cognitive systems*. A transition from pre-engineering to true-engineering methods is in need. The construction of a cognitive system (CS) in the past has always been a form of craftsmanship and not an instantiation of proven, repeatable engineering processes. Most CSs have not gone beyond the research phase. In a sense, all CSs that we have seen are one-of-a-kind systems. The customer-oriented repeatability of engineering methods has not been manifested at all in the domain of CS construction. Even the reuse of "commercial" cognitive technologies as IBM's Watson is still a kind of hacking. This being said, the achievement of a systematic use of engineering-grade methods for CS implementation is very important, because it could enable the strict fulfilment of user-centric requirements for their operation. This is a need for real-world deployable systems, from the dual perspective of their *capability* and their *dependability*.

This is a position paper in which we analyse the role that ontologies can play in this transition. Anticipating the final conclusion of our analysis we believe that, to achieve the potential improvement provided by using a rigorous engineering process to build CSs, we are in *need of a deeper, unified cognitive science* —a solid theory of mental processes— that could sustain such endeavour. Engineering methods based on this science of the mind will lead to the synthesis of the two classes of engineering assets that are necessary for CS engineering (CSEng): *design patterns* —structural/behavioural aspects for cognitive architectures— and *ontologies* —concepts to bind i) the minds of the engineers and system stakeholders; ii) the mind of the engineer to the CS under construction; and iii) the mind of the CS to its world and the world of its user.

The paper is structured as follows: first we analyse the concepts of *system*, *mission* and *cognition* in the domain of CSs; then we provide an overview of the engineering task of building CSs; after this, we offer a brief analysis of fundamental architectural aspects of these systems; a section on the question of dependability, resilience and trust follows; next a discussion of the need of CS-specific engineering methods; finally we include the core analysis of the need of a solid theory of cognition —based on patterns and ontologies— to support the engineering life-cycle of CSs. The paper ends with some conclusions and future perspectives.

[1] Universidad Politécnica de Madrid, Autonomous Systems Laboratory, Spain, email: ricardo.sanz@upm.es
[2] TU Delft Robotics Institute, Delft University of Technology, The Netherlands c.h.corbato@tudelft.nl
[3] https://www.partnershiponai.org/

2 ENGINEERING COGNITIVE SYSTEMS

The issue that we stress in this paper is the need of having systematic engineering methods to build custom cognitive systems and the critical role that ontologies will play in achieving this.

2.1 Systems and Missions

AI is artificial not because it is not natural, but because, as described by Simon [63], things are *built to satisfy a purpose*. Real-word systems[4] should be useful to people; their users. Users expect that the systems will be able to fulfil their missions as planned.

When we ask Siri to find a nearby Spanish restaurant we expect i) to get an answer; ii) that the restaurant suggested is close to our position; and iii) that it is indeed a Spanish restaurant and not a Mexican one or a New-yorkean tapas bar. Good engineering is always concerned about building *systems* that reliably fulfil their *missions* [39].

The use of AI techniques in any kind of systems help these systems deal with complex problems and situations. We call these systems *cognitive* because they "know", *i.e.* hey exploit some form of knowledge[5] in the performance of their missions[6]. AI systems have been able to progress in problem solving steadily [46], overcoming many of the predicted limits [24]. Most of the envisioned limits of AI are related to aspects of the human mind that are considered particularly peculiar like creativity and consciousness. These and other, simpler aspects of human minds are the essential focus of Cognitive Science (CogSci) and are explored in the heterogeneous ways and domains that characterise CogSci [71].

Computer implementations of CSs have always been focused on essentially two parallel but different purposes: i) the evaluation by computer simulation of cognitive theories of the biological mind (esp. human); and ii) the construction of artefacts able to intelligently deal with their complex worlds. In some cases these two threads are mixed —e.g. when implementing cognitive anthropoids— leading to confusion about the purposes and the degree of success of the development of the CS.

Precise statement of system goals and operational requirements is a strong principle (almost a dogma) of systems engineering and is badly needed in the context of CS construction; otherwise projects get lost into disparaged explorations sometimes concerned about the mission of the system, sometimes about properly mimicking human features, and sometimes about the exploration of the elusive landscapes of cognitive capabilities.

In the domain of Cognitive Systems Engineering (CSEng)[7] it is necessary to focus on the mission that the CS shall perform, trying to avoid vagaries into its human resemblances. On the other hand, mimicking humans is a perfectly valid class of mission. So, we can indeed undertake the mission of modelling the human mind using engineering-grade methods, but this requieres a precise statement of this goal and a method of verification, *i.e.* his needs *an objective test of human-likeness*. However, this specification of an *objective* Turing test is however still a dream.

In essence, CSEng seeks to properly use artificial intelligence at the service of improving mission-level capability. Cognitive systems are able to see, perceive objects and affordances, making good decisions and acting properly in the world in the service of a pre-specified mission —all this driven by knowledge available to the agent or gained through the senses. Besides having intrinsic capability —e.g. the capability of following a path— cognitive systems must also be dependable. Trust and usability rely both on dependability.

Robustness —the capability of tolerating disruption— and resilience —the capability of functional recovery— [35] are hence critical aspects for real-world systems and one the main negative aspects of research-grade CSs.

2.2 Engineering Life-cycles for Cognitive Systems

The expression *Cognitive Systems Engineering (CSE)*[8] has been used by other authors to refer to the engineering of systems that include humans performing cognitive tasks. In essence it is used to refer to a special discipline of *human factors* in systems development that addresses the design of human-cognitive socio-technical systems (a system in which humans are the providers of essential capability related to perceiving, evaluating, deciding, planning and executing [36, 55, 72]. CSE is in essence the system analysis and design effort necessary to support the cognitive requirements of human work inside socio-technical systems [41, 48].

In this paper, however, we will use the term *Cognitive Systems Engineering* to refer to the wider discipline of building systems that are cognitive (e.g. including humans, AIs or both) in support of the *mission itself*. Inside this large domain, we are specifically interested in artificial systems that are cognitive by themselves, not by being in relation with humans who are the thinkers. Obviously, the characterisation of what is cognitive and what is not deserves special attention [1], and, while it is not the central aspect of this paper, it will be necessarily addressed later in the section on the need of a deeper cognitive science (see Section 3).

Cognition is a general capability (*i.e.* ot exclusively human or biological) that can be provided by machines, hence the inclusion of humans is not a necessary condition for a system to be cognitive. Nor it is necessary for a system to behave like a human to be cognitive. In our work at the UPM Autonomous Systems Laboratory, we are specifically interested in CSs that do not depend on cognitive capabilities of other agents —esp. humans— to fulfil their missions. We are interested in autonomous CSs. Cognition in artefacts can span a wide range: from minimal cognition [8] to human-like performance in specific tasks [66].

This wide spectrum of cognition apparently implies that there is no possibility of having a single, established engineering process and cognitive system life-cycle. This process and life-cycle would simplify the construction process and provide improvements concerning the predictable fulfilment of user requirements. A systematic engineering methodology for CSEng is needed [14] and it shall be based on the right architectural patterns and reusable assets [19].

Figure 1 summarily depicts a part of the engineering processes and system assets that are performed and appear along the life-cycle of a cognitive system. It shows the two/three major phases in a CS life-cycle: i) engineering —design and construction— and ii) operation.

[4] We use the term *real-world* to refer to systems that are deployed to serve a specific purpose, in contrast to research-oriented systems.

[5] While knowledge and belief are obviously different things, for practical reasons in this paper we do not make a strong distinction between them.

[6] Note that this conception of *cognitive* somehow departs for a widespread understanding of the term that equates it with *having some resemblance of human mental processes*. This conception of cognition is excessively anthropomorphic and lessens the possibility of advancing in a general cognitive science.

[7] The term *Cognitive Systems Engineering* has been used by other authors [36, 55, 72] to refer to systems that include humans doing cognitive work —see next Section— but we use this term here to refer to the engineering of (maybe artificial) cognitive systems like those mentioned before —e.g. Siri or the Tesla autonomous car.

[8] Note that this is different from CSEng described before.

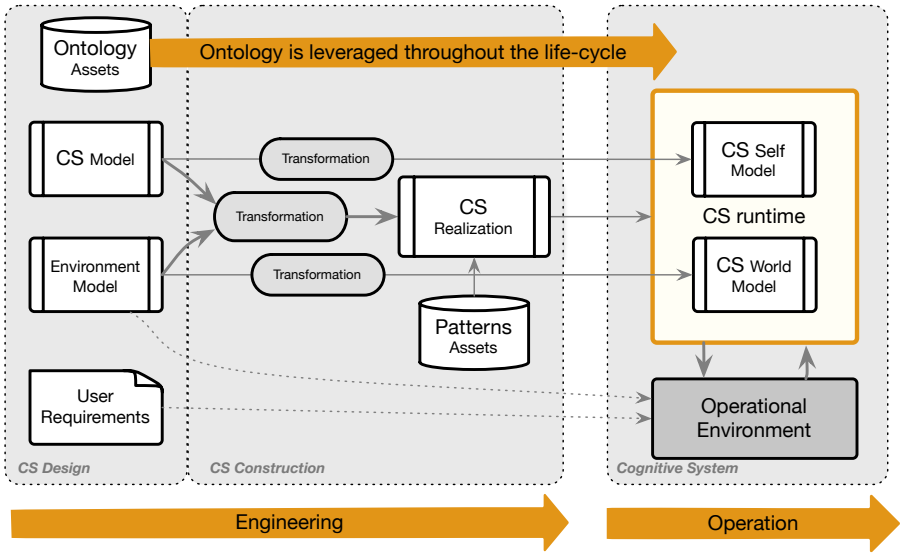

Figure 1. *The cognitive system engineering life-cycle. Systems concepts support the whole process, providing a framework for analysis, design, construction and operation of targeted cognitive systems. Ontologies capture the concepts in the mind of engineers and their formalisations can be used to synthesise parts of the architecture of the target cognitive system including the representations used for its mental content: perceptions and reflections of the CS's world.*

Resilient CSs use runtime models of their operational environment and models of themselves to sustain their activity. The CS is built from assets realised by transformation of the CS engineering models and reusable assets in the form of ontologies and patterns. In this paper we focus on the role of ontologies. Details will be given later (in Section 4).

2.3 Cognitive Systems Designs

The literature on cognitive systems is full of examples of architectures for cognition [43]. Some of them are focused on very specific tasks while others are postulated as almost universal [42, 4]. Some of these architectures come with associated engineering methods that can be applied to build custom systems using the reusable assets that the architecture provides [2]. However, from a perspective of general CSEng, these architectures come with excessive a priori commitments that limit the decisions that a cognitive system architect may take [62].

There is a need of lowering the sophistication and completeness of descriptions of generic cognitive systems, going back from full creatures [12, 22] to focused functional patterns [3, 15, 49] that can provide specific capabilities (e.g. recognition, learning, metacognition, etc.). From the CSEng perspective, there is a need of moving back from the search of *unified theories of cognition* proposed by Newell [50] into a search of *unifiable patterns of cognition*.

This means that cognitive systems designs shall be described at two levels: i) at the *whole system level* in the form of reference architectures and ii) at the *functional subsystem level* in the form of specific cognitive patterns [17, 18].

Reference architectures and cognitive patterns are the necessary design assets for the realisation of the engineering process suggested in the previous section. Note that the patterns are what provide the specific capabilities while the architectures provide the framework for their integration [60]. Note also that autonomous cognitive systems require both direct capabilities —e.g. face recognition— and meta capabilities —e.g. learning— to provide the required levels of autonomy [16]. Note also that patterns for general resilience —necessary to build up trust— are mostly based on reflection and cognitive self-penetrability of the agent itself [54].

3 A DEEPER COGNITIVE SCIENCE?

The previous analysis implies that there is a need of clarifying "cognition" from the perspective of elementary capabilities. The *unifiable patterns of cognition* mentioned earlier should be general systems' generalisations of the micro-theories that psychology has been building about cognitive phenomena in humans [51]. Stork [67] claims for a scientific foundation for engineering cognitive systems but, for the reasons explained before, this scientific foundation shall also encompass humans —and animals. We need a deeper cognitive science that can ground the rigorous engineering of cognitive systems.

Identification of common concepts between humans and machines is critical to provide a grounded integration of humans in cognitive socio-technical systems [53]. The conventional approach of providing conceptual mappings through the human-machine interface (HMI) between extremely different human/machine ontologies does not provide the constructibility, flexibility, and dynamic allocation of tasks that is needed for mixed autonomy resilient systems [23].

The real possibility of a scientific, unified cognitive science is still a theme of debate [70, 25]. There is a common thought that "cognitive science" was a different discipline to different groups of scholars, and that even fundamental aspects would remain open forever. For example, Peterson [52] bewares of unification risks, pointing at the possibility of empty relabelling of mental phenomena done by "representational cognitivism" and the practical neglect of different viewpoints (like enactivism or embodiment). In the same vein, Husbands says that "More than a decade ago, in his fine exploration of the then burgeoning cognitive science movement [30], Gardner asked if neuroscience would devour the entire field. His answer was no, for the same reasons that he could see no prospects for a single unified discipline: there will always be separate subject areas, we will always need diverse multi-level descriptions and explanations." Dale

[20] warns us that "cognitive scientists should take seriously the possibility that a single, unified framework for all of cognition is an unrealistic expectation for its diverse interdisciplinary goals and subject matter".

However, there is also a generalised feeling of an intrinsic, underlying reality of cognition that shall be captured by a deep cognitive science [59]. For example, Husbands [37] says that "solutions to the problem of how brains produce adaptive behaviour would necessarily entail (at least implicit) specifications for building artificial brains [...] being potentially useful practical devices". This points to the cognitive patterns mentioned in the previous section.

Cognitive science shall be a science of cognition and not only a science of human cognition. Obviously, human minds are specially capable in some contexts and hence their preeminence in cognitive science. However non-human minds[9] are also very relevant for a universal science, non anthropocentric view [33, 40]. This is of special importance on the coming years where cognitive artefacts will progressively take roles reserved to humans. A solid theory of mind —esp. the artificial mind— is needed to fulfil requirements and to avoid emergent, disruptive phenomena [11, 44].

The possibility of a unified science of mind may seem remote, but note that the "computer" mind model seems distant to human minds due to its digital nature. The sentence "thinking is computation" may hence be seen as plainly false. It is necessary to stress the fact that "computation" is not necessarily "digital computation". A paradigmatic example of this case are analog computers. Note also that the brain is both discrete and continuous. In the words of Schadé and Smith [61]: "Thus one single neuron possesses the qualities of a hybrid computer."

There is a strong need of finding the right framework where to build the *definite theory of mind*[10]. The theoretical frameworks we are using today to formalise system structure and behaviour are quite disjoint. For example, when reasoning about temporal dynamics of system state, consider the difference between representing change by means of temporal logic [27] or systems dynamics [45]. The models that we use are sometimes seen as mere rough approximations [64].

While it may seem difficult to achieve a theoretical agreement on a *universal theory of cognition*, most present theories are not too far from each other. The cognitive science debates are grounded not on theoretical issues but on attachments to disciplinary practices and languages, excessive targeting of full-fledged human cognition or excessive narrowing on specific cognitive tasks and feats. For example, when analysed in terms of general systems theory, classical cognitivism, embodied cognitive science or enactivist cognition are not so far from each other [47]. In fact, if we eliminate representation, what remains is mindless body; it may be performant, but only in a singular task[11]. If there is no information, there is no adaptation. We must be able to separate *what minds are* from *what happens when minds are attached to a particular environment*.

4 ONTOLOGIES AS BACKBONE

The essential question is: What are the bricks to build the edifice of the deep cognitive science described in the previous section? What is the theoretical substrate that grounds the description of architectures and patterns needed for cognitive systems engineering? The answer is simple: concepts [5, 29]. Fundamental concepts about mental phenomena will provide the necessary assets to build the cognitive capability theories and the patterns and architectures that reify them. Ontologies —as conceptual structures— will be the backbone that will sustain the cognitive systems engineering processes of the future, because they will both support the human engineering activities and the operation of the cognitive systems themselves.

It must be noted that these ontologies shall not be restricted to mind internals, but must also capture the realities outside cognitive agents [**?**]. The CS *environment* shall be formally addressed in ontological work to enable both the engineering of single-CS activity and multi-CS cooperation. Note also that these ontologies and their use must be flexible enough as to be able to accommodate differences and inconsistencies between different CS (or even between the the ontologies themselves). Ontologies shall move to use less rigid logical frameworks (capable of addressing *belief*, *uncertainty* and emphinconsistency).

Within computer science and software engineering, an ontology defines a set of ontological elements as representational primitives that can be used to model a domain of knowledge [31], [32]. The underlying idea is that ontologies provide a common vocabulary with explicit semantics [65].

Ontologies would act as a foundation for system science, where they facilitate organising the domain knowledge and formalising the different engineering artefacts or entities (e.g stakeholder, requirement, function, scenario, etc.) and their relationships, to resolve ambiguities and check for consistency and completeness [10], [38]. As examples of this approach, ontologies have been used to: assist the systems engineering process when it comes to establishing requirements [28]; drive the specification of the structural, functional, behavioural knowledge for a domain–specific system design [6]; or establish a set of common and shared concepts identified and agreed upon for autonomous systems representation and engineering [7].

When ontologies are reified as mental assets for autonomous CSs they will play both roles of defining the meaning of terms and grounding the understanding of the world (*i.e.* he CS itself and its environment including other CSs). Ontologies would act as a representation-based mechanism expressible in a computational language, to describe the different entities participating in the design and operation of the system. Concepts will then become computational elements in knowledge models that would exist through their representations in software, such as UML diagrams [26] or ontology-related languages such as OWL [68]. Concepts in engineering minds and engineering models will also be reified as physical elements in the CS or in its environment. This accounts for the modelling relation that underlies life and cognition [56]. As a consequence, the knowledge models will no longer be characterised by the usual high level of arbitrariness or difficult reuse, as it happens when knowledge is formalised as vocabularies expressed in natural language —the usual practice in cognitive science. How the meanings of concepts defined in the ontologies are understood by different components of the cognitive system is as much the focus as how the different actors in the CS life-cycle can use a common conceptualisation of the problem under scrutiny, to come up with a possible solution.

The complementary element of this ontological vision is the use of design patterns for description and engineering if system capabilities, as they provide solutions to design problems that happen repeatedly [34]. Such design patterns will act as the cornerstones of reusing CS architectures [60]. These patterns could be either archi-

[9] In this context, "mind" shall not be understood as a full-fledged human psyche. A mind is an information-driven controller for a system. These appear in all the biological and artificial spectrum in different degrees of capability, consciousness and self-awareness [57].

[10] A single account for human, animal and artificial minds.

[11] Arguments about the possibility of non-representational minds (e.g. [13]) are easily deflated by simple pointing at the representations used in the proposed "representationless" architectures.

tectural or domain patterns [15]. *Architectural patterns* express the structural organisation of a cognitive system, *i.e.* hey realise its architecture. *Domain patterns* describe a mechanism to solve a concrete but recurring problem in a particular context. These point towards the cognitive micro-theories that populate the literature. Note that architectural and domain patterns are not usable in isolation; in a concrete CS both shall be merged to *put mind —architecture– to a task —domain*.

It is worth pointing out that these patterns will not be used independently but having a domain ontology acting as backbone (see Figure 1). The architectural patterns will describe the cognitive system internal organisation and dynamics, based on the interactions between the ontological elements that describe the system elements themselves. Domain patterns in turn will describe the interactions among the cognitive system components, and with the environment, by using the underlying conceptualisation specified by the ontology, that represent design solutions so that the behaviour of the cognitive system fulfils the engineering requirements. Thus all system patterns will not only be specified from the ontological concepts, but eventually will become part of the ontology itself, modelling the relations and interactions between them as designed by the engineers [9].

5 CONCLUSIONS

Dale, Dietrich, and Chemero [21] defend the idea that " the 'framework debate' in cognitive science is unresolvable. The idea that one theory or framework can singly account for the vast complexity and variety of cognitive processes seems unlikely if not impossible.".

In the context of cognitive systems engineering —with and without humans— this idea is untenable for two reasons, one theoretical, one practical:

- Science advances by unification [69]. This can be considered a theoretical dogma or just a leitmotif of science. In any case, it is inalienable in the opinion of these authors because it is needed to support engineering-grade CS construction (see Section 2).
- Engineering works by using commensurable designs. There are not —there cannot be— different theoretical frameworks underlying systems of "vast complexity and variety" no matter what level of complexity they have. This does not mean that there cannot be different realms, but all them shall integrate seamlessly under a common framework. For example, mechanical engineering and electrical engineering address different domains of knowledge and practice. However, both fall under the physical framework. This is what enables the construction of solid electromechanical systems (e.g. the concepts of energy, force and torque are shared in both realms; this is what enables electrical motors moving machines).

Science and engineering are necessarily bound together under a common framework. The differences that manifest in different systems or theories shall be sought for not in the theoretical framework but in the concrete instantiations of that framework on specific entities. For example, electrical systems theory is a *single theory* even when the electrical systems in the US and in Europe have very important differences. Having a common theoretical framework implies that differences can be overcome; electrical systems can be interconnected and interoperate as a system-of-systems able to fulfil a single mission. The same can be said about cognitive systems: differences shall be found in concrete realisations, not in the theoretical concepts that underlie them.

Intense work shall be done to identify, clarify and formalise the concepts and patterns that underlie all classes of cognitive behaviour beyond the anthropomorphic trap [58]. Their expression in the form of formal ontologies will then constitute core assets for the engineering of the trustable autonomous cognitive systems of the future. The fundamental concepts used by engineers to think about the cognitive system under construction and the fundamental concepts used by the cognitive system itself in its interaction with its world —including itself— will be the same and captured in a formal representation. This will open a new world of capabilities for cognitive systems that will be able to deeply adapt to changing worlds in pursue of their missions.

ACKNOWLEDGEMENTS

We acknowledge the support of the Spanish Government through grant RTC-2016-5191-8 Sistema Autónomo de VIgilancia y SEguridad basado en multirotores (ADVISE).

REFERENCES

[1] Frederick Adams and Rebecca Garrison, 'The mark of the cognitive', *Minds and Machines*, **23**, 339–352, (2013).
[2] James S. Albus and Anthony J. Barbera, 'RCS: A cognitive architecture for intelligent multi-agent systems', *Annual Reviews in Control*, **29**(1), 87–99, (2005).
[3] Christopher Alexander, Sara Ishikawa, and Murray Silverstein, *A Pattern Language: Towns, Buildings, Construction*, Oxford University Press, 1977.
[4] John L. Anderson, 'Autonomous systems intelligence', in *Proceedings of the 1983 Annual Conference on Computers : Extending the Human Resource*, ACM '83, pp. 229–233, New York, NY, USA, (1983). ACM.
[5] A. T. Bahill and B. Gissing, 'Re-evaluating systems engineering concepts using systems thinking', *IEEE Transaction on Systems, Man and Cybernetics, Part C: Applications and Reviews*, **28**(4), 516–527, (1998).
[6] Shreya Banerjee and Anirban Sarkar, 'Ontology–driven approach towards domain-specific system design', *International Journal of Metadata, Semantics and Ontologies*, **11**(1), 39–60, (2016).
[7] Behzad Bayat, Julita Bermejo-Alonso, Joel Carbonera, Tullio Facchinetti, Sandro Fiorini, Paulo Goncalves, Vitor A. M. Jorge, Maki Habib, Alaa Khamis, Kamilo Melo, Bao Nguyen, Joanna Isabelle Olszewska, Liam Paull, Edson Prestes, Veera Ragavan, Sajad Saeedi, Ricardo Sanz, Mae Seto, Bruce Spencer, Amirkhosro Vosughi, and Howard Li, 'Requirements for building an ontology for autonomous robots', *Industrial Robot: An International Journal*, **43**(5), 469–480, (2016/09/14 2016).
[8] Randall D. Beer and Paul L. Williams, 'Information processing and dynamics in minimally cognitive agents', *Cognitive Science*, **39**(1), 1–38, (2015).
[9] Julia Bermejo, Carlos Hernández, and Ricardo Sanz, 'Model-based engineering of autonomous systems using ontologies and metamodels', in *IEEE International Symposium on Systems Engineering 2016 (IEEE ISSE 2016)*, Edinburgh, Scotland, (2016).
[10] Ralph Bogusch, 'Ontology-based systems engineering - the smart way of realizing complex systems'. Keynote Lecture, 13 November 2015.
[11] Nick Bostrom, *Superintelligence: paths, dangers, strategies*, Oxford University Press, 2014.
[12] Rodney A. Brooks, 'The whole iguana', in *Robotics science*, ed., Michael Brady, 432–458, MIT Press, Cambridge, MA, USA, (1989).
[13] Rodney A. Brooks, 'Intelligence without representation', *Artificial Intelligence*, **47**(1-3), 139–159, (1991).
[14] Dennis M. Buede, *The Engineering Design of Systems*, Wiley, 2009.
[15] Frank Buschmann, Regine Meunier, Hans Rohnert, Peter Sommerlad, and Michael Stal, *Pattern Oriented Software Architecture. A System of Patterns*, John Wiley & Sons, Chichester, UK, 1996.
[16] Nicholas L. Cassimatis, Paul Bello, and Pat Langley, 'Ability, breadth, and parsimony in computational models of higher-order cognition', *Cognitive Science*, **32**, 1304–1322, (2008).
[17] Robert Cloutier, Gerrit Muller, Dinesh Verma, Roshanak Nilchiani, Eirik Hole, and Mary Bone, 'The concept of reference architectures', *Systems Engineering*, **13**(1), 14–27, (2010).

[18] Robert J. Cloutier and Dinesh Verma, 'Applying the concept of patterns to systems architecture', *Systems Engineering*, **10**(2), 138–154, (2007).
[19] Edward Crawley, Bruce Cameron, and Daniel Selva, *Systems architecture: Strategy and product development for complex systems*, Prentice Hall,, Hoboken, NJ, USA, 2015.
[20] Rick Dale, 'The possibility of a pluralist cognitive science', *Journal of Experimental & Theoretical Artificial Intelligence*, **20**(3), 155–179, (2013/04/30 2008).
[21] Rick Dale, Eric Dietrich, and Anthony Chemero, 'Explanatory pluralism in cognitive science', *Cognitive Science*, **33**(5), 739–742, (2009).
[22] Daniel Dennett, 'Why not the whole iguana?', *Behavioral and Brain Sciences*, **1**, 103–104, (1978).
[23] A. Dominguez-García, G. Hanuschak, S. Hall, and E. Crawley, 'A comparison of GN&C architectural approaches for robotic and human-rated spacecraft', in *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit*, pp. 20–23, (2007).
[24] Hubert L. Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason*, The MIT Press, 2 edn., 1992.
[25] Barbara Von Eckardt, *What is Cognitive Science?*, Bradford Books, MIT Press, 1995.
[26] James Rumbaugh et al., *The Unified Modeling Language Reference Manual*, Addison-Wesley, 1999.
[27] *Handbook of Temporal Reasoning in Artificial Intelligence*, eds., M. Fisher, D. Gabay, and L. Vila, Elsevier, 2005.
[28] Anabel Fraga, Juan Llorens, Luis Alonso, and José M. Fuentes, 'Ontology-assisted systems engineering process with focus in the requirements engineering process', in *Complex Systems Design & Management: Proceedings of the Fifth International Conference on Complex Systems Design & Management CSD&M 2014*, ed., Frédéric et al. Boulanger, 149–161, Springer International Publishing, (2015).
[29] Peter Gardenfors, *Conceptual Spaces. The Geometry of Thought*, MIT Press, Cambridge, MA, 2000.
[30] Howard Gardner, *The Mind's New Science. A History of the Cognitive Revolution*, Basic Books, 1985.
[31] T. Gruber, 'Ontology', in *Encyclopedia of Database Systems*, eds., L. Liu and M. Tamer Ozsu, 1963–1965, Springer US, (2009).
[32] Nicola Guarino, Daniel Oberle, and Steffen Staab, 'What is an ontology?', in *Handbook on Ontologies*, 1–17, Springer Berlin Heidelberg, (2009).
[33] Richard Harper, Alex Taylor, and Micheal Molloy, 'Intelligent artefacts at home in the 21st century', in *Material Agency. Towards a Non-Anthropocentric Approach*, eds., Carl Knappett and Lambros Malafouri, 97–120, Springer, (2008).
[34] Carlos Hernández, Julita Bermejo-Alonso, Ignacio López, and Ricardo Sanz, 'Three patterns for autonomous robot control architecting', in *The Fifth International Conference on Pervasive Patterns and Applications - PATTERNS 2013*, pp. 44–51, Valencia, (May 27 - June 1 2013).
[35] Kjell Jørgen Hole, *Anti-fragile ICT Systems*, SpringerOpen, 2016.
[36] Erik Hollnagel and David D. Woods, *Joint cognitive systems: foundations of cognitive systems engineering*, Taylor & Francis, Boca Raton, FL, 2005.
[37] Phil Husbands, 'Cognitive science: Real or imaginary?: Review of the MIT encyclopedia of the cognitive sciences edited by robert a. wilson and frank c. keil', *Artificial Intelligence*, **130**(2), 191–194, (8 2001).
[38] INCOSE, *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities, version 4.0*, John Wiley and Sons, Inc, Hoboken, NJ, USA, 2015.
[39] Scott Jackson, *Architecting Resilient Systems: Accident Avoidance and Survival and Recovery from Disruptions*, Wiley, Hoboken, New Jersey, USA, 2010.
[40] Owain Jones and Paul Cloke, 'Non-human agencies: Trees in place and time', in *Material Agency. Towards a Non-Anthropocentric Approach*, eds., Carl Knappett and Lambros Malafouri, 79–96, Springer, (2008).
[41] G. Klein, S. Deal, and S. Wiggins, 'Cognitive systems engineering: The hype and the hope', *Computer*, **41**(3), 95–97, (March 2008).
[42] John E. Laird, Allen Newell, and Paul S. Rosenbloom, 'Soar: an architecture for general intelligence', *Artificial Intelligence*, **33**(1), 1–64, (1987).
[43] Pat Langley, 'Cognitive architectures and general intelligent systems', *AI Magazine*, **27**(2), 33–44, (2006).
[44] Nancy Leveson, *Engineering a safer world: systems thinking applied to safety*, Engineering systems, The MIT Press, Cambridge, Mass., 2012.
[45] Lennart Ljung, *System Identification: Theory for the User*, Prentice Hall PTR, 2nd edn., December 1998.
[46] Stephen Lucci and Danny Kopec, *ArtificiAl intelligence in the 21st century: A Living Introduction*, Mercury Learning and Information, 2nd edition edn., 2016.
[47] Mohan Matthen, 'Debunking enactivism. debunking enactivism: a critical notice of hutto and myin's radicalizing enactivism', *Canadian Journal of Philosophy*, **44**(1), (2014).
[48] Laura G. Militello, Cynthia O. Dominguez, Gavan Lintern, and Gary Klein, 'The role of cognitive systems engineering in the systems engineering design process', *Systems Engineering*, **13**(3), 261–273, (2010).
[49] Michael S. P. Miller, 'Building minds with patterns'. ISBN-10: 0-692-54140-1, 2016.
[50] Allen Newell, *Unified Theories of Cognition*, Harvard University Press, 1990.
[51] Allen Newell and Herbert A. Simon, *Human Problem Solving*, Prentice-Hall, Englewood Cliffs, 1972.
[52] Donald M. Peterson, 'The mind's new labels?: Review of R.A. Wilson and F.C. Keil (eds.), the MIT Encyclopedia of the Cognitive Sciences', *Artificial Intelligence*, **130**(2), 213–216, (8 2001).
[53] *Human-System Integration in the System Development Process: A New Look*, eds., Richard W. Pew and Anne S. Mavor, The National Academies Press, 2007.
[54] Zenon Pylyshyn, 'Is vision continuous with cognition? the case for cognitive impenetrability of visual perception', *Behavioral and Brain Sciences*, **22**, 341–364, (1999).
[55] Jens Rasmussen, Annelise Mark Pejtersen, and L.P. Goldstein, *Cognitive Systems Engineering*, Systems Engineering, Wiley-Interscience, 1994.
[56] Robert Rosen, *Anticipatory Systems. Philosophical, Mathematical, and Methodological Foundations*, volume 1 of *IFSR International Series on Systems Science and Engineering*, Springer, 2nd edn., 2012.
[57] Ricardo Sanz, 'Machines among us: Minds and the engineering of control systems', *APA Newsletters - Newsletter on Philosophy and Computers*, **10**(1), 12–17, (2010).
[58] Ricardo Sanz, 'Escaping the human constraint. towards a general theory of consciousness', in *Proceedings of the 16th Conference of the Association for the Scientific Study of Consciousness*, Brighton, UK, (July 14-17 2012).
[59] Ricardo Sanz, 'Reducing behaviorism and cognitivism', in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, eds., M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth, Berlin, (July 31-August 3 2013).
[60] Ricardo Sanz and Janusz Zalewski, 'Pattern-based control systems engineering', *IEEE Control Systems Magazine*, **23**(3), 43–60, (June 2003).
[61] J.P. Schadé and J. Smith, 'Computers of the brain and brainmade computers', in *Computers and Brains*, eds., J.P. Schadé and J. Smith, volume 33 of *Progress in Brain Research*, Elsevier, (1970).
[62] Daniel Selva, Bruce Cameron, and Ed Crawley, 'Patterns in system architecture decisions', *Systems Engineering*, n/a–n/a, (2016).
[63] Herbert A. Simon, *The Sciences of the Artificial*, MIT Press, Cambridge,USA, 3rd edn., 1996.
[64] Herbert A. Simon and Allen Newell, 'Models : their uses and limitations', in *The state of the social sciences*, ed., Leonard D. White, 66–83, University of Chicago Press, Chicago, (1956).
[65] Ben Smith, Valentina Tamma, and Michael Wooldridge, 'An ontology for coordination', *Applied Artificial Intelligence*, **25**, 235–265, (2011).
[66] James J. Staszewski, 'Models of human expertise as blueprints for cognitive engineering: Applications to landmine detection', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **48**(3), 458–462, (2017/01/26 2004).
[67] Hans-Georg Stork, 'Towards a scientific foundation for engineering cognitive systems –a european research agenda, its rationale and perspectives', *Biologically Inspired Cognitive Architectures*, **1**(0), 82–91, (7 2012).
[68] W3C. OWL - web ontology language - semantic web standards, 2012.
[69] Edward O. Wilson, *Consilience. The Unity of Knowledge*, Alfred A. Knopf, New York, USA, 1998.
[70] R. A Wilson, 'The cognitive sciences: A comment on 6 reviews of the mit encyclopedia of the cognitive sciences', *Artificial Intelligence*, **130**(2), 223–229, (8 2001).
[71] *The MIT Encyclopedia of Cognitive Sciences*, eds., Robert A. Wilson and Frank C. Keil, MIT Press, Cambridge, MA, 1999.
[72] David D. Woods, 'Cognitive technologies: The design of joint human-machine cognitive systems', *AI Magazine*, **6**(4), 86–92, (1985).