

7. Ontology Evaluation

7.1 Introduction

In this chapter we describe an experiment aimed at the validation of the ontological categories presented in chapter 5. As previously mentioned, the most important contribution of the ontology we developed is constituted by the philosophy-related categorical structures: in particular, the classes describing philosophical *ideas* and philosophical *events*. Accordingly, with the following experiment we want to evaluate such classes by means of a more empirical methodology.

By using a well known knowledge acquisition technique (card sorting) we have elicited a group of philosophers' implicit cognitive constructs *about* philosophy. As a result, we have been able to perform an initial evaluation of the *correctness* and *completeness* of our ontological categories against the philosophers' categorizations.

The method we have chosen can be described as a 'bottom-up' strategy, in so far as it relies on interviewing experts and then organizing the derived data by means of statistical methods. The choice of a bottom-up strategy is a deliberate attempt to contrast the main methodology we used in developing the ontology (as described in section 5.2.1). The latter was indeed the result of a 'top-down' approach, even though we took advantage of several discussions with domain experts and other informal knowledge acquisition techniques (such as the

consultation of books and encyclopedias). In a nutshell it is fair to say that all the knowledge we identified had been ‘mediated’ and re-organized by us, with the explicit purpose of creating a coherent formal model of the domain.

The approach used in this evaluation study is thus the following: we want to counterbalance the knowledge engineers’ results with the ‘direct’ and non-mediated representations of the domain experts. This can be obtained by means of an appropriate knowledge elicitation (KE) technique.

As we will see, our findings tend to align with the previous ontological definitions. Moreover, they also confirm one of the initial insights presented in the domain analysis section (5.2), namely that the philosophical domain is highly ambiguous and slippery. This feature makes the task of organizing it quite a controversial and debatable activity.

The chapter is organized as follows: in section 7.2 and 7.3 we describe the scope and approach of the experiment; section 7.4 describes the details involving the design of the experiment; section 7.5 outlines and categorizes the results, which are then discussed in section 7.6. Finally, section 7.7 concludes the discussion and outlines a few possible research directions that could improve on our findings.

7.2 Aims of the evaluation

Evaluating an ontology is process that stands out for both its importance and its difficulty. This situation is well described by Brewster and colleagues (Brewster, 2004):

There are inherent problems in trying to evaluate an ontology as it is not clear what exactly one is trying to evaluate. An ontology is a representation or model of knowledge, a “formal, explicit specification of a shared conceptualisation” according to (Gruber, 1993), and this means that however ‘shared’ it may be it is still extremely subjective, representing the time, place and cultural environment in which it is created. A particular ontology reflects the interests of the knowledge users, which must be captured in the design criteria for ontology construction. The labels of the concepts picked out to describe the concepts of interest to a user or application context, is an act of interpretation over the information available.

Despite these difficulties, several ontology evaluation approaches have been proposed. For example, a first major distinction can be made between a *quantitative* and a *qualitative* approach. A *quantitative* approach might involve mathematical and statistical methods for formally assessing the logical consistency or completeness of a model. Gomez-Perez (Gomez-Perez, 2001) and Gangemi (Gangemi et al., 2005) provide useful overviews of such methods. At the other end of the spectrum, a *qualitative* approach might involve the direct evaluation of an ontology by a chosen group of people, who are asked to rate it using different scales and methods. Often this second approach is more precise but, almost inevitably, more time-consuming. For example, the Ontoclean methodology (Guarino and Welty, 2002) defines a set of ‘identity criteria’ that one can use to analyze an ontology, so to reveal various modeling errors (such as inconsistencies in its taxonomical structure).

Another useful classification of evaluation strategies is the one proposed by Yu and colleagues (Yu et al., 2007). The authors highlight three common approaches to evaluating ontologies, which altogether lie on a plane that is perpendicular to the one of the qualitative/quantitative distinction introduced above:

- 1) Gold standard evaluation. This approach compares an ontology with another ontology that is considered as a benchmark. This is described, for example, in the work of Staab and colleagues (Maedche and Staab, 2002). Often this type of evaluation can be carried out by means of computational methods (i.e., quantitatively).
- 2) Criteria based evaluation. In this case the ontology is evaluated according to a set of proposed criteria. These criteria include *clarity*, *consistency*, *completeness*, *conciseness*, *expandability*, *correctness*, *coverage*, *minimal ontological commitment*, *minimal encoding bias*. Table 7-1 summarizes these criteria, as they were outlined in a precedent publication by Yu and colleagues (Yu et al., 2005). Usually this approach requires a degree of precision and expertise that only humans are currently able to do (i.e., it requires a qualitative approach).
- 3) Task-based evaluation. In this last case what is evaluated is the *competency* of the ontology in completing tasks. Usually, a task-based evaluation involves measuring the performance of the ontology in a quantitative manner within the context of a specific application. The main disadvantage of this approach is that an evaluation for one application may not be comparable with another one. Consequently, it is harder to generalize the results of an evaluation.

In the case of the PhiloSurfical ontology, we opted for the second type of evaluation, for the following reasons: first, because a gold standard evaluation (point 1) would not have been possible, since there are no other philosophical ontologies usable for performing an extensive comparison (as discussed in the literature review, chapter 3); second, because although the task-based evaluation (point 3) could have been an interesting alternative it would have required the PhiloSurfical tool's functionalities to be much more extensively developed and tested (cf. also section 8.6 in the next chapter).

1. Clarity	A clear ontology should effectively communicate the intended meaning of defined terms and where possible the definition should be stated formally. Definitions in an ontology should be clearly specified such that there is no doubt or ambiguity.
2. Consistency	Consistency, also called coherence, describes the logical consistency of an ontology. For a given ontology's set of definitions and axioms (explicit or inferred), there should be no contradictions.
3. Conciseness	An ontology is concise if it does not store any unnecessary or useless definition, declared explicitly or inferred.
4. Expandability	Expandability, also called extendibility, relates to the ability of the ontology to be extended further to describe specific application domains in a way that does not change the current definitions within the ontology.
5. Correctness	Correctness refers to whether the representational choice made when entities and their properties are modeled correlate with entities in the world being modeled. Correctness depends on the frame of reference that the ontology is based on.
6. Completeness	Whether an ontology or its individual definitions are complete cannot be proven, however we can deduce an incomplete ontology by detecting individual definitions that may be incomplete or that at least one definition is missing from an ontology. More specifically, we refer to a frame of reference to determining an ontology's incompleteness by using an ontology's set of competency questions or with its reference to the real world itself.
7. Coverage	Coverage can be intended in two distinct ways: coverage of terms over the domain from concepts identified in the domain; or coverage or completeness of instances. Other authors also refer to coverage as the congruence or fit of an ontology with the domain represented by the corpus of information. The difference here is that the text corpus is taken to represent the domain itself.

8. Minimal ontological commitment	Ontological commitment refers to an ontology being able to be agreed upon by users or ontology adopters. Minimal ontological commitment refers to minimizing the ontological commitment of an ontology to allow more freedom in an ontology's usage. It is about not over-defining terms that may impede some potential users of the ontology.
9. Minimal encoding bias	An encoding bias occurs when representational choices are made purely for the convenience of notation or implementation. This should be minimized to prevent an ontology to be specific only to a particular ontology notation, language specification or implementation.

Table 7-1- Criteria for evaluating ontologies (adapted from Yu, 2005)

Thus, we decided to evaluate our ontology *qualitatively* against a set of predefined criteria, with the help of a group of domain experts. In particular, we aimed at evaluating the **correctness** and **completeness** of our ontology (points 5 and 6 in table 7-1). As mentioned above, this was achieved by analyzing the results of a knowledge elicitation experiment aimed at discovering patterns of knowledge organization among experts in the philosophical domain. Thus, in general, we looked for similarities and differences in their perception of philosophy-related concepts so to be able to:

- a) highlight the common criteria and categories philosophers employ for organizing their domain;
- b) determine to what degree such criteria can be correctly translated into our ontological representation (*correctness* - in particular of the classes on the philosophical-idea branch), and whether other classes are needed to perform this operation (*completeness*).

- c) suggesting the design of new formal models and more specific evaluation studies for this domain.

For that regards the other possible criteria for evaluating an ontology (cf. table 7-1), we intend to examine them with greater detail in future work. It is worth remembering though that the *expandability* aspect is partly guaranteed by the fact that we are ourselves extending the CIDOC-CRM model, a well-known standard for modeling cultural heritage domains.

For that regards the *coverage* criterion, we share the view that this aspect is mostly related to the instances of an ontology. Therefore, since in our subsequent analysis we are dealing uniquely with classes, we believe that their capacity to map out the philosophical domain can be better expressed in terms of the aforementioned *completeness* criterion.

Finally, we reckon that the investigation of aspects such as *clarity*, *conciseness* and *minimal ontological commitment* of our ontology could largely benefit from a more mature community of researchers with expertise in both philosophy and knowledge representation techniques. As described in the literature review (cf. chapter 3), the formalization of philosophical and humanistic domains through SW technologies is still in its very infancy. Thus we believe that these types of evaluations are likely to produce more useful results once a more interdisciplinary community will have developed.

Instead, a separate discussion is needed for that regards the aspect of *logical consistency* of the ontology. In fact, the precise reader may ask a crucial question: why not employing well known ontology evaluation techniques aimed

at clearly separating out classes from properties, such as Ontoclean (Guarino and Welty, 2002)?

The answer we matured is that the inherent *abstractness* of a domain such as the philosophical one makes it not suitable for this type of ontology-evaluation techniques. In fact, approaches such as Ontoclean take a *realistic* stance towards the representation of knowledge. That is, they propose a series of schematic and precise principles (e.g., *rigidity*, *identity*, etc.) which aim at revealing the *objective* ontological status of the entities we want to describe.

In the case of philosophy, however, we recognized how its characteristic abstractness is inevitably related to a high degree of *subjectivity*. As a consequence, the objective principles of Ontoclean are not much of use here. For example, we cannot give a precise definition of what a ‘theory’ is, but we can easily make reference to a group of philosophers who are discussing about a theory.

In a nutshell, we realized that the ‘realms of abstractness’ are closer to the ever-changing spheres of *culture* and *debate*, rather than the more rigid ones of *science* and *analytical methods*. The incredible variety of philosophical positions competing with each other during the course of human history, in our opinion, has to be taken as an argument supporting this view.

As a result, when building the ontology our purpose shifted from an objective representation of the philosophical domain to a representation which could support the construction of learning narratives (see also section 6.6).

7.3 On the knowledge elicitation method chosen

For that regards the decision about which specific KE technique to use, the literature presents a variety of options (Gammack, 1987) which have often been ‘borrowed’ from other disciplines, such as cognitive science, anthropology, ethnology etc.. In general, KE techniques vary in the degree of structure they impose on the interview process. For example, on the ‘less structured’ end of the scale we can have *observation*, *interviews* and *focus groups*; on the other hand instead we have methods such as *sorting techniques* (Rugg and McGeorge, 2005), *laddering* (Corbridge et al., 1994, Rugg and McGeorge, 1995) and *repertory grids* (Rugg and Shadbolt, 1991, Shaw, 1980).

Among the various options, we decided to use a sorting technique, firstly because of its ease of use (for both the respondent and questioner) and implementation, secondly because, as shown in previous research (Rugg and McGeorge, 2005) it is suitable for non-scalar categories, such as the ones characterizing the philosophical domain:

Repertory grid technique encounter particular problems when dealing with nominal values, i.e. data which do not form any sort of semantic scale, and which are divided into non-scalar categories. Such categories are well-handled, however, by the sorting techniques.

This KE strategy is defined by Rugg and McGeorge (Rugg and McGeorge, 2005) as follows:

The basic idea behind the sorting techniques is simply to ask respondents to sort things into groups. The things may be *objects*,

such as different type of mouse, or *pictures*, such as screen dumps of various screen layouts, or *cards*, with the name of objects or situations on the cards, such as the names of different editors. The groups may be ones chosen by the questioner, or ones chosen by the respondent, or a mixture of both. The sorting techniques are a useful way of eliciting respondents' groups, and of finding out how much agreement and disagreement there is between respondents about the categories.

The theoretical foundations for this method refer to Kelly's Personal Construct theory (PCT) (Kelly, 1955). This constructivist theory emphasizes the role of our interpretations in defining our basic nature: to put it simply, it claims that the way we make sense of the world (i.e. the way we *interpret* it) is by constructing categories or theories that organize our experiences and sense-data. In particular, an important corollary of this theory is that when considering two (or more) persons, an existing similarity in the experience-constructs (i.e. the *interpretations*) is likely to suggest a similarity also at the level of their psychological processes. Therefore, by using a KE technique we attempt to reach this 'deeper' level of the understanding of reality; in particular, this is achieved by means of a statistical analysis of the categories that a set of subjects employ for interpreting a world's sub-domain.

During a card-sorting session, once a respondent has generated a number of groups, the researcher writes down the information about the *sorting criterion* used, the *categories* that have been created, and the *entities* belonging to each group. The process is then repeated until the respondent runs out of ordering criteria.

At the end of the experiment, the data can be analyzed by looking at the number and nature of the criteria generated (also called ‘constructs’). Criteria can then be organized by type or because they are common among respondents. Often, since the choice of wording may vary among respondents (even if they refer to the same criteria) it is useful to group these ‘verbatim criteria’ into some meta-criteria. This will help highlighting similarities. In the next sections we will see in more detail how this strategy has been applied in our scenario.

7.4 Experiment design

7.4.1 Specific method and procedure

We decided to implement the sort methodology presented above by using a set of paper cards carrying the name of philosophy-related concepts (card-sorting technique, CS). The use of named paper cards was motivated mainly by the fact that they could be quickly and easily created by us. Moreover, since most of the philosophical entities we are investigating are abstract in nature, it would have been impossible to represent them by means of physical objects or cards carrying drawings or photos.

The authors of (Rugg and McGeorge, 2005) enumerate various types of sorting techniques: according to their classification our experiment can be seen as an example of both *repeated single-criteria sorts* and *hierarchical sorts*.

The former, because we asked the respondents to sort the cards repeatedly, to create groups representatives of a different single attribute (i.e. a criterion) each time.

The latter, because we chose cards representing entities at different semantic levels (e.g., theories, people, events). This is motivated by the fact that, as discussed previously in the context of entities such as ‘rationalism’ (cf. section 5.3.5.1), in the philosophical domain the task of defining the ‘type’ of an entity is not at all obvious. Thus, through the card-sorting exercise we also aimed at establishing a basic semantic hierarchy of philosophical entities.

We asked the volunteers to sort the cards repeatedly so to group them each time according to a different and unique criterion of choice. No criteria were provided by the researcher. At the end of the sort, the researcher would write down the results using a pre-formatted table with columns for criteria, categories and card-numbers (similar to the tables in Appendix A, section 2). The results were recorded verbatim as given by the respondents. Before starting the session, the volunteers were given a verbal example of what a sort could be like, in the following form:

“Imagine you had a set of cards representing cars: you might sort them according to their color; or also according to their style. In the first case, you would probably generate groups such as ‘white cars’, ‘red cars’ etc.. In the second case instead you may create groups such as ‘sport cars’, ‘4x4 cars’ etc..”

The session was concluded by annotating the nationality of the volunteers, and asking them to assess their level of philosophical knowledge on a three-valued scale ('low', 'medium' or 'high').

7.4.2 Respondents

For the experiment we gathered 12 volunteers. 5 of them were british and english-speaking, while the remaining 7 were italian and italian-speaking. All the respondents were aged between 20 and 40, 8 were males and 4 females.

Respondents can be divided into two groups of equal number (6), according to their declared level of expertise. The first group (expertise level high or medium-high) comprises 6 males linked to the Open University teaching and research environment. All of them have a philosophical teaching-related position (professor, lecturer, assistant lecturer).

The second group (expertise level medium or low) comprises 4 females and 2 males attending a PhD course at the University of Bologna (Italy). The important feature of the people in this group is that philosophy is not their primary research subject but just a subsidiary topic. In fact, for 2 of them the main research topic is psychology and cognitive sciences, while the remaining 4 are studying semiotics. Although these areas are not purely philosophical, it is fair to say that they involve a large deal of philosophical knowledge. Therefore, by forming this second group we tried to gather people with a less technical but still academic understanding of the philosophical domain. In order to avoid the

situation where respondents of this group had scarce or no philosophical knowledge at all, during the selection phase we asked all of them to specify whether they were familiar with philosophy or not, and especially with the philosophy of language. All the volunteers gave a positive response.

The table below sums up the information about the people involved in the experiment.

Volunteer	Language	Declared knowledge level
VOLUNTEER-1	ENGLISH	HIGH
VOLUNTEER-2	ENGLISH	HIGH
VOLUNTEER-3	ENGLISH	HIGH
VOLUNTEER-4	ENGLISH	HIGH
VOLUNTEER-5	ENGLISH	MEDIUM-HIGH
VOLUNTEER-6	ITALIAN	HIGH
VOLUNTEER-7	ITALIAN	MEDIUM
VOLUNTEER-8	ITALIAN	MEDIUM-LOW
VOLUNTEER-9	ITALIAN	LOW
VOLUNTEER-10	ITALIAN	LOW
VOLUNTEER-11	ITALIAN	LOW
VOLUNTEER-12	ITALIAN	LOW

Table 7-2. Volunteers' nationality and declared philosophical expertise level

7.4.3 Materials

We used 23 cards representing philosophy-related concepts (see Table 7-3).

The cards have been created on a computer using a commonly available graphics software, then printed on white paper sized 10x5 cm. Cards were randomly numbered from one to twenty-three for recording purposes. An example of how the graphical cards looked like can be found in Appendix A, section 1.

Given the double nationality of the respondents, we created two equivalent sets of cards, one in english and one in italian. The translation of the concepts from english to italian was straightforward and did not cause any ambiguity problem. For that regards the cards' concepts, we selected them in order to have a representative sample of the various aspects of the ontology. Even if we did not carry out any formal evaluation of the cards' concepts (for example, in relation to people's ability to understand their meanings) during the experiment none of the respondents showed significant difficulties in using them. As a conclusion, we believe that the cards' choice was positive.

Card Number	Value (english)	Value (italian)
1	Wittgenstein's 2nd philosophy	2nda filosofia di Wittgenstein
2	nominalism	nominalismo
3	epoche'	epoche'
4	monism	monismo
5	dialectic	dialettica
6	mathematics	matematica
7	enlightenment	illuminismo
8	Aristotle's philosophy	filosofia di Aristotele
9	Ludwig Wittgenstein	Ludwig Wittgenstein
10	philosophy of religion	filosofia della religione
11	substance problem	problema della sostanza
12	a posteriori	a posteriori
13	theory of descriptions	teoria delle descrizioni
14	truth table method	metodo delle tavole di verita
15	Vienna circle	circolo di Vienna
16	analytic philosophy	filosofia analitica
17	meaning	significato
18	Wittgenstein's ontology	ontologia di Wittgenstein
19	Plato	Platone
20	logical atomism	atomismo logico
21	ontology	ontologia
22	Continental philosophy	filosofia continentale
23	world	mondo

Table 7-3. Cards used in the experiment

7.4.4 Software support

In order to analyze the data we made use of a purposely-built software tool that allows the visualization of multiple perspectives on the data. This is an ontology-based web-application which supports the navigation of the data and the automatic creation of html tables aimed at facilitating the interpretation of the results. For example (cf. figure 7-1) by clicking on a respondent name it is possible to see all of his/her sorts and categories in one table, while, by clicking on a meta-criteria we can see which criteria and categories it comprises, etc..

The application relies on a simple ontology with five classes (namely *volunteer*, *criterion*, *category*, *card* and *meta-criteria*), which has been instantiated using the data from the experiment. The application uses OCML (Motta, 1999) for the knowledge modeling and data-storage, the lisp programming language (Lispworks, 2008) for the server and back-end functionalities, html and ajax techniques (Holdener, 2008) for the user interface.

Although it has been implemented only as an *ad-hoc* solution for card sorting data visualization, the application turned out to be of fundamental importance for the success of the experiment. Hence we do not exclude to develop it further and produce a generic framework for supporting the visualization of data elicited through card-sorting.

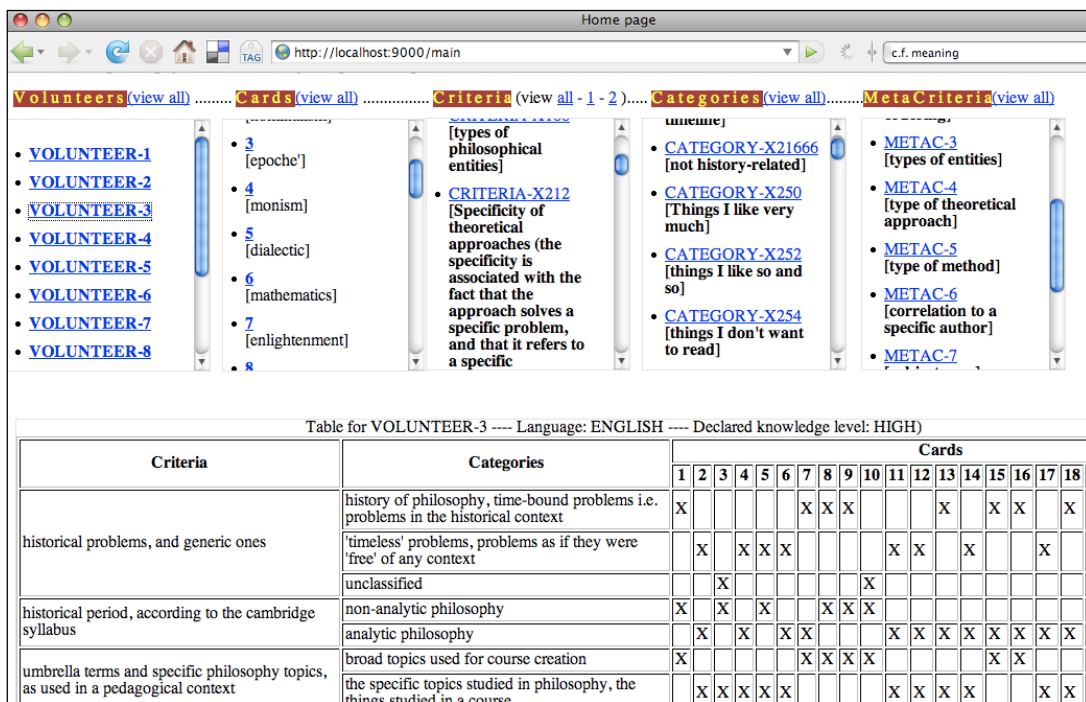


Figure 7-1. Screenshot of the software application built for visualizing the experiment data

7.5 Experiment results

In general, respondents did not encounter any difficulty in understanding the task and sorting the cards multiple times. Often they found the experiment a fun activity to do and gained new ideas about novel ways to organize the cards after the first or second sort.

All the results recorded for each respondent can be found in Appendix A, section 2. In each table each appearance of a criterion corresponds to a new sort; the categories, instead, are representatives of the groups formed during each sort.

7.5.1 Comments on respondents' native language

As already mentioned, the results of each sort have been recorded manually on pre-formatted sheets. Since more than half of the respondents (seven out of twelve) were not native english speakers, their sessions have been conducted in italian. Accordingly, the results of such sessions were recorded *verbatim* using the italian language.

Sometimes, given the respondents' proficiency in the english language, the categories and criteria created have been annotated in both languages, under their supervision. However, in the majority of cases we translated the results from italian to english, usually after the session. We always aimed at producing a literal translation, paying extra attention not to modify the original authors' intended meaning. The translation process has always been simple and straightforward, mainly because the original verbatim results were always clear and unambiguous.

7.5.2 Number of criteria and categories

In total a number of 52 criteria were generated. The total number of criteria generated by each respondent varied overall from 2 to 8. In group A this value varied from 2 to 6, with an average of 4.33 criteria per person. In group B the value varied from 2 to 8 with an average of 4.33.

Table 7-4 shows the criteria (26) created by group A, the philosophy experts (namely, volunteers 1 to 6); these are compared with the criteria generated by

group B, the people with less philosophical knowledge (namely, volunteers 7 to 12).

	GROUP A	GROUP B	TOTAL
CRITERIA	26 (avg. 4.33 per person)	26 (avg. 4.33 per person)	52
CATEGORIES	104 (avg. 17.3 per person)	72 (avg. 12 per person)	176

Table 7-4 - Summary of the experiment results

For that regards the categories generated, they are in total 176. Group A generated 104 categories; the average number of total categories per person is 17.3. The number of categories per criterion varied from 2 to 11.

Group B generated a total of 72 categories; the average number of categories per person is 12, while the number of categories per criterion varied from 2 to 7.

Criteria	No of Categories	Volunteer
history of philosophy related	2	1
Things I like reading	3	1
chronological ordering	7	1
Pedagogical criterium - things a philosopher should know, resembling a curriculum-creation criteria	3	1
types of philosophical entities	7	1
Specificity of theoretical approaches (the specificity is associated with the fact that the approach solves a specific problem, and that it refers to a specific philosopher)	4	1
Type of entities and connection to Wittgenstein	5	2
Abstract/concrete distinction	2	2
Types of entities	7	2
Relation to philosophy	1	2
Types of entities, alternative classification	6	2
Degree of connection to Wittgenstein	2	2
historical problems, and generic ones	3	3
historical period, according to the cambridge syllabus	2	3
umbrella terms and specific philosophy topics, as used in a pedagogical context	2	3
importance of topics, according to a specific school of philosophy (cambridge)	3	3
By subject areas	3	3
chronological ordering	4	4
philosophies' approach / problems investigated	3	4
type of entities	4	4
kind of entities	4	5
different ways to look at the world, approaches and schools	5	5
type of generic philosophical approach	2	5
entities' type in relations to their respective philosophical approach	6	5
type of things	11	6
types of problems, problem areas	3	6

Table 7-5. Verbatim criteria generated by group A, namely volunteers 1 to 6 (the experts)

In tables 7-5 and 7-6 we can see the list of criteria generated by the two groups, together with an indication of how many categories there are for each criterion. At first sight, it is not difficult to identify in both groups some constructs which are the same (i.e. 'chronological ordering') and several constructs which are very similar (i.e. 'type of entities' and 'type of things'). Using these similarities, in the following sections we will gain more insights about the significance of the data collected.

Criteria	No of Categories	Volunteer
closeness of perspective, or of the problems tackled	5	7
type of philosophical entities	7	7
proximity to schools of thought	3	8
proximity to Wittgenstein's views	3	8
problem areas, questions asked	4	8
chronological ordering	3	9
types of philosophical things	5	9
concepts which are specific of philosophy only, or of other disciplines too	2	10
ordering based on the type of methods usually associated to a concept	3	10
types of ideas	3	10
criteria based on the dichotomy container/contents	2	10
concepts specific to an author or group of authors, intended as a single author	2	10
things linked to Plato or Aristotle or not	2	11
things related to Wittgenstein, or not	3	11
philosophy-types vs the rest	2	11
themes related to enlightenment, or not	2	11
things I know or not	2	11
problems and non problems	2	11
people vs all the rest	2	12
stuff related to the 'reference' problem, or not	2	12
philosophical schools of thought, or not	2	12
correlation to the semiotic/meaning problem	2	12
stuff related to Wittgenstein, or not	2	12
things that belong to antiquity, or not	2	12
historical ordering	3	12
concepts which are familiar to me, or not	2	12

Table 7-6. Verbatim criteria generated by group B, volunteers 7 to 12 (the non-experts)

7.5.3 Commonalities of verbatim criteria

In order to give an account of the commonalities of criteria, we grouped the verbatim criteria according to their similarities, thus generating a small set of *meta-criteria*. This was necessary for two main reasons, which are:

- a) because respondents' verbatim criteria often present wording differences, while expressing the same meaning (e.g., 'type of entities' and 'type of things');
- b) because almost half of the criteria were originally expressed in the Italian language and only subsequently translated into English.

In total we identified 9 meta-criteria (see table 7-7 for a summary; also, a complete illustration of the criteria/meta-criteria mappings can be found in appendix A, sections 3 and 4. The number of criteria subsumed by each meta-criterion varies from 1 to 15, with an average of 6.3 criteria per meta-criterion (note that in certain cases it was necessary to allow a single criterion to belong to more than one meta-criterion - we will discuss this later). The meta-criteria generated are the following:

- 1) **Types of entities.** This is the meta-criterion with the highest number of criteria. It contains a total of 15 criteria, 9 specified by volunteers in group A and 6 in group B. In general, it refers to all the sorts which attempted to organize entities according to their 'ontological' type, that is, their *principium individuationis* (what makes them essentially unique and different from the others). Examples of categories generated within such meta-criterion can be 'people' or 'problems' (cf. Appendix A, table A4-3).
- 2) **Time-based ordering.** With this meta-criterion we grouped 8 criteria, 5 specified by volunteers in group A and 3 in group B. It refers to all the orderings which are based on chronological time or, in general, on an

- historical approach - e.g., '*medieval philosophy*', '*modern philosophy*' etc. (cf. Appendix A, table A4-1)
- 3) **Subject area.** With this meta-criterion we grouped 8 criteria, 3 specified by volunteers in group A and 5 in group B. It identifies criteria highlighting whether or not an entity is usually associated to one of the classical subject areas in philosophy - e.g., '*ontology*' - or to any defined topic intended as an area of investigation - e.g., '*meaning*' (cf. Appendix A, table A4-2).
- 4) **Type of problems tackled.** With this meta-criterion we grouped 7 criteria, 3 specified by volunteers in group A and 4 in group B. It refers to criteria that focus on the kind of problems tackled in order to organize the philosophical entities - e.g., '*substance problem*' (cf. Appendix A, table A4-4).
- 5) **Correlation to a specific author.** With this meta-criterion we grouped 6 criteria, 2 specified by volunteers in group A and 4 in group B. It identifies criteria that highlight the proximity of an entity to a specific author's philosophy, or, in general, to one of his/her views. For example, we can have '*degree of connection to Wittgenstein*' (cf. Appendix A, table A4-5).
- 6) **Type of theoretical approach.** With this meta-criterion we grouped 6 criteria, 4 generated by volunteers in group A and 2 in group B. The specificity of the criteria in this group resides in the fact that they all attempt to organize the cards' entities by putting the emphasis on the 'philosophy' or 'school of thought' they belong to (or are normally associated with). So, for example, we can have categories such as '*linguistic approaches*' or '*analytic philosophy*'. If compared to meta-criterion 1, the difference is that in this case the criteria gathered represent instances of a specific ontological type, that is, 'theoretical approaches' or 'views' (cf. Appendix A, table A4-6).

Meta-criteria	Description	Tot. No. Of Criteria
1) Types of entities	It refers to all the sorts which attempt to organize entities according to their 'ontological' type	15
2) Time-based ordering	It refers to all the orderings which are based on chronological time or, in general, on an historical approach	8
3) Subject area	It identifies criteria highlighting whether or not an entity is usually associated to one of the classical subject areas in philosophy, or to any defined topic intended as an area of investigation	8
4) Types of problems tackled	It refers to criteria that focus on the kind of problems tackled in order to organize the philosophical entities	7
5) Correlation to a specific author	It identifies criteria that highlight the proximity of an entity to a specific author's philosophy	6
6) Type of theoretical approach	It refers to criteria putting the emphasis on the 'philosophy' or 'school of thought' the cards' entities belong to	6
7) Importance/ pedagogical perspective	it refers to sorts by which respondents wanted to highlight the relevance of an entity with respect to a 'socially accepted scale of values'	3
8) Things I like/know	It refers to criteria expressing personal preferences and familiarity with certain topics	3
9) Type of method	It refers to types of methods used in deriving the entities described in the cards	1

Table 7-7. The nine meta-criteria we identified

- 7) **Importance / pedagogical perspective.** With this meta-criterion we grouped 3 criteria, all of them specified by volunteers in group A. Essentially it refers to sorts in which respondents wanted to highlight the relevance of an entity with respect to a 'socially accepted scale of values' (e.g., '*importance according to the Cambridge curriculum*'). The pedagogical perspective (which appears in two cases) was treated here as an alternative manifestation of the 'importance' meta-criterion (cf. Appendix A, table A4-7).
- 8) **Things I like/know.** With this meta-criterion we grouped 3 criteria, 1 specified in group A and 2 in group B. It refers to criteria expressing personal preferences and familiarity with certain topics. It is different from the meta-criterion 7 because it is not meant to represent any principle that is objectively and socially shared (cf. Appendix A, table A4-8).
- 9) **Type of method.** This is the meta-criterion containing the fewest criteria, i.e. only 1 verbatim criterion which is '*ordering based on the type of methods usually associated to a concept*'. We decided to introduce this single-item meta-criterion because, even after repetitive considerations, we could not make it fit anywhere else without forcing the original author's intended meaning. The categories associated with it (which are 'rational or demonstrative method' and 'discursive, qualitative or observational method') explicitly make reference to types of methods used in deriving the entities described in the cards. Therefore, they could not be inserted in any of the other groups. It is worth noting that also this meta-criterion, if compared to meta-criterion 1, also this meta-criterion (similarly to what said above for meta-criterion 6) appears to gather elements belonging to *one type* of entities only - i.e. the 'methods' (cf. Appendix A, table A4-9).

A final issue needed being dealt with. In fact in certain cases the verbatim criteria were not specific enough in order to be unambiguously positioned within one group only. For example, the verbatim criterion '*historical problems, and generic ones*' mixes a time-based perspective with a problem-centered one. We admit that this is a shortcoming not of the respondents' constructs, but instead of the questioner's methodology. In other words, even if we asked respondents to use only one criterion per sorting, we did not manage to get them to apply such principle consistently so to completely avoid the usage of 'double criteria'. As a solution, in four cases we decided to position such 'double criteria' within more than one meta-criteria. More specifically, this happened in the following situations:

- 1) with the just mentioned '*historical problems, and generic ones*' (criterion generated by volunteer 3), which we grouped under meta-criteria number 2 and 4.
- 2) With the criterion '*philosophies' approach / problems investigated*' (generated by volunteer 4) which we grouped under meta-criteria 4 and 6.
- 3) With the criterion '*type of problems, problem area*' (generated by volunteer 6) which we grouped under meta-criteria number 3 and 4.
- 4) With the criterion '*closeness of perspective, of the problems tackled*' (generated by volunteer 7) which we grouped under meta-criteria number 3 and 4.

It is interesting to look at these double criteria not just as ‘mistakes’ but instead as implicit indications of categories which people tend to associate or even mesh together. In the discussion section we will suggest how these details could become another source of insight from a knowledge representation perspective.

To conclude, table 7-8 shows the total number of criteria per meta-criterion, their distribution among the two groups, an example of the subsumed criteria and the total number of categories which are indirectly linked to each meta-criterion.

Meta-criteria	ex. verbatim criteria	Tot. No. Of Criteria	Group A	Group B	Tot. No. of Categories
1) Types of entities	“Philosophical schools of thought, or not”	15	9	6	73
2) Time-based ordering	“Historical ordering”	8	5	3	26
3) Subject area	“Themes related to enlightenment, or not”	8	3	5	22
4) Types of problems tackled	“Correlation to the semiotic/meaning problem”	7	3	4	20
5) Correlation to a specific author	“Stuff related to Wittgenstein, or not”	6	2	4	16
6) Type of theoretical approach	“Proximity to Wittgenstein’s views”	6	4	2	20

Meta-criteria	ex. verbatim criteria	Tot. No. Of Criteria	Group A	Group B	Tot. No. of Categories
7) Importance/pedagogical perspective	“Importance of topics, according to a specific school of philosophy (Cambridge)”	3	3	0	8
8) Things I like/know	“Concepts which are familiar to me, or not”	3	1	2	7
9) Type of method	“Ordering based on the type of methods usually associated with a concept”	1	0	1	3

Table 7-8. Meta-criteria and criteria distribution per groups

7.5.4 Commonality of verbatim categories

Given the high number of categories generated in total (176), we investigated the commonality of categories only in the context of the meta-criteria previously identified. In general, we hardly found *verbatim* categories which were the same. Nonetheless, in many cases it was still easy to determine the equivalence of the categories because they had only minimal wording differences (e.g., ‘*techniques, methods*’ and ‘*methods*’). In the following paragraphs we will outline our findings, which are ordered according to the meta-criteria organization previously presented.

7.5.4.1 Meta-criterion 1: *types of entities*

This is the group of criteria with the highest number of similar verbatim categories. In particular, the terms which have been used by different people

were the following: 'doctrines', 'problems', 'methods', 'people', 'movements', 'positions', 'topics', 'concepts', 'themes' and 'schools of thought'.

This apparently high degree of similarity motivated us in carrying out a very detailed examination of the way the cards were classified, hoping to find some commonalities also at that level (e.g., we expected to find out that the majority of people classified card 20 as a 'school of thought').

However, this turned out an incredibly difficult task to do, because respondents' results did not seem to have any consistency or stability. As an example, we report in detail a specific attempt to trace a similarity of categorizations at the cards level. Concretely, we checked wherein card 4 (i.e. 'monism') had been repeatedly categorized using the same construct. It turned out that people in different sessions classified this same card as a 'method', a 'doctrine', a 'movement' and a 'problem' (cf. table 7-9). Clearly, there is not much overlapping among those terms, at least not without stretching their meanings quite considerably.

As a consequence of these negative results, we abandoned this strategy and opted instead for a more qualitative analysis of the data, aimed at re-constructing the implicit model each of the respondents' made use of when organizing the cards' entities by type. We will say more about this second approach in the discussion section (cf. section 7-6) .

In conclusion, we can say that the categories in this group exhibited only a superficial similarity, because respondents mostly used them to express different meanings.

Ontological Requirements for Supporting Smart Navigation of Philosophical Resources

Table for CARD-4: 'monism' according to Meta Criteria: 'types of entities'		
Category	of Criteria	Volunteer
doctrines	Types of entities	2
contents: the thing analysed	criteria based on the dichotomy container/contents	10
externalist problems	entities' type in relations to their respective philosophical approach	5
movements - more specific than the philosophies	types of ideas	10
things which are not schools of thought	philosophical schools of thought, or not	12
positions - within the strands of philosophy	type of entities	4
Theoretical approaches, ways of solving problems (more specific than schools of philosophy: one school can contain many of them!)	types of philosophical entities	1
philosophical positions, 'answers' to the problems (the positions can be sustained by the epochal movements)	type of things	6
Not connected to Wittgenstein	Type of entities and connection to Wittgenstein	2
groups of approaches	types of philosophical things	9
abstract entities	Abstract/concrete distinction	2
not people	people vs all the rest	12
currents of thought traversing the history of philosophy	type of philosophical entities	7
methods, ways to break things down or specific approaches	kind of entities	5

Table 7-9. All the categories used for card 4, in the context of meta-criterion 1.

7.5.4.2 Meta-criterion 2: time-based ordering

In this group the only two categories described using the same terms were 'antiquity' and 'modernity'.

7.5.4.3 Meta-criterion 3: subject area.

Quite surprisingly, we could not find any overlapping verbatim category within this group. We were not expecting such a large variety of results here, specifically for two reasons: the first one is that we initially had chosen most of the cards exactly because of their connection to the philosophical areas of logic and philosophy of language. The second one is more general, and it refers to the fact that although philosophy can have many sub-disciplines, its 'core' areas are not many (traditionally, textbooks might highlight metaphysics, logic, epistemology, ethics and a couple of others). For these reasons we were expecting respondents to have more similar results.

7.5.4.4 Meta-criterion 4: *type of problems tackled*

Also this meta-criterion did not comprise any common or overlapping categories. For the same reasons exposed in the precedent paragraph (meta-criterion 3) we found such results quite unexpected.

7.5.4.5 Meta-criterion 5: *correlation to a specific author*

In this group the only categories exposing a high degree of similarity are the ones describing a connection to Wittgenstein: for example, the verbatim categories can be 'connected to Wittgenstein', 'things related to Wittgenstein', 'stuff related to Wittgenstein' etc.

7.5.4.6 Meta-criterion 6: *type of theoretical approach*

This group did not present any common category, except for the appearance twice of the notion of 'analytic' philosophy in the verbatim categories 'analytic and empiricist approaches' and 'analytic philosophy'.

7.5.4.7 Meta-criterion 7: *importance / pedagogical perspective*

We did not find any common category in this group.

7.5.4.8 Meta-criterion 8: *things I like/know*

Also in this group we did not find any common category.

7.5.4.9 Meta-criterion 9: *type of method*

This group gathered only results from a single person's session, thus we (obviously) could not find any similar categories.

7.6 Experiment discussion

In this final section we will go through the experiment results with the purpose of understanding better what are the categories used by our volunteers for organizing the philosophical concepts on the cards. In particular, we will compare such results with the classes in the PhiloSufical ontology, paying special attention to those in the `philosophical-idea` branch (as shown below, fig. 7-2).

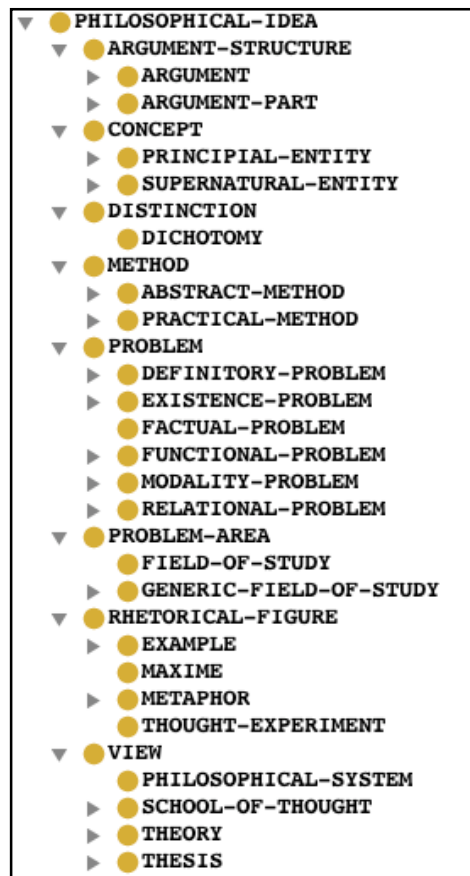


Figure 7-2. The 'philosophical' branch of the PhiloSufical ontology, as it appears in Protege'

In general we can say that most of the respondents' constructs could be matched to some classes in our model (*completeness* aspect). Also, what emerged from the respondents' (more or less explicit) cognitive structures

seemed to prove the ontological structures created by us (*correctness* aspect). The rationale for these claims are discussed extensively in the following sections.

However, there is also a vast number of philosophy-related classes in the ontology that remained unevaluated, since respondents did not mention them at all. For example, as we will see, none of the respondents attempted to classify *problems* according to their logical structure, which is instead the solution we have chosen in the ontology. We believe that this *partial* lack of coverage could be mostly related to the fact that the philosophical branch of our ontology is way too vast and detailed to be covered with a single card-sorting experiment.

As a preliminary step towards the solution of this problem, in the final section of this chapter we discuss the design of future and more specific experiments, so that they will complement and build on our results.

It is also important to remember that although the card sorting technique has been successfully used in many domains (e.g., software requirements' definition (Nurmuliani et al., 2004) or for investigating websites' usability (Maurer and Warfel, 2007)), we did not find any scientific report of experiments investigating the type of abstract entities which are important in the philosophical domain.

For this reason, we acknowledge that the usage of card-sorts in such an unexplored domain should be considered more an exploratory technique, rather than a well-established knowledge elicitation strategy. Thus we must consider this experiment only as a *first step* in the right direction; the task of acquiring formal or semi-formal knowledge in a domain as fluid as the philosophical one undoubtedly appears to be a worthy enterprise for future research.

In the following sections we highlight the main themes emerging from the analysis of the experiment results. A complete list of the experiment results, ordered according to the classification of meta-criteria, can be found in appendix A, section 4.

In general, we interpreted the data by going through all the various criteria comprised within each meta-criterion, so to gather more insight about their role and significance. Often, in order to better understand what respondents intended to mean with a criterion we analyzed in more detail also the single categories created within that criterion. This ‘extra’ analysis have been particularly useful in the case of meta-criterion 1 (cf. section 7.5.3, point 1). In fact, the ‘*type of entities*’ meta-criterion appears to be a ‘*meta*’ categorization also with respect to the classes in our ontology. Precisely, it is exactly the kind of perspective we adopted when creating the ontology. Thus in this case, in order to find interesting results in the experts’ constructs, it has been necessary to examine carefully all the *categories* comprised by each single ‘type of entities’ criterion, as created by the respondents .

7.6.1 Experts vs non-experts

First of all, by looking again at table 7-8, we can see how the distribution of meta-criteria per groups reveals an interesting trend: the meta-criteria no. 1, 2, 6 and 7 (e.g., ‘importance/pedagogical perspective’, ‘time-based ordering’, ‘types of entities’ and ‘type of theoretical approach’) have the biggest

contributors in the respondents of group A (the experts). We can interpret this as an indication that those classifications are more difficult to produce for a non-expert, because they require a more systematic vision of the field.

7.6.2 Philosophers like ontology

This is a conclusion that we drew only by looking at the number of categories created. In particular, we are referring here to the ones defined in the context of the '*type of entities*' meta-criterion. Our respondents seemed to have a natural predisposition for sorting the cards according to their ontological type. In total, under this kind of organization we had 15 criteria (the group with the highest number) and 73 categories (cf. table 7-8).

Thus we believe it is fair to draw the conclusion that *philosophers tend to look at their domain using an ontological schema*.

This insight is maybe not so surprising, if we consider that the job of categorizing knowledge and building ontologies is eminently philosophical. But on the other hand, from a software engineering perspective this result could be a hint of the fact that ontology-driven navigation mechanisms are likely to be well-understood and positively received by philosophers. Although it is clear that we do not have any compelling evidence for this last claim, we believe that this is an inspiring result that calls for more research on the topic. In fact, other positive findings in this direction could be used to validate our ontology-based learning narratives approach (cf. section 4.3).

7.6.3 People vs Ideas

Seven subjects (respondents n. 1, 2, 4, 5, 6, 7, 12) out of twelve agreed about the fact that '*individuals*' and '*groups of people*' stand out against the rest of the entities, all of which are abstract in nature (see also section 7.6.9). Nonetheless, there is also some contradictory evidence about another connotation of the '*group of people*' category. In fact, sometimes this is seen as equivalent to the '*school of philosophy*' or '*school of thought*' category (cf. Volunteers 1 and 7); in other cases instead, '*school of thought*' respondents mainly wanted to refer to an abstract and impersonal entity (cf. Volunteers 2 and 5).

In our opinion these results confirm our view in the ontology that groups of people, schools of thought and intellectual movements are *fundamentally* related to each other. This ambiguity is dealt with in the ontology thanks to the modeling pattern discussed in section 5.3.5.2. This model provides a way to formalize the aforementioned three types of entities in such a way that each of them becomes representative of a different aspect (i.e., the abstract side, the people-related side, or the temporal side) of a more complex phenomenon. In conclusion, we take this as an encouraging result that proves the correctness of the modeling pattern.

7.6.4 Time-related entities

Although our ontology covers various aspects linked to temporal knowledge, in the `philosophical-idea` branch represented above we do not make use of any time-related class. This was done purposely, since we made a clear distinction

between *continuants* and *occurrents* (and philosophical ideas are sub-types of continuants).

On the contrary, many criteria and categories created by the respondents (in particular, those ones grouped under meta-criterion 2, cf. table A4-2) organize philosophical concepts using temporal constraints. For example, we can have ‘*greek philosophy*’ or ‘*enlightenment period*’. We take this as an indication that time-based representations are a natural way for people to ‘index’ philosophical entities, even if such entities are not events but abstract ideas. Moreover, it is important to underline that from the respondents’ results’ sometimes it is not clear whether, by producing a ‘chronological ordering’, they refer to a sequence of *events* or to a a set of *subject-areas* that also happen to be sequenced according to their time occurrence (cf. for example the categories generated by volunteer 3 and 12 in table A4-2).

Whatever the interpretation of these results, we believe that in both cases they are confirming our ontological decisions. In the first case, in the philosophical ontology we can represent all those entities as various occurrences of *events*, that is, time-frames more or less linked to the philosophical world (for example, they could be represented as ‘philosophical movements’). In the second case instead they could be represented as linked to the *field-of-study* class; more precisely, as instances of one of its subclasses, *philosophical-discipline*.

7.6.5 Historical vs timeless

Another interesting point regarding the just mentioned time-related categories is the distinction various respondents made between things which *are in history*

(e.g., with the verbatim category '*greek philosophy*', table A4-2, volunteer 1) and things which are *outside history* (e.g., with the verbatim category '*timeless problems, problems as if they were free of any context*', table A4-2, volunteer 3). We believe this is a subtle but fundamental feature of the domain of abstract ideas. In fact, although none of the respondents seemed to agree on *what* are the historical or timeless entities (e.g., by indicating consistently the same cards as 'historical'), they all agreed on the existence of the dichotomy. For example, the same idea could be considered as linked to the authors or events associated with its creation (the 'historical' way), or in terms of its logical characteristics (the 'timeless' way).

We therefore drew the conclusion that the historical/timeless feature of ideas must be related to some intrinsic (= essential) properties of them; in other words, ideas seem to be rightfully described under both the 'historical' and 'outside history' perspectives (of course, each perspective entails a different description). This feature stands in opposition with other types of properties: for example when we say, of an entity, that it must be *either* material *or* abstract.

The same phenomenon was recognizable in the sorts organized under meta-criterion 1 ('*type of entities*', table A4-3). Also here, the spatial and historical dimensions appear to be a constant criteria of identification: so we can have '*doctrines tight to particular individuals and times*' as opposed to a more generic '*doctrines*'; or similarly, we have '*broad areas of enquiry, with no historical dimension*' as opposed to '*topics Wittgenstein worked on*' (examples taken from volunteer 2).

Therefore, as a conclusion, we believe that respondents tend to manifest a somehow 'double-sided' representation of philosophical entities: the underlying

principle is that abstract things can always be easily ‘specified’ by associating them to concrete ones, and vice-versa. Our ontology provides a mechanism for dealing with this feature of ideas, which is the *school-of-thought / contextualized-school-of-thought* modeling pattern (cf. section 5.5.5.4). In brief, we created a formal model for describing schools of thought at various levels of abstraction. This mechanism has been specifically created for one type of philosophical idea only. Thus, given the results of this experiment, in future work we intend to extend this modeling approach also to the other types of ideas.

7.6.6 Problems

Ten (volunteers 1, 2, 3, 4, 5, 6, 7, 9, 11, 12) out of twelve respondents identified ‘*problems*’, which are often also associated to ‘*themes*’, ‘*topics*’ and even ‘*concepts*’ (see also the sorts under meta-criterion 1). In general, they looked at these entities as the ‘*explananda*’ (a term used by volunteer 2) which are tackled by ‘philosophical approaches’ of varying granularity and abstractness.

However, while in the ontology we aimed at organizing problems by looking at their internal structure (e.g., problems of the kind ‘what is X?’) or at their philosophical ‘purpose’ (e.g., problems investigating the ‘function’ or the ‘mode of being’ of an entity), the respondents’ results did not go in this direction. Instead, they tended to define problems according to the specific topics they are about (e.g., ‘*meaning*’, table A4-4, volunteer 6) or by means of their correlation to an author (e.g., ‘*Wittgenstein’s perspectives and problems*’, table A4-4, volunteer 7).

It remains unclear though how such terms should be situated on a semantic scale.

In conclusion, although the ontological existence of a ‘problem’ class needs no discussion, the experiment cannot provide any evidence about the correctness of the ontological classes, which are subtypes of ‘problem’.

7.6.7 Subject areas

Ten (volunteers 1, 2, 3, 4, 6, 7, 8, 9, 10, 11) out of twelve respondents generated various criteria which can all be related to the *problem-area* class in the ontology (cf. section 5.5.2).

Similarly to our definition of *problem-areas* (which span from a simple and arbitrary agglomerate of problems to the more institutional fields of study), also the experts’ results tended to group together quite different things under the ‘subject area’ criterion. For instance, the ‘*meaning problem*’ and ‘*mathematics*’ are equally treated as problem areas (cf. table A4-1). But unfortunately respondents never produced an in-depth classification so to help disambiguate the relationships between, for example, the previous two entities.

Moreover, by looking at the sorts grouped under the meta-criterion 1 (‘*type of entities*’, table A3-1) we can see how respondents have been referring to subject-areas also with ‘*sub-disciplines of philosophy, topics*’, ‘*domains of enquiry*’, ‘*broad areas of enquiry*’, ‘*disciplines*’ and ‘*groups of problems*’.

In general, we believe there is enough evidence for saying that ‘*subject areas*’ can be defined *at least* in terms of the problems they group together - in fact

they are often described as the areas where the ‘enquiry’ or the ‘investigation’ takes place.

This characterization complies with the one we proposed in section 5.5.2. However we cannot derive any conclusion on whether a ‘discipline’ is just that, or also something else (for example, the underlying theories which are defining its boundaries, or its established methods of enquiry). Notably, none of the respondents used the term ‘field of study’.

Hence, at this stage we cannot properly evaluate the entire set of formal specifications presented in section 5.5.2.

7.6.8 Overlap between problems and subject areas

Another interesting result is that in general respondents tended to mix and put together ‘problems’ and ‘problem areas’ by co-locating them within the same criterion (at times also calling them ‘*themes*’, ‘*topics*’, ‘*subject areas*’ and ‘*disciplines*’ - cf. tables A4-1 and A4-4). This differs from our model, where the two entities, although related, are also quite distinct in their meaning.

We believe that such results are motivated by the fact that the cards only presented a small variety of problems, and never in the form of a specific question (which is probably the most typical form a problem takes, in philosophy). Therefore it was much easier to misinterpret problems as umbrella terms referring to sets of interrelated topics (e.g., ‘*meaning*’ problem intended as an area of enquiry). Or, to treat certain areas of enquiry (e.g., ‘*mathematics*’) as the problems which are investigated.

Whatever the reason, for the moment this remains an open issue that will have to be further investigated. At this stage we have to acknowledge though that there seems to be a blurry borderline between the two entities.

7.6.9 A people-centric world-view

This thesis sums up the respondents' tendency to identify things (= ideas) through associating them to their authors, or, in general, to some of the people who are related to them. This type of constructs have been created by six volunteers (2, 6, 7, 10, 11, 12) out of twelve.

This phenomenon emerged mainly in the context of the criteria within meta-criterion 5 (*'correlation to a specific author'*, table A4-5). These criteria do not find a direct translation into the ontological classes depicted in figure 7-2, since they rely on a transversal and often blurry connection between a person (of group of persons) and one or more ideas.

However, our ontology does provide support for defining such correlations elsewhere, that is, by means of the `interpretation` class, and more precisely of the `propositional-content interpretation`. For example, by using these classes (which are subtypes of `event`) it is possible to say of any idea that it is *typical of* an author, it is *used by* him/her or it is *studied by* him/her (for a complete list of these properties, cf. section 5.3.5).

In other words, the 'correlation' can be easily defined by means of an event that links actors and abstract contents via a specific set of properties.

Nonetheless, we must acknowledge what has emerged in the experiment. Namely, the fact that users will be likely to find as familiar and easy to understand an organizing principle of philosophical ideas which is based on the simple '*correlation to a specific author*'. This may contrast with the event-based design used in the ontology, which can be quite complex at times. In the future, we intend to investigate more this issue by looking at one of the following two options:

- a) simplifying the ontological representations by means of a more direct connection between people and ideas (i.e. without having to instantiate events);
- b) drawing specific requirements for the user interface of our ontology-based system, so that it would hide the 'confusing' and unnatural details from the users, while in the background it would keep taking advantage of a highly structured event-based representation mechanism.

Please notice that also the tool's evaluation confirmed the thesis just mentioned. As it will be discussed later (section 8.5), when searching for an item such as 'Frege's philosophy' through the PhiloSurfical interface, users tended to choose the pathway starting from the person (e.g., searching for 'Frege' first, secondly then for his intellectual products), instead of the one departing from the ideas (e.g., searching directly for the available 'philosophies').

7.6.10 Different types of philosophical theories

Analogously to what happened with problems (section 7.6.6), with the criteria in the group '*type of theoretical approach*' (table A4-6) respondents always attempted to refer to the *contents* of a theoretical approach, rather than at its *structural* or *logical* properties. For example, they generated constructs such as '*idealistic approaches*' or '*internalist positions*' (volunteer 5). On the contrary, in the ontology the branch starting from `view` classifies entities according to their generality and exhaustivity (e.g., `thesis`, `school-of-thought`, `theory` etc.).

It is instead possible to find some more useful hints in this direction by looking at the categories emerged in the context of meta-criterion 1 (table A4-3).

In general, from these results we can say that respondents agreed about the existence of a double layer of abstraction for this category. According to six respondents (out of the eight emerged as relevant in the context of meta-criterion 1), we can have the more abstract '*groups of approaches*' (also called '*schools of thought*' and '*doctrines*') and the more specific '*theoretical approaches*' which are used for solving a problem.

Usually, the second category is associated to a single author's philosophy. Only in one case (cf. volunteer 6) a triadic categorization was made: the '*philosophical position*' (i.e. what answers the problem) was differentiated from both the '*philosophies of an author*' and the more abstract '*doctrines*'. The new category in this case is the middle one, which refers to a group of '*philosophical positions*', usually constituting a coherent system of tenets.

In conclusion, these results fit nicely with our ontology but unfortunately they represent only a subset of the categories we are using.

7.6.11 Methods

As discussed in section 7.5.3, meta-criterion 9 (*'type of method'*, table A4-9) gathers only one *verbatim* criterion which is not of much help for discussing how it could map to our ontology. It is instead more useful to consider the results within meta-criterion 1 (table A4-3): here we can find that six out of eight respondents mentioned *'methods'*, which are also called *'conceptual tools'*, *'tools'* and *'techniques'*.

Consequently, we can justify the existence of a `method` class in the ontology, but we do not possess any evidence for validating any of its subclasses. To this end, a more focused experimentation is needed.

Moreover, if we attempt to apply the conclusions described in section 7.1.4.9 with respect to the *'method'* entity, a problem may arise. In fact, following the thesis that all ideas can have a double-sided interpretation (i.e. a *historical* versus a *timeless* one), we should be able to replicate this phenomenon also with methods, for example, by pointing at the space/time instantiation of a *'timeless'* method.

Nonetheless, it is not easy to determine what this instantiation would be, for the respondents never suggested categories such as *'historical'* methods versus *'non-historical'* ones (on the contrary, they did mention the existence of *'historical'* versus *'a-historical'* problems).

Unfortunately, given the scarce evidence we gathered in this respect, the problem remains open.

7.6.12 Importance of ‘what is important’

This final theme is related to meta-criterion 7 (*‘importance/pedagogical perspective’*, table A4-7) and 8 (*‘things I like/know’*, table A4-8). In total, four out of twelve volunteers produced this type of constructs (volunteers 1, 3, 11, 12).

This theme reflects the relevance respondents put on categories which, in general, could be characterized by their being more or less *subjective*. Precisely, such categories are used to represent a scale of values which is accepted by a social group (i.e. with meta-criterion 7), or they can just represent some personal preferences (i.e. with meta-criterion 8).

In the first case, the *importance* criterion is something we have not taken into consideration when building the ontology because we thought it would have been dealt with at the ‘learning narratives’ level. In other words, since a ‘pedagogical organization’ is a meta-description of the entities in the philosophical domain (more precisely, it is a description of how they can be sequenced in order to fulfill a certain pedagogical design) we reckoned it would have been adequately represented by using a knowledge model (or any other similar functionality) which is *external* to the philosophical ontology.

The second case, instead, refers to criteria which are sound only *ad personam*: also this time there is no correlative in the ontology, simply because such

criteria go against the intended purpose of the ontology (i.e. modeling the important 'objective' dimensions of the philosophical domain).

Nonetheless, it is important to emphasize that it is possible to provide some generic mechanisms for letting users *review* or, in general, express preferences towards some philosophical contents; but again, this is a functionality lying on a plane which is orthogonal to the one of representing the philosophical domain.

7.7 Conclusions and future work

As the main purpose of this evaluation experiment was to examine how experts' categorizations could evaluate our ontological categories, in the following table we are providing a short summary of these results. In particular, in the table we are referring to the *correctness* criterion described at the beginning of this chapter. As already mentioned, we are focusing on the `philosophical-idea` branch of the ontology, because that is where our model stands out for its novelty.

Instead, for that regards the *completeness* criterion, it has been shown in the previous sections how it is possible to express all of the volunteers' constructs by means of our ontology. Therefore, in the light of this initial evaluation experiment, we can say that the ontology is complete.

In the first column of table 7-10 it is possible to see the name of the class we wanted to evaluate, together with the details of its subclasses. In the second

and third columns we indicate whether the results are positive or negative, and the reasons for this decision.

Ontology Class	evaluation	rationale
METHOD PRACTICAL-METHOD SCIENTIFIC-METHOD PRECEPT ABSTRACT-METHOD ARGUMENTATIVE-METHOD RULE-OF-INFERENCE LOGICAL-MATHEMATICAL-METHOD METHOD	<i>positive</i>	We gathered evidence about ‘methods’ with meta-criterion 1 and 9. However, we do not have any evidence about the correctness of the subclasses of method.
PROBLEM FACTUAL-PROBLEM MODALITY-PROBLEM RELATIONAL-PROBLEM FUNCTIONAL-PROBLEM DEFINITORY-PROBLEM COMPOSITION-PROBLEM EXISTENCE-PROBLEM	<i>positive</i>	Various people used the ‘problem’ category, especially within meta-criterion 1 and 4. However, we do not have any evidence about the correctness of the subclasses of ‘problem’.
PROBLEM-AREA GENERIC-FIELD-OF-STUDY SCIENTIFIC-DISCIPLINE HUMANISTIC-DISCIPLINE FIELD-OF-STUDY	<i>positive</i>	We gathered evidence for the ‘problem-area’ category, especially within meta-criterion 1 and 3. However, we do not have any evidence about the correctness of its subclasses.
CONCEPT	<i>neutral</i>	Some of the cards could have been modeled as ‘concepts’. But the respondents used this category rarely and inconsistently.
VIEW SCHOOL-OF-THOUGHT CONTEXTUALIZED-SCHOOL-OF-THOUGHT PHILOSOPHICAL-SYSTEM THEORY SCIENTIFIC-THEORY PHILOSOPHICAL-THEORY THESIS PRINCIPLE SELF-EVIDENT-PRINCIPLE LAW	<i>positive</i>	We gathered evidence for the ‘view’ category, although respondents did not call it so. This happened within meta-criterion 1 and 6. Also three of its subclasses have been highlighted (‘school of thought’, ‘philosophical system’ and ‘theory’).

Table 7-10. Ontological classes matching the evaluation results

As we can see, 4 out of 8 *direct* subclasses of `philosophical-idea`, according to our analysis, have a direct match with the criteria used by respondents for organizing the cards.

As for the others, in most cases we did not gather enough evidence so to draw any conclusion. We believe that this is not a shortcoming of the methodology but just a limitation imposed by the cards' scope and variety. For example, if we performed another experiment using cards carrying mostly names of philosophical problems or arguments, we will likely get some hints about the categories experts use for *those* entities.

In section 7.7.3 we propose some research directions to tackle this issue.

7.7.1 Attempt at a synthesis of respondents' implicit mental models

Finally, as a further result of the discussion of section 7.6, we would like to present the reader with a more *systematic* model of respondents' results. This model has been created as an attempt to combine all the categories emerged during the experiment into a more comprehensive schema (see figure 7-3).

Let us emphasize that this is only *our interpretation* of the results. As such, the model below can also be described as a *visual* summary of what we thought were the respondents' implicit viewpoints when they ordered the cards. In general, the key features emerging from this analysis are the following ones:

- 1) the *explananda/explanations* dichotomy;
- 2) the triadic division within both the *explananda* and the *explanations*;
- 3) the central positioning of *methods*;

- 4) the vertical dimension of space/time specificity, which applies to all ideas;
- 5) the horizontal dimension of general/specific, which applies to the various types of *explananda* and *explanations*.

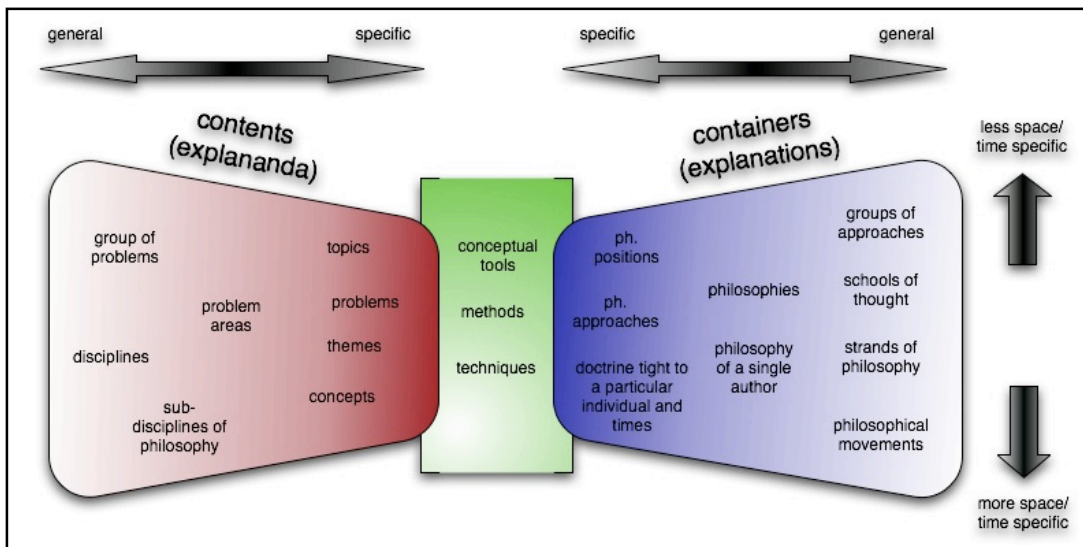


Figure 7-3. Schema showing our synthesis of the respondents' categories

7.7.3 Future work

An implicit risk associated to the type experiment we conducted comes from the fact that, given the vastness and novelty of the domain, we may have chosen the wrong methodology or taken too broad a scope for the investigation.

As a first step towards the solution of these and related issues we will now enumerate a series of alternative research directions. In general, we think that these approaches could complete and possibly further extend our exploratory results.

A first complementary research experiment would entail running other sorting sessions, but using cards which are specific to a particular philosophical topic only. For example, it would be interesting to investigate:

1. How experts differentiate among views of different granularity (e.g., schools of thought, theories or philosophies). For example, from our investigations it often appeared that we do not have only one class categorizing the so-called -isms (e.g., rationalism, monism etc.), but instead a hierarchy of theory-types comprising -isms of varying granularity (e.g., rationalism is sometimes considered a type of monism, etc.). How do experts proceed in order to make sense and manipulate these hierarchies of ideas? We believe that the use of a more fine-grained set of cards could shed light on this issue.

2. How experts would organize topics, problems and disciplines into a consistent representation; we have seen from our previous results that these entities are sometimes used interchangeably, but we did not have enough evidence in order to individuate which are the dimensions that make them overlap so easily. Also, another aspect of this problem that could be investigated further regards the clarification of the relationship between disciplines (i.e. subject areas, fields of study) and views.

3. What typologies of methods do exist, and how they relate to views: do they derive from them? Or are they used by them? Or both? Also, it is interesting to investigate further the relationships between methods and problems: e.g., according to the experts, do methods solve problems, or do views solve problems?

4. The effective role of what, in the ontology, we defined as 'concepts' (that is, the atomic elements defined and used by theories). This could be achieved by using a set of cards representing entities related to one single philosophy (or theory) only, so to derive categorizations highlighting the structural role of the various elements (and among them, the concepts) within a more circumscribed context.

Secondly, we also think it would be interesting to experiment with other knowledge acquisition methodologies, such as:

5. Laddering (Corbridge et al., 1994, Rugg and McGeorge, 1995). This technique aims at the creation of a hierarchical structure of the domain, by asking questions that force respondents to specify the position of the categories on a hierarchy. Similarly to the sorting technique, also laddering can be implemented using cards representing domain entities. The results could be a concept ladder, which is a tree diagram representing the different levels of abstraction of the knowledge items; or also an attribute ladder, which is instead obtained by asking respondents to classify the entities also with respect to the properties they have. We believe that the laddering technique could be very useful especially for investigating the hierarchies of views which exist in the philosophical domain.

6. Repertory grid analysis (Rugg and Shadbolt, 1991, Shaw, 1980). For this method, the domain expert is usually presented with a matrix representing concepts versus attributes; the task involves the rating of each concept against each attribute, using values from a continuous

scale (e.g., a numerical scale). The ratings can then be examined using a statistical calculation (cluster analysis) so to group the concepts and attributes having a similar score. We believe that this technique could be useful for the analysis of several types of philosophical entities (in primis, views, problem areas and methods). However, repertory grid analysis can be used only at an advanced stage of investigation; mainly, because it is necessary to elicit an adequate number of valid concepts and attributes to put in the grid. This could be done by using the other KE methods previously described.

7.8 Summary

In this chapter we presented and discussed an evaluation of several key-classes of the PhiloSurgical ontology. Precisely, we aimed at providing an empirical validation of the classes representing the various types of philosophical ideas. The evaluation design is the following: by means of a knowledge elicitation experiment, we gathered data on how philosophy experts tend to categorize their domain; subsequently, we analyzed such results and compared them to the ontological model of our creation. In particular, we used a card-sorting experiment which involved twelve philosophy experts; in order to gather the data, we asked them to arrange a set of philosophical concepts into meaningful groups, by highlighting their commonalities and differences. The results partly support our ontological choices, partly push for further and more specific experimentation that we plan to carry on in future research.