

Symbol Grounding for the Semantic Web

Anne M. Cregan^{1,2}

¹ National ICT Australia (NICTA)

² CSE, University of New South Wales, Australia

Anne.Cregan@nicta.com.au

Abstract. A true semantic web of data requires dynamic, real-time interoperability between disparate data sources, developed by different organizations in different ways, each for their own specific purposes. Ontology languages provide a means to relate data items to each other in logically well-defined ways, producing complex logical structures with an underlying formal semantics. Whilst these structures have a logical formal semantics, they lack a pragmatic semantics linking them in a systematic and unambiguous way to the real world entities they represent. Thus they are intricate "castles in the air", which may certainly have pathways built to link them together, but lack the solid foundations required for robust real-time dynamic interoperability between structures not mapped to each other in the design stage. Current ontology interoperability strategies lack such a meaning-based arbitrator, and depend instead on human mediation or heuristic approaches. This paper introduces the symbol grounding problem, explains its relevance for the Semantic Web, illustrates how inappropriate correspondence between symbol and referent can result in logically valid but meaningless inferences, examines some of the shortcomings of the current approach in dealing effectively at the level of meaning, and concludes with some ideas for identifying effective grounding strategies.

Keywords: Ontology Alignment, Semantic Interoperability, Semantic Web, Symbol Grounding.

1 Introduction

The purpose of the World Wide Web is to share and leverage information. But information is only ultimately useful if it produces some result in the real world, either in the physical environment, or in someone's state of understanding. Raw unprocessed data is not very helpful in this regard, as it requires significant human effort, and the application of implicit human knowledge to understand it and process it appropriately to produce tangible benefits. It is generally agreed that machines should be doing more of the work of turning data into knowledge in a way that supports the production of results for human benefit. The purpose of the Semantic Web is to address this; its stated objective being to make information more easily shared and applied, by making its meaning explicit [1]. The implicit assumption is that once meaning is represented explicitly, machines will be able to align and process data according to its meaning, thus turning it into knowledge, and supporting web services and intelligent agents to produce real-world results on our behalf.

However, this implicit assumption has not yet been thoroughly investigated. To date, information processing has been based on a symbolic processing paradigm, and to process information at a semantic level requires a fundamental paradigm shift. New methodologies, processes, and criteria for judging success are needed. Many of the techniques for aligning or reconciling meaning are already known from programming, but not at a mature level where meaning is made explicit and machine processing does the rest: it requires a human being to analyze the meaning and devise and implement appropriate code to do the necessary transformations.

How are we to start making inroads into this new semantic territory? As an initial step, taking a good look at the really hard questions should help focus the effort, and provide foundations for this new information processing paradigm.

These hard questions include but are not limited to the following:

1. What is meaning?
2. What do we need to do to make meaning explicit?
3. What is the appropriate way to process meaning?
4. How can we judge whether we have been successful in representing and processing meaning at a semantic level?
5. Will the current Semantic Web approach, based on the Web Ontology Language OWL [10], produce the right kind of representations, and support the right kinds of processes, to achieve the results being sought, or is a key component of the solution missing?

Spanning from the very philosophical to the very practical is necessary because the issue of meaning is a fundamental philosophical issue, whilst the goals of the Semantic Web are very practical. By their very nature as a “specification of a conceptualization” [3], creating ontologies involves bridging between the realm of IT/Engineering, and the realm of Cognitive Science/Philosophy. It is hoped that such an investigation can uncover the foundations for such a bridge, providing a basis not only for the Semantic Web but for the Pragmatic Web it will ultimately support.

Organization

The paper is organized as follows:

- **Section 1** introduces the challenge being undertaken
- **Section 2** relates meaning to both entailment and designation, looks at symbolization and introduces the symbol grounding problem
- **Section 3** explains why symbol grounding is relevant for the Semantic Web in its aim to achieve dynamic real time interoperability, and extensional approaches and URIs are not sufficient in themselves to provide adequate symbol grounding.
- **Section 4** considers next steps in identifying suitable symbol grounding strategies for the Semantic Web and concludes.

2 Meaning and Symbol Grounding

What is meaning? The greatest philosophers and thinkers have considered this question for the last several thousand years, but as yet there seems to be no definitive

answer. What are the implications for the Semantic Web, which is being built around the keystone of making meaning explicit and machine-processable? Is it ever really going to get off the ground, or perhaps do so initially but quickly collapse under its own weight for lack of good foundations? It seems somewhat foolhardy to attempt to devise explicit well-defined procedures for operating at the level of meaning, without attempting to lay good foundations by stating what meaning is taken to be.

Whilst a conclusive answer to the question is unlikely (isn't that what makes a good philosophical question after all?) and Semantic Web researchers are, generally speaking, practical people who want results in reasonable timeframes and certainly don't want to get bogged down in the vagaries of philosophy, I believe that as part of the construction of Semantic Web technologies, for purely practical reasons, there should be some attempt to state what we take meaning to be for the purposes of the Semantic Web.

A clear conception of meaning for the purposes of the Semantic Web should, at the very least, assist researchers in devising appropriate and precise procedures and methods for making meaning explicit, which then has the flow-on effect of supporting practitioners to build appropriate semantic models representing their respective domains, and will make such models better suited for interoperability. It also provides a theoretical basis for semantically processing the information captured by such models.

Whilst many modeling errors have been identified and are well understood e.g. [8], there is still quite a spectrum of "correct" models available for modeling any given domain. The ontology builder has considerable discretion in make design choices. Are some of the resulting models better than others? Intuitively the answer is yes, and depends on the intended function of the ontology. However, we are still seeking a more precise understanding of the nature of this dependence, and at the moment there is no one clear guiding methodology for building domain models. Whilst there are obviously several factors at play, the model's effectiveness in making meaning explicit should certainly be considered a key criterion.

Beyond this, an analysis of meaning also offers insights into the overall Semantic Web approach and whether it will ultimately be able to deliver on its promises. Capturing meaning is clearly a fundamental component, but are the current suite of Semantic Web standards and technologies adequate to the task of capturing machine-processable meaning to produce the outcomes being sought, or will they ultimately fall short? If we want to ultimately build a "Pragmatic Web", that delivers tangible real-world benefits, we need to make sure the foundations are firm enough to support this.

2.1 What Is Meaning?

Without getting too bogged down in philosophy, let's take a practical approach to home in on what meaning is, by identifying what it is that we really want when we ask the meaning of something. In everyday life, we generally don't ask the meaning of concrete things like a chair, or a train, or a person, or a pet. Such things have no meaning: they just are. We ask the meaning of actions and events, policies and such like, in which case we are generally trying to identify the relevant entailments, **or** we ask the meaning of symbols, in which case we want to know what they designate, or stand for. When we ask about meaning, we are usually asking for one of two things: either for entailment, or for designation.

2.2 Entailment and Designation

Entailment: What are the logical consequences of some action, event or state?

Examples:

- *If I take this promotion does it mean I will be able to afford the house?*
- *If Serena wins this point, does that mean she wins the match?*
- *If my business is registered as a public company, does that mean we are required to have annual audits conducted?*

Designation: What is being referred to? What does the symbol symbolize?

Examples:

- *What's the meaning of "verisimilarity"?*
- *What does that sign mean?*
- *What do you mean by giving me that wink?*
- *What does the green line on the graph mean?*

Designation gives the referent being represented by some kind of symbol: a word, a street sign, a gesture, a line on a graph. It uses symbols to point to something; a convenience originating from the need to identify and communicate something that does not have a local physical existence, is abstract, or is an internal state and not directly accessible. Designation is the back end of symbolization: it establishes the referent, or what the symbols symbolize.

Symbolization

Note that there are (at least) two related senses of symbolization. In the first sense, a recognizable concrete thing is used to stand for a more abstract intangible thing e.g. a dove is used to symbolize peace. It is usually chosen as a symbol because it has some kind of real-world historical or mythological relationship with the abstract thing, or evokes it through some other kind mental or perceptual association, or it can simply be a matter of convention. In this sense, a non-verbal meaning relation is pre-established and the symbolization makes use of it to evoke the intended referent. This should not be confused with symbolization as used in this setting, which is being referred to as "designation" for the purposes of clarity within this paper.

In our setting, symbolization refers to the scenario where a mark, character, sound, avatar or some such arbitrary thing is used to designate some physical or conceptual thing. In this case, the symbol is an arbitrary physical token, designed by humans specifically for the purpose of representation, and does not usually have a meaning in and of itself (Although in the case of avatars, some recognizable topographical resemblance may exist, and thus their form may be argued not to be completely arbitrary). In this kind of symbolization (designation), the essential question is how the relationship between an arbitrary symbol and its intended referent is to be established. This question has been identified in Artificial Intelligence Research as the "Symbol Grounding Problem".

2.3 Denotation and Connotation

Designation itself has two aspects: denotation and connotation, a distinction introduced by J.S. Mill [5]. To illustrate by example, the denotation of a term such as ‘woman’ refers to all the individuals to which may correctly be applied, whilst the connotation consists of the attributes by which the term is defined e.g. being human, adult and female. Connotation determines denotation, and in J.S. Mill, is taken to be meaning, whereas terms like proper names e.g. ‘Mary’, which have denotation if there is someone so called, are taken to lack meaning as they have no connotation, as no attributes define ‘Mary’.

2.4 Relevance to the Semantic Web

Both entailment and designation have relevance for the Semantic Web: entailment relating to what can be concluded from what is already known, and designation relates to establishing the connection between symbols in a formal system and what they represent. There is already a very significant body of work around entailment for the Semantic Web [10], based on description logics providing an underlying formal semantics for the various flavours of OWL.

However, designation has had less attention to date. OWL’s formal semantics have a set-theoretic basis, where a set (‘concept’ or ‘class’ in DLs) is essentially defined by its extension - clearly a denotational approach. However, meaning based on denotation is less than adequate for the needs of the Semantic Web, as will be explained below. The consequence is that the entailment parts of the Semantic Web have no theoretical basis for anchoring to anything in the real world, and are thus floating castles in the air.

To explain: an OWL ontology is made up of a set of logical axioms, themselves composed of primitive objects, predicates and operators, combined via formation rules into well-formed formulae. Unless some kind of faithful and appropriate correspondence is established between the primitives and whatever they are intended to represent outside the formal logical system, any entailment produced by the system will not result in reliable conclusions that correspond to the actual state of affairs in the real-world domain of interest. Establishing a correspondence between the primitives (which are effectively just symbols or symbol strings once they are inside the logical system), and the domain is an extra-logical consideration. The question of how the relationship between the symbol and the referent is to be established has been identified in Artificial Intelligence Research as the “Symbol Grounding Problem”.

2.5 The Symbol Grounding Problem

The Symbol Grounding Problem, as described, for instance, by Harnad [4] relates to the inadequacy of defining symbols using only other symbols, as is commonly done in a dictionary or a formal logical system. In his exposition, Harnad takes Searle’s [9] famous Chinese Room scenario, originally used by Searle to illustrate the difference between mechanical symbol manipulation, which merely simulates mind, and a true understanding of intrinsic meaning, which necessarily involves processing at the semantic level.

The scenario involves a machine hidden inside a room, which is given a set of Chinese language inputs and produces a set of Chinese language outputs. Searle points out that a machine using only symbolic manipulation to match a list of pre-defined inputs with a list of pre-defined outputs may be capable of simulating conversation with a Chinese speaker well enough to pass the Turing test. However, Searle argues, such a machine cannot be said to understand Chinese in any sense, any more than a human who uses such a list to produce statements in Chinese can be said to understand Chinese. Searle ultimately concludes that meaning is in the head, not in the symbols, and furthermore that cognition cannot be just symbol manipulation, as it clearly requires some activity to take place at the semantic level.

Harnad puts an alternate spin on Searle's Chinese Room scenario, asking the reader to imagine having to learn Chinese as a second language, where the only source of information available is a Chinese/Chinese dictionary. He observes that *"The trip through the dictionary would amount to a merry-go-round, passing endlessly from one meaningless symbol or symbol-string (the definiens) to another (the definienda), never coming to a halt on what anything meant."* He then presents a second variant, where one has to learn Chinese as a first language, and again the only source of information available is a Chinese/Chinese dictionary. He argues that if the first variant is difficult, then the second must be impossible, relating it to the task faced by a purely symbolic model of the mind, and asking *"How can you ever get off the symbol/symbol merry-go-round? How is symbol meaning to be grounded in something other than just more meaningless symbols? This is the symbol grounding problem."*

How indeed, are we to get off the Symbol/Symbol merry-go-round? Firstly though, let us consider in detail how the symbol grounding problem is relevant for the Semantic Web.

3 Why the Semantic Web Needs Symbol Grounding

The ultimate vision of the Semantic Web is a web of data connected by meaning which is machine processable. The idea is to get meaning out of the technologists and domain expert's heads, and into some explicit, machine processable representation which defines how to link it up appropriately, in real-time, without reference to human mediators. But the current Semantic Web building blocks are a long way from achieving this vision. Let's take a look at why this is.

In building an ontology, the designer chooses terms for classes, instances and properties, and builds axioms/structure linking them. The terms are usually chosen for their meaning in some natural or domain-specific language. Additional annotations may explain the meaning of the term, using more natural and/or domain specific language. But natural language is notoriously ambiguous and slippery. Its symbols/semantic units are imperfectly grounded, as we will explain in the following section. And whilst domain specific-terminology may be unambiguous within the domain, it is not necessarily unambiguous when linking across domains.

If the basic terms used for ontologies are ambiguous, then having a well-defined structure that supports entailment is of dubious benefit. The structure by itself is not the meaning: as discussed, meaning requires both logical structure for the purposes of entailment, and grounding for the purpose of establishing correspondence between the domain and the logical structure. Only then can entailments made by virtue of the logical structure be guaranteed to be an accurate reflection of the real-world state. Garbage in, garbage out, as the old saying goes.

3.1 Meaningfulness

As an example of this principle, there is a considerable body of work e.g. [6] in Mathematical Psychology around determining which kinds of variables can be subjected to which kinds of mathematical operations, in order to produce only meaningful results and avoid meaningless conclusions. Considerable effort has gone into investigating “meaningfulness” to avoid the inappropriate use of statistics. In a classic example, the school football team are assigned numbers to wear on their football jerseys. This numerical assignment is simply to give each football player a unique label for the purpose of identification. However, this assignment does not support taking the average of those numbers, and asserting that this average reflects some meaningful property of the football team. This is because the numbers have no numerical properties attached to them - they are just labels, and could equally well be any other arbitrary symbol (letters of the alphabet, pictures of animals) - the only important factor is that each player is designated by a unique symbol. The underlying variable being represented (identity) is not a quantitative variable, so any mathematical inferences derived from the football jersey numbers are simply meaningless. This is not the case for a quantitative variable like the heights of the football players, where it is perfectly appropriate to represent heights as real numbers and calculate average and standard deviations in the height of this population.

Note that the “meaningfulness” criteria is not necessary because of any problem to do with numbers themselves, or with mathematical reasoning. The problem is that the real-world dimension being represented does not have the same properties as the chosen representation. The representation is richer and more structured than what is being represented, and thus permits reasoning and inferences which have no correspondence with the real-world. Inferences which are perfectly valid inside the representation symbol are thus meaningless when we attempt to map them back to the domain of reference. As a formal logical system without appropriate grounding strategies to connect it to the real-world, the Semantic Web faces a similar problem.

3.2 Semantic Interoperability Problems

Pollock and Hodgson’s analysis of types of semantic conflicts [7] identified eleven kinds of semantic level clashes: DataType, Labeling, Aggregation, Generalization, Value Representation, Impedance Mismatch, Naming, Scaling & Unit, Confounding, Domain, and Integrity. This analysis has been adapted and re-organized to fit the Semantic Web and the focus of the current investigation.

Semantic Conflict:	Manifests As:	Example:
Terminology	The same term is used to mean different things (homonyms) or different terms are used to mean the same thing (synonyms).	- "mouse" as a hardware peripheral vs a rodent - "Holiday" vs "Vacation" : different terms, same meaning
Representation: Instance Level	The same information is being referred to at the meaning level but is being represented differently.	- Fahrenheit vs Celsius temperature scales.
Representation: Concept Level	Concepts have been abstracted differently	- StartTime and Duration vs StartTime and EndTime
Representation: Structural Level	Different choices about the division of the domain into instances, classes and properties, and/or different choice of axioms	- Modeling a Person's educational institution as an ObjectProperty connected to a Class, or as a DataProperty connected to string values.
Representation: Superstructure Level	Different modeling constructs or paradigms used, based on fundamentally different representation methodologies.	- Relational DB vs Object-Oriented (Impedance Mismatch) - Entity-Relationship model vs first order logic model
Granularity	The same information is represented at different levels of granularity.	- Daily Sales vs Monthly Sales - Temperature information: 39.3 degrees Celsius vs "hot"

Semantic Conflict:	Manifests As:	Example:
Perspective/ Context	The information may be from the point of view of a particular part of the supply chain, a particular business process or application and does not apply universally.	- Whilst related, "Cost" from a supplier's point of view is the cost of production, whilst to the consumer it is the cost of purchasing the finished product.
Underlying conceptualization	The information may be based on a different kind of conceptualization, theory or ideology	- Linnaeism vs Cladism: different criteria for classification of classes of animals - A commonsense classification vs a technical one (eg tomato as a vegetable vs a fruit)
Origin	Whilst the information is ostensibly the same, the purpose the data was collected for, or the way it was collected is different, creating bias.	- Information about income collected for tax purposes vs collected in a credit application.

As semantic conflicts, every one of these problems requires reference to the meaning level of the information for its resolution. The following table looks at the resolution method for each class of semantic conflict and indicates how Symbol Grounding and Meaning is relevant in each case.

Semantic Conflict:	Resolution Method:	How Symbol Grounding / Meaning is relevant:
Terminology	Merge or align the two if the two terms have the same meaning Separate or treat separately if the two terms have distinct meanings.	Joining or separating based on meaning is determined by the identity or divergence of the real-world entity the terminology designates ie what the symbol is symbolizing.
Representation	Transformation of content or structure, or mapping relations based on meaning	Need to determine real-world relationships between entities being represented to identify equivalences if any and determine which transformations would be valid.
Granularity	Identify how the two fit together and create some transformation and/or mapping of values if possible	Need to know what the underlying real-world dimension is? What is the base/lowest level of granularity? If aggregated, from what base and by what criteria?

Semantic Conflict:	Resolution Method:	How Symbol Grounding / Meaning is relevant:
Perspective/ Context	Make perspective explicit and align into a superstructure	Meaning is relative to perspective and context but how does this affect it? How do we specify the context and how it affects the meaning?
Underlying conceptualization	Make underlying assumptions explicit and align into superstructure that states these explicitly	What are the underlying assumptions and methodology? How do we represent these?
Origin	Make underlying assumptions explicit and align into superstructure that states these explicitly	What is the source and how does it affect the data? How do we represent this and adjust for it?

3.3 Current Support for Semantic Interoperability Conflict Resolution

As the analysis of the previous section shows, symbol grounding and meaning is at the heart of these interoperability problems. Resolving these kinds of problems are common occurrences in mapping and aligning ontologies as they exist today. Whilst

tools and heuristics are available to humans to assist the process, it is a problem essentially being addressed by human beings, rather than machines. Somehow, the underlying meaning needed to resolve these interoperability problems is not being explicitly represented in the Semantic Web, and/or there are not sufficient tools and techniques available for resolving it through automated processing. The remainder of this section makes an initial pass at identifying where the shortcomings are, starting with specifics of the OWL language and broadening out from there.

Mapping Constructs Provided by OWL

OWL provides only very limited constructs for mapping ontologies, and essentially none for transformations. The OWL language has four constructs specifically for use in mapping ontologies for the purposes of merging. These constructs are intended for use when two or more ontologies have been built independently of each other and later need to be merged or linked, i.e. they are being mapped after the initial design phase has taken place. They are:

<code>owl:equivalentClass</code>	asserts that two or more classes have exactly the same instances
<code>owl:equivalentProperty</code>	asserts that two or more properties are equivalent
<code>owl:SameIndividual</code>	asserts that two or more individuals are identical
<code>owl:DifferentIndividuals</code>	asserts that two or more individuals are identical

Whilst these constructs provide the means to specify that two or more classes, properties or individuals are equivalent / identical, or that two or more individuals are different, they are only useful once the equivalence, identity or difference is determined. This determination is outside of OWL, and the Semantic Web technologies do not provide any formal basis for a machine to determine this without human guidance (albeit supported by tools and heuristics). The use of class extensions and URIs is insufficient, as explained below.

Extensive vs Intensive Class Definitions

Some may argue that if two classes contain the same set of individuals, they must be the same, and machine processing can determine this. However, having the same extension does not necessarily prove that two classes are the same: they may happen to contain the same individuals, but have different intensive meanings: that is, the criteria for membership of the class is different. Philosophically, this is the denotation vs connotation distinction.

For example, the extension of the members of the school basketball team in a Sports Ontology and the extension of the school's Grade A students in an Academic Ontology may conceivably, at some point in time, consist of exactly the same set of individuals, and machine processing may determine on the basis of these equal extensions that the two classes must be equivalent, and map them using `owl:equivalentClass`.

The possible ramifications of such an ill-advised mapping are obvious: if a student drops his grade, he will find he is no longer classified as a member of the basketball team. If she drops out of the basketball team, she will no longer be classified as a Grade A student. If a new student is added to the class of Grade A students, he will find himself automatically in the basketball team. A new member of the basketball team will find she is automatically classified as a Grade A student.

Clearly, the extensive approach to class definition and mapping is inadequate, especially when changes in the variables used for classification occur, or new, unclassified instances are encountered, simply because the classification criteria are not adequately captured. A complete specification of meaning needs to support a decision procedure which determines whether a previously unseen instance qualifies for membership of the class or not, by making the membership criteria explicit. We also note that it is the nature of some classes to have fuzzy boundaries [2] (e.g. the colour red), and support may be needed for graded class membership in such cases.

Representing Mappable Differences/Transformations

OWL also lacks the means to specify semantic differences and the transformations needed in a way that support interoperability. Programming code can get around this, but has to be written for each specific situation, straying from the Semantic Web ideal of explicit representation enabling automated processing at the meaning level.

As a simple example, a temperature in Fahrenheit and a temperature in Celsius can easily be converted either way via a simple arithmetic equation. The underlying measurement scales are interoperable, but there is no support for representing such a relationship within an ontology. Assuming that both ontologies map temperature as a `DataProperty` to a real number value, the numerical values of the respective properties differ, but have a well-defined arithmetic conversion. However, OWL provides no mechanism for specifying the scale, the properties of the scale or for doing such a transformation. Meaning that is available to the designers, and could be made explicit in building the respective ontologies, subsequently supporting interoperability via machine processing, is currently not able to be represented within the ontology.

Independent and Dependent Ontologies

The mapping constructs in OWL for mapping independently designed ontologies were considered above. However, because OWL ontologies are built from URIs, the components can reside anywhere. The principle of composability in OWL means that when an ontology is being built or revised, the designer can freely use any constructs from any other available ontology. This results in one ontology having logical dependence on another. Interoperability is an issue for both dependent and independent ontologies, and some of the broader interoperability concerns are addressed below.

Dependent Ontologies: If a base ontology changes, other dependent ontologies are affected. This is potentially much more serious than just a “broken link” because it can potentially change the inferences made. In the World Wide Web, broken links and web page changes are not critical, they are simply a dead-end and one can always look for information elsewhere. However, when an external URI is an essential part of

a logical structure, a change or deletion can have serious real-world consequences, such as an incorrect classification as an illegal alien for example.

Independent Ontologies: In the case where two independent ontologies need to be aligned or mapped and have no common constructs other than the Ontology language itself, currently the options for resolution are either heuristic approaches with varying success rates, or the human designers of the respective ontologies can communicate with each other or check other sources to establish the meanings of terms devise an appropriate mapping. The problem here is not only that this mapping problem is potentially in the order of N^2 to achieve interoperability across the semantic web, but the problem is an N^2 human-to-human problem. The resulting reward to effort ratio causes pause for reflection.

3.4 Why Using URIs Is Not a Sufficient Grounding Strategy

A commonly encountered argument is that Unique Resource Identifiers (URIs) can be used for any disambiguation needed for the semantic web. This is Sir Tim Berners-Lee's own view (conversation with author, November 2006). After all, anything can have a URI, why not simply use that as a unique identifier? If two concepts, individuals or properties have the same URI, they must therefore be the same. Problem solved!

Whilst this approach has its merits, it is not sufficient in itself to resolve all the kinds of semantic interoperability problems. Granted, it can identify cases where two ontologies are referring to the same thing, but it cannot identify what that same thing is that they both refer to, or that either of them are representing or processing it appropriately based on its meaning. This is because the URI does not have a grounding mechanism to connect it to anything outside the information system: A textual description residing at the URI or within the URI itself is natural language and thus subject to ambiguity and vagueness.

The exception, of course, is when the thing being referred to is, exactly, the information that resides at the URI. For instance, an identifier for a particular tax law can be grounded to a URI that contains the exact text of the tax law, and thus there is no need to go further. But if the URI is intended to reference a real world physical thing, like a person or a building, or something else outside the information system itself, it needs a symbol grounding strategy.

A further consideration when considering an information system on the intended scale of the Semantic Web, is who is going to check that every URI maps to one and only one real-world referent? No doubt many things will have more than one URI, in which case we still have the human problem discussed earlier, of determining that the URIs have the same referent and mapping them, which obviously cannot be resolved using the URI itself. And on the flip side, there will no doubt be many real-world things that have no corresponding URI, so the system is incomplete.

4 Next Steps and Conclusions

If Searle is right, cognition cannot be reduced to symbol manipulation. Semantic processing therefore requires an understanding of cognition in regard to meaning.

The next steps underway in this line of research are the investigation of a wide range of grounded symbol systems, such as Musical Notation, Cartography, Chess Notation, Circuit Diagrams, Barcodes and even Knitting Patterns. These are being analyzed to determine the grounding strategies used, and how and why they are effective or ineffective. The analysis will identify the kinds of grounding strategies available, and determine appropriate criteria for assessing them, and it is hoped it will provide a theoretical basis for constructing Symbol Grounding strategies for the Semantic Web will be identified. Following this, the question of devising appropriate processing and procedures to produce meaningful results will be addressed.

In conclusion, this paper has put forward some of the hard questions the semantic Web needs to answer, examined some of the pitfalls that may occur if they are not addressed, and explained the relevance of the symbol grounding problem for the kinds of semantic interoperability issues commonly encountered. Some insights from measurement theory in Mathematical Psychology were briefly covered to illustrate how inappropriate correspondence between symbol and referent can result in logically valid but meaningless inference. Some of the shortcomings of the current Semantic Web technologies in dealing effectively at the level of meaning level were investigated. The arguments that set extensions and URIs can provide an appropriate basis for grounding the Semantic Web were considered and found wanting. Finally, next steps for identifying effective grounding strategies and doing meaning-level processing were briefly discussed.

Acknowledgments. To my infant nephew, Gaelen Raphael Bonenfant, for providing a beautiful example of how the grounding process gets started.

Research reported in this paper has been partially financed by NICTA (<http://www.nicta.com.au>). NICTA is funded by the Australian Government's Department of Communications, Information Technology and the Arts, and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence program. It is supported by its members the Australian National University, University of NSW, ACT Government, NSW Government and affiliate partner University of Sydney.

References

1. Berners-Lee, T., & Fischetti, M. (1999). *Weaving the Web*. San Francisco, Harper.
2. Gärdenfors, P. *Conceptual Spaces*. (2000) MIT Press, Cambridge MA.
3. Gruber, T. R. (1993) A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2), 199-220.
4. Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42: 335-346.
5. Mill, J.S. (1843) *A System of Logic*, London.
6. Narens, L.E (2002) *All you ever wanted to know about meaningfulness*. Volume in the Scientific Psychology Series, Lawrence Erlbaum Associates, Mahwah, NJ.
7. Pollock, J.T. and Hodgson, R. (2004). *Adaptive information: Improving business through semantic interoperability, grid computing, and enterprise integration*. Wiley Series in Systems Engineering Management, John Wiley & Sons, Inc.

8. Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H, Stevens,R., Wang, H. and Wroe, C (2004). OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns. *EKAW2006 Proceedings*, (pp. 63-81). Available at www.co-ode.org/resources/papers/ekaw2004.pdf
9. Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3: 417-457.
10. Smith, M.K., Welty, C., & McGuinness, D.L (eds) (2004). *OWL Web Ontology Language Guide*, W3C Recommendation, 10 February 2004. Available at <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>. Latest version available at <http://www.w3.org/TR/owl-guide/>