Relative Definability in Formal Ontologies

Brandon Bennett*
School of Computing, University of Leeds, Leeds LS2 9JT, England
brandon@comp.leeds.ac.uk

Abstract. The paper investigates technical meta-logical notions and theorems relating to definitions and relative definability and shows how these are relevant to the concerns of knowledge representation. In particular they can be used to identify sets of primitives that are sufficent to fully describe a given domain. Fundamental definability theorems of Tarski are examined, and it is shown how these can adapted to be applicable not only to precise and categorical mathematical theories but also to more loosely formalised ontological theories.

1 Introduction

The purpose of the current paper is to examine in some detail the notion of *relative definability* and explore its relevance to the construction of formal ontological theories. Definitional languages, such as description logics [1, 2] play a very significant role in the construction of many ontologies. Nevetheless, the theory of definability itself, seems to have been seldom applied to ontology design.

Though there is a considerable mathematical literature on definability, most of it is concerned with what concepts, functions and relations are definable in a given logical language (such as 1st-order logic). In the present paper I am not concerned with the absolute definitional power of a language, but rather with the question of whether, in the context of a given formal theory, a given non-logical symbol (i.e. a predicate, relation or function symbol) is definable by means of some set of other non-logical symbols of the theory. I shall refer to this as *relative definability*.

The notion of relative definabilty bears directly on the role of 'primitive' concepts in ontology construction. The methodology of representing knowledge in terms of a small number of fundamental primitives has been adopted in many AI formalisms. For example the 'conceptual dependency' diagrams of [3] were used to represent a wide range of natural language verbs in terms of a small number of action primitives. Another example is spatial representation languages such as the Region Connection Calculus [4]. All ontologies will implicitly have primitive concepts that are not introduced by definitions. However, the number of primitives and whether they have special status within the formal framework will vary from ontology to ontology. Most large ontologies (e.g. CYC [5] and Ontolingua [6]) are not based on an explicit set of primitives (apart from the logical and quasi-logical primitives that occur in the representation language for the ontology). But there are some developments where ontologies are based on a restricted set of primitives. For example, [7] describes an architecture in which concepts are specified in terms of a library of basic pre-axiomatised components.

*This work was partially supported by the Ordnance Survey, as part of a project researching foundations for an ontology of the built environment.

In [8] I argued the case for much more emphasis on definitions and primitives in ontology development. It should be noted that this approach does not mean that an ontology should be dogmatically committed to any specific set of primitives. Any theory can be formulated in many different ways, which can take different sets of concepts as primitive. Some choices may be easier to work with than others, depending on the conceptual vocabulary one wants to formalise within the theory (for a large, multi-faceted ontology, it may actually be useful to identify several sets of primitives, which give alternative perspectives on the theory). Nevertheless, the possibility of defining one concept in terms of others gives a very powerful mechanism for organising and streamlining ontology development. If one is extending an ontology with a new concept then it is very useful to know whether it is already definable from existing vocabulary. If so, specifying a definition will enable its meaning to be fully formalised without the need for additional axioms. Of course, it may still be useful to record other logical properties of the concept; but these will have the status of lemmas rather than new axioms — they will be provable from the axioms governing the primitves.

The present paper concentrates on theoretical properties and theorems relating to relative definability and the identification of sets of primitives that are adequate to articulate a theory. I shall be concerned to show how theorems, which were originally applied to very strict mathematical theories, can be adapted to provide useful tools for analysing more loosely formulated ontological theories.

In the next section I examine the *method of Padoa* and Tarski's definability theorems [9] which provide fundamental theoretical tools for the analysis of definitions, and definability. In particular, I shall examine the notion of a *conceptually complete* theory, which is one that cannot be extended with new concepts except those that are already definable in terms of its existing vocabulary. This can be used to identify sets of primitives that are sufficient to fully describe a given domain. Tarski's conceptual completeness theorem only applies directly to theories, which are both categorical and 'monotransformable', thus the remainder of the paper is concerned with adapting it to theories which do not have these rather strict properties. In section 3 I show that the monotransforabilty requirement can be relaxed to give a more general theorem. Section 4 examines non-categorical theories and explains how they can be extended to categorical theories that explicitly axiomatise the space of interpretations of the original theory. These are then amenable to the application of Tarsk's conceptual completeness theorem to identify fully expressive sets of primitives. I conclude with a few further remarks about the problems and prospects of my approach.

2 The Meta-Theory of Relative Definability

2.1 The Method of Padoa

At the beginning of the last century, the mathematician Alessandro Padoa, identified a method for determining whether a given conceptual term of a theory is definable from a set of other terms of that theory [10]. Suppose we have a formal theory Θ formalised in terms of symbols $\sigma, \tau_1, \ldots, \tau_n$ and possibly other some symbols. Padoa suggested that to show that σ is *not* definable in terms of $\tau_1, \ldots, \tau_n$, it suffices to give two interpretations of all the non-logical symbols of Θ (i.e. models of Θ), which agree on the denotations of each of the symbols $\tau_1, \ldots, \tau_n$, but give different denotations for the symbol σ. Later, Tarski [9] proved rigorously that this indeed gives a necessary and sufficient condition for definability within a formal theory.

The essence of Padoa's method lies in the realisation that where there are dependencies among concepts, this is manifest both semantically, in terms of dependencies among interpretations and syntactically, in terms of formal definability. Although its statement and proof are somewhat subtle, Padoa's method can actually be employed rather easily to any kind of formal theory. Hence, it provides a valuable tool for answering questions of relative definability in ontological theories.

2.2 Categoricity and Montransformability

The most important meta-theoretic concepts in the theory of definability are categoricity and monotransformability.

Definition 1. *A theory (i.e. a set of classical formulae) is* categorical *iff and only if all its models are isomorphic.*[1]

Definition 2. *A theory is* monotransformable *if given any two isomorphic models of the theory there is a* unique *isomorphism between them.*

Monotransformability of a theory is equivalent to the condition that for each of its models, the only automorphism (i.e. mapping onto itself) on the domain that preserves the denotations of the non-logical symbols is the identity transform. Theories that are categorical but not monotransformable have some kind of symmetry or self-similarity in their models. For instance, consider a categorical theory of n-dimensional Euclidean point geometry based only on the 4-place equidistance primitive $xy = zw$. Though all models of this theory are isomorphic, there are many possible mappings between two models. This is because the structure is highly self-similar: it is invariant under transformations of translation, rotation and scaling.

But it is very simple to augment such a geometry to get a monotransformable theory. What we need to do is add some extra constants that enable us to fix a specific coordinate system within the space. For example, in the 2D case we might add constants c_o, c_x and c_y denoting respectively the origin point and the points lying at a unit distance from the origin along the x and y axes. We then add the axioms: $c_o c_x = c_o c_y$ and $\mathsf{RA}(c_x, c_o, c_y)$, where RA expresses the relation holding between three points that form a right-angle (this relation is definable from equidistance). The resulting theory is now monotransformable because once we equate the denotations of c_o, c_x and c_y in one model with those in another, the mappings of all other points are completely determined.

Though they are not commonly considered, it is possible to construct theories which are monotransformable but not categorical. Such theories have multiple non-isomorphic models but all of these are free of symmetry. For example, start with a 2-dimensional Euclidean geometry without a fixed coordinate system and add three point constant symbols: a, b, c. Add axioms to ensure that these points are linearly independent but apart from that leave their geometry indeterminate. Clearly the resulting theory is not categorical since there is a model for every assignment of linearly independent points to the constants a, b and c. However, all these models are free of symmetry since the identity mapping is the only transform on the domain that will preserve the denotations of all non-logical symbols including the constants a, b and c

2.3 Definability and Conceptual Completeness

We now come to the formal definition of definability and the key theorems which tell us when a concept is definable within a theory and under what circumstances a theory can be augmented with new undefinable primitives.

[1] Note that, because of the Skolem-Lowenheim theorem, 1st-order theories with infinite domains cannot be categorical because they always have non-denumerable models. However, if they have a unique denumerable model they are called $\aleph_0$-categorical. In the current paper I am *not* restricting to 1st-order theories except where I explicitly say so. Thus I use a strong and simple definition of 'categorical'.

Definition 3. *The meta-functions* $\mathsf{Voc}(\psi)$ *and* $\mathsf{Voc}(\Theta)$ *denotes the set of the non-logical symbols occurring respectively in the formula* ψ *and the theory* Θ*.*

Definition 4. $\cong$ *is a polymorphic equivalence relation which can hold between constants, functions or relations (where symbols of the latter two types must have the same arity). According to the type of its arguments* $\cong$ *is interpreted as follows*

- $(a \cong b) \;\equiv\; (a = b)$
- $(f \cong g) \;\equiv\; \forall x_1 \ldots x_n[f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n)]$
- $(P \cong Q) \;\equiv\; (P \leftrightarrow Q)$
- $(R \cong S) \;\equiv\; \forall x_1 \ldots x_n[R(x_1, \ldots, x_n) \leftrightarrow S(x_1, \ldots, x_n)]$

Definition 5. *A* possible definition *of a non-logical symbol* σ*, which is called the* definiens *of the definition, is a formula of the form* $\forall x[(x \cong \sigma) \leftrightarrow \Phi(x)]$ *, where* Φ *does not contain* σ*.*

Definition 6. *A non-logical symbol* σ *is* definable *within a theory* Θ *in terms of vocabulary* V *if and only if there is some possible definition* $\forall x[(x \cong \sigma) \leftrightarrow \Phi(x)]$*, where* $\mathsf{Voc}(\Phi) \subseteq V$ *and* $\Theta \models \forall x[(x \cong \sigma) \leftrightarrow \Phi(x)]$*.*

Note that this form of definitional formula is in general 2nd-order, since the universal quantifier ranges over denotations of the the definiens σ, which may be a predicate or relation. For practical purposes we shall usually be concerned with 1st-order definitions:

Definition 7. *1st-order definitional formulae take one of the following forms, depending on the type of symbol defined:*

- $\forall x_1 \ldots x_n[R(x_1, \ldots, x_n) \leftrightarrow \Psi(x_1, \ldots, x_n)]$
- $\forall x y_1 \ldots y_n[(x = f(y_1, \ldots, y_n)) \leftrightarrow \Psi(x, y_1, \ldots, y_n)]$
- $\forall x[(x = c) \leftrightarrow \Psi(x)]$

The definienda *(terms defined) by such formulae are respectively: the relation* R*, the function* f *and the constant* c*. The expression* $\Psi(x_1, \ldots, x_n)$ *is called the* definiens*.*

A 1st-order definition *is any formula of one of the above forms in which the definiendum does not occur in the definiens.*

Happily, if we are dealing with a 1st-order theory, we can be sure, because of Beth's definability theorem, that any definable symbol can be given a 1st-order definition.

The notion of definability allows us to identify subsets of the vocabulary of a theory which can be taken as 'primitive' concepts:

Definition 8. *A* primitive vocabulary *for a theory* Θ *is a set* $S \subseteq \mathsf{Voc}(\Theta)$ *such that all symbols in* $\mathsf{Voc}(\Theta)$ *are definable within* Θ *from the symbols in* S*.*

Definition 9. *A* minimal primitive vocabulary *for a theory* Θ *is a primitive vocabulary for* Θ *of which no proper subset is also a primitive vocabulary for* Θ*.*

Now we come to a property that give us a useful handle on the definitional power of theories.

Definition 10. *A theory* Θ *is* conceptually complete[2] *iff there is no categorical theory* Θ' *such that* $\Theta \subset \Theta'$ *and* $\mathsf{Voc}(\Theta')$ *contains at least one symbol that is not definable within* Θ' *in terms of the vocabulary* $\mathsf{Voc}(\Theta)$*.*

[2]Tarski actually uses the phrase 'complete with respect to its specific terms' rather then 'conceptually complete'.

So a ‘conceptually complete’ theory is one of which no *categorical* extension contains any new concept that is not definable (in the extended theory) from the vocabulary already present in the original theory. Note that if we dropped the restriction of considering only categorical extensions then no theory would be conceptually complete: it is always possible to add extra undefinable terms to a theory if we don’t care about exactly specifying their meaning.

Though conceptual completeness is concerned with definability in categorical extensions of a theory it is still a useful notion for theories that we do not intend to axiomatise up to categoricity. Given that any categorical axiomatisation of a new concept within a conceptually complete theory would entail a definition, it will normally be more sensible to define the concept rather than give a partial axiomatisation.

It is worth noting that conceptual completeness of a theory does *not* mean that it is capable of defining all possible relationships over its domain. Indeed, if the domain is infinite there will be an uncountable number of possible predicate extensions, so these cannot all be defined within a language with a recursively enumerable set of well defined formulae. The property of conceptual completness just means that the definitional power of the theory cannot be extended by adding more primitives.

2.4 Tarski’s Definability Theorems

Now we come to the key theorems of [9]. The first specifies conditions for definability and the second gives a sufficient condition for conceptual completeness of a theory.

Theorem 1. *For any theory $\Theta(\sigma, \pi_1, \ldots, \pi_m, \rho_1, \ldots, \rho_n)$, where $\sigma, \pi_1, \ldots, \pi_m, \rho_1, \ldots, \rho_n$ are all the non-logical symbols of the theory, the symbol σ is definable purely in terms of the symbols $\pi_1, \ldots, \pi_m$ just in case the following equivalent conditions hold:*

a) $\Theta(\sigma, \pi_1, \ldots, \pi_m, \rho_1, \ldots, \rho_n) \rightarrow (\forall x[(x \cong \sigma) \leftrightarrow \exists y_1 \ldots y_n[\Theta(x, \pi_1, \ldots, \pi_m, y_1, \ldots, y_n)]]])$

b) $\forall x x' y_1 \ldots y_m z_1 \ldots z_n z'_1 \ldots z'_n[$
$$(\Theta(x, y_1, \ldots, y_m, z_1, \ldots, z_n) \wedge \Theta(x', y_1, \ldots, y_m, z'_1, \ldots, z'_n)) \rightarrow x \cong x']$$

Since the symbols $\sigma, \pi_1, \ldots, \pi_m, \rho_1, \ldots \rho_n$ may be of any logical type, these definability formulae are in general higher order. However, it follows from basic logical principles that condition b) can be proved just in case:

c) $(\Theta(\sigma, \pi_1, \ldots, \pi_m, \rho_1, \ldots \rho_n) \wedge \Theta(\sigma', \pi_1, \ldots, \pi_m, \rho'_1, \ldots \rho'_n)) \rightarrow \sigma \cong \sigma'$

So, to check whether σ is definable in Θ in terms of symbols $\pi_1, \ldots, \pi_m$, we construct theory Θ' which is exactly like Θ except that all non-logical symbols apart from the $\pi_1, \ldots, \pi_m$ are replaced by new symbols. In particular σ is replaced by some symbol σ'. Then σ is definable in Θ in terms of symbols $\pi_1, \ldots, \pi_m$ just in case $\Theta \wedge \Theta'$ entails that σ is equivalent to σ'. If Θ is 1st-order this gives a 1st-order test for definability.

Theorem 2. *Every monotransformable theory is conceptually complete.*

Proof sketch:[3] Let Φ be a monotransformable theory with $\mathsf{Voc}(\Phi) = \{\pi_i\}$ and Φ^+ be a categorical extension of Φ, containing some additional vocabulary including a symbol α and possibly some additional axioms. Thus $\Phi^+ = \Phi \wedge \Psi(\alpha)$. Appealing to theorem 1c we see that α is definable from symbols $\{\pi_i\}$ in Φ^+ just in case $(\Phi \wedge \Psi(\alpha) \wedge \Psi(\alpha')) \rightarrow \alpha \cong \alpha'$. Now consider a model $\mathcal{M} = \langle D, \ldots, \pi_i, \ldots, \alpha, \alpha' \rangle$ which satisfies $(\Phi \wedge \Psi(\alpha) \wedge \Psi(\alpha'))$. This has sub-models $\langle D, \ldots, \pi_i, \ldots, \alpha \rangle$ and $\langle D, \ldots, \pi_i, \ldots, \alpha' \rangle$. We know that $(\Phi \wedge \Psi(\alpha))$

[3]See [9] for detailed proof.

and likewise $(\Phi \wedge \Psi(\alpha'))$ are categorical so these structures must be isomorphic, modulo the renaming of α to α'. But these models also satisfy Φ which ensures monotransformability. Thus the only isomorphism between these models is the identity mapping on the domain D. Consequently, α and α' must have exactly the same denotations on D. Thus $\mathcal{M}$ must satisfy $\alpha \cong \alpha'$. ■

3 Generalising the Definability Theorem

Theorem 2 applies to theories that are 'monotransformable' a property that is rarely mentioned ontology circles. However, it is quite easy to generalise the theorem to apply to any theory whose primitives exhibit symmetries with respect to model structures. The generalised theorem is:

Theorem 3. *If all automorphisms on the models of a theory Θ belong to a class $\mathcal{T}$ then Θ is conceptually complete with respect to all concepts that are invariant under the transforms in $\mathcal{T}$.*

Proof sketch:[4] The theorem can be proved by modification of the proof of Theorem 2. As before we consider a model of a theory containing symbols α and α' which satisfy exactly the same constraining formulae. We again note that the two sub-models obtained by dropping one of the symbols from the signature are isomorphic modulo renaming. This time the isomorphism between these sub-models is not necessarily the identity mapping but it must be in the class $\mathcal{T}$. Nevertheless, since we require that the denotation of α must be invariant under mappings in this class, we can still conclude that α and α' have the same denotation and hence that α is definable. ■

This generalisation means that a theory may exhibit a weakened form of conceptual completeness even though its models exhibit certain symmetries. For many applications we will be concerned with concepts that are invariant under certain kinds of transform on their models. For instance qualitative spatial relationships of the type corresponding to descriptive terms of natural language (e.g. 'between', 'touching', 'inside') are nearly always invariant with respect to translation and rotation (and usually also re-scaling) within space; and temporal relations such as Allen's interval relations are invariant under translation of the underlying time series.

Theorem 3 gives us an extremely powerful tool in deciding what can be defined from what primitives. In order to decide whether a concept α is definable from primitives $\{\pi_i\}$ we need only consider whether there is some transform, which if applied to any model of our theory would preserve the extensions of all concepts in $\{\pi_i\}$ and yet would not preserve the extension of α.[5] So, for example, suppose I have a spatial theory based on the primitive of 'betweenness'. This primitive is preserved under all and only *affine* transforms, which re-scale, rotate, translate or linearly skew the space. Since 'convexity' is preserved by all these transforms, I immediately know that it is definable from betweenness. But 'sphericity' is not preserved by skewing the space, so a 'sphere' predicate is not definable. On the other hand,

[4]A more detailed proof can be obtained by modifying Tarski's proof of his THEOREM 4 in [9]. Here we need to replace the condition $R = I$ constraining the isomorphism to be the identity by the condition $R \in \mathcal{T}$.

[5]Here I am assuming that the concept intended as the referent of α is axiomatisable within the formal system we wish to employ. If it is not axiomatisable, then clearly it will not be definable. If we have a theory with an infinite domain then there are an uncountable number of possible extensions of concepts. So, if we restrict our axiomatic system to having some finite representation, then clearly there are possible extensions that we could never capture as the extension of a defined (or categorically axiomatised) symbol. It seems to me unlikely that we would ever find a practical use for theories including partial axiomatisations of such 'inaccessible' concepts.

the concept of a 'point' is preserved by affine but also many non-affine transforms, which do not preserve convexity, so 'convexity' is not definable from 'point'.

4 Non-Categorical Theories

The method of Padoa applies to any kind of theory, so can be used to detect interdefinability within the vocabulary of any ontology. While this is useful, it is not sufficient to establish conceptual completeness and thereby identify a set of primitives as being fully expressive over a domain. To find such primitives we need to look for concepts that are sufficient to provide a monotransformable axiomatisation, which allows no symmetries (i.e. non-identical automorphisms) in its models.

An apparent limitation of the notion of conceptual completeness is that it only ensures that we can define all concepts in *categorical* extensions of the theory. But categoricity is not normally a property of the kinds of theory that are currently being constructed as ontologies for use in computer systems. The consensus among computer-oriented ontology developers seems to be that the categoricity is only useful in pure mathematical theories. In more pragmatic contexts it is generally believed that the attainment of categoricity is both infeasible and undesirable. In this section, I shall demonstrate how non-categorical theories can be regarded as partial formalisations of larger categorical theories. This means that, even where we cannot or do not wish to achieve categoricity, we can employ Tarski's definability theorem and its generalisation to identify expressively complete sets of primitives.

There are several reasons, both practical and theoretical, why a formal theory may not be categorical. Specifically, I distinguish the following:

Incompleteness: The axiom set is incomplete relative to the intended meaning of the vocabulary of the theory.[6]

Contingency: The theory contains concepts that refer to contingent properties and relations, whose extension will vary according to different possible situations that can be described within the theory.

Vagueness: The theory contains vague or ambiguous concepts, which can be interpreted in different ways.

In general non-categoricity may be due to a combination of several or all of these factors; and, in order to get a clear picture of the status of a theory, an ontology developer should be aware the extent to which each comes into play.

4.1 Incompleteness

Ideally we would like the axiomatisation of an ontology to capture all the intrinsic semantic relationships between concepts. But there are two reasons why this may not be so:

a) Axioms been unintentionally overlooked due to limitations of the ontology developers.
b) Axioms have been intentionally omitted — perhaps to help make reasoning more tractable.

Unintentional incompleteness may be seen simply as an inadequacy in the axiomatisation of an ontology. However, most large ontologies admit to some incompleteness and indeed regard a degree of incompleteness as an inevitable consequence of design decisions upon which the ontology is based. In particular, incompleteness often arises from limited expressiveness

[6] Note that we are concerned here with semantic rather than proof theoretic incompleteness. A 2nd-order theory may completely characterise its intended models and thus be semantically complete, although it does not have a complete proof theory (e.g. 2nd-order Peano arithmatic [11]).

of the underlying logical language used to represent the ontology. This language (for instance a particular description logic) will have been chosen as a tradeoff between expressive power and tractability of inference (see e.g. [12]).

Unfortunately the practical concern of achieving tractability tends to lead to a bluring between intended and unintended incompleteness. If semantic analysis is simplified for the sake of tractability, one risks compromising the precise and unbiased specification of concept meanings that an ontology is supposed to provide. Rather than developing ontologies within languages of limited expressiveness, a cleaner methodology would be to separate the representation and reasoning problems: on the one hand one should aim for a complete axiomatisation of the concepts in the ontology; but on the other hand in order to compute inferences one will need to use an inference mechanism that is incomplete. Such an inference mechanism may well work in terms of computationally oriented languages such as description logics, which would form sublanguages of the more expressive language needed for the complete axiomatisation.

Demonstrating completeness is a complex theoretical task and becomes increasingly hard to achieve for theories with a large number of independent concepts. As I argued in [8], this is a motivation for adopting an architecture in which concepts are defined in terms of as few primitives as possible. As long as we establish a complete axiomatisation for the primitive concepts we guarantee complete axiomatisation of the defined concepts. But even if we fail to achieve completeness, we must accept that ideally we would like our theory to be complete with respect to the intended meanings of its terms. This means that in identifying definability among concepts it is reasonable to consider whether a definition would be possible in an axiomatically complete extension of our imperfect theory. Hence in applying Tarski's definability theorem we may overlook any lack of categoricity that arises simply through incomplete axioms.

4.2 Contingency, Categoricity and Possible Worlds

The state of the world is largely a matter of contingent circumstance: its stuff could be arranged in many different ways. Surely each of these possible states of the world should correspond to some model of a general ontological theory. Thus the theory should have many different models: it certainly should not be categorical otherwise it would limit us to just one world state. Moreover having a class of possible models is fine for reasoning, since a valid inference is one that holds in all models of the theory.

This point is true as far as it goes; but if we consider a bigger picture we can see a set of possible models as constituents of a larger model. In an informal sense, a complete theory becomes categorical as soon as one states that its models correspond to all and only those structures that are genuinely possible — i.e. the formal models are exactly the 'intended' models. If this is so, we can conceive of these possible models as embedded structures within a categorical super-theory, whose unique model incorporates all possible states of the world. This gives us a way of examining definability properties of theories, which are not in their original form categorical.

This idea can be made fully formal. Consider a theory $\Theta(\pi_1 \ldots \pi_n)$ with many models. We shall transform this into a categorical theory, $\mathsf{CPC}(\Theta)$, the *categorical possibilistic closure* of Θ. Each n-ary symbol π_i in $\mathsf{Voc}(\Theta)$ will correspond to a new $n{+}1$-ary symbol $\hat{\pi}_i$ in $\mathsf{CPC}(\Theta)$. This vocabulary of new symbols will be referred to as $\mathsf{Voc}^{\wedge}(\Theta)$. The extra argument of the new symbols is for a possible world operand. If $R(x, y)$ is a binary relation in Θ, the new theory $\mathsf{Voc}^{\wedge}(\Theta)$ will include a ternary relation $\hat{R}(w, x, y)$, which can be read as: "$R(x, y)$ is *realised* at world w".

First, I construct an intermediate theory $\mathsf{PC}(\Theta)$, which contains the combined vocabulary of Θ and $\mathsf{CPC}(\Theta)$ — i.e. has predicates both with and without the additional world argument. $\mathsf{PC}(\Theta)$ includes a suitable set of axioms Θ^Σ to ensure that the elements that can satisfy the possible world argument places of all the new $\hat{\pi}$ symbols are disjoint from those satisfying any other argument places (so the worlds are distinct sort of entity).

To help describe how the new vocabulary relates to the old I introduce the following notation. If $\hat{\pi}$ is an n-ary symbol in $\mathsf{Voc}^\wedge(\Theta)$ then $(\hat{\pi}\downarrow_w)$ refers to that $(n-1)$-ary relation or function obtained by fixing the 'world' argument to the value of w. Thus $(\hat{\pi}_1\downarrow_w) \cong \xi$ states formally that the denotation of variable ξ has the same extension as the relation, function or constant denoted by $\hat{\pi}_1$ has at world w. Using this notation I define $\mathsf{PC}(\Theta)$ as follows:

Definition 11. *The* possibilistic closure *(*PC*) of a theory* Θ *is given by:*

$$\begin{aligned}\mathsf{PC}[\Theta] \equiv_{\text{def}} \ & \Theta \wedge \Theta^\Sigma \wedge \\ & \forall \xi_1 \ldots \xi_n[\Theta(\xi_1 \ldots \xi_n) \leftrightarrow \exists w[((\hat{\pi}_1\downarrow_w) \cong \xi_1) \wedge \ \ldots \ \wedge ((\hat{\pi}_n\downarrow_w) \cong \xi_n)] \wedge \\ & \quad \forall w_1 w_2[(((\hat{\pi}_1\downarrow_{w_1}) \cong (\hat{\pi}_1\downarrow_{w_2})) \wedge \ \ldots \ \wedge ((\hat{\pi}_n\downarrow_{w_1}) \cong (\hat{\pi}_n\downarrow_{w_2}))) \rightarrow w_1 = w_2]\end{aligned}$$

Here, the third conjunct ensures that for all possible values of relation, function, and constant symbols that satisfy the constraints imposed by Θ on its vocabulary, there is some world in $\mathsf{PC}(\Theta)$ that realises this valuation. The final conjunct says that there cannot be two distinct worlds that realise exactly the same extensions.

Clearly $\mathsf{PC}(\Theta)$ contains a unique world corresponding to each model of Θ and fixes exactly the extensions of all the $\hat{\pi}$ predicates. However, $\mathsf{PC}(\Theta)$ is not itself categorical because it does not fix the extensions of the original symbols of Θ. However the subtheory of $\mathsf{PC}(\Theta)$ that involves only concepts in $\mathsf{Voc}^\wedge(\Theta)$ is clearly categorical. Hence, I define:

Definition 12. *The* categorical possibilistic closure *(*CPC*) of a theory* Θ *is given by:*

$$\mathsf{CPC}(\Theta) \ = \ \{\phi \mid \mathsf{PC}(\Theta) \models \phi \quad \textit{and} \quad \mathsf{Voc}(\phi) \subset \mathsf{Voc}^\wedge(\Theta)\}$$

For any theory Θ, $\mathsf{CPC}[\Theta]$ is categorical and hence by theorem 3 is conceptually complete with respect to any concepts that are preserved by the symmetries of its models.

The CPC construct is satisfactory only if the vocabulary of the original theory is adequate to specify all relevant distinctions that could be made between possible worlds. This construction should not be applied if one wants to allow multiple distinct worlds realising the same denotations for the vocabulary of the original theory. Moreover, if such multiplicity does exist, there may be scope for adding useful new primitives that are not definable from the initial vocabulary.

The CPC construct is only an artificial device to illustrate the fact that theories of contingent properties can be made categorical within a broader possibilistic framework. I do not suggest that it should be actually used for constructing practical ontological theories. Having said this, explicit reference to possible worlds may well be a reasonable way of modelling contingency within a categorical framework. But, whereas CPC invokes heavy meta-level machinery to get a categorical possible worlds theory, it is much preferable to find a way of specifying the domain of possible worlds directly within the ontology.

The main point of the CPC construction is that it shows that contingency can be modelled within a categorical framework. As with the ideal complete extension of an incomplete theory, the CPC shows how theories that deal with contingent concepts can also be made categorical. Hence, again we can use the Tarski' definability theorem to establish fully expressive sets of primitives for a contingent theory. However, in order to do this we will in general need to augment our theory with some mechanism of refering to possible worlds — either explicitly (as in the CPC) or implicitly (perhaps using a modal logic).

4.3 Vagueness

The issue of definability is somewhat complicated by the pervasive presence of vagueness and ambiguity in natural language. Since a vague concept can be interpreted in various different ways it is difficult to imagine how such concepts could be formalised within a categorical theory. Moreover, it seems unlikely that a vague concept could be adequately defined from the primitives of a categorical theory. Nevertheless, it is instructive to apply the method of Padoa to vague concepts.

For example, let us consider the possible interpretation of 'good deed' relative to a categorical materialistic theory of the spatio-tempoal history of the universe. Such a theory would contain geometric, temporal and material primitives sufficient to completely describe the physical structure of any spatio-temporal history (and thus its CPC would describe all possible spatio-temporal theories). Clearly there will be no uncontroversial way of interpreting 'good deed' relative to a material specification of the universe. However, suppose we have two models that agree exactly on the physical properties of the universe. Can they then differ on the interpretation of 'good deed'? Could exactly the same sequence of physical configurations constitute in one case a good deed but in another not so?

The point I am making is that the undefinability of 'good deed' does not come from the nature of what it refers to but rather from its vagueness. If we had a perfectly clear idea of what constitutes a good deed then, given a complete material description of the history of the universe. we would be able to precisely identify all the instances of good deeds; and thus, by application of the method of Padoa, we would find that 'good deed' is indeed definable from material primitives.

However, it seems that for many natural concepts, such as 'good', vagueness is intrinsic and cannot be eliminated except by artificial regimentation. Thus, if we want concepts in an ontology to correspond to natural language concepts we must allow for this source of undefinability.

We can treat vague concepts is as loosely constrained relative to the precise primitives of a theory. Thus, although there is a unique precise model (assuming we have eliminated contingency as discussed in the previous subsection), there are many interpretations of the vague concepts. However, it may still be useful to apply the method of Padoa to discover definitional dependencies among vague concepts. Moreover, we can apply a form of closure much the same as the CPC to regain categoricity. I will call this the *categorical supervaluationistic closure* (CSC) because of its relationship to the 'supervaluation' semantics for vagueness proposed by Fine [13]. The extra variables added in constructing the CSC now correspond to possible interpretations (or *precisifications*) rather than possible worlds.

In general a theory may involve both vagueness and contingency. In such a case one would first apply CPC to the precise concepts of the theory. Then add the axioms for the vague concepts and finally form the CSC. The extension of each concept in the theory constructed in this way would then be evaluated relative to both a possible world and a precisification. Within such a theory we can apply Tarski's definability theorem to identify conceptually complete sets of primitives.

5 Conclusion

I have examined the notion of definability and considered its relevance to ontology construction. Tarski's theorems and the method of Padoa give us a semantic route to determining definability and conceptual completeness. I suggest that these notions can play an important role in organising the formalisation of a large vocabulary. Of course there are still many choices open to the ontology designer. Just because one concept is theoretically reducible to certain other concepts does not mean it must necessarily be treated as defined within an ontology. Indeed, if the ontology is formulated in a restricted logical language chosen for inferential properties, this may not be possible. However, if our concern is primarily with representing meaning, rather than computing inferences, defining the concept is likely to be preferable to axiomatising it as if it were a new independent concept.

In the current paper I have said relatively little about actual sets of primitives that can form the basis of conceptually complete theories. A preliminary study of this was presented in [8], which explored the expressive power of geometrical primitives and, in particular the theory of *Region Based Geometry* [14]. The analysis given there suggests that from a theoretical point of view it may be possible to define very wide ranging vocabularies from rather small sets of primitives. Spatial and temporal primitives are almost certainly necessary to formalise any reasonably expressive vocabulary. To capture concepts whose meaning does not concern only actual states of affairs, one will also need some primitive relating to alternative possibilities. And, to handle concepts relating to intelligent agents, one may need further primitives relating to mental states and intensionality.

A collection of very expressive foundational theories based on spatial and temporal primitives can be found on the *Foundational Ontology* web site at `http://www.comp.leeds.ac.uk/ontology` and a core ontology of matter and material objects is described in [15].

References

[1] F. Donini, M. Lenzerini, M. Nardi, and A. Schaerf. Reasoning in description logics. In G. Brewka, editor, *Principles of Knowledge Representation*, Studies in Logic Language and Information. CSLI Publications, Stanford University, 1996.

[2] D. Calvanese, M. Lenzerini, and D. Nardi. Description logics for conceptual data modelling. In J. Chomicki and G. Saake, editors, *Logics for Databases and Information Systems*, chapter 1. Kluwer, 1998.

[3] R. C. Schank. *Conceptual Information Processing*. North-Holland, Amsterdam, 1975.

[4] D.A. Randell, Z. Cui, and A.G. Cohn. A spatial logic based on regions and connection. In *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*, pages 165–176, San Mateo, 1992. Morgan Kaufmann.

[5] D.B. Lenat. CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 1995.

[6] R. Fikes and A. Farquhar. Distributed repositories of highly expressive reusable ontologies. *IEEE Intelligent Systems*, 14(2):73–79, 1999.

[7] Ken Barker, Bruce Porter, and Peter Clark. A library of generic concepts for composing knowledge bases. In *Proceedings of the First International Conference on Knowledge Capture (K-CAP-01)*, pages 22–29, Victoria, October 2001. ACM.

[8] B. Bennett. The role of definitions in construction and analysis of formal ontologies. In *Sixth Symposium on Logical Formalizations of Commonsense Reasoning*, AAAI Spring Symposium, Palo Alto, 2003. `http://www.comp.leeds.ac.uk/qsr/pub/BennettAAAsymp03a.ps`.

[9] A. Tarski. Some methodological investigations on the definability of concepts. In *Logic, Semantics, Metamathematics*. Oxford Clarendon Press, 1956. trans. J.H. Woodger.

[10] Alessandro Padoa. Un nouveau système irréductible de postulats pour l'algèbre. In *Proceedings of the Deuxième Congrès International des Mathématiciens*, pages 249–256, Paris, 1902.

[11] Richard Dedekind. *Was sind und was sollen die Zahlen?* Braunshweig, 1888.

[12] I. Horrocks and U. Sattler. Ontology reasoning in the $\mathcal{SHOQ}(d)$ description logic. In B. Nebel, editor, *Proc. of the 17th Int. Joint Conf. on Artificial Intelligence (IJCAI 2001)*, pages 199–204. Morgan Kaufmann, 2001.

[13] K Fine. Vagueness, truth and logic. *Synthèse*, 30:263–300, 1975.

[14] B. Bennett. A categorical axiomatisation of region-based geometry. *Fundamenta Informaticae*, 46(1–2):145–158, 2001.

[15] B. Bennett. Space, time, matter and things. In C. Welty and B. Smith, editors, *Proceedings of the 2nd international conference on Formal Ontology in Information Systems (FOIS'01)*, pages 105–116, Ogunquit, 2001. ACM. `http://www.comp.leeds.ac.uk/qsr/pub/bennett-fois01.ps`.