Partial Reductive Paraphrase: Toward More Transparent Requirements

Kimberly S. Wasson
Department of Computer Science
University of Virginia
151 Engineer's Way
Charlottesville, VA 22904-4740
kwasson@cs.virginia.edu

Abstract

Partial Reductive Paraphrase (PRP) is a method for constructing explications, a specific kind of definition, for use in requirements documents. Explications avoid common weaknesses of definitions produced by other methods and encourage effective representation of the information most critical to the application domain at hand, allowing better understanding on the part of requirements users. This paper introduces PRP and demonstrates how it is used to construct explications of concepts and present them for accessibility to users. PRP explications are provided for concepts drawn from industrial requirements documents and are then evaluated relative to analogous definitions produced by other methods. The PRP explications are shown to be free of specific faults that otherwise result. Finally, means by which PRP can be integrated into the requirements process are addressed.

1. Introduction

Achieving validity in a software artifact depends critically on the ability of requirements consumers to correctly interpret the intentions of requirements producers. Properties of natural language allow miscommunication among even parties with similar backgrounds. Further, the fact that much of the material to be communicated via a software requirements specification (SRS) is specific to the application domain, attention must be paid to the presentation of this material such that it is made as accessible as possible, to non-experts specifically, and to all consumers of an SRS generally.

In support of this goal, it is commonly advised that an SRS include a glossary or other presentation of terminology important to comprehension of material in the document. Though such a provision is generally made, and made in good faith and by smart people, the definitions provided are virtually guaranteed to be insufficient to the task. Specifically, the methods by which terms and phrases are chosen for definition and the methods by which they are defined are guided not by systematic invocation of knowledge about how people interpret meaning, but by ad hoc and intuitive processes rooted in the definition writer's experience of other definitions and best guesses about what might be important. These methods produce definitions that are vulnerable to a set of weaknesses which can threaten or prevent correct interpretation, allowing misconceptions to enter a downstream system design, implementation or delivered product if not discovered.

Linguistics includes study of the difficulties inherent in definition writing, and provides insights into the properties of good and bad definitions that can be exploited to improve the quality of SRS glossaries. In particular, an initiative to develop a universal and structurally sound dictionary offers a framework adaptable to applications of non-universal scope. The Natural Semantic Metalanguage (NSM) project aims to build this dictionary via the ordered compounding of a set of empirically supported semantic primitives. We take some of the tenets and methods of NSM and refine them to a use appropriate to our goals.

Specifically, we take knowledge about how particular properties of definition structure encourage consumers of a definition to interpret it as intended, while others result in misinterpretation or render meaning completely inaccessible. We assert that provocation of a correct interpretation is structurally directable, and have developed a method that offers such direction.

In this work, we provide an analysis of the problem of definition and a method, Partial Reductive Paraphrase (PRP), for systematically and efficiently constructing definitions of higher quality. We demonstrate that PRP results in representations free of specific faults that otherwise result and that hinder arrival at the intended interpretation.

It is further shown how PRP can be integrated into a greater requirements process.

2. The Problem of Communicating Word Meanings

To characterize the problem effectively, it is first necessary to dispel a common misconception regarding communication. We then use the corrected model as a basis upon which to frame the specific challenges to be overcome by our method.

2.1 Words are not Containers

The *container metaphor* is the name for the phenomenon that allows people to conceive of words as containers and meaning as that which is contained in them. That is, it is a mapping from our notion that objects in one location can be packaged and sent to another location onto a conception that ideas in one's mind can be packaged up in words and delivered, via speech, writing or otherwise, to another person, who, upon hearing or reading them, unwraps the meaning contained therein. [7] The metaphor is convenient and natural, and is apparent in such utterances as "Do you get what I'm saying?" or "He didn't catch the meaning."

However, while it is a useful conceptual shorthand, the container metaphor does not in fact accurately model communication. Rather, meaning does not move at all; it is entirely an emergent property of cognition and is thus wholly conceived and processed by an individual, and not itself something that can be given, accepted or recorded. An individual builds up his own idiosyncratic set of form-meaning associations over his lifetime of experience, and interpretation is the process of activating possessed meanings in response to presentation of the forms with which he has associated them.

The actual mechanism by which communicating parties understand what is being communicated is via either or a combination of two processes. First, a speaker can provoke the activation of a meaning already (assumed to be) possessed by an individual via presentation of a word that the hearer (is assumed to have) previously associated with that meaning. Second, a speaker can provoke the compounding of such activated meanings into a more complex one for which the speaker introduces a new word in order that the hearer can create a new association between it and the complex meaning. The important point is that while each individual does possess a set of word-meaning associations, in communication, it is only the words and not the meanings that can be traded as objects. The degree to which the communication is successful is the degree to which the communicating parties either possess sufficiently similar word-meaning associations to begin with, or can use their intuition about each other's experience and knowledge in order to scaffold the construction of new meanings via the invocation of existing associations and compounding and labeling of the result.

Thus the problem of communication is not an issue of determining the best way for one person to *convey* meaning *to* another, but rather one of determining the best way to *provoke* meaning *in* another. Understanding this difference is critical to being able to strategize the generation of more correct interpretations in the minds of those on the receiving end of a SRS.

2.2 Definitions are Descriptions

A definition is a device intended to allow an individual who does not possess a word-meaning association involving a word in question to generate one. Specifically, a definition presents a collection of words that if interpreted as intended, allow the construction of a new meaning that can be associated with the new word.

Entering a word in a dictionary implies that the builders of the dictionary believe there exists a relevant speech community in possession of a common meaning for a defined word. For example, a bilingual English-French dictionary makes assertions about the word-meaning associations shared by the population of native English speakers when it offers definitions of English words for French users. The motivating population for our purposes is the community of people who contribute to the production of a SRS, and specifically, the sub-populations that possess expertise in areas not familiar to those who will use the SRS.

A good definition can be characterized in terms of how well it represents the codified word-meaning association shared by the source community. "[A]n accurate definition will predict the appropriate range of use of a word. [T]he reader should be able to trust that the definition is a reliable guide" to how the word is used within the community in question. "[S]uch a definition is said to be *descriptively adequate*." [2]

Achieving descriptive adequacy is a two part problem that must take into account both the integrity of the description and the integrity of the described. The described is the concept for which a lexicographer wishes to produce a definition; he must correctly ascertain whether the concept as he conceives it is representative of the concept as conceived by the source community. If he continues with a faulty concept, then even the best description will describe the wrong thing. The description, on the other hand, is the representation used to provoke understanding of the concept in the mind of the user. If the lexicographer

is in possession of the correct concept, a faulty description will result in either no provocation of meaning at all in the mind of the user, if the description is simply intractable, or worse, provocation of incorrect meaning, which the user is unlikely to question, since we are conditioned to believe that dictionaries are the authority on word meanings.

In this work, we are primarily concerned with the integrity of the description, that is, the effectiveness with which the definition promotes comprehension of the described entity. For treatment of the integrity of the described, as well as treatment of the greater methodology of which the method to be presented here is a part, see section 8, as well as [13].

2.3 Requirements Producers vs. Lexicographers

While on the face of it, those charged with producing glossaries for SRSs have a job similar to that of the lexicographer, there are some important differences. Requirements producers and lexicographers both (should) have the goal of descriptive adequacy, and both generally produce well-intentioned definitions that are nonetheless insufficient to this task. Each can benefit from the methods to be described.

However, requirements producers further have to achieve this goal under greater resource constraints than do lexicographers, and with greater needs for success, under the threat of potentially severe consequences. Consequences to a software project that derive from requirements misinterpretation on the part of SRS users range from rework at various degrees of expense, to late or undelivered products, to products delivered faulty with the potential for causing harm. Lexicographers do not have these constraints or consequences, and since the definitions they produce are generally good *enough* for their intended purpose, they need not adjust their methods (but they still could, to better result). Requirements producers, on the other hand, have much more to gain from an adjustment to existing methods. We next characterize the ways that the integrity of a description can be violated, in order that we can strategize methods for prevention.

2.4 Common Weakness in Definitions

As discussed in section 2.2, achieving descriptive adequacy is a two part problem consisting of maintaining both the integrity of the description as well as the integrity of the described. We are concerned in this work with a method for maintaining the integrity of the description, and here present a typology of faults that violate this integrity.

Obscurity. An obscure definition presents a description that uses words no more significant to the user than the word being defined. [2] Since the user is lacking a word-meaning association allowing him to understand the target word to begin with, a definition that relies on words requiring additional word-meaning associations not either already possessed by the user or themselves otherwise appropriately defined is ill-equipped to support the desired outcome. In other words, an obscure definition might well make a true statement regarding the range of use of the defined, but the description is not useful since it cannot successfully provoke an interpretation by the user, correct or otherwise.

Scientific definitions are particularly vulnerable to obscurity. [9] defines ‘air’ as “the mixture of invisible odorless tasteless gases (as nitrogen and oxygen) that surrounds the earth.” A user not familiar with the term is not likely to be any more familiar with terms such as ‘nitrogen’ than with ‘air’ and thus will be limited in what he can interpret from the given description.

A more general source of obscurity in definitions is the confusion of meaning with knowledge, in which the lexicographer offers information that while true of the target in the encyclopedic sense, is superfluous to the goal of describing its meaning. [15] gives an example definition of ‘dentist’, drawn from [8]: “[A] person who is skilled in and licensed to practise the prevention, diagnosis, and treatment of diseases, injuries, and malformations of the teeth, jaws, and mouth and who makes and inserts false teeth.” Confusing meaning with knowledge contributes to more obscure definitions because the provision of additional knowledge beyond what is descriptively adequate increases the likelihood of reliance on more complex terms. Further, it hinders predictability, because with superfluous information, it is not clear what is predictable.

Obscurity can be eliminated to the degree that for every word used in a description of a target word, an effective word-meaning association is either already possessed by the user or can be constructed via support from additional fault-free definitions also available in the dictionary.

Circularity. A circular definition presents a description that uses one or more words that themselves are defined in terms of the target word. [2] A circular definition is non-interpretable because like an obscure definition, it relies on the availability of a further effective word-meaning association, but in the circular case, the association required is not forthcoming because it is for the target term itself, now recognized to be ineffectively defined.

Circularities can be divided into two classes: obvious circularities and insidious circularities. An obvious circularity uses the target term or a version of it in the description, such as “red: having the property of redness”. Understanding ‘redness’ presupposes understanding ‘red’, but understanding ‘red’ cannot be presupposed on the part of a user in search of a definition for ‘red’. Obvious circu-

larities are obvious because the circularity is apparent from the definition itself without reference to other definitions. Insidious circularities, on the other hand, are those for which there is hidden indirection that nevertheless results in a term relying eventually on itself for its own definition. For example, the [10] definition for 'question' relies on 'request', which relies on 'ask', which relies on 'answer', which relies back again on 'question'.

The reader might recognize that to break circularity completely, it will be necessary to presuppose a set of indefinables. This matter will be addressed in sections 4 and 5.

Otherwise non-predictive. Otherwise non-predictive definitions present descriptions that rely on devices which, while convenient to the lexicographer, explicitly hinder the ability for a user to predict appropriate range of use of the term. Such devices include the disjunction 'or', subjunctives 'may', 'might' and 'can', umbrellas like 'etc.' and 'esp.' and hedges like 'usually' and 'generally'. It is asserted in [2] that by resorting to devices like 'etc.', "...the lexicographer makes the definition untestable." The issue with such devices is that without criteria for the scope they intend to cover, it is impossible to know what is included and what is not. Thus a user is unable to predict whether the term is appropriate in a given situation.

Strategies for removing such devices are suggested by [15] and we will appropriate them in sections 5 and 6.

2.5 Synthesis of the Problem

Thus from the premises that words are not containers and that meaning must provoked rather than conveyed arises the problem of structuring a description such that the meaning described has the best chance of being correctly interpreted. Threats to the integrity of the description include obscurity, circularity, and the use of devices that render definitions otherwise non-predictive.

The task is to develop a structured and guided method of lexicography suited to the purposes of requirements producers, that takes into account both the more critical need to manage common definition faults, as well as the limits imposed by the necessity of practical application.

3. Related Work

Related work addresses a number of areas surrounding the problem and certain results provide bases we will appropriate for our goals. None, however offers a method sufficient for the task described above.

3.1 Domain Description Mechanisms

Careful understanding of the connection between a system and its application domain has been an object of attention for both Jackson and Heninger. The work of each has addressed to some degree the need for explicit content to be ascribed to some identifiers within a requirements specification or other project documentation. Heninger asserts the need for definitions of critical terms, but offers no advice on definition character or construction. [4]

Jackson goes further in detailing a class of description mechanisms, separating them into *designations*, *definitions*, *refutable descriptions* and *rough sketches*, each serving a different purpose for use in requirements and specification generally. [6] Of these mechanisms, the one that comes closest to serving the purpose we are here concerned with is a *designation*, or "...a recognition rule for recognizing some class of phenomenon that you could observe in a domain." [6] While a recognition rule would go far in achieving predictive power in a definition, no explicit guidance is provided for achieving the power of recognition. Further, there is no explicit mechanism to prevent designations from suffering the weaknesses of definition described in section 2.4.

3.2 Data Dictionaries

Data dictionaries provide descriptions intended to allow accurate prediction of the range of use of a set of identifiers. However, they are not concerned with the description of elements in the real world of the application domain. Rather, data dictionary entries pertain to named entities in a design model and the names with which they are associated. [12] The point of a data dictionary is explicit recording of the associations constructed within the modeled world, not primarily the accessibility of any real world meaning required to conceive of the model in the first place. Data dictionaries serve to manage namespace, for example, to codify naming and enforce consistent name use. [12, 11] They do not serve, other than incidentally, to provide domain or other meaning, as the glossaries of SRS documents are intended to. It should be noted, however, that the same structural issues in definition can arise in data dictionaries as well, and the method presented in this work can be likewise be exploited to construct data dictionary entries of greater value.

3.3 Lexicography and Lexicology

Lexicography is the study and activity of dictionary making, that is, the recording and collecting of representations of semantics for practical purposes, namely, introducing new word-meaning associations to readers lacking associated meanings for words in question. Lexicology (lexical semantics) is the study of semantics for linguistic

theory purposes, to make and test hypotheses about how and what humans understand via analysis and empirical study of the meaning of words. Both have an interest in the generation of a reliable theory and method of analyzing and representing word meanings, and requirements producers share this interest with these communities.

[15] suggests a gap exists between the knowledge possessed by lexicographers and lexicologists in service of their related goals and further, that lexicographers can learn a lot from lexicologists (recall the definitions provided in section 2.4) and still produce commercially viable dictionaries. It is an assertion of the present work that requirements producers lack even the tools of traditional lexicographers. While lexicographers at least have a wealth of experience in writing definitions and the commensurate ability to apply good judgment and founded intuition, requirements producers are limited to good intentions and common sense, varying knowledge of the domain, little to no experience in lexicography, potentially insufficient time to do the job, and arguably insufficient recognition of its importance.

Lexicology recognizes generally that there is often a need for better definition construction methods than are available, but it is not generally in the business of application. Work in lexicology has produced some interesting results regarding representation of the meaning of words, but these results require refinement for productive use in practical environments. One particular initiative that offers a number of potentially effective strategies for our purposes is the body of work on the Natural Semantic Metalanguage (NSM). Since we use this work as the basis for our own, we treat it in relative detail in the next section.

4. The Natural Semantic Metalanguage

The Natural Semantic Metalanguage (NSM) is the product of a research area in linguistics aimed at determining whether there exist primitive semantic units that are lexicalized universally across natural languages, and if so, what they are, and of what value it might be to characterize them. That is, NSM seeks to discover whether there are concepts innate enough to all humans that they can provide a set of *indefinable* building blocks out of which all more complex meanings can be constructed and represented. [3]

NSM makes use of the notion of a *metalanguage*, a language used to describe or represent another language. In a dictionary, the language used to define the entries is a metalanguage. The metalanguage of NSM is the set of semantic primitives that may be combined in building up more complex concepts, as well as the principles for their combination. To date, the NSM initiative has conducted "...intensive empirical and descriptive work on the semantics of a wide range of languages. This work has led to a set of highly concrete proposals about a hypothesized irreducible core of all human languages." [1]

The NSM initiative begins from the following premises:

- "[S]emantic analysis must be conducted in natural language, rather than in terms of technical formalisms [...], if only because technical formalisms are not clear until and unless they are explained in ordinary language." [1]
- "[T]he full meaning of any semantically complex expression can be stated in terms of a reductive paraphrase in ordinary language, i.e., an equivalent expression composed exclusively of simpler meanings than the original." [1]
- "It follows from these premises that every language must have an irreducible semantic core consisting of a mini-lexicon of indefinable expressions (semantic primitives) and a mini-syntax governing how they can be combined." [1]

On this hypothesis, the NSM project has conducted investigation into the membership and nature of the set of semantic primitives via empirical work in constructing legal reductive paraphrase *explications*. An explication is a special case of definitional description that results from the process of reductive paraphrase and thus consists only of terms simpler than that being explicated. The process of reductive paraphrase continues until terms are reached that can not be explicated. These become candidates for the metalanguage. To date, the metalanguage consists of approximately 60 concepts that are hypothesized to be primitive. Until candidate primitives are disconfirmed, they "are held to designate meanings which are impervious to (non-circular) definition and universal in the sense of having equivalents in all languages." [1] The 60 that are currently hypothesized have proven recalcitrant to disconfirmation throughout the work.

Thus the main elements of the NSM project are the process, reductive paraphrase, and the product, the metalanguage of semantic primitives. The goal they serve is a linguistically sound theory of word meaning.

4.1 Explication Example

To demonstrate the value of reductive paraphrase explications, we now analyze an example. The following is a full reductive paraphrase explication for the word 'broke'. The example shows only the end result and not the intermediate reductions.

X *broke* Y =
X did something to thing Y
because of this, something happened to Y at this time
because of this, after this Y was not one thing any more [1]

Note that the explication is textual and uses grammatical arguments (i.e. X and Y) to scaffold the paraphrase.

These arguments serve the RP doctrine of substitutability. This requires that for maximum effectiveness at descriptive adequacy of the concept in question, an explication must be able to stand in for its explicand without a change in meaning. Thus the arguments provide anchors by which a substitution can be effected. Given an utterance "Joe broke the glass", the explication should be able to substitute for 'broke', contextualized by arguments 'Joe' and 'the glass':

Joe did something to the glass
because of this, something happened to the glass at this time
because of this, after this the glass was not one thing any more

4.2 Explications vs. Definitions

Explications differ from traditional definitions in a number of ways. First, because of the substitutability criterion, "explications are essentially 'texts' composed of a specified subset of ordinary language. They are not 'lists of necessary and sufficient conditions', or 'bundles of features'" as are more common definition patterns and representations offered by other frameworks in lexical semantics. [1] In this way, descriptive adequacy can be easily checked through a determination of whether the substitution maintains the original meaning.

Most importantly, explications uphold the integrity of the description by design. A legal explication contains no obscurity since reduction is by definition to simpler terms. A legal explication contains no circularity because the reduction cannot reach the metalanguage if circularity is present. Finally, the metalanguage contains none of the devices that render a definition non-predictive and thus a full explication is free of them. The process of reaching the explication filters them out since the entities represented by, for example, an 'etc.' cannot be fully known and thus the 'etc.' in context cannot be reducible (of course, 'etc.' as the context-free head of an explication can be reduced in the normal way). The analyst is forced by the method not to lean on such devices.

Thus the NSM initiative shows that the issues of definition in section 2.4 can be systematically avoided in the construction of definitional descriptions, however, the investment required to do full reductive paraphrase on a non-trivial number of words is large. Further, it is not necessary for the purposes of requirements production to fully explicate. A representation in semantic primitives is more representation than is needed; reduction to a *relatively* primitive set would suffice. It remains, then, to determine whether a refined reductive paraphrase process can be developed that circumvents the definitional issues in section 2.4 while limiting the extent of the reduction commensurate with available resources and descriptive goals.

5. Defining Partial Reductive Paraphrase

The goals of the NSM project and the goals of requirements producers overlap in the desire of each to be able to systematically produce high-quality definitional descriptions of words. NSM achieves this through reductive paraphrase explication via the metalanguage of semantic primitives. Our first point of departure from NSM proper occurs with prioritization of the elements of the method. While NSM seeks to build a primitive set as much for the sake of its own intrinsic linguistic interest and value as for executing reductive paraphrases, our goal specifically is obtaining the comprehension value provided by the reduction. This requires that a metalanguage be generated as a side-effect, however, it is not the metalanguage in and of itself in which we are interested. In fact, we expect that the metalanguages generated will not be uniform across projects, while NSM specifically seeks one universal metalanguage.

5.1 Elements Reused from NSM/RP

To begin to define Partial Reductive Paraphrase (PRP), it is necessary to determine what we can use from full RP proper and what we must change or invent. The first notion we would like to use is that of reduction. It has been recognized elsewhere that a reduction need not be full to be useful: "It is not always necessary to resolve an explication right down to the level of semantic primitives. An explication can still be reductive--and still be valuable--even while containing some semantically complex terms, provided that none is more complex that the original term being defined and provided none needs to be defined in terms of the original word." [2]

The constraints on an incomplete reduction described above reference two other notions we would like to reuse from full RP. First, the provision that no term contained in an explication is more complex than the original term being defined ensures explications that are free of obscurity. Second, the provision that no term in an explication needs to be defined in terms of the original word ensures explications that are free of circularity.

Finally, by using a reductive approach, we will also benefit from the incompatibility between explications and devices that render definitions otherwise non-predictive. As explained in section 4.2, the entities represented by a hedge or other device cannot be determined and thus a valid reduction of such a hedge is impossible.

5.2 Elements Created to Define PRP

From full RP proper we take the notion of reduction and

its attendant structural barriers against obscurity, circularity and devices that render descriptions otherwise non-predictive. We leave the notion of exhaustive reduction to semantic primitives for reasons addressed in section 4.2. The implication from these decisions is that we must otherwise define a point at which to stop a reduction. Thus the additional infrastructure we have to create in order to satisfy our needs is systematic direction in scoping explications to a level we will define as *relatively* primitive. Relatively primitive is equated to the level of complexity at which intersubjective agreement about the meanings of words occurs spontaneously (without need of paraphrase or other description) among members of the speech community in question. In this case, the speech community in question is the community of consumers of an SRS as opposed to the community of producers, since the consumers are the community to which explications must be accessible.

So far we know that we want Partial Reductive Paraphrase (PRP) explications to be accomplished via a metalanguage of words for which the associations are intersubjectively agreed among the population in question. While this clarifies the definition of PRP, it is not instructive in recognition of relative primitives when they are reached in an explication. The strategy for knowing when to stop is detailed below.

5.3 Stopping the Reduction

It is impossible a priori to select a set of words and declare that they are relatively primitive for a speech community in question. There are no logical selection criteria that can be applied systematically. For example, if one were to claim that the basis for such a set might be all of the terminology, specialized and general, that a professional software developer is expected to know in order to be able to do his job, there are still no valid criteria by which one might ascertain this set.

Rather, we will take the premise that relative primitives are emergent as part of the process of PRP and our approach will be to allow candidates to emerge and attempt to disconfirm them as relatively primitive.

Specifically, PRP consists of an indefinite number of rounds to explicate a glossary, after each of which candidate primitives are evaluated, accepted, or subjected to further reduction. The point at which all candidates are accepted as relatively primitive is the stopping point of the reduction. The set of accepted relative primitives constitutes the metalanguage for the glossary.

Round 0 begins with the initial set of terms chosen to be explicated. This set can be chosen according to existing methods of term selection for glossary construction, which tend to be ad hoc and based on requirements producers' best guesses about what might be important. Alternatively, term selection can be accomplished via a more systematic and theoretically grounded method developed to complement PRP and addressed in section 8 and [13].

The first explication task is to attempt a reductive paraphrase for each term in the set. Recall this is not a full reductive paraphrase, but just a paraphrase in which the analyst attempts to use only simpler words. To attempt this reductive paraphrase, the analyst must have confidence in his own understanding of the meaning of the term; if this is not the case, an expert from the source domain should be consulted in this effort. Alternatively, confidence in one's understanding of the meaning of a term can be achieved via a systematic and theoretically grounded method developed to complement PRP and addressed in section 8 and [13].

All of the words used in the first set of explications constitute the candidate metalanguage for the glossary. At this point, it must be determined which candidates are in fact relatively primitive and which require further explication. That is, it must be determined whether the candidate terms have intersubjectively agreed meanings. This is analogous to the problem of selecting which terms to include in the glossary to begin with, and thus there are two ways in which this can be accomplished, one informal and intuitive, and one scaffolded by additional knowledge. The first entails the analyst's educated guess about what the SRS consumers know, and this guess can be strengthened through consultation regarding the list with representatives of this group. The second option is to use the facilities for term selection provided in the umbrella methodology of which PRP is a part (see section 8 and [13]) and apply these facilities to the new list as if it were the initial body for which terms in need of definition were to be selected. The terms that by either account are not in need of further explication enter the set of relative primitives (the metalanguage). Those remaining present for another round of explication. The rounds continue until no term needs further explication.

During the PRP process it will be necessary to both address structural issues as they arise and to perform tests to locate others. Obscurity is handled by the reductive nature of the paraphrase, and hedges and other devices that render definitions non-predictive present for correction as it becomes their turn to be explicated. The analyst is forced to clarify what is intended by the use of such devices or the reduction cannot continue. Obvious circularities can be detected as they are committed; an explication that uses any version of its explicand is disallowed. Insidious circularities require more complex measures for their detection.

To detect insidious circularities, it is useful to maintain a tree sketch indicating the dependencies among terms as they are explicated. Thus an explicand is a root and all of the terms upon which it depends are first entered as leaf

nodes. Of the terms that require further explication, their nodes become internal and new leaves are added to represent the terms upon which those terms depend. Any time a new term is added to the tree, it must be checked that it does not duplicate a node already present; this signifies an insidious circularity. If such a circularity is detected, the relationships among the terms in question need to be renegotiated and some explications might need to be redone. If no term needs further explication and no duplications were found, there are no insidious circularities.

Clearly the tree management activities described above can benefit from the provision of tool support. Such support exists for earlier related work and is detailed in [14]. Full support for management of explications, glossaries, and additional relationships among these entities and others is characterized in [13].

6. Executing Partial Reductive Paraphrase

In this section, we demonstrate PRP on a collection of concepts drawn from a set of industrial requirements.

6.1 Demonstration: Maritime Track Control

The terms explicated in this section are drawn from a set of requirements obtained from an industrial collaborator. The requirements describe an international standard with which maritime track control systems must comply in operation and performance. [5]

The terms chosen are ‘heading’, ‘bearing’, ‘track’ and ‘course’. They were chosen for explication because they have meanings specific to the domain of application as well as in everyday English, and thus they require explicit definition in order to prevent assumption on the part of non-experts that the everyday meaning is what is intended. Further, they have relationships to one another that allow better illustration of the method.

The first step in achieving PRP explications for these terms is an initial reductive pass:

X’s heading =
X’s direction as measured in degrees from due north
X’s bearing to Y =
direction to Y as measured in degrees from X’s heading
X’s track =
path in space that X is supposed to follow
X’s course =
series of tracks that lead to X’s destination

These explications provide a first indication of the relationships among these terms. Note that the explication for ‘bearing’ uses ‘heading’ and that the explication for ‘course’ uses ‘track’. Further, both ‘heading’ and ‘bearing’ use ‘direction’, ‘measured’ and ‘degrees’. Before going on, we first check for obvious circularity and there is none, and construct tree sketches that will allow the discovery of insidious circularities as they grow. There are no hedges or other devices to correct, nor is there any apparent obscurity, as the method is biased against its entering.

Next, we choose from among the terms used in the explications which to present for further explication. It is decided in consultation with a representative member of the population of potential glossary users that ‘due north’ is the only term in need of further explication. Thus:

due north =
direction to the north pole, defined to be 0 degrees

To the relative primitive set we add any terms in the above explication not already accepted to be relatively primitive, that is, we add ‘north pole’, ‘defined’ and ‘degrees’ (we have already discounted prepositions, numerals, etc.). Thus the metalanguage for this mini-glossary consists of these terms plus all terms in the previous explications with the exception of ‘due north’. Were we to want to enlarge the glossary, the explication of a new term could be checked against this set for determination of which additional terms might need explication. The glossary can be enlarged in this manner indefinitely, as long as the tree is maintained such that insidious circularities cannot enter and the required validation steps are taken on every iteration.

7. Evaluating Partial Reductive Paraphrase

The PRP explications elaborated in section 6 can be compared with the analogous definitions produced by the original means used in construction of the glossary that actually accompanies the maritime track control requirements. This comparison provides a clear demonstration of the value of PRP for managing the challenges of definition writing as described in section 2.4.

7.1 Term 1: Heading

The definition for ‘heading’ from the glossary of the original document is given as: “The horizontal direction in which a ship actually points or heads at any instant, expressed in angular units from a reference direction, usually from 000 degrees at the reference direction clockwise through 360 degrees.” [5]

The definition as provided suffers from obscurity via confusion of knowledge with meaning as described in section 2.4, and possibly also through specific terms such as ‘reference direction’. It also makes use of devices that render it non-predictive, e.g. ‘usually’. Further, it contains an instance of obvious circularity through the use of ‘heads’ to define ‘heading’.

By contrast, the PRP explication has none of these

weaknesses and meets the criteria set forth in [2] for descriptive adequacy.

7.2 Term 2: Bearing

The glossary of the original document contained no definition for 'bearing', and so we must compare the presence of an explication with the absence of any definition at all.

In this case the document provides no guidance in understanding the range of use of the term, while by contrast, the PRP explication meets the criteria for descriptive adequacy.

7.3 Term 3: Track

The definition for 'track' from the glossary of the original document is given as: "Path to be followed over ground." [5]

While this definition contains no instances of the particular definition problems about which we are concerned, it does violate what is termed a *logical entailment* in lexical semantics. If the object to which this system relates is a ship, this definition violates what the requirements users are likely to know about ships, in particular that they move over water.

The PRP explication contains no instances of the definition problems described in section 2.4 either, nor does it violate any logical entailments.

An alternative analysis for this term holds that 'over ground' is actually a domain specific term and should be explicated as such. In this case, a descriptively adequate explication can be produced, but the original definition would be rendered obscure. It is of note that in consultation with an expert, it was decided that the particular medium was immaterial to the meaning of the term. Thus 'space', as used in the explication, sufficed.

7.4 Term 4: Course

The definition for 'course' from the glossary of the original document is given as: "For marine navigation, course is the horizontal direction in which a vessel is steered or intended to be steered, expressed as angular distance from north, usually 000 degrees at north, clockwise through 360 degrees." [5]

The definition as provided suffers from obscurity via confusion of knowledge with meaning as well as through specific terms such as 'vessel' and 'angular distance'. It also makes use of devices that render it non-predictive including 'or' and 'usually'. Further there is the possibility of an insidious circularity since while 'steer' does not have its own entry in the original glossary, a version specific to track control would intuitively make use of the concept of 'course' (and a PRP explication would employ it as such).

In contrast, the PRP explication has none of these weaknesses and meets the criteria for descriptive adequacy.

7.5 Synthesis

PRP explications are shown to eliminate the problems of circularity, obscurity and devices that render definitions otherwise non-predictive. In each of the cases examined, the explications achieved descriptive adequacy and did so using resources appropriate to the constraints of an industrial requirements process. While no data are available on the investment necessary to construct the original definitions, the author invested less than two person hours in doing the explications for the terms addressed, including drafting and revision of the explications and the circularity management tree, as well as conferring with both an expert and a non-expert at the necessary points in the precess.

8. Integrating PRP into the Requirements Process

PRP can be used as presented at the point in any requirements process when terms must be defined. It need not create or adjust dependencies that add complexity to processes as they are currently practiced in industrial environments.

However, it was indicated in section 2.2 that PRP is part of a greater requirements methodology that also takes into account the integrity of the described, to complement the value of PRP to maintain the integrity of the description. Further, it was indicated in section 5.3 that more systematic alternatives exist for both selecting the input to PRP as well as for identifying the correct conception of an entity in order to explicate it (in further support of the integrity of the described).

This greater methodology is CLEAR, a requirements method under development at the University of Virginia. [13] CLEAR, Cognitive Linguistic Elicitation and Representation, exploits results from linguistics and cognitive psychology in the provision of well-defined methods and artifacts for capturing, storing and representing domain knowledge such that it is accessible to the consumers of requirements. CLEAR enables the construction of an organized repository of domain semantics critical to the production and validation of software systems.

CLEAR has three phases: pre-processing, elicitation, and representation. Pre-processing effects the selection of the phrases under consideration for explication in the repository. Elicitation effects collection of the semantics intended by the producers of requirements documents to be

associated with these phrases, that is, it addresses the issues of accuracy and precision of the content to be represented. Representation effects the organization and presentation of the collected semantics such that the phrases are most likely to provoke the intended meanings. The representation phase integrates the content collected via elicitation with the structure provided by PRP to maintain the integrity of both the described as well as the description.

9. Summary

Though glossaries are necessary components of SRSs, guidance for the construction of definitions within them is unavailable in the software and requirements research and instruction literature. The default methods generally employed produce definitions that suffer from a number of critical flaws. These flaws impair the ability of requirements consumers to access the intended meanings of concepts necessary to valid implementation of software systems.

The Natural Semantic Metalanguage project provides a method, reductive paraphrase, for explicating any concept in any language such that these flaws are circumvented. Full reductive paraphrase is both too resource-intensive and exhaustive for the needs of requirements producers and consumers, however, it is adaptable in ways that meet our resource constraints while exploiting the power of full reductive paraphrase.

We developed Partial Reductive Paraphrase (PRP), a refinement that employs the value of reduction and provides direction in managing faults in definitions, within the constraints implied by application in industrial environments. PRP guides a requirements producer in the construction of definitions free of specific faults known to hinder correct interpretation. It was demonstrated that PRP explications compared favorably to analogous definitions produced by other means, and it was further addressed how PRP can be integrated into a greater requirements process.

10. Acknowledgements

This work was funded in part by NASA under contracts NAG-1-02103 and NAG-1-2290, and by NSF under contract CCR-0205447.

11. References

[1] Goddard, C., “Bad Arguments Against Semantic Primitives,” *Theoretical Linguistics*, 24(2-3):129-156, 1998.

[2] Goddard, C., Semantic Analysis, Oxford University Press, Oxford, 1998.

[3] Goddard, C. and Wierzbicka, A., eds., Semantic and Lexical Universals, John Benjamins Publishing Company, Amsterdam, 1994.

[4] Heninger, K., “Specifying Requirements for Complex Systems: New Techniques and Their Applications”, *IEEE Transactions on Software Engineering*, SE-6(1):2-12, 1980.

[5] International Electrotechnical Commission, Maritime Navigation and Radio Communication Equipment and Systems - Track Control Systems, “Operational and Performance Requirements, Methods of Testing and Required Test Results,” Project number 62065/Ed. 1, 2000.

[6] Jackson, M., Software Requirements and Specifications, Addison-Wesley, Wokingham, 1995.

[7] Lakoff, G., Women, Fire and Dangerous Things: What Categories Reveal about the Mind, The University of Chicago Press, Chicago, 1987.

[8] Longman Dictionary of the English Language, Longman, London, 1984.

[9] Merriam-Webster’s Collegiate Dictionary (10th ed.), Merriam-Webster Incorporated, Springfield, 1993.

[10] Oxford Paperback Dictionary, Oxford University Press, Oxford, 1979.

[11] Pressman, R., Software Engineering: A Practitioner’s Approach (5th ed.), McGraw-Hill, Boston, 2001.

[12] Sommerville, I. and Sawyer, P., Requirements Engineering: A Good Practice Guide, John Wiley and Sons, Chichester, 1997.

[13] Wasson, K., The CLEAR Requirements Method: Cognitive Linguistic Elicitation and Representation, Dissertation, The University of Virginia, 2004 (forthcoming).

[14] Wasson, K., Knight, J., Strunk, E., and Travis, S., “Tools Supporting the Communication of Critical Domain Knowledge in High-Consequence Systems Development,” Proceedings: 22nd International Conference in Computer Safety, Reliability and Security (SafeComp2003), Springer-Verlag, New York, 2003.

[15] Wierzbicka, A., Semantics, Oxford University Press, Oxford, 1996.