CLEAR Requirements: Improving Validity Using Cognitive Linguistic Elicitation and Representation

A Dissertation
Presented to
the Faculty of the School of Engineering and Applied Science
University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy (Computer Science)

by

Kimberly S. Wasson

May 2006

'We must do something' is the unanimous refrain. 'You begin' is the deadening refrain.
--Walter Dwight

Abstract

Effectively communicating requirements is necessary to the success of every software system, especially those that are safety-critical or otherwise high-consequence. High-consequence systems demand the most valid requirements achievable, but validity is compromised by habits and limits of the human linguistic endowment. Certain properties of this endowment interfere with the creation and transfer of knowledge necessary to produce such systems, making effective communication of requirements very difficult to achieve. This work presents and evaluates a theoretically-grounded methodology for improving communicative fidelity in requirements environments.

Existing methods for communicating requirements are based on ad hoc and intuitive approaches. These approaches are able to neither predictably identify nor demonstrably correct faults that allow miscommunication.

Cognitive Linguistics provides a foundation of results that can be used to both explain undesired phenomena observed in requirements environments and direct how they might be avoided. It elaborates cognitive states that must exist and events that must take place in order for communication to be pragmatically sufficient. In everyday life, such states and events arise spontaneously. In environments that require the accuracy and precision necessary for the correct implementation of high-consequence systems, and where communication is often across domain boundaries, they generally do not.

Cognitive Linguistic Elicitation and Representation (CLEAR) is built on the proposition that these necessary states and events can be chaperoned into existence and occurrence in environments where their spontaneous creation is not the common case. CLEAR accommodates natural human communicative habits and limits by defining, and guiding the construction of, explicit processes and artifacts that allow necessary cognitive states to be achieved and necessary events to occur.

CLEAR was evaluated using a two-part approach designed to achieve both experimental control and environmental realism. A formal experiment first demonstrated the value of central CLEAR activities and products with statistical significance. A larger-scale case study of its application to an actual safety-critical medical device in development then both replicated earlier results as well as extended them.

CLEAR makes three main contributions to the study and practice of requirements engineering. First, it provides a linguistically-grounded analysis of communication deficiencies and their sources. Second, it provides a mechanism for systematic detection and correction of several classes of such deficiency. Finally, analysis of the products resulting from CLEAR provides for the first known mechanism for quantitatively estimating the miscommunication risks of a project's critical notions, thereby offering guidance in the allocation of resources for the preservation of their integrity throughout a project's duration. Together, these contributions characterize an advancement beyond the current state of the art in achieving high-integrity communication for requirements engineering.

APPROVAL SHEET

This dissertation is submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy (Computer Science)

Kimberly S. Wasson (Author)

This dissertation has been read and approved by the examining Committee:

John C. Knight (Advisor)

Worthy N. Martin (Chair)

David E. Evans

Jack W. Davidson

Mark J. Elson (Minor Representative)

Accepted for the School of Engineering and Applied Science:

James H. Aylor (Dean,
School of Engineering and Applied Science)

May 2006

Acknowledgements

This point is never reached alone, and a host of individuals supported and encouraged me in many ways.

The extended SIGBeer universe is the coolest group of people I could ever have the privilege of calling my friends. My early days in CS were tickled by game nights, happy hours and random debates with clever, intelligent and funny fellow travelers. Thanks Doug Blair, John Regehr, John Karro, Sean McCulloch and Gabe Ferrer, among others. My later CS days have been warmed by SIGBeer all-grown-up, through dinner parties, wedding receptions, baby showers and play dates. Thanks to the Coppits, Englers, Beekwilders, Hopes, Natrajans, Nguyen-Tuongs and Karpoviches. And of course “the boys”, Mark Morgan and John Jones.

A number of department staff were indispensable. Ginny Hilton, Brenda Perkins, Peggy Reed and especially Kim Gregg have between them always had every answer I ever needed to any administrative or financial question. The systems staff have handled my equipment emergencies and non-emergencies with efficiency and friendliness.

The National Science Foundation and the National Aeronautics and Space Administration generously supported this work through a number of grants, and the personnel of the University of Virginia Walker Project gave me their time, materials and expertise, allowing me to evaluate my method on a real application.

Robyn Lutz and Steve Easterbrook provided thoughtful feedback, advice and mentoring. Robyn stretched and bent to accommodate my constraints when our collaborations overlapped my maternity leave. A number of other RE denizens have offered a variety of adventures on a yearly basis.

My dissertation committee was administratively cooperative and intellectually insightful. Mark Elson went above and beyond the call of duty and has provided me guidance in some form or another for 14 years. John Knight went *way* above and beyond the call of duty, absorbed my tantrums and misgivings without letting me off the hook, and saw me through to the end.

Phinished (www.phinished.com) reflected back to me a community of people just like me, unintentionally doing this the hard way, and helping each other lighten that unnecessary load. Finding them two years ago was a key turning point.

Glenn gave, and gives me every day, a sounding board, comfort, courage and stamina. Sofia taught me how to do what needs to be done. Together, they give me the greatest joy I have ever known, and with the perspective that brings, made the dissertation tractable.

Table of Contents

Don't ever make the wrong mistake.
--Yogi Berra

1 Introduction

Software systems that function other than as intended threaten lives, property and societal well-being. This work concerns some of the challenges inherent in transforming an informal idea into a software system that fulfills intentions. Specifically, it examines the mechanisms by which system requirements are communicated among stakeholders of a system, and how these mechanisms can be improved.

The majority of faults and failures in high-consequence systems can be traced to deficient requirements [38, 61]. Faults and failures indicate that a system does not meet the purpose for which it is intended, and thus, is not a valid system. Validity is a relation on intentions and instantiations. Therefore for a system to be valid, critical information must be communicated with fidelity between at least two parties: those that generate requirements, that is, those who formulate intent, and those that consume them, that is, those who execute downstream stages of a development process to instantiate a software artifact. If this communicative fidelity is lacking, the validity of a system is at risk.

This work employs results from disciplines that study communication to explain the mechanisms by which communication breaks down in requirements engineering. It further describes and evaluates a methodology that provides procedural support for the elicitation and representation of software requirements in which specific threats to communicative fidelity are demonstrably reduced. It contributes to the theory and practice of requirements engineering specifically, and thus to the theory and practice of computer science generally.

The remainder of this chapter establishes the context of and basis for the work and previews the methodology and contributions.

1.1 Scope of the Work: Software Requirements Engineering

Requirements engineering (RE) is the first stage in a canonical software development process. Its activities are concerned with finding out and recording what the system under development is supposed to do. RE is characterized by specific goals, as well as commonly performed tasks constructed to achieve those goals. RE is further characterized by challenges in the execution of those tasks and thus in the realization of the goals the tasks support.

1.1.1 Goals of Requirements Engineering

The foremost goal of the RE component of software development is to provide for system builders an adequate understanding of a system to be built. "Adequate" indicates a level of understanding that supports the construction of a valid system. This means that at the core of RE is the necessity of discovering, creating and disseminating knowledge among parties with various roles in relation to the system, such that each has the understanding necessary to make his contribution.

The goal of knowledge assembly and transfer is realized through execution of a number of common tasks by the participants in the development process. These tasks include:

- the acquisition and/or elicitation of needs from clients
- the acquisition and/or elicitation of domain knowledge pertaining to those needs from experts and/or clients
- analysis and modeling of the functional and non-functional activities and properties that the system, as elicited from the clients and experts, must express
- negotiation among different stakeholders in the system regarding prioritization of these activities and properties
- validation of the analyses, models and prioritizations arrived at, and
- recording of the requirements in forms supporting these tasks throughout their execution and the remainder of the development process.

The execution of these tasks may be explicit or implicit, according to the goals, resources and level of maturity of an organization. In addition, these tasks are subject to iteration as new information is learned and new decisions are made.

1.1.2 Challenges to Achieving Goals of Requirements Engineering

Brooks has stated that "[t]he hardest single part of building a software system is deciding precisely what to build" [8], as requirements provides many of the most recalcitrant challenges in all of software development.

The most significant challenge faced by agents of the requirements process is achieving validity. Validity refers to whether a software artifact intended for a given purpose is "the right thing," that is, whether it fulfills its intended functionality under intended constraints and without unintended consequences.

More abstractly, validity is a relation on any intention and any instantiation of that intention, whether the instantiation be in the form of requirements documentation, formal specification, design, source code, or a running system. The validity of intermediate instantiations is a precondition to the validity of a deployed system.

Achieving validity is a challenge because as a measure of the goodness of a mapping to human intent, it is necessarily an informal property. Personality, politics, knowledge and skill sets, and the organic bases of human cognition all introduce potentially arbitrary variation into the fidelity of a given mapping, through the variation they introduce into the tasks that contribute to that mapping. In other words, the influence of these variables on the tasks of elicitation, analysis, and recording, etc., translates to influence on the provision of the necessary understanding of the system to its builders. The number of variables and the challenges and impossibilities in controlling or measuring them render validity very difficult to test for and demonstrate with assurance and rigor. Validation as practiced is thus more an art than a science.

1.2 The Target Challenge: Communication

Primary among the complex variables that affect our ability to achieve validity is communication. Since the goal of RE is to provide for system builders an adequate understanding of a system to be built, RE fundamentally relies on communication, because what the builders have to understand is a product of the intenders, and this intention must be communicated in order for the understanding to be achieved. All of the tasks necessary to achieve this goal, and all of the knowledge assembly and transfer involved in accomplishing those tasks, is dependent on informal ideas being communicated among human agents.

Communicative transactions take place among all sets of stakeholders in a software development process, and all transactions have some effect on the validity of the product. However, of heightened interest in this work are the communicative transactions that take place between producers of requirements and consumers of requirements, since these are the transactions upon which the goal of requirements mainly rests. “The success of the design process depends critically on the designers' understanding of the problem” [14], and the problem to be understood is that as posed by the requirements producers.

1.2.1 Deficient Communication

Deficient communication is characterized by any of a number of properties that compromise understanding of the message, in this case a set of requirements materials. Properties often cited in the literature as markers of deficient requirements include ambiguity, incompleteness, inconsistency and general obscurity. Though recognized as a problem, these deficiencies have proven recalcitrant to straightforward correction; some existing approaches and their merits and limits will be addressed in Chapter 3.

In the common case of everyday language and communication, such deficiencies are well-tolerated. This is in part because our cognitive heuristics are tuned, for example, to fill in missing information pragmatically and to resolve inconsistencies, and in part because our needs for accuracy and precision are not generally exceedingly high.

However, our needs for communicative accuracy and precision in requirements environments are uncommonly high, higher than our cognitive endowment is naturally built to handle. Further, our needs for cross-domain communication, that is, communication between agents with varying levels of expertise with regard to a domain, are also uncommonly high, another purpose that our innate mechanisms do not serve well. Thus reliance on heuristics that work for the common case is an insufficient strategy for communication in requirements environments. In addition, the number of opportunities currently afforded to miscommunicate is unnecessarily multiplied by the implicit and ad hoc methods by which many requirements tasks are accomplished. Often the same information is requested and provided numerous times and transacted between multiple sets of parties. This further favors the influence of natural variation and hinders our goals.

Thus communicating semantics with sufficient fidelity in requirements environments is very hard. However, it is also critical to the reliability of high-consequence systems, as well as to the efficiency of the software process. We next examine the consequences of deficient communication.

1.2.2 Consequences of Deficient Communication

Poor communication of requirements contributes to faults and failures in the software systems they describe. Faults and failures in software systems, especially high-consequence ones, threaten safety, property, and the dependability of critical infrastructures, and are further a source of unnecessarily high costs due to rework and project failure.

Errors introduced into software during the early stages of the lifecycle pose two significant challenges to the development of high quality systems. First, errors introduced early are correspondingly more difficult and expensive to correct [4, 42]. In the worst case, they are left uncorrected and for some systems can result in significant loss. Second, there are more of them [4, 46]; the requirements stage in particular is implicated as the locus of introduction of more defects than any other. One study found that the majority of safety-critical defects in the systems under examination derived from poor requirements [38], and another found that the majority of *all* defects found were created during the requirements stage [61].

Results such as this suggest that substantial reductions in safety-, mission- and infrastructure-critical defects, as well as in costs due to rework and process inefficiency, might be achieved if deficiencies in requirements could be reduced.

1.2.3 Implications for Research

To a first approximation, the greatest improvement to high-consequence software quality and to the efficiency of the development process is to be gained if a substantial advance in requirements validity can be achieved. Thus to a first approximation, the problem of reducing requirements errors is equivalent to the problem of improving communicative fidelity in requirements environments.

The difficulties that must be managed for this information to be communicated with the necessary accuracy and precision and across domain boundaries are beyond the capabilities of our main strategy for communicating it: natural language. However, natural language must be used; whether or not any of various techniques, for example, formal methods using languages such as Z and VDM, or semi-formal ones using, for example, UML, are then applied to aid in structuring the information and making it easier to manipulate, the initial transfer and any negotiation of a subsequent representation is accomplished through the medium of natural language. We must therefore find a systematic and predictable way to better wield natural language as the primary tool of communication for requirements.

Thus, arguably the most significant barrier to the improvement of poor requirements is our limited ability to effectively manage the use of natural language. If we could improve this state of affairs, we could realize not only a reduction in rare but catastrophic failures of high-consequence systems, but importantly, we could also realize an improvement in the overall efficiency of producing such systems, as gained through a reduction in rework, a reduction of schedule and cost overruns, and a reduction of failed projects that never enter deployment.

The implication is that if we could improve communication, we could improve requirements validity. More specifically, if we could systematically accommodate human cognitive habits and limits in the use of natural language, and thus demonstrably reduce the potential for misunderstandings, requirements validity would be more readily achieved.

1.3 Solution Strategy and Preview of Contributions

In order to systematically accommodate human cognitive habits and limits in the use of natural language, this work takes the following approach. First, we recognize that despite the informalities and complex variables that characterize the human element of requirements, there is actually significant regularity to human behavior and cognition, especially in the use of natural language. Then, given the presence of this regularity, the question becomes, is it characterizable enough to exploit?

The strategy embodied by the solution presented in this work is to draw on previously untapped resources in guiding our understanding and management of the requirements communication problem, namely, research results from the General and Cognitive Linguistics literature. These results allow the communication problem to be broken down to lower level mechanics that can be examined and treated more directly.

1.3.1 Nature and Use of Linguistics Resources

Linguistics is the study of language as a scientific object. It provides theories, observations, and validated results describing the structure and use of natural language by humans. Of particular value to us is that it identifies cognitive mechanics that characterize, at a low level, the communicative activities of humans. These mechanics not only explain how it is that language functions, but in addition, and more importantly for us, how it breaks down. Our current lack of understanding of how communication breaks down precludes us from addressing this phenomenon directly; understanding of some of the mechanics can allow more strategic construction and application of accommodations.

The theories and results offered by linguistics describe the abstract entities and processes manipulated by humans in the process of achieving communication. Access to the organization and behavior of these entities and processes allows us to characterize more precisely what happens when two parties undertake and execute a transaction. Whereas at a higher level of examination, we might recognize the emergent property that a transaction has failed or was achieved with insufficient fidelity, a lower-level view allows further specificity regarding which entities and processes contributed to this outcome and in what ways. In the first case, intuition might suggest additional care, but in the second, directed compensatory actions become definable. Thus we will use linguistic theory to bring into better focus the building blocks of high-fidelity communication and direct creation of the accommodations that allow us to achieve it. These activities are integrated into a methodology called CLEAR: Cognitive Linguistic Elicitation and Representation.

1.3.2 Goal in Using Linguistics Resources

Ideally, we would like to create a situation in which the difficulties in RE that derive from properties of human communication are completely eliminated. Ultimately, achieving this goal is impossible (except in the degenerate case of giving up the goal of building software). Humans and their language are organic, stochastic systems and their inherent unpredictability in an absolute sense cannot be controlled. This leads to the essential indeterminacy of meaning: the fact that a given instance of meaning is a product of one individual's consciousness, and is shaped by that individual's experience, personality, skills and other environmental factors. The essential indeterminacy of meaning means that we can never know with absolute certainty, for example, what a requirements producer intends by a given statement, or what a requirements consumer interprets in response to it.

However, while meanings as conceived by requirements producers and consumers are unpredictable and uncontrollable in the absolute sense, we assert that the degree of possible variation between what a producer intends and what a consumer interprets can be demonstrably narrowed. While natural language is not formally regular, it *is* rule-governed. Our goal will be to use linguistics results about the regularity that *is* present to guide construction of methods that define a particular type of communicative transaction. This type of transaction will be characterized by our ability reliably demonstrate the presence of properties in it, that limit the potential for, and the effects of, variation in important and desirable ways. Our method will in effect provide a scaffold for the systematic construction of transactions that reliably avoid particular and otherwise common sources of communicative deficiency.

1.3.3 Thesis Statement

Our solution strategy distills to the following thesis for the work:

Communication between requirements producers and requirements consumers can be practicably improved to a level where desirable properties necessary to communicative fidelity are demonstrably present.

1.3.4 Preview of Contributions

The CLEAR Method makes three main contributions to the study and practice of requirements engineering. First, it provides a linguistically-grounded analysis of requirements communication deficiencies and their sources. Second, it provides a mechanism for systematic detection and correction of several classes of such deficiency. Finally, analysis of the products resulting from CLEAR provides for the first known mechanism for estimating the miscommunication risk of a project's critical notions, thereby offering guidance in the allocation of resources for the preservation of their integrity throughout a project's duration. CLEAR's use of a theoretical foundation to analyze and reduce the incidence of communicative deficiencies and to build a model of the semantic complexity of a domain is shown in the remainder of this work to constitute an advancement beyond the current state of the art in achieving high-fidelity communication of requirements for software systems. In so doing it advances the state of the art of both software engineering specifically and computer science generally.

1.4 Organization of the Work

The remainder of the work is organized as follows:

Chapter 2 elaborates the details and boundaries of the problem of improving communicative fidelity in requirements engineering. It addresses elements of the problem that require clarification and definition in order to render the problem actionable.

Chapter 3 addresses the related literature with regard to managing communication in requirements engineering. It assesses the current state of the art and asserts that the problem as posed remains open in the community.

Chapter 4 introduces the theoretical bases that allow explanation of the miscommunication phenomena observed and provide the foundation for components of the methodology. Results from Linguistics and Psychology together ground the majority of CLEAR, while areas of Information Retrieval are also called upon in service of some support activities that CLEAR employs.

Chapter 5 presents the CLEAR Method in detail, elaborating the four component phases of the methodology. The constructs and activities associated with each phase are defined, and their derivation from the theoretical foundation is traced.

Chapter 6 describes empirical assessment activities undertaken to demonstrate the value of the methodology. The CLEAR Method is applied in two separate environments, one enabling experimental controls, the other providing environmental realism. Together they enable replication of results as well as a broader picture of the influence of the method than either study does alone.

Chapter 7 reviews the methodology, synthesizes findings and summarizes the contributions of this work. It further sets forth an agenda for future activity that will extend community understanding of the role of communication in engineering and promote the development of additional methods and tools that can be usefully applied in industry.

Then anyone who leaves behind him a written manual, and likewise anyone who receives it,
in the belief that such writing will be clear and certain,
must be exceedingly simple-minded.
--Plato

2 The Problem

In the previous chapter, we introduced the problem of improving communication among parties in a requirements process such that more valid requirements can be achieved. In this chapter, we define and clarify elements of this problem in order to render it actionable.

Chapter 1 argued that the problem of reducing requirements errors is to a first approximation equivalent to the problem of improving communicative fidelity in requirements environments. To make a contribution to the improvement of communicative fidelity in requirements environments, this work argues the thesis that communication between requirements producers and requirements consumers can be practicably improved to a level where desirable properties necessary to communicative fidelity are demonstrably present. Several elements of this statement require elaboration in order that the thesis may be argued.

First, communicative fidelity must be defined with parameters that allow characterization of potential improvements. Second, desirable properties that contribute to these improvements must be defined such that their existence may be demonstrated in a solution construct. Third, criteria for practical utility must be presented such that a candidate solution can be evaluated with respect to them.

The characterization of potential improvements and the identification of contributing properties can be advanced through examination of the linguistic mechanisms of communication. In addition, properties of requirements environments interact with these mechanisms to provide further constraints to the kind of communication improvement we seek. Environmental issues also impose qualifications on the methods proposed to achieve such improvements.

In the remainder of this chapter, we elaborate several influences that parameterize the problem statement and thus shape candidate solutions. These influences are the characteristics of requirements environments that impose particular needs on communication that must be met by a solution, linguistic phenomena that characterize the notion of a communicative transaction and thus provide an object of improvement, and methodological issues giving rise to practical constraints that a candidate solution must possess in order to have potential for industrial adoption.

2.1 The Needs of Requirements Environments

This section addresses properties of requirements environments that impose particular needs on communication and thus qualify the kind of improvement we seek.

Communication in requirements environments shares some properties with everyday communication and is distinguished by others. While these two environments employ the same cognitive substrate to fulfill the same goal of transfer of thoughts between parties, they are distinguished primarily by the fact that in the common case, our communicative needs are sufficiently met, while in requirements environments, they are not. Thus, in order to strategize for improvement, we must examine the factors of requirements environments that impede function of that natural cognitive substrate.

Incident to any communication event are communicating parties, information to be communicated, and activities that instantiate a communicative transaction involving these parties and this information. These entities are not specific to any environment, but simply indicate the phenomena in play when communication takes place.

Specific to some environments, however, are qualifications on these entities that result in emergent phenomena when the entities interact. In requirements environments, each of these entities is qualified by particular properties that together, render communication within this environment more challenging than in the common case.

The qualifications on these entities that separate requirements environments from the common case are as follows. First, the parties to communicative transactions in these environments vary widely in background and expertise. It is virtually guaranteed in most software projects of any complexity that an application domain expert, for example, an avionics engineer (without software training), will have to communicate productively in some way with a software development expert (without avionics training). Further, the information to be communicated is very often application-dependent and therefore domain-specific. A software artifact is only useful insofar as it fulfills its intended role within a greater system. This relationship to a system context implies an application domain, which in turn implies application domain knowledge that must be communicated with sufficient fidelity to non-experts with respect to that domain, for example, developers. Likewise developers must be able to communicate back to application domain experts regarding, for example, limitations of what can be done in software with respect to the application needs.

The emergent fact of such transactions is that they can be characterized as trans-boundary. This means that experts in different domains must communicate with each other such that each achieves a necessary understanding of concepts originating in either's domain, that they must both ba able to productively manipulate without unintended consequences. This inevitability does not characterize the common case of everyday communication.

Further, in addition to trans-boundary communication, requirements environments impose greater needs for accuracy and precision in the message than does the common case. Requirements communication provides the foundation for instantiation of an informal idea in a digital system and as such it must be sufficiently precise. These digital systems are built to serve purposes and as such they must model relevant parts of the world accurately or risk not meeting these purposes. As discussed in Chapter 1, exactly the right thing must be built or tremendous cost in many forms might result.

Properties that are specific to requirements environments therefore interact with our natural cognitive processes in ways that have thus far limited our ability to communicate sufficiently to realize our engineering goals. Since our environmental needs are not met by the normal processes with which we are equipped, one constituent of the improvement we seek can be viewed as a complement of custom equipment. This equipment must serve as a scaffold between the communicative results we get naturally and those we need in order to achieve our goals in these environments.

2.2 The Object of Improvement

The complement of custom equipment indicated in the previous section is required because our innate communication mechanisms are not sufficient to meet the needs of requirements environments. This raises the question of the nature of what we *are* able to achieve naturally, how we do it, and how we might use this information in order to direct accommodations. The need for custom equipment implies the need to understand what we are customizing, that is, the need to determine what is fundamental to the system and which parts of it we want to enhance or change. To better understand what we want to improve, we first review some of the basic principles of human communication. Later in Chapter 4, we elaborate the low-level cognitive mechanisms of interest upon which our methodology is based.

Effective communication in any but the least consequential environments is hard. Natural language has such power and flexibility that there are any number of ways to get some piece of information across, and at least as many ways to render that transfer deficient. Deficiency is a function of two properties of a communication event. Any communication event has both purpose and mechanism, and if the mechanism as invoked does not meet the purpose as conceived by the communicating parties, then the communication is deficient.

The purpose of communication is to effect behavior in another party by causing him to have a thought. Often, parties have a shared goal, and they must provoke and build on thoughts collaboratively in order to attain this goal. In the context of this work, the goals of communicating parties involve rendering individuals capable of further action incident to a development process, by providing them with necessary information they otherwise do not possess.

The mechanism by which this takes place is the communicative transaction, a collection of cognitive operations that allow a thought to be presented by one party for interpretation by another. Complementary operations must be executed successfully by each party in order for a transaction as a whole to be successful. Thus if a communicative transaction is executed among parties in a development process and that transaction fails to provide the information necessary to enable appropriate individuals to undertake appropriate further action, then the communication has been deficient.

In the remainder of this section, we will first examine the structure of a communicative transaction and then consider how this mechanism might fail to serve its purpose. Next, the linguistic principles that interact to influence its fidelity are introduced, pragmatic sufficiency and the linguistic sign. From this discussion are derived a pair of desirable properties of transactions that a candidate solution to the requirements communication problem should support.

2.2.1 The Communicative Transaction

Since the purpose of communication is to effect behavior in another party by provoking him to have a particular thought, we now consider the mechanism by which such provocation is possible, the communicative transaction.

A communicative transaction consists of a complementary set of cognitive operations executed between two parties to achieve a communicative goal. The complementary sets of operations correspond to the complementary roles the parties fulfill; in any given transaction, one party is the source of the thought to be communicated and the other is the target in which the thought is to be provoked. The operations allow the thought to be given a representation by the source party, that is then made available for interpretation by the target party. The operations of the source party canonically transform a mental conception into a physical signal given through speech, that the target perceives aurally and then interprets back into a mental conception, ideally matching that which was intended by the source.

Alternatively, speech and aural perception may be replaced by various other output and input channels, for instance hand signals and visual perception of those signals as in sign language. Note further that a transaction need not come in the form of dynamic conversation; any construction crafted by one party for the purposes of provoking a thought in the mind of another, combined with the interpretation of that construction, is a communicative transaction. In the environments we are concerned with, the most common channels are written recordings of representations, that are at some later point processed visually, since our focus is on the requirements documentation materials generated as part of a development process, and their interpretation by parties downstream in that process.

If a provoked thought does not sufficiently match that intended by the source, then the communicative transaction lacks sufficient fidelity. Fidelity of a communicative transaction is thus a special case of validity as introduced in Chapter 1. Validity is a relation on the intentions held by a party and the instantiations of those intentions. Instantiations might take any of a number of forms, for example, prose description, formal specification, or running system; the validity of each is a measure of how well it matches what was intended for that form. The fidelity of a communicative transaction is then a relation on the intention of the source, and the thought as instantiated in the target. In fact, since the foremost goal of requirements is to provide for system builders an adequate understanding of a system to be built, fidelity of the communicative transaction is actually the root case upon which all other levels of validity of a system are predicated. We are thus very interested to know what influences the fidelity of a communicative transaction, since supporting this fidelity provides the basis for successful requirements communication and thus for requirements, and ultimately system, validity.

The fidelity of a communicative transaction is a function of the interaction of two linguistic principles, pragmatic sufficiency and the linguistic sign. One concerns how much fidelity is necessary given the purpose of a transaction, and the other concerns where this fidelity comes from, that is, the constituents that must be addressed in order to achieve it.

We next discuss the role of pragmatic sufficiency, followed by consideration of the linguistic sign. Implications of these principles are then integrated into a statement of the approach to improvement of communicative transactions taken by this work.

2.2.2 Pragmatic Sufficiency

The principle of pragmatic sufficiency says that source parties in a communicative transaction strive to say (or sign, or write) *what* they believe necessary and sufficient, *in a manner* necessary and sufficient, to achieve their aims [7]. Pragmatic sufficiency is a direct outgrowth of the notion of language as purposive; it reflects choices on the part of the source party made according to the purpose of the transaction.

The principle of pragmatic sufficiency helps to scope the problem of improving requirements communication since it invalidates any goal of achieving perfect fidelity. Recall the essential indeterminacy of meaning: the fact that a given instance of meaning is a product of one individual's consciousness, and is shaped by that individual's experience, personality, skills and other environmental factors. The essential indeterminacy of meaning means that we can never know with absolute certainty, for example, what a requirements producer intends by a given statement, or what a requirements consumer interprets in response to it, and further, that there can never be an exact match between them, since the experience of each individual is idiosyncratic to him. However, what pragmatic sufficiency brings to this discussion is the insight that the fidelity of communication need not ever be perfect; it need only be sufficient to the situation. This insight permits the requirements community to focus its efforts, since it indicates that any goal of closing entirely the gap between concept as intended and as interpreted is actually far from necessary, independent of its impossibility.

Thus perfect communication is neither possible nor necessary, but sufficiency for purpose is the target quality of every communicative transaction. However, insufficiency happens, and in environments of any consequence, undesirable outcomes result from this insufficiency. The needs for fidelity in requirements environments exceed those of the common case, and these needs are clearly not being met. The goal then is to demonstrably achieve higher levels of fidelity, thereby better approaching sufficiency for purpose in these environments. To do this, communicating parties must be supported in making their choices about what they want to say and how they want to say it.

To strategize support for pragmatic sufficiency, we next consider the influences that, depending on their manifestations, either contribute to or threaten it. The degree of adherence of parties to two main principles determine whether a transaction will be sufficient. The first of these, the Cooperative Principle, is addressed in the general linguistics literature and we review it below. The second, the Capability Principle, though naturally implicated, is first set forth in this work. The possible manifestations of each in a communicative transaction and the resulting effects on fidelity and thus sufficiency are discussed.

2.2.2.1 Cooperation and Capability

In order to accomplish a pragmatically sufficient communicative transaction, the parties to the transaction must uphold two criteria in executing their respective activities. In the most basic sense, the parties must be both *willing* and *able* to carry out their responsibilities to the transaction. While the existence of these criteria makes clear common sense, achieving them is actually non-trivial, and knowledge of the ways parties can fail to achieve them can be used to direct compensatory action. We first examine the *willingness* component.

Grice's Cooperative Principle. The criterion of *willingness* indicates that in order to approach pragmatic sufficiency, parties must buy in to a cooperative philosophy that supports attainment of the pragmatic goal. In codified form, the Cooperative Principle as stated by Grice says that parties implicitly agree to cooperate under the following terms: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" [22]. In other words, parties buy in to a goal of "maximally effective exchange of information" [22], that is prerequisite to the greater pragmatic goal of the circumstance, and shape their actions to the best of their ability to cooperate toward this goal.[1] Thus the Cooperative Principle sets forth the rules of engagement for pragmatically sufficient communication.

The Cooperative Principle is detailed by a taxonomy of constituents called the Conversational Maxims. The Maxims are a set of norms and assumptions about the speech act held by natural language speakers. More specifically, the Maxims prescribe what a party is supposed to do in order to uphold his part of a communicative contract. The Maxims concern a source party's choice regarding the quantity of information presented, the quality of that information, the relevance of that information to the purpose of the transaction, and the manner in which that information is presented. In particular, a source party is prescribed to target effective and appropriate manifestations of each in his delivery [22].

One threat to pragmatic sufficiency derives from failure to uphold one or more of the Maxims. The Maxim of Quantity, for example, can be violated through the provision by the source party of either too much or too little information for the purpose at hand. The target party then either lacks the necessary information to influence his behavior appropriately, or is placed in the position of needing to filter information, possibly without guidance, before determining, correctly of incorrectly, which information might be relevant to subsequent behavior. This raises the question of how to determine what the appropriate quantity for a purpose is; we will revisit this question in Chapter 4.

Grice provides a taxonomy of failures to fulfill the Maxims, but focuses on those that are used as devices to realize other communicative aims. For example, a party might violate a given maxim, but in so doing offer information that he believes will allow the other party to derive an implication. In answer to the question *Where does C live*, for instance, he might say *Somewhere in the South of France*, which violates the Maxim of Quantity, since presumably this response is less informative than the requesting party desired. This allows him to intuit, however, that the questioned party is not in possession of the requested information [22].

Note that the violation just described is intentional on the part of the violator, as are the majority of violations upon which Grice focuses. However, unintentional failures to uphold the Maxims are virtually ignored, in particular, those that derive from *inability* as opposed to *impossibility*. That is, while Grice recognizes, for example, the case of a clash,

1. Note that the goal of "maximally effective exchange of information" is not universal. Other goals of communication exist, for example social bonding through conversation, in which the information content is secondary to the act in accomplishing the goal. However, information exchange is the most common goal and the one relevant to the greater problem of improving requirements communication.

in which a party has no choice but violate one Maxim in order to uphold another, thus rendering complete adherence impossible, he does not address violations caused by the inability to execute cognitive functions at a level of fidelity sufficient to the purpose. However, this is precisely the case that concerns us. Clearly, in addition to willingness to cooperate, parties must also possess sufficient facility with necessary language functions to achieve a given communicative goal. In the common case, limitations are well-tolerated, but the level of facility required in the environments with which we are concerned is uncommonly high.

Further note that in the violation most recently described, both parties are *aware* of the violation, and this fact is a prerequisite of the intended implication. However, parties in both roles are not always aware of violations, and in the cases with which we are concerned, the state of awareness of each party regarding possible violations contributes additional complexity to a characterization of deficient transactions.

This places us in a position of being able to extend Grice's taxonomy to describe scenarios previously unaccounted for in the literature. Specifically, we are concerned with cases in which parties are attempting to cooperate in good faith, but in which they are unable to uphold a Maxim due to natural cognitive limitations to achievable fidelity, with the result that fidelity sufficient to the purpose is not reached. Further, there are four awareness scenarios that additionally qualify such a violation: either both parties are aware of a violation, only the source is aware, only the target is aware, or neither is aware.

These four cases stage different responses by cooperating parties with respect to transactional repair. In the first case, the parties recognize that further information is required from an outside source and can initiate activities to acquire it. In the second, the source recognizes that the intended concept has not been provoked in the target and source-initiated repair ensues, during which the source tries alternative presentations in order to achieve the desired result. In the third case, the target recognizes that the concept provoked in his mind is something other than what was likely intended (or he is unable to conceive of something coherent at all), and target-initiated repair ensues, during which the target requests alternative presentations.[1]

All of the previous cases have in common the fact that at least one party recognizes the violation and actions take place to accommodate it in order to achieve the transactional purpose. These activities can drain resources, especially given the overhead associated with communication and repair through asynchronous document exchange, as is the case in requirements environments, but they do usually result in achievement of transactional purpose.

In the fourth case, however, the violation is recognized by neither party. That is, a concept other than what was intended is provoked in the mind of the target and each party continues on to his subsequent activities unaware that the goal of the transaction has not been achieved. This is by far the most critical case, since first, it is a source of greater resource drain than the previous cases, because if the miscommunication is eventually recognized later in the development process, the cost of repair is multiplied in accordance

1. Note that upon reparative action, any case might become instance of another case. Reparative action continues until one of two states is reached: pragmatic sufficiency, or the fourth and most critical case, to be described below.

with the delay. In addition, however, if the miscommunication is never recognized, it poses risk to the validity of the deployed system.

Thus the cooperation of communicating parties does not in itself provide for pragmatically sufficient communication. The issue appears to be rather with limitations on the *ability* of those parties to fulfill cognitive operations at a level of fidelity sufficient to the communicative purpose. That is, the problem lies with *unintentional* failure to uphold Grice's Maxims, and is further elaborated by the differences that come with varying awareness of this failure on the parts of the communicating parties. We therefore turn next to a discussion of the mechanics of ability, in order that we might be able to use this information to scaffold transactions that more readily avoid both the Case 4 situation of communicative failure that can threaten the validity of a system, as well as the less severe, but still costly, use of resources for reparative activity characteristic of the other cases.

The Capability Principle. In order to approach pragmatic sufficiency, parties must not only be willing to cooperate, but must also possess communication skills commensurate with the level of fidelity required in a given environment. In analogy to the Cooperative Principle, then, we can state the Capability Principle: communicating parties must have facility with the cognitive and linguistic mechanics required to satisfactorily complete the communicative transaction. Thus whereas the Cooperative Principle is a set of norms and assumptions that set forth the rules of engagement for pragmatically sufficient communication, the Capability Principle is rather a set of functions that must be successfully executed, describing the cognitive generation and interpretation of the form and content of the engagement and the properties this form and content must have in order to achieve the communicative goal.

The Capability Principle is detailed by a complementary set of operations taking part in the minds of producers and consumers of communication, respectively. A producer is any source party, such as a speaker, signer, or writer, while a consumer is any receiving party, such as a listener, observer, or reader. The operations are (usually unconscious) mental manipulations of semantic abstractions executed in the minds of natural language speakers. More specifically, the operations describe what we know about what the respective parties do cognitively in order to accomplish communicative transactions. In order to elaborate these operations further, however, we must first dispel a common misconception regarding communication.

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. It maps 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 to another person, who, upon hearing or reading them, unwraps the meaning contained therein [33]. 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 of and possessed by an individual, and not itself something that can literally be given, accepted or recorded.

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 communicative transaction. The implication is that we must continually keep in mind that the processes to be scaffolded (procedural accommodations to operations) will *not* include anything having to do with packaging and delivery. Rather, they have to do with stimulus and provocation.

By presenting a chosen aural or visual stimulus, a producer hopes to receive a particular response, the provocation of a particular thought in the mind of a consumer. In other words, he is relying on the existence of a codified *association* between a particular stimulus and a particular response. This codified association maps a thought to a representation and vice versa; each party executes the mapping from his own point of view in order to complete the transaction. This execution is characterized at a minimum by the following processes: one that turns something a producer thinks into something he says, writes, or otherwise represents, and another that turns something a consumer hears, sees, or otherwise senses into something he thinks.

An individual builds up his own idiosyncratic set of form-meaning associations over his lifetime of experience. Many become shared in a community, and it is on this basis that a producer may begin to rely on such codifications. Given a set of sufficiently codified associations, the complementary operations that define the communicative transaction are the following. Articulation is the process of representing thoughts possessed by a producer in forms that can be transacted. Interpretation is the process within the consumer of activating possessed meanings in response to the presentation of forms with which he has associated them.

Further, parties can cooperatively build new associations from a common basis. In the general case, a producer can provoke the activation of a meaning already (assumed to be) possessed by an individual via presentation of a form that the consumer (is assumed to have) previously associated with that meaning. However, it is sometimes the case that a producer wants to communicate a meaning for which he believes the consumer might not already possess an association to a form. That is, he believes that in order to provoke the thought in a consumer most efficiently, he would otherwise use a particular form, but in the case that he believes the consumer to not have a preexisting association between the form and thought, he must find an alternative means. In this case, it is possible to build new associations on the basis of existing ones. Specifically, a producer can provoke the activation of a set of meanings using preexisting associations, and then direct the consumer to integrate that set of activated meanings into a more complex one, for which the producer introduces a new form in order that the consumer can create a new association between it and the complex meaning. This is one way that people accumulate knowledge.

The point to remember is that while each individual does possess a set of form-meaning associations, in communication, it is only the forms 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 form-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.

Deficiencies in the execution of the complementary producer and consumer operations threaten the achievement of pragmatically sufficient communication. If any link in the chain of processing comprising articulation and interpretation is lacking in integrity or broken, the result will be something other than what the producer intended. Since in any case of attempted communication, the producer must accomplish his own operations with sufficient integrity to cause the operations of the consumer to result in the intended thought, then assuming cooperating parties (as we do), the problem of improving integrity of the transaction lies largely with the producer. It is the producer's job to provide the stimulus that has the highest chance of provoking the correct cognitive response on the part of the consumer. If the stimulus presented by the producer is deficient, it may provoke, instead, no response at all (insufficient to complete consumer operations), multiple competing responses (ambiguity), or an erroneous response (creation of an unintended thought).

Consider a communication event in which the producer is an expert in a given field and the consumer is not. The producer intends to help the consumer to create a new association via the provocation and integration of a set of shared meanings, for which he will offer the consumer a new form to associate to the integrated meaning. A break that might occur in the chain of articulation and interpretation in this scenario is the failure of the producer to provoke one of the several meanings to be integrated, for example, because it represents tacit knowledge he possesses by virtue of being an expert. The consumer then interprets the set of forms as provided and creates a new meaning from their integration, but it is underspecified relative to the producer's intention. The result is an ambiguity, since the underspecification allows more than one interpretation of the remaining set of forms; thus the meaning constructed by the consumer might be what was intended or it might not. Further, if there are multiple consumers, their interpretations might be inconsistent. If neither party recognizes the ambiguity, no attempt at resolution can be made and behavior of the consumer ensues on the basis of the faulty understanding the parties have established. In the case of requirements engineering of a high-consequence system, this behavior might include, for example, the design of a safety-critical system component.

2.2.2.2 Pragmatic Sufficiency and Fidelity of the Communicative Transaction

The Cooperative and Capability Principles indicate that in order to achieve pragmatically sufficient fidelity in a communicative transaction, cooperating parties must be able to execute the required cognitive operations with skill commensurate to the task. Since the skill level required in the environments with which we are concerned is so high, we want to accommodate natural habits and limits in order to reach this level of fidelity. To do this, we want to target elements of the transaction that can be systematically improved.

The natural choice is to more closely examine the two entities under manipulation by the operations, that is, the complementary elements of the association. The association and its elements is codified by a linguistic abstraction called the Sign. Definition of this abstraction will allow the definition of desirable properties that would allow the further definition of procedural accommodations to the operations in order to achieve those properties and thus an improved transaction. We next address this abstraction, and define those properties.

2.2.3 The Linguistic Sign

To achieve pragmatically sufficient communication, parties must achieve pragmatically sufficient manipulation of a linguistic system. Linguistic systems are representation systems, sets of associations between meanings and forms that are codified within a speech community. The thought-representation dichotomy is so foundational to language that it is defined by its own abstraction in the linguistics literature; this association is called the Linguistic Sign. Better understanding of the structure of its components will allow us to develop methods to more consciously and directly manipulate it, in order to elevate the level of fidelity in a transaction.

The Linguistic Sign is composed of two parts, the signified and the signifier. The signified is the meaning part of the association; it is the notion as conceived by a party, to either be communicated, or resulting from communication. The signifier, on the other hand, is the physical pattern of sound or light that has been codified in a speech community to be associated with that meaning, for example, a word, executed in speech, writing or another format. The sign is the abstraction that establishes a bidirectional pointer between these two, from form to meaning and vice versa. The signifier and signified are mutually defining for a native speaker in a speech community in which the sign exists; each component is provoked by the other's presentation [55].

The association that forms the sign is intrinsically arbitrary, in other words, the sound pattern associated with a given meaning has no actual relationship to that meaning other than the one imposed by a speech community [55]. This arbitrariness is what allows members of different speech communities to communicate among themselves with regard to the same concepts communicated in other communities, with different forms; in other words, this is the basis for the existence of different languages. Because of this arbitrariness, however, meaning can only be interpreted from a given form based on a system of contrasts between it and other forms. As a simple example, consider the difference between the forms *book* and *books*; the additional sound in the latter is what indicates the difference in meaning.

Thus the ability to distinguish between two or more meanings is a function of representational context, in particular, the evidence of contrast it provides. A producer must provide sufficient context for a consumer to identify the factors that contrast the meaning he eventually interprets with all other possible meanings. If sufficient contrast was in fact provided, the consumer will arrive at an interpretation matching that intended by the producer.

It can be the case (and is the case for our problem) that the producer and consumer belong to speech communities that only partly overlap, for example, they speak the same native language, but differ in their level of expertise in a given domain. To achieve their engineering goals, new associations, that is, new signs, must be scaffolded from the common bases of these parties. In order to scaffold new signs, it is required that the producer in any transaction provide the consumer with the right amount of context to establish all relevant contrasts, that is, to distinguish the intended new meaning from any other. Further, this must be done with the necessary accuracy and precision of requirements environments.

The problem then is how to establish the level of context necessary to distinguish meanings with the necessary accuracy and precision, and how to guide the producer in determining and providing it. This is the goal of the Method presented in this work. We next define the desirable properties of communicative transactions that the Method will bring into existence.

2.2.4 Implications for a Candidate Solution

The ability of parties to successfully execute a pragmatically sufficient communicative transaction is a function of the skill with which they manipulate the entities of the Linguistic Sign. In particular, the producer must identify with sufficient accuracy and precision the signified that he wishes to communicate. Then, in the absence of a pre-existing association shared with the consumer, he must construct a new signifier that will provoke a set of meanings in the mind of the consumer such that these meanings can be integrated into a new signified. This signified may then be associated with a new simple signifier if appropriate in order to complete the generation of a new sign for the consumer.[1]

The foundational steps of identifying the signified to be communicated and constructing an alternative signifier to present to the consumer motivate two desirable properties that we would like to establish and safeguard via the procedural support of an engineering method. These properties are the integrity of the signified and the integrity of the signifier. The integrity of the signified is the degree to which the necessary contrasts that distinguish this signified from any other have been determined. The integrity of the signifier is the degree to which these contrasts are represented in an alternative signifier such that they are interpretable as intended by the consumer.

Functionally, then, the problem is to design procedural support to accommodate human cognitive habits and limits such that these properties can be maximized in any communicative transaction. In the next section, we look at additional issues that shape our solution: those that place practical constraints on the shape of the method.

2.3 Practical Constraints on a Candidate Solution

A candidate solution to the problem under consideration will in this work take the form of a methodology encompassing processes and artifacts that support the target aspects of requirements engineering. A methodology with potential to be used effectively will have both content components that address deficiencies in communicative transactions, as well as structural components that aid the embodiment of the content components in a method that has practical utility.

Since requirements engineering fundamentally involves the collection and transformation of cognitive products of humans, direct interaction among humans and between humans and tools is necessary, and no requirements process can ever be completely automated. Thus any candidate solution must accommodate practical constraints implied by the need to interact with humans and do so in ways that give confidence in desirable outcomes. The marshalling of human resources is a challenging problem in its own right, and we must therefore attend to more diverse issues of practicability than if we were produc-

1. Note that this process as presented canonically describes single-concept definition tasks, but in fact applies to any communicative transaction, for example a sentence in which the parties share each component concept, but their particular combination asserts new knowledge. That is, every communicative transaction is characterized by an element of novelty as the point of its existence. In these cases, of course, the entire combination of meanings is not associated with a new signifier; these are reserved for signifieds that encapsulate individual concepts and recur as such.

ing, for example, an automated tool that needed at most a human operator. This imposes methodological constraints on the shape of any candidate solution.

To constitute an attractive method, a candidate solution should have a number of properties in addition to those implicated in the previous sections. In particular, any candidate method should necessarily be capable of application in the intended environment, that is, it should be feasible. A method as defined cannot be inexecutable given environmental parameters including humans as resources. In addition, the method must provide some assurance of predictability of its outcome despite variation injected by human agents. A method that does not reliably produce a desirable outcome is of questionable utility. Together, these properties indicate that a candidate method is capable of being productively practiced. We thus refer to them together as the practical utility of a method.[1]

Several characteristics contribute to the feasibility and predictability, and thus the practical utility of a method. Feasibility is supported by the provision of realistic direction in how tasks may be accomplished, guided by realistic expectations about what is possible given environmental parameters. Evidence should be provided in support of the executability of all directed actions. Predictability of outcomes is supported by repeatability of the process producing them, despite variation across executing parties and execution environments. This repeatability comes from consistency in application of the method, which necessitates sufficient guidance to parties completing tasks that are otherwise completed implicitly and/or in an ad hoc manner. That is, predictability of the outcome is supported by constraints on and direction of the behavior of humans in executing the method. Further, the particular constraints and direction applied should not only restrict variation, but should also ensure that the outcome can be demonstrated to be desirable. That is, the process should enforce that the product has identifiable properties associated with the improvements we seek, in other words, that the product is evaluable.

Thus, for a method to have practical utility, its users must both be able to execute it given environmental constraints, as well as have confidence that executing it will reliably produce desirable results. We combine these needs with those implicated in earlier sections to formulate a refined, actionable statement of the problem in the following section.

2.4 Refined Statement of the Problem

The goal of this work is to improve the fidelity of communication in requirements environments such that requirements validity can be more readily achieved. To make progress with regard to this problem, we more fully specified its elements in the preceding sections in order to more directly address these elements in the construction of a candidate solution.

The problem of improving communication in requirements environments is in this work delimited by the following three parameters:

---

1. The cost to an organization of applying the method is obviously a practical issue as well, but one that is beyond the scope of this work. The prospect of comprehensive cost measures is addressed in Chapter 7 under Future Work. While cost measures are not directly addressed in this work, Chapter 6 does offer suggestions for efficient use of the method, based on empirical observations collected during its evaluation.

1. The needs particular to requirements environments for trans-boundary communication as well as for accuracy and precision exceed those of the common case of everyday communication. A solution must therefore address these increased needs.

2. Communication is equated for our purposes to fidelity of the communicative transaction, which is a product of the interaction of pragmatic sufficiency and the linguistic sign, which are each products of further phenomena. A solution must therefore address this fidelity and its constituent influences.

3. Any proposed mechanism of improvement to communication is subject to factors that qualify the practical utility of that mechanism. A solution must therefore accommodate such factors to preserve its practical utility.

These parameters integrate into the following refined problem formulation: To improve communication in requirements environments, the fidelity of communicative transactions must be improved such that environmental needs for trans-boundary communication, accuracy and precision are demonstrably supported, and environmental constraints affecting practical utility are accommodated.

The following logic thus defines the greater argument of the work. The validity of requirements reflects the fidelity of communication undertaken in the presentation and reception of those requirements. The fidelity of communication is a function of the interaction between the pragmatic sufficiency of the communication for the circumstance and the facility of communicators in manipulating the linguistic sign. To reduce threats to pragmatic sufficiency, deficiencies in manipulation of the sign must be accommodated. To accommodate these deficiencies, the construction and function of each component of the sign must be explicitly addressed. Two desirable properties are thus indicated: integrity of the signified and integrity of the signifier, of any given target of a communicative transaction. A method must support the chaperoning of these properties into existence. Further, it must do so at a level appropriate to environmental needs, which exceed those of the common case, and in a manner of demonstrably practical utility, or its potential value will not be realized.

2.5 Chapter Summary

In this chapter we presented three main sets of issues that further shape the problem under consideration and thus impose constraints on candidate solutions. Specific challenges to communication imposed by the environment, the basic structure of communication generally, and practical constraints imposed by the need to apply a candidate method productively, all contribute to the formulation of an actionable goal. Thus, the problem revisited is to support communicative fidelity, to do so at a level appropriate to environmental needs, and to do so in a manner amenable to practical use. This work provides a mechanism to accomplish these goals and a demonstration of its efficacy.

The remainder of the work assesses the state of the art with regard to this problem, reviews and develops as necessary the underlying theory, presents a theoretically-grounded methodology conforming to the indicated needs and constraints, and reports on evaluation activities designed to show that the presented methodology does in fact achieve the goals elaborated in this chapter.

In the next chapter, we argue that the problem as posed is open in the community and thus that the contributions to be made herein constitute an advancement to the state of the art.

If an elderly but distinguished scientist says that something is possible, he is almost certainly right; but if he says that it is impossible, he is very probably wrong.
--Arthur C. Clarke

3 State of the Art and Practice

This chapter reviews the literature addressing improvements to communication in requirements environments. In particular, it concerns what approaches exist for detecting, correcting, and preventing errors in requirements that compromise the communicative fidelity of concepts critical to applications in development. It assesses their contributions and limitations, in order to determine the boundaries of current knowledge in this area. As a result, it is argued that the problem of achieving pragmatically sufficient communication between producers and consumers of requirements remains open in the community, and that the contributions to be made in this work thus constitute an advancement to the state of the art.

3.1 Work Treating Requirements Communication

Because they have such great influence on software validity, a number of other researchers have investigated improvements to the language and communication of requirements. This universe of work can be divided along multiple dimensions. In this discussion we first address those approaches that deal primarily with the *form* of requirements representations, including formal methods and natural language processing approaches. These are followed by an overview of approaches that are more focused on the *meaning* intended by requirements representations, including mechanisms for ascribing meaning to identifiers, psychological and social factors affecting the understanding of requirements, and various forms of guidance in ambiguity detection. From these, a description of the state of the art is then synthesized.

3.1.1 Form-oriented Approaches

This section addresses approaches to improving requirements communication that focus primarily on the form of representations. Formal languages and the formal methods built around them offer the opportunity to construct mathematically verifiable models by virtue of logic rules that representations must abide by in order to be correct. Natural language processing approaches perform automated mathematical and other types of analyses of natural language texts in order to uncover interesting patterns.

3.1.1.1 Formal Methods

Many approaches to reducing common language and communication challenges in software requirements advocate the practice of translating requirements representations into formal languages, for example, Z or PVS, or those that mimic natural language in order to be more accessible but are still definitionally formal, such as Attempto Controlled English (ACE) [15].

The motivation for the use of formal languages is clear: they are precise, unambiguous, and mechanically analyzable for various sorts of consistency and completeness. However, despite these advantages, formal languages have a number of limitations that preclude their address of exactly the issues with which we are concerned in this work.

First, errors that can be detected through analysis of a formal representation are only those that are possible to detect within the closed world of the model being analyzed. Analysis of formal representations can verify adherence to a set of logical rules operating over a set of formally-defined entities, but it can not, for example, provide evidence for completeness of the model or validity of its elements. Arguments of validity, in particular, are fundamentally outside of the scope of formal methods precisely because of their formality. The integrity of a statement in a formal language is entirely a function of its syntax, while validity explicitly concerns that which a representation is intended to mean.

Second, for formalization to even be possible, the communication issues addressed by this work must be successfully managed in order to know with confidence what entities and rules to propose in a formalization. "...[A]ny formalism is parasitic upon the ordinary everyday use of language, in that it must be understood intuitively on the basis of ordinary language" [39], that is, formalism is bootstrapped by informalism. Every software artifact to be built begins, by its very nature, as an informal cognitive concept in the mind of a client. While any useful solution must eventually be formalized in code in order to be executable, formalization does not answer the question of determining and communicating what is to be formalized.

Further, there is support for the notion that representations that address meaning are desirable even once a model *has* been successfully formalized. In a case study of 12 different formal specifications for the same problem, Wing stated "One lesson I learned...is that injecting domain knowledge adds reality and complexity to a specificand. If such knowledge exists and if it can be added systematically, then incorporating a knowledge-based specification technique...in the overall software-development process would be beneficial" [69].

We thus find that in order to address the meanings intended to be communicated in the development and refinement of software requirements, the use of informal languages, in particular the (ideally native) natural language of the communicating parties, is necessary. First, it provides by definition the starting currency, since it is all that is available to any parties beginning a communicative relationship, regardless of what later might be built upon it. Second, because the use of natural language is the primary mechanism for execution of a communicative transaction and its attendant operations, it is provides the the most accessible indicators to meaning as intended and as interpreted, whether on its own or in support of a formalization.

3.1.1.2 Natural Language Processing

Other methods recognize the presence of natural language as a necessary fixture in the development process and seek to address some of its challenges through primarily syntactic means, such as mechanical analysis of text. For example, the Automated Requirements Measurement (ARM) tool aids the detection of certain communicative deficiencies in requirements expressed in natural language. Ambiguities and incompletenesses, for example, are operationalized via indicators called *weak phrases*, which are located in the text and counted as a measure of overall requirements quality. *Weak phrases* are defined as phrases argued to "...cause uncertainty and leave room for interpretation" [68].

However, the method of generation of the set of weak phrases is not provided; we do not know exactly why particular phrases were chosen and others not. In fact, by the definition provided, any phrase at all is weak, since every phrase leaves room for interpretation according to the linguistic principles governing communication. Further, we do not know to what degree the use of these phrases as indicators is valid, that is, to what degree the presence of such phrases positively correlates with the measured presence of the named deficiencies. Finally, the ARM tool aids detection of deficiencies, but offers no guidance for their effective correction. Ultimately, such deficiencies must be successfully corrected in order that they not compromise requirements validity.

Another approach combines both NLP elements and formal representations to address requirements communication issues. Burg's COLOR-X provides for the coding of natural language requirements in a linguistics-inspired formal language, such that automated processing technology can extract models from these requirements. Entities in the models are then connected to definitions in the WORDNET lexicon, a lexical database that relates words to each other through characteristics such as synonymy.

Burg's approach does address some instances of ambiguity and inconsistency, in particular, those that are removed by conversion to a formal representation. Further, despite being primarily a syntactic processing tool, it addresses the need to reference meanings, and contracts out this job to an existing general-purpose lexical database. However, in both cases, goals in which we are interested are not met. Specifically, since the requirements representation is first converted to a formal language, the language deficiencies detected and corrected are the same set addressed by formal methods generally, while the question of the validity of the representation before conversion is still in question. Further, while entities in the extracted models are connected to descriptions of their meanings via WORDNET when available, the organization of WORDNET and its entries not only does not meet our particular purposes, but it in addition does not include the vast majority of the domain-specific notions particular to the applications we are interested in supporting.

3.1.2 Meaning-oriented Approaches

A number of additional approaches recognize the inability of primarily-syntactic methods to address significant communication and validity issues in requirements. Some contributions that more focally address semantics and provide context for this work concern mechanisms for connecting meaning to representations, psychological, and social influences on requirements communication, and common sense heuristics that reduce the incidence of communication deficiencies such as ambiguity.

3.1.2.1 Connecting Meanings to Representations

Careful understanding of the connection between a system and its application domain has been an object of attention for Zave and Jackson, Jackson, Heninger, and Leite and Franco. The work of each addresses 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 and proposes the notion of *text macros* to meet this need, but offers little advice on their character or construction [28].

Zave and Jackson similarly propose the notion of *designations*, with the intention of linking a form of definition to terms that might otherwise be opaque. They base this proposal in fact on the insight that formal representations are semantically void, insofar as the correctness of a formal representation is a property only of its syntax, and therefore, that designations must be provided for the formal symbols in a specification [71].

Jackson further elaborates an entire class of description mechanisms, separating them into *designations*, *definitions*, *refutable descriptions* and *rough sketches*, each serving a different purpose for semantic description in requirements and specification generally [31]. Of these mechanisms, the one that comes closest to serving our definitional purposes is the *designation*, here defined as "...a recognition rule for recognizing some class of phenomenon that you could observe in a domain" [31]. While a recognition rule would serve our purposes well, no explicit guidance is provided for constructing one. Further, there is no explicit mechanism to prevent designations from suffering common weaknesses of definitions described later in 4.2.1.2. In addition, none of these approaches addresses the semantic dependencies possible among entities, for example, that text macros and designations could be self-referential (or indirectly so) and that this fact complicates the integrity of the communication.

Leite and Franco, in contrast, recognize such dependencies, and in their Language Extended Lexicon (LEL) approach, advocate exploiting this fact by providing domain term definitions that maximally make use of other domain terms [35]. The approach is based on the strategy of exposing all possible dependencies, but it has a limitation. By encouraging maximal use of other domain terms, they encourage the construction of potentially circular chains of reference that are never guaranteed to be grounded in accessible meaning. If the grounding problem is not addressed, then consumers have no means of access to the semantics of terms for which they also do not know the semantics of their components.[1] Further, as with the other proposed mechanisms, there is little guidance (other than maximal use of other domain terms) in the structure and construction of these definitions.

Finally, data dictionaries also 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 [56]. 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 [56, 49]. They

---

1. Consider using a dictionary printed in a foreign language you do not speak. It may be possible to determine the indexing scheme and to find a definition for a term, but if the definition is not itself in your native language, then you are not any better off. You could then look up each of the terms in the definition, but with the same result: another innaccessible definition.

do not serve, other than incidentally, to provide domain or other meaning, as the glossaries of requirements documents are intended to.

3.1.2.2 Psychological and Social Factors

Other work that has addressed the communicative aspects of requirements takes human psychological and social behavior into account and concerns the influence they have on the way we understand information. Leveson, for example, proposes *intent specifications*, designed to enhance the way humans perform tasks and solve problems, based on what is known about these activities from work in cognitive psychology and human-computer interaction [36]. Leveson recognizes the need to capture expert intent and proposes structures in which to record it, but applies psychology only at the level of compound, higher-order activities such as an expert's reason for preferring an option, and does not address the more foundational issues of communicating meaning with fidelity. In this sense, the purpose of intent specifications is more to record "Why?" than "What?", with regard to the content of requirements.

Potts and Newstetter, and Goguen both advocate forms of naturalistic inquiry as it is practiced in social and anthropological endeavors. Naturalistic inquiry involves a data collection process whereby an analyst observes participants in a community and pieces together a picture of their interaction through various techniques including interviewing and artifact collection [47]. The goal is generally to expose "...the meanings and tacit understandings that participants in social contexts negotiate and derive from interactions with the artifacts and other participants there" [47]. Goguen goes so far as to say that all information is relative, shaped by the context in which it is used, and by the filters through which for example, users, developers, and others understand it, and concerns himself with the theory of this contextual aspect and the organizational dynamics associated with it [21]. Potts, together with Takahashi and Anton, further uses observations of human interaction to motivate strategies to enhance communication. For example, they propose a structured conversation model for communicating requirements based on what they term the *inquiry cycle*, the patterns according to which humans iteratively request and process information in constructing new understanding [48].

All of these approaches contribute in various dimensions and at various levels of abstraction to the goal of improving requirements communication. While they exploit psychological and sociological knowledge in this endeavor, the low level cognitive linguistic mechanics of communicative transactions remain uninvestigated.

3.1.2.3 Heuristic Ambiguity-Detection Methods

Several additional approaches offer heuristic guidance built from codified common sense and other knowledge to support detection of ambiguities in natural language requirements.

Gause and Weinberg, for example, demonstrate with a well-known nursery rhyme the number of ambiguities that can exist in even supposedly innocuous places, and how their cumulative effect can greatly distort the eventual interpretation they provoke. To combat this, they suggest the Mary Had a Little Lamb heuristic, which directs us to take each element in turn and explicitly consider the alternatives to the first assumed meaning [16].

Similarly, Berry advocates the inclusion in any requirements team of a person to fulfill the role of a *smart ignoramus* [3, 2]. This smart ignoramus is defined as someone who is: (1) generally intelligent and, in particular, (2) has a demonstrated facility with the use of language, but (3) who does not have any knowledge of the domain of the system for which a piece of software is being developed, and (4) is not afraid to ask questions. The point of including such a person derives from Berry's personal experience. More valid system models resulted when he happened to serve in roles where the above properties held. The intuitive explanation for this effect is that these properties drove him to be generally more inquisitive and detail oriented, and to ask all the questions he felt necessary until he believed he constructed the intended understanding.

Kamsties, Berry and Paech provide further codification to intuition, combined with some philosophical and grammatical analysis of ambiguity. Having created a detailed taxonomy differentiating types of ambiguity relevant to requirements engineering, they propose an inspection process based on checklists and scenario-based reading, that guides the posing of ambiguity-exposing questions of particular areas of and statements in requirements [32].

Gause and Weinberg, Berry, and Kamsties, Berry and Paech all offer heuristic guidance in reducing ambiguity in requirements with the added bonus of a low investment bar for getting started. However, ambiguity is only one of several manifestations of representation insufficiency. Further, while each supports ambiguity detection, no guidance is offered for effective correction or prevention.

3.2 Synthesis and Implications

To review, then, a number of approaches address the challenges of improving communication for requirements. Form-oriented approaches include formal methods and natural language processing. Formal methods are free of deficiencies like ambiguity and are mechanically analyzable for other types of deficiency, but are unable to detect a different set of deficiencies and fundamentally can not address validity. Natural language processing and other automated text processing methods accept natural language as a fixture in the development process, but the patterns they expose in representations are only as interesting as the indicators used are valid. Further, neither formal methods nor natural language processing offer rigorous guidance in correction or prevention of the kinds of deficiencies in which we are interested.

Of the approaches that focus on the meaning component of representations, some work takes into account the need to provide definitions or other meaningful descriptions for identifiers, but does not recognize the implications of dependencies among these descriptions, or the need for accessibility of these descriptions to consumers. Other work focuses on psychological and social factors in communication, but generally treats the issues at a higher level of abstraction than that required to systematically improve the fidelity of communicative transactions. Some heuristic approaches provide easy-to-use guidance in detecting ambiguities, but do not address several other types of deficiency, and like the form-oriented approaches, offer little guidance for correction and prevention.

We thus assert that the problem of achieving pragmatically sufficient communication between producers and consumers of requirements is open in the community.

3.3 Chapter Summary

This chapter examined engineering approaches that have addressed the goal of improving communication for requirements. It reviewed their contributions as well as assessed their limitations with respect to this goal. Assessment of the boundaries of knowledge in this area allowed the argument that the problem of achieving pragmatically sufficient communication between producers and consumers of requirements remains open in the community. Thus the contributions made in the remainder of this work, insofar as they extend the boundaries of knowledge in this area, constitute an advancement to the state of the art. We next outline the cognitive and linguistic foundations of the methodology we propose to further advance this agenda.

If you want to make an apple pie, you must
first create the universe.
--Carl Sagan

4 Theoretical Foundations of CLEAR

This chapter provides theoretical foundation for the development of procedural support for improving the fidelity of communicative transactions in requirements environments.

The fidelity of communicative transactions is a product of the integrity of the components being transacted. That is, to improve transactional fidelity, we seek to improve the skill with which we manipulate the elements of the linguistic sign. Specifically, we are interested in exactly how signifieds are cognitively distinguished from one another, and further, in the qualities of signifiers that encourage or hinder correct interpretation on the part of a consumer. This chapter investigates the linguistic bases for both what we need to represent, and how we need to represent it, in order to improve transactional fidelity.

In the remainder of this chapter, we introduce cognitive categories, the constructs by which humans store and organize semantics, and the ways that factors in requirements environments interact with categories to limit the efficiency of innate cognitive heuristics. This leads to a foundation for the systematic determination of the boundaries of a signified in a communicative transaction. We further introduce reductive paraphrase, an approach from the linguistics literature that can provide the basis for the construction of high-quality signifiers by which to scaffold new form-meaning associations for consumers. Together, they compose the basis for the core of the procedural support encapsulated by CLEAR.

4.1 Accounting for the Signified

In order to maximize the fidelity of the communicative transaction, it is necessary to determine and then effectively represent the set of attributes that allow pragmatically sufficient distinction of the concept intended by the producer from all other concepts. This section sets forth the linguistic foundations for that part of CLEAR that will support such determinations.

4.1.1 Primer on Human Semantic Organization

The root objective of any attempt at communication is to recreate the particular semantics conceived of by one person in the mind of another. In order to understand how the semantics conceived of by one person may be recreated in the mind of another, it is first useful to have a notion of how these semantics are stored in a mind in the first place. Research results from linguistics and cognitive psychology demonstrate that the universe

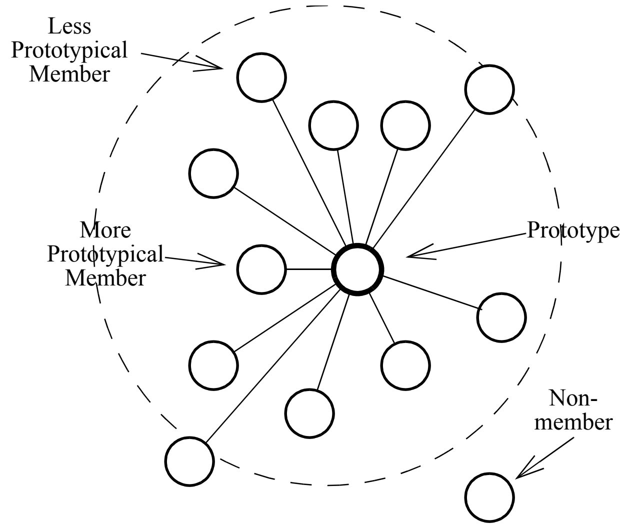


Figure 1. Cognitive Categories

of semantics stored by any person is organized into collections called *categories*, which individually have internal structure and are related to each other structurally and behaviorally in a number of ways [52, 41, 60, 34, 58]. The remainder of this section defines these entities and their role in communicative transactions that both succeed and fail.

4.1.1.1 Cognitive Categories

Cognitive categories can be defined for our purposes as follows:

> *Cognitive categories are collections of mental representations of entities encountered or imagined by an individual that are judged by that individual to be sufficiently similar to each other to count in some partitioning of reality as being the same.*

Since there are many possible partitionings that are useful to us in our interaction with the world, an entity may be a member of many categories depending on the factors considered to be salient for the task or experience at hand. For example, a tree stump might be identified by an observer as what is left of the trunk of a tree that has been cut down, but it may simultaneously be identified as a welcome seat by a tired hiker.

Categories, in addition to being collections, have internal structure to organize the collection of which each one is comprised (Figure 1). They can be visualized as radial arrangements of instances that bear degrees of resemblance to central prototypes. Instances closely clustered around the prototype bear stronger resemblances to it and instances further away are less prototypical. The tree stump above may indeed be a member of the *seat* category, but for most individuals it would be a non-prototypical member, whereas a chair such as is found in an office or living room would be more representative. Further, a given subject's prototype might manifest itself for the subject as a gestalt with a number of representative attributes, but which is not equivalent to any particular instance of the category. Consider the category *airplane*; one subject might consider a Boeing 777 to be prototypical, but another might posses a prototype that is an abstract amalgam of many airplanes, characterized by a set of attributes that are representative for the informant.

Determination of which instances are prototypical or peripheral for a given category is an individual judgment call, based on accumulated past experience of the category possessor. In fact, membership itself is an individual determination: peripheral members of a category might be so peripheral that they might be judged as in or out of the category inconsistently by different individuals or even by the same individual as his experience changes. He may even be unable to decide at a given instant, allowing categories to overlap. Membership in cognitive categories, unlike membership in formal sets, is not definitionally binary. Specifically, two main principles govern category membership, one based on the presence of Necessary and Sufficient Conditions, and the other based on the presence of Family Resemblances.

Categories that obey the Necessary and Sufficient Conditions model have a common set of attributes that are universally associated with all members of the category by most to all informants in a speech community. For example, for a year to be categorized as a leap year, it is necessary to be both divisible by 4 and not divisible by 100, unless also divisible by 400; evaluation of these conditions together is sufficient to perform the categorization, and all members share both attributes. Judgments of membership in categories governed by the Necessary and Sufficient Conditions model are straightforward given the ability to determine the presence of critical attributes and have little variation across informants within a speech community in which the category is codified. Further, such notions are easier to formalize since there is little room for differing experience that can lead to ambiguity and assumption in communication that uses these categories. However, categories that obey this model are relatively rare in even specialized domain-specific communication.

In contrast, categories that obey the Family Resemblances model have a set of attributes that are not universally shared by all members of the category, but rather serve as more and less frequent markers, perhaps clustering together in various combinations. For example, *investigation*, a category prominent in the safety-critical literature, is characterized by a large number of attributes, many of which are shared by many to most investigations, but few of which are universal among them[1]. In the case of Family Resemblances, membership judgments tend to feel more intuitive to the maker, even though in theory they are made by a usually subconscious probabilistic process involving the calculated similarity of the entity at hand to previously experienced instances, and the frequency of prior encounters with the instance at hand or others like it [50]. The set of attributes relevant in making these membership distinctions, and especially the set as conceived by the

1. A problem sharing commonalities with the communication of requirements for software systems is the problem of communicating to accident investigators the process of conducting an investigation within a given environment. Such activities often suffer from the lack of explicit definition of terms like *investigation* in guidelines that prescribe such processes. Hanks, Knight and Holloway argued that the lack of explicit restriction on the interpretation of this and other terms limited the predictability and repeatability of investigation processes and thus the value of the data generated by them; if two investigation teams interpret the details of their tasks differently, then their data are not comparable and are therefore incapable of exposing patterns and trends that could motivate appropriate corrective action [24].

producer in a transaction, is what we are interested to more fully determine and represent than what is otherwise managed naturally.

Some categories additionally have a further organizing dimension representing semantic **extensions** to the category. Extension refers to the selective appropriation of a subset of category semantics for use in classifying another group of entities. For example, before computers existed, the category *memory* was arguably restricted to representing a faculty of sentient life forms; its use in computer terminology abstracts certain, but not all, of its prototypical properties and exploits their commonality with aspects of human memory to allow an intuitive understanding of what computer memory does. Computer memory is not a peripheral member of the base category *memory*, rather, it is more like a category in its own right with a prototype and peripheral members (consider how *core* has gone from one to the other over time), but with a semantic origin and continued mapping to a preexisting category. This is essentially the cognitive view of metaphor.

In addition to internal organizing principles characterizing cognitive categories, the collective set of categories possessed by an individual has been shown empirically to obey certain principles. Among these is *cognitive economy.*

Cognitive Economy. *Cognitive economy* is the ability of humans to quickly associate with an entity a large number of attributes that might not be readily observable [52, 60]. This association is made through the identification of an entity with a category as described above, which leads to the heuristic ascription to the entity of additional attributes highly correlated with other known instances of that category. For example, a consumer, upon receiving a message with the signifier *nest* in it, would link that signifier to the category represented by it in his inventory. Heuristically he would then ascribe to the instance attributes that might not be apparent but are highly correlated with other nests he has experienced, for instance that it is made of twigs and dirt, and that it might contain eggs or birds. If the message was, in fact, "Look at that nest!", the consumer would likely look up, possibly into a tree, because those locations highly correlate with the locations of nests he has previously experienced. The consumer would already know what to look for and where; the message producer need not have indicated any description or direction. In this way, the consumer in a communicative transaction is able to conceive of rich and usually very accurate semantics while the message size is kept much smaller than it might otherwise need to be.

Hierarchical Organization. Another principle collectively obeyed by the set of categories in an individual's inventory is *hierarchical organization*. Empirical evidence suggests that the categories we possess are organized into a hierarchy of specificity, with more general categories at higher levels and more constrained categories at lower levels. This hierarchy is one of inclusion, meaning that many low level, highly constrained categories are collected under the umbrellas of more general categories, several of these more general categories are collected into still more general categories, and so on (Figure 2). Such hierarchies offer a range of options for conceiving of or referring to an entity, according to the level of specificity that matters for the communication. For example, a particular object in the world might be referred to as a piece of furniture, a chair, or a recliner according to the role of the object in the context of the transaction.

Evidence also suggests that a particular level of this hierarchy has a special salience in our perception of the world [53, 52, 60]. This level is intermediate; it is neither a very

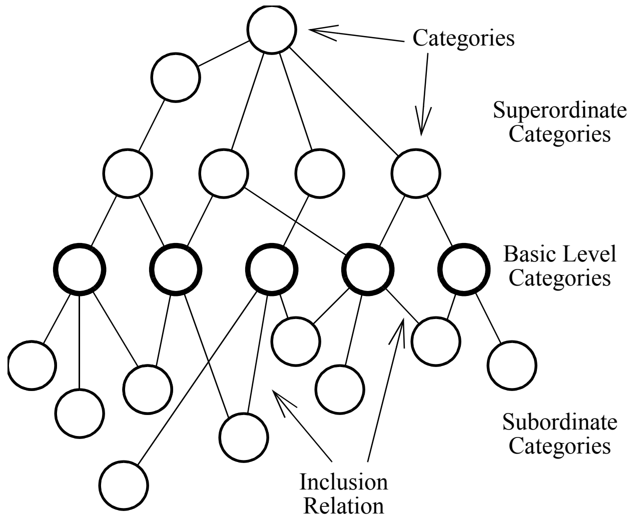


Figure 2. Hierarchical Structure

general way to describe a thing nor very specific, for example, *dog* rather than *mammal* or *retriever*, and its special role is evidenced by features particular to it in our acquisition and use of categories. It is termed the *basic level*[1], and it represents the categories first acquired by children, the categories we call up first when classifying a newly encountered entity, and the categories we use when introducing a new category into conversation.

Importantly, analysis of empirical data has demonstrated a further property of the basic level that supports its apparent salience: *the basic level is that level of the hierarchy at which elements of any given category share the most attributes with each other and the fewest with members of other categories* [53, 52, 60]. For example, *chair* is a basic level category; at a higher level, e.g. *furniture*, entities have fewer attributes in common, while at a lower level, e.g., *rocking chair* or *office chair*, fewer attributes serve to distinguish. Basic level categories simultaneously maximize within-category sameness and extra-category differentness.

In the common case, cognitive categories and their organizational mechanisms and the principles they collectively obey allow the conceptualization and communication of semantics among humans. However, there are a number of cases for which these organic mechanisms are not evolutionarily tuned, and the fidelity of the communicative transaction is at risk. We next examine several of these cases.

4.1.1.2 Communicative Breakdown

Heuristics such as cognitive economy that allow humans to communicate large volumes of semantics using little bandwidth are tuned by evolution to be most effective in the common case of everyday communication. However, since they are heuristics, there of course exist cases in which they are less effective than desired. Linguistic analysis suggests, for example, that cognitive economy actually backfires under certain circumstances. Unfortunately, they are circumstances that characterize requirements environments: first,

1. Note that the basic level does not refer to the base of the hierarchy.

higher than common needs for accuracy and precision, and second, the high incidence of trans-boundary communication.

These facts of requirements environments also provide further challenges for our natural cognitive endowment. In particular, the hierarchical organization of categories interacts with the high incidence of trans-boundary communication to create relatively more opportunities for misinterpretation. This section elaborates the challenges originating from the interaction of environmental properties with our innate language processing mechanisms. It accounts for many of the communication difficulties seen in these environments and provides a basis for the development of procedural accommodation.

Challenges to Cognitive Economy. Cognitive economy works by exploiting the fact that in the common case, parties with shared experience will make similar assumptions regarding the attributes of members of a category. One form of pathological case that defeats the purpose of cognitive economy is the event in which a consumer identifies the wrong category given its signifier. This occurs when a signifier signifies more than one category or extensions thereof, i.e., there are name collisions. For example, a *monitor* can refer to the display on one's desk, a synchronization primitive, a minimal operating system, or even the child elected to manage traffic in the school hallway. While these are all extensions of a general *monitor* base category, each independently possesses its own internal categorical structure. If there is insufficient context to aid the consumer in discriminating the intended extension, or if he simply does not possess certain extensions, it is possible or even likely that he will ascribe to the entity in question a whole host of attributes that do not match those that were intended. His ensuing behavior will then be conducted on the basis of this incorrect understanding. Such misidentification can happen any time the communicating parties have sufficiently different experience contextualizing the transaction, but transactions in which one party is an expert in the area and the other is not provide a particularly clear illustration of this differential.

Another form of pathological case that defeats the purpose of cognitive economy is only subtly different than the above, but far more insidious. In this case, the consumer identifies the correct category or extension, that is, the signifier has the same basic referent for both parties. However, either the producer or consumer has accumulated past experience that renders his version of the category more dense with attributes, and their relative correlations with the prototype more tempered by experience and less open to adaptive reinterpretation. This difference in quality and density of experience causes the corresponding versions of the category possessed by two different people to have different topologies, i.e., different loci of prototypicality and peripherality, different relations and distances between members, and different levels of precision and focus on categorical elements. Any time two such people use this category, each assumes the topology and constellation of attributes consistent with his own experience. Depending on the relative expertise of the consumer, the result is that he constructs an understanding that is both qualitatively misaligned as well as either over- or under-specified, relative to that which was intended. Again, the consumer's ensuing behavior will be conducted on the basis of an incorrect understanding, and decisions and actions dependent on it will be invalid.

Many such categories exist where the denser, more fixed, and more precise category is possessed by an expert in a domain and a less constrained, more dynamic version is held by non-experts. Consider, for example, *energy*: in physics, this category is constrained and well defined, with a precise set of attributes. It is *the capacity to do work*, where *work* has itself a specific meaning, and *capacity* is specified in measurable units. In lay usage, *energy* is still a source of power, but the work it allows is less specified than its physics counterpart and its measurement is also variable. This situation poses a communicative risk any time one party in a transaction is significantly more expert in the area in question than is the other party, an event that is common in requirements environments.

A further challenge arises from the interaction of trans-boundary communication and the hierarchical organization of categories. Domain experts have accumulated experience that results in their association of more attributes with certain categories (specifically those in their domain) than a non-expert would associate with the same categories. These attributes provide more dimensions along which to potentially collect and differentiate entities. This results in certain of these categories bearing the above mentioned mark of the basic level: that level at which members of any given category share the most attributes with each other and the fewest with members of other categories.

The implication is that experts often tend to see, within their domain, lower-level, more constrained categories as basic, and therefore use them in ways that basic-level categories are used. On being presented with a new entity to be classified, provided the entity has some role within his domain, an expert is likely to associate it with a more constrained category than would a non-expert. Similarly, on introducing an entity into conversation, he is likely also to invoke a more constrained category.

This means that in addition to experts and non-experts possessing more and less constrained versions, respectively, of certain categories (as in the *energy* example above), the denser expert versions are more likely to come up in discussions relative to more general categories because, to the expert, they are at the basic level. This translates to proportionally more opportunities for misalignment between the categories transacted between producers and consumers than would occur because of the backfiring of cognitive economy alone.

Thus some of the distinguishing characteristics of requirements environments are exactly those that cause our innate cognitive machinery to fail; it is fundamentally not in our nature to communicate with fidelity with very high accuracy and precision and across domain boundaries. This accounts for many of the communication problems we see in software development. In addition, it provides direction for accommodation and ultimately improvement. We next describe the methodological implications deriving from this analysis.

4.1.2 Implications for a Method

The circumstances of requirements environments that provide challenges to our innate cognitive machinery have in common the fact that they make use of assumptions on the part of communicating parties. When communicating parties share significant relevant experience, these assumptions are more likely to be correct, but in requirements environments, this prerequisite is often missing. In the absence of sufficient common experience, accommodation must be made such that the opportunity for faulty assumption is reduced.

The main trigger for assumption at the level of the communicative transaction is the presentation of a signifier without elaboration, which provokes for the consumer a category or subcategorical elements that might or might not be as intended. Therefore one possible counter to the making of these assumptions would be to provide explicit elaboration with presentation of the signifier of the details that might otherwise be assumed with questionable validity, such that many of the assumptions never have the chance to arise. In other words, a method that directed the determination and presentation of the set of categorical attributes that chaperoned a correct interpretation without resort to the risks of cognitive economy would contribute to the ability to communicate concepts with pragmatic sufficiency in requirements environments.

The *determination* of this set of attributes is one main goal of CLEAR; it picks out with guided care and direction the thing in world intended by the producer, in so doing safeguarding the integrity of the signified in a transaction. The *representation* of these attributes is the other main goal of CLEAR, and the challenges of the construction of good representations are addressed next.

4.2 Accounting for the Signifier

In order to maximize the fidelity of the communicative transaction, in addition to determining the set of attributes that influence judgments regarding category membership, it is further necessary to effectively represent this information, such that the representation has the best chance of provoking the intended interpretation in the mind of the consumer. This section sets forth the linguistic foundations for that part of CLEAR that will support the construction of such representations.

4.2.1 The Purpose and Challenges of Definitions

In the case of a shared sign, each party in a communicative transaction associates the same meaning with a given form. In the case of a non-shared sign, for which a consumer needs to construct new association, an alternative to the original signifier must be provided. This alternative is a representation consisting of a collection of signifiers that ideally *are* correctly interpretable by the consumer, such that the set of meanings provoked characterizes the signified for the new association to be constructed. The common form for such representations is a definition.

In order to construct representations, or definitions, that have the best chance of provoking intended interpretations in the minds of consumers, we next discuss the properties such constructions are desired to have, as well as weaknesses that are commonly identifiable in these constructions and therefore in need of address if improvement to communicative fidelity is to be realized.

4.2.1.1 Definitions are Descriptions

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

Entering a word in a dictionary or glossary implies that the builders of the dictionary or glossary believe there exists a coherent speech community in possession of a common meaning for a word of interest to another speech community. For example, a bilingual English-French dictionary makes assertions about the form-meaning associations shared by the population of native English speakers when it offers definitions of English words for French users. The motivating populations for our purposes are the producers and consumers in communicative transactions executed in the development of requirements.

A good definition can be characterized in terms of how well it represents the codified meaning associated with a form in question. "[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*" [19].

Achieving descriptive adequacy is a two part problem that must take into account both the integrity of the description under construction and the integrity of that being described. The described is the concept for which a lexicographer wishes to produce a definition, that is, the signified concept. 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 conception, then even the best description will describe the wrong thing. This is the part of the problem addressed by the determination of categorical attributes discussed earlier.

The description, on the other hand, is the structural representation used to provoke understanding of the concept in the mind of the consumer, that is, it is the collection of signifiers presented as an alternative to the original. 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 consumer, if the description is simply intractable, or worse, provocation of incorrect meaning, which the consumer is unlikely to question, since we are conditioned to believe that dictionaries and glossaries are authorities on the meanings associated with given forms

Writing good definitions, however, is a non-trivial task. We next discuss some common weaknesses that limit their descriptive adequacy.

4.2.1.2 Common Weaknesses in Definitions

The integrity of a definitional description can be violated in a number of ways. We here present a typology of common weaknesses that threaten the descriptive adequacy of definitions.

Obscurity. An obscure definition presents a description that uses words no more significant to the consumer than the word being defined [19]. Since the consumer is lacking a form-meaning association allowing him to understand the target word to begin with, a definition that relies on words requiring additional associations not either already possessed by the consumer or themselves otherwise appropriately defined is ill-equipped to support descriptive adequacy. In other words, an obscure definition might in fact 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 consumer, correct or otherwise.

Scientific definitions are particularly vulnerable to obscurity. '[A]ir' has been defined as "the mixture of invisible odorless tasteless gases (as nitrogen and oxygen) that surrounds the earth" [40]. A consumer 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.

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

Circularity. A circular definition presents a description that uses one or more words that themselves are defined in terms of the target word [19]. A circular definition is non-interpretable because like an obscure definition, it relies on the availability of a further effective form-meaning association, but in the circular case, the association required is not forthcoming because it is for the target word 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 word 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 consumer in search of a definition for *red*. Obvious circularities 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, a published definition for 'question' relies on 'request', which relies on 'ask', which relies on 'answer', which relies back again on 'question' [43].

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 4.2.2.

Otherwise non-predictiveness. Otherwise non-predictive definitions present descriptions that rely on devices which, while convenient to the definition writer, explicitly hinder the ability for a consumer to predict appropriate range of use of a word. Such devices include the disjunction 'or', subjunctives 'may', 'might' and 'can', umbrellas like 'etc.' and 'esp.' and hedges like 'usually' and 'generally'. By resorting to devices like 'etc.', "...the lexicographer makes the definition untestable" [19]. The issue with such devices is that without criteria for the scope they are intended to cover, it is impossible to know what is included and what is not. Thus a consumer is unable to predict whether the word is appropriate in a given situation.

The next section reviews an approach from the linguistics literature that addresses the construction of definitions that avoid these common weaknesses.

4.2.2 Reductive Paraphrase

The construction of good definitions is important to a number of other areas including our own. In addition to their use in law, standards, and obviously dictionary making, the production of high-quality descriptions of the meanings of words is also a sub-goal of a linguistic initiative that has produced an approach to definition from which we can benefit. This initiative is called the Natural Semantic Metalanguage (NSM) project.

NSM is the product of a research agenda 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 [20].

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 descriptions of 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" [18].

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.
- "[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.
- "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" [18].

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 only terms remain 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" [18]. The 60 that are currently hypothesized have proven recalcitrant to disconfirmation throughout the work.

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 [18]
```

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
```

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 [18]. 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 signifier 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 and must, for example, enumerate instead the cases covered.

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. NSM shows that the issues of definition discussed in 4.2.1.2 can be systematically avoided in the construction of definitional descriptions. The next section addresses the possibility of appropriating elements of this initiative to improve the fidelity of communicative transactions in the development of requirements.

4.2.3 Implications for a Method

The goal of NSM is determination of the metalanguage, and reductive paraphrase is the process by which it is determined. For our purposes, however, the metalanguage is of little interest while the value of reductive paraphrase for rendering definitions more capable of provoking their intended meanings is attractive. 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 development to fully explicate. A representation in proper semantic primitives is more representation than is needed, since the speech communities in question for requirements communication problem do have *some* common experience that can be leveraged. In other words, 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 4.2.1.2 while limiting the extent of the reduction commensurate with available resources and descriptive goals.

A method that directed the qualified reductive paraphrase of signifiers, such that the sets of attributes previously determined to be relevant were represented in an accessible way, would contribute to the ability to communicate concepts with pragmatic sufficiency in requirements environments.

4.3 Summary

This chapter introduced cognitive categories and described their role in communicative transactions that both succeed and fail. Examination of breakdown in transactions allowed the identification of properties a method should have in order to support the ability of a consumer to distinguish among concepts. This chapter further introduced reductive paraphrase, an approach from the linguistics literature for constructing high-quality representations of word meanings. While reductive paraphrase proper is not suitable for our needs, modifications were suggested that would allow the approach to benefit our purposes. Together, the structure and organization of cognitive categories and the process of reductive paraphrase provide a basis from which to develop a method to determine both what is to be represented, and how to represent it, in service of improving the fidelity of communicative transactions in requirements environments.

In the next chapter, we present the definition of the CLEAR Method as derived from these foundations.

The priest persuades a humble people to endure their hard lot, a politician urges them to rebel against it, and a scientist thinks of a method that does away with the hard lot altogether.
--Max Percy

5 The CLEAR Method

This chapter presents a solution to the problem of improving the fidelity of communication in requirements engineering of high-consequence software systems. Cognitive Linguistic Elicitation and Representation (CLEAR) was developed to address the incidence of miscommunication between producers and consumers of requirements in a software development project. We here define the constructs and activities that constitute the method to solve the problem posed by this work.

The problem as set forth earlier is to improve communication among parties in a requirements process such that more valid requirements can be achieved. In particular, we seek to provide procedural support for communicative fidelity at a level appropriate to environmental needs and in a manner amenable to practical use. To do this, CLEAR will answer the questions, “How can it be decided which concepts are in need of special attention?” and “What kinds of special attention should they be given?”

To solve this problem, we rely on linguistics as a source of insight regarding how communication works at a low level. Specifically, we exploit results from linguistics and cognitive psychology in the provision of well-defined procedural support for capturing, storing and representing the intent of requirements producers such that it is accessible to requirements consumers. CLEAR enables the construction of an organized repository of domain semantics critical to the production of valid software systems.

The solution takes the shape of a serial manipulation of linguistic objects such that desirable properties are enforced. The solution methodology consists of four phases, the end result of which is a body of entries characterized by these desirable properties. We call this product a CLEAR Knowledge Base. The value of the knowledge base is twofold: the nature of its construction provides assurance that particular communicative hazards with regard to the information contained therein have been avoided, and analysis of its content in concert with its organization offers insight that was not previously available regarding the application domains at hand.

The chapter is organized as follows. Section 5.1 defines the Method. In particular, section 5.1.1 presents an introduction to the methodology as a whole, including the main elements and their relationships to one another, and section 5.1.2 focuses on low-level details of the constructs that facilitate the phases and the processes governing their generation. Section 5.2 provides a brief chapter summary.

5.1 Definition of the CLEAR Method

CLEAR is composed of a number of phases. Some definitions, goals, and strategies hold for the entire method, while others are applicable only to particular phases. We first introduce CLEAR in the large, and present bases that apply to the method globally. We then treat the individual phases in detail.

5.1.1 CLEAR in the Large

This section describes the CLEAR Method at the level of the whole. It presents global definitions applicable throughout as well as the set of specific deficiencies the method was built to address. It further overviews how the parts of CLEAR combine to meet its overall goals.

5.1.1.1 Entities of Interest

This section defines the entities over which the CLEAR Method operates.

The **communicative transaction** was previously defined as an event between two (or more) parties in which the first produces a **signifier** intended to provoke a particular **signified** in the mind of the second, who interprets it. The first party we call the **producer**, and the second, the **consumer** in the transaction. The **signifier** and **signified**, recall, are complementary parts of the **linguistic sign**, an association between a form and a meaning. To meet the goal of improving the fidelity of the **communicative transaction**, we need to increase the likelihood that the same **signified** possessed by the **producer** is provoked in the **consumer** when presented with a **signifier**. We thus want to manipulate elements of the **sign** such that this is more likely to occur.

In order to create a **signifier** that is more provocative of its **signified**, we first determine as closely as possible what the **signified** in question is. This is the **concept** as intended by the producer. We then want to create an alternate **signifier** that is ideally more provocative of the intended **concept** than the original. The new **signifier** we call a **representation**, and the original, a **phrase**. A **phrase** is simply the word or string of words by which the **concept** is conventionally named, while a **representation** is a compound of other **phrases**, that provoke a compound of other **concepts**, in order to provoke the intended **concept** through construction when the conventional **phrase** is insufficient. Thus definitions are a kind of **representation**, as are **explications**, the products of Reductive Paraphrase as reviewed in the previous chapter.

Given unlimited resources, we would address every **sign** used in the requirements. In practice, however, we partition this set into **targets** and non-targets. A **target** is a **sign** with which we are concerned. A **target concept** is then the **signified** of this **sign**, a **target phrase** is the original **signifier** as used by the producer, and a **target representation** is a definition, paraphrasing, or other description composed of **phrases**, either originally available or purpose-built to scaffold the construction of this **sign** for consumers.

5.1.1.2 Properties vs. Quantities

This section explains how improvement to the fidelity of the communicative transaction is qualified in this work.

The goal of improving fidelity of the communicative transaction leads to the question of how to measure how well a signifier provokes an intended signified. Ideally, we want to know what percentage of the intended meaning is provoked, that is, we want to know the goodness-of-fit between the signifier and the signified. However, goodness-of-fit can not be quantified; one can not measure how removed a signifier is from its signified because signifieds are physically inaccessible for observation. It is not possible to look at, for example, the exact pattern of neural firings that instantiate one.

This means that goodness-of-fit can not be used as a comparator for alternate signifiers. Instead, we look to properties of signifiers that are known to advance or compromise communicative fidelity; we can assess signifiers for the presence or absence of such properties. For example, we can compare two representations with regard to the presence of well-defined deficiencies and say that we have improved the state of affairs if we create representations with fewer such deficiencies. Further, note that while we can not directly quantify goodness-of-fit, we *can* quantify deficiency counts.

Thus to improve communicative fidelity, CLEAR locates and removes pre-existing instances of specific deficiencies, as well as guides the creation of new representations free of them. We next present the typology of deficiencies with which CLEAR is concerned.

5.1.1.3 Typology of Deficiencies Addressed

Recall from the previous chapter that the process of Reductive Paraphrase was designed to construct explications, special kinds of definitions free of three particular types of common definitional deficiency [19]. The deficiencies addressed by Reductive Paraphrase are obscurity, circularity, and otherwise non-predictiveness, structural characteristics that limit the ability of a consumer to predict the accurate range of use of a term or phrase. These three make up one subset of the deficiencies addressed by CLEAR.

A complementary subset that focuses on the amount and quality of the *content* of a definition can also be postulated based on linguistic category theory. Recall from the previous chapter that linguistic category theory provides for a logic of category membership, albeit graded; thus in theory there exists a set of semantic constraints governing inclusion in and exclusion from any category. An ideal definition would therefore provide valid information in neither excessive nor insufficient quantity to communicate those constraints. Too much information risks overconstraint; too little, underconstraint, that is, these affect precision in the identification of category boundaries. Erroneous information compromises accuracy in identifying these boundaries. We therefore add these three semantic deficiencies to the three provided above.

The full set of definitional deficiencies that CLEAR addresses consists then of obscurity, circularity, otherwise non-predictiveness, the provision of too much information, the provision of too little information, and the provision of erroneous information. Any representation constructed according to CLEAR guidance will as far as possible be free of these deficiencies.

5.1.1.4 Mechanism of Address

The specific activities of CLEAR are built around the goal of safeguarding the integrity of the linguistic sign. In particular, CLEAR provides a set of steps by which to determine with high accuracy and precision the category topology of the signified. Further, it provides another set of steps by which to create an alternate signifier for this signified such that no deficiencies of the types elaborated above are present. The first set constitutes a phase called Elicitation, the second a phase called Representation. When Elicitation and Representation are executed on a target, the result is a representation that can be demonstrated to compare favorably to definitions produced by other means, with regard to both the content represented, and the presence of common definitional deficiencies that inhibit fidelity of the communicative transaction. That is, the activities involved scaffold a path to the achievement of desirable properties that otherwise do not spontaneously occur.

5.1.1.5 Encapsulation into the Method

While Elicitation and Representation provide the foundational phases of the CLEAR Method, two other phases serve to provide both reasoned support for these as well as to exploit their results in order to derive knowledge not previously available.

Selection precedes the other phases and is comprised of the set of steps by which targets are separated from non-targets for further processing. Selection combines the strengths of both manual inspection and statistical document analysis to generate a set of targets that can be argued to be in greater need of address than non-targets, thereby providing guidance in resource allocation.

Integration follows the other phases and is comprised of the set of steps by which the representations produced are collected together and organized into a greater structure that records their relationships to each other. The relationships thus instantiated then provide a substrate for analysis that leads to the visualization of interesting and useful properties of application semantics that were not otherwise available.

Thus four phases in total combine to define the canonical CLEAR Method, Selection, Elicitation, Representation, and Integration. The entire method can be visualized as in Figure 1, and the individual parts are elaborated in the following sections.

5.1.2 CLEAR in the Small

This section describes the CLEAR Method at the level of the parts. It elaborates in detail each of the four component phases and how their serial manipulations of entities result in the product of CLEAR, a CLEAR Knowledge Base (CKB). It further describes how the CKB may be analyzed to generate further contributions to a given application.

5.1.2.1 CLEAR Phase 1: Selection

The first phase, Selection, exists in practical support of the remainder of CLEAR. While the other phases perform manipulations of representations of target concepts, the Selection phase defines what those targets will be. The Selection phase of CLEAR partially automates, with foundations drawing from the Information Retrieval literature, the task of selecting out from a set of materials what is of interest for further processing. This section addresses the goals, construction, and use of the products associated with Phase 1, Selection.

Selection Goals. This section states the goals of the phase, describes their contribution to solving the greater research problem, and introduces the strategy to be used in realizing the goals.

Statement. The goal of Selection is to provide reasoned guidance in the separation of targets from non-targets for further processing.

Contribution. This goal contributes to the ultimate improvement in fidelity of the communicative transaction in two ways. First, it provides a means by which some phrases may be prioritized over others, and thereby guide resource allocation. Executing CLEAR on every possible phrase is prohibitive; Selection allows focus on higher-priority items.

Second, the guidance that allows attainment of this goal enforces a degree of consistency across multiple executions of CLEAR. Since targets are separated from non-targets according to a set of criteria, targets selected for any application have a particular set of properties in common with targets selected from another application. This provides a basis for possible comparison across target sets from multiple projects.

Strategy. The strategy for realizing this goal is to combine the strengths of manual and automated approaches for identifying phrases of interest. Manual approaches offer the ability to exploit reader intuition about what content is most important in a text, as well as subjective judgment regarding the accessibility of given phrases. A phrase poorly understood by a reader is a good candidate for further processing, as is one that intuitively seems to be central to description of safety-critical parts of the system. Manual approaches further offer the ability to inspect the output of automated approaches and adjust it as appropriate for concerns that are not manageable by automated means.

Automated approaches, on the other hand, offer the ability to quickly and accurately collect statistics on measurable properties of the text that may be of interest. These objective measures complement the subjectivity of manual approaches and provide a different view of the text with which to compare.

In particular, CLEAR Selection makes use of the notion of *indicators* from the Information Retrieval literature. An indicator is a pattern in text that correlates with the presence of a property of interest. For example, if two phrases frequently occur within some limited distance of each other, it is likely their meanings have some relation in the content of the text. The indicators used in the automated part of CLEAR Selection are the frequency with which a phrase occurs in the set of materials under consideration, and its distribution across documents or other partitions within that set. These two measures comprise the occurrence properties of a phrase, and their implications and use are elaborated in the Selection Definition below.

By executing versions of both manual and automated approaches, and comparing and selectively combining the results, we have a good idea of which phrases offer the most to gain from further processing.

Selection Definition. This section describes the sources, products, constructs, and activities defining the Selection phase of CLEAR.

Sources. CLEAR is designed to allow starting points from various forms of requirements documentation, purpose-built or embedded in other materials. The only constraint is that the materials under consideration exist in a form amenable to automated text processing if that option is to be made use of.

Products. The product of the Selection phase is a set of target phrases to be further processed during later phases of CLEAR.

Constructs. Execution of the Selection phase is facilitated by a table of the occurrence properties of all phrases in the source materials. The Selection Table is sortable according to the following parameters:

- *Frequency* with which the phrase occurs in the source materials. This is the total count of individual occurrences of the phrase or its grammatical variations. Grammatical variations include different tenses, numbers or word orders (where possible) of the phrase. High relative frequency indicates that the concept marked by the phrase participates in a relatively large percentage of the interactions set forth by the requirements, and thus the potential for misunderstandings to result in faults in core or foundational dimensions of the system.
- *Distribution* of the phrase across multiple documents or alternate functional divisions of the source materials, such as system components. This is the total count of documents or alternative functional divisions in which a phrase or its grammatical variations occurs. High relative distribution indicates wide influence on multiple aspects of a system and thus the potential for misunderstandings to result in non-localized or uncontained faults that affect the system in pervasive ways.
- *Intersection* of frequency and distribution. This is the set of phrases that simultaneously meet chosen frequency and distribution thresholds. Membership in this set indicates that the influence of the concept marked by this phrase is both wide and deep relative to phrases not in the set, and thus that the potential exists for the misunderstanding of one of these phrases to result in relatively many and diverse system faults.

These orderings allow prioritization decisions based on the capacity for influence that the understanding of a given phrase has on the validity of a system.

Activities. In order to select targets for further processing from among all possible phrases in the requirements, the following four steps are executed.

1. Scoping

CLEAR is designed to allow starting points from various forms of requirements documentation, purpose-built or embedded in other materials. This being the case, it is first necessary to constrain the set of materials to that to which manual and automated processing activities will be applied. Several concerns can influence the choice of scope for a given application of CLEAR, for example, whether the material in question actually represents requirements, whether it represents safety-critical requirements, or whether it represents core or pervasive system functionality. The ultimate determination of scoping boundaries is made according to priorities of the development organization. In practice, it is recommended that the chosen scope is as large as resources allow[1], and focused on the collection of material for which the consequences are thought to be highest if communication errors are not minimized. For example, the Selection scope for the requirements for a commercial aircraft might include the flight control software for takeoff and landing (the statistically most dangerous portions of a flight), but not the software that manages the in-flight entertainment.

It should be noted that determination of the in-scope material does not remove the remainder of the material from consideration during later steps. The remainder serves as a reference and resource as necessary during later stages of execution of CLEAR. Determination of the scope is simply a prioritization mechanism for constraining the area that will be processed during Selection. In the Representation Definition below, we describe how the remainder of the material may be further used.

2. Manual Inspection

The second step in execution of the Selection phase is the manual inspection by one or more analysts of the in-scope material for phrases meeting either or both of two criteria only possible through a manual approach. The two criteria are:

- The phrase marks a concept that intuitively appears to be central to correct functioning of the system, and/or
- The phrase provides difficulty to the analyst in reaching a confident understanding of its meaning.

All phrases meeting either or both of these criteria are added to a set of candidate targets awaiting further evaluation. If more than one analyst is available, all unique phrases selected by any analyst are made candidates. Unique in this case indicates that any selected phrase is only added to the set once, and grammatical variations of a phrase count as instances of that phrase.

3. Automated Processing

The third step in execution of the Selection phase is automated text processing of the in-scope material in order to generate a prioritization of phrases, based on their occurrence properties. This step consists of the following substeps:

- Generate the list of phrases for which occurrence properties will be measured. For purposes of automated processing, a phrase is defined to be any string of words that does not cross a clause boundary and does not begin or end in a part of speech that precludes its reasonable interpretation as coherent concept. Using this definition, together with a stoplist of potential phrases to ignore, such as numerals, month and day names, etc., all phrases in the materials are collected into a working list. Adjustments are then applied to this list before occurrence properties are computed. First, grammatical conflation is executed such that all grammatical variations of a phrase constitute one unique entry. Second, the list is manually inspected for phrases to be discounted due to excessive generality or being nonsense collocations; these phrases are removed from the list.

---

1. Determination of what resources will allow is currently educated guesswork accompanied by the ability to adjust scope as appropriate during execution. It is expected that as experience in applying CLEAR accumulates, better heuristics will become available for this task.

- Compute occurrence properties for every phrase in the list. In this substep, the frequency and distribution of every phrase is computed and recorded in the Selection Table.
- Effect a Prioritization. In this substep, the phrases can be ordered according to their occurrence properties to enable visualization of their relative influence in the associated requirements. In addition, thresholds may be set such that those meeting a chosen frequency threshold and those meeting a chosen distribution threshold can be intersected to produce and equivalence class of relatively-more influential phrases. The absolute size of the set of phrases selected in this way is arbitrary and open to adjustment. In practice, the aim is to make the set large enough such that when it is coanalyzed with the results of manual selection, the final set of targets is of a size appropriate to available resources.

4. Coanalysis

The final step in execution of the Selection phase is coanalysis of the manually- and automatically-generated phrase lists in order to produce a single set of target phrases to forward on to Elicitation. The obvious first step is to intersect the lists; any phrase that occurs on both lists has been indicated by both subjective and objective means to be in likely need of further address. This might or might not generate a set of sufficient size; if not, further additions may take into account other subsets of either list deemed reasonable by the analyst, perhaps in consult with experts. For example, remaining phrases that were identified as having both high frequency and high distribution would be prioritized over those that had only one of these properties. Further, any phrases in the manually-generated list that were selected based on subjective difficulty of interpretation might be prioritized since there already exists evidence of their opacity. In this way, the final set of targets can be constructed to be of the desired size.

Use of Selection Products. The target list ultimately generated during Selection serves the remainder of the process by defining the canonical starting point for linguistic manipulation. In the course of later steps, targets will be both added and removed according to provisions of the method, such that the set of entries in the final CLEAR Knowledge Base is not entirely the same as that in the original Selection set, but the Selection set allows the core of CLEAR processing to begin.

5.1.2.2 CLEAR Phase 2: Elicitation

The second phase, Elicitation, begins the manipulative processing of the targets selected during Phase 1. The elicitation phase of CLEAR provides procedural support for improving the integrity of a target signified. Elicitation guides 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. This section addresses the goals, construction and use of the products associated with Phase 2, Elicitation.

Goals. This section states the goals of the phase, describes their contribution to solving the greater research problem, and introduces the strategy to be used in realizing the goals.

Statement. The goal of Elicitation is to determine with high accuracy and precision what the signified is for a given target.

Contribution. This goal contributes to the ultimate improvement in fidelity of the communicative transaction by providing a clear set of semantics to be accounted for in a representation. Elicitation addresses the integrity of the meaning component of the linguistic sign as one determiner of the pragmatic sufficiency of a communicative transaction; if a structurally perfect representation signifies an inaccurate or imprecise set of semantics for the target, then the consumer of this representation can not gain a pragmatically sufficient understanding of this target.

Strategy. The strategy for realizing this goal is to use linguistic category theory as presented in the previous chapter to guide the construction of a template for organizing category topology information for a target. This template can then be filled through consult with application domain informants. Since such informants are not generally familiar with category theory, an auxiliary construct is also required in order to collect this information on informants' own terms. The information can then be converted or transferred intact as appropriate into the respective fields of the template. A completed template becomes a *category record* for the target. The structure of both the template and its supporting collection form are elaborated in the Elicitation Definition below.

Elicitation Definition. This section describes the sources, products, constructs, and activities defining the Elicitation phase of CLEAR.

Sources. The Elicitation phase takes as input the set of target phrases generated during Selection.

Products. The product of the Elicitation phase is a set of completed category templates, or category records, for the targets, to be forwarded for further processing during the next phase of CLEAR.

Constructs. Execution of the Elicitation phase is facilitated by two constructs. The first is a category record, which organizes the category topology information for a target. The second is a collection form that supports the gathering of this information from informants.

The notion of the category record derives from the motivation to better determine the accurate range of use of a given phrase, such that this information may later be appropriately reflected in a representation. To better determine the accurate range of use, we use the direction provided by linguistic category theory in characterizing how people create and distinguish cognitive categories. Recall from the previous chapter the properties of cognitive categories; their structure and organization and the resulting implications for the fidelity of communication are used here as a basis for directing the information gathering process for CLEAR targets.

The category template is a structure designed to allow recording of the information used by an expert informant to determine category membership for a given target. It includes fields that when completed record elements of a hypothetical "categorization function" that contributes to the determination of the accurate range of use of any signifier associated or to be associated with the signified in question. In bringing into better focus what is a member and what is not, the "categorization function" embodied by the template provides raw material to be later encoded into a representation with more predictive power than that arising through other means.

The main fields of note in the template are:

- An expert definition. This provides the starting point for conception of the category, to be refined by information from additional fields.
- Most salient attributes. These are any attributes that immediately arise in the mind of an informant when asked what features allow him to determine whether two entities are instances of the same concept or not[1].
- Unexpectedly salient attributes. These are attributes that might not be salient in a lay or other relative of the category but are salient for the domain-specific use at hand. Attention to such attributes allows, for example, predictions that instances otherwise thought to be outliers might in fact be more central members of a category.
- Unexpectedly unsalient attributes. These are attributes that might be salient in a lay or other relative of the category but are not salient for the domain-specific use at hand. Attention to such attributes allows, for example, predictions that instances otherwise thought to be central members of a category might in fact instead be outliers.
- Supercategories, subcategories, and siblings. These relations locate the category within its category hierarchy and thereby allow implications to be derived based in this location. For example, since category hierarchies are hierarchies of inclusion, a given category shares all of the attributes of its supercategories in addition to the refinements that specialize it.

A copy of a full Category Template as used in a particular application of CLEAR is located in the Appendix.

By explicitly provoking this information from an informant, we can then more directly work with the constraints defining a concept. However, to get this information, a support construct is required since informants do not generally have sufficient familiarity with linguistic category theory to provide this information directly. We thus also define a collection form that encodes the questions implied by the category template into forms more accessible to informants. For example, to find out the most salient attributes of a concept, we present the following example scenario and question:

```
Among its many characteristic properties, certain
properties of wood are especially noteworthy with
regard to the domain of building and crafting. For
example, the hardness and durability of many woods
provide value to this domain, while the low cost
of others is attractive.
```

---

1. This does not imply adherence to the deficient Necessary and Sufficient Conditions model of category membership discussed in the previous chapter, since there is no requirement that all category members actually share a salient attribute. Rather, it is consistent with the Family Resemblances Model, in which certain features are shared by some or most members, and those that do the most work in allowing membership distinctions are the most salient.

```
With regard to the [current application domain],
what are the attributes of the concept represented
by [the phrase in question] that are particularly
important or noteworthy?
```

The answer to the above question can then be transferred with conversion if necessary by the analyst into the appropriate field in the category template.

A copy of a full collection form as used in a particular application of CLEAR is located in the Appendix. We next detail the activities facilitated by these constructs.

Activities. In order to determine with high accuracy and precision what the signified is for a given target, the following steps are executed. For each target selected in phase 1:

1. Collect information via completion of the collection form by expert consult. Minimally, one informant is required per phrase, but more than one allows their responses to be compared.

2. Convert responses to template content. Using the information provided in the collection form, in combination with context from source materials and a working understanding of linguistic category theory, fill in appropriate fields in the category template.

Note that it is not necessary, nor is it uniformly recommended, to fill in a category template exhaustively. The fields organize the available relevant information, but force-filling the structure would defeat its purpose. For example, a distinction in salience of attributes should not be forced if the informant has not provided evidence of one. Likewise, hierarchy information should not be filled in for the sake of filling it in, even if such relations exist, if these relations are not relevant to the application domain.

Use of Elicitation Products. The set of category records ultimately generated during Elicitation serves the remainder of the process by providing the source content for the representations to be constructed in the next phase. Further, organizational information that they record, such as hierarchies in which the targets participate, can be instantiated in the eventual CLEAR Knowledge Base, allowing visualization of these relationships in addition to others shared among entries. The records can additionally serve as an archival record for potential later use in future CLEAR and other agendas. In particular, it would be interesting to investigate determination of the location of the basic level in a recorded category hierarchy; if reliably possible, this could add another dimension to the value of CLEAR results.

5.1.2.3 CLEAR Phase 3: Representation

The third phase, Representation, continues the manipulative processing of the targets selected during Phase 1 and processed to an intermediate state in Phase 2. The representation phase of CLEAR provides procedural support for improving the integrity of a target signifier. Representation is executed via a structured definition process in which the recordings of expert conceptions generated during elicitation are iteratively converted into descriptions that abide by specific form and content rules, such that they are more likely to provoke the semantics intended. This section addresses the goals, construction and use of the products associated with Phase 3, Representation.

Goals. This section states the goals of the phase, describes their contribution to solving the greater research problem, and introduces the strategy to be used in realizing the goals.

Statement. The goal of Representation is to construct an alternate signifier for a given target, that meets particular form and content constraints.

Contribution. This goal contributes to the ultimate improvement in fidelity of the communicative transaction by providing guidance for the realization of desirable properties in a representation. The Representation phase addresses the integrity of the form component of the linguistic sign as the other determiner of the pragmatic sufficiency of a communicative transaction, along with the meaning component addressed by Elicitation. If a semantically accurate and precise conception of a target is signified by a structurally deficient representation, then a consumer of this representation can not gain a pragmatically sufficient understanding of the target.

Strategy. The strategy for realizing this goal is to adapt to our purposes Reductive Paraphrase (RP) as reviewed in the previous chapter, in creating a variant of explication that is free of an extended set of definitional deficiencies. Further, this adaptation loosens other constraints of RP proper that are less important to an engineering undertaking than to the linguistic goals RP was conceived to advance. The adaptation is called Partial Reductive Paraphrase (PRP), and when executed for a given target, it results in a PRP explication for that target. PRP and its resultant explications are further elaborated in the Representation Definition below.

Representation Definition. This section describes the sources, products, constructs, and activities defining the Representation phase of CLEAR.

Sources. The Representation phase takes as input the set of category records generated during Elicitation. In addition, Representation uses as needed and appropriate the set of original materials that provided the starting point, informant consult, as well as outside references such as Wikipedia [67].

Products. The product of the Representation phase is a set of Partial Reductive Paraphrase explications to be further processed during the final phase of CLEAR.

Constructs. Execution of the Representation phase is facilitated by the explication construct, a special kind of definition designed to avoid common definitional faults through the enforcement of particular properties. A CLEAR explication is based on the explication construct defined by the Natural Semantic Metalanguage (NSM) project as reviewed in the previous chapter. Recall that an NSM explication is a definition abiding by the following constraints:

- It is free from obscurity.
- It is free from circularity.
- It is free from otherwise non-predictiveness.
- It is written entirely in semantic primes.

CLEAR explications are adaptations of NSM explications such that the first three constraints remain, three new constraints are added, and the final one is relaxed. A CLEAR explication is a definition abiding by the following constraints:

- It is free from obscurity.
- It is free from circularity.
- It is free from otherwise non-predictiveness.
- It does not contain excessive information. This means that it contains no information that could overconstrain the concept being explicated, relative to that provided during Elicitation.
- It does not contain insufficient information. This means that it lacks no information that by its absence could underconstrain the concept being explicated, relative to that provided during Elicitation.
- It does not contain erroneous information. This means that it contains no information that is not consistent with that provided during Elicitation.
- It is written entirely in relative semantic primes. "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.

We next present the process for creating a CLEAR explication.

Activities. In order to construct a CLEAR explication for a given target, the following steps are executed.

1. Draft a definition using available content sources.

The content sources available for this step are the category records produced in the previous phase, the original materials, outside references when useful, such as Wikipedia [67], and informant consult when necessary.

The drafting process is as follows. First, attention is paid at this point only to the content constraints of explications, that is, the amount and integrity of the information provided. Given the available sources, the set of attributes that must be accounted for is assembled. This set should include everything there is evidence for in these sources, and exclude anything for which there is not evidence. If inconsistencies or other issues are apparent, informant consult should be sought, now and/or during a later validation pass. The set of supported attributes is then refined into a grammatically correct statement that accounts for them all, which should at this point be a semantically faithful description of the signified, relative to the available sources, awaiting further manipulation to abide by other constraints.

2. Reduce this definition in accordance with specified structural constraints.

The draft is now reduced in the style of Reductive Paraphrase as reviewed in the previous chapter, abiding by its structural constraints, except for the easement in the requirement that the explication reach the level of semantic primes. It has been recognized 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" [19]. Since reduction to absolute primes gains us no advantage, and adds unnecessary effort, we therefore reduce only to relative primes, or the level at which intersubjective agreement about the meanings of words occurs spontaneously among members of the speech community in question. Since there is no way to know a priori where intersubjectivity occurs, this level is reached in practice through trial and error, accompanied by informant consult. Since the reduction is only taken to the level of relative rather than absolute primes, we call this process Partial Reductive Paraphrase (PRP).

Two kinds of reductions are necessary to abide by the constraints of PRP explications. First, obscurity must be removed, that is, an explication may contain no terms or phrases less significant to the user than the explicand [19]. To remove obscurity, rephrase any term or phrase in the draft in simpler terms. Second, otherwise non-predictiveness must be removed, that is, an explication may contain no terms or phrases that by their nature hinder prediction of accurate range of use of an explicand [19]. To remove otherwise non-predictiveness, rephrase any occurrences of the devices associated with it by, for example, enumerating the class covered by “etc.”, or committing to the sets of events covered and not covered by subjunctives such as “may” and “might”.

Finally, as additional levels of reductive paraphrase are completed, it must be verified that the resulting explication tree remains free of cycles. Cycles in the tree indicate circularity, or a description that directly or indirectly uses one or more terms or phrases that themselves are defined in terms of the target term or phrase [19]. 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. Obvious circularities 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. The presence of either implies that the level of intersubjective agreement has not been reached. To detect circularities, the explication tree can be traversed in search of cycles. To remove them, rephrase the offending words or phrases appropriately.

Use of Representation Products. The set of explications ultimately generated during Representation serves the remainder of the process by providing the basis for construction of the CLEAR Knowledge Base, to be performed during Integration. Since the explication of a target may contain other targets, it is possible to conceive of a graph that records these interdependencies. The CKB instantiates them and allows their analysis.

Alternatively, even if Integration were not to be undertaken, the explications could be collected into a more traditional requirements glossary, but with fewer of the definitional deficiencies otherwise common to the entries in such collections.

5.1.2.4 CLEAR Phase 4: Integration

The fourth phase, Integration, concludes the manipulative processing of the targets selected during Phase 1 and processed to intermediate states in Phases 2 and 3. The Integration phase of CLEAR guides the assembly of the refined and enhanced representations into an integrated structure permitting collective manipulation and analysis. This section addresses the goals, construction and use of the products associated with Phase 4, Integration.

Goals. This section states the goals of the phase, describes their contribution to solving the greater research problem, and introduces the strategy to be used in realizing the goals.

Statement. The goal of Integration is to collect and organize the representations resulting from the previous phase into a single structure that instantiates their relationships to each other and permits analysis.

Contribution. This goal contributes to the ultimate improvement of the fidelity of the communicative transaction in two ways. First, it provides a navigable repository for users such that they may check their understanding directly, or link phrases in development materials to entries in the structure. Second, it provides a mechanism by which to measure and calculate properties of the included targets that allow previously unavailable insight regarding their individual risks of miscommunication. This information can be used to preferentially allocate resources to, for example, maintenance and improvement of particular areas of the requirements, or to increased testing of particularly affected parts of the system.

Strategy. The strategy for realizing this goal is to create individual entries in a database for each of the targets, with a field in each entry to hold the target's explication. Then every invocation of a target within any explication is linked back to the entry for that target. Instantiation of these connections results topologically in a graph upon which measurements may be taken. We call the structure a CLEAR Knowledge Base (CKB); its definition is further elaborated in the Integration Definition below.

Integration Definition. This section describes the sources, products, constructs, and activities defining the Integration phase of CLEAR.

Sources. The Integration phase takes as input the set of Partial Reductive Paraphrase explications generated during Representation, as well as the set of category records generated during Elicitation.

Products. The products of the Integration phase are a CLEAR Knowledge Base and the results of its analysis.

Constructs. Execution of the Integration phase is facilitated by the CLEAR Knowledge Base (CKB) construct. The CKB collects and organizes a set of explications and instantiates relationships among them thereby enabling analysis.

Specifically, a CKB is a set of entries, where an entry contains the following fields.

- The target phrase. This is simply the original signifier for the concept.
- The target explication. This is the representation as constructed during the previous phase.
- Category Hierarchy information. This includes supercategories, subcategories, and siblings of the category associated with the concept. This provides orientation information for users regarding how the concept might fit into their larger sense of the domain. It further provides an alternative organization to alphabetical order for presentation of the entire set.
- Additional notes. This field holds any observations made by the analyst that are outside the scope of CLEAR, but nonetheless potentially valuable. For example, information regarding a concept that is not appropriate to an explication because

it does not serve prediction of the accurate range of use, but that is still noteworthy in the pursuit of other system goals, may be recorded here.

In addition to collecting a set of entries together, the CKB establishes links among them. If an explication for a given target contains any phrases that are also targets, these target phrases are linked to their respective entries. In this way a graph may be visualized in which every target is a node and every dependency is an edge. Edges are unidirectional, in the direction of the reduction.

Given this graph, a number of interesting measurements become possible. First, the depth and breadth of each explication can be determined. The depth of an explication is the number of total levels in its explication tree, while the breadth is the number of other target phrases upon which the explication immediately depends. This information is useful in characterizing the skills needed to gain working control of a lexicon, insofar as depth is a measure of semantic distance, or remoteness, from intersubjective agreement or common understanding, while breadth is an indicator of the variety of notions that must be integrated together in the understanding of a concept. This indicates, for example, that a concept with many levels in its explication is quite removed from spontaneous intersubjective agreement and therefore more prone to miscommunication than is a concept with few.

More interestingly, the number of unique nodes per explication, as well as the number of ways a concept might be encountered in the lexicon of the domain, can also be determined. The number of unique nodes per explication represents the number of unique explicated concepts that participate in the explication of a given concept. Complementarily, the number of ways a concept might be encountered in the lexicon is the number of explications in which an explicated concept participates, including its own. The implications of these two measures are significant. Specifically, the number of unique nodes in an explication is an indicator of the number of possible ways to misunderstand the concept represented by that explication, since each node represents a critical piece of the whole.[1] Further, the number of explications in which a concept participates is an indicator of its pervasiveness and influence throughout the critical domain knowledge of the application. If a concept is misunderstood, it affects the validity of potentially much wider partitions of the requirements than just those in which it is directly referenced, since it compromises the understanding of as many other concepts as rely on it, as well as any that rely on those, and so on.

These measurements are analogous to the arguments in the standard formula for the computation of risk, that is, the probability of an event, multiplied by the severity of damage should that event occur. The probability of misunderstanding of a concept can be analogized to the number of possible ways there are to misunderstand it, and the severity of damage can be analogized to the pervasiveness and therefore influence of that concept throughout the domain knowledge of the application. Thus these observations allow us to propose for the first time a quantitative measure of the risk of miscommunication of a critical concept with reference to its application domain:

1. It is recognized that concepts on which a given concept depends can be misunderstood in combinations in addition to one at a time; this does not change the general argument.

The Risk of Miscommunication of Concept

is equal to

the Number of Unique Nodes in that Concept's Explication,

multiplied by

the Number of Explications in which that Concept Participates

This formula results in a natural number we define to be the Risk Index of a concept. The Risk Index allows us to see, for example, that while explication A might have more nodes in its explication than explication B, it poses less of a risk than explication B because explication B pervades more of the total domain knowledge of the application, participating in sufficiently more other explications to have relatively more influence.

Activities. In order to construct and analyze a CLEAR Knowledge Base for a given set of targets, the following steps are executed.

1. Create each entry. This entails completing each field using the products of previous phases as appropriate.

2. Instantiate dependencies among the entries. This entails linking each target used in an explication to its own entry.

3. Perform the defined graph measurements. This entails the required graph walks that enable the necessary counts as defined above.

Use of Integration Products. The CLEAR Knowledge Base ultimately generated and analyzed during Integration serves the method's goals by providing a repository of domain application concepts chosen according to influence and provided with representations that reduce the risk of misinterpretation by reducing the incidence of phenomena that cause it. The CKB can be used as a standalone resource to serve as the central authority on local usage to be consulted when necessary, or can be directly linked to system documentation, such that any instance of an included target can invoke the target's entry in the structure.

In allowing analysis, the CKB further offers previously unavailable insights regarding the semantics of the application domain that can be used to guide the allocation of resources to areas most likely or most severely to be affected by potential miscommunication. The Risk Index allows resources to be directed preferentially, for example, at validation and testing of system functions relying on concepts with high Risk Indices. It further provides the basis for prioritization of corrective action in requirements and other system documentation.

Finally, analyses from different projects can be compared to generate meta-analyses of domain commonalities and differences that could shed light on, for example, the typology of miscommunication events generally. Knowledge of any universals could further guide the minimization of such events in the engineering of safety-critical systems.

5.2 Chapter Summary

This chapter presented the canonical definition of the CLEAR Method. Specifically, it provided global definitions governing the process as a whole. Details were then presented of the goals, constructs, and activities associated with each of its four component phases, Selection, Elicitation, Representation, and Integration. The next chapter evaluates CLEAR in application.

Today's scientists have substituted mathematics for experiments, and they wander off through equation after equation, and eventually build a structure which has no relation to reality.
--Nikola Tesla

6 Empirical Assessment of CLEAR

In this chapter we evaluate whether the Method as presented in Chapter 5 productively addresses the Problem as presented in Chapter 2. Recall that to improve requirements communication, the fidelity of communicative transactions must be improved such that environmental needs for trans-boundary communication, accuracy and precision are demonstrably supported, and environmental constraints affecting practical utility are accommodated. Our task then is to determine whether use of CLEAR helps reach these goals. This chapter details evaluation activities designed to assess whether application of CLEAR results in the kinds of improvements in which we are interested. In particular, we present an evaluation strategy and then report on the empirical activities executed in order to realize it. The results from each exercise are then synthesized into the main findings of the combined undertaking.

6.1 The Evaluation Strategy

This section describes the strategy for evaluating CLEAR via empirical investigation using two complementary research methods. Specifically, we want to demonstrate improved fidelity of communicative transactions, achieved such that constraints of the requirements environment are respected. We do this in two parts. First, we seek a theoretical result: to show that CLEAR enables better comprehension of domain semantics. We then seek to repeat and extend this result in practice: to demonstrate improved quality and organization of domain semantics, at a greater scale, in a realistic environment. To do this, we want to execute CLEAR and observe its effects with regard to these goals; we thus require empirical methods.

6.1.1 Empirical Methods

Empirical methods provide frameworks for making observations in a variety of environments such that knowledge claims inferred from those observations are rationally supportable. There exists a choice of evaluation methodologies, environmental control generally traded off with environmental realism among them, since the more controls that are placed, the less natural and genuine are the interactions among multiple variables in any situation. This implies that different options are appropriate to different purposes and environments. In the case of this work, we want to demonstrate contributions with several desired properties. In particular, we want to both demonstrate with high confidence the theorized existence of an effect, but also show that this effect persists and scales in reality.

We thus design two of these activities, a formal experiment and a case study, with the goal of complementation. These approaches are introduced below.

6.1.1.1 Formal Experiments

Formal experiments are empirical observation frameworks in which variables of interest are isolated through the controlling away of potential confounds, and statistical tests of hypotheses are thereby made possible that can deliver formal mathematical support or refutation with high confidence. They have a narrow focus in that most of the context of any variable must be controlled away in order to examine it independently and any findings concern it and it alone, but within this narrow focus, they can make statistically significant claims. Formal experiments are useful in environments where the strength of statistically-grounded results is desirable and the necessary controls are possible; in such environments they can make very strong arguments for the explanation of an observed effect.

6.1.1.2 Case Studies

In contrast to formal experiments, case studies use analytical rather than statistical means to support inferences made based on structured observation. They begin from a research question and a set of propositions and use a careful, guided process of evidence collection and argument construction in order to support or refute those propositions. They further have a broad focus in that they attend to potentially many dynamics at once, rather than isolating individual variables. They are useful in a number of environments where formal experiments are not possible or practical, and preserve in these contexts the opportunity for rigorous support of an argument in the absence of formal controls [44, 70].

In particular, case studies are appropriate in the following settings:

Controls are impracticable. In many environments, social, psychological and other organizational factors interact in unique and complex ways, with "(m)any more variables of interest than data points" available [44]. The interactions of these variables are too many and too diverse to be able to control out confounds and isolate variables of interest. However, we still want to rationally investigate phenomena in such environments because they are interesting, and observation without statistically viable controls is often the only available option.

Controls are detrimental. Further, these environments represent the realism of situations that is difficult to achieve in controlled environments. In order to maintain realism in many objects of study, they cannot be examined independently of their context and the many complex influences that shape them, since their behavior might then change [70].

Interest in side-effects. In addition, since the phenomena user study are often actual natural or engineered systems or processes in real environments, there can be dynamic implications for related or interacting systems and processes. We are interested in these side-effects as well, and the broad focus of case study observation attends to them in ways that the narrow focus of formal experimentation does not.

In other words, while controlled experiments can provide theoretical demonstration of an interesting result, we can not be, nor do we want to be, thoroughly reductionist with regard to phenomena involving the behavior of humans in real environments. A well-designed case study allows the rational investigation of such phenomena and rigorous support of resulting inferences.

6.1.1.3 Multi-Method Research

More interestingly, the existence of both of these methods allows their combination in multi-method research such that a macro-study can complement controls with realism, narrow with broad focus, and theoretical with practical results [12]. In fact, the combined effect is greater than the power of either alone, because the results can be triangulated. "Triangulation is the application and combination of several research methodologies in the study of the same phenomenon" [51], and concerns the determination of which elements of the findings of each investigation are congruent with those of the other. For multi-method research, this kind of congruence constitutes empirical replication [44], reinforcing and strengthening each of the claims that has a congruent analog.

6.1.2 The Strategy

The evaluation strategy adhered to for this work was based on the desire for the above described complementation of statistical control and environmental realism. In practice, two versions of CLEAR at different levels of maturity were evaluated. We first performed a formal controlled experiment to determine the value of core CLEAR constructs. This experiment answered questions, both specifically regarding the supportability of the experimental hypothesis, and more generally as to whether we were on the right track in the development of the method. It also provided insight for method refinement as a result of experience in actually applying some of its elements; the experiment in effect told us that even though we had made a theoretical contribution, we would have to both make adjustments to CLEAR, as well as rationally structured observations of those adjustments in action, for CLEAR to eventually demonstrate practical utility.

We thus refined CLEAR accordingly and then took it into a more realistic environment, for evaluation that sought to deliver both empirical congruence as well as answers to an extended set of research questions. The details of both evaluation experiences are presented below.

6.2 The Evaluation Activities

This section reports on the design, execution, outcomes and interpretation of empirical activities conducted to evaluate CLEAR.

The previous chapter introduced CLEAR as a response to recalcitrant communication issues in requirements engineering. CLEAR exploits results from linguistic analysis to provide procedural support for accommodation of human habits and limits in the communication of application domain semantics with pragmatically sufficient accuracy and precision.

In order to demonstrate that CLEAR has the kinds of value in which we are interested, we must demonstrate a causal relationship between the interventions we propose and desirable outcomes. We first demonstrate such a relationship in a controlled environment via formal test of a hypothesis. We then demonstrate it in a realistic environment via an explanatory case study using a refined and enhanced version of the method.

6.2.1 Activity 1: The Formal Experiment

We here describe a formal experiment undertaken to test a hypothesis regarding the utility of core CLEAR concepts for improving the comprehensibility of application domain knowledge.

6.2.1.1 Methods

This section details the context, design and execution of the experiment.

Context. The target application for this experiment was an international standard with which maritime track control systems must comply in operation and performance. It contains both functional and non-functional requirements, including several relating to the fact that such systems are safety-critical.

For purposes of the experiment, the requirements consumers were made up of the set of students in an advanced undergraduate software engineering class at the University of Virginia on a particular day during a past semester. This population allowed both the size of subject pool needed (each group with n >= 30 to enable use of desired statistical measures), as well as the enforcing of required controls via the enclosed classroom environment and limited time period.

Design. The formal experiment approach begins from a research questions and a hypothesis. The research question characterizes the topic or phenomenon under investigation while the hypothesis makes a formally testable prediction within the scope of the research question. These are followed by the design type, which characterizes what kinds of controls are in place in the observation environment, the operationalizations, which define the variables under observation and how they are to be manipulated and measured, and the instrumentation, which defines the collection mechanism for the data.

Research Question. How does the use of a protoCKB affect subjects' comprehension of application domain semantics intended to be communicated by requirements?

Experimental Hypothesis. Use of a protoCKB will improve subjects' comprehension of application domain semantics intended to be communicated by requirements, compared to the comprehension achieved given the requirements alone.

Null Hypothesis. The null hypothesis is then that no significant difference in comprehension will be demonstrated with use of the protoCKB. In this experiment, we will reject the null hypothesis if improvement in comprehension is seen at the $p = .05$ level of significance or better.

Design Type. We next describe a two-group true random experiment designed to assess the supportability of the hypothesis. A two-group true random arrangement is one in which two different randomly assigned subject groups receive two different interventions, one of which may be degenerate, that is, no intervention. The differential effects of the interventions on the two groups are then measured with respect to variables of interest. This is a design that has been shown to counter a large number of threats to validity, and is therefore a useful tool for assessing causal relationships between the application of an intervention and an observed effect [59].

In this case, the two groups were made up of the students present in the previously described software class on the experiment day. 114 total subjects participated, and were randomly assigned to the two groups of 57 subjects each. The groups were thus statistically equivalent with regard to social factors such as age, gender and level of education, as well as with regard to level of prior exposure to the domain.

Operationalizations. There are two variables of interest in this experiment. The independent variable is application of the intervention, and the dependent variable is its observed effect. In this case, the independent variable is manipulated by providing the treatment group with a protoCKB in addition to the natural language requirements for use in completing a comprehension task, while the control group receives only the natural language requirements.

The dependent variable is the level of comprehension subjects demonstrate in the completion of this task. Comprehension is measured by performance of subjects on a diagnostic test of the application domain semantics necessary to apply the given requirements with validity.

Thus if a statistically significant causal relationship between the independent variable, that is, use of a protoCKB, and the dependent variable, that is level of comprehension, can be demonstrated, then we will reject the null hypothesis and say that the intervention was the statistically supportable source of the improvement.

Instrumentation. Three pieces of instrumentation are necessary to instantiate this experiment: a set of source materials constituting the original requirements to be provided to both groups, a protoCKB generated by application of protoCLEAR to those requirements, to be provided to the treatment group, and a diagnostic test of comprehension of the requirements to be taken by both groups using their respective provided materials.

The original sources provided to both groups consisted of a self-contained excerpt of the track control standard described above. The excerpt focuses on functional requirements of maritime vessels, and the domain knowledge under consideration is constrained to that directly associated with these requirements.

The protoCKB provided to the treatment group was generated by an early version of CLEAR in which selection was entirely by manual inspection, the content of explications was informed by expert consult, and the structure of explications was constrained by an early version of Partial Reductive Paraphrase that included the no-cycle rule as well as the requirement for intersubjectivity of leaf-node concepts.

The diagnostic test of comprehension of the requirements consisted of 8 closed-form questions, which were closed-form in that they had clearly correct answers, as opposed to being open-ended, in which a subject might be asked to discuss an opinion or argue a point. The test was constructed in consultation with an expert in the application domain, familiar with the standard, in order that this intent be represented with validity. The expert confirmed that the questions included in the final version of the test did in fact concern application domain knowledge that any developer tasked with working on such a system should know in order to be able to model entities within the domain with validity and manipulate these modeled entities in a system design and implementation. The eight question of the test follow:

1. **What is an active track? How does it relate specifically to motion of a ship?**
2. **Which factors must be taken into account in order to determine whether an approach maneuver to a track is safe?**
3. **What information does the track control system need in order to calculate dead reckoning position, for example, in the case of position sensor failure?**
4. **Which entities/quantities must be known and/or measured in order to accurately model a ship's bearing?**
5. **What is provided by an EPFS? From where does it get this information?**
6. **What is the system expected to do if a course difference limit is exceeded? For whom is any resulting information intended, that is, who is the user?**
7. **What is the difference between an alarm and an indication?**
8. **If a track control stopped alarm is initiated, to what kind of control does the ship revert? Briefly explain how this other form of control works.**

Execution. This section recounts the instantiation of the design in the experimental environment. It describes both how the necessary controls were achieved as well as the activities that unfolded within the confines of those controls.

Randomization and Physical Arrangement. To accomplish randomization of the subject pool, the subjects were assigned natural numbers in order from 1 through 114 as they entered the lecture hall in which the experiment took place. Once all students had arrived, students assigned even numbers were separated from students assigned odd numbers to opposite sides of the hall, with a wide aisle in between. The randomization mechanism controlled for potential confounds with test performance, such as the possibility of groups of friends who might both study together and sit together, which might lead to the possibility of a cluster of subjects in one group with convergent skills, or the possibility of differential performance achieved by early arrivals versus late arrivals, which could correlate with conscientiousness. The physical separation controlled for the possibility of subjects from one group seeing source content or test answers belonging to subjects of the other group.

Test Completion. A coin flip determined which group became the treatment group and which became the control group. The groups were then given their corresponding sets of materials, including the test and the source requirements for both groups, as well as the protoCKB for the treatment group. Subjects were instructed not to look at the materials until the start of the test period. Prior to the start, subjects were given instructions to complete the diagnostic test as accurately as possible within a 45 minute time limit, and to use the materials at their disposal in whatever manner and to whatever degree they found them useful. During the 45 minute period, 3 proctors were available to monitor the enforcement of experimental controls and to answer questions regarding the instructions. However, the proctors were not allowed to assist in any way with subjects' considerations of their answers to the test questions.

Grading. The structure of each of the closed-form questions was such that a complete answer included two main pieces of information. For example, a correct and complete answer to "What is the system expected to do if a course difference limit is exceeded? For whom is any resulting information intended, that is, who is the user?", must include indication that an alarm is to be sounded as well as indication that the user in question is the officer of the watch on the ship. Thus, each question was graded out of a potential 2 points, with each subject receiving a 0, 1, or 2 as appropriate, according to the corresponding number of elements of the correct answer contained in his response. For purposes of consistency in scoring, a number of additional constraints were applied. Specifically, the test was graded by one person, one question across the entire subject pool at a time, with the tests in random order such that the grader did not know to which group the subject belonged.

6.2.1.2 Results and Discussion

This section presents the results of the domain-knowledge comprehension test and their interpretation.

Results. Table 1 displays the test scores resulting from the experiment. It indicates how each of the two subject groups performed on each question, as well as on the test as a whole. The performance values are presented as both raw scores (out of 2 points for each question and out of 16 for the overall test), as well as percentages. In addition, the performance difference between the groups is indicated in the final column.

Specifically, the treatment group scored 74.34% overall while the control group scored 44.96%, indicating that the treatment group demonstrated sufficient comprehension of an additional 29.39% of the domain knowledge under consideration, relative to the control group. Further, the treatment group achieved higher scores on 7 of the 8 questions individually, demonstrating sufficient comprehension of greater than 50% more of the domain knowledge under consideration in 3 cases and greater than 40% more in an additional case.

To determine the likelihood of achieving these results by chance variation in the performance of the randomized groups, we computed this probability via t-test, which was used to accommodate the fact that this is a discrete rather than continuous data set. This

Table 1: Test Performance of the Subject Pool

	Control Group	Treatment Group	Difference: Treatment less Control
Overall	7.193 (44.96%)	11.895 (74.34%)	29.39%
Q1	1.070 (53.51%)	1.877 (93.86%)	40.35%
Q2	1.561 (78.07%)	1.316 (65.79%)	-12.28%
Q3	.544 (27.19%)	1.596 (79.82%)	52.63%
Q4	.439 (21.93%)	1.526 (76.32%)	54.39%
Q5	1.018 (50.88%)	1.509 (75.44%)	24.56%
Q6	1.228 (61.40%)	1.421 (71.05%)	9.65%
Q7	.474 (23.68%)	1.526 (76.32%)	52.63%
Q8	.860 (42.98%)	1.123 (56.14%)	13.16%

computation allowed us to reject the null hypothesis with $p < .01$, that is, the hypothesis that our intervention would result in the predicted improvements is upheld.

Discussion. For our definition of comprehension, it is clear from the test data that the treatment group performed better, that is, the group that had access to a protoCKB in addition to the original natural language requirements was better able to correctly and completely answer questions that were confirmed by a consulting domain expert to be representative of domain knowledge necessary to perform tasks associated with development of systems that conform to this standard.

We interpret these results to indicate that the protoCKB did in fact provide necessary information with sufficient accuracy and precision better than did the natural language requirements alone, as opposed to leaving omissions and assumptions about developers' individual knowledge of the domain unaddressed. For example, to answer a question regarding the pieces of information necessary to calculate dead reckoning position, subjects in the treatment group had access to a definition of what dead reckoning meant in the domain in question, this definition was located alphabetically in a comprehensive list of domain concepts, and other domain concepts upon which its definition relied were indicated and themselves defined. No such definition was provided in the natural language requirements[1], and comprehension of this information was left to observation of use of the term in context, and/or prior knowledge that the developer was assumed to possess.

Indeed, many of the subjects in the control group did in fact get this answer completely correct, since they were able to use either the context, prior knowledge, or a combination, in deducing what must be necessary to calculate dead reckoning position. Several subjects indicated as much in their answers, using qualifiers such as "Presumably", "As far as I understand", and "I would think that". The issue, however, is precisely the unpredictability of this activity. The fact that many *other* subjects in the control group showed *very little* comprehension on this question demonstrates the diversity of domain exposure and reasoning skills possessed by any body of potential developers.

This variety, often estimated to be at least a factor of 5 difference between extremes [42, 5] has been corroborated as a highly influential factor in software productivity and quality [13]. A useful approach to coping with this variety and its attendant influence is one that in a systematic way reduces the reliance on developers' prior knowledge of domain-specific information and more importantly, on their assumptions regarding what terms mean. Recall that a main finding of the linguistic analysis was that the heuristic of cognitive economy, which works by exploiting the value of making assumptions in known domains, breaks down when communication is across domains, because assumptions naturally made become less and less valid as the domains grow further apart. CLEAR is designed to systematically reduce the need to make assumptions, thereby reducing the incidence of invalid ones. Our data support the assertion that use of a protoCKB in conjunction with this set of requirements reduced the need for subjects to make assumptions or rely on prior domain knowledge, in effect rendering the diversity among subjects less influential in their comprehension of domain semantics.

The reader might assert that since the treatment group had more information, it was a foregone conclusion that they would perform better. However, as illustrated by the data from question 2, the treatment group *can* in fact do worse. On this question, the treatment group underperformed relative to the control group by demonstrating comprehension of 12% less of the information in question. On reexamination of the experiment materials, we believe there is a plausible explanation. A correct and complete answer to question 2 was actually clearly and explicitly available in the natural language requirements. This by itself indicates that all subjects had the potential to achieve full credit for this question based on information explicitly provided, with no recourse to assumption or prior knowledge. However, the treatment group in this case was also faced with additional information, meaning more raw quantity to process before making a decision regarding the answer, and thereby the potential for a lack of clarity as to which information was most important or relevant to the answer. Worse, duplicate attention in the documents opens the door to the potential for conflicting information. We believe that the underperformance of the experimental group on this question can be attributed to these factors, and we note that more information does not directly imply better performance. Quality and structure of the information are the key factors, and CLEAR attends to exactly these.

1. The requirements from which the test excerpt was taken did include a short glossary, but neither dead reckoning position, nor most of the other domain concepts brought out by the process, were included. Further, dependencies among domain concepts were not addressed. The excerpt provided to subjects did not include the original glossary, however, this glossary included only one term that might have made a difference to the performance of the control group on one question, had the definition provided essential information. However, that definition was circular, and explicitly included none of the information necessary for a correct answer, that is, subjects would still have needed to resort to assumption or prior knowledge.

6.2.1.3 Threats to Validity

This section describes the known threats to validity of the inferences drawn by the experiment and how they were mitigated or accommodated.

Construct Validity. Construct validity concerns the degree to which the chosen measurements support the inferences made by the experiment [70]. The primary measure used in this experiment was the level of comprehension of application domain knowledge demonstrated by subjects.

Comprehension was operationalized by performance on a test of the domain knowledge in question, and via this definition, we were able to observe and quantitatively measure the effect of our intervention and compare it to a control situation. In this regard, the operationalization of comprehension allowed the test and ultimate support of the experimental hypothesis.

Internal Validity. Internal validity concerns the degree to which a causal relationship can be established between two events, in this case, the intervention and the result [70]. If the result might have arisen because of some third, unconsidered factor, for example, internal validity would have been compromised.

There are three criteria for establishing a causal relationship. Specifically, the intervention must temporally precede the effect attributed to it, the cause and effect must covary, and there must exist no plausible alternative explanation for the effect [11, 59].

The temporal precedence criterion necessary for establishing a causal relationship is met in this experiment by design: subjects had access to the materials and were able to use them to whatever degree they wished before answering any question. In addition, if a subject answered a question and later decided to change the answer based on newly discovered information, he was permitted to do so.

The covariance criterion is met if the performance of the treatment group can be demonstrated to be better than the performance of the control group; this would indicate a positive correlation between the independent and dependent variables. This differential was in fact observed.

A number of alternative explanations for the observed effect are ruled out by design. For example, a diffusion threat arises when members of one group learn details about the experimental circumstances of the other group and knowledge of this information has the potential to affect their performance [59]. In the case of this experiment, if there were a break and members of the two groups were not prevented from discussing with each other the materials to which they had access, information provided within restricted materials might diffuse to those not supposed to have it. This could affect the answers subjects provide on the test. This threat was minimized through the enforcing of a no communication rule between the groups for the duration of the experiment. Further, the groups were physically separated by a distance greater than that at which any content in one group's materials could be discerned by a subject in another group.

Further, the operationalizations and controls in place allowed the data to be subjected to a test of statistical significance, which enabled rejection of the null hypothesis.

We thus conclude that the internal validity this experiment was not compromised and that the intervention applied in the treatment group was in fact the cause of the observed improvement.

External Validity. External validity concerns the generalizability of the findings of the experiment to a wider context than that in which the experiment took place [70]. In this case we are interested in whether our results using advanced undergraduate software engineering students generalize to a wider body of professional software developers.

In doing empirical research, a tradeoff exists between the ability to control the environment and the ability to obtain environmental realism. Generally, if statistical controls are desired, environmental realism suffers, and vice versa [70, 44]. In the case of this experiment, statistical controls were desired, and since it was not possible to achieve the controls and size of subject pool we needed for the experiment in a professional development environment, we used advanced undergraduate software engineering students. This allowed the experiment to be executed efficiently under the indicated constraints, ultimately enabling a statistical test of a hypothesis, an activity not generally otherwise possible in software engineering research of this nature.

The fact that the subjects were students means that the experiment results can only be statistically generalized to other students, in particular, advanced undergraduate software engineering students, possibly only others at the same institution. However, analytic generalization allows some amount of support for the assertion that the results do in fact apply to professional developers. First, the subject pool has several characteristics in common with the target population, namely field and level of education, in fact given their majors, these students are as likely as anyone, and perhaps more, to be in the position of developing software in the near future.

We recognize that the subject pool does have on average less experience using requirements and developing software than does the target population, and that this could have a potential effect on, for example, how familiar they are with variations on project documentation, how to efficiently navigate it, and related issues of experience. However, we do not believe this difference to have had an influence on the quality of answers we received on the diagnostic test. Specifically, the materials provided to each group were presented in accessible formats that filtered out extraneous infrastructure and boilerplate from the original requirements. That is, the focus was on reading comprehension of a passage, and as such was made orthogonal to software engineering experience. In addition, since the focus was in fact reading comprehension, the main characteristic of interest in a potential subject pool such that it can represent a target population is that they share the natural human linguistic endowment, which they do. Thus we do not believe the external validity of this experiment was severely compromised by having used students as subjects.

Reliability. Reliability concerns the extent to which an experiment can be repeated with the same results [70]. In the case of this experiment, the key threats to reliability are the subjectivity of the experimenter in development of the protoCKB and the diagnostic test, and in execution of the grading methodology.

The protoCKB had the potential for variation in the set of concepts chosen for inclusion as well as in the content and structure of definitions. We mitigated this potential by using selection criteria for inclusion during manual inspection, using expert consult combined with CLEAR principles on the content of definitions, and by using further CLEAR principles in the form of syntax requirements on the structure of definitions.

The diagnostic test had potential for variation in the set of questions chosen for inclusion and their wording. We mitigated this potential by using expert consult in the determination of areas of the domain semantics that were more critical than others, as well as in validation that the questions chosen reflected those areas, and worded the questions such that a format was adhered to that allowed consistency of shape and therefore comparability of the answers sought.

The execution of the grading methodology had potential for variation insofar as grading by multiple graders and/or under multiple sets of environmental conditions can affect grades. We mitigated this potential by using a single grader, who graded one question at a time across the whole subject pool, in random subject order, and finished the grading of each question across the pool in a single session, such that environmental stability was maintained.

We thus do not believe that the reliability of this experiment has been severely threatened by experimenter subjectivity.

6.2.1.4 Experiment Summary

The hypothesis of this investigation was that use of a protoCKB in conjunction with the standard requirements documentation of an application enables improved comprehension of application domain knowledge in comparison to the comprehension achieved using the requirements documentation alone. This hypothesis was tested via a two group true random experiment involving 114 subjects and an international standard of requirements with which the control systems of maritime vessels must comply in operation and performance. It was shown that the mean performance of the treatment group, which was provided a protoCKB, was statistically higher than the mean performance of the control group in a test of comprehension of application-specific information. This result was supported at the $p < .01$ level of significance. We thus conclude that the kind of systematic compensation for normal human variability embodied in CLEAR can promote the comprehensibility of application domain knowledge with higher fidelity.

We next report on a case study that provides a complementary and extended set of CLEAR results.

6.2.2 Activity 2: Case Study

Whereas the experiment just presented was designed to formally test a hypothesis, the case study described below was designed to investigate a related research question in a complementary way. Recall that a formal experiment can deliver statistical significance regarding the supportability of a hypothesis in a controlled environment, while case studies provide for the generation of rational inferences when controls are both not possible and even undesirable. This case study focuses on the more diverse, subtle, and complex interactions set in motion among entities participating in an application of CLEAR.

6.2.2.1 Methods

This section details the context, design and execution of the case study.

Context. The target application for this case study was a software-controlled, safety-critical medical device. Based on the notion of a walker as used by the elderly and others with ambulatory difficulty, it provides enhanced physical support and guidance to persons with limited ability in dynamic situations involving balance and motion. We will refer to the target application henceforth as the Walker Project.

The Walker Project is a research and development initiative spanning two engineering departments and a medical center at the University of Virginia. While not housed in a canonically industrial environment, the Walker Project nonetheless has a number of properties in common with industrial software development. Its focus is the development of a new product from conception through design and implementation, with the ultimate goal of public commercial consumption. Further, its development brings together individuals with varying backgrounds and multiple domains of expertise who must all communicate their contributions sufficiently among each other for the device to be realized. In addition, its documentation practices represent one of many variations present across software development organizations. In particular, the requirements for the Walker Project are spread informally across a number of documents with varying purposes, as well as located implicitly in the minds of project personnel.

In the execution of this study, we were allowed access to all existing documentation pertaining to the project, as well as consultation with all of its personnel in order to collect information required to build our artifact and to validate it once it was completed.

Design. The case study approach begins from a research question and a set of propositions. The research question characterizes the topic or phenomenon under investigation, while the propositions provide guidance regarding the kinds of observations to make in order to answer the research question. These are followed by the Unit of Analysis, which defines the boundaries of observation for the study, and the Study Questions, which define that being observed [70]. The design elements of this study are presented below.

Research Question. How does the execution of a CLEAR process affect the quality and organization of representations of critical concepts in an application domain?

Proposition 1. Execution of a CLEAR process will improve the integrity of both the form and the content of representations of critical domain concepts, thereby reducing opportunities for miscommunication involving these concepts.

Proposition 2. The products of a CLEAR process, by virtue of their organization, will further provide a substrate for analysis that generates insights not previously available regarding semantics of the application domain.

Unit of Analysis. The boundaries of observation for this study are defined by the entities and activities involved in a single application of the CLEAR Method. Thus a number of physical and virtual artifacts are of interest, as well as the manipulations undertaken on those artifacts, together with selected behavior of individuals with relationships to those artifacts. Ultimately, we begin observation with the assembly of the set of source materials, apply the method-defined manipulations to those materials and their derivatives, in combination with the solicitation of further information from personnel, and produce a CLEAR Knowledge Base for the application. These entities and activities bound the scope of observation.

Study Questions. The propositions presented above indicate two sets of derivative study questions. The first set concerns the method of comparison between representations available in the original resources and those produced by CLEAR. The second set concerns the measures possible given a CLEAR Knowledge Base.

The criteria for comparison between originally available representations and those constructed by CLEAR are as follows. For each phrase determined to be of interest by CLEAR Selection, in the original materials as compared to the CLEAR Knowledge Base,

Is there a representation?
Does the representation contain obscurity?
Does the representation contain circularity?
Does the representation contain otherwise non-predictive elements?
Does the representation contain too much information?
Does the representation contain too little information?
Does the representation contain erroneous information?

The organization of a CLEAR Knowledge Base lends itself to a number of measures that under analysis are proposed to lead to new insights regarding the application semantics. Following are the measures to be taken on the CLEAR Knowledge Base.

What is the distribution of explication depths?
What is the distribution of explication breadths?
What is the distribution of number of unique nodes per explication?
What is the distribution of number of explications in which each target phrase participates?

The execution of this study design is detailed next, followed by the presentation and discussion of results.

Execution. This section recounts the execution of the four phases of a CLEAR process on the Walker Project application. It includes detail of both the instantiation of canonical Method activities with real application data, as well as description of several decisions that were made in the process, along with the rationales for the resolutions chosen. Figure 3 abstracts the process as instantiated for this study.

Selection. In the case of the Walker Project, since no formal requirements documents exist and the requirements are distributed across other materials, the starting point for this application of CLEAR was the collected total of materials published or in submission regarding the project. This set included five separate documents, consisting of one funded

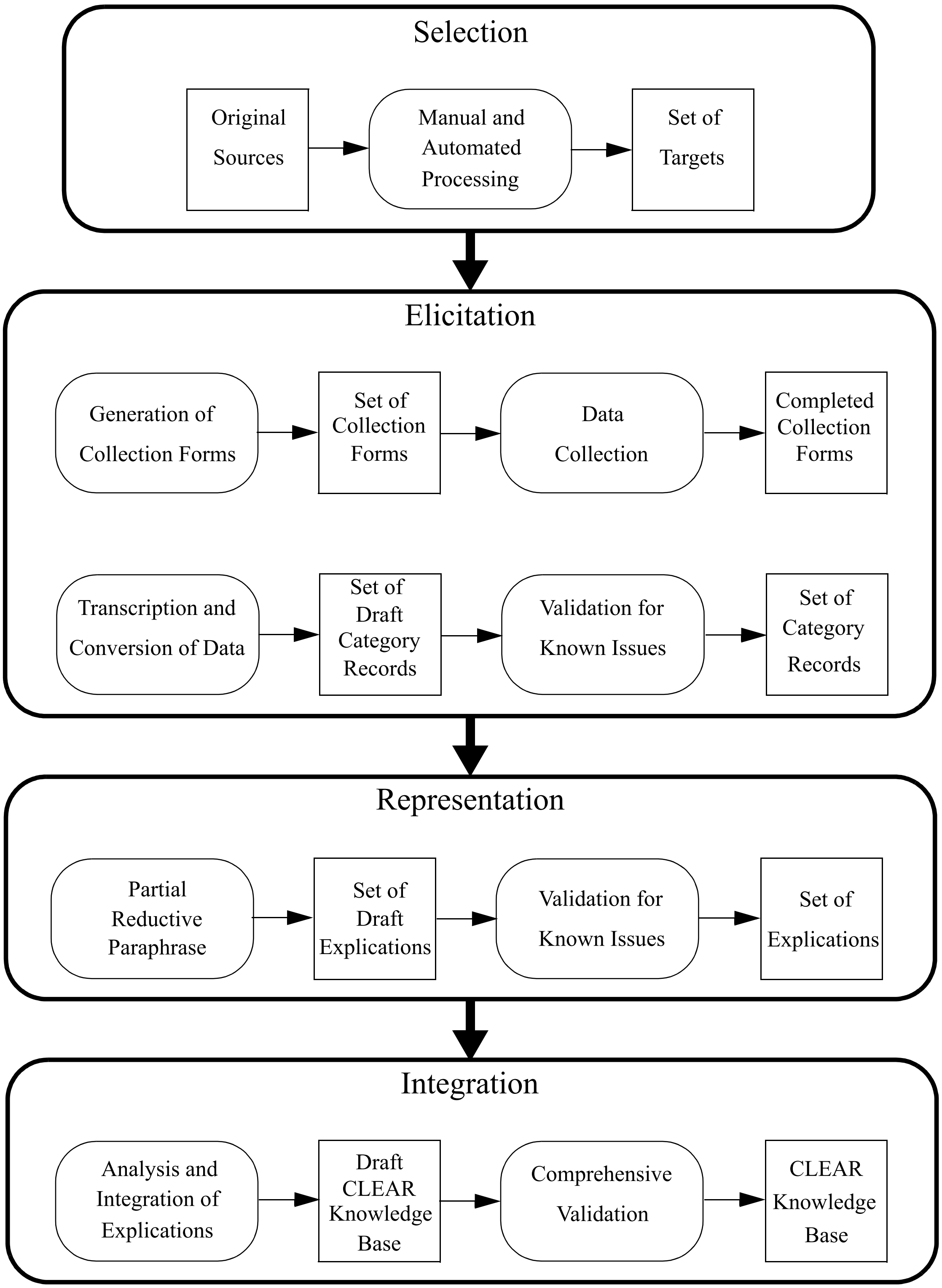


Figure 3. The CLEAR Method as Instantiated in the Case Study

grant proposal, three published research papers, and one research paper in submission (later accepted for publication). In the absence of purpose-built requirements materials, this set represents the most definitive record of the research and development agenda and goals of the project at the point in time when this process began. In particular, these materials record the fullest available specification for the device to be designed and built. This set provided the source material from which explicit or implied requirements were derived and upon which a CLEAR process was undertaken.

The first manual step was to separate in-scope material from out-of-scope material within the available documents. For purposes of this study, in-scope material was defined to be any text stating or suggestive of requirements for the device, in particular, for its control system, while remaining material was ignored. Thus sections of the documents on technical approach, technical descriptions, goals, and constraints of the device were selected, while sections of related work, for example, were not. The selected portions were then forwarded for further manual processing as well as for automated processing.

Next, a candidate set of target phrases was generated via inspection. This was accomplished by attending to two criteria during a thorough reading of the materials. The criteria for manual selection of a candidate target phrase were that it mark a concept that intuitively appears to be central to correct functioning of the system, and/or, that its meaning not be immediately transparent to one or more of a set of analysts, coming from outside of the project if possible. This second criterion models the premise that programmers who will implement the software either come from outside the application domain, or, if they are from within, then they are necessarily fulfilling multiple roles and are at best expert in only a fraction of the multiple areas brought together to realize the project goals.

Once the set was generated, certain adjustments were made for ease of management as well as to aid visualization of particular conceptual relationships. First, any grammatical variations of the same concept were conflated. Grammatical variations are variations of a phrase that differ in, for example, tense, number or word order. Phrases that differed in only these ways, such as “virtual moment” and “virtual moments”, or “obstacle avoidance” and “avoidance of obstacles”, were merged into single unique entries.

Second, selected phrases that participated in modificatory relationships, either as modifier or modified, were more closely examined in order to collect other candidates that shared dependency relationships with phrases of interest. For example, selection of “user intent” and “walker intent” might lead to consideration of other types of “intent”, and in fact “navigational intent” was selected in this manner. Further, “intent” supported the consideration of “intended path”, which itself supported the consideration of “possible path” and “predicted path”. The manual inspection process is necessarily a subjective activity, and though this fact offers advantages discussed in the previous chapter, the current step provides a form of guidance in bringing out phrases of interest that might have been overlooked during manual inspection despite the presence of selection criteria.

Once grammatical conflations were applied and modificatory relationships were examined and accounted for, a refined set of target phrases resulted and was held pending the results of automated processing.

Automated processing was then undertaken on the same portions of text that were considered manually. In this step, occurrence properties of all phrases present in the materials were computed. For purposes of automated processing, a phrase was defined to be any string of words that did not cross a clause boundary and did not begin or end in a part of speech that precluded its reasonable interpretation as coherent concept. In addition, the same grammatical conflations applied in the manual generation step held here as well, such that all variations of the same phrases were counted together. The occurrence properties of interest were total frequency with which a phrase occurred in the combined text under consideration, and how widely the occurrences of the phrase were distributed across the five separate documents of which the source materials were comprised. Recall that high relative frequency indicates that the concept participates in a relatively large percentage of the interactions represented by requirements, while high relative distribution indicates wide influence on multiple aspects of a system and thus the potential for misunderstandings to result in non-localized or uncontained faults that affect the system in pervasive ways [64]. In this case, the partitions reflected by the distribution measure were the different dimensions of the walker system introduced and defined by each original document, for example, how it uses the physics of the user's motion to predict his navigational goals, and how its control system actuates responses based on those predictions.

Pilot experiences with drafting small numbers of category records and explications, combined with knowledge of available resources for this effort, led to the goal of selecting 50 to 100 phrases for further processing. This would allow for both the predicted increase in number of phrases addressed in the eventual knowledge base that would derive from inclusion as a result of later steps, as well as for the additional steps of integration and validation that had not been undertaken before. To generate this set of 50 to 100 phrases, we applied a ranking scheme to the set of phrases for which occurrence properties were generated, and combined this information with knowledge of the set of phrases that was previously chosen for consideration manually.

Specifically, we first rank-ordered the phrases by distribution as well as by frequency, and then discounted phrases determined to be too common or abstract for consideration, such as “take”, “try”, and “general”. Recall as well that prepositions, articles, etc., were also accounted for through having been selectively defined out of the notion of a phrase. We then created the set comprised of those phrases distributed across at least 2 of the 5 original documents and having a frequency of at least 2 total occurrences. These thresholds reflect the fundamental difference between one and more than one, that is, instances, in the case of frequency, and locations, in the case of distribution. Intersecting them reflects that all members of this set simultaneously reach both of these standards, implying that the concepts they represent share a non-trivial level of criticality in the domain.

This set was then intersected with the manually-generated one, producing a preliminary set to be forwarded for further processing. This preliminary set was insufficiently large, so we then adjusted selection parameters to return a greater number. In particular, we prioritized distribution over frequency, because we judged the possibility of a pervasive fault to be more undesirable than a severe, but contained one[1], so we adjusted the thresholds such that all phrases of any frequency occurring across 3 or more of the original documents were returned. We then added to these several manually-selected phrases that

1. Alternatives to this decision may be chosen in environments where factors clearly motivate them, for example, when an organization chooses to perform CLEAR on only a small but critical portion of functionality; in this case distribution is less illuminating as an indicator of influence since only a small area is being examined to begin with.

did not meet occurrence property thresholds but were nonetheless judged to be in need of closer attention, resulting in a final set of 64 phrases that would be forwarded on for further processing. The 64 phrases were:

active behavior	active control
active mode	active steering
active walker	behavior based control system
behavior based system	biomechanical stability
bounded repulsive virtual moment	clearance maintenance
control laws	control rules
control system	control system state
dangerous situation	difficult situation
environmental context	environment map
environment perception	environment state
force sensing	goal directed navigation
handle forces	head angle
heuristic logic	historical data
historical state	instability
intended path	local environment
navigational intent	nonholonomic constraint
obstacle avoidance	opening passage
oscillation between behaviors	oscillations in navigational control
passive device	passive mode
passive robot	passive robotics
passive walker	possible path
postural stability	predicted path
propulsive force	propulsive moment
repulsive moment	repulsive virtual moment
roll	safety
safety hazard	shared control
slide	slip
steering angle	user intent
virtual moment	walker environment
walker frame	walker heading
walker intent	walker interface
walker state	walker steering

Elicitation. Five informants participated in the elicitation phase of this CLEAR process, all of them personnel of the Walker Project. The five informants included an electrical engineering professor, two electrical engineering graduate students, a computer science research scientist, and a medical automation research scientist.

The informants completed response forms that altogether addressed 62 of the 64 target phrases selected by the previous activities. Two of the phrases were unintentionally overlooked; explication was later completed for these two based only on information contained in the original materials, however, particular attention was paid to these two phrases during validation meetings in order to assure that the same character of information was eventually acquired for these as for the rest of the phrases.

The content of each response form was then analyzed and transcribed into a category template for each phrase. Recall from the previous chapter that category templates are intermediate artifacts that organize the content relevant to each element of the category definition for a concept, to be used in later construction of explications. In addition, notes on inconsistencies detected and other issues to be addressed during validation were recorded at this time, for example, it was unclear from the combination of source materials and collected information whether "walker environment" as conceived by the project personnel included the user or not. Also recorded at this time were notes regarding information of potential interest to eventual users that fell outside of intended use in explications. This information was ultimately preserved in the notes fields of the entries in question in the Knowledge Base. For example, information was sometimes available regarding how a concept in question interacted with other domain concepts to produce larger complexes of semantics that recurred together to play a collective role in the system. This information does not properly belong in an explication, the job of which is definitional and not encyclopedic (recall the discussion of the previous chapter), however, it is recognized that this information is indeed potentially useful and it was thus recorded.[1]

Copies of the category template and collection form used in this CLEAR application are included in the Appendix.

Representation. The analyst then drafted explications for all of the target phrases, using the content of their category records to guide the semantics to be included in each explication. For example, the category records supplied some basic definitional information as well as indication of dimensions of the concept that were critical to the application, and ways that the domain-specific version of the concept differed from possible other versions. Using this information together with the original source materials and occasionally other sources such as Wikipedia, the analyst sought to create a semantically faithful rendering of the appropriate range of use of the concept within the application domain, while also abiding by the structural constraints of legal explication.

To assure the structural integrity of the explications, explication trees were maintained in the process for the purpose of cycle detection, which reflected explication circularities. Draft explications were revised as necessary to remove these as well as obscurities and otherwise non-predictive elements.

Explication often resulted in the addition of new candidate concepts to the set, as they were required to adequately explicate previously-selected phrases, that is, to reach the analyst-estimated (and later to be validated) level of intersubjective agreement for each explication. When the need to consider a new concept arose, it was first determined whether it met certain criteria for candidacy, which was decided by exactly how its consideration was prompted. For example, concepts that were provoked through the explication of the originally selected phrases were entered as candidates, while those that were later provoked by explications of *those* phrases were not; this was entirely a scoping decision and alternate resource availability conditions would dictate alternate candidacy rationales. If the concept was not made a candidate, it was left as a leaf node in the explication that provoked it. However, if it was confirmed as a candidate, its official inclusion in the

1. In fact, representation of the *encyclopedic* semantics of the application domain are part of our agenda for future work; see Chapter 7 for more information.

Knowledge Base was either decided by the analyst on the basis of assembled information, or was decided later during validation according to information then collected from project personnel.

Integration. The concepts were then organized into a CLEAR Knowledge Base, in which each was given an entry as defined in the previous chapter. Further, all uses of an included phrase in any other explication were linked to the entry for that phrase, resulting in the underlying directed acyclic graph structure of a CLEAR Knowledge Base, the property that makes it amenable to certain kinds of desirable analysis.

Validation. Prior to analysis of the CLEAR Knowledge Base, the structure and content of the artifact and its elements were validated.

A number of content issues were handled as they arose in the explication phase during short consultations with project personnel. In these consultations, clarification of pieces of information from the collection forms was sought and tentative theories of meaning that the analyst had reached through the integration of multiple sources were confirmed or disconfirmed before couching those theories into explications. Structural issues for individual explications such as circularity or the presence of a non-predictive element or any other explicitly disallowed device were also addressed through refinement of the draft explications.

Following integration of these partially validated explications into the Knowledge Base, the entire artifact was validated with and without informant consultation as appropriate. First, decisions on the inclusion or exclusion of candidate Knowledge Base additions that arose during previous steps was accomplished with informant consult. The main criteria for inclusion were that the candidate be judged by a member of the project personnel to be both meaningful in its own right within the application domain, i.e., not either a lexical variation of another phrase or a nonsense construction, and, that it be intuitively relevant within the application domain, i.e., that the concept has a functional role in the system, and is not simply a notion the personnel are aware of.

Next, each explication was examined with informant consult, and both specific questions that had been previously recorded by the analyst, as well as the content of every explication regardless of the existence of specific questions, were addressed. At this time both the information included in and excluded from the explications as well as the levels to with the reductive paraphrase process was carried were assessed for appropriateness.

Once content changes resulting from these validation steps were applied, the structure of the complete Knowledge Base was validated both in the large, in that it was determined that every phrase of interest had a complete entry and that there were no duplicate or inconsistent entries, as well as in the small, in that each explication was rechecked after revision to confirm that it met the structural requirements of legal explications. The Knowledge Base was then frozen for analysis.

6.2.2.2 Results and Discussion

The previous steps ultimately resulted in a CLEAR Knowledge Base containing 125 unique entries representing 125 concepts judged to be critical to validity of the requirements and therefore in need of fidelitous communication. This section presents the measurements taken on the Knowledge Base and interpretation of the data thereby acquired.

The results of the study can be broken out into three themes. First, we address the data comparing our treatment of the phrases of interest to the way they are treated in the original sources. Next, we discuss the observations that become possible through analysis of the CLEAR Knowledge Base as a whole. Finally, we elaborate additional findings generated through use of the case study approach.

Theme 1: Comparison. Application of the CLEAR Method resulted in representations for concepts of interest that differed from the representations available in the original sources in a number of dimensions. In particular, we detected and corrected deficiencies present in the original sources, and produced representations that by design were free of specific faults.

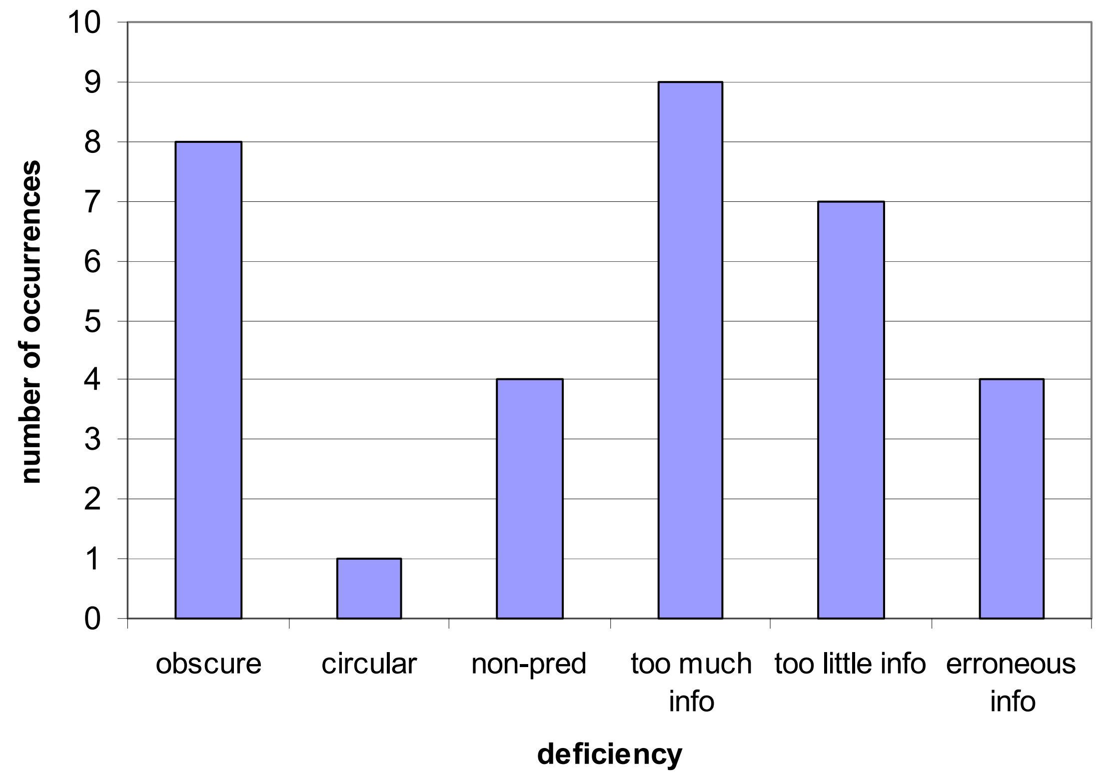
First, of 125 critical concepts addressed, only 11 had definitions in the original materials, while 114 did not[1]. This in part reflects the documentation practices of the organization, which correlate to its level of maturity, but these practices are not unrepresentative among all organizations developing software.

Of the 11 concepts for which definitions *were* provided, we detected and corrected a total of 33 faults of the types elaborated in Chapter 5. For example, *sensor data* was defined in the original materials as "consist[ing] primarily of the forces and moments the user exerts on the walker's handles during the natural assisted walking process, as well as the user's local environment." However, the use of "primarily" in the original definition renders it non-predictive. Further, following elicitation of the relevant semantics of this concept from informants, it was determined that there is both obscurity as well as the provision of too much information represented in the original. Specifically, a number of the notions used, such as "forces and moments", are themselves in need of explication (and as it turns out are superfluous to the ultimate explication), and almost all of the information provided is beyond that implicated by the category record. In fact, *sensor data* was ultimately explicated as simply ***data*** *collected through* ***sensors***, where boldface indicates that *data* and *sensor* are themselves further explicated in the CKB. The tree representing the explication dependencies characterizing *sensor data* is contained in the Appendix.

Note that these definitions, created by project personnel who are all experts in their respective areas of contribution to the project, had an average of three faults each. The distribution of fault types is displayed in Figure 4; it indicates, for example, that obscurity, as well as the provision of too much or too little information were the most common fault types detected. This finding may later be combined with analogous findings from other applications in order to establish a typology of the most common fault types within and

1. To determine whether a concept had a definition in the original materials, the context of all of its occurrences was examined and if in any case there was definitional material, as evidenced, for example, by a statement such as "An X is...", then credit was given for the presence of a definition. If more than one definition was present, we took the best possible combination of the definitions to be the definition of record for comparison with our explication. We believe these criteria were generous since we did not require any definitions be contained in a glossary or other collection in order to count, although this significantly aids locatability for users. Further, in taking the best possible combination for comparison when more than one definition was available, we discounted the real possibility that multiple definitions can be inconsistent, and simply attended to what definitional information was technically available in the original materials.

Figure 4: Deficiency Distribution



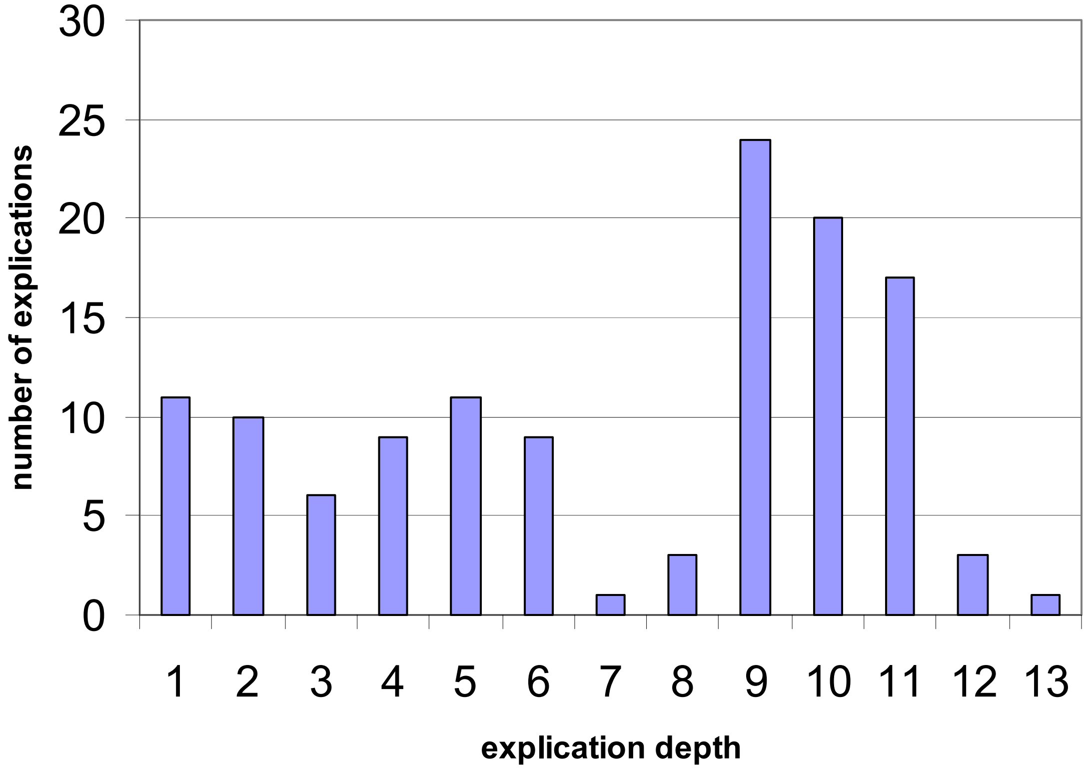
across domains. This information could be used to further inform the study and practice of engineering communication generally.

Of note is that among the 114 concepts *lacking* definitions in the original materials, 105 were each characterized through analysis of the collected data to provide an increased risk of miscommunication, either because they represented a domain-specific restriction of a common concept, e.g., *behavior*, they had multiple available senses, e.g., *angle*, or no common version existed to provide a meaning at all, e.g., *virtual moment*. Some phrases, such as *state* had two of these risk factors at once, in this case domain specific restriction of one of multiple available senses. Only 9 of the undefined concepts were analyzed to have no *apparent* risk factors.

Thus CLEAR compared favorably to the original materials in addressing all of the concepts indicated to be critical and providing explications for these concepts free of the structural and semantic deficiencies described in Chapter 5.

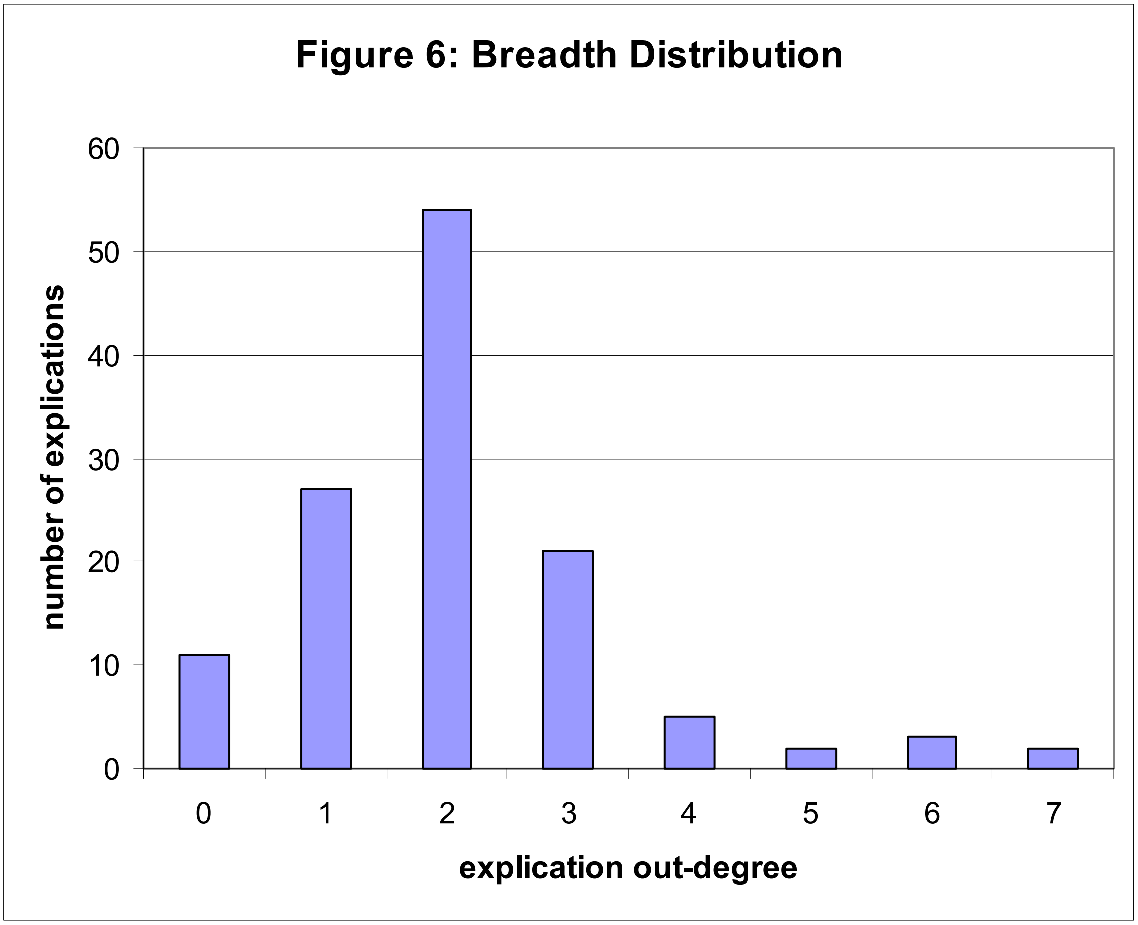
Theme 2: Knowledge Base Analysis. Organization of the explications into a single structure allowed analysis of the many conceptual relationships that were uncovered in the process. Specifically, the Knowledge Base is structurally a directed acyclic graph, and as such, analysis of the patterns of nodes and edges in its subgraphs gives rise to several implications for more effective use of this lexicon in its source project.

Figure 5: Depth Distribution



First, the depth and breadth of each explication can be determined. Recall from the previous chapter that the depth of an explication is the number of total levels in its explication tree, while the breadth is the number of other target phrases upon which the explication immediately depends. From Figures 5 and 6, we see that the deepest of the explications included 13 levels, and the broadest explications included 7 direct dependencies, while the most common values for these variables were 9 and 3, respectively. This information is useful in characterizing the skills needed to gain working control of a lexicon, insofar as depth is a measure of semantic distance, or remoteness, from intersubjective agreement or common understanding, while breadth is an indicator of the variety of notions that must be integrated together in the understanding of a concept. This indicates, for example, that a concept with 13 levels in its explication is quite removed from spontaneous intersubjective agreement and therefore more prone to miscommunication than is a concept with an explication tree depth of 4.

More interestingly, the number of unique nodes per explication, as well as the number of ways a concept might be encountered in the lexicon, can also be determined. Recall that the number of unique nodes per explication represents the number of unique explicated concepts that participate in the explication of a given concept. For example, the explication for *intent* as meant in the walker domain includes the other targets *navigation*, *control*, *behavior*, *device*, and *kinematics*, as well as itself, resulting in a node count of 6. Complementarily, the number of ways a concept might be encountered in the lexicon is



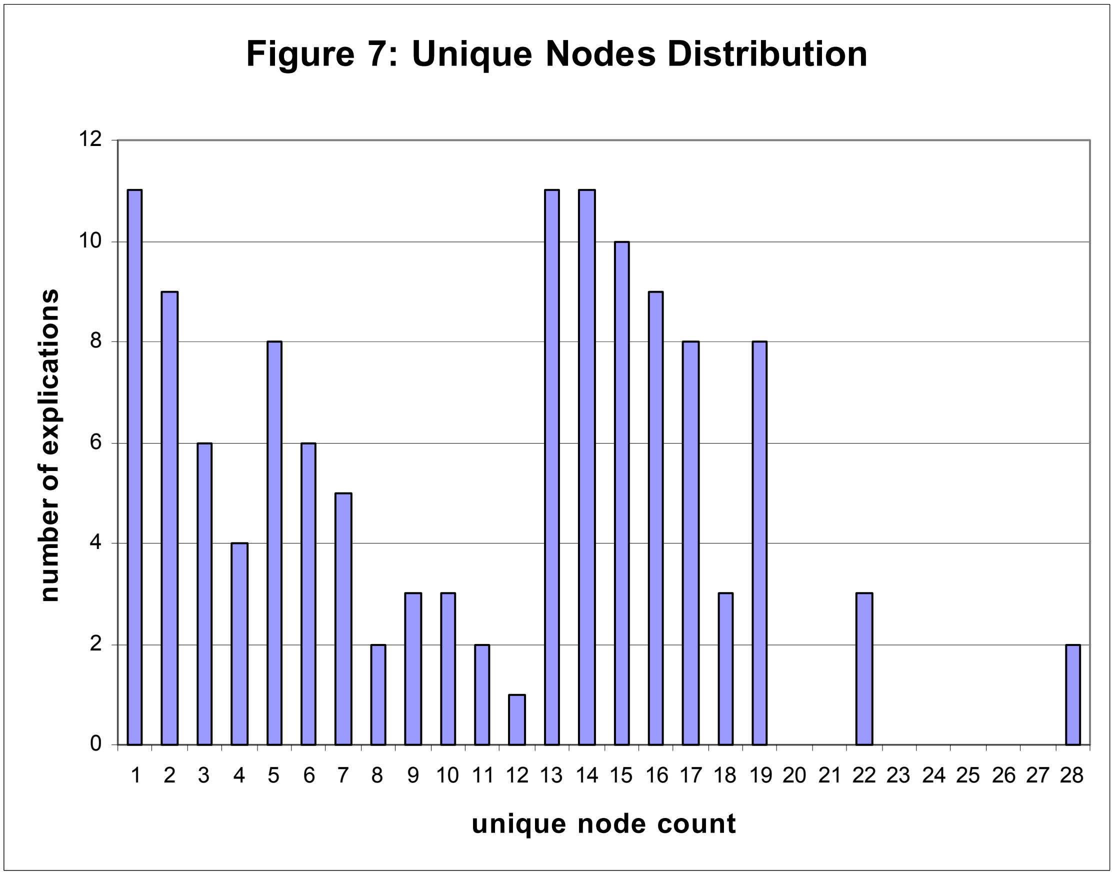
the number of explications in which an explicated concept directly participates, including its own. For example, *active* participates in the explications of 6 other concepts besides itself, resulting in an encounterability measure of 7. From Figures 7 and 8 we see further that 8 concepts included 5 unique explicated critical concepts each in their own explications, and that 18 explicated critical concepts participated in the explications of 2 concepts each.

The implications of these two measures are significant. Specifically, the number of unique nodes in an explication is an indicator of the number of possible ways to misunderstand the concept represented by that explication, since each node represents a critical piece of the whole.[1] Further, the number of explications in which a concept participates is an indicator of its pervasiveness and influence throughout the critical domain knowledge of the application. If a concept is misunderstood, it affects the validity of potentially much wider partitions of the requirements than just those in which it is directly referenced, since it compromises the understanding of as many other concepts as rely on it, as well as any that rely on those, and so on.

With these two measurements, we can now compute the risk of miscommunication of each concept in the Knowledge Base as defined by the formula in Chapter 5.

---

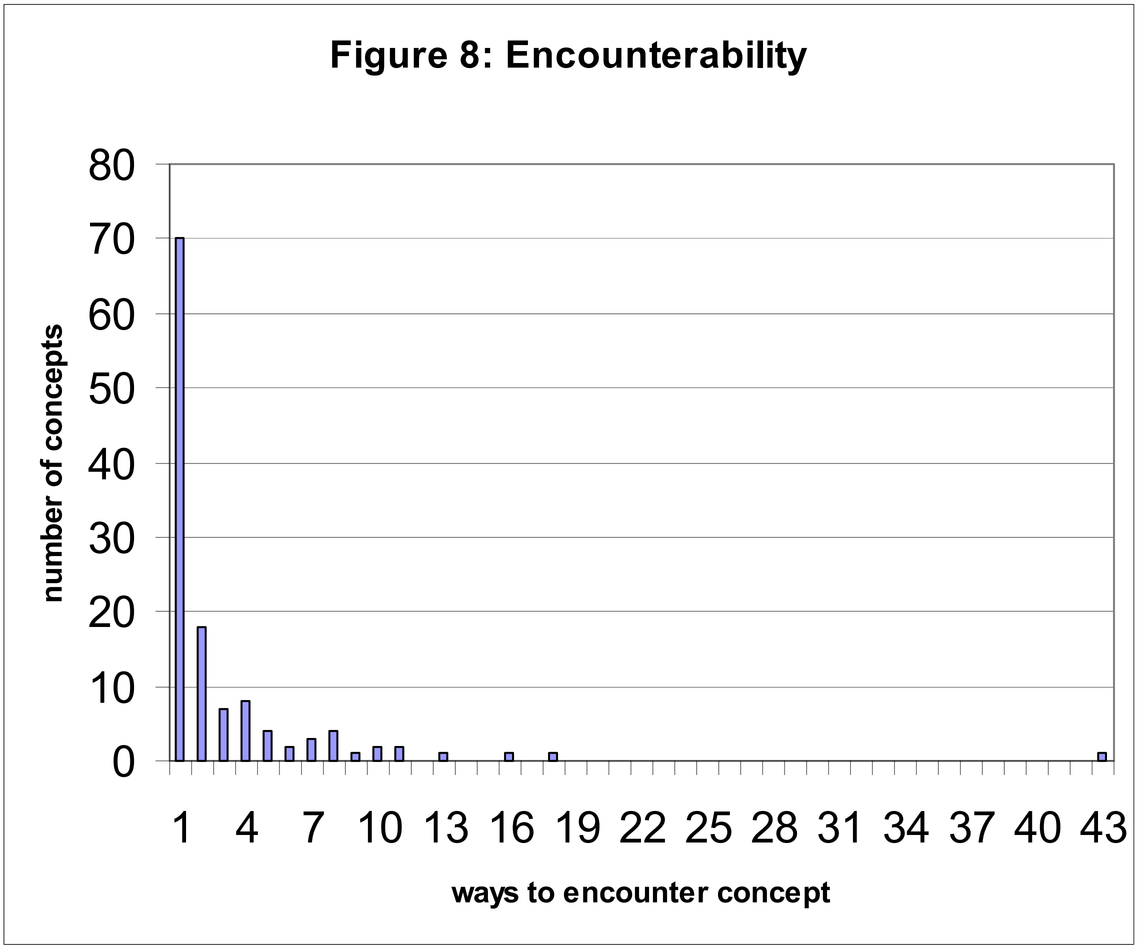
1. It is recognized that concepts on which a given concept depends can be misunderstood in combinations in addition to one at a time; this does not change the general argument.



Multiplying the number of unique nodes per explication by the number of explications in which a concept participates for each explication in the CLEAR Knowledge Base for the Walker Project yields the Risk Indices presented in Table 2, which has been sorted by Risk Index from highest to lowest. The Risk Indices allow us to see, for example, that while *radial depth map* has more nodes in its explication (17) than *virtual moment* (10), it poses less of a risk (RI: 34) than does *virtual moment* (RI: 50) because *virtual moment* pervades more of the total domain knowledge of the application, participating in 5 explications, to *radial depth map*'s 2. This allows resources to be directed preferentially, for example, at validation and testing of system functions relying on concepts with relatively high Risk Indices. It further provides the basis for prioritization of corrective action in requirements and other system documentation.

Finally, as with the findings on Distribution of Fault Types from Results Theme 1, all of the Knowledge Base Analysis findings can be compared to those of other applications to generate meta-analyses of domain commonalities and differences, that could shed light on, for example, the typology of miscommunication events generally.

Results Theme 3: Additional Findings. In addition to the quantitative data already presented, our use of the case study approach allowed a number of observations outside of the frame of a traditional controlled experiment. Since the unit of analysis of the study was all



of the entities and activities involved in the execution of a CLEAR process, a wide range of experiences was available for consideration. We note below additional findings that complement the previously presented results.

Semantic non-compounding. Patterns in explications demonstrate the need for even apparently trivial ones to exist. Consider the phrase *walker heading*, a noun-noun construction. It is explicated by “The *heading* of the *walker*”, where italics represent the existence of further explications for italicized phrases. At first, it might seem that this explication provides no information not already provided by the original phrase, but in fact, it serves at least two purposes. First, it provides the mechanism by which the depended-on explications of *heading* and *walker* can be examined. More importantly, however, it defines exactly what the relationship between *walker* and *heading* is. Specifically, it is modificatory, and indicates that the heading in question is the one possessed by the entity represented by *walker*. In contrast, *drive system*, an apparently similar noun-noun construction, is explicated as “The *system* whose components enable turning of the front wheel of the *walker*”. The relationship in this case is not one of possession, but rather one of facilitation. The difference in these relationships, represented by apparently similar syntactic constructions, is brought out by the process of explication. In other words, explication accounts for and records evidence of the fact that when semantic entities combine, the result is not always a straightforward compound of the respective concepts; to assume that two phrases represent analogous relationships because they are syntactically similar is to risk misunderstanding.

Table 2: Risk Indeces for the Entries in the Walker Project CKB

Risk Index	Entries with this Risk Index
516	walker
169	environment
117	data
104	sensor
91	map
90	control system
75	obstacle
68	user intent
60	environment state
56	perception
50	virtual moment
48	active
44	control system state
40	navigation
38	historical data, predicted path
36	intended path, shared control
35	active behavior, steering
34	radial depth map
33	control
32	possible path, walker state
30	environment map, force, infrared sensor
28	behavior based control, historical state, likely path, walker heading, wheel heading
26	drive system
24	robot
22	propulsive virtual moment, repulsive virtual moment
20	mode, moment

Table 2: Risk Indeces for the Entries in the Walker Project CKB

Risk Index	Entries with this Risk Index
19	clearance maintenance, handle sensor, instantaneous map, obstacle avoidance, opening passage, user intent perception
18	control rule, force sensing, intent, steering angle
17	force data, handle force, handle moment, head angle sensor, propulsive force, walker intent
16	behavior, dangerous situation, device, difficult situation, docking, environment perception, ground reaction force, sensor data, shared
15	behavior based control system, environment data, local environment, map data, perception system, walker steering
14	environmental context, global environment, grid style map, local map, nonholonomic constraint, oscillation, walker safety
13	active walker, brake system, interface, passive walker, walker frame, walker system
12	passive robot
11	bounded propulsive virtual moment, bounded repulsive virtual moment, system
10	mobile robot, passive, robotics, shared control device, shared control system
9	active mode, active steering
8	head angle, heading, passive robotics
7	active control, heuristic logic, passive mode, passive steering, path
6	acceleration, goal directed navigation, robotic system, state
5	autonomous robot, passive control
4	gravitational force, normal force, safety, stability
3	angle, instability, kinematics, pan, passive device, tilt, velocity
2	bearing, dynamics, safety hazard, user safety
1	roll, slip/slide/skid

Inconsistencies discovered and reported. The process uncovered a number of instances of inconsistent lexical use, either of multiple unintended phrasal variations used to represent the same concept, or of individual phrases having inconsistent meaning across informants, original materials, or a combination. For example, *environment* sometimes appeared to include the user and sometimes not, evidencing the kind of situation that can arise when assumptions about meanings are not examined, and which, left unchecked, can lead to unintended system behaviors should such inconsistencies persist in the deployed system. In this case, the "official" meaning of *environment* was decided in consultation with personnel during validation for purposes of producing an explication, but this is only the first step in propagating this codification through the project. Thus in addition to a copy of the complete CLEAR Knowledge Base, the Walker Project is consulted regarding specific issues that were located, so that the project leads can disseminate policy decisions regarding their usage in development materials and publications.

Further types of inconsistency noted were represented by *navigational intent*, which was determined to be synonymous with *user intent*, and the semantics covered by *heading*, which needed to be split into two concepts, *heading* and *bearing*, to adequately differentiate the intended meanings. Interestingly, this is not the first time we have come across communication issues involving *heading* and *bearing*; they were singled out as well in our analysis of the language of a maritime track control standard, for miscommunication risks involving their use [26]. This reinforces the possibility that meta-analysis of CLEAR products from different applications can contribute to a potential typology of miscommunication.

General noise in the collected information. Finally, the process also told us some useful things about human nature and the efficacy of our instruments. In particular, experience of how informants interacted with the collection form gives us ideas for how to improve it. For example, they sometimes ignored the provided structure and seemed to prefer to write everything they knew in one place. In addition, it was noted that they might not offer some particular piece of information unless provoked, that they might come up with contradictory answers when asked the same questions in different ways, and in one case an informant even responded differently to the same questions regarding the same concept when he accidentally treated it twice on different days. In addition, idiosyncrasies of individual informants played a role: different informants provided answers at different levels of abstraction and comprehensiveness, and different levels of English fluency required some amount of interpretation on the part of the analyst.

In revision of the collection instrument we plan to do two things. First, we will adjust the questions in order to reduce some of the noise described above. Second, a light version of CLEAR is planned in which the initial explications will be constructed completely on the basis of otherwise available resources, without an elicitation round. This change will push more effort into validation of the artifact, since presumably the draft explications will require more adjustment, but will require fewer meetings and less informant time overall, as well as reclaim the effort spent in responding to noise. These savings are predicted to lower the bar for an organization to invest in undertaking a CLEAR process.

6.2.2.3 Threats to Validity

This section describes the known threats to validity of the inferences drawn by the case study and how they were mitigated or accommodated.

Construct Validity. Construct validity concerns the degree to which the chosen measurements support the inferences made by the study [70]. The measures used in this study were the coding scheme for classifying deficiencies, and the graph metrics used to characterize the Knowledge Base.

The structural classes of the deficiency coding scheme are based on a typology of errors of definition that is founded in linguistic theory [19]. We extended this typology with semantic classes that accounted for additional observable phenomena. Characterization of deficiencies via this coding scheme allowed us to directly compare differences between the original materials and our artifact with respect to these phenomena. In this regard, the coding scheme enabled the inferences presented in Results Theme 1.

The graph metrics used to characterize the Knowledge Base are straightforward counts of elements of a data structure. The phenomena we suggest they imply derive from linguistic theory, in the case of properties of explications with regard to accessibility of meaning, as well as intuitive analogy to existing constructs in the case of the Risk Index. Risk is commonly abstracted as the probability of an event multiplied by its severity, and parameterized for given purposes in different environments; we have done the same. In these regards, the graph metrics enabled the inferences presented in Results Theme 2.

Internal Validity. Internal validity concerns the degree to which a causal relationship can be established between two events, in this case, the intervention and the result [70]. If the result might have arisen because of some third, unconsidered factor, for example, internal validity would have been compromised.

In the natural and social sciences, many naturally occurring phenomena are eventually explained by factors or interactions of factors that were not at first considered. In the case of this study, however, the CLEAR Method was *engineered* to produce an artifact with certain properties that do not naturally arise. Further, the appearance of the artifact and its properties satisfies the criteria for establishing a causal relationship [11, 59]. Specifically, the intervention temporally preceded the effect attributed to it, the cause and effect covary, insofar as these properties are only observed when these activities are undertaken, and there exists no plausible alternative explanation for the effect, since to favor a rival explanation for the existence of this artifact and its properties following execution of the process would violate Occam's Razor. We thus conclude that the intervention of the CLEAR Method resulted in the existence of the artifact and its properties.

External Validity. External validity concerns the generalizability of the findings of the study to a wider context than that in which the study took place [70]. In this case we are interested in whether our results from collaboration with the Walker Project generalize to a wider set of software development projects and organizations.

Since this is a case study and not a formal experiment, we rely on analytical rather than statistical generalization. This means that because statistical sampling is not possible, we must make an analytical argument that the properties of the study environment that are relevant to the intervention are representative of environments targeted by the CLEAR Method, and that other properties are orthogonal to its use. The results of this study can then be argued to generalize to other environments sharing these properties, as no others can be supportably linked to action of the CLEAR Method.

The organization with which we collaborated had several characteristics of interest to us as described in 6.2.2.1. Most importantly, the Walker Project is comprised of the research and development activity surrounding a safety-critical, software controlled medical device, with personnel from several domains of expertise needing to communicate with each other with pragmatic sufficiency in order to reach the goal. However, we recognize that the Walker Project documentation practices represent the less mature (though no less realistic) end of the software engineering spectrum. This likely had an effect on the *percentage* of critical concepts for which definitions were available in the original materials, as compared to organizations with practices that encourage inclusion of a glossary or similar in a more rigorous set of requirements documentation. Further, since the number of personnel on the Walker Project is relatively small (5), several members each perform in multiple roles, and thus are necessarily more and less expert at some of their tasks as compared to others. However, we do not believe these facts to have had an influence on the *quality* of the definitions originally provided. In our experience, expertise in the domain of the concept being defined is less a factor than more universal cognitive habits and limits, in the construction of good definitions. Specifically, we have found similar types and quantities of faults across other of applications [26, 24], constituting replication. Further, the results that go beyond comparison between our artifact and the original material address issues not related to organization size or process maturity, so we do not believe these factors severely compromise the external validity of this study.

Reliability. Reliability concerns the extent to which a study can be repeated with the same results [70]. In the case of this study, the key threat to reliability is the subjectivity of the analyst. This subjectivity affects, for example, the exact wording chosen for explications, as well as judgments of the types of faults represented in a pre-existing definition, or regarding whether a candidate phrase should be included in the Knowledge Base.

To mitigate this threat, all activities with non-trivial judgment components were governed by clear, mechanical operationalizations, such that the opportunity for interpretation was minimized within the limits of the informal nature of the environment. For example, explication as an activity is defined by constraints regarding what semantics are included and excluded from an explication and what syntactic rules explications must obey. From this we gain confidence that the relevant components of explications, i.e., their syntactic structure and the set of concepts upon which they rely, are likely to be largely comparable across analysts.[1] Similarly, the deficiency coding scheme defines 6 separate classes of syntactic and semantic faults with criteria that constitute demonstration, and candidacy decisions were made according to another set of rational, observable criteria. Finally, a large portion of the results do not rely on analyst judgement but are rather directly quantitatively measurable. We thus do not believe that the reliability of this study has been severely threatened by analyst subjectivity.

1. Note that empirical demonstration of this comparability is an avenue for future work.

6.2.2.4 Case Study Summary

The research question providing the scope for this investigation concerned how the execution of a CLEAR process affects the quality and organization of representations of critical concepts in an application domain. Within this scope, we made two propositions, first, that execution of CLEAR would improve the integrity of both the form and the content of representations of critical domain concepts, and second, that analysis of the products of CLEAR would generate insights not previously available regarding the semantics of the application domain. These propositions were investigated via a case study involving application of CLEAR to the domain semantics associated with a software-controlled, safety-critical medical device. It was shown that the products of CLEAR both compared favorably to the existing representations of semantics, and further, that analysis of the resulting CLEAR Knowledge Base allowed pattern visualization enabling, for example, quantitative assessments of the risk of miscommunication of concepts addressed. We thus conclude that the use of CLEAR improves the fidelity of communication while respecting constraints of realistic environments.

6.2.3 Conclusions of the Macro-Study

Consideration of the two investigations together leads to the following further conclusions: complementation provided us a more comprehensive picture of the benefits of CLEAR than did either study alone, selected results of the case study constituted empirical replication of the results of the experiment, and the case study results in addition extended beyond the results of the experiment.

Complementation. Given the results of case study, we can make a more diverse set of claims regarding the properties of CLEAR demonstrated in the experiment. The experiment told us that comprehensibility could be improved within the controlled environment in which we tested CLEAR components, but the context allowed by the case study format enabled broader observation opportunities, which in turn enabled not only attention to a wider set of propositions and study questions, but also led to additional findings not available in the experimental format. Thus the complementation provided both narrow depth as well as observational richness.

Replication. The results of the case study further provided empirical replication of the experiment results through triangulation. The contributions from each study regarding the linguistic properties of CLEAR products and their value are consistent with each other across environments, applications, and empirical research designs. This congruence constitutes replication in multi-method approaches and allows the results of each study to reinforce each other. The combination thus makes a stronger case for the claim that CLEAR improves the fidelity of communicative transactions.

Extension. Finally, the case study results extend the experiment results in that they make new contributions not delivered by the experiment. In particular, the case study demonstrated productive application of CLEAR on a much greater scale, in terms of scope of domain semantics addressed, at a greater level of method maturity, in terms of theoretical and procedural additions and refinements, and at a greater level of comprehensiveness, in terms of execution of the entire method, rather than of selected components.

We thus conclude that CLEAR advances the state of the art with respect to the problem of improving communication in requirements engineering.

6.3 Chapter Summary

This chapter detailed the evaluation of the CLEAR Method as introduced earlier in this work. In particular, we first presented a strategy for combining two empirical research methods in order to address multiple dimensions of the value of CLEAR. We then reported on the execution and results of a formal experiment, in which the experimental hypothesis was statistically supported. This was followed by a report on the execution and results of a case study, in which the study propositions were also upheld. These findings were then synthesized into the conclusions of the macro-study, which characterize the effects of CLEAR in application and provide a balanced picture of what it can be demonstrated to do in both theory and practice.

To make an end is to make a beginning. The end is where we start from.
--TS Eliot

7 Conclusion

This work examined the problem of achieving high-integrity communication in requirements engineering. The choice of topic was driven by two main factors: the importance to the community that there be progress in this area, and the promise embodied by an approach that had not before existed and could rationally be predicted to provide a benefit.

The approach involved looking to pertinent results in fields that study communication, based on the recognition that there is significant regularity in language that might be amenable to exploitation for our purposes. The starting point was to study what is known about the low-level mechanics of communicative transactions, in order that we might be able to better characterize the mechanisms of miscommunication in requirements environments and treat them more directly.

These characterization efforts resulted in the first contribution of this work: a more detailed analysis than was previously available of the issues and entities in play during communicative transactions that occur in requirements environments.

This analysis further provided the foundation for the main contribution of this work, the methodology. The CLEAR Method, Cognitive Linguistic Elicitation and Representation consists of a set of artifacts and processes that guide the scaffolding of communicative transactions that are demonstrably and reliably free of important classes of deficiency.

The methodology was evaluated in both controlled and realistic environments, and it was argued that the thesis of the work was upheld, that is, communication between requirements producers and requirements consumers can be systematically improved to a level where desirable properties necessary to communicative fidelity are reliably present.

The remainder of this chapter reviews the central argument of the work, elaborates its contributions and limitations, and describes avenues for future work deriving from the foundation it provides.

7.1 Summary of the Work

The argument began with the identification of an area of activity within the requirements engineering community in which a number of related issues are currently recognized as in need of address. We then focused on a core objective within that scope and undertook to make it actionable. Following analysis of lower-level components of the problem, we presented a solution in response to it, and evaluated this solution. Each of these elements is reviewed below.

7.1.1 Scope

The context of the work is software requirements engineering; the focal challenge within this context is achieving requirements validity.

The informal and especially social components that influence a requirements engineering process have been recognized as a central challenge to the reliable generation of valid requirements. These components are characterized by the interplay of a large number of unpredictable and uncontrollable variables that derive from the heuristic and stochastic nature of processes of human cognition, operating within individuals who are distinguished by an arbitrarily wide diversity of experience to begin with.

The difficulties and impossibilities in controlling or measuring these variables imply that validity is necessarily an informal property, and its presence is therefore difficult to demonstrate with assurance and rigor. However, some assurance that a measure of validity has been achieved is critical to the deployment and safe use of high-consequence systems.

This work was concerned at a high level with better understanding the influence of social factors at play in achieving requirements validity, for the purpose of better management of these factors.

7.1.2 Objectives

It was argued that primary among these social factors were the ways in which communication is used by agents of a requirements process. Since validity is a mapping of intentions to instantiations, validity depends on the correct and complete understanding of those intentions by those who will instantiate them. Barring the case where the intender is also the instantiator (a non-existent situation in complex systems development), this understanding can only come from communication between these parties.

The consequences of poor communication are well-known and motivate the addressing of this problem. Faults and failures that can be traced to invalid requirements threaten safety, security, property and the integrity of critical infrastructures. Further, they are a source of significant unnecessary cost in form of rework and failed projects that never enter deployment.

Responses to the communication issues inherent in requirements engineering address their manifestations in requirements materials, such as ambiguity, incompleteness and inconsistency in the information presented. Existing methods are varied in approach, but uniformly address these notions holistically, that is, without breaking them down. In addition, the derivation of existing methods is generally intuitive rather than theoretically-grounded. Our first insight was that better characterization of communication and how it breaks down in requirements environments was needed.

This led to the preliminary objective of this work: to provide a more comprehensive and theoretically-grounded analysis of the communication deficiencies that compromise requirements activities.

This preliminary objective was designed to provide support for the main objective. Using knowledge gained in the above analysis, we sought to build a method that addressed lower-level mechanics of communication more directly. The main objective of this work was to demonstrate that the fidelity of communication between requirements producers and requirements consumers could be reliably improved. In support of this main objective, we formulated a more actionable statement of the problem that focused on supporting both the integrity of the signified and the integrity of the signifier in a given communicative transaction such that more pragmatically sufficient transactions could be realized.

7.1.3 Review of the Solution

Given the parameters of the refined objective, we first sought to examine and characterize the entities and processes at play during communicative transactions that impede the achievement of pragmatically sufficient communicative transactions, particularly in requirements environments. This examination drew on results from disciplines that study the structure and use of human language, and allowed the identification of elements contributing to insufficient transactions that could be treated more directly than by methods that respond instead at the emergent level of underdifferentiated miscommunication.

The solution offered specifically addresses the stated objectives of directly supporting both the integrity of the signified and the integrity of the signifier. It further addresses the issue that an ordering of targets for processing provides valuable guidance in allocation of resources in execution of later CLEAR activities, and also allows the collective analysis of the resulting representations together, in order to permit visualization of insights not otherwise available.

The CLEAR Method, Cognitive Linguistic Elicitation and Representation consists of a set of artifacts and processes that guide the scaffolding of communicative transactions that are demonstrably and reliably free of important classes of deficiency that threaten the integrity of the signified and the integrity of the signifier. The method is divided into four-phases.

7.1.3.1 Phase 1 of CLEAR

Phase 1 of the CLEAR Method is Selection. It accomplishes the goals of scoping of a set of requirements materials and prioritization of material for further processing, and thus supports better allocation of development resources. The input to the phase is the set of existing requirements for the project, in whatever form that existence takes. Selection activities first enable assembly of materials that can be automatically processed. Automatic processing based in principles of Information Retrieval is combined with manual refinement to produce a prioritized list of target phrases that are passed on to the next phase for further processing.

7.1.3.2 Phase 2 of CLEAR

Phase 2 of the CLEAR Method is Elicitation. It accomplishes the goal of supporting integrity of the signified in a communicative transaction. Elicitation activities are based in principles of Cognitive Linguistics that describe how semantics is stored and manipulated in the minds of users of a language. The primary mechanism for semantic organization is a mental abstraction called the cognitive category, which has an empirically validated structure characterized by specific properties. Knowledge of this structure allows the definition of an intermediate archival artifact called the category record, a storage mechanism that organizes semantic information elicited through a complementary collection form. The combination allows elicitation of the semantics of an application concept from a client or expert, according to the ways such individuals store those semantics naturally. This enables a more accurate and complete picture of what the producer intends by a given concept under consideration than that gained without such guidance. The product of this phase is a set of category records that capture the semantic topologies of prioritized concepts with a better fit to their actual conception than otherwise achievable.

7.1.3.3 Phase 3 of CLEAR

Phase 3 of the CLEAR Method is Representation. It accomplishes the goal of supporting integrity of the signifier in a communicative transaction. Representation activities are based in principles of Cognitive Linguistics that describe how complex concepts are built up out of simpler ones in the minds of users of a language. We adapted an existing method for constructing specialized definitions called explications that avoid particular faults common to traditional definitions. Our adaptation, Partial Reductive Paraphrase (PRP), renders complex concepts accessible to those without prior experience of them or with insufficient or otherwise flawed experience of them, while taking into account needs and constraints of requirements environments. The content of PRP explications is guided by the material assembled in the category records of Phase 2, in support of integrity of the signified, and the structure of PRP explications explicitly disallows faults that limit the accessibility of traditional definitions, in support of integrity of the signifier.

7.1.3.4 Phase 4 of CLEAR

Phase 4 of the CLEAR Method is Integration. It accomplishes the goal of supporting collective organization and analysis of the explications generated in Phase 3. The explications form the basis for entries in a CLEAR Knowledge Base (CKB), the product of Representation and primary artifactual result of a CLEAR Method application. The CKB organizes entries such that certain relationships among them are made explicit, and provides the basis for various analyses that allow further insight into the nature and complexity of the concepts associated with a given application. Importantly, it allows for the first time a quantitative computation of the risk of miscommunication of a concept, based on the mechanisms and opportunities for miscommunication evidenced by the graph properties of that concept in the CKB. This information can further direct the allocation of resources within a project, and when combined with analogous information from other projects, can allow generalizations that inform the practice of requirements engineering more generally.

7.1.4 Argument Summary

Via execution of the four phases of CLEAR and generation of the CLEAR Knowledge Base, the two main communicative elements of the problem are addressed, that is, integrity of the signified and integrity of the signifier for a given target concept. In addition, the order in which targets are addressed is prioritized such that resources may be allocated accordingly. Finally, the CKB artifact provides a persistent resource that informs both the content and the process of engineering requirements for a given system.

The CLEAR Method was evaluated via a two-part strategy designed to achieve both experimental control and environmental realism. In the first part, we conducted a controlled formal experiment to assess the differential levels of comprehension associated with artifacts embodying core CLEAR concepts, and those produced by current operating procedures of a given organization, respectively. The resulting data allowed us reject the null hypothesis, that the apparent desirable effects of CLEAR were achieved by random chance, at the .01 level of significance. In the second part, we conducted a case study designed to demonstrate that refinements and enhancements made as a result of the formal experiment were feasible and had desired results when applied in an actual development environment. This experience allowed us to duplicate and extend the results of the initial experiment, and the experiences together provided a broader picture of the effects of CLEAR.

As a result we argued that the thesis of the work, that communication between requirements producers and requirements consumers can be reliably and demonstrably improved, was upheld.

7.2 Contributions and Limitations

This section addresses the impact of this work, as well as limitations that qualify that impact. It characterizes the change this work makes to the landscape of requirements engineering research and lays the foundation for additional work that will extend its contributions.

7.2.1 Contributions

This work makes three main contributions to the study and practice of requirements engineering. First, it provides a linguistically-grounded analysis of communication deficiencies and their sources. Second, it provides a mechanism, the CLEAR Method, for systematic detection and correction of several classes of such deficiency. Finally, analysis of the products resulting from CLEAR provides for the first known mechanism for estimating the miscommunication risk of a project's critical notions, thereby offering guidance in the allocation of resources for the preservation of their integrity throughout a project's duration.

7.2.1.1 Theoretically-Grounded Analysis of Deficiencies and Their Sources

The first contribution of this work is a more comprehensive and theoretically-grounded analysis of communicative deficiencies and their sources. This analysis exploited what is known by linguists about the low-level mechanics of communicative transactions, and extended this knowledge to better characterize the mechanisms of miscommunication in requirements environments. These characterization efforts resulted in a clearer picture than was previously available of the issues and entities in play during communicative transactions that occur during requirements activities. The main finding was that cognitive heuristics are tuned to a common case, and our minds naturally operate in one basic way. However, the efficacy of these heuristics is only possible because of properties of common case communication environments, such as the background context or experience that communicating parties share. A fundamental insight that leads to the first contribution of this work is that the properties of communication environments that allow innate heuristics to work are not present in requirements environments. First, the requisite shared experience is absent, and second, the needs for accuracy and precision are far higher. But communicating parties still, and have no choice but to, continue to use the cognitive heuristics with which they are naturally equipped. Communicative breakdown originates here; the assumptions allowed by innate heuristics are not valid in requirements environments.

This analytic result provides a previously absent foundation for the construction of methods that accommodate the habits and limits of innate cognitive processes in operation during requirements activities.

7.2.1.2 Detection and Correction of Deficiencies

The second contribution of this work is the CLEAR Method, Cognitive Linguistic Elicitation and Representation. CLEAR uses the knowledge gained from the linguistic analysis of deficiencies and their sources as the basis for artifacts and processes that more directly address the low-level mechanics of miscommunication.

CLEAR provides guidance in both the directed elicitation of the semantics of application concepts to better match the intentions of experts and clients. It further provides guidance in the construction of representations of those concepts that are demonstrably free of specific definitional faults. These representations are related in a CLEAR Knowledge Base, and artifact that gathers and organized the processed information for further analysis and use as a development resource.

CLEAR's use of a theoretical foundation to analyze and reduce the incidence of communicative deficiencies was shown to constitute an advancement beyond the current state of the art in achieving high-integrity communication for requirements engineering. Specifically, CLEAR enables desired properties not otherwise achieved or maintained to be upheld with respect to a body of knowledge that is shared among project stakeholders.

7.2.1.3 Assessment of Miscommunication Risk

The third contribution of this work is the definition of a set of measurements on CKBs that allow for the first time a founded quantitative assessment of the risk of miscommunication of a concept in question. The Risk Index is the product of two other measurements that respectively indicate properties analogous the probability and severity of miscommunication of a concept; multiplied together, they result in a natural number whose relative magnitude indicates its level of miscommunication risk, as compared to that of other concepts in the CKB.

The benefit of this information is twofold; first, it provides guidance in allocating limited resources in the correction and maintenance of requirements for complex high-assurance systems. Second, when combined with similar data from other projects and domains, it will allow meta-analysis that is expected to result in the identification of cross-domain patterns in the ways engineers use language, further informing our needs to improve such activities.

7.2.2 Limitations

The value of this work and its products is qualified in at least two ways.

First, the problem it addresses can never be absolutely solved. It is thus impossible to quantify the improvement delivered by CLEAR relative to an optimal solution. That is, the essential indeterminacy of meaning prevents us from knowing with absolute certainty what a requirements producer intends by a given statement, or what a requirements consumer interprets in response to it. Thus, given a CLEAR representation, we cannot quantify its validity as a percentage of perfect fit. This limitation drove the search for alternative means of demonstrating the value of CLEAR. As detailed previously, we instead demonstrate both quantitative improvements relative to an instance of a currently active requirements process, as well as both quantitative and qualitative improvements evidenced by the presence of desirable properties in our representations that are otherwise absent.

Another limitation of this work is that the objects of analysis and processing handled by CLEAR are restricted to terms and phrases representing individual concepts within a system under consideration. More complex interactions among concepts, such as further characterize the semantics of these systems, are not addressed. There are many hierarchical levels of conceptual understanding that go into cognition of a complex entity like a software system. Individual concepts are just the most basic building blocks of these complexes, but in isolation they do little. Rather it is in their interactions that the dynamics of the world unfold. In this work, we have only looked at the level of the individual concepts, the level of the basic building blocks. This provides a critical foundation for being able to understand the more complex interactions of these building blocks, but such an examination was beyond the scope of this work. In the next section, we describe other potentially interesting hierarchical levels of conception and our plans for addressing them, as well as other lines of inquiry for which this work sets the foundation.

7.3 Future Work

This section addresses further research sub-agendas that derive from this work and its results.

7.3.1 Cost Measures

This work focused on the invention, feasibility of use, and demonstrable benefit afforded by a methodology. Though it provided some guidance in allocation of resources, it was not a main objective of this work to undertake assessment of the costs incurred in using this methodology. Clearly, however, rigorous cost assessment is important to achieving the eventual goal of technology transfer.

To this end, one avenue for future work involves the design and execution of empirical studies that observe and measure the resources required to accomplish applications of CLEAR in industrial environments. Of particular interest would be data comparing applications of CLEAR that have been parameterized in various ways. For example, applications of CLEAR that were controlled for the size of the phrase list might be expected to require comparable levels of investment to complete, given an equivalent number of phrases to process. Alternatively, they might be expected to vary widely as a function of a number of variables such as diversity of experts required for consultation or the complexity of the domain lexicon. Similarly, either of these variables could be controlled for and the variation in others measured. A rigorous characterization of the costs incurred by elements of the method individually and together, perhaps in a wider context of return-on-investment studies, would allow organizations to evaluate the potential value of using CLEAR for their projects. Additionally, it would provide the foundation for strategic construction of "light" versions of CLEAR that allowed organizations to balance power with cost-effectiveness. The possibility of light versions is further discussed below.

7.3.2 Tool Support

Also important to achieving the goal of technology transfer is comprehensive industrial tool support. Tool support for CLEAR will automate much of the word processing, file management, and CKB construction, management and measurement tasks. Further, it will allow search and query functions on the knowledge base, as well as graphic visualization of many of the relationships instantiated within it. Finally, we also envision provisions for electronically linking requirements documents into the CKB, including the development of automatic integration modules for requirements development support tools like RationalRose and Doors.

7.3.3 Methodological Enhancements

Two possible methodological enhancements to the existing CLEAR facilities are the construction of light versions and other variations, and the treatment of conceptual compounds.

7.3.3.1 Light Versions and Other Variations

In some circumstances, it might be desirable to focus on one or another of the CLEAR Phases and devote fewer resources to the others. To this end, we plan a research sub-agenda in which Heavy CLEAR and Light CLEAR are differentiated, and may be chosen and applied independently at Phase boundaries. The determination of which features will characterize heavy vs. light versions will be influenced by a number of factors such as the results of return-on-investment studies as well as the availability of other methodological enhancements not yet developed that might be more reasonably associated with one or the other version.

Another variation on CLEAR that is already in development substitutes for Phase 1 information gained from defect reports generated during testing to construct the list of phrases targeted for processing. This version has the working title of RetroCLEAR, since it takes information provided from downstream stages of a development process and uses it retroactively to correct and maintain upstream requirements materials. This version of the method was evaluated via application to the requirements and defect report materials for the software for a deep space telescope, and the patterns that emerged from the analysis allowed insights not previously available [64].

7.3.3.2 Treatment of Conceptual Compounds

As noted above, the objects of analysis and processing handled to date by CLEAR are restricted to terms and phrases representing individual concepts within a system under consideration. Complex conceptual compounds that reside at higher hierarchical levels of conceptual understanding are not addressed. Another area of future work focuses on examination of the linguistic theory that operates over such compounds and the integration of this theory into methodological enhancements.

In particular, cognitive linguists have empirically demonstrated the existence of mental abstractions they call frames, scripts and schemas, that govern the ways we conceive of and talk about compound semantic entities and activities. These mechanisms have the potential to provide insight into the ways that requirements agents communicate about interacting collections of concepts that recur and are meaningful as collections (and thus as new building blocks in their own right) within systems and within domains.

7.3.4 Application Beyond Requirements

Finally, the needs for accuracy and precision and for the communication of information across domain boundaries are not unique to requirements engineering. Closely related is the process of standards development, in which the problem of demonstration of compliance with a standard is dependent on understanding what is communicated by that standard. Further, consistent application of a standard is dependent on consistent interpretation of it by its appliers.

Similarly, many other fields such as accident investigation, areas of medicine, and areas of law are have needs in common with requirements engineering, especially of high-consequence systems. In addition to the RetroCLEAR activities described earlier, we have also pursued the application of certain CLEAR themes to accident investigation, and more generally, forensic engineering [24]. All of these areas together with the work on standards are expected to be united in future under the umbrella of a wider agenda investigating the role and productive management of communication in engineering generally.

7.4 Conclusions

In this work we posed two objectives: to better characterize nature of communicative breakdown in requirements engineering, and to use the knowledge gained by this analysis to construct a method to more directly and systematically manage and avoid deficiencies in the communication of requirements.

By at least two different measures, we have done this. The analysis formed the basis for the method, and the method was successful under evaluation. We can say with strong statistical support that application of core CLEAR facilities improved comprehensibility of a body of knowledge by those without prior experience of that knowledge, and we can say that in application to an actual development project, CLEAR supported the construction of representations of knowledge pertaining to that project that were characterized by demonstrable desirable properties that discourage miscommunication.

Thus we assert that the thesis of the work, that communication between requirements producers and requirements consumers can be demonstrably and reliably improved, is upheld. This constitutes a qualitative advancement in the field of requirements engineering and provides the foundation for a wide and novel research agenda.

The wisdom of the wise and the experience of
the ages are perpetuated by quotations.
--Benjamin Disraeli

References

[1] Babbie, E. The Practice of Social Research, 8th ed. Wadsworth Publishing Company, Belmont, CA, 1998.

[2] Berry, D. Formal Methods: the Very Idea, Some Thoughts about Why They Work When They Work. Science of Computer Programming, 42(1):11-27, 2002.

[3] Berry, D. The Importance of Ignorance in Requirements Engineering. Journal of Systems and Software, 28:179-184, 1995.

[4] Boehm, B. Software Engineering. IEEE Transactions on Computers, C-25(12):1226-1241, 1976.

[5] Boehm, B. Software Engineering Economics. Prentice Hall, Englewood Cliffs, NJ, 1981.

[6] Boehm, B. Verifying and Validating Software Requirements and Design Specifications. IEEE Software, 1(1):75-88, 1984.

[7] Bogushevich, D. Discussion of pragmatic sufficiency in message with subject "[M]uch less." E-mail to CogLing Mailing List. Date e-mailed: 23 Feb. 2004.

[8] Brooks, F. No Silver Bullet: Essence and Accidents of Software Engineering. IEEE Computer, 20(4):10-19, 1987.

[9] Browne, G., and M. Rogich. An Empirical Investigation of User Requirements Elicitation: Comparing the Effectiveness of Prompting Techniques. Journal of Management Information Systems, 17:4, 2001.

[10] Burg, J. Linguistic Instruments in Requirements Engineering. IOS Press, Amsterdam, 1997.

[11] Cook, T. and D. Campbell. Quasi-Experimentation: Design and Analysis for Field Settings. Rand McNally, Chicago, 1979.

[12] Creswell, J. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 2nd ed. Sage Publications, Thousand Oaks, 2003.

[13] Curtis, B., H. Krasner, and N. Iscoe. A Field Study of the Software Design Process for Large Systems. Communications of the ACM, 31(11):1268-1287, 1988.

[14] Easterbrook, S. Elicitation of Requirements from Multiple Perspectives. PhD Thesis, Imperial College of Science, Technology and Medicine, University of London, Sep. 1991.

[15] Fuchs, N. Attempto Controlled English (ACE). Internet WWW page, at URL: <http://www.ifi.unizh.ch/attempto/> (version current as of March 17, 2006).

[16] Gause, D. and G. Weinberg. Exploring Requirements: Quality Before Design. Dorset House, New York, 1989.

[17] Gervasi, V. and B. Nuseibeh. Lightweight Validation of Natural Language Requirements. Software Practice and Experience, 32:113-133, 2002.

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

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

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

[21] Goguen, J. Formality and Informality in Requirements Engineering. Proceedings: 2nd International Conference on Requirements Engineering (ICRE’06), 1996.

[22] Grice, P. Studies in the Way of Words. Harvard University Press, Cambridge, 1989.

[23] Hanks, K. and J. Knight. Improving Communication of Critical Domain Knowledge in High-Consequence Software Development: an Empirical Study. Proceedings: 21st International System Safety Conference (ISSC'03), 2003.

[24] Hanks, K., J. Knight and C. Holloway. The Role of Natural Language in Accident Investigation and Reporting Guidelines. Proceedings: 2002 Workshop on the Investigation and Reporting of Incidents and Accidents (IRIA’02), 2002.

[25] Hanks, K., J. Knight and E. Strunk. A Linguistic Analysis of Requirements Errors and Its Application. University of Virginia Department of Computer Science Technical Report CS-2001-30, 2001.

[26] Hanks, K., J. Knight and E. Strunk. Erroneous Requirements: A Linguistic Basis for Their Occurrence and an Approach to Their Reduction. Proceedings: 26th NASA IEEE Software Engineering Workshop, 2001.

[27] Hayhurst, K. and C. Holloway. Challenges in Software Aspects of Aerospace Systems. Proceedings: 26th NASA IEEE Software Engineering Workshop, 2001.

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

[29] International Electrotechnical Commission, Division of 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.

[30] Jackson, M. The Meaning of Requirements. Annals of Software Engineering, 3:5-21, 1997.

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

[32] Kamsties, E., D. Berry and B. Paech. Detecting Ambiguities in Requirements Documents Using Inspections. Proceedings: 1st Workshop on Inspection in Software Engineering (WISE'01), 2001.

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

[34] Langacker, R. Concept, Image, and Symbol: The Cognitive Basis of Grammar. Mouton de Gruyter, Berlin, 1990.

[35] Leite, J. and A. Franco. A Strategy for Conceptual Model Acquisition. Proceedings: 1st IEEE International Symposium on Requirements Engineering, 1993.

[36] Leveson, N. Intent Specifications: an Approach to Building Human-Centered Specifications. Proceedings: 3rd International Conference on Requirements Engineering, 1998.

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

[38] Lutz, R. Analyzing Software Requirements Errors in Safety-Critical, Embedded Systems. Proceedings: 1st IEEE International Symposium on Requirements Engineering, 1993.

[39] Lyons, J. Semantics. Cambridge University Press, Cambridge, 1977.

[40] Merriam-Webster's Collegiate Dictionary, 10th ed. Merriam-Webster Incorporated, Springfield, 1993.

[41] Mervis, C.and E. Rosch. Categorization of Natural Objects. Annual Review of Psychology, 32:89-115, 1981.

[42] Mizuno, Y. Software Quality Improvement. IEEE Computer, 16(3):66-72, 1983.

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

[44] Perry, D, S. Sim and S. Easterbroook. Case Studies for Software Engineers. Tutorial Proceedings: 26th International Conference on Software Engineering (ICSE'04), 2004.

[45] Poesio, M. Semantic Ambiguity and Perceived Ambiguity. In K. van Deemter and S. Peters, eds, Semantic Ambiguity and Underspecification. CSLI Publications, 1996.

[46] Potter, B., J. Sinclair and D. Till. An Introduction to Formal Specification and Z. Prentice Hall, London, 1996.

[47] Potts, C. and W. Newstetter. Naturalistic Inquiry and Requirements Engineering: Reconciling their Theoretical Foundations. Proceedings: 3rd International Symposium on Requirements Engineering, 1997.

[48] Potts, C., K. Takahashi and A. Anton. Inquiry-Based Requirements Analysis. IEEE Software, 2(11):21-32, March 1994.

[49] Pressman, R. Software Engineering: A Practitioner's Approach, 5th ed. McGraw-Hill, Boston, 2001.

[50] Reason, J. Human Error. Cambridge University Press, Cambridge, 1990.

[51] Research Methods - Triangulation in Research Course Syllabus. Internet WWW page, at URL: <http://www.tele.sunyit.edu/traingulation.htm> (version current as of February, 2006).

[52] Rosch, E. and B. Lloyd, eds. Cognition and Categorization. Lawrence Erlbaum Associates, Hillsdale, 1978.

[53] Rosch, E., C. Mervis, W. Gray, D. Johnson and P. Boyes-Braem. Basic Objects in Natural Categories. Cognitive Psychology, 8:382-439, 1976.

[54] Ryan, K. The Role of Natural Language in Requirements Engineering. Proceedings: 1st IEEE International Symposium on Requirements Engineering, 1993.

[55] de Saussure, F. Course in General Linguistics. C. Bally, A. Sechehaye and A. Riedlinger, eds., Wade Baskin, tr. McGraw-Hill, New York, 1966.

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

[57] Taboada, M. Functional Linguistics Course Syllabus. Internet WWW page, at URL: <http://www.sfu.ca/cwil/docs_content/fac_docs/w_courses/LING_481_syllabus.pdf> (version current as of September, 2003).

[58] Taylor, J. Linguistic Categorization: Prototypes in Linguistic Theory. Clarendon Press, Oxford, 1989.

[59] Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: <http://trochim.human.cornell.edu/kb/index.htm> (version current as of August 16, 2004).

[60] Ungerer, F. and H. Schmid. An Introduction to Cognitive Linguistics. Longman and Longman, London, 1996.

[61] USAF Software Technology Support Center. Guidelines for Successful Acquisition and Management of Software Intensive Systems. Version 3.0. Internet WWW page, at URL: <http://www.stsc.hill.af.mil/stscdocs.asp> (version current as of May 2000).

[62] Wasson, K. Partial Reductive Paraphrase: Toward More Transparent Requirements. University of Virginia Department of Computer Science Technical Report CS-2004-16, 2004.

[63] Wasson, K., J. Knight, E. Strunk and S. Travis. 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.

[64] Wasson, K., K. Schmid, R. Lutz and J. Knight. Using Occurrence Properties of Defect Report Data to Improve Requirements. Proceedings: 13th IEEE International Requirements Engineering Conference (RE'05), 2005.

[65] Weiss, K., N. Leveson, K. Lundqvist, N. Farid, and M. Stringfellow. An Analysis of Causation in Aerospace Accidents. Proceedings: Space 2001, 2001.

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

[67] Wikipedia. Internet WWW page, at URL: <http://www.wikipedia.org> (version current as of January 10, 2006).

[68] Wilson, W., L. Rosenberg and L. Hyatt. Automated Quality Analysis of Natural Language Requirement Specifications. Proceedings: Pacific Northwest Software Quality Conference, 1996.

[69] Wing, J. A Study of 12 Specifications of the Library Problem. IEEE Software, 5(4):66--76, July 1988.

[70] Yin, R. Case Study Research: Design and Methods, 3rd ed. Sage Publications, Thousand Oaks, 2003.

[71] Zave, P. and M. Jackson. Four Dark corners of Requirements Engineering. ACM Transactions on Software Engineering and Methodology, 6(1):1-30, January 1997.

[72] Zowghi, D. and V. Gervasi. The Three Cs of Requirements: Consistency, Completeness, and Correctness. Proceedings: 8th International Workshop on Requirements Engineering: Foundation for Software Quality (REFSQ'02), 2002.

The covers of this book are too far apart.
--Ambrose Bierce

Appendix

This appendix contains copies of instrumentation used in CLEAR (reformatted for presentation here). It further contains a depiction of the dependency subtree for *sensor data* as constructed during the case study.

A.1 Category Template

Below are the fields of a category template, prior to completion by a CLEAR analyst. Once completed, the template is called a category record.

Category Template

Note: () indicates question(s) in collection form that informs field.

Note: templates are filled by direct quote from collection forms when appropriate, and unless indicated otherwise by [], except Synthesis and Notes section, which is analyst generated.

PHRASE:

IDENTIFYING THE CATEGORY

*A first cut expert definition. (Q1)

LOCATING THE CATEGORY

*Any domain-specificity of which the expert is aware. (Q2)

*Relevant supercategories, i.e., sub-of; relevant subcategories, i.e., super-of. (Q3, Q4) sub-of:

super-of:

*Evidence of basicness of category, within and outside of domain. (Q5)

CLARIFYING THE CATEGORY

*Most salient attributes. (Q6)

*Unexpectedly or uncommonly salient or unsalient attributes. (Q7a and Q7b)

a.

b.

SYNTHESIS AND NOTES

*What does the above evidence indicate wrt risks to the integrity of the signified?

*What accommodations can be implied with regard to supporting the integrity of the signified?

*Additional observations with potential bearing on analysis. (Q8+analyst insights)

*Clarification issues for validation

A.2 Collection Form

Below is a copy of the collection form as used to elicit category topology information from informants during the case study on the Walker Project.

Collection Form

For each phrase or collection of phrases you address, enter the phrase (or collection) where prompted below by "Phrase(s):". Then, answer the remaining questions to the best of your knowledge and understanding with regard to the concept represented by that phrase (or phrases). Not all questions will make sense for all phrases. Please consider each question carefully before making such a judgment, but if this is the case, or if a question falls outside of your domain of expertise and/or ability to answer with confidence with regard to the phrase(s), please write N/A.

Phrase(s):

1. Within the context of the walker project, what does this phrase (or phrases) mean?

2. Sometimes terms and phrases that are used by the general population are used in specialized ways within particular domains of discourse. For example, the term "clutter" as used by Air Traffic Controllers is related to, but more specialized than the version used by the general population. Similarly, the term "traffic" as used by network engineers is related to, but more specialized than the version used by the general population.

To your knowledge, are there ways in which this phrase (or phrases) as used in the walker project might differ from a lay understanding? If you know of no lay version, please indicate so.

3. Many concepts participate in conceptual relationships with other concepts. An example is supercategorization. For instance, with regard to building and crafting, wood is a member of the supercategory of raw materials out of which items can be made, along with metal, plastic, clay, etc. Similarly, running falls into the supercategory of natural locomotion for many organisms, alongside walking, crawling, hopping, etc.

Does this phrase represent a concept that falls under any supercategories relevant to the project? If so, which ones?

4. Continuing with the examples used above, in addition to participating in supercategories, wood breaks into its own subcategories. For example, oak, maple, hickory, pine and walnut are all varieties of wood heavily used to build items such as furniture. Likewise, running can be further distinguished into multiple forms such as sprinting, jogging, trotting or loping.

Does this phrase represent a concept that breaks into any subcategories relevant to the project? If so, which ones?

5. When you first encounter a concept, it may be introduced to you as a subcategory of another concept, as a supercategory of another concept, or as a first-class concept in its own right. For example, cell phones were probably introduced to you via predication on the concept of phones, while your first encounter with wood likely had little influence from its super- or subcategories.

If you were to introduce the concept in question to someone unfamiliar with it, would you be more likely to introduce it via a supercategory, a subcategory, or at its own level of abstraction? Why?

6. Among its many characteristic properties, certain properties of wood are especially noteworthy with regard to the domain of building and crafting. For example, the hardness and durability of many woods provide value to this domain, while the low cost of others is attractive.

With regard to the walker project, what are the attributes of the concept represented by this phrase that are particularly important or noteworthy?

7. Sometimes those attributes of a concept that are important in the application domain are atypical of the concept generally. For example, comparatively low cost is not a common property of woods chosen for building and crafting. However, some woods are chosen specifically for their cost, like pine. Likewise, there can be attributes of a concept that are typical, but are not important in the domain or are overshadowed by the importance of other attributes. For example, hardness and durability are typical of building woods, but pine is comparatively soft and less durable than these while still serving a clear purpose for this domain.

a. Are there any attributes of the concept in question that are atypical of the concept but important for the walker project?

b. Conversely, are there any attributes of the concept in question that are typical for the concept, but are not important (or are less important) to the walker project?

8. Having considered these issues surrounding the meaning and use of this phrase, are there any additional points you feel would be of value or interest regarding its meaning or use in the context of the walker project?

A.3 Subtree for *sensor data*

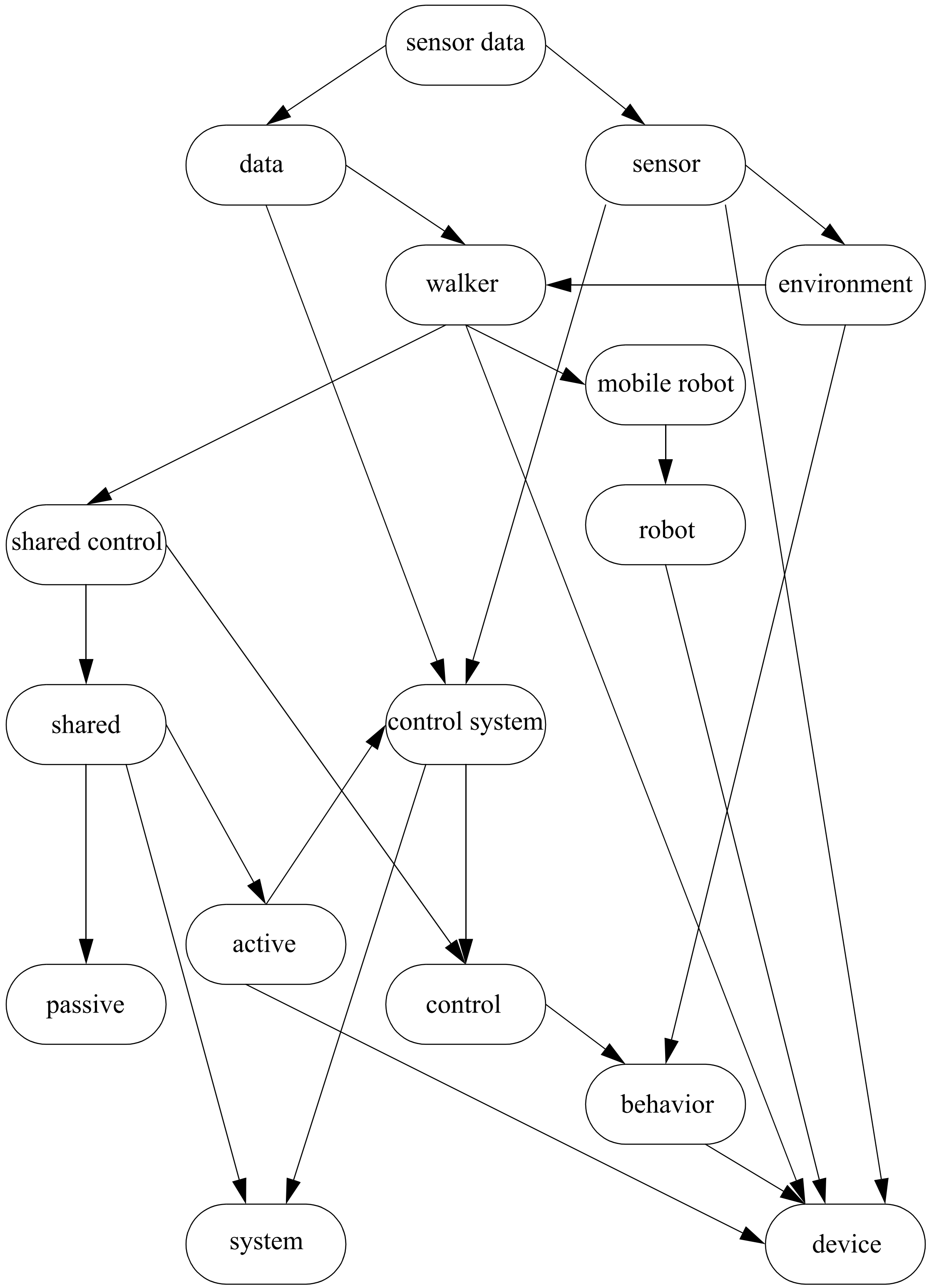


Figure 9: Explication dependencies for *sensor data*