Design and Evaluation of a Method to Reduce the Lexical Ambiguity of Requirement Specifications

Stephen Boyd (B.Eng(Comp.Sys))

A Thesis Submitted for the Degree of

DOCTOR OF PHILOSOPHY

University of Technology, Sydney
Faculty of Engineering and Information Technology

2009

Certificate of Authorship/Originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree, except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student

Production Note:
Signature removed prior to publication.

Acknowledgement

The author would like to thank three people for their support throughout this research:

1. I would like to thank my primary supervisor Prof. Didar Zowghi from the University of Technology, Sydney for her regular face-to-face discussions, problem solving advice, literature recommendations, and detailed review comments throughout the years of my research.

2. I would like to thank my secondary supervisor A/Prof. Vincenzo Gervasi from the University of Pisa, Italy for his in-depth review and accurate comments on my design solution and evaluation method.

3. Most importantly, I would like to thank my wife Alia for her understanding, patience, guidance, and continual support.

Preface

The author has previously published sections of this thesis in conference proceedings:

- Boyd, S., Zowghi, D., and Farroukh, A., 2005, 'Measuring the Expressiveness of a Constrained Natural Language: An Empirical Study', *Proceedings of the 13th International Conference on Requirements Engineering (RE'05)*, Paris, France, pp.339-349.
- Boyd, S., Zowghi, D., and Gervasi, V., 2007, 'Optimal-Constraint Lexicons for Requirement Specifications', *International Working Conference on Requirements Engineering Foundations for Software Quality (REFSQ '07)*, Trondheim, Norway, pp.203-217.

The author has been industry-supervisor for four engineering Capstone Project Reports that relate to this PhD thesis:

- Adi-wijaya, A., 2003, 'BOYDA: Object Oriented Requirements Engineering', University of Technology, Sydney, Unpublished Capstone Project Report.
- Farroukh, A., 2005, 'To define and use expressiveness to empirically derive the verbs of BOYDA for a particular domain', Faculty of Engineering, University of Technology, Sydney, Unpublished Capstone Project Report.
- Saeed, M., 2005, 'To define an unambiguous, disciplined and repeatable process for the coupling of BOYDA with an existing system modelling language', University of Technology, Sydney, Unpublished Capstone Project Report.
- Selvarajan, R., 2005, 'To develop a software tool that implements the BOYDA language and BOYDA modelling rules', University of Technology, Sydney, Unpublished Capstone Project Report.

Table of Contents

Abstract

The requirements engineering process has been criticised for its immaturity. Firstly, in the context of safety-critical systems, missing, misunderstood, and erroneous requirements have been attributed as the cause of many safety-system faults; and secondly, in the context of project success factors, many IT projects have identified requirement defects as a primary cause of being over-time or over-budget. Ambiguity is a requirement defect that is commonly associated with challenged IT projects, however there are but few empirical studies on how ambiguity can be reduced or eliminated from requirement specifications.

Eliminating the ambiguity inherent within a requirement specification is the seemingly unattainable ambition of the systems engineering zealot. This is because ambiguity is considered an unavoidable side-effect of using natural language, and most requirement specifications are written in natural language. One proposed solution to the ambiguity problem is to express requirements in Controlled Natural Language (CNL). CNLs enforce grammatical and/or lexical constraints to reduce the inherent ambiguity of natural language without sacrificing correctness, readability, or expressiveness. There is, however, a view in the literature that CNLs are overly restrictive and unnatural to read and write. Furthermore, the design and development of CNLs is both labour-intensive and time-intensive.

This thesis describes how a requirements specification can be automatically re-expressed in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality. The thesis specifically focuses on lexical ambiguity, since this is the form of ambiguity most attributable to the lexicon used to express the specification. The term re-expression is used to distinguish this approach from that of CNLs, since the lexicon is not static, but is optimally selected on a word-by-word basis such that lexical ambiguity is minimised, whilst correctness and conventionality are maximised. Fundamental to the optimal word selection is a new concept: replaceability(W_1, W_2), which is the degree to which word W_1 can replace word W_2. The replaceability equation developed within this thesis is a function of semantic similarity, polysemy, frequency, and lexical width.

We implement a software prototype, and execute it on an existing industry-specification. A controlled experiment is used to measure the effects of the re-expression in terms of correctness, conventionality, and lexical ambiguity. Data are collected from project stakeholders using a questionnaire-style approach, and hypothesis testing is used to decide whether or not the optimal re-expression has significantly reduced lexical ambiguity without significantly reducing correctness or conventionality.

Chapter 1 Introduction

1.1. Research Problem

Systems Engineering has been defined as "an engineering discipline whose responsibility is creating and executing an interdisciplinary process to ensure that the customer and stakeholder's needs are satisfied in a high quality, trustworthy, cost efficient and schedule compliant manner throughout a system's entire life cycle" (International Council on Systems Engineering 2005). EIA-632 (2003) and IEEE-Std-1220 (1994) are two of the most widely invoked standards for engineering a new system or modifying an existing system. Both standards recognise that establishing and evolving a complete and consistent set of requirements that will enable delivery of a feasible and cost-effective system solution is a primary activity of systems engineering.

A requirement is a condition, capability, or characteristic of a system or service that is needed, wanted, or desired by some user or stakeholder (Davis 2005; Sommerville & Sawyer 1997). The definition of a requirement is achieved by the requirements engineering process, which is comprised of eliciting, analysing, specifying, agreeing, and evolving requirements (Nuseibeh & Easterbrook 2000). Requirement specifications become larger and more complex as more advanced capabilities and characteristics are needed, wanted, or desired by the users or stakeholders of systems and services.

The requirements engineering process has been criticised for its immaturity (Faulk 1997). Firstly, in the context of safety-critical systems, missing, misunderstood, and erroneous requirements are causal to many safety-related software faults (Lutz 1993; Marszal & Scharpf 2002); and secondly, in the context of project success factors, only 16.2% of IT projects are successfully completed on-time and on-budget. A large share of the remaining 83.8% attribute their challenges to requirements defects (The Standish Group 1994).

One of the most recognised requirements defects is *ambiguity*. Ambiguity is characteristic of natural language, and natural language is used to express 71% of requirement specifications (Mich, Franch & Novi Inverardi 2002). Ambiguity is a major problem in requirements specifications because the different readers of the requirements specification may understand different things. If the implementers' understanding of the document differs from that of the customer or users, then the customer and the users are likely not to be satisfied with the implementation produced by the implementers. Many times, ambiguity is not noticed by anyone looking at the requirements document. Very often, each subconsciously disambiguates the document to the first interpretation he or she finds and he or she thinks that this first interpretation is the *only* interpretation (Berry & Kamsties 2003). Furthermore, a case study has shown that ambiguous requirements, when detected, lead to requirements volatility (Nurmuliani 2007) since the authors want to disambiguate. However, requirements change is characteristic of challenged projects (The Standish Group 1994), so there is a need to detect and resolve ambiguity problems very early in the requirements engineering process.

Berry and Kamsties (2003) suggest three possible solutions to the ambiguity problem: 1) Learn to write less ambiguously and less imprecisely, 2) Learn to detect ambiguity and imprecision, and 3) Use a CNL which is inherently less ambiguous and more precise. We have decided to focus on 3) CNLs, since this is the solution that appears to have the least empirical evaluation and the most speculation in the literature. CNLs aim to express text using a grammar and/or lexicon that will reduce the inherent ambiguity of natural language without sacrificing correctness or readability (Fuchs & Schwitter 1996; Schwitter 2002). However CNLs are often criticised for being overly restrictive and unnatural to read and write (Goyvaerts 1996; Somers 2003). Furthermore, the development of CNLs is both labour-intensive and time-intensive, for example at least 5 years from inception to industry adoption (Allen 2004; Nyberg, Mitamura & Huijsen 2003). Nyberg, Mitamura and Huijsen (2003) suggest that "perhaps an ideal situation for CNLs is for the machine to rewrite texts automatically into CNL without changing the meaning expressed by the sentence. For example, vocabulary selection could be done automatically when the author uses a term outside the controlled vocabulary. Sentences would be rewritten if the author uses expressions outside the CNL grammar. Furthermore, disambiguation would be done automatically with no author interaction. After the machine's rewrite is completed, the author would just read the text to confirm that it still expresses the original intention and that there are no major stylistic errors. Such a rewriting system could help to maximize author productivity and minimize training problems, while taking full advantage of the benefits of CNLs".

In summary, the problem addressed by this thesis is *ambiguous requirements*. Our motivation for solving this problem is driven by its significance, specifically the impact on project costs, schedule, and potential safety consequences. We are inspired by the vision of Nyberg, Mitamura and Huijsen (2003), and so focus this thesis on automating the re-expression of requirements in a way that significantly reduces their lexical ambiguity, without significantly reducing their correctness or readability. For reasons that we explain later in Section 2.6.1, we have bound the research question to *conventionality* rather than *readability* since it is extremely difficult to build an automated system that predicts human understanding.

1.2. Research Question

How can a requirements specification be automatically re-expressed in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality?

1.3. Motivation to Answer Research Question

Throughout the past 10 years I have been challenged with requirements-related issues almost daily. I have worked for the Royal Australian Air Force, BAE Systems, ADI-Limited, Raytheon Australia, Thales Australia, and Downer EDI Rail. I have elicited, analysed, specified, negotiated, and evolved requirements using a range of Requirements Management tools, e.g. DOORS, CORE, RequisitePro. I have been involved in all lifecycle stages of engineering a system, from the operational concept through to user acceptance testing. I have experienced requirements problems first hand, and cannot over-emphasise, the enormous amount of time that is spent trying to understand the meaning of poor quality requirements.

The project I am on at the time of writing this thesis has over 200 requirement changes in process. At least half of these are classed as ambiguous requirements. The time spent negotiating changes at preliminary design review is significantly more than what it would have been back at system requirements review, and in almost every case, the requirement could have been identified as ambiguous from the outset. The problem as I see it is that we allow our projects to live with ambiguity, without really understanding the consequences. I am motivated to change this mindset because I have seen the problems created by ambiguity in the design, implementation, test, and acceptance phases of a project.

1.4. Research Scope

The scope of our research is necessarily limited. Such limitations are progressively introduced throughout the course of this thesis, however, to assist the reader, we summarise the key boundaries from the outset.

1. There are numerous quality factors considered relevant to requirements and requirements specifications, e.g. (Davis et al. 1993). This research focuses primarily on reducing ambiguity. Specifically we focus on reducing lexical ambiguity, which occurs when a word has several meanings, also referred to as polysemy and homonymy. This research does not focus on reducing other forms of ambiguity such as syntactic, semantic, and pragmatic ambiguity as defined by Berry and Kamsties (2003).

2. There have been various methods proposed to reduce ambiguity, such as inspections (Kamsties, Berry & Paech 2001), style guides (Buley, Moore & Owess 1988; Kovitz 1998; Oriel 1999), sentence patterns (Adi-wijaya 2003; Rolland & Proix 1992; Rupp & Goetz 2000), and words-to-avoid lists (Gray 2000), however, in this thesis we focus primarily on Controlled Natural Languages (CNLs). Furthermore, we focus on controlled lexicons rather than controlled grammars due to our previous decision to focus on lexical ambiguity.

3. We limit our consideration of CNL side-effects to correctness and conventionality. We do not consider the multitude of other possible quality factors affected, such as those defined by Davis et al. (1993). We justify this decision in the thesis.

4. We develop artifacts in accordance with the design-science paradigm (March & Smith 1995). Thus the artifacts are limited to constructs, methods, models, and instantiations.

5. The scope of our evaluation is limited to a single industry specification from a single application domain, i.e. the rail-industry. We do not evaluate other specifications from other application domains.

6. The scope of our evaluation is also limited to the re-expression of verbs within requirements, and not other parts of speech. This limitation is made to control confounding effects.

7. We develop a software prototype for proof-of-concept. Since the software is a prototype, software testing is limited and informal, i.e. not documented in this thesis.

1.5. Research Contributions

This research combines design-science and natural-science using the framework and method described by March and Smith (1995), thus the artifacts produced are: constructs, models, methods, and instantiations. There are five key contributions of this research:

1. The Model: the model is a set of unique propositions that express relationships amongst domain constructs. This thesis defines four new propositions: a) Similar b) Correct, c) Conventional, and d) Unambiguous. The novel contribution is the formal definition we have provided for these concepts, which we consider to be an extension upon the literature reviewed e.g. our proposition for *conventional* is an amalgamation of Fellbaum's *richly-lexicalised* (Fellbaum 1990) and Green's *frequently occurring* (Green 2006) concepts. The synthesis of new definitions and propositions from existing literature is the reason that the Model presented in Section 4.3 is considered to be a unique contribution to the body of knowledge.

2. The Method: the method defines *how* we intend to calculate each function used by the model. This thesis defines six new functions: a) Meaning, b) Senses, c) Frequency, d) Width, e) Similarity, and f) Replaceability. The novel contribution is the formal definition we have developed for each function. The method also defines the exact sequence of activities that need to be followed to enable optimal re-expression to occur. We believe the largest single contribution within the Method is the *replaceability* metric (Boyd, Zowghi & Gervasi 2007) which aims to provide the optimal trade-off between correctness, conventionality, and lexical ambiguity. It is for these reasons that the Method presented in Section 4.5 is considered to be a unique contribution to the body of knowledge.

3. The Software Prototype: the software prototype is a re-usable and fully-functional implementation of the method. The purpose of the software prototype is to demonstrate proof-of-concept and to provide a tool to facilitate evaluation. The novel contribution is the source code itself. Other researchers may wish to re-use our software prototype to extend the evaluation to other requirement specifications; or to develop a plug-in

validation tool for existing requirements management tools. It is for these reasons that Appendix F: Prototype Software is considered to be a valuable contribution.

4. The Evaluation Method: the evaluation method defines a set of new metrics that we have designed for comparing the correctness, conventionality, and lexical ambiguity of optimally re-expressed specifications to their original NL source specification. The literature review confirmed that such metrics did not previously exist. Since there is recognition within the literature that more empirical validation of CNLs is needed to justify their merit, we believe the evaluation method and metrics that we have created in Chapter 5 are a unique contribution to the body of knowledge that can be re-used by future researchers.

5. Empirical Evidence: this thesis provides valuable empirical evidence in support of the claim that constraining natural language, in particular the lexicon, can improve the quality of requirement specifications. This evidence is a novel contribution since there is a lack of empirical evidence to support the claims made in the CNL literature.

1.6. Roadmap

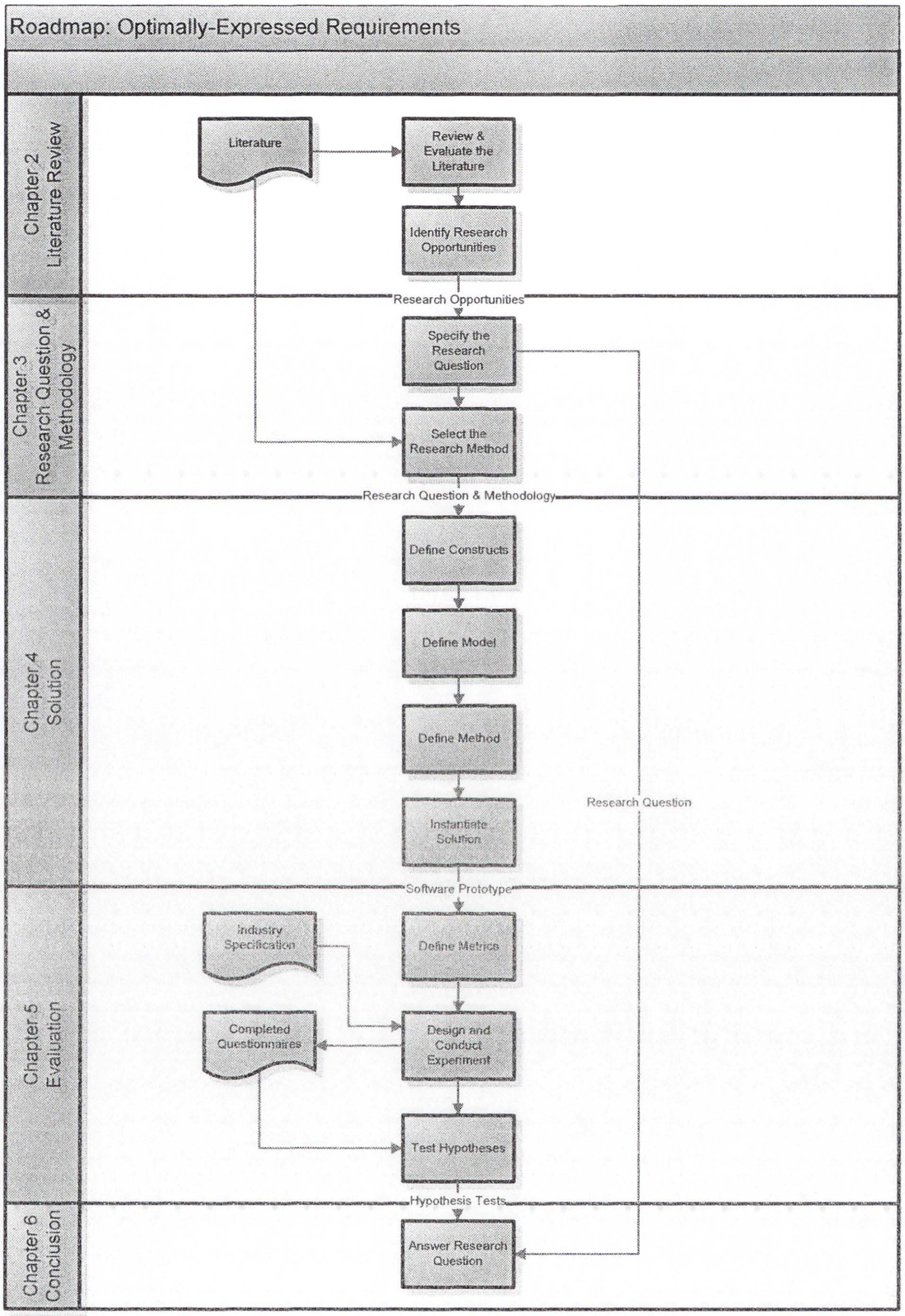

Figure 1 - Thesis Roadmap

In Chapter 2 we review and evaluate the literature relating to the topic of reducing ambiguity in requirements, and we identify our research opportunity. In Chapter 3 we transform the research opportunity into the research question addressed by this thesis.

We select March and Smith's (1995) strategic combination of design-science and natural-science as the appropriate research methodology to answer the research question. In Chapter 4 we present the design of the artifacts to answer the research question in terms of constructs, models, methods, and instantiations. One key instantiation is a software prototype that we use to evaluate our solution to the research question.

In Chapter 5 we define the metrics to collect to measure the dependent variables: correctness, conventionality, and lexical ambiguity. We nominate the evaluation method and design the controlled experiment to collect the required data using questionnaires. We use an industry-based specification as the sample specification. We define hypothesis tests for determining the *significance* in change to correctness, conventionality, and lexical ambiguity when compared to the original unconstrained industry-specification.

In Chapter 6 we answer the research question by determining whether it is possible to automatically re-express a requirements specification in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality. We consider the internal and external validity of our answer, and we highlight key implications for future research and industry practice with respect to our topic. We conclude the thesis by providing our suggestions for future research.

Chapter 2 Literature Review

2.1. Purpose

The purpose of the literature review as defined by Leedy & Ormrod (2005) is:

- to establish the importance of the problem;
- to demonstrate familiarity with significant research related to the problem;
- to describe theoretical perspectives and previous findings on the problem
- to compare and contrast existing research on the problem
- to highlight significant research opportunities relating to the problem

2.2. Systems Engineering

Systems Theory first originated in biology in the 1920s out of the need to explain the interrelatedness of organisms in ecosystems (Bale 1995). The term *systems engineering* can be traced back to Bell Telephone Laboratories in the 1940s (Schlager 1956). The need to identify and manipulate the properties of a system as a whole, which in complex engineering projects may greatly differ from the sum of the parts' properties, motivated the Department of Defense, NASA, and other industries to apply the discipline (Hall 1962). That is, systems engineering was born out of a need to manage the growing complexity of systems (Sage 1992). Systems Engineering is an interdisciplinary approach and means for enabling the realization and deployment of successful systems (Bernhard 1993).

The International Council on Systems Engineering (INCOSE) (2005) define a system as "a construct or collection of different elements that together produce results not obtainable by the elements alone. The elements, or parts, can include people, hardware, software, facilities, policies, and documents; that is, all things required to produce systems-level results. The results include system level qualities, properties, characteristics, functions, behaviour and performance. The value added by the system as a whole, beyond that contributed independently by the parts, is primarily created by the relationship among the parts; that is, how they are interconnected". The Institute of Electrical and Electronics Engineers (IEEE-Std-1220 1994) defines a system as "the top element of the system architecture, specification tree, or system breakdown structure that is comprised of one or more products and associated life-cycle processes and their products and services". The Electronic Industries Association (EIA-632 2003) defines a system as "an aggregation of end products and enabling products to achieve a given purpose". The primary difference between these definitions is that INCOSE defines a system as inclusive of *people*, however both IEEE-Std-1220 and EIA-632 define a system as an aggregation of products, where products are defined as physical items: hardware, software, documents, or services, i.e. excluding people.

A consensus of INCOSE *fellows* has defined Systems Engineering as "an engineering discipline whose responsibility is creating and executing an interdisciplinary process to ensure that the customer and stakeholder's needs are satisfied in a high quality, trustworthy, cost efficient and schedule compliant manner throughout a system's entire life cycle" (INCOSE 2005). Noting that INCOSE *fellows* are described as "individuals with significant verifiable contributions to the art and practice of Systems Engineering in industry, government or academia" (INCOSE 2005), we should have some confidence in the validity of this definition. IEEE-Std-1220 (1994) defines Systems Engineering as "an interdisciplinary collaborative approach to derive, evolve, and verify a life-cycle balanced system solution that satisfies customer expectations and meets public acceptability". We do not find there is any major conflict between these two definitions.

Systems engineering is not new. Figure 2 has been extracted from Martin (2008) and shows the historical evolution of standards and models that describe the systems engineering process. Furthermore, in the authors' experience, such standards and models often form part of a developer's contract, and can therefore be considered reflective of industry practice. In particular standards IEEE-Std-1220 (1994) and EIA-632 (2003) are a foundation to the latest standards, and tend to be mandated in many recent defence and rail industry contracts (authors' experience).

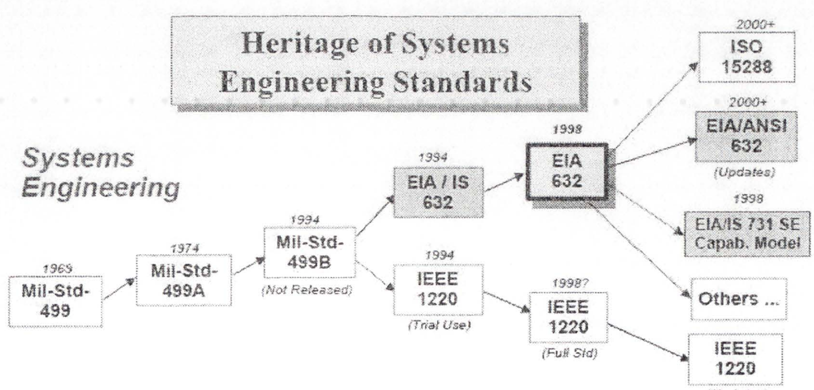

Figure 2 - Heritage of Systems Engineering Standards (from (Martin 2008))

IEEE-Std-1220 (1994) describes the process of defining the consumer product *requirements* in terms of functional and performance requirements, quality factors, producibility, supportability, safety, and environmental impacts. EIA-632 (2003) describes the process of establishing and evolving a complete and consistent *set of requirements* that will enable delivery of feasible and cost-effective system solutions. I.e. both standards focus heavily on the importance of requirements engineering.

This thesis intentionally addresses the higher-level topic of *system* requirements, as opposed to the lower-level topic of *software* requirements. This is because the authors' background and experience is as a systems engineer (recall Section 1.3), and it is often the case that requirements ambiguities originate at the system level. That being said, it is worth highlighting the fact that Sommerville (1992), Davis (1993), and Boehm (1975) all concur that the establishment, definition, analysis, and evolution of requirements is fundamental to the development of a software system. Furthermore, Brooks (1987) has claimed that "defining precisely what to build" is the *single hardest part* of building a software system.

2.3. Requirements Engineering

2.3.1. What is a Requirement

The IEEE Standard Glossary (IEEE Std 610.12 1990) defines a requirement to be:

1) A condition or capability needed by a user to solve a problem or achieve an objective.
2) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents.
3) A documented representation of a condition or capability as in 1) or 2).

Sommerville and Sawyer (1997) define a requirement as a capability that stakeholders want or a statement of a system service or constraint. Davis (2005) defines a requirement as externally observable characteristics of a desired system.

For simplicity, and to demonstrate consistency, we amalgamate these definitions as follows: *a requirement is an externally observable behaviour or property of a system or service that is needed, wanted, or desired by some user or stakeholder*. The complete set of requirements for a system or service is captured in a requirement specification.

2.3.2. Requirements Engineering Process

The requirements engineering process is responsible for producing the requirement specification. Nuseibeh and Easterbrook (2000) explain the RE process consists of:

1) Eliciting requirements,
2) Analysing requirements,
3) Specifying requirements,
4) Agreeing requirements, and
5) Evolving requirements.

Zave (1997) agrees, and provides a software-oriented definition of requirements engineering as "the branch of software engineering concerned with the real-world goals for, functions of, and constraints on software systems. It is also concerned with the relationship of these factors to precise specifications of software behaviour, and to their evolution over time and across software families".

Hsia, Davis and Kung (1993) describe RE as one of the most crucial steps in the process of creating a high-quality software product, explaining that RE consists of all activities relating to 1) identifying and documenting customer and user needs, 2) creating a document that describes the external behaviour and the associated constraints that will satisfy those needs, and 3) analysing and validating the requirements document to ensure consistency, completeness and feasibility, and satisfying evolution needs. The RE process described by Nuseibeh and Easterbrook (2000) is consistent with the RE process described by Hsia, Davis and Kung (1993), which indicates some level of maturity in the RE process. Interestingly, Faulk (1997) criticises the RE process as being "characterised by its immaturity".

2.3.3. Properties of a High-Quality Requirements Specification

IEEE Std 830 (1993) states that a requirements specification should be: a) Correct, b) Unambiguous, c) Complete, d) Consistent, e) Ranked for importance, f) Verifiable, g) Modifiable, and h) Traceable.

EIA-632 (2003) states that a requirements specification should be: a) Correct, b) Feasible, c) Focused (what not how), d) Complete, e) Modifiable, f) Unambiguous, g) Verifiable, and h) Consistent.

Other sources (Boehm 1975; Davis et al. 1993, Davis 2005; Zowghi & Gervasi 2002) suggest one or more of the following: a) Understandable, b) Unambiguous, c) Complete, d)

Consistent, e) Correct, f) Feasible, g) Minimal, h) Not Redundant, i) Agreed, j) Under/Over-Specified, k) Quantified/Precise, l) Closed, and m) Verifiable.

This highlights two key points:

1) There is no single agreed list of quality factors for requirements, and
2) There are a large number of quality factors to consider.

2.3.4. Important Quality Factors

We consider quality factors causal to schedule-overruns, cost-overruns, and safety-related accidents to be the most important quality factors.

The U.K Health and Safety Executive conducted a study of thirty-four accidents in different industries that were all caused by control and safety system failures. The results indicate the main cause of these accidents was errors in the system specification (Marszal & Scharpf 2002). Lutz (1993) agrees, and identifies that in the context of safety-critical systems, missing (incomplete), misunderstood (unclear), and erroneous (incorrect) requirements are causes of safety-related software faults.

The CHAOS Report (The Standish Group 1994) found that only 16.2% of projects are successfully completed on-time and on-budget, 52.7% of projects are completed late, run over budget, or deliver less features than originally specified, and 31.1% of projects are cancelled. In surveying 365 IT executive managers about why projects succeed and why they fail. The report found that:

The top three reasons why a project will succeed are:

1. User involvement,
2. Executive management support, and
3. A clear statement of requirements.

The top three reasons why a project will fail are:

1. Lack of user input,
2. Incomplete requirements, and
3. Changing requirements.

Bell and Thayer's historical report (Bell & Thayer 1976) on the analysis of requirement problem reports concluded that the most common types of requirement problems detected were incorrect, incomplete and unclear requirements. Interestingly, Nurmuliani's much more recent empirical case study into reasons for requirements volatility (Nurmuliani 2007) shows that unclear (ambiguous) requirements were the second most significant reason for requirements change (15.7%), missing (incomplete) requirements were the fourth most significant reason for requirements change (11.2%), and erroneous (incorrect) requirements were not listed in top five significant reasons for requirements change. Of interest, the case study found that functionality enhancements were the most significant reason for requirements change (21.8%). Thus, after 30 years of RE, it appears that requirements defects continue to trigger requirements change, and that requirements change continues to have adverse effects on software project attributes, such as schedule and cost (Stark, Skillicorn & Ameele 1999), performance (Zowghi & Nurmuliani 1998), and development effort (Pfahl & Lebsanft 2000).

Thus, whilst there are many quality factors relevant to requirement specifications, for the reasons outlined above, we consider incorrectness, incompleteness and ambiguity to be important, since they have been linked to schedule-overruns, cost-overruns, and safety-related accidents. Note however, that since the scope of this research is to focus on optimal-constraint, and since a primary goal of constrained languages is to reduce ambiguity (Berry & Kamsties 2003; Fuchs & Schwitter 1996; Huijsen 1998), we focus the remainder of this literature review on ways to reduce ambiguity, rather than on improving correctness or improving completeness.

2.4. Requirements Ambiguity

There is no single comprehensive definition of ambiguity in the software engineering literature. Berry and Kamsties (2003) derive two possible meanings for ambiguity from the Merriam Webster English Dictionary (2004), highlighting that ambiguity is itself ambiguous! (Berry, Kamsties & Krieger 2003):

1) "the capability of being understood in two or more possible senses or ways"; and
2) "uncertainty".

Berry and Kamsties (2003) explain that ambiguity is a much deeper concept than suggested by the dictionary, and present a taxonomy of ambiguity types. Bach (2000) also proposes a model of ambiguity i.e. lexical and structural, where structural in (Bach 2000) appears to be equivalent to syntactic in (Berry & Kamsties 2003). Hence the latest model proposed in (Berry & Kamsties 2003) and shown in Figure 3, also appears to be the most comprehensive and granular.

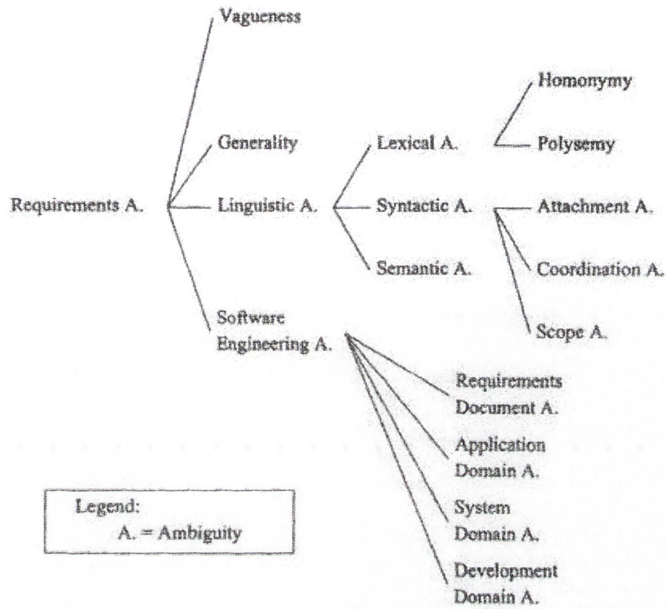

Figure 3 - Taxonomy of Ambiguity Types (from Berry & Kamsties (2003))

Note that whilst Figure 3 shows vagueness and generality as a type of requirements ambiguity, in a separate section of (Berry & Kamsties 2003), it is stated that vagueness and generality are a form of indeterminacy, which is closely related, but different from ambiguity. Boehm (1975) suggests that indeterminacy (due to missing information) is indeed a form of incompleteness rather than ambiguity. Furthermore we note that the following definitions of ambiguity by Schneider, Martin, and Tsai (1992), and Gause and Weinberg (1989) tend to relate more to the topic of indeterminacy and incompleteness than to ambiguity.

- Schneider, Martin, and Tsai (1992) define ambiguity as: "An important term, phrase, or sentence essential to an understanding of system behaviour has either been left undefined or defined in a way that can cause *confusion and misunderstanding*. Note, these are not merely language ambiguities such as an uncertain pronoun reference, but ambiguities about the actual system and its behavior"
- Gause and Weinberg (1989) define ambiguity as relating to *missing information* and *communication errors*. Missing information has various reasons. For instance, humans make errors in observation and recall, tend to leave out self-evident and other facts, and generalize incorrectly. Communication errors that occur between the author and the reader are typically due to expression inadequacies in the writing.

2.4.1. Linguistic Ambiguity

Berry, Kamsties and Krieger (2003) state that linguistic ambiguity is investigated in linguistics and in related fields, namely, computational linguistics, e.g. (Allen 1995; Hirst 1987; Lyons 1977; Voorhees, Claudia and Geoffrey 1995). There is no universally accepted definition of linguistic ambiguity. We believe the following definitions from the literature relate to linguistic ambiguity.

- IEEE-Std-830 (1993) states "An SRS is unambiguous if, and only if, every requirement stated therein has *one interpretation*."
- Davis et al. (1993) explain ambiguity: "Imagine a sentence that is extracted from an SRS, given to ten people who are asked for an interpretation. If there is *more than one interpretation*, then that sentence is probably ambiguous."
- Harwell et al. (1993) define unambiguity as: "A requirement must be unambiguous in the sense that different users with similar backgrounds would give the *same interpretation to the requirement*"
- Kamsties, Berry and Paech (2001) define a requirement as ambiguous "if it has *multiple interpretations* despite the readers knowledge of the context". Kamsties and Paech (2000) also state: "a requirement is genuinely ambiguous if it has a discrete number of interpretations, no general meaning is available which covers the distinct readings, and clarification is required to make sense of it."
- Berry, Kamsties and Krieger (2003) state "an ambiguity is *anything that causes different people to understand differently*".
- Hooks (1994) states that "terms are ambiguous because they are subjective -- they mean something different to everyone who reads them."
- Chantree (2004) says that ambiguity means "*capable of being understood in more than one way*".
- Ide and Véronis (1998) state "Ambiguity is the property of being ambiguous, where a word, term, notation, sign, symbol, phrase, sentence, or any other form used for communication, is called ambiguous if it *can be interpreted in more than one way*. Ambiguity is context-dependent: the same communication may be ambiguous in one context and unambiguous in another context. For a word, ambiguity typically refers to an unclear choice between different definitions as may be found in a dictionary. A sentence may be ambiguous due to different ways of parsing the same sequence of words."

- Bach (2000) states "ambiguity is, strictly speaking, a property of linguistic expressions. A word, phrase, or sentence is ambiguous if it has *more than one meaning*."

These definitions of ambiguity all relate to having *more than one interpretation*. We note that whilst the taxonomy of ambiguity types in Figure 3 shows three types of linguistic ambiguity 1) Lexical, 2) Syntactic, and 3) Semantic. The more recent work of Berry and Kamsties (2003) extends this to a fourth type, pragmatic ambiguity. Interestingly, this aligns with the syntactic-structural-semantic-pragmatic quality model proposed by Fabbrini et al. (1998) where syntactic is equivalent to lexical, and structural is equivalent to syntactic. Conversely, we note that Bach (2000) uses a more limiting definition, restricting linguistic ambiguity to lexical ambiguity and syntactic ambiguity (note that Bach refers to syntactic ambiguity as structural ambiguity). I.e. Bach excludes semantic ambiguity and pragmatic ambiguity from his definition of linguistic ambiguity. We now explain these forms of linguistic ambiguity.

2.4.1.1. Lexical ambiguity

Berry and Kamsties (2003) say that lexical ambiguity occurs when a word has several meanings. Ide and Véronis (1998) agree that for a word, lexical ambiguity typically refers to an unclear choice between different definitions as may be found in a dictionary. Lexical ambiguity can be further subdivided into 1) homonymy and 2) polysemy.

1) Homonymy: occurs when two different words have the same written and phonetic representation, but unrelated meanings and different etymologies, i.e. different histories of development (Berry & Kamsties 2003). For example, *bank* has several distinct definitions, including financial institution and the edge of a river.

2) Polysemy: occurs when a word has several related meanings but one etymology (Berry & Kamsties 2003). For example, the word *gas* in the requirement *there shall be a warning light telling the driver when they need more gas* could be interpreted as:
    a. gasoline, gasolene, *gas*, petrol -- (a volatile flammable mixture of hydrocarbons (hexane and heptane and octane etc.) derived from petroleum; used mainly as a fuel in internal-combustion engines)
    b. accelerator, accelerator pedal, gas pedal, *gas*, throttle, gun -- (a pedal that controls the throttle valve; "he stepped on the gas")

Of course interpreting the *gas* warning light as a need to "step on the gas" is unlikely, yet possible and highly dangerous!

The distinction between polysemy and homonymy is often unclear: is it one word with two related meanings, or two completely separate words that just happen to have the same form? (Frath 2001). Polysemy is a much larger contributor to lexical ambiguity than homonymy, since related meanings are harder to distinguish than unrelated meanings (Berry & Kamsties 2003). There are a few exceptions to this, for example, words like *light* and *bear* are homonyms that induce ambiguity in phrases or sentences in which they occur, such as *light suit* and *the duchess can't bear children.*

Veale (2004) states "Ambiguity is such a vexing problem in natural language processing that it is easy to forget that, like cholesterol, lexical ambiguity comes in both a good form and a bad form. Homonymy, the bad form, is ambiguity arising from historical coincidences of language that do not follow any predictable conceptual patterns, and which generally serve no useful purpose beyond the generation of puns. The study of homonymy may illuminate in some small way the diachronic development of language but sheds no light at all on its conceptual underpinnings. Polysemy, on the other hand, arises when two or more related meanings are shoehorned into the same lexical form for reasons of linguistic economy or creativity. Polysemy is thus a good form of lexical ambiguity. Its presence, if detected, can reveal the workings of a systematic conceptual trend at work, or a relational similarity between senses that has not been explicitly marked in the lexicon". The key phrase here is "if detected", thus Veale argues that ambiguity is acceptable in specifications, providing it is in the form of polysemy and providing the ambiguous words are detected and not misinterpreted.

Bach (2000) explains there are various tests for detecting lexical ambiguity. One test is having two unrelated antonyms, as with *hard*, which has both *soft* and *easy* as opposites. Another is the conjunction reduction test. Consider the sentence, "The tailor pressed one suit in his shop and one in the municipal court". Evidence that the word *suit* is ambiguous is provided by the anomaly of the crossed interpretation of the sentence, on which suit is used to refer both to an article of clothing and to a legal action.

2.4.1.2. Syntactic ambiguity

Berry and Kamsties (2003) say *syntactic ambiguity* (also called *structural ambiguity* (Bach 2000)), occurs when a given sequence of words can be given more than one grammatical structure, and each has a different meaning. In the terminology of compiler construction, syntactic ambiguity occurs when a sentence has more than one parse. Syntactic ambiguity can be further subdivided into 1) attachment ambiguity and 2) coordination ambiguity.

1) Attachment ambiguity: occurs when a particular syntactic constituent of a sentence, such as a prepositional phrase or a relative clause, can be legally attached to two parts of a sentence. E.g. "The police shot the rioters with guns" which can be interpreted as "The police shot [the rioters with guns]" or "[The police shot] the rioters [with guns]" (Hirst 1987), also, "Tibetan history teacher", which can be parsed in two structurally different ways, e.g., "[Tibetan history] teacher" and "Tibetan [history teacher]" (Berry, Kamsties & Krieger 2003).

2) Coordination ambiguity: occurs when a) more than one conjunction (i.e. and/or) is used in a sentence, or b) when one conjunction is used with a modifier. E.g. (a) "I saw Peter and Paul and Mary saw me". E.g. (b) "young man and woman" (Berry, Kamsties & Krieger 2003).

We have discovered two articles of literature that appear to specifically define ambiguity as syntactic ambiguity rather than the more general linguistic ambiguity.

- Grover et al. (2000) believes that if an input sentence results in more than one formalism then the sentence is ambiguous.
- Gillon (1990) says that an expression is ambiguous iff the expression can accommodate more than one structural analysis.

2.4.1.3. Semantic ambiguity

Berry and Kamsties (2003) say *semantic ambiguity* occurs when a sentence has more than one way of reading it within its context although it contains no lexical or structural ambiguity. They define semantic ambiguity as being inter-twined with syntactic ambiguity and pragmatic ambiguity. We focus here on the aspect of the definition that is not inter-twined with other definitions, that is, *scope ambiguity.*

Berry and Kamsties (2003) say *scope ambiguity* occurs when using quantifier operators such as "every", "each", "all", "some", "a", etc. and the negation operators such as "not". E.g. "all linguists prefer a theory" generates the semantic ambiguity as to whether all linguists prefer the same one theory, or whether that they prefer a perhaps different theory. We assume that Bach (2000) is considering scope ambiguity as another form of syntactic ambiguity since he does not explicitly recognise semantic ambiguity as a type of ambiguity.

Further confusing the definition of semantic ambiguity, Ravin and Leacock (2000) define semantic ambiguity as "the ways in which multiple meanings, or senses, are represented in a dictionary or lexicon and related to each other; the principles that govern these relations and the mechanisms that allow the creation of new senses." We ignore this definition because it does not make sense in the context of Berry and Kamsties (2003) *taxonomy of ambiguity types*.

2.4.1.4. Pragmatic ambiguity

Pragmatics is the study of the relations between language and context (Levinson 1983). Berry and Kamsties (2003) say pragmatic ambiguity occurs when a sentence has several meanings in the context in which it is uttered. The context comprises the language context, i.e. the sentences uttered before and after, and the context beyond language, i.e. the situation, the background knowledge, and expectations of the speaker or hearer and the writer or reader.

1) Referential ambiguity: occurs when an anaphor can take its reference from more than one element, each playing the role of the antecedent. E.g. "The trucks shall treat the roads before *they* freeze". The antecedent of the anaphor *they* can be either trucks or roads.

2) Deictic ambiguity: occurs when pronouns, time and place adverbs, such as now and here, and other grammatical features, such as tense, have more than one reference point in the context.

We now compare Berry and Kamsties (2003) definition of pragmatic ambiguity to the definition of pragmatic quality provided by Fabbrini et al. (1998). Fabbrini et al. (1998) state that pragmatic quality means 1) *valid comprehension*: everything that has been understood by the reader is actually a statement in the requirements, and 2) *complete comprehension*: every statement in the requirements has been understood by the reader. It seems that Berry and Kamsties define *pragmatic* in a more limited sense than Fabbrini et al., since *referential ambiguity* and *deictic ambiguity* relate to *valid comprehension*, but *complete comprehension* does not appear to be within the scope of Berry and Kamsties (2003) definition.

2.4.2. Why are Requirements Ambiguous?

Mich, Franch and Novi Inverardi (2002) conducted an online survey to compare natural, restricted, and formal specification languages, and they discovered that 71.8% of requirement specifications are written in Natural Language. We have noticed that this one survey has been cited by many other researchers, e.g. (Berry, Kamsties & Krieger 2003; Kiyavitskaya et al. 2008), perhaps highlighting the lack of empirical evidence on the subject. It is not clear from this study whether the trend is heading towards more formal methods, or whether NL is becoming even more popular for writing requirement specifications.

According to the Oxford English Dictionary (2005), the 500 words used most in the English language each have an average of 23 different meanings. The word *round*, for instance, has 70 distinctly different meanings. The variance of word meanings in natural language has always posed problems for those who attempt to construct an unambiguous and consistent statement. It is often the case that a written statement could be interpreted in several ways by different individuals, thus rendering the statement subjective rather than objective.

Byrd et al. (1987) finds that out of approximately 60,000 entries in Websters Seventh Dictionary, 21488, or almost 40%, have two or more senses. Moreover the most commonly used words tend to be the most polysemous. This finding supports Zipf's law (Zipf 1949), which states that "the number of meanings of a word is correlated with its frequency. Zipf argues that conservation of speaker effort would prefer there to be only one word with all meanings while conservation of hearer effort would prefer each meaning to be expressed by a different word. Assuming that these forces are equally strong, Zipf (1949) argues that the number of meanings '*m*' of a word obeys the law: $m \propto \sqrt{f}$", where '*f*' is the frequency of occurrence.

Berry and Kamsties (2003) claim "we all know that natural language is *so* imprecise, so ambiguous, and so inherently so!" IEEE-Std-830 (1993) states that "requirements are often written in natural language (for example, English). Natural language is inherently ambiguous". Kiyavitskaya et al. (2008) agrees that ambiguity is an intrinsic phenomenon of natural language. Requirements written in NL are often ambiguous, inaccurate, inconsistent and therefore very difficult to process automatically (Fraser, Kumar & Vaishnavi 1991; Schwitter 2002; Sommerville 1992). Yet 71.8% of requirement specifications continue to be written in Natural Language (Mich, Franch & Novi Inverardi 2002). Jackson (1995) suggests this is because "Requirements Engineering is where the informal meets the formal", Berry and Kamsties (2003) explain that "there is no escaping NL requirement specifications". They believe "ambiguity in natural language specifications is inescapable when producing computer-based system specifications".

Davis and Rauscher (1979) explain that "one of the most commonly used languages for expression of requirements is English. Although the freedom of expression available in English makes it easy to use and inexpensive to generate, it suffers from high inherent ambiguity level. This leads to misunderstandings between customers and suppliers, requirements writers and software designers, software designers and management, and software designers and test generation teams."

We have explained that the English lexicon is lexically ambiguous, and since most requirements are written in the English language they will inherit the same lexical ambiguities.

2.4.3. Reducing Ambiguity in Requirements

Berry and Kamsties (2003) suggest three possible solutions to the ambiguity problem: 1) Learn to write less ambiguously and less imprecisely, 2) Learn to detect ambiguity and imprecision, and 3) Use a restricted natural language which is inherently less ambiguous and more precise. So 1) and 2) reduce the disadvantage of existing natural language, and 3) restricts the natural language to disallow the disadvantages. Berry and Kamsties (2003) omit *modelling* as a solution to reducing ambiguity. For instance, the Unified Modelling Language (UML) is a mainly graphical way of expressing problems and their solutions (Fowler & Scott 2000). This is relevant because some things that can be difficult to express textually, such as multiplicity, are easily handled by modelling languages. We have decided not to investigate modelling languages any further in this literature review, since our research question pertains to reducing lexical ambiguity which is textual, not graphical.

2.4.3.1. Learn to write less ambiguously and less imprecisely

Berry, Kamsties and Krieger (2003) explain that at least three distinct strategies can be identified to avoid ambiguity in the written requirements. First, the precision of natural language can be increased. Second, more contextual information can be provided in order to allow the reader to resolve ambiguities herself. Third, conventions on how ambiguous phrases shall be interpreted can be set up between the writer and the reader. We now consider glossaries, style guides, sentence patterns, and poor word lists as potential solutions.

A glossary or dictionary defines important terms and phrases used in a requirements document. Thus, it helps avoid lexical ambiguity. It requires considerable effort to create and validate a glossary, but the initial effort pays off since it can be reused for future projects within the same application domain. For example, Kovitz (1998) gives detailed guidelines for creating a glossary. Designations by Jackson (1995) and the Language Extended Lexicon by Leite and Franco (1993) are other techniques for grounding terms in reality.

A style guide is a set of rules on good practices in requirements writing. Ben Achour (1998), Buley, Moore and Owess (1988), Hull, Jackson and Dick (2002), Kovitz (1998), Oriel (1999) have all proposed style guides to assist requirements authors in writing better quality requirements. An example of a rule that is found in many style guides is "use active voice rather than passive voice".

Sentence patterns have been proposed to give the requirements author support in articulating requirements. Alexander and Stevens (2002), Rolland and Proix (1992), Rupp and Goetz (2000), Wiegers (2006), and Adi-Wijaya (2003) have all proposed sentence patterns for writing requirements. Typically, the requirement must be rewritten slightly to fit the pattern. E.g.:

```
[when?] [under what conditions?] THE SYSTEM SHALL | SHOULD | WILL
<process> <thing to be processed> [<process detail>*].
```

Poor word lists provide a collection of ambiguous words that should be avoided. For example Gray (2000) provides a collection of ambiguous words that should be avoided. Similarly, Hooks (1994) lists ambiguous and vague words to avoid e.g. "acceptable", "accurate", "appropriate", "easy", "efficient", "essential", "immediately", "minimum", "maximum", "periodically", "sufficient", "user-friendly", etc.

Thus, we have found that glossaries, style guides, sentence patterns, and poor word lists are already well addressed by the literature.

2.4.3.2. Learn to detect ambiguity and imprecision

Berry, Kamsties and Krieger (2003) explain that there are a few different strategies for detecting ambiguities.

1. Search for particular patterns of ambiguity, as done in reading techniques for requirements inspections, e.g. (EIA-632 2003; Kamsties, Berry & Paech 2001;

Kamsties 2001; Gause & Weinberg 1989). Some ambiguity detection techniques have been automated by natural language processing tools, e.g. (Fabbrini et al. 2001; Gervasi & Nuseibeh 2000; Ishihara, Seki & Kasami 1993; Mich & Garigliano 2000).

2. Compare the interpretations of a document by different stakeholders; if they differ, there is an ambiguity in the original document, e.g. (Easterbrook & Callahan 1998; Lami 2005; Shubert et al. 1995; Stewart 1998).
3. Communicate an interpretation back to the requirements author, after which she can easily point out misinterpretations, e.g. (Lami 2005; Wilson, Rosenberg & Hyatt 1997).

The last two are perhaps the most effective strategies for finding ambiguities in requirements specifications, but they demand more resources than other strategies and are therefore applicable only in situations in which the cost of not finding ambiguities is at least the cost of the resources required to find the ambiguities (Berry, Kamsties & Krieger 2003).

2.4.3.3. Use a restricted natural language

We have explained that requirements are ambiguous because they are typically written in natural language, and that natural language is inherently ambiguous. Formal Languages (FL) such as Z, B, and LOTOS have been introduced to combat the ambiguity of NL. In theory, formal languages should be the preferred choice of developers since they aim to be unambiguous, clear, concise, precise, and computer processable (Schwitter 2002; Sommerville 1992), however, Fraser, Kumar & Vaishnavi (1991) explain that formal languages typically require a degree of mathematical sophistication to learn and understand. Schwitter (2004) acknowledges that FLs are difficult for non-specialists to apply. It can be very cumbersome to describe the total behaviour of a complex software system because of the vast number of possible uses of the future system. Neither is it clear what kind of language should be used to describe the behavioural requirements so that a specification can be derived in a systematic way from these requirements (Jackson 1995).

Berry and Kamsties (2003) state "There is that old trade-off: requirements specifications written in natural languages versus mathematics-based formal languages. On one hand, natural language is inherently ambiguous, but there is always someone who can write it, and it is always more or less understood by all stakeholders, albeit somewhat differently by each. On the other hand, mathematics-based formal language is inherently unambiguous, but there is not always someone who can write it, and it is not understood by most stakeholders, although all that do understand it understand it the same".

Davis and Rauscher (1979) state "as requirements become more formal, there are fewer inter-party disagreements and misunderstandings. As a requirements language becomes more applications-oriented and better human-engineered, there is more customer understanding. Ideally, we would like to have a language which is the ultimate in both aspects." They go on to say "The ultimate situation is to have different requirements languages for different application areas". Whilst this may seem unrealistic, impractical and highly-costly, it highlights the key point that requirements languages must be adaptable to be useful. This important point influences the construction of our research question.

Dodd (1990) sums it up: "Somewhere between ridiculous pedantry and erroneous formulation there presumably exists a reasonably precise way of specifying a problem in English".

2.5. Controlled Natural Language

IEEE-Std-830 (1993) states that "one way to avoid the ambiguity inherent in natural language is to write the requirements in a particular requirements specification language". Berry, Kamsties and Krieger (2003) define a controlled (restricted) language as "a precisely defined subset of natural language for the use in specific environments". Schwitter (2002) agrees, and clarifies that a CNL is a subset of a NL, such as Standard English (Quirk & Greenbaum 1996) that has been restricted with respect to its grammar and/or lexicon. Huijsen (1998) describes a CNL as "...an explicitly defined restriction of a natural language that specifies constraints on lexicon, grammar, and style. The overall aim here is the reduction in ambiguity, redundancy, size and complexity".

Grammatical restrictions result in less complex and less ambiguous sentences. Lexical restrictions reduce the size of the vocabulary and the meaning of the lexical entries for a particular application domain. Mich, Franch and Novi Inverardi (2002) estimate that 15.9% of specifications are written in structured natural language. It is unclear whether this number is on the rise, or whether CNLs have lost popularity since the survey. We were unable to find any other surveys to use as a source of comparison.

There appears to be conflict within the literature about the effects of CNLs. On the one hand, there are claims that CNLs "increase the readability and understandability" (Fuchs & Schwitter 1996; Grover et al. 2000; Pulman 1996). I.e. CNLs have been proposed to solve the problem of over-flexibility of natural language, and to improve the quality of specifications without losing their readability (Schwitter 2004; Sommerville 1992). On the other hand, there are claims that CNLs tend to be unnatural to read and write (Somers 2003). IEEE-Std-830 (1993) states that a disadvantage in the use of CNLs is the length of time required to learn them. Goyvaerts (1996) claims that writing requirements in controlled languages is 20% more time consuming that writing requirements in unrestricted NLs. Unfortunately, we were unable to ascertain the specific method Goyvaerts used to make such an estimate.

Nyberg, Mitamura and Huijsen (2003) provide a detailed description of CNL evolution and highlight that there is no standardised CNL, say for English, which is approved by some global authority. Rather than simply repeating their excellent historical account, we have attempted to add value by summarising the key information, and by citing other relevant work that was for some reason omitted by Nyberg, Mitamura and Huijsen.

2.5.1. Evolution of Controlled Natural Languages

The notion of CNLs can be traced back to the 1930s (Ogden 1932) and 1940s (Ogden 1942). Ogden's Basic English consists of 850 words and few rules that describe how to inflect these words and how to derive other words. The first CNL put to actual use was Caterpillar Fundamental English (CFE), used by Caterpillar Inc. in the 1970s. CFE was to enable non-English speakers to read technical manuals after some basic training. The intention of CFE was to eliminate the need to translate documents into other languages. However, the vocabulary was a mere 743 words in 1972. CFE was abandoned in 1982 primarily because equipment became too complex to express with such a limited vocabulary (Kamprath et al. 1998). As a part of a major modernization effort by Caterpillar Inc., and to overcome the limitations of CFE, Caterpillar Technical English (CTE) was created. CTE was designed to be an Enforceable Controlled English – in contrast to CFE, which was not enforced by Caterpillar Inc. at its time of use.

Rather than training technicians in CFE, CTE was designed to be machine translatable into 35 different languages. This was a fundamental shift from Caterpillars initial approach which was to write all manuals in *fundamental* English such that there was no need for translation into other languages. CTE was developed from 1992-1997 by 5 full-time people, specifically, linguists and domain experts. CTE contains over 70,000 words that were designed to be unambiguous to the human reader.

The large volume of documentation (over 100,000 new pages each year) and the requirement of translation in up to 35 languages necessitate heavy automation of the translation process. To this end, Caterpillar engaged Carnegie Group Inc. and the Center for Machine Translation (CMT) at Carnegie Mellon University to develop and deploy a combined authoring and translation system based on CMT's KANT technology. The KANT Machine Translation (MT) system (Nyberg & Mitamura 1992) produces high-quality translation provided that the source language is strictly controlled for both vocabulary and grammar.

In 1979, the Douglas Aircraft Company constructed a dictionary of about 2,000 words which it uses for technical manuals. In the UK, Perkins Engines Ltd. introduced Perkins Approved Clear English (PACE) (Douglas & Hurs 1996) to simplify their publications and to aid translation in the 1980s. In Sweden, the Scania company, a leading manufacturer of heavy trucks, has defined ScaniaSwedish (Alqvist & Sagvall Hein 1996), a CNL for the automatic translation of truck maintenance documentation. Documentation is written by technical writers at Scania in Swedish, and is then translated in its full versions into seven languages: English, German, Dutch, French, Italian, Spanish and Finnish.

One of the best-known CNLs is Simplified English (SE), a human-oriented CNL for aircraft-maintenance documentation (AECMA 1986). SE includes a limited basic vocabulary of approximately 3,100 words and a set of 57 writing rules. The guiding principle in the SE lexicon, as in any other CL lexicon, is "one word one meaning". The vocabulary can be extended with aircraft-industry terminology as needed: technical names, which are nouns denoting specialized aircraft entities (e.g. fuselage, air-traffic control), and manufacturing processes, which are verbs denoting industrial activities (e.g. anneal, polish). The writing rules pertain to punctuation, word choice, sentence length, syntactic constructions, text structure, style, and layout.

Very few researchers have tried to employ CNLs for writing requirements specifications; we found (Fuchs & Schwitter 1996; Grover et al. 2000; Schwitter 2002). There has been limited research into the formalisation of NL requirements; we found (Ishihara, Seki & Kasami 1993; Macias & Pulman 1993; Pulman 1994).

Attempto Controlled English (ACE) (Fuchs & Schwitter 1996) is designed specifically for specifying software requirements such that the requirements can be translated into first-order predicate logic. The translated document can be verified for completeness and consistency by querying it in ACE, thus validation and prototyping in concepts close to the application domain become possible and the results can be understood by all parties concerned. Schwitter defines PENG (Processable ENGlish) as a computer-processable CNL (Schwitter 2002). PENG covers a well defined subset of standard English and is precisely defined by a controlled grammar and a controlled lexicon. The controlled language contains domain specific content words (nouns, verbs, adjectives, adverbs) and predefined function words (pronouns, prepositions, conjunctions, determiners, and interrogatives). Like ACE, PENG is designed for specifying software requirements (in fact Schwitter was heavily involved in the development of ACE). There are no obvious differences between the controlled grammar and controlled lexicon of PENG versus the controlled grammar and controlled lexicon of ACE.

According to Schwitter, the primary advantage PENG has over ACE is ECOLE – a look ahead text editor that guides the requirements engineer through the grammar of the CNL. In this way, the domain specialist does not need to learn and to remember the restrictions of the controlled language. The development of ECOLE was based on the premise that if the syntactic and lexical restrictions are too hard for the author to remember, or if it takes too long for the author to come up with a sentence that conforms to the controlled language definition, then such controlled languages will not be acceptable. This view is shared by Berry & Kamsties (2003) who state "there is a genuine need for tool support for writing requirements documents in order to enforce the grammar and the fixed vocabulary of the controlled language."

Grover et al. (2000) have developed a controlled language for the specification and verification of hardware designs is used in interactive, dialogical settings, without extensive redrafting of the input documents and without requiring its users to be trained in the use of the language. In this way Grover's hardware language is similar to PENG with ECOLE.

Table 1 was first presented in (Boyd, Zowghi & Farroukh 2005) and provides a comprehensive list of the CNLs we have encountered throughout our review of the literature. We later discovered that Spaggiari, Beaujard and Cannesson (2003) created a similar list – importantly, the two lists overlapped, giving a greater degree of confidence in the completeness of Table 1.

Table 1 - CNL vs. Domain

Controlled Natural Language (CNL)	Domain
Alcatel COGRAM (Adriaens & Schreurs 1992)	Telecommunications
Attempto Controlled English (ACE) (Fuchs & Schwitter 1996)	General
Caterpillar Fundamental English (CFE) (Kamprath et al. 1998)	Heavy Equipment
Caterpillar Technical English (CTE) (Kamprath et al. 1998)	Heavy Equipment
Simplified English (SE) (AECMA 1986)	Aerospace Technical Manuals
Controlled Automotive Service Language (CASL) (Allen 2004)	Auto Manufacturing
Grover et al Hardware Specification Language (Grover et al. 2000)	Hardware Specifications
IBM EasyEnglish (Bernth 1997)	Information Technology Services
Kodak International Service Language (KISL) (Allen 2004)	Photographic/Digital Technology
Perkins Approved Clear English (PACE) (Douglas & Hurs 1996)	Diesel Engine Manufacturing
Processable ENGlish (PENG) (Schwitter 2002)	General
ScaniaSwedish (Alqvist & Sagvall Hein 1996)	Truck Maintenance Manuals
Sun Controlled English (Allen 2004)	Computer Hardware
Xerox Multilingual Customized English (Allen 2004)	Printing Equipment
Airbus A3XX (Spaggiari, Beaujard & Cannesson 2003)	Aircraft Warning Messages

2.5.1.1. Development of Controlled Natural Language

In order to define a controlled vocabulary for a particular application domain, pre-existing documents are analyzed as an initial source of vocabulary (Grover et al. 2000; Nyberg, Mitamura & Huijsen 2003). Domain experts are engaged, and automatic parsers are utilised to extract domain keywords and re-occurring phrases respectively (Kamprath et al. 1998). The literature confirms that:

- ScaniaSwedish was derived from 15,000 pages of text (Alqvist & Sagvall Hein 1996),
- Caterpillar Technical English from 50Mb of data (Kamprath et al. 1998),
- A3XX from 3000 sentences and 700 words (Spaggiari, Beaujard & Cannesson 2003)
- SE was derived from four large aircraft maintenance manuals (AECMA 1986).

Note that in all cases it is unclear whether these sample sizes were decided to support an external validity argument or whether this was all that was available at the time. We note this because in our research we need to make an external validity claim, and we are interested in how other researchers have approached the sample size question. Unfortunately the sample sizes were all defined in different and non-comparable units.

Nyberg, Mitamura and Huijsen (2003) state that when defining a controlled English for a new domain, there are two key rules: 1) limit one meaning per word–part-of-speech pair, and 2) replace polysemous words with their monosemous synonyms. Applying these rules was a manual process for all CNLs reviewed. For example, in CTE, an instruction was given to the lexicographer to "identify ambiguous terms and eliminate polysemy by restricting each term to a single meaning" (Kamprath et al. 1998).

Due to the manual effort involved in the development of a CNL, the resources required are not insignificant. For example, the personnel required at Caterpillar for CTE development, pilot, and training, averaged about five full-time equivalent employees per year for five years (Nyberg, Mitamura & Huijsen 2003). Allen states that creating a domain specific CNL can take from 1 month to 2-4 years in research, and 5-6 years to implement (Allen 2004).

We notice that all of the CNLs designed for specifying requirements, i.e. ACE, PENG, and the Grover et al. Hardware Language, constrain the grammar and function words (pronouns, prepositions, conjunctions, determiners, and interrogatives) but do not constrain the content words (nouns, verbs, adjectives, adverbs). Their reason for not constraining content words is to maintain domain independence, i.e. content words are those words that typically reflect the domain – e.g. proper nouns. However, since content words are the most polysemous parts of speech (Merriam Webster English Dictionary 2004; Oxford English Dictionary 2005), and since content words represent more than 99% of the English lexicon (Oxford English Dictionary 2005), the effect of constrained function words on the lexical ambiguity of a requirements specification is arguably negligible.

2.5.1.2. Claimed Quality Improvements

We have encountered many *claims* in the literature about the advantages of various CNLs. We have found that in the majority of cases, the claims appear to be unsubstantiated in terms of any empirical evidence. In this section we briefly describe these claims.

- Kamprath et al. (1998) state that CTE is "*designed to be unambiguous to the human reader*". This is because in the event of polysemy, the linguists chose the most appropriate sense of the word for a particular part of speech. Kamprath et al. limit the definition of ambiguity to polysemy, i.e. lexical ambiguity. We note there is no reference to any empirical evidence to validate whether CTE actually is *unambiguous to the human reader*. They also explain that "CFE was abandoned in 1982 primarily because equipment became too complex to *express* with such a limited vocabulary", however it is not clear how CTE was designed to be more expressive.

- Alqvist and Sagvall Hein (1996) claim that ScaniaSwedish applies "Systematic restrictions with the aim of eliminating unnecessary linguistic variation whilst preserving the *expressive power* that is required", however it is not explained what "expressive power" means.
- Grover et al. (2000) claims to "...finding an optimal trade-off between *expressiveness* and *tractability*" but it is not clear how her CNL has achieved the optimal trade-off.
- Spaggiari, Beaujard and Cannesson (2003) claims that "reducing the complexity of syntactic structures of a text increases its readability", but it is not explained how reducing the complexity increases readability.
- Reuther (1998) states that "it is a well known and indisputable fact within the CL community that the use of a Controlled Language (CL) in technical documentation leads to quality improvement with respect to readability, consistency and translatability". But clearly, as evidenced by this literature review, it is disputable.
- Fuchs and Schwitter (1996) make a number of unsupported claims about ACE.
    - "...specifications can be written in (controlled) natural language, so that they are *readily understood*, while their interpretation *is unambiguous*."
    - "ACE can be accurately and efficiently processed by a computer, but is *expressive enough* to allow natural usage."
    - "ACE enforces writing standards that restrict the grammar and the vocabulary, thus leading to documents containing *more predictable* and *less ambiguous* language"
    - "ACE is *sufficiently expressive* to write specifications of *high quality* and *high readability* and *understandability*"
    - "the right trade-off between *expressiveness* and *processability*"
    - "The meaning of the specification text can finally be *represented unambiguously*"
- Schwitter (2002) makes similar unsubstantiated claims about PENG.
    - "Grammatical restrictions result in *less complex* and *less ambiguous sentences*"
    - "By reducing the size of the vocabulary, texts become *easier to read and to understand for humans*, and easier to process for machines."

ACE and PENG constrain the grammar and function-words (i.e. determiners, prepositions, conjunctions, pronouns). ACE and PENG do not constrain the content-words (i.e. nouns, verbs, adjectives, adverbs). Since nouns and verbs are the most polysemous parts of speech (Merriam Webster English Dictionary 2004; Oxford English Dictionary 2005), Fuch's and Schwitter's claim to *unambiguity* is considered to be weak. It seems that *syntactic ambiguity* was meant instead of the more general ambiguity, since ACE and PENG focus on the grammar, whereas lexical ambiguity is expected to be largely unimproved since there is no constraint on the use of polysemous content words.

2.6. Quality Model

In Section 2.3.3 we highlighted two key points:

1) There is no single agreed list of quality factors, although there is considerable overlap between existing lists and essentially no conflict between them, and
2) There are a large number of quality factors to consider.

In Section 2.3.4 we explained that whilst incorrect, incomplete and ambiguous requirements are reported as contributors to schedule-overruns, cost-overruns, and safety-related accidents, we decided to limit the scope of the literature review to ways of reducing ambiguity in requirement specifications.

In Section 2.5 we explained that there appears to be conflict within the literature about the effects of CNLs. On the one hand, there are claims that CNLs increase readability (Fuchs & Schwitter 1996; Grover et al. 2000; Nyberg, Mitamura & Huijsen 2003; Pulman 1996; Reuther 1998; Schwitter 2002; Sommerville 1992; Spaggiari, Beaujard & Cannesson 2003), and on the other hand there are claims that CNLs reduce readability (Huijsen 1998; IEEE-Std-830 1993; Somers 2003). Furthermore there is a tendency to assume that CNLs reduce the expressiveness (Nyberg, Mitamura & Huijsen 2003; Somers 2003), and are more time consuming to write (Goyvaerts 1996; IEEE-Std-830 1993; Nyberg, Mitamura & Huijsen 2003; Somers 2003). In Section 2.5.1.2 we highlighted that many of these claims are speculatory because they lack any empirical support.

We therefore believe that the minimum set of quality factors that must be addressed when measuring the effectiveness of a CNL are:

- Ambiguity,
- Readability,
- Conventionality
- Writability,
- Expressiveness, and
- Correctness

Since ambiguity has already been discussed at length in Section 2.4, we now continue with covering the remaining quality factors.

2.6.1. Readability

Readability is regularly mentioned in the CNL literature, but is rarely defined by the authors that use it. Perhaps this is because it is an idea that is prevalent in general language, and authors assume that readers understand the concept as "the ease with which written language can be read with understanding" (Crystal 1992). Klare (1974) states that when we talk of readable writing, "...we mean that the intended readers are able to read it quickly, understand it clearly, and accept it readily (i.e. persevere in reading it)".

Sommerville (1992) explains that CNLs try to reduce the ambiguity of specifications without losing their readability. Somers (2003) explains that one of the biggest criticisms of CNLs is that they tend to be unnatural to read. IEEE-Std-830 (1993) states that a disadvantage in the use of CNLs is that many non-technical users find the resulting specification unintelligible. Huijsen (1998) says many controlled languages risk reducing readability by replacing ambiguous words with unconventional alternatives. For example, authors may be told to use the word "right" only in the sense of "right-hand side" and not in the sense of "correct". This can create problems when the reader is familiar with the word "right" meaning "correct". On the other hand, Spaggiari, Beaujard and Cannesson (2003) demonstrate that reducing the complexity of syntactic structures of a text increases the readability. Nyberg, Mitamura and Huijsen (2003) agree that a reduction in the complexity of the lexicon and the adherence to writing rules may improve the readability.

Hence the literature is divided about whether constraining a language has positive or negative effects on readability. Nyberg, Mitamura and Huijsen (2003) find that there are but a few empirical studies on this subject. Knops (2000) highlights that: Generally speaking, there is an urgent need for facts and figures obtained in experimental situations and real-life production environments and relating to the effects of particular CL standards, rules and rule sets on readability. We found the following:

- Chervak, Drury and Ouellette (1996), Holmback, Shubert and Spyridakis (1996), Kincaid (1997), and Shubert et al. (1995) have all studied the effects of Simplified English on the readability of maintenance documentation for native English speakers versus non-native English speakers. The results of these studies suggest that the improvements in readability and ambiguity gained by using SE are much more pronounced for non-native English speakers than for native English speakers.

- Spaggiari, Beaujard and Cannesson (2003) interviewed 64 people, and concluded that aircraft warning texts rewritten in A3XX appear to be more readable than those written naturally. So this provides some empirical evidence that constraining text can improve its readability.
- Cadwell (2008) aims to answer the question "*is readability increased by applying controlled-language rules to texts?*" Cadwell conducts a survey of text-user attitudes and finds that a majority of participants find the text written in CNL is more readable that the text written in NL. We note that Cadwell's survey includes CTE, PACE, Grover, SE, A3XX but excludes RE specific CNLs such as ACE and PENG.

Interestingly, Cadwell (2008) finds that extra-linguistic variables including technical expertise, participant profile, and familiarity with the terminology and subject matter, have a high impact on readability and must be taken into account. In particular, familiarity with domain terminology appeared to have a strongly positive impact on readability. This would seem to suggest that being familiar with the domain of a text, even if you do not necessarily comprehend it, or use it effectively, makes that text seem more readable to you. The view of Cadwell (2008) is consistent with that of Swaffar, Arens and Byrnes (1991) which suggests that what makes text readable is that it "deals with topics of interest or familiar to the intended readers (so that it allows for communication and expressions from within the readers' frame of reference)".

Thus we understand from Cadwell (2008) and Crystal (1992) that the readability of a text is essentially a function of 1) whether the words within the text are conventional to the reader, and 2) whether the reader is able to understand the meaning of the text. In our opinion, whether the words within a text are conventional to a reader, is more controllable than whether the reader is able to understand the meaning of the text. This is because reader understanding is largely a function of reader background, education, intelligence, etc (Cadwell 2008), whereas conventionality is largely a function of word frequency and word position within a lexical hierarchy (as we explain in the next section). Cadwell (2008) has empirically shown that terminology that is conventional in the domain is more likely to be readable; however, it is difficult to build an automated system that predicts human understanding. This explains our bounding of the research question to *conventionality* rather than *readability*.

2.6.2. Conventionality

A word hierarchy provides a logical arrangement of words, e.g. from the most general to the most specific as found in WordNet (2003). Words are likely to be more familiar to the reader, and therefore more readable, when they are expressed at a hierarchical level known as the *basic level* – than is the case for superordinate or subordinate levels (Berlin 1992). When we refer to a specific entity, we usually do so by using a label for a class the entity is a member of (Brown 1958; Lakoff 1987; Rosch et al. 1976). For example, we seldom bother to refer to an automobile by its Vehicle Identification Number (VIN), which would enable us to name it uniquely, but, given a neutral context, we typically refer to it as a *car*. But the class of cars is not the only class that could be used for referring to a specific automobile; both more specific (e.g., *sedan*) and more general (e.g., *vehicle*) classes/names are also available. Importantly, more specific and more general categories are more difficult to comprehend than basic level categories (Brown 1958; Rosch et al. 1976).

Fellbaum (1990) suggests the basic-level can be identified as a *bulge*, that is to say, a level far more richly lexicalized than the other levels in the same hierarchy. The extent to which a word is *richly lexicalised* is measured as the number of coordinate (sibling) terms in a word hierarchy (i.e. the size of the set of words that have the same parent concept). Fellbaum calls the most richly lexicalised layer L0, the layer above it L+1, and the layer below L-1.

Alternatively, Green (2006) proposes a set of linguistically-oriented heuristics for automatically identifying *basic-level* words in a noun hierarchy. We note that one of Green's linguistically-oriented heuristics suggests that the frequency of occurrence of a word is some measure of the *conventionality* of the word. So the conventionality of a word can be measured both in terms of how richly lexicalised the word is in some hierarchy (Fellbaum 1990), and in terms of how frequently used the word is in some textual corpus (Green 2006).

Since readability is a function of both *understanding* and *conventionality* (Cadwell 2008), it seems that an appropriate goal for CNLs is to replace words with more conventional alternatives, or at least to not replace with less conventional alternatives.

2.6.3. Writability

To be *writable* means: "capable of being put in writing" (Merriam Webster English Dictionary 2004). Somers (2003) explains that one of the biggest criticisms of CNLs is that they tend to be unnatural to write. Being unnatural tends to reduce the authors' capability to put their intended meaning into writing.

IEEE-Std-830 (1993) states that a disadvantage in the use of CNLs is the length of time required to learn them. Goyvaerts (1996) claims that writing requirements in controlled languages is 20% more time consuming that writing requirements in unrestricted NLs. Nyberg, Mitamura and Huijsen (2003) agree that the writing task may become more time-consuming. It can take more concentration to write documents if they must conform to the rules of a CNL, which can slow down the writing process. Berry and Kamsties (2003) ask "is writing in such a [controlled] language really different from writing in a mathematics-based formal language?" Thus we have found consensus in the literature that writing in CNL is more time consuming than writing in NL.

In Section 1.1 we quoted a vision expressed by Nyberg, Mitamura and Huijsen (2003) which stated that "perhaps an ideal situation for CNLs is for the machine to rewrite texts automatically into CNL without changing the meaning expressed by the sentence. For example, vocabulary selection could be done automatically when the author uses a term outside the controlled vocabulary. Sentences would be rewritten if the author uses expressions outside the CNL grammar". Writability would no longer be a problem, since the author would write naturally, and the machine would re-express the authors' original text into CNL. The key message here is that by automatically translating NL into CNL, the single most agreed disadvantage of CNLs could be overcome. This explains why our research question focuses on *automatic re-expression* rather than *writability*.

2.6.4. Expressiveness

We have found three definitions of expressiveness:

1) Nyberg, Mitamura and Huijsen (2003) believe the expressiveness of a language is some measure of the variety of lexical and grammatical constructions it allows.
2) Gnesi et al. (2005) believe the expressiveness of a language relates to its ability to convey meaning such that the meaning is understood by the human recipient.
3) Fabbrini et al. (1998) imply that the expressiveness of a controlled language is different for every domain/problem, and is defined as the ability of the CL to express the problem.

There are fundamental differences between these three definitions. Gnesi et al. (2005) believe expressiveness is dependant on whether the human actually understands what is expressed, whereas Nyberg, Mitamura and Huijsen (2003) and Fabbrini et al. (1998) base their definitions on what can be expressed, and exclude the human understanding in their definitions. Furthermore, Fabbrini et al. believe expressiveness is a domain or problem specific measurement, whereas Gnesi et al. and Nyberg, Mitamura and Huijsen maintain a more general domain independent definition.

It is common for CNL developers to try and justify the expressiveness of their CNLs (Alqvist and Sagvall Hein 1996; Fuchs and Schwitter 1996; Grover et al. 2000). We assume this is because reduced expressiveness is perceived as an unavoidable consequence of constraining the English language, i.e. the expressiveness of a language is some measure of the variety of lexical and grammatical constructions it allows (Nyberg, Mitamura & Huijsen 2003). Since a CNL constrains such lexical and grammatical constructions – the subsequent expressiveness of the language is expected to decline (Boyd, Zowghi and Gervasi 2007). Berry and Kamsties (2003) state that "by restricting the natural language, it is not so natural anymore. Then the question of expressibility comes up: Have we lost something valuable in the restrictions?" Nyberg, Mitamura & Huijsen (2003) explain that authors may experience a reduction in the power of expression if words that express the meaning they want to convey are unapproved and no good alternatives are provided. This is important since our research question relates to *automatic re-expression.*

2.6.5. Correctness

Perhaps one of the earliest definitions of the correctness of a requirement specification comes from Davis and Rauscher (1979) who state "correctness in requirements specifications is defined as the absence of incompleteness, redundancy, inconsistency, or ambiguity". Zowghi and Gervasi (2002; 2004) admit that "correctness by itself is a vague concept" but that "from a formal point of view, correctness is usually meant to be the combination of consistency and completeness. Consistency refers to situations where a specification contains no internal contradictions, whereas completeness refers to situations where a specification entails everything that is desired to hold in a certain context." We note the key difference between these definitions is that Zowghi and Gervasi do not consider "ambiguity" as a cause of incorrectness. We would also question Davis and Rauscher (1979) on whether redundancy (e.g. repeated statements of the same requirement) is a cause of incorrectness. Instead, it might be argued that redundancy is a risk to consistency in the advent of change. I.e. changing one requirement in a set of redundant requirements may create inconsistency within the specification.

Wilson, Rosenberg and Hyatt (1997) and Stokes (1991) agree that for a requirements specification to be correct it must accurately and precisely identify the individual conditions and limitations of all situations that the desired capability will encounter and it must also define the capability's proper response to those situations. In fact, this agrees with Zowghi and Gervasi (2002; 2004) in that "all situations that the desired capability will encounter" infers completeness; furthermore consistency and unambiguity are implied.

Zowghi and Gervasi (2004) suggest that "from a practical point of view, however, correctness can be more pragmatically defined as satisfaction of certain business goals. This indeed is the kind of correctness which is more relevant to the customer, whose goal in having a new system developed is to meet his overall business needs." Jackson (1995) has formalised this so-called "practical" definition of correctness. That is, a specification (S) is correct if when deployed in the domain (D), satisfies the customer's business goals (B).

$$S \cup D \models B$$

If we refer to the original requirements specification as S_o, where $S_o \cup D \models B$, and the optimal re-expression of S_o as S_r, then S_r is a correct re-expression of S_o iff $S_r \cup D \models B$. In other words, this thesis is about optimal re-expression of requirements, so the definition of correctness that we adopt, is that the re-expressed specification (S_r) is correct if it continues to express the business goal when deployed in the domain of S_o.

2.7. Summary and Research Opportunities

In Section 2.2 we summarised the process for engineering a system, we highlighted that requirements engineering is one phase of the systems engineering lifecycle. In Section 2.3 we defined what is meant by a requirement, we summarised the requirements engineering process, and identified the important quality factors of a requirement. We explained that specifying an unambiguous statement of requirements is critical to the success of a project, and to the safety of people and the environment. In Section 2.4 we presented a taxonomy of ambiguity types and looked at how ambiguity could be avoided in requirement specifications. We identified that controlled natural languages are one of three primary solutions to the ambiguity problem. In Section 2.5 we delved into the history and presented the most popular controlled natural languages. We discovered that historically, the CNL development process has been ad-hoc, time-consuming and preference-based – rather than structured, automated and optimised. We found a distinct lack of empirical evidence to support the claim that CNLs actually improved the quality of requirement specifications. In

Section 2.6 we investigated the quality of a requirements specification. We found that there is no universal quality model applicable to requirement specifications, but that there is some consistency across the CNL literature in terms of the quality factors described.

In summary, this literature review has revealed three important findings:

1. Historically, the CNL development process has been ad-hoc, time-consuming and preference-based,
2. There is no trade-off between important quality factors when developing a CNL,
3. There is a distinct lack of empirical evidence to support the claim that CNLs actually improved the quality of requirement specifications

Each finding presents a research opportunity that we discuss in the next chapter.

1. Instead of humans trying to pick *high-quality* words for the CNL lexicon, why not develop a metric that identifies the *optimal* word to express any concept, where optimal is defined in terms of key quality factors?
2. Instead of humans having to learn to write in the CNL, why not design a machine-automatable way of re-expressing natural English requirements?
3. Instead of speculating about quality improvements when writing in CNLs, why not re-express a specification using *optimal* words, and then empirically measure the change in key quality factors?

Chapter 3 Research Question and Methodology

3.1. Purpose

The purpose of this chapter is to convert the research opportunities identified from the previous chapter into a research question that we answer in this thesis. We also describe and justify the research methodology we have chosen to answer the research question.

3.2. Research Question

The literature review identified three key research opportunities.

1. Develop a metric that identifies the *optimal* word to express any concept, where optimal is defined in terms of key quality factors.
2. Design a machine-automatable way of re-expressing natural English requirements.
3. Re-express a specification using *optimal* words, and then empirically measure the change in key quality factors.

We have transformed these research opportunities into a single research question:

> *How can a requirements specification be automatically re-expressed in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality?*

3.2.1. Explanation of Research Question

The research question has been carefully constructed:

- The first word "*how*" is used to suggest that a design solution needs to be developed, which subsequently implies a research methodology such as design science.
- The word "*automatically*" has been used, since recall from Section 1.1, Nyberg, Mitamura and Huijsen (2003) state that "Perhaps an ideal situation for CL is for the machine to rewrite texts *automatically* into CL without changing the meaning expressed by the sentence". Also Davis and Rauscher (1979) stated "The ultimate situation is to have different requirements languages for different application areas". Thus we highly encourage the development of an *automated* solution. Furthermore, as discussed in Section 2.6.3, this explains why our research question excludes reference to *writability*,

i.e. the intention is to write naturally, but to *automatically* re-express in a way that is less ambiguous.

- The word "*re-expressed*" is used instead of "*constrained*" or "*CNL*" to demonstrate openness to solutions other than the traditional static CNL. That is, there may be more dynamic ways to constrain a requirements specification that are not bound by a fixed lexicon, perhaps involving the use of Natural Language Processing (NLP) techniques. This would also support the "*automatically*" keyword.
- The word "*significantly*" is used to mean statistical significance. This implies some kind of formal hypothesis test is required. More importantly, the word "*significantly*" implies some trade-off, perhaps between correctness, conventionality, and lexical ambiguity is required to optimise the re-expression.
- The phrase "*reduces its lexical ambiguity*" is referring to the overall reduction in lexical ambiguity of the requirements specification. We explained in Section 2.3.4 that the focus of this thesis is on ways to reduce ambiguity. We have further limited ambiguity to lexical ambiguity since grammatically related ambiguities have been more comprehensively addressed in other studies (refer Section 2.4.1.2 for syntactic, 2.4.1.3 for semantic, and 2.4.1.4 for pragmatic). We are looking for a statistically *significant* reduction in lexical ambiguity of the requirement specification. The evaluation section defines precisely how lexical ambiguity will be quantified for a specification.
- The phrase "*without significantly reducing its correctness or conventionality*" is referring to a statistically significant reduction in correctness and conventionality of the requirements specification as a result of the re-expression. Again, the evaluation section needs to define precisely how correctness and conventionality will be quantified for a specification. As explained in Section 2.6.1, we have bound the research question to *conventionality* rather than *readability* since it is difficult to build an automated system that predicts human understanding.

3.2.2. Relationship to Contributions

It is imperative that solving the research question produces the research contributions outlined in Section 1.5. Here we briefly explain how this occurs.

1. We first need to develop a model to describe the causal relationships between correctness, conventionality, and lexical ambiguity. This produces the first contribution: the Model.
2. Next we develop the method of optimal re-expression. In doing so we devise a new measure called *replaceability*, which is used to trade-off correctness, conventionality,

and lexical ambiguity and to identify the most optimal replacement. Additionally we make decisions about how the method is to be implemented, and in what sequence the activities occur. This produces the second contribution: the Method.

3. Next we develop the software prototype by developing software to implement the method. This produces the third contribution: the Software Prototype.
4. We then define a set of metrics and variables that are used to develop an evaluation method. The evaluation method explains how data are to be collected, analysed, and interpreted to determine the significance of any change in correctness, conventionality, and lexical ambiguity resulting from the optimal re-expression. This produces the fourth contribution: the Evaluation Method.
5. Finally, the execution of the Software Prototype on an existing industry specification provides empirical evidence to support/refute the claim that constraining the lexicon of a natural language improves the quality of requirement specifications. This produces the fifth contribution: Empirical Evidence.

3.3. Research Methodology

Hevner et al. (2004) find that two paradigms characterize much of the research in the Information Systems discipline: natural science and design science. The natural science paradigm seeks to develop and verify theories that explain or predict human or organizational behaviour. The design-science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts. March and Smith (1995) agree and describe the two fundamentally different research methodologies as follows:

1) Natural science is concerned with how and why things are. Natural science includes traditional research in physical, biological, social, and behavioural domains. Such research is aimed at understanding reality. Natural science is often viewed as consisting of two activities, discovery and justification. Discovery is the process of generating or proposing scientific claims (e.g., theories, laws). Justification includes activities by which such claims are tested for validity. Theories - deep, principled explanations of phenomena are the crowning achievements of natural science research.

2) Design science attempts to create things that serve human purposes. It is technology-oriented. Its products are assessed against criteria of value or utility - does it work? is

it an improvement? Rather than posing theories, design scientists strive to create models, methods, and implementations that are innovative and valuable. One issue that must be addressed in design-science research is differentiating routine design or system building from design research. The difference is in the nature of the problems and solutions. Routine design is the application of existing knowledge to organisational problems, such as constructing a financial or marketing information system using best practice artifacts including constructs, models, methods, and instantiations that are existing in the knowledge base. On the other hand, design-science research addresses important unsolved problems in unique or innovative ways or solves problems in more effective or efficient ways.

March and Smith (1995) explain that natural science aims at understanding and explaining phenomena; design science aims at developing ways to achieve human goals. Typically, design science creates the methods that natural scientists use. March and Smith (1995) argue that an appropriate framework for IT research lies in the interaction of design and natural sciences. IT research should be concerned both with utility, as a design science, and with theory, as a natural science. The theories must explain how and why IT systems work within their operating environments. They propose Figure 4 as a strategic combination of design and natural sciences as a framework appropriate to IT research.

Research Activities

Research Outputs		Build	Evaluate	Theorize	Justify
	Constructs				
	Model				
	Method				
	Instantiation				

Figure 4 - IT Research Framework (from (March & Smith 1995))

Figure 4 is a four-by-four framework that produces sixteen cells, each cell describing a viable research effort. The first dimension of the framework is based on design science research outputs or artifacts: constructs, models, methods, and instantiations. The second dimension is based on broad types of design science and natural science research activities: build, evaluate, theorize, and justify. Building and evaluating IT artifacts have design science intent. Theorizing and justifying have natural science intent. Research efforts often cover multiple cells – but are not expected to cover all cells.

In Section 3.2 we anticipated that answering the research question would involve the development of a model, method, instantiation, and an evaluation. Thus it is clear that this research should follow a design science research methodology. March and Smith's (1995) argument for a research framework that interacts with both design and natural sciences is valid in the context of our research – in that we build and evaluate artifacts *as well as* theorise and justify. We have therefore decided to adopt the research framework defined by March and Smith (1995) for our research.

3.3.1. Research Outputs

The research outputs are: constructs, models, methods, and instantiations. These research outputs are described by March and Smith (1995) and are included within this thesis as follows:

3.3.1.1. Constructs

Constructs or concepts form the vocabulary of a domain. They constitute a conceptualization used to describe problems within the domain and to specify their solutions. They form the specialized language and shared knowledge of a discipline or sub-discipline. Later, in Section 4.2, we define the constructs in terms of notation, universe, constants, variables, functions, and predicates applicable to this research.

3.3.1.2. Models

Models are a set of propositions or statements expressing relationships among constructs. A model can be viewed simply as a description, that is, as a representation of how things are. Later, in Section 4.3, we define the model in terms of propositions applicable to this research.

3.3.1.3. Methods

Methods are sets of steps, such as algorithms and guidelines that are used to perform a task. Methods are based on a set of underlying constructs (language) and a representation (model) of the solution space. Later, in Section 4.5, we define the individual methods applicable to this research. Essentially, we define a method for each function.

3.3.1.4. Instantiations

An instantiation is the realization of an artifact in its environment. Instantiations operationalize constructs, models, and methods. In much of the computer science literature it is realized that constructs, models, and methods that work on paper may not necessarily work in real world contexts. Consequently, instantiations provide the real proof. Later, in Section 4.5, we provide a cross-reference between the design methods and the software prototype.

3.3.2. Research Activities

The research activities are: build, evaluate, theorize, and justify. These research activities are described by March and Smith (1995) as follows:

3.3.2.1. Build

We build an artifact to perform a specific task. The basic question is, does it work? Building an artifact demonstrates feasibility. These artifacts then become the object of study. Later, in Section 4.2 we build constructs, Section 4.3 we build a model, Section 4.5 we build methods, and in Appendix F: Prototype Software we build a software prototype.

3.3.2.2. Evaluate

We evaluate artifacts to determine if we have achieved our research objectives. The basic question is, how well does our design work? Evaluation requires the development of metrics and the measurement of artifacts according to those metrics. Later, in Chapter 5 we explain that our evaluation method has been guided by the quantitative research methodology described by Leedy and Ormrod (2005). We have also been guided by the preliminary guidelines for empirical research prepared by Kitchenham et al. (2002), for example, we have considered key concepts:

1. hypotheses (Section 5.3.3);
2. sample and sampling method (Section 5.3.7);
3. variables (Section 5.2.2)
4. confounding effects (Section 5.3.1)
5. questionnaire design (Section 5.3.6)
6. reliability and validity (Section 5.3.10 and 5.3.11)

3.3.2.3. Theorize

Given an artifact whose performance has been evaluated, it is important to determine why and how the artifact worked or did not work within its environment. Such research applies natural science methods to IT artifacts. Theories explicate the characteristics of the artifact and its interaction with the environment that result in the observed performance. Whilst the concept of theorizing post-evaluation makes sense in principle, we argue that in practice, it is sometimes necessary to bound the build and evaluation phases by theorising early on in the process. This means our approach differs slightly from that explained by March and Smith (1995) in that we theorise, and then build, evaluate, and justify the original theory. Hence, we believe it is possible to theorise before we have collected metrics, by basing the theory on literature and practical experience. Later, in Section 4.4 we present our theories with respect to *ideal* versus *optimal* replacements.

3.3.2.4. Justify

Given a generalization or theory we must justify that explanation. That is, we must gather evidence to test the theory. The justification requires empirical and/or theoretical research to test the theories posed. The aim is to use the metrics collected during the evaluation phase to justify the base theories. In doing so, we have an increased level of confidence in our methods and instantiations. Later, in Section 5.4.6 we test and justify our theories.

3.4. Summary

In Section 3.2 we presented the research question, and in Section 3.3 we explained our reasons for adopting March and Smith's (1995) research framework. Whilst this chapter may be brief, it provides the focus and framework for the remainder of the thesis. In Chapter 4 we present the solution to the research question in terms of constructs, models, methods, and instantiations, and in Chapter 5 we evaluate the efficacy of the solution to the research question in terms of key quality factors: correctness, conventionality, and lexical ambiguity.

Chapter 4 Solution

4.1. Purpose

The purpose of this chapter is to present our solution to the research question within the research framework of March and Smith (1995). Specifically, this chapter defines the constructs, establishes the model, and designs the method of optimal re-expression. This chapter provides the basis for the implementation, and thus each method described in this chapter has been cross-referenced to the corresponding function in the source code.

4.2. Constructs

March and Smith (1995) explain that constructs or concepts form the vocabulary of a domain. They constitute a conceptualization used to describe problems within the domain and to specify their solutions. They form the specialized language and shared knowledge of a discipline or sub-discipline.

4.2.1. Notation

Table 2 provides the notation used to define the model and method.

Table 2 - Notation

Symbol	Definition	Explanation
Connectives		
x	assertion	"x is true"
$\neg x$	negation	"x is not true"
$x \vee y$	disjunction	"either x is true, or y is true, or both"
$x \wedge y$	conjunction	"both x and y are true"
$x \rightarrow y$	implication	"if x is true, then y is true"
$x \leftrightarrow y$	equivalence	"x and y are either both true or both false"
Quantifiers		
$\forall_x$	universal quantifier	"for all p", "for each p", "for every p" in the universe
$\exists_y$	existential quantifier	"there exists a p", "there is at least one p" in the universe"
Set Theory		
{ }	set	collection of distinct objects (no duplicates)
[ ]	collection	collection of objects (allows duplicates)
$x \in A$	inclusion	"x is a member of A"
$x \notin A$	exclusion	"x is not a member of A"
$A \subseteq B$	subset	"A is a subset of B"
$\lvert A \rvert$	cardinality	number of elements in set (or collection) A

4.2.2. Universe

The universe (also called the universe of discourse) bounds the values attributable to the variables. The propositions in the predicate logic are statements on objects of a universe. The universe is thus the domain of the individual variables. We define the universe of *words* as the *set of all Natural English* words, which may be a superset of those words defined in a dictionary, e.g. by including proper nouns specific to an application domain.

4.2.3. Constants

Constant	Definition
S_o	[W_{o1}, W_{o2},... W_{on}] words comprising the original specification
D	{W_{d1}, W_{d2},... W_{dn}} words comprising the dictionary

4.2.4. Variables

Variable	Definition
W_o	$W_o \in S_o$ i.e. each original word (W_o) exists in original spec (S_o)
W_r	$W_r \in D$ i.e. each replacement word (W_r) exists in the dictionary (D)
S_r	[W_{r1}, W_{r2},... W_{rn}] words comprising the re-expressed specification
similarity_threshold	similarity(W_1, W_2) $\geq$ similarity_threshold $\rightarrow$ W_1 is similar to W_2

4.2.5. Functions

Function	Returns
context(W)	[words before W] W [words after W]
meaning(W)	meaning of (word) W in context(W)
senses(W)	possible meanings of (word) W
frequency(W)	number of occurrences of (word) W
width(W)	number of coordinates of (word) W
similarity(W_1, W_2)	the degree to which (word) W_1 is similar to (word) W_2
replaceability(W_1, W_2)	the degree to which (word) W_1 can replace (word) W_2

4.2.6. Predicates

Predicate	Definition
similar(W_r,W_o)	(word) W_r is similar to (word) W_o
correct(W_r, context(W_o))	(word) W_r is correct in context(W_o)
conventional(W_r, context(W_o))	(word) W_r is conventional in context(W_o)
unambiguous(W_r, context(W_o))	(word) W_r is unambiguous in context(W_o)
ideal(W_r, context(W_o))	(word) W_r is ideal in context(W_o)
optimal(W_r, context(W_o))	(word) W_r is optimal in context(W_o)

4.3. Model

March and Smith (1995) define a model as a set of propositions or statements expressing relationships among constructs. A model can be viewed simply as a description, that is, as a representation of how things are; it is not a model of our solution, but is a descriptive model of our world.

4.3.1. Similar(W_r, W_o)

$\forall W_o . \exists W_r$: similarity(meaning(W_o), senses(W_r)) $\geq$ similarity_threshold $\rightarrow$ similar(W_r,W_o)

For each original word (W_o) in the original specification (S_o) there exists a set of similar words $\{W_r\}$, such that (W_r) is similar to the original word (W_o) when the similarity between the disambiguated meaning of (W_o) and any dictionary-defined sense of (W_r) is greater than some threshold. We refer to this variable threshold as the *similarity_threshold.*

4.3.2. Correct(W_r, Context(W_o))

$\forall W_o . \exists W_r$: Max(similarity_threshold) $\wedge$ similar(W_r,W_o) $\rightarrow$ correct(W_r, context(W_o))

For each original word (W_o) in the original specification (S_o) there exists a set of correct replacement words $\{W_r\}$, where (W_r) is considered correct in the context of the original word (W_o) when (W_r) is similar to (W_o) and when the similarity_threshold maximised.

Figure 5 is a UML class diagram that has been extracted from (Boyd, Zowghi & Farroukh 2005) that we use to illustrate the concept of correctness. A requirement is an expression that is comprised of *n* words, and conveys *M* meanings, where M_i is the authors' intended meaning and if M > 1 then the Requirement Expression is ambiguous. If we can replace a single word (W_o) within the Requirement Expression with another word (W_r) such that the stakeholders continue to interpret the authors' intended meaning (M_i), then W_r is a correct replacement for W_o. For this to occur, we expect that the similarity threshold would need to be maximised in order for the predicate *similar*(W_r, W_o) to be true only when W_r and W_o are very similar (e.g. synonym, direct hypernym, etc). Note that other aspects of Figure 5 are not relevant to this discussion on correctness, and so we do not elaborate any further on the detail.

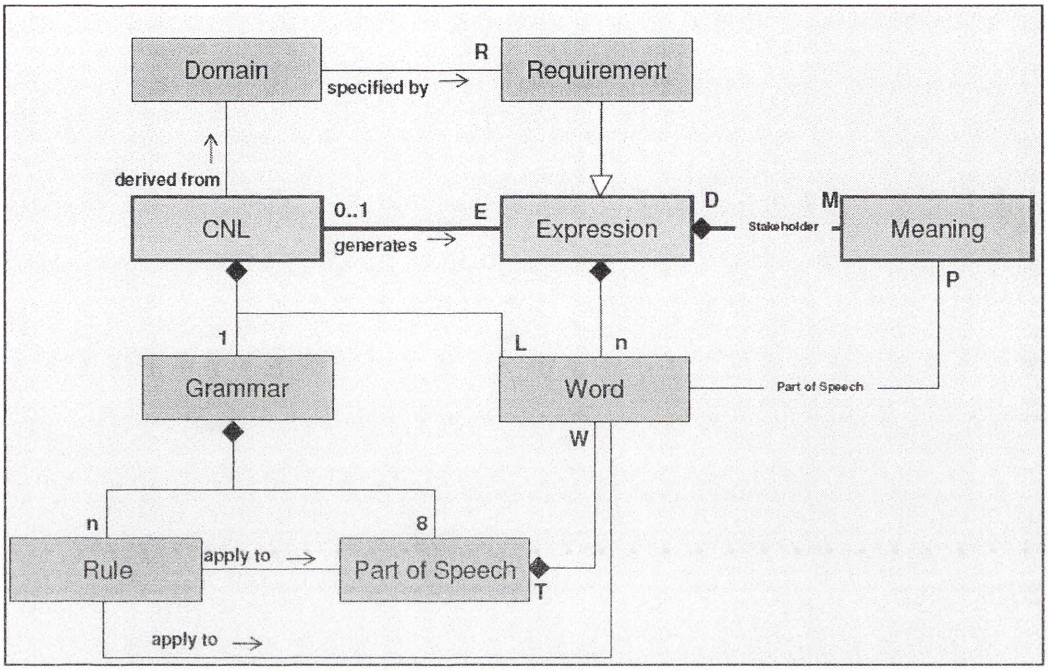

Figure 5 - Semantic Expressiveness (from (Boyd, Zowghi & Farroukh 2005))

4.3.3. Conventional(W_r, Context(W_o))

$\forall W_o.\exists W_r$: correct(W_r,W_o) $\wedge$ frequency(W_r) $\geq$ frequency(W_o) $\wedge$ width(W_r) $\geq$ width(W_o) $\rightarrow$ conventional(W_r, context(W_o))

For each original word (W_o) in the original specification (S_o) there exists a set of conventional replacement words $\{W_r\}$, where (W_r) is conventional in the context of the original word (W_o) when:

- (W_r) correctly expresses (W_o); and
- (W_r) is more frequently occurring than (W_o) in (S_o); and
- (W_r) is defined on a more richly lexicalised level than (W_o) in (D).

This is based on the findings of Fellbaum (1990) and Green (2006) that are described in Section 2.6.2.

4.3.4. Unambiguous(W_r, Context(W_o))

$\forall W_o.\exists W_r$: correct(W_r,W_o) $\wedge$ $|$senses(W_r)$|$ = 1 $\rightarrow$ unambiguous(W_r, context(W_o))

For each original word (W_o) in the original specification (S_o) there exists a set of unambiguous replacement words $\{W_r\}$, where (W_r) is an unambiguous replacement for the original word (W_o) when:

- (W_r) correctly expresses (W_o); and
- (W_r) is monosemous (has only one meaning).

This is based on the definition of lexical ambiguity provided by Berry and Kamsties (2003) in Section 2.4.1.1

4.4. Theories

March and Smith (1995) explain that typically theories explicate the characteristics of the artifact and its interaction with the environment that result in the *observed* performance, however, as explained in Section 3.3.2.3 we argue that in practice, it is sometimes necessary to bound the build and evaluation phases by theorising early on in the process. This means our approach differs slightly from that explained by March and Smith (1995) in that we theorise, and then build, evaluate, and finally, justify our theories. Hence, we theorise based on literature and our practical experience, and we subsequently confirm the theories by collecting and analysing relevant data.

Our theories are conjectures: i.e. statements or ideas that are unproven, but are thought to be true. Each conjecture below is stated as a satisfiable well-formed formula (wff), in that we expect the constants and variables can be defined in such a way as to yield an interpretation that makes each wff true.

4.4.1. Ideal(W_r, Context(W_o))

$$\forall W_o . \exists W_r : \text{correct}(W_r, W_o)) \wedge \text{conventional}(W_r, \text{context}(W_o)) \wedge \text{unambiguous}(W_r, \text{context}(W_o)) \rightarrow \text{Ideal}(W_r, \text{context}(W_o))$$

For each original word (W_o) in the original specification (S_o) there exists a set of *ideal* replacement words $\{W_r\}$, where (W_r) is ideal in the context of the original word (W_o) when:

- (W_r) correctly expresses (W_o);
- (W_r) is conventional in the context of (W_o);
- (W_r) is unambiguous in the context of (W_o);

Note that due to the cause-and-effect relationships that exist between the quality factors correctness, conventionality, and lexical ambiguity (to be explained later in Section 4.5.6), we do not always expect to find *ideal* replacements (i.e. Ideal(W_r, context(W_o)) may return an empty set). This is because it won't often be the case that we can find a correct, conventional, and unambiguous alternate way of expressing the original word (W_o). Thus we theorise about *optimal* replacements in the next section.

4.4.2. Optimal(W_r, Context(W_o))

$\forall W_o . \exists W_r$: optimal(W_r, context(W_o)) $\rightarrow$ {$\forall W_r$: argmax(replaceability(W_r, W_o))}

For each original word (W_o) in the original specification (S_o), there exists an optimal replacement word (W_r), where (W_r) is optimal in the context of the original word (W_o) when the replacement word (W_r) maximises replaceability(W_r,W_o). Whilst the concept of replaceability is yet to be defined (later in Section 4.5.6), it is useful to highlight the fact that it may not always be possible to find an *ideal* replacement, and so by maximising replaceability, we expect to find the most optimal replacement. In other words, maximising replaceability should result in a replacement word (W_r) that significantly reduces the lexical ambiguity of the requirement without significantly reducing its correctness or conventionality.

4.5. Method

The Method defines how we intend to calculate each predicate and function so that we can subsequently instantiate the method, and interpret and confirm the theories through an evaluation. That is, the Method defined in this chapter is specific to our solution, meaning that other researchers could potentially calculate the predicates and functions differently. Note that since the predicates in Section 4.2.6 were each defined in terms of the functions of Section 4.2.5, this section focuses solely on the method applicable to each function (i.e. the predicates are derived from the functions).

This section also defines the sequence of activities in converting an original word (W_o) into a replacement word (W_r). The sequence of activities is later defined in terms of a sequence diagram that clarifies how the re-expression occurs in a way that W_r is a:

- Correct replacement for W_o;
- Conventional replacement for W_o; and
- Less ambiguous replacement for W_o.

4.5.1. Meaning(W)

Function	Returns
meaning(W)	meaning of (word) W

4.5.1.1. Background

Word Sense Disambiguation (WSD) involves the association of a given word in a text or discourse with a definition or meaning (sense) which is distinguishable from other meanings potentially attributable to that word (Ide & Véronis 1998). This is done by looking at the context of the word's use, i.e. "You shall know a word by the company it keeps" (Firth 1957).

Manning and Schütze (1999) explain a word is assumed to have a finite number of discrete senses, often given by a dictionary, thesaurus, or other reference source, and the task is to make a forced choice between these senses for the meaning of each usage of the word, based on the context of use. Context is the only means to identify the meaning of an ambiguous word. Therefore, all work on sense disambiguation relies on the context of the target word to provide information to be used for its disambiguation. Therefore WSD is the process of discriminating the meaning of the original word (W_o) in the context of specification (S_o) using the sense definitions of dictionary D.

The task therefore necessarily involves two steps: (1) the determination of all the different senses for every word relevant to the text or discourse under consideration (e.g. a list of senses such as those found in everyday dictionaries); and (2), the assignment of words to senses. This is accomplished by reliance on two major sources of information: the context of the word to be disambiguated, in the broad sense: this includes information contained within the text or discourse in which the word appears, and external knowledge sources, including lexical, encyclopedic, etc. Ide and Véronis (1998) explain that there are two mainstream approaches to WSD: 1) Supervised Disambiguation, and 2) Dictionary-Based Disambiguation.

1) *Supervised disambiguation* is disambiguation based on a labelled training set (e.g. Brown et al. 1991). Manual sense-tagging to create a labelled training set is extremely costly, and at present very few sense-tagged corpora are available: the *Linguistic Data Consortium* distributes a corpus of approximately 200,000 sentences from the *Brown Corpus* and the *Wall Street Journal* in which all occurrences of 191 words are hand-tagged with their *WordNet* senses (Neff et al. 1993). Also, the *Cognitive Science*

Laboratory at Princeton has undertaken the hand-tagging of 1000 words from the *Brown Corpus* with *WordNet* senses (Miller et al. 1993). Sense-tagged corpora that are available are far smaller than those typically used with statistical methods. The problem of data sparseness, which is common for much corpus-based work, is especially severe for work in WSD. First, enormous amounts of text are required to ensure that all senses of a polysemous word are represented, given the vast disparity in frequency among senses. In addition, the many possible co-occurrences for a given polysemous word are unlikely to be found in even a very large corpus, or they occur too infrequently to be significant (Ide & Véronis 1998).

2) *Dictionary-based disambiguation* methods calculate a score based on the number of stemmed words that are shared by the sense definition and the context, e.g. (Lesk 1986). There have been several experiments using pre-existing lexical knowledge resources for WSD, e.g. Cowie, Guthrie and Guthrie (1992) using LDOCE; Yarowsky (1992) using Roget's International Thesaurus; and Resnik (1995), Richardson and Smeaton (1994), Sussna (1993), Voorhees, Claudia and Geoffrey (1995) using WordNet. The success of dictionary-based WSD is adversely affected by unknown words such as proper nouns. E.g. *Navratilova* is an excellent indicator of the category *sports* – yet we would probably not find *Navratilova* in a general dictionary or thesaurus. Dictionary-based WSD often uses the definition of the words in the context window. Importantly, words that do not occur in the dictionary (e.g. proper nouns) are ignored (Ide & Véronis 1998).

We immediately rule-out the *supervised disambiguation* method since the few sense-tagged corpora that exist are not reflective of the technical domain of requirements. Even the *dictionary-based disambiguation* method is problematic. The requirements for a bespoke or legacy system would typically specify some combination of *new things*; *new concepts*; *new features, new processes,* and therefore we expect unknown words to be a common occurrence in NL requirements (e.g. new system name, new person role, new organisation, new function, etc.). Lexical resources have to be updated to keep pace with these changes, and evolving static dictionaries to include domain-specific terms can be very expensive. Neff et al. (1993) estimates it takes on average half-an-hour to create a single lexical entry from scratch. Fundamentally, a sense cannot be assigned to a word if the word does not exist in the chosen dictionary. Furthermore, since each word provides the context for disambiguating surrounding words, the effect of an unknown word can propagate to the incorrect disambiguation of the surrounding words.

4.5.1.2. Method

We use a *dictionary-based disambiguation* method and assume the meaning of each original word (W_o) in the specification (S_o) is defined in the dictionary (D).

$\forall W_o : W_o \in S_o \rightarrow meaning(W_o) \in D$

In other words, for each word (W_o), if W_o is a word that exists in specification (S_o), then we assume the meaning of W_o is defined in dictionary (D). Thus we need to ensure the dictionary chosen supports this assumption, otherwise we risk unknown words being ignored (Ide & Véronis 1998).

4.5.1.3. Implementation

People have been writing programs for automatic disambiguation for around 50 years and still the validity of this task remains in doubt (Ide & Véronis 1998). The difficulty of WSD stems from the fact that although computers are best at following fixed rules, it is impossible to create a set of fixed rules that can accurately disambiguate any word in any context. The problem of WSD has been described as AI-complete, that is, a problem which can be solved only by first resolving all the difficult problems in artificial intelligence, such as the representation of common sense and encyclopedic knowledge (Manning & Schütze 1999). Ide and Véronis (1998) agree that there has been relatively little progress made in WSD development in nearly 50 years. Even though much recent work cites results at the 90% level or better, these studies typically involve a very few words, most often only nouns, and very frequently concern very broad sense distinctions. With sufficiently greater resources and enhanced statistical methods at their disposal, researchers in the 1990's have obviously improved on earlier results, but it appears that we may have reached near the limit of what can be achieved in the current framework.

We informally tested two freely available auto-WSD tools: WordNet::SenseRelate (Pedersen, Patwardhan & Michelizzi 2004) and Sense Learner 2.0 (Mihalcea & Faruque 2004) with both tools failing to accurately disambiguate the requirement text in the majority of cases. The disappointing results are believed to occur due to the presence of words that are unknown to the dictionary that is used in *dictionary-based disambiguation*. The following example illustrates the point with "rolling stock" as the unknown word.

"The rolling stock shall have comfortable seating"

[NP<rel="SBJ" of="s1_1"> **The**DT<lem="the"> **rolling**VBG<lem="roll"> **stock**NN<lem="stock">]

[VP<id="s1_1"> **shall**MD<lem="shall"> **have**VB<lem="have">]

[NP<rel="OBJ" of="s1_1"> **comfortable**JJ<lem="comfortable"> **seating**NN<lem="seating">]

Whilst the natural language understanding system may have correctly identified "have" as a verb, "comfortable" as an adjective, and "seating" as a noun; the phrase "rolling stock" has been *incorrectly* tagged as "rolling" (verb gerund) and "stock" (noun).

Interestingly, had "rolling stock" been expressed "rolling-stock", the shallow parser would have *incorrectly* considered it an adjective.

[NP<rel="SBJ" of="s1_1"> **The**DT<lem="the"> **rolling-stock**JJ<lem="rolling-stock">]

[VP<id="s1_1"> **shall**MD<lem="shall"> **have**VB<lem="have">]

[NP<rel="OBJ" of="s1_1"> **comfortable**JJ<lem="comfortable"> **seating**NN<lem="seating">]

Furthermore, had "rolling stock" been *incorrectly* spelled "rollingstock" the shallow parser would have *correctly* considered it a noun.

[NP<rel="SBJ" of="s1_1"> **The**DT<lem="the"> **rollingstock**NN<lem="rollingstock">]

[VP<id="s1_1"> **shall**MD<lem="shall"> **have**VB<lem="have">]

[NP<rel="OBJ" of="s1_1"> **comfortable**JJ<lem="comfortable"> **seating**NN<lem="seating">]

This simple example illustrates the type of POS tagging errors that occur when unknown words (e.g. "rolling stock") are encountered by the natural language understanding system. Such tagging errors can then snowball into larger problems if humans are not engaged in the process. For example if the "rolling" part of"rolling stock" is identified as a verb, then the WSD tool will attempt to assign a verb sense. The base verb, "roll" might then be automatically replaced by a synonym verb such as "spin", leading to the eventual replacement of "rolling stock" with "spinning stock". Therefore, unknown words in any part of speech can infect the optimal re-expression process. Auto-WSD tools rely on contextual information to make a probabilistic determination on the sense of each word. That is, to disambiguate a content word that has *N* possible senses, the auto-WSD tool would determine a probability for each *n* sense – given the surrounding content words. When the surrounding content words are unknown words (e.g. proper nouns), the WSD tool often fails to accurately disambiguate the requirement text, which may then escalate into an incorrect replacement.

On the other hand, human beings are particularly good at WSD. For example, given the sentence *The bark of the dog was very loud*, it is immediately obvious to a human that the word *bark* is referring to the sound made by a dog. In the sentence *The dog scratched its back on the bark of the tree* we know the word bark here means tree covering (Ide & Véronis 1998). Choueka and Lusignan (1985) show that humans do surprisingly well at sense discrimination if only a few words of adjacent context are shown – giving more context contributes little to human disambiguation performance. Human performance is typically considered the upper bound for WSD (Manning & Schütze 1999) and therefore we have decided to human-disambiguate each W_o in S_o in order to control WSD as a confounding variable in our optimal re-expression process. In other words, since unknown words tend to be prevalent in requirements, and since unknown words tend to result in auto-WSD tools making errors in the WSD, and since errors in WSD degrade the optimal re-expression process, we have decided to minimise this confounding effect by using humans to conduct the WSD of requirements.

Additionally, we have decided to use Web WordNet (2003) as the dictionary (D) of senses for Word Sense Disambiguation. We understand that WordNet is at present the best known and the most widely utilized on-line lexical reference system. WordNet's design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, and adjectives are organized into synonym sets (synsets), each representing one underlying lexical concept (Miller 1990). Synsets are then associated with other synsets sets via lexical relationships (e.g. synonymy, antonymy, hyponymy (is a), meronymy (part of), and morphological relationships). We should note that we also considered the option of using Wikipedia: The biggest multilingual free-content encyclopedia on the Internet (Wikimedia Foundation 2008). Wikipedia contains over 7 million articles in over 200 languages, and is growing every day. The one key advantage Wikipedia has over WordNet is that it is growing every day with new content words. Wikipedia contains many domain specific words that would not normally be included in a standard dictionary (e.g. WordNet). Our primary reason for choosing WordNet over Wikipedia was the availability of relevant software tools for automatically working with the dictionary (e.g. WordNet has freely available tools for PoS tagging, Similarity calculations, etc).

We acknowledge that WordNet is not a perfect resource for word sense disambiguation. The most frequently cited problem is the fine-grainedness of WordNet's sense distinctions, which are often well beyond what may be needed in many language processing applications.

However, we would prefer *too many* sense definitions rather than *not enough* sense definitions, since we are assuming meaning(W_o) $\in$ D. Since we have decided to manually disambiguate word senses, the fine-grainedness of WordNet's sense distinctions is expected to become more of an efficiency problem, with the human's having to read many sense definitions before making his or her choice, than an accuracy problem, since he or her should still choose the correct sense once the selection is made..

4.5.2. Senses(W)

Function	Returns
senses(W)	senses of (word) W

4.5.2.1. Background

Whilst there is psychological validity to the notion of senses (Jorgensen 1990; Simpson & Burgess 1989), lexicographers themselves are well aware of the lack of agreement on senses and sense divisions. Manning and Schütze (1999) find that dictionaries often differ greatly in the number and kind of senses they list. One of the foremost problems for sense discrimination is to determine the appropriate degree of sense granularity. Several authors, e.g. Slator and Wilks (1987) have remarked that the sense divisions one finds in dictionaries are often too fine for the purposes of WSD work, therefore overcomplicating the disambiguation process. Additionally, there are often obscure senses that would never lead to the realisation of lexical ambiguity in a particular domain. In other words, whilst a word may have many possible meanings in a dictionary – we are interested in the number of meanings actually exhibited by the word in a usage context, such as might be used in a requirements specification. That is, *actual* sense usage is more interesting that *potential* sense usage.

4.5.2.2. Method

$\forall W : \{senses(W)\} = \{\forall W_s : W_s \in [S_o] : (W_s = W) \rightarrow meaning(W_s)\}$

That is, for any word (W), the set of senses of word (W) is the set of meanings of word (W) that actually occur in the original specification (S_o). I.e. the *actual* sense usage: senses(W) may be a subset of the *potential* senses of word (W) as defined by dictionary (D).

4.5.2.3. Implementation

Refer to the `measureLexicalFeatures` function of Appendix F: Prototype Software for the C++ implementation of senses(W).

4.5.3. Frequency(W)

Function	Returns
frequency(W)	number of occurrences of (word) W

4.5.3.1. Background

Green's (2006) linguistically-oriented heuristics suggest that frequently used words are more conventional than infrequently used words. For example, *petrol* conventionally refers to *gasoline* in the UK, whereas *gas* conventionally refers to *gasoline* in the US. We intend to use frequency information to help ensure that conventional terms are not replaced by unconventional terms. Frequency of word occurrence requires some consideration. For instance, does it mean how frequently the disambiguated *word meaning* occurs or how frequently the *word term* regardless of the word meaning occurs? Furthermore, which corpus should be used to count word occurrences? (e.g. so that *gas* does not replace *petrol* in UK specifications).

4.5.3.2. Method

$\forall W$: frequency(W) = $\lvert [\forall W_s : W_s \in [S_o] : (W_s = W) \wedge (\text{meaning}(W_s) = \text{meaning}(W))] \rvert$

That is, for any word (W), the frequency of occurrence of word (W) is defined as the cardinality of the collection of words (W_s) that occur in specification (S_o) and have the same word term and the same word meaning as word (W). In other words, we have decided that frequency is a count of disambiguated *word meanings* as they occur within the original specification (S_o) corpus. Therefore $\forall W_o$: frequency(W_o) ≥ 1.

4.5.3.3. Implementation

Refer to the `measureLexicalFeatures` function of Appendix F: Prototype Software for the C++ implementation of frequency(W).

4.5.4. Width(W)

Function	Returns
width(W)	number of coordinates of (word) W

4.5.4.1. Background

Fellbaum (1990) states that *L0* is the most richly lexicalised level since the *width* of *L0* is greater than the *width* of both the direct hypernym and direct hyponym. Recall from Section 2.6.1 that replacements are more likely to be conventional if the replacement word W_r exists on a more richly lexicalised level of the semantic hierarchy than the original word W_o. By counting the number of coordinate terms of W_o and W_r, we can determine whether W_r is more (or less) richly lexicalised than W_o in the hierarchical dictionary (D). As with senses(W), whilst a word may theoretically have many coordinate terms defined in a hierarchical dictionary, we are interested in the number of coordinate terms that actually occur in a usage context. Coordinate term usage information provides us with a more appropriate measure of width.

4.5.4.2. Method

$\forall W: width(W) = \lvert\{\forall W_s : W_s \in [S_o]: hypernym(W_s,D) = hypernym(W,D)\}\rvert$

That is, for any word (W), the width of word (W) is defined as the cardinality of the set of words (W_s) that occur in specification (S_o) and have the same hypernym as word (W) defined in the hierarchical dictionary (D). In other words, we have again decided to use the original specification (S_o) as the corpus of words eligible to count as coordinate terms – rather than the full extent of coordinate terms defined in the dictionary (D). Therefore $\forall W_o$: $W_o \in S_o$: width(W_o) ≥ 1.

There is an implicit limitation here that *L0* cannot occur at the top-level of a hierarchy since unique beginners do not have any hypernym and therefore no coordinate concepts. Whilst in most cases this limitation will not manifest itself, there are some *L0* terms (such as the verb to *be*) which exist at the very top of their part of speech hierarchies. In these cases, the *L0* term may be replaced with a less conventional term on a more richly lexicalised level of the hierarchy.

4.5.4.3. Implementation

WordNet comes with an Application Programmers Interface (API) providing a variety of ANSI C functions that can query the WordNet database. For example, the *findtheinfo_ds* function returns a pointer to the disambiguated word in WordNet. The *traceptrs_ds* function returns the target of a lexical relationship (e.g. synonym, hyponym, hypernym, etc). Refer to the `measureLexicalFeatures` function of Appendix F: Prototype Software for the C++ implementation of width(W).

4.5.5. Similarity(W_1, W_2)

Function	Returns
similarity(W_1,W_2)	the degree to which (word) W_1 can be substituted for (word) W_2

4.5.5.1. Background

Miller and Charles (1998) claim that *semantic similarity* provides some measure as to the degree of *contextual interchangeability*, or the degree to which one word can be substituted by another in context. This is because measures of similarity quantify how much two concepts are alike, based on information contained in an is–a hierarchy. For example, an automobile might be considered more like a boat than a tree, if automobile and boat share vehicle as a common ancestor in an is–a hierarchy (Pedersen, Patwardhan & Michelizzi 2004).

Agirre and Rigau (1996) explain there are a variety of different methods for measuring the similarity between two concepts:

- The length of the shortest path that connects W_1 and W_2,
- The depth in the hierarchy: concepts in a deeper part of the hierarchy should be ranked closer,
- The density of concepts in the hierarchy: concepts in a dense part of the hierarchy are relatively closer than those in a more sparse region, and/or
- The information content of the Lowest Common Subsumer (LCS) of two concepts. The LCS is the most specific concept that is an ancestor of both W_1 and W_2.

Pedersen, Patwardhan and Michelizzi (2004) categorise similarity measures as either path-based, or IC-based. Path-based measures determine similarity by evaluating the distances between the two concepts being compared. The shorter the path from one concept to another concept, the more similar the two concepts are. Information Content (IC) for concepts is determined from a sense-tagged corpus (such as SemCor, Brown, Penn Treebank, British National Corpus, etc). Since a sense-tagged corpus is required, and since existing sense tagged corpora will not be reflective of domain-specific requirement specifications, we have discounted IC-based methods in favour of path-based methods.

Is–a relations generally do not cross part of speech boundaries, so path-based similarity measures are typically limited to making judgments between noun pairs (e.g., *cat* and *dog*), verb pairs (e.g., *run* and *walk*), etc. However, this is desirable since we would not want replacement word (W_r) to be of a different part of speech to the original word (W_o).

4.5.5.2. Method

Based on the discussion above, we have decided to exclude the IC-based similarity measures as an option in our method. We have also decided to exclude the path-based measure Leacock and Chodorow (Miller 1990) because it is not normalised to unity. We consider the following similarity measures relevant to our method.

Similarity Measure	Type	Calculation
Path Length (Miller 1990)	Path	$sim(W_1, W_2) = 1/ length(W_1, W_2)$
Wu and Palmer (Wu & Palmer 1994)	Path	$sim(W_1, W_2) = 2*depth(lcs) / (depth(W_1) + depth(W_2))$

Where,

- $sim(W_1, W_2)$ is the similarity between W_1 and W_2.
- $length(W_1, W_2)$ is the shortest path length (i.e., having minimum number of nodes) between W_1 and W_2.
- lcs is the most specific concept that is an ancestor of both W_1 and W_2.

4.5.5.3. Implementation

WordNet is particularly well suited for path based similarity measures, since its parts of speech are organised into hierarchies of is–a relations. Pedersen, Patwardhan and Michelizzi (2004) have developed WordNet::Similarity, a set of Perl modules that implement the selected similarity measures of Path Length, and Wu and Palmer (1994) amongst many others.

Quantifying the similarity between two words can be achieved in two ways: 1) develop a Perl module that calls the WordNet::QueryData and WordNet::Similarity modules, or 2) use the command line interface from the command prompt. Since the researchers are not familiar programming with Perl, and since Pedersen, Patwardhan and Michelizzi (2004) have specifically developed WordNet::Similarity with a command line interface for ease of use, and since the command line interface enables the input of a text file, and enables the output to be re-directed to another text file, we have decided to implement the function $similarity(W_1,W_2)$ via a system call to WordNet::Similarity. The command line interface allows the user to request the similarity between a stated pair of words, or between all pairs of words in a text file. For example, assuming we have FILE.txt containing:

- cat#n#1 dog#n#1
- cat#n#1 bird#n#1

Then perl similarity.pl –type WordNet::Similarity::wup –file FILE.txt returns:

- cat#n#1 dog#n#1 0.85
- cat#n#1 bird#n#1 0.72

Indicating that when executing the similarity measure of Wu and Palmer (*wup*) within the semantic network of WordNet, a cat is considered more similar to a dog than it is to a bird.

Performance consideration: If there are N unique words for a particular part of speech from specification (S_o), then there are N^2 unique combinations of words. Similarity is bi-directional: i.e. similarity(W_1,W_2) = similarity(W_2,W_1), and the diagonal of the N^2 matrix will compare the word with itself, so there are (N^2 - N) / 2 unique word combinations that need to be calculated by WordNet::Similarity. Figure 6 shows that a one order of magnitude increase in N (e.g. from 100 to 1000 words) results in a quadratic increase in the number of calculations required by WordNet::Similarity (e.g. from 5000 to 500,000). We must be cognisant of the number of comparisons that need to be performed to enable us to develop an efficient software solution. In particular, it would be extremely inefficient for us to calculate the similarity between each pair of words using the pair-wise command line system call. It is much more efficient for us to use a single command line system call that reads input from a predefined text file as described above.

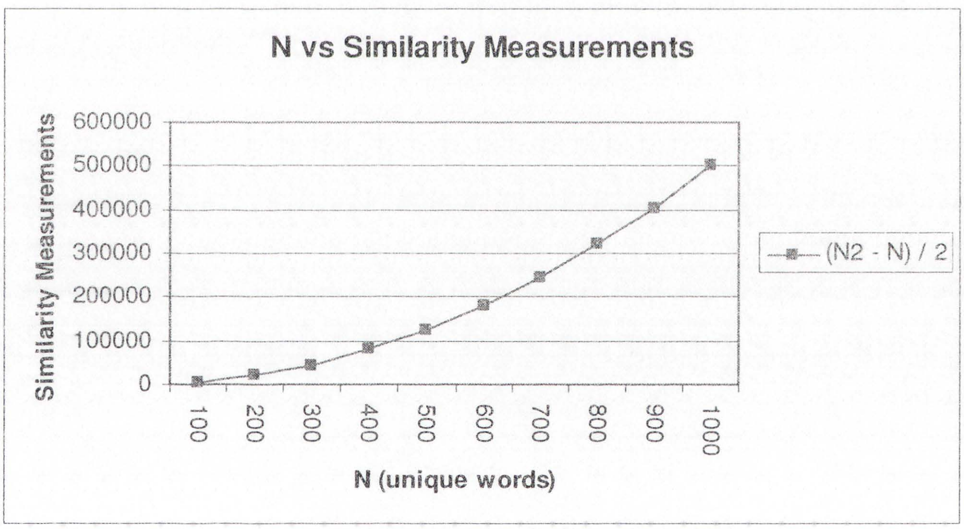

Figure 6 - N vs. Similarity Measurements

4.5.6. Replaceability(W_1, W_2)

Function	Returns
replaceability(W_1, W_2)	the degree to which (word) W_1 can replace (word) W_2

4.5.6.1. Background

Section 4.4.1 conjectured that an ideal replacement is one which is correct, conventional, and lexically unambiguous. We explained that it is often not possible to find an ideal replacement so we conjectured in Section 4.4.2 that an optimal replacement would be the W_r that maximizes the *replaceability* between W_o and W_r. However, we did not explain how replaceability(W_1, W_2) would be calculated – that is the subject of this section.

Given replaceability(W_r,W_o) is necessarily a function of correct(W_r,context(W_o))), conventional(W_r,context(W_o)), and unambiguous(W_r,context(W_o)), we know that replaceability(W_r,W_o) must be increased when:

a) similarity(meaning(W_o), senses(W_r)) $\geq$ similarity_threshold;
b) $|$senses(W_r)$|$ / $|$senses(W_o)$|$ < 1;
c) frequency(W_r) / frequency(W_o) ≥ 1; and
d) width(W_r) / width(W_o) ≥ 1.

Reducing the similarity threshold from sim_thr_1 to sim_thr_2 (where $sim\_thr_1 > sim\_thr_2$) is expected to expand the set of potential replacement words (i.e. $\{W_{r1}\} \subseteq \{W_{r2}\}$). As a result, we expect the following effects:

- $\{W_{r1}\} \subseteq \{W_{r2}\} \rightarrow \forall W_{r1}.\exists W_{r2}$: frequency(W_{r2}) $\geq$ frequency(W_{r1})
- $\{W_{r1}\} \subseteq \{W_{r2}\} \rightarrow \forall W_{r1}.\exists W_{r2}$: width(W_{r2}) $\geq$ width(W_{r1})
- $\{W_{r1}\} \subseteq \{W_{r2}\} \rightarrow \forall W_{r1}.\exists W_{r2}$: $|$senses(W_{r2})$|$ $\leq$ $|$senses(W_{r1})$|$

We expect these effects because increasing the set of potential replacements increases the likelihood of finding 1) a replacement that is more frequently occurring in specification (S_o), 2) a replacement that is defined on a more richly lexicalised level in dictionary (D), and 3) a replacement that has fewer sense usages in specification (S_o).

Furthermore, when the normalised similarity threshold < 1, W_r may not be correct since recall from Section 4.3.2 that $\forall W_o.\exists W_r$: Max(similarity_threshold) $\wedge$ similar(W_r,W_o) $\rightarrow$ correct(W_r, context(W_o)). In other words, if the similarity_threshold is not maximised (because it is < 1) then W_r may not be correct in the context of W_o.

This expected reduction in correctness means we cannot presume that replaceability will be increased by reducing the similarity threshold. That is, we have a trade-off to deal with:

- $\downarrow$similarity threshold $\rightarrow$ $\downarrow$correctness $\wedge$ $\uparrow$conventionality $\wedge$ $\downarrow$lex. ambiguity

- ↑similarity threshold → ↑correctness ∧ ↓conventionality ∧ ↑lex. ambiguity

Figure 7 graphically demonstrates this cause-and-effect relationship using a Causal Loop Diagram (CLD). CLDs are an appropriate way of documenting variables and the causal relationships between them. The relationships in CLDs are annotated with polarity (+/–). The polarity tells whether a dependency has positive (+) polarity, whereby increasing the cause increases the effect, or negative (–) polarity, whereby increasing the cause decreases the effect (Hall 1962; Randers 1980).

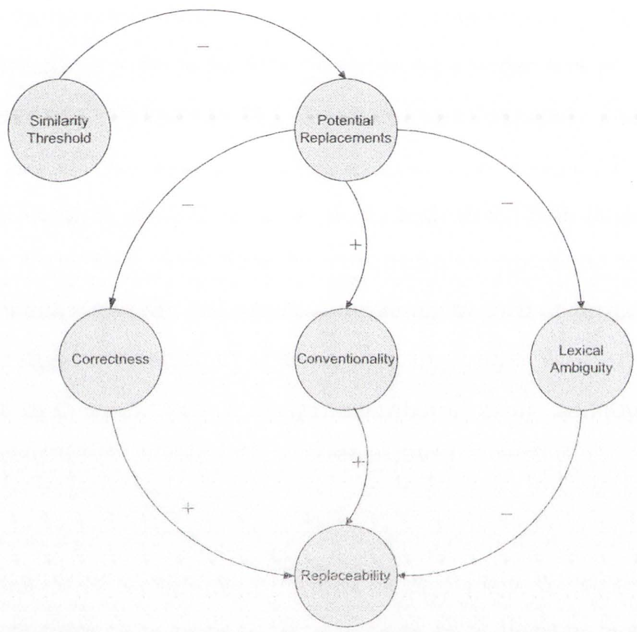

Figure 7 - Causal Loop Diagram

Figure 7 illustrates the inherent trade-off between correctness, conventionality, and lexical ambiguity. If the similarity threshold decreases, the number of potential replacements increases (i.e. negative polarity). If the number of potential replacements increases, then correctness decreases, conventionality increases, and lexical ambiguity decreases. Figure 7 also illustrates the desired relationship between our new measure *replaceability* and the quality factors: correctness, conventionality, and ambiguity. That is, we want replaceability to increase when correctness and conventionality increase, and when lexical ambiguity decreases:

- ↑replaceability → ↑correctness ∧ ↑conventionality ∧ ↓lex. ambiguity

Due to the mixture of positive and negative causal effects, combining a) b) c) and d) (above) with a Boolean AND is expected to result in extremely few replacements. We need to combine a) b) c) and d) in a way that results in W_r being *significantly* less ambiguous than W_o but without being *significantly* less correct and without being *significantly* less conventional, i.e. we acknowledge that W_r is unlikely to simultaneously satisfy all three predicates: correct, conventional, and unambiguous. To ensure the trade-off is equitable amongst quality factors, we must consider the relative magnitude of each contributor, such that scaling factors can be introduced to balance the equation.

The first scaling factor we introduce is based on the assumption that the more frequently occurring a word (frequency), the more likely it exists on a wider lexical level (width). I.e. we assume that width(W) $\propto$ frequency(W). This is because frequently occurring words are more conventional and are therefore more likely to exist on or around L0 in the hierarchy. Fallbaum finds that levels around L0 are more richly lexicalized than other levels, i.e. they represent the *bulge*. Considering that there are two measures pertaining to conventionality (i.e. frequency and width) and only one measure pertaining to each of correctness and lexical ambiguity, we have decided to exploit the assumption that width(W) $\propto$ frequency(W) and combine frequency and width into a single contributor using addition, i.e. $\forall W_r.\forall W_o$: similar(W_r,W_o) $\wedge$ ((frequency(W_r) + width(W_r)) $\geq$ (frequency(W_o) + width(W_o))) $\rightarrow$ conventional(W_r, context(W_o)).

The second scaling factor we introduce is based on Zipf's law. Recall from Section 2.4.2 that Zipf's law states that "the number of meanings of a word is correlated with its frequency". Zipf argues that conservation of speaker effort would prefer there to be only one word with all meanings while conservation of hearer effort would prefer each meaning to be expressed by a different word. Assuming that these forces are equally strong, Zipf argues that the number of meanings m of a word obeys the law: $m \propto \sqrt{f}$ (Zipf 1949), or in our language: |senses(W)| $\propto$ $\sqrt{}$frequency(W). Combining this with our assumption that width(W) $\propto$ frequency(W) gives |senses(W)| $\propto$ $\sqrt{}$(frequency(W) + width(W)).

The third scaling factor we introduce is based on the assumption that each original content word (W_o) is correct (Section 4.3.2 shows that when W_r=W_o, correct(W_r, context(W_o)) is always true), then we are interested in preserving that correctness (we assume if W_o is correct, then W_r cannot be *more* correct – only *less* correct). As similarity(meaning(W_o), senses(W_r)) reduces, so to does correct(W_r,W_o). We have therefore decided to attenuate replaceability(W_r,W_o) with similarity(meaning(W_o), senses(W_r).

In summary, we have decided that replaceability(W_r, W_o) will be defined as a function of:

- $similarity\left(meaning\left(W_o\right), senses\left(W_r\right)\right)$
- $\sqrt{\dfrac{freq\left(W_r\right)+width\left(W_r\right)}{freq\left(W_o\right)+width\left(W_o\right)}}$, and
- $\dfrac{|senses\left(W_o\right)|}{|senses\left(W_r\right)|}$

4.5.6.2. Method

$$repl\left(W_r,W_o\right)=similarity\left(meaning\left(W_o\right), senses\left(W_r\right)\right)\times\sqrt{\frac{freq\left(W_r\right)+width\left(W_r\right)}{freq\left(W_o\right)+width\left(W_o\right)}}\times\frac{|senses\left(W_o\right)|}{|senses\left(W_r\right)|}$$

Thus:

$\forall W_o.\exists W_r$: repl(W_r, W_o) > 1 → replaces(W_r, W_o)

$\forall W_o.\exists W_r$: repl(W_r, W_o) ≤ 1 → $W_r = W_o$ → ¬replaces(W_r, W_o)

$\forall W_o.\exists W_r$: $W_r = W_o$ → replaceability(W_r, W_o) = 1

Replaceability(W_r,W_o) is a measure of the ability of an original word (W_o) to be re-expressed by another word (W_r). Replaceability(W_r,W_o) is asymmetric because there is no guarantee that the inverse replacement will be valid, i.e. replaceability(W_r,W_o) ≠ replaceability(W_o,W_r).

4.5.6.3. Implementation

Refer to the `calculateReplaceability` function of Appendix F: Prototype Software for the C++ implementation of replaceability(W_r,W_o).

4.5.7. Sequence of Activities

Figure 8 is a UML sequence diagram that serves to illustrate the method of optimal re-expression. Importantly, Figure 8 provides functional sequencing and functional allocation to responsible entities including the human user, WordNet, WordNet::Similarity, and the Prototype Software that we have included in Appendix F: Prototype Software.

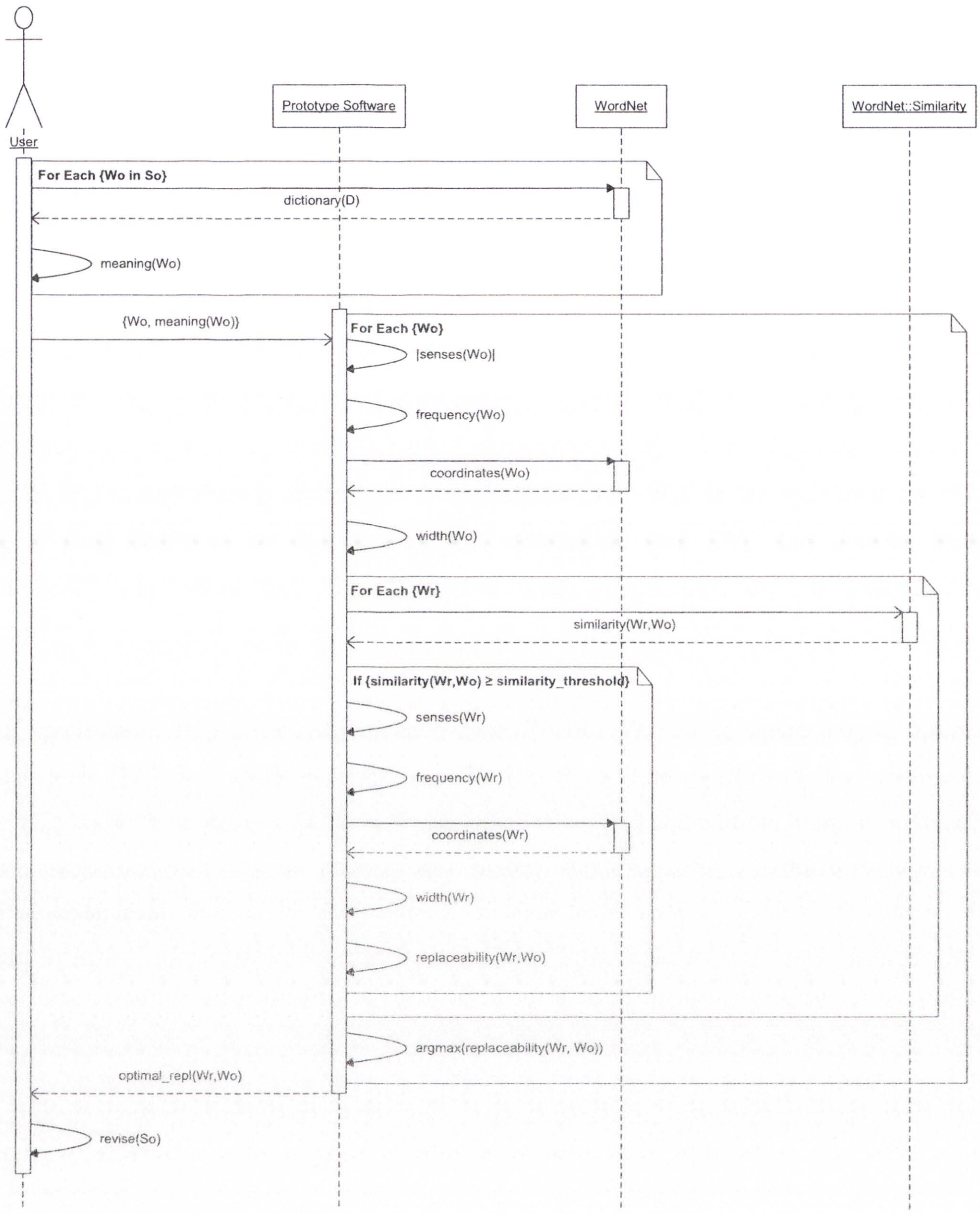

Figure 8 - UML sequence diagram

Figure 8 illustrates that the process of optimal re-expression begins with the human user disambiguating the meaning of each original word (W_o) from specification (S_o) using WordNet sense definitions as the dictionary (D). Once the meaning of all original words have been disambiguated, the set {W_o, meaning(W_o)} is provided as input to the Prototype Software in a text file. For each (W_o), the prototype software then determines a) the different senses of (W_o) in use in (S_o), b) the number of occurrences (frequency) of (W_o) in (S_o), and

c) the number of words occurring in (S_o) that are defined as coordinates terms of (W_o) in the WordNet hierarchy.

Then, for each potential replacement (W_r), where similarity(W_r,W_o) $\geq$ similarity_threshold, the prototype software determines a) the different senses of (W_r) in use in (S_o), b) the number of occurrences (frequency) of (W_r) in (S_o), and c) the number of words occurring in (S_o) that are defined as coordinates terms of (W_r) in the WordNet hierarchy. Given repl(W_r,W_o) = similarity(meaning(W_o), senses(W_r) * √ ((frequency(W_r) + width(W_r)) / (frequency(W_o) + width(W_o))) * |senses(W_o)| / |senses(W_r)|, the Prototype Software calculates repl(W_r,W_o) for every potential replacement (W_r). The (W_r) that maximises repl(W_r,W_o) is then chosen as the optimal replacement for (W_o). In theory, this is the (W_r) that is most likely to be *significantly* less ambiguous than (W_o) but without being *significantly* less correct than (W_o) and without being *significantly* less conventional than (W_o).

If specification (S_o) is revised with each original word (W_o) being replaced by its optimal replacement (W_r), we hypothesise that specification (S_r) will be *significantly* less ambiguous than (S_o) but without being *significantly* less correct than (S_o) and without being *significantly* less conventional than (S_o). We evaluate the validity of this hypothesis in the next chapter on design evaluation.

Chapter 5 Evaluation

5.1. Purpose

The purpose of this chapter is to define an evaluation model, develop an evaluation method, and to present the results of conducting the evaluation. The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods to claim a design artifact as a research contribution (Hevner et al. 2004).

The design artifacts we are evaluating in this chapter are:

- The Model (refer Section 4.3),
- The Method (refer Section 4.5), and
- The Software Prototype (refer Appendix F: Prototype Software)

These three design artifacts are not independent of one another. The model is the basis of the method, and the software prototype is an instantiation of the method. Therefore, we justifiably use the software prototype as the basis of our evaluation since this encapsulates the method and model. This chapter also produces the two remaining research contributions:

- The Evaluation Method, and
- The Empirical Evidence.

5.2. Evaluation Model

Before we evaluate our design solution to the research question, we must first establish an evaluation model to define and relate the key concepts of the evaluation. The extent to which a specification is considered correct, conventional, and lexically unambiguous is, to some extent, dependent upon the background of the reader. In Section 4.4.2 we theorised that it would be possible to optimally re-express each original word (W_o) with a replacement word (W_r) to significantly reduce the lexical ambiguity of the requirement without significantly reducing its correctness or conventionality. We now need to evaluate this theory, and specifically we want to evaluate whether a selection of human stakeholders agree that the optimally re-expressed specification has indeed significantly reduced lexical ambiguity without significantly reducing correctness or conventionality.

5.2.1. Metrics

Easterbrook et al. (2007) claim that Software Engineering evaluation typically involves human interaction. Since humans are the end-users of requirements (e.g. customers and developers), we believe it is essential to define our metrics in a way that enables humans to contribute to the evaluation. The metrics must somehow enable human participants to assess the correctness, conventionality, and lexical unambiguity of the optimal re-expressions. Fundamentally, the type of data that we can collect from human participants is, very different from the parameters of the model and method which is intended to be implemented by machine. For this reason, we have defined two evaluation metrics that can be easily extracted from human participants, and can be related to the model that we have defined in Section 4.3. The metrics we have defined are:

1. makes_sense(P, W_r, context(W_o))
2. would_use(P, W_r, context(W_o))

5.2.1.1. Makes_sense(P, W_r, context(W_o))

In Section 4.3.2 we defined the predicate correct(W_r, context(W_o)) as a function of another predicate similar(W_r,W_o) and the chosen similarity_threshold. Answering the question *similar(W_r, W_o)* may be entirely appropriate for a machine given W_r, W_o, a similarity measure, and a similarity threshold (recall method from Section 4.5.5), but is not an appropriate question to be answered by a human participant. A more appropriate question for a human might be whether replacement word (W_r) *makes sense* to them given the context of original word (W_o). This is a question a human participant should be able to answer if they are familiar with the business goal (B), and the application domain (D).

Predicate	Definition
makes_sense(P, W_r, context(W_o))	(participant) P thinks (word) W_r makes sense in context(W_o)

$$\forall W_o . \exists W_r : makes\_sense(P, W_r, context(W_o)) \rightarrow W_r \cup context(W_o) \cup D \models B \rightarrow correct(W_r, context(W_o))$$

$$\forall W_o . \exists W_r : Max(similarity\_threshold) \wedge similar(W_r, W_o) \rightarrow makes\_sense(P, W_r, context(W_o))$$

In other words, if participant (P) thinks replacement word (W_r) *makes sense* in the context of original word (W_o), then participant (P) thinks the re-expressed requirement continues to express the business goal (B) and is therefore correct (recall Jackson's definition of correct in Section 2.6.5). Furthermore, applying our predicate for correct from Section 4.3.2 implies that when the similarity_threshold is maximised, there exists a set of correct replacement words {W_r} that we expect will make sense to the human participants. This implies that one of the senses of W_r is the disambiguated meaning of W_o. As the similarity_threshold is reduced (i.e. no longer Max), the replacement words $\{W_r\}$ may no longer make sense to the human participants in the context of the original word (W_o), implying the similarity_threshold is an independent variable.

The *makes sense* evaluation metric can also be used to determine unambiguity. Recall our predicate for unambiguous from Section 4.3.4 of the model:

$\forall W_o.\exists W_r$: correct(W_r,W_o) $\wedge$ |senses(W_r)| = 1 $\rightarrow$ unambiguous(W_r, context(W_o))

Thus, if participant (P) determines that there are multiple (>1) dictionary senses of replacement word (W_r) that *make sense* in the context of original word (W_o), then we can interpret this to mean that participant (P) considers W_r to be an ambiguous replacement for W_o. In fact this implies that participant (P) is unsure of the business goal (B) since they thought that multiple different meanings of W_r satisfied the business goal.

5.2.1.2. Would_use(P, W_r, context(W_o))

In Section 4.3.3 we defined the predicate conventional(W_r, context(W_o)) as a function of another predicate correct(W_r, context(W_o)) and two parameters (frequency and width). In the previous Section 5.2.1.1 we explained that the correct(W_r, context(W_o)) predicate as defined in Section 4.3.2 of the model could be answered by machine, but is not appropriate to be answered by a human participant. Similarly, the number of occurrences of a word (frequency) and the number of coordinates of a word (width) could be answered by machine (recall methods from Sections 4.5.3 and 4.5.4 respectively), but is not an appropriate question to be answered by a human participant. A more appropriate question for a human might be whether they themselves *would use* the replacement word (W_r) within the context of original word (W_o). This is a question a human participant should be able to answer if they are familiar with the business goal (B), and the application domain (D).

Predicate	Definition
would_use(P, W_r, context(W_o))	(participant) P would use (word) W_r in context(W_o)

$\forall W_o.\exists W_r$: would_use(P, W_r, context(W_o)) $\rightarrow$ conventional(W_r, context(W_o))

$\forall W_o.\exists W_r$: correct(W_r,W_o) $\wedge$ frequency(W_r) $\geq$ frequency(W_o) $\wedge$ width(W_r) $\geq$ width(W_o) $\rightarrow$ would_use(P, W_r, context(W_o))

In other words, if participant (P) *would use* replacement word (W_r) in the context of original word (W_o), then participant (P) thinks the re-expressed requirement is conventional.

Furthermore, applying our predicate for conventional from Section 4.3.3 of the model implies that when W_r correctly expresses W_o, and the number of occurrences of W_r exceeds the number of occurrences of W_o (i.e. when W_r is more frequently occurring that W_o), and when the number of coordinate-terms of W_r exceeds the number of coordinate-terms of W_o (i.e. when W_r is on a more richly lexicalised level of the hierarchy than W_o), then we expect the human participants would use W_r to express W_o.

Since the original specification S_o is a constant textual corpus (refer Section 4.2.3) that is used as the reference text against which to measure both frequency and width (recall methods from Sections 4.5.3 and 4.5.4 respectively), we are left with the similarity_threshold as the only independent variable since it influences predicate correct(W_r, context(W_o)).

5.2.2. Variables

A variable is any quality or characteristic in a research investigation that has two or more possible values. When we investigate *cause-and-effect* relationships, we are looking at the extent to which one variable (the *cause*) influences another variable (the *effect*) (Leedy & Ormrod 2005).

5.2.2.1. Independent Variable(s)

A variable that the researcher studies as a possible *cause* of something else, in many cases, this is one that the researcher directly manipulates, is called an independent variable (Leedy & Ormrod 2005). The model in Section 4.2.4 identifies four variables: W_o, W_r, S_r, and similarity_threshold. Considering the causal loop diagram presented in Section 4.5.6, it should be evident that the similarity_threshold is the only variable that we can directly manipulate to *cause* a change in correctness, conventionality, and lexical ambiguity. I.e. the similarity_threshold is our independent variable. Manipulating the Similarity Threshold (by changing its value between 0 and 1) influences replaceability(W_r,W_o) and therefore optimal(W_r,context(W_o)).

Independent Variable	Definition
similarity_threshold	$0 \leq$ similarity_threshold ≤ 1

5.2.2.2. Dependent Variable(s)

A variable that is potentially *affected* by the independent variable is called a dependent variable (Leedy & Ormrod 2005). The research question implies that we have to measure the significance of change in the correctness, conventionality, and lexical ambiguity of the replacement word (W_r) compared to the original word (W_o). In particular, we are interested in comparing the proportion of replacement words $p(W_r)$ that are correct ($p(W_r)_{correct}$), conventional ($p(W_r)_{conventional}$), and unambiguous ($p(W_r)_{unambiguous}$) to the proportion of original words $p(W_o)$ that are correct ($p(W_o)_{correct}$), conventional ($p(W_o)_{conventional}$), and unambiguous ($p(W_r)_{unambiguous}$) respectively.

Dependent Variable	Definition
$p(W)_{correct}$	|makes_sense(P, W, context(W_o))| / |optimal(W_r,context(W_o))|, where $W_r \neq W_o$
$p(W)_{conventional}$	|would_use(P, W, context(W_o))| / |optimal(W_r,context(W_o))|, where $W_r \neq W_o$
$p(W)_{unambiguous}$	|makes_sense(P, W, context(W_o))| / |optimal(W_r,context(W_o))|, where $W_r \neq W_o$ and |senses(W)| = 1

Explanation:

- The definition of optimal(W_r,context(W_o)) in Section 4.4.2, and repl(W_r, W_o) in Section 4.5.6.2 means that when repl(W_r, W_o) ≤ 1 the original word is retained (i.e. $W_r = W_o$ so there is no noticeable replacement). Therefore, when counting proportions, we want the denominator to represent the total number of optimal replacements where $W_r \neq W_o$.
- *W* is used in the definitions to represent either W_r or W_o since both sets of proportions are needed for comparison purposes. Importantly, whilst we might expect the participants to consider the original words to be correct and conventional, we do not make this assumption in the evaluation.

5.3. Evaluation Method

A method is a set of organizing principles around which empirical data are collected and analyzed. The selection of evaluation methods must be matched appropriately with the designed artifact and the selected evaluation metrics (Hevner et al. 2004). Hevner et al. (2004) present five design evaluation methods appropriate to the design-science paradigm (Table 1 of (Hevner et al. 2004))

1. Observational (Case Study, Field Study)
2. Analytical (Static Analysis, Optimization)
3. Experimental (Controlled Experiment, Simulation)
4. Testing (Black Box, White Box)
5. Descriptive (Informed Argument, Scenarios)

We evaluate the integrated design artifact at multiple similarity thresholds to determine whether the dependent variables: correctness, conventionality, and lexical ambiguity have been affected in the ways predicted in Chapter 4. We are therefore interested in the *cause-and-effect* relationship between the similarity threshold (independent variable) and our dependent variables: correctness, conventionality, and lexical ambiguity.

Leedy and Ormrod 2005 state that "as a general rule, qualitative studies do *not* allow the researcher to identify *cause-and-effect* relationships – to answer questions such as *what caused what? Why did such-and-such happen?* You will need quantitative research, especially experimental studies, to answer questions of this kind." Qualitative studies involve studying phenomena in all their complexity without simplifying to a subset of specific quality attributes. Clearly our research does simplify to a subset of quality attributes, by solely focusing on correctness, conventionality, and lexical ambiguity rather than attempting to cover the full-range of possible quality attributes, e.g. per Davis et al. (1993).

We have chosen to evaluate the design solution using a Controlled Experiment for the following key reasons derived from Leedy and Ormrod (2005):

1. The research process considers a limited set of known variables with a detached view – rather than unknown variables with a personal view.
2. The data are numerical, representative and collected via questionnaire – rather than textual/image based, collected via personal interviews.
3. The data are to be analysed using inferential statistics – rather than inductive reasoning.
4. Findings are to be communicated formally using statistics – rather than through narratives, individual quotes in a literary style.

A controlled experiment is an investigation of a testable hypothesis where one or more independent variables are manipulated to measure their effect on one or more dependent variables. Controlled experiments allow us to determine in precise terms how the variables are related and, specifically, whether a cause-effect relationship exists between them. Each combination of values of the independent variables is a treatment (Leedy & Ormrod 2005).

Leedy & Ormrod (2005) claim "A researcher can most convincingly identify *cause-and-effect* relationships by using an experimental design. In such a design, the researcher considers many possible factors that might cause or influence a particular condition or phenomenon. The researcher then attempts to control for all influential factors *except* those whose possible effects are the focus of investigation".

The term experimental design refers to a plan for assigning subjects to treatment conditions. A good experimental design serves three purposes.

- Causation. It allows the experimenter to make causal inferences about the relationship between independent variables and a dependent variable.
- Control. It allows the experimenter to rule out alternative explanations due to the confounding effects of extraneous variables (i.e., variables other than the independent variables).
- Variability. It reduces variability within treatment conditions, which makes it easier to detect differences in treatment outcomes.

5.3.1. Confounding Effects

To maximise internal validity when identifying *cause-and-effect* relationships, the researcher needs to control confounding variables so that these variables are ruled out as explanations for *effects* observed (Leedy & Ormrod 2005). We have controlled the confounding effects as follows:

5.3.1.1. Multiple Parts of Speech

Given that we have defined predicates for correct, conventional and unambiguous in terms of the context, and given that a single requirement statement is comprised of multiple words from multiple parts of speech, there is a risk that the context for a single word in a single requirement will be contaminated by other replacements associated with other words. This is confirmed by Ravin and Leacock (2000), who state "context alters the senses of the words found in it". In other words, a context comprised of replacements (W_r) will likely yield different results for predicates correctness, conventionality, and unambiguity than a context comprised purely of original words (W_o) from S_o.

Since requirements should be written as simple sentences rather than compound or complex (Hooks 1994; Hull, Jackson and Dick 2002) and since simple sentences contain a single subject and predicate, and describe only one thing, idea or question, and have only one verb (Learn English, 2008), we have decided that by limiting our evaluation to optimally re-expressing verbs, and no other part of speech, then in the case of requirements written as simple sentences, we will likely limit re-expression to one word per requirement, and therefore the results will not be confounded by a context comprised of substitutions associated with other optimal replacements. It is also worth noting that verbs are the most lexically ambiguous part of speech averaging 2.11 senses per verb (Oxford English Dictionary 2005), and therefore potentially the part of speech with most to gain from our research. Furthermore, verbs are more common than nouns, in the sense that verbs tend to be less domain-specific (or proper), and so our assumption in Section 4.5.1.2 that words contained in the original specification (W_o) will also be defined in the dictionary (i.e. $\forall W_o : W_o \in S_o \rightarrow meaning(W_o) \in D$) is more likely to be valid when W_o is a verb.

5.3.1.2. Time

Whilst time is often a confounding variable in research, this is not the case with our research. This is because requirements are static textual statements that do not *naturally* evolve with time (i.e. change is always due to intervention and not due to ageing). Therefore, by preventing human intervention, we can effectively eliminate *time* as a confounding variable in our design evaluation. In doing so, we can simultaneously measure the *before* and *after* effects in a single application of a measuring instrument. I.e. there is no need for the *before* and *after* measurements to be separated in time.

5.3.2. Constants

Constants were identified in Section 4.2.3 as 1) S_o, the original specification, and 2) D, the dictionary. To ensure the results of the experiment are valid, the original specification (S_o) should be a specification that is representative of the intended population. The specification used for the evaluation is selected and described in Section 5.4.1. WordNet was already nominated in Section 4.5.1.3 as the constant dictionary (D).

5.3.3. Hypotheses

Given that we have decided on a controlled experiment, it is normal to take a post-positivist stance, in that we "tend to accept the idea that it is more productive to refute theories than to prove them, and we increase our confidence in a theory each time we fail to refute it, without necessarily ever proving it to be true" (March and Smith 1995).

We believe the design solution proposed in Chapter 4 *automatically re-expresses natural English requirements in a way that significantly reduces lexical ambiguity, without significantly reducing the correctness or conventionality*. That is, we hypothesise that our design solution answers the research question, and we intend to formally test our belief by establishing one or more formal research hypotheses. Thus, we expect that the:

1) Proportion of replacement terms that are correct $p(W_r)_{correct}$ in S_r is not significantly less than the proportion of corresponding original terms that are correct $p(W_o)_{correct}$ in S_o.

2) Proportion of replacement terms that are conventional $p(W_r)_{conventional}$ in S_r is not significantly less than the proportion of corresponding original terms that are conventional $p(W_o)_{conventional}$ in S_o.
3) Proportion of replacement terms that are unambiguous $p(W_r)_{unambiguous}$ in S_r is significantly more than the proportion of corresponding original terms that are unambiguous $p(W_o)_{unambiguous}$ in S_o.

Every hypothesis test requires the analyst to state a null hypothesis (labelled H0) and an alternative hypothesis (labelled HA). The hypotheses are stated in such a way that they are mutually exclusive.

Hypothesis	**Definition**
$H0_{correct}$	$p(W_r)_{correct} \geq p(W_o)_{correct}$
$HA_{correct}$	$p(W_r)_{correct} < p(W_o)_{correct}$
$H0_{conventional}$	$p(W_r)_{conventional} \geq p(W_o)_{conventional}$
$HA_{conventional}$	$p(W_r)_{conventional} < p(W_o)_{conventional}$
$H0_{unambiguous}$	$p(W_r)_{unambiguous} \leq p(W_o)_{unambiguous}$
$HA_{unambiguous}$	$p(W_r)_{unambiguous} > p(W_o)_{unambiguous}$

- Our expectation that *re-expression has not sacrificed correctness* is supported with failure to reject the null hypothesis ($H0_{correct}$). Rejection of ($H0_{correct}$) indicates that re-expression has sacrificed correctness.
- Our expectation that *re-expression has not sacrificed conventionality* is supported with failure to reject the null hypothesis ($H0_{conventional}$). Rejection of ($H0_{conventional}$) indicates that re-expression has sacrificed conventionality.
- Our expectation that *re-expression has significantly reduced lexical-ambiguity* is supported with rejection of the null hypothesis ($H0_{unambiguous}$). Failure to reject ($H0_{unambiguous}$) indicates that re-expression has not significantly reduced lexical-ambiguity.

5.3.4. The Data

The data, in this case, is the opinion of the human research participants, that is 1) whether the human research participants think the replacement word (W_r) *makes sense* in the context of the original word (W_o), i.e. a) Yes, I think W_r makes sense in the context of W_o, or b) No, I think W_r does not make sense in the context of W_o, and 2) whether they *would use* the replacement word (W_r) in the context of the original word (W_o), i.e. a) Yes, I would use W_r in the context of W_o, or b) No, I would not use W_r in the context of W_o. Thus, the data will comprise two binary Yes / No answers per optimal replacement (W_r).

5.3.5. Data Collection

The purpose of this section is to nominate an appropriate data collection instrument and to design that instrument to ensure the data collected is reliable. We have so far determined that:

- The evaluation method is a controlled experiment,
- The data are subjective but quantified,

Since the data to be collected is subjective and opinion-based, we believe that a survey represents the most appropriate data-collection instrument. In a survey, the researcher poses a series of questions to willing participants – perhaps about their characteristics, opinions, attitudes, or previous experiences (Leedy & Ormrod 2005). Surveys may be implemented as a face-to-face interview, a telephone interview, or as a written questionnaire.

We have decided to collect the data by questionnaire, since questionnaires are the most efficient and least intrusive data collection method. Leedy and Ormrod (2005) describe two disadvantages of questionnaires being 1) "low return rate" and 2) "potential for misinterpretations". In our Ethics Application (refer to Appendix A: Ethics Approval) we additionally recognised 3) "risk of embarrassment" where the participant is uncomfortable providing an answer. We believe 1) is less of an issue since this researcher is employed within the domain and expects reasonable participation from colleagues (participants are described later in Section 5.3.9), 2) is controlled by carefully designing the questionnaire and conducting a pilot test, 3) is minimised with a questionnaire rather than via face-to-face or telephone.

5.3.6. Questionnaire Design

Since we eliminated *time* as a confounding variable in our design evaluation (refer 5.3.1.2), we designed the questionnaire to simultaneously measure the *before* and *after* effects. I.e. there was no need for the *before* and *after* measurements to be separated in time since requirements do not age and only change due to human intervention.

5.3.6.1. Questions

The questions were designed to individually collect a complete set of data. This means that each single question collects data pertaining to correctness, conventionality, and lexical ambiguity across the range of similarity thresholds (i.e. treatments). Recall from Section 5.2.1 that by collecting the metric *makes sense* from the participants, we can calculate the correctness and lexical ambiguity variables, and by collecting the metric *would use* from the participants, we can calculate the conventionality variable.

5.3.6.1.1. Makes Sense?

In this question, we present to the research participant in random order, the set of Senses(W_r) using the sense definitions from WordNet, and we ask the research participant to choose the sense(s) that *make sense* to them in the Context(W_o). We ask this question for each sense of each optimal replacement (W_r), where different optimal replacements (W_r) may occur at different similarity thresholds (treatments).

The context(W_o) is defined as the words surrounding W_o, but excluding W_o:

- context(W_o) = [W_1, W_2, … ,"____", …, W_{N-1}, W_N]

Depending on the similarity_threshold selected, we can have a different optimal W_r:

- similarity_threshold = i → optimal($W_{r(i)}$, context(W_o))
- similarity_threshold = j → optimal($W_{r(j)}$, context(W_o))

Each word has one or more senses defined in the dictionary (D):

- senses(W_o) = {$sense_{o.1}$, $sense_{o.2}$, …, $sense_{o.N}$}
- senses($W_{r(i)}$) = {$sense_{i.1}$, $sense_{i.2}$, …, $sense_{i.N}$}
- senses($W_{r(j)}$) = {$sense_{j.1}$, $sense_{j.2}$, …, $sense_{j.N}$}

We can now generate questions asking which individual senses in senses(W_o), senses($W_{r(i)}$), senses($W_{r(j)}$), etc, *makes sense* in place of the context blank "____":

1. $sense_{i.1}$ *makes sense* in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)
2. $sense_{o.N}$ *makes sense* in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)
3. $sense_{j.N}$ *makes sense* in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)
4. $sense_{o.1}$ *makes sense* in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)
5. $sense_{i.N}$ *makes sense* in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)
6. $sense_{j.1}$ *makes sense* in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)

Notes:

[1] question order is random so there are no hints to i, j or W_o, etc.

[2] $sense_{o.2}$, $sense_{i.2}$, $sense_{j.2}$ not shown for brevity.

We would expect the research participant to nominate $sense_{o.1}$ or $sense_{o.2}$ or $sense_{o.N}$ since these are the senses of the original word (W_o). We expect (from the causal loop diagram in Section 4.5.6) that as the similarity threshold reduces (i.e. when i > j) the research participant is more likely to nominate meanings that occur within senses($W_{r(i)}$) than meanings that occur within senses($W_{r(j)}$) as those meanings that *make sense*. If more than one sense of the same W_r *makes sense* in Context(W_o), then W_r is ambiguous in Context(W_o). I.e. both dependent variables $p(W)_{unambiguous}$ and $p(W)_{correct}$ are based on the *makes sense* metric.

5.3.6.1.2. Would Use?

In this question, we present to the research participant in random order, the set of optimal replacements (W_r) that occur at different similarity thresholds (treatments) and we ask the research participant to choose the sense(s) that they *would use* in the Context(W_o).

The context(W_o) is defined as the words surrounding W_o, but excluding W_o:

- context(W_o) = [W_1, W_2, … ,"____", …, W_{N-1}, W_N]

Depending on the similarity_threshold selected, we can have a different optimal W_r:

- similarity_threshold = i → optimal($W_{r(i)}$, context(W_o))
- similarity_threshold = j → optimal($W_{r(j)}$, context(W_o))

We can now generate questions asking which words out of W_o, $W_{r(i)}$, $W_{r(j)}$, etc, the research participant *would use* to fill the context blank "____":

1. I *would use* $W_{r(j)}$ in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)
2. I *would use* W_o in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)
3. I *would use* $W_{r(i)}$ in [W_1, W_2, … ,"____", …, W_{N-1}, W_N]? (True / False)

Note question order is random so there are no hints to i, j or W_o, etc.

We would expect the research participant to use W_o since this was the original word. In Section 4.5.6 we explained that as the similarity threshold reduces (i.e. when i > j) the research participant is more likely to use $W_{r(j)}$ than $W_{r(i)}$.

5.3.6.1.3. Similarity Threshold

Recall from Section 4.5.6 that changing the similarity threshold widens (or narrows) the scope of potential replacements, and thus:

- ↓similarity threshold → ↓correctness ∧ ↑conventionality ∧ ↓lex. ambiguity
- ↑similarity threshold → ↑correctness ∧ ↓conventionality ∧ ↑lex. ambiguity

We are interested in discovering evidence to support the conjecture that there is a *cause-and-effect* relationship between the similarity threshold and the dependent variables: correctness, conventionality, and lexical ambiguity. Importantly, we are interested in discovering the range of similarity thresholds that yield a desired outcome to the hypothesis testing, specifically:

1. No significant reduction in correctness ($H0_{correct}$ supported)
2. No significant reduction in conventionality ($H0_{conventional}$ supported)
3. A significant reduction in lexical ambiguity ($H0_{unambiguous}$ rejected)

The similarity threshold is a real number between 0 and 1. Thus, with every decimal place, the range of possible values increases by an order of magnitude. Given that each similarity threshold may result in a different optimal(W_r, context(W_o)), and given the need to evaluate the predicates *makes sense* and *would use* for each optimal(W_r, context(W_o)), we need to carefully control the granularity of the similarity threshold so as to prevent the questionnaire and resultant data from becoming unnecessarily voluminous.

Considering the Path Length measure of similarity is calculated as sim(W_1, W_2) = 1/ length(W_1, W_2), (refer Section 4.5.5) and given that length(W_1, W_2) is an integer, we speculate that the range of similarity thresholds (0, 0.25, 0.5, 0.75, 1) provides adequate granularity for the purpose of our evaluation. Furthermore, this range of similarity thresholds aligns with our experience running WordNet::Similarity with both the Path Length and Wu and Palmer similarity measures.

So that there is no bias in the responses, we hide the similarity threshold from the respondent when presenting each question. This was demonstrated by randomly ordering the questions in Sections 5.3.6.1.1 and 5.3.6.1.2, i.e. we don't want the situation where the respondent *expects* the replacement word W_r not to make sense *because* the similarity threshold was low (or vice-versa).

5.3.6.1.4. Length

With a similarity threshold in the set 0, 0.25, 0.5, 0.75, 1, and assuming that each similarity threshold yields a different optimal(W_r, context(W_o)) then there are five words that influence the answers to each question. Assuming average senses per word = 2, then there are 10 *makes sense* decisions to be made per question. Assuming on average there are 25 words per sense definition in WordNet, this gives 250 words to read per question. The international average reading speed is accepted to be about 250 words per minute (Hladczuk & Eller 1992). Therefore, reading each question can be expected to take approximately one minute. If we allow an additional two minutes for decision making and choosing answers, then each question is expected to take roughly three minutes to complete. In accordance with our ethics application (refer to Appendix A: Ethics Approval), we limit the questionnaire to 120 minutes duration (max). This gives 40 questions per questionnaire. This estimation was validated by pilot test.

5.3.6.1.5. Format and Structure

The questionnaire needs to be formatted and structured to enable efficient codification. Leedy and Ormrod (2005) claim electronic questionnaires can be highly effective when the participants feel comfortable with computers. Our respondents are to be knowledgeable stakeholders of the specifications being analysed. In most cases, the types of people will be engineers and design managers with a strong computing background, especially familiar with spreadsheets and word processors. We therefore decided to use Microsoft Excel™ as the spreadsheet tool to present questions and collect data.

One advantage of this decision is that we can make use of Excel's *data validation* feature such that respondents are forced into providing a valid response for every question via a drop-down single-selection box. Another advantage is that Excel™ enables us to hide information from the respondent. For example, we hide the correlation between sense definitions and words (W_o, $W_{r(0)}$, $W_{r(0.25)}$, $W_{r(0.5)}$, $W_{r(0.75)}$, $W_{r(1)}$), but then instantly reveal these correlations during post-test-analysis to run hypothesis tests on proportions ($p(W)_{correct}$, $p(W)_{conventional}$, $p(W)_{unambiguous}$).

5.3.6.1.6. Validation

Leedy and Ormrod (2005) define twelve guidelines for developing a questionnaire that encourages people to be co-operative and yield responses that can be used and interpreted. Table 3 provides traceability from each of the twelve guidelines to the section within this chapter that addresses the guideline.

Table 3 - Questionnaire Validation (from (Leedy & Ormrod 2005))

Check	Refer Section
Keep it short	Section 5.3.6.1.4: Length
Use simple, clear, unambiguous language	Appendix D: Questionnaire Pilot Test (Sheet 1) Appendix E: Questionnaire Pilot Test (Sheet 2)
Check for unwarranted implicit assumptions	Appendix D: Questionnaire Pilot Test (Sheet 1) Appendix E: Questionnaire Pilot Test (Sheet 2)
Don't hint at preferred response	Section 5.3.6.1: Questions
Check for consistency ("countercheck")	Section 5.3.10.1: Interrater Reliability
Structure for codification	Section 5.3.6.1.5: Format & Structure
Keep respondent's task simple	Appendix C: Questionnaire Instruction Sheet
Provide clear instructions	Appendix C: Questionnaire Instruction Sheet
Give a rationale when the purpose is unclear	Appendix B: Research Participant Information Sheet
Looks attractive and professional	Appendix D: Questionnaire Pilot Test (Sheet 1) Appendix E: Questionnaire Pilot Test (Sheet 2)
Conduct pilot test	Appendix D: Questionnaire Pilot Test (Sheet 1) Appendix E: Questionnaire Pilot Test (Sheet 2)
Scrutinize the almost-final product	Appendix D: Questionnaire Pilot Test (Sheet 1) Appendix E: Questionnaire Pilot Test (Sheet 2)

5.3.7. Sample Size

In Section 5.3.6.1.4 we decided that there would be 40 questions in the questionnaire. Since a single question represents one original word W_o being replaced ($W_{r(0)}$, $W_{r(0.25)}$, $W_{r(0.5)}$, $W_{r(0.75)}$, $W_{r(1)}$), this limitation imposes a sample size of 40 replacements. Experimental testing of the software prototype has shown that if we have 1000 W_o, we can expect between 50 and 200 unique W_r (depending on the similarity threshold). Thus, a sample size of 40 replacements represents from 80% (50 unique W_r) to 20% (200 unique W_r) of the total number of unique replacements expected to occur. Gay and Airasian (2003) suggest that a sample size of 20% should be sufficient.

Therefore, a single sample of 40 replacements were randomly drawn from the population of all optimal replacements W_r. In Section 5.3.11.2 we use the concept of proximal similarity to generalise the results from the sample to a bounded population of requirement specifications.

5.3.8. Hypothesis Testing

In Section 5.3.3 we defined the null hypotheses as:

Null Hypothesis	Definition
$H0_{correct}$	$p(W_r)_{correct} \geq p(W_o)_{correct}$
$H0_{conventional}$	$p(W_r)_{conventional} \geq p(W_o)_{conventional}$
$H0_{unambiguous}$	$p(W_r)_{unambiguous} \leq p(W_o)_{unambiguous}$

In Section 4.5.6 we explained that the similarity threshold influences the replacement word (W_r), therefore we expand the above hypotheses into a hypothesis per similarity threshold. Thus we have:

Sim. Thres.	Correctness	Conventionality	Unambiguity
0	$H0_{correct(0)}$	$H0_{conventional(0)}$	$H0_{unambiguous(0)}$
0.25	$H0_{correct(0.25)}$	$H0_{conventional(0.25)}$	$H0_{unambiguous(0.25)}$
0.5	$H0_{correct(0.5)}$	$H0_{conventional(0.5)}$	$H0_{unambiguous(0.5)}$
0.75	$H0_{correct(0.75)}$	$H0_{conventional(0.75)}$	$H0_{unambiguous(0.75)}$
1	$H0_{correct(1)}$	$H0_{conventional(1)}$	$H0_{unambiguous(1)}$

We therefore test the following 15 hypotheses:

Null Hypothesis	Definition
$H0_{correct(0)}$	$p(W_r)_{correct(0)} \geq p(W_o)_{correct}$
$H0_{correct(0.25)}$	$p(W_r)_{correct(0.25)} \geq p(W_o)_{correct}$
$H0_{correct(0.5)}$	$p(W_r)_{correct(0.5)} \geq p(W_o)_{correct}$
$H0_{correct(0.75)}$	$p(W_r)_{correct(0.75)} \geq p(W_o)_{correct}$
$H0_{correct(1)}$	$p(W_r)_{correct(1)} \geq p(W_o)_{correct}$
$H0_{conventional(0)}$	$p(W_r)_{conventional(0)} \geq p(W_o)_{conventional}$
$H0_{conventional(0.25)}$	$p(W_r)_{conventional(0.25)} \geq p(W_o)_{conventional}$
$H0_{conventional(0.5)}$	$p(W_r)_{conventional(0.5)} \geq p(W_o)_{conventional}$
$H0_{conventional(0.75)}$	$p(W_r)_{conventional(0.75)} \geq p(W_o)_{conventional}$
$H0_{conventional(1)}$	$p(W_r)_{conventional(1)} \geq p(W_o)_{conventional}$
$H0_{unambiguous(0)}$	$p(W_r)_{unambiguous(0)} \leq p(W_o)_{unambiguous}$
$H0_{unambiguous(0.25)}$	$p(W_r)_{unambiguous(0.25)} \leq p(W_o)_{unambiguous}$
$H0_{unambiguous(0.5)}$	$p(W_r)_{unambiguous(0.5)} \leq p(W_o)_{unambiguous}$
$H0_{unambiguous(0.75)}$	$p(W_r)_{unambiguous(0.75)} \leq p(W_o)_{unambiguous}$
$H0_{unambiguous(1)}$	$p(W_r)_{unambiguous(1)} \leq p(W_o)_{unambiguous}$

We are interested in determining the range of similarity thresholds that give:

- No significant reduction in correctness ($H0_{correct}$ supported), and
- No significant reduction in conventionality ($H0_{conventional}$ supported), and
- A Significant reduction in lexical ambiguity ($H0_{unambiguous}$ rejected).

5.3.8.1. Z-Test for Two-Proportions

Each hypothesis that we have defined is a comparison of two proportions, i.e. a comparison between $p(W_r)$ and $p(W_o)$. The appropriate statistical inference test to apply when comparing two proportions is the *Z-Test for Two Proportions*. The first proportion ($p(W_o)$) is from the original specification (pre-treatment), and the second proportion ($p(W_r)$) is from the optimally-expressed specification (post-treatment).

To answer our research question we need to determine whether or not the reduction in quality factors: correctness, conventionality, and lexical ambiguity is statistically *significant* or not. As explained in the previous section, we hope that there is no significant reduction in correctness and conventionality, but that there is a significant reduction in lexical ambiguity. Given that we are only interested in determining whether or not there is a *reduction* in quality factors, we have by definition, a "one-tailed" test.

5.3.8.1.1. Significance Level (α)

We have chosen a significance level of 99%, corresponding to an alpha (α) of 0.01 for this research. We understand that significance levels of 95% and 99% are commonly applied in research without detailed rationale. We do however recognise that in our case:

- the higher the significance level, the bigger the difference needed in proportions to reject the null hypothesis, so it is more likely $H0_{correct}$ and $H0_{conventional}$ will be supported, and it is less likely $H0_{unambiguous}$ will be rejected. Conversely,
- the lower the significance level, the smaller the difference needed in proportions to reject the null hypothesis, so it is less likely $H0_{correct}$ and $H0_{conventional}$ will be supported, and it is more likely $H0_{unambiguous}$ will be rejected.

In other words, no matter what significance level (and α) we choose, there will be both a positive and negative effect in terms of design evaluation.

5.3.8.1.2. Critical Value

The critical value(s) for a hypothesis test is the threshold against which the Z-score is compared to determine whether or not the null hypothesis is rejected. The critical value for any hypothesis test depends on the significance level at which the test is carried out, and whether the test is one-tailed or two-tailed. Our Z-tests are one-tailed since the alternative hypothesis is "< / >" and not "= / ≠". Table II from (Fisher & Yates 1974) shows the critical value (Z_{crit}) = +/- 2.33 for a one-tailed test ($\alpha = 0.01$).

5.3.8.1.3. Test Statistic (Z-Score)

The z-score test statistic (z) is computed for each hypothesis tested as follows:

$$S_{p1-p2} = \sqrt{\hat{p}\,\hat{q}}\,\sqrt{\frac{n_1 + n_2}{n_1\, n_2}} \quad \text{where, } \hat{p} = \frac{n_1\, p_1 + n_2\, p_2}{n_1 + n_2} \text{ and } \hat{q} = 1 - \hat{p}$$

$$Z = \frac{p_1 - p_2}{S_{p1-p2}}$$

Where,

- p_1 = proportion 1 (e.g. $p(W_o)_{correct}$)
- p_2 = proportion 2 (e.g. $p(W_r)_{correct(0)}$)
- $n_1 = n_2$ = sample size

If there is evidence that:

- $p(W_r)_{correct}$ is significantly less than $p(W_o)_{correct}$
- $p(W_r)_{conventional}$ is significantly less than $p(W_o)_{conventional}$
- $p(W_r)_{unambiguous}$ are significantly more than $p(W_o)_{unambiguous}$

As determined by a comparison of the magnitude of the test statistic (z-score) with the critical values (Z_{crit}) of the probability distribution, we reject the respective null hypothesis in support of the alternative hypothesis.

5.3.9. Participants

Since the questionnaire requires the respondents to read technical requirements and make decisions on 1) whether dictionary-defined meanings *make sense* in a particular context, and 2) whether the respondent *would use* a particular word in a particular context, it is imperative that the respondents be domain knowledgeable people with at least one year of experience working in the domain. In accordance with our ethics application (refer Appendix A: Ethics Approval), we anticipate five domain knowledgeable respondents. Their employer has limited the maximum number of participants to be five, given the estimated 120 minute duration. We have decided that it is preferable to conduct the experiment with five domain knowledgeable people, than to conduct the experiment with a larger less-experienced group (such as university students). It is our intention that the five respondents each be from a different team on the project (e.g. systems engineering, test engineering, software engineering, requirements engineering, etc). With 5 respondents each answering 40 questions, this generates 200 unique answers, with each individual question being answered 5 times. The information sheet that we used to invite people to participate in the research is provided in Appendix B: Research Participant Information Sheet. Relevant respondent details are provided in Table 7.

5.3.10. Reliability

Leedy and Ormrod (2005) explain we can measure something accurately only when we can also measure it consistently. In other words, in order to have validity, we must also have reliability.

5.3.10.1. Interrater Reliability

Interrater reliability is the extent to which two or more individuals evaluating the same product or performance give identical judgements. We have 40 questions to be answered, with each question being answered by five different people. We expect interrater reliability to be high given that all respondents are domain-knowledgeable with similar backgrounds and experience levels. In other words, we *expect* each of the five respondents to answer the same questions in the same way.

5.3.10.2. Test-retest Reliability

Providing *interrater reliability* is high (expected), we do not consider it necessary to have the same respondent re-conduct the questionnaire (i.e. *test-and-retest*). High *interrater reliability* means different participant answers are generally in agreement, and therefore re-testing the same participant on the same questionnaire is not likely to change the result. In addition, since this research require approximately 120 minutes of paid employee-time, re-testing the same participant on the same questionnaire may be perceived as a waste of company time. Therefore, we do not re-test participants unless interrater reliability proves lower than expected.

5.3.11. Validity

5.3.11.1. Internal Validity

Internal Validity is the approximate truth about inferences regarding cause-effect or causal relationships. Thus, internal validity is only relevant in studies that try to establish a causal relationship (Trochim 2006).

The key question in internal validity is whether observed changes can be attributed to your program or intervention (i.e., the cause) and *not* to other possible causes (sometimes described as alternative explanations for the outcome). In our case, whether the change in dependent variables: correctness, conventionality, and lexical ambiguity can be attributed to the change in similarity threshold (independent variable).

5.3.11.1.1. Establishing a Cause & Effect (Causal) Relationship

There are three criteria that must be met to claim a causal relationship exists (Trochim 2006).

1. **Temporal Precedence:** this means the cause must happen before the effect. In Section 5.3.6.1.3 we explain that optimal(W_r, context(W_o)) depends on the similarity threshold, and predicates *makes sense* and *would use* depend on optimal(W_r, context(W_o)). I.e. our evaluation method is to change the similarity threshold before we measure the effects on correctness, conventionality, and lexical ambiguity, thus we have demonstrated temporal precedence.
2. **Covariation of the Cause and Effect:** this means there must be a defined relationship between the cause and effect. In Section 4.5.6 we present a Causal Loop Diagram that shows the causal relationships between the similarity threshold (independent variable) and dependent variables: correctness, conventionality, and lexical ambiguity, thus we have demonstrated covariation of cause and effect.
3. **No Plausible Alternative Explanation:** this means that there are no other missing variables causing the outcome. In Sections 5.3.1 and 5.3.2 we defined the "confounding variables" and "constants" respectively, thus we have considered and controlled other possible causes.

5.3.11.1.2. Threats to Internal Validity

With a pretest-posttest single-group design, there are certain threats to internal validity that must be identified and controlled. We believe that all threats to the internal validity of our research have been adequately controlled in the design of the experiment.

1. **History Threat:** this means it is not the intervention that caused the outcome, but rather some historical event that occurred after the intervention and before the measurement. In Section 5.3.1.2 we explained that time is not a confounding variable for our research. In Section 5.3.6 we design the one questionnaire to collect data on the correctness, conventionality, and lexical ambiguity of both the original (W_o) and replacement (W_r) words. I.e. one questionnaire collects both pre- and post-test data, thus there is no separation in time and the *history threat* has been controlled.

2. **Testing Threat:** this means that taking the pre-test affects how participants respond to the post-test. The same argument from the history threat with respect to time and using the single questionnaire for pre- and post- test results applies here. In addition, Sections 5.3.6.1.1 and 5.3.6.1.2 explain that the participant have no *hints* about whether they are answering a pre- or post- test question for correctness, conventionality, and lexical ambiguity respectively. I.e. one questionnaire collects both pre- and post-test data, and participants are unaware whether they are answering a pre- or post-test question, thus the *testing threat* has been controlled.
3. **Instrumentation Threat:** this means that the pre-test measuring instrument differs in some way to the post-test measuring instrument. In Section 5.3.6 we explained that a single questionnaire is used for collecting pre- and post- test data (refer to Appendix D: Questionnaire Pilot Test (Sheet 1) and Appendix E: Questionnaire Pilot Test (Sheet 2)). In Section 5.3.10 we considered the reliability of the measuring instrument. I.e. the measuring instrument is identical for pre- and post- test, thus the *instrumentation threat* has been controlled.
4. **Mortality Threat:** this means that people conducted the pre-test, and then dropped out of the experiment before conducting the post-test. In Section 5.3.6 we explained that a single questionnaire is be used for collecting pre- and post- test data (refer to Appendix D: Questionnaire Pilot Test (Sheet 1) and Appendix E: Questionnaire Pilot Test (Sheet 2)). I.e. the number of participants that answer pre-test questions should be identical to the number of participants that answer post-test questions, thus the *mortality threat* has been controlled.

5.3.11.2. External Validity

External validity is related to generalizing, and is the degree to which the conclusions in our study would hold for other persons in other places and at other times. We contribute more to humanity's knowledge about the world when we conduct research that has implications that extend far beyond the specific situation studied (Leedy & Ormrod 2005). When the result is externally valid, we can generalize this result that we obtained for the sample to a larger population.

There are two major approaches to how we provide evidence for a generalization 1) the Sampling Model, and 2) the Proximal Similarity Model (Campbell 1986).

1) The Sampling Model conducts the research using a representative sample from the population. Since the sample is representative of the population, it is possible to

generalise the results of the study to the population. The Sampling Model can only be used when the population is known *a priori*. External validity is threatened whenever the sample is not representative of the population.

2) The Proximal Similarity Model conducts the research first, and then identifies other contexts that are considered similar to the study. This approach means generalising the results of our study to other persons, places or times that are more like (that is, more proximally similar) to our study. Notice that here, we can never generalize with certainty -- it is always a question of *more* or *less* similar.

We have opted to apply the Proximal Similarity Model since it seems more sensible to evaluate the optimal constraint method on a specification, and then determine how the method could be generalised to other contexts such that similar results could be expected, rather than trying to start with some unknown and inaccessible population of requirement specifications, for instance, would the population be all specifications for a particular organisation? particular country? particular industry? etc.

5.3.11.2.1. Limitations of Generalisation

Threats to external validity relate to how we might incorrectly generalise. For instance, we conclude that the results of our study, which was done in a specific place, with certain types of people, and at a specific time can be generalized to another context, i.e. another place, with slightly different people, at a slightly later time. In other words, once we answer the question: *How can a requirements specification be automatically re-expressed in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality?* What is the extent to which we can generalise the answer? 1) to other parts of speech, other than verbs?, 2) to other domains, other than the Rail Industry?, and 3) to specifications written in other languages, other than English?

1) The optimal re-expression method described in Chapter 4 is not limited to verbs. The evaluation in Chapter 5 was limited to verbs for reasons explained in Section 5.3.1.1 (i.e. to control the confounding effects that might be created with multiple word replacements in the one requirement). There is nothing that limits the application of the optimal re-expression method to verbs. Other parts of speech (e.g. nouns, adjectives, adverbs, etc) can be sense disambiguated, measured in terms of word occurrences, lexical width, and polysemy. Furthermore, the similarity measures described in Section 4.5.5.2 are equally applicable to any part of speech. Despite the theoretical expectation that the optimal re-expression method could be generalised to

work for any part of speech, there are practical current limitations of technology that prevent this. Specifically, our chosen semantic network *WordNet* does not contain hypernymic relationships for parts of speech other than nouns and verbs. According to Miller (1990), this extension is a work in progress, particularly for adjectives and adverbs. Therefore, due to technological limitations in semantic networks, we can only generalise this method to the nouns and verbs in requirements. Fortunately nouns and verbs are the most polysemous parts of speech (Oxford English Dictionary 2005), and are therefore most likely to be ambiguous.

2) In Section 1.4 we limited the scope of our evaluation to a single industry specification from a single application domain (i.e. the Rail Industry). We do not evaluate other specifications from other application domains – yet we claim that the optimal re-expression method described in Chapter 4 is not limited to the Rail Industry. Fundamentally, the method is not tied to any particular domain, but can be more effective if two key factors are considered when generalising to other application domains.

   a. The size of the specification, in terms of the number of requirements is a key factor to consider. This is because a requirement is defined in Section 2.3.1 as *an externally observable behaviour or property of a system or service that is needed, wanted, or desired by some user or stakeholder*, and since a requirement should not be duplicated within a specification, we might assume that the more requirements, the more nouns to describe the *systems or services*; the more verbs to describe the *externally observable behaviour* and the more adjectives to describe the *properties*. Thus, the more requirements, the more potential replacements, and the more effective the optimal re-expression method. Also the overhead of applying the optimal re-expression method is probably only justified for larger projects with many requirements.

   b. The domain-specificity of the words used in requirements is a key factor to consider. This is because the more domain specific the words, the less likely the words will exist in standard dictionaries, and so the word cannot be sense tagged or compared to other words for similarity measurements. Thus the less domain-specific the language used in requirements, the more effective the optimal re-expression method.

3) Not being familiar with many other languages, we are careful not to claim that the optimal re-expression method can be generalised to specifications written in other languages, other than English (e.g. other indo-european languages). However, we do not see any reason why we should not make such a claim on the basis that the method is insensitive to word position in a sentence, so variations in grammatical structure between English and other languages are not important. In fact, variations in grammar between English and other languages are more likely to affect the success of auto-WSD tools since context windows are important, so our decision in Section 4.5.1.3 to manually disambiguate word sense has perhaps improved our ability to generalise the method to specifications written in other languages. Note also that WordNet is now available in other languages, which further simplifies and encourages this generalisation, since this means sense tags, hypernyms, synonyms, etc, are available, and presumably WordNet::Similarity and the prototype software we have developed would continue to operate, however this is not within the scope of our evaluation, but is postulated as an argument for generalisation.

5.3.12. Evaluation Limitations

Pragmatic limitations have been applied to evaluate whether the design artifact is indeed *capable* of solving the research problem. Without these self-imposed restrictions, the scope of design evaluation would become unwieldy – jeopardising the integrity of the results.

1) We do not empirically evaluate whether optimal(W_r, context(W_o)) is indeed the most optimal replacement. I.e. we could evaluate different variations of the replaceability equation to empirically confirm that W_r is indeed the most optimal replacement for W_o. Instead we limit the evaluation to the detailed theoretical rationale that was provided in Section 4.5.6 on the calculation of replaceability(W_1, W_2).

2) We do not evaluate the effect of simultaneously constraining multiple parts of speech. We made this decision in Section 5.3.1.1 to ensure that the evaluation of a single replacement is made in the context of a static requirement, rather than in the confounding context of a requirement comprised of other replacements. We have decided to limit the evaluation to an optimal re-expression of verbs.

3) We do not observe whether other people, other than the researcher can perform the optimal-constraint process. Instead, we validate that the senses manually disambiguated

by the researcher *make sense* to the questionnaire respondents. This supports the claim that other people would probably achieve the same result as the researcher. Refer to Section 5.4.2.1 for the result.

4) We do not formally test (e.g. white-box, black-box) the implementation. Since the implementation is a software prototype, we have performed informal unit testing, but have not applied a formal test regime. We suggest formal testing would be a necessary activity in the process of converting the software prototype into a commercial product.

5.4. Evaluation Results

This section reports on, and discusses the results of the evaluation. The evaluation has been executed in accordance with the evaluation method described in Section 4.5. Specifically, this section reports on:

- the results of manual word sense disambiguation
- the results of executing the software prototype
- the results of the questionnaire
- the results of the hypothesis testing

5.4.1. Specification (S_o)

For reasons of convenience, accessibility, variety, currency, interest, and generalisability of results (refer Section 5.3.11.2.1) we selected the specification for an 8-car air-conditioned suburban electric double-deck passenger train to operate on the Sydney electrified rail network as the natural English requirement specification to be optimally re-expressed and evaluated. The trains are being procured by the Rail Corporation (RailCorp) of New South Wales under a Public Private Partnership (PPP) effective from 7th Dec 2006. The Train Performance Specification (TPS) (RailCorp Contract C01645 2006) contains some 500+ functional and non-functional requirements that altogether specify the required characteristics, behaviour, and performance of the new suburban electric trains.

We decided to use the TPS as the requirement specification to support the evaluation of our optimal-constraint process. We believe this decision is justified for the following key reasons:

1) Depth and Breadth of Specification: The TPS specifies a complex integration of electrical, mechanical, pneumatic, communications, and software related systems for the new passenger train.

2) Multi-Authored: The TPS was authored and reviewed by a variety of Subject Matter Experts (SMEs), thus a random sample will likely yield data from a number of requirements authors in the rail industry.
3) Specification Accessibility: The TPS is freely available for download in the public domain (refer hyperlink in (RailCorp Contract C01645 2006)).
4) SME Accessibility: The researcher is employed by the core contractor for this project and therefore has relatively unrestricted access to SMEs for the purpose of clarifying the meaning of requirements, conducting questionnaires, etc.
5) Currency: This is a relatively recent project awarded in 2006 and due for completion in 2013, so can be expected to reasonably reflect current best practice in requirements specifications in a particular industry (i.e. the rail industry).

5.4.2. Results of Manual Word Sense Disambiguation

We decided in Section 4.5.1.3 that words would be disambiguated manually by domain knowledgeable humans using WordNet to provide sense definitions. We decided in Section 5.3.1.1 that we would limit the evaluation to a single part of speech – verbs. We decided in Section 5.4.1 that we would use the Train Performance Specification (RailCorp Contract C01645 2006) as S_o for the purpose of evaluating the design solution. The results were as follows:

Table 4 - Disambiguation Metrics

Disambiguation Metric	Value
Total # of requirements	561 requirements
Total # of verbs disambiguated	1025 verbs
Avg. # of verbs per requirement	1.83 verbs / requirement
Total time for verb sense disambiguation	100 hours
Avg. time per verb sense disambiguation	5.85 minutes

We prepared a tab delimited text file in the format of Table 5 for input to the Software Prototype. Note: Table 5 is non-exhaustive and is indicative of format only. It took the primary researcher roughly 100 hours to prepare Table 5 for the complete TPS. The majority of this time was spent disambiguating word senses. Thus our decision in Section 4.5.1.3 "to minimise this confounding effect by using humans to conduct the WSD of requirements" has added approximately 100 hours to the optimal re-expression process. We remind the reader, however, that CNLs can take 5-6 years to implement (refer 2.5.1.1), so the additional time is not considered to be significant.

Table 5 - File Format for Input to Software Prototype

Requirement ID	Word	PoS	Sense*
TPS.3.1.Safety.5	access	v	2
TPS.3.1.Safety.5	operate	v	3
TPS.3.2.1.Trackside.Noise.Levels.1	reduce	v	1
TPS.3.2.1.Trackside.Noise.Levels.2	exceed	v	1
TPS.3.2.1.Trackside.Noise.Levels.3	exceed	v	1
TPS.3.2.2.Vibration.1	minimise	v	2
TPS.3.2.2.Vibration.1	meet	v	4
TPS.3.2.3.Materials.and.End.of.Life.Disposal.1	use	v	1
TPS.3.2.3.Materials.and.End.of.Life.Disposal.1	cause	v	1

...

*Senses from WordNet (2003)

5.4.2.1. Validation of Manual Word Sense Disambiguation

In Section 5.3.12 we stated that we do not observe whether other people, other than the researcher, can perform the optimal-constraint process. Instead, we validate that the senses manually disambiguated by the researcher *make sense* to the questionnaire respondents. This supports the claim that other people would probably achieve the same result as the researcher. Section 5.3.6.1.1 explains how the questionnaire collects data on what makes sense to the respondent. We can therefore validate that the researcher's disambiguation of W_o makes sense to the questionnaire respondents in the context of W_o.

In Section 5.3.7 we decided on a sample size of 40 replacements. Each of these 40 replacements can be traced back to an original word (W_o) that was manually disambiguated by the researcher before being optimally replaced. In Section 5.3.6.1.1 we ask the questionnaire respondents to identify the senses of W_o that *make sense* to him or her, given the context(W_o). Therefore, for each of the 40 replacements in the sample, we can compare the sense manually disambiguated by the researcher, to the sense(s) that *make sense* to each of the five questionnaire respondents. Figure 9 illustrates that in most cases the questionnaire respondents agreed that the researcher's disambiguation of W_o makes sense in context(W_o):

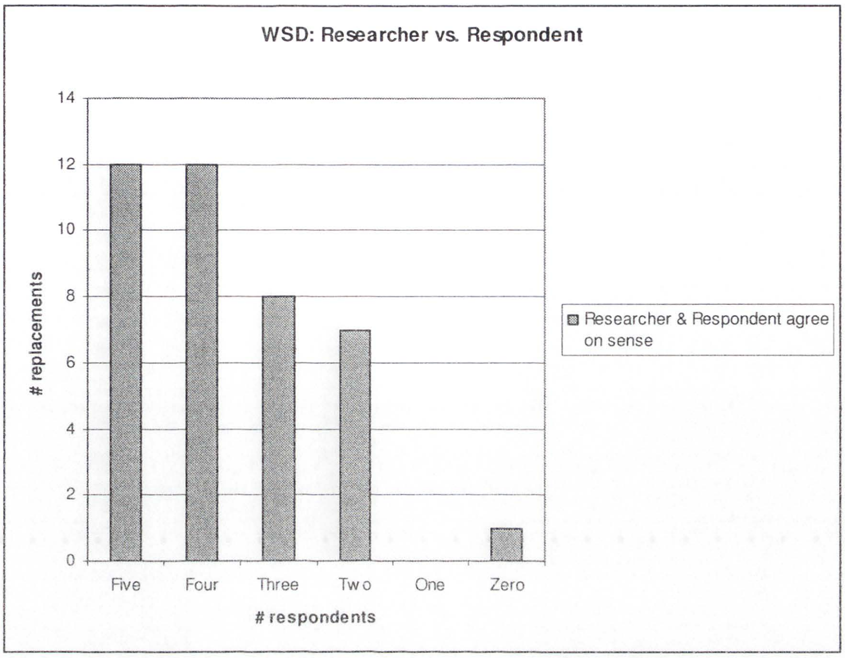

Figure 9 - WSD: Researcher vs. Respondent

We have therefore validated the researcher's sense disambiguations:

- The majority of questionnaire respondents (i.e. at least 3 out of 5) *agreed* with 80% (i.e. 12+12+8/40) of the researcher's sense disambiguations.
- Only one of the researcher's sense disambiguations did not make sense to any of the respondents. Incidentally, this was the disambiguation of the verb "*disconnected*" in the requirement "The EAPS shall be *disconnected* from the Main Power Supply in Fault or Failure conditions." None of the participants agreed with the researcher's chosen sense: "*of electrical appliances*". In hindsight, the researcher agrees with the respondents that a better sense definition would have been: "*make disconnected, disjoin or unfasten*". I.e. this exception relates to a human-error in the manual-WSD and is not a defect of the model, method, or implementation.

We acknowledge that this validation is limited. In particular the researcher was limited to one disambiguated meaning per word, whereas the respondent was unlimited in the number of meanings that could make sense to them. This means that although the respondent might agree with the researcher that a particular meaning makes sense, the respondent might not have picked that meaning had they of been limited to one meaning per word. Thus this validation is really only useful in telling us where the respondents did not agree with the researcher's WSD (i.e. identifying the exceptions).

5.4.3. Execution of the Software Prototype

The software prototype (refer to Appendix F: Prototype Software) was executed with:

1) Requirement Specification: Verbs from TPS (refer 5.4.2)
2) Semantic Network: WordNet (refer 4.5.1.3)
3) Similarity Measure: Wu & Palmer (refer 4.5.5.2)
4) Similarity Threshold: 0, 0.25, 0.5, 0.75, and 1.0 (refer 5.3.6.1.3)

The full population of replacements that result from executing the software prototype on some 500+ verbs across numerous similarity thresholds is too voluminous to duplicate in this Thesis. Instead, we present the summary metrics, and report in detail on the random sample of replacements that has been selected as the basis of our questionnaire and hypothesis testing.

Figure 10 graphically reports four key metrics on the output of the software prototype.

1) *total repl.* shows the total number of replacements that occurred at each similarity threshold. The *total repl.* metric appears to support our expectation (in Section 4.5.6) that the lower the similarity threshold, the more potential replacements, the more likely we will find an alternative word with replaceability greater than 1. Note that *total repl.* reports the total number of original verbs (W_o) that were optimally replaced (W_r). This includes duplicates – i.e. where the same (W_o) is being replaced with the same (W_r), but the context requirement is different.
2) *total repl. (unique)* is similar to above, however if there are repeat occurrences of an original verb (W_o) being replaced by a replacement verb (W_r), it is only counted once regardless of the number of times this same replacement occurs across the set of requirements.
3) *lexicon size* is the total number of verbs needed to express (S_r) the replacement specification. The *lexicon size* metric appears to support our expectation (in Section 4.5.6) that as the similarity threshold is reduced, the number of replacements increases, and the lexicon size decreases. I.e. the replaceability equation (refer 4.5.6) works to reuse the more conventional replacement verbs (W_r) (e.g. by considering frequency) rather than introducing more unique replacements.
4) *Non-occurring dictionary* is the number of replacements that are dictionary words that did not occur anywhere in the original specification (S_o). This can be compared to *total repl.* to determine the ratio of occurring vs. non-occurring replacements at each similarity threshold.

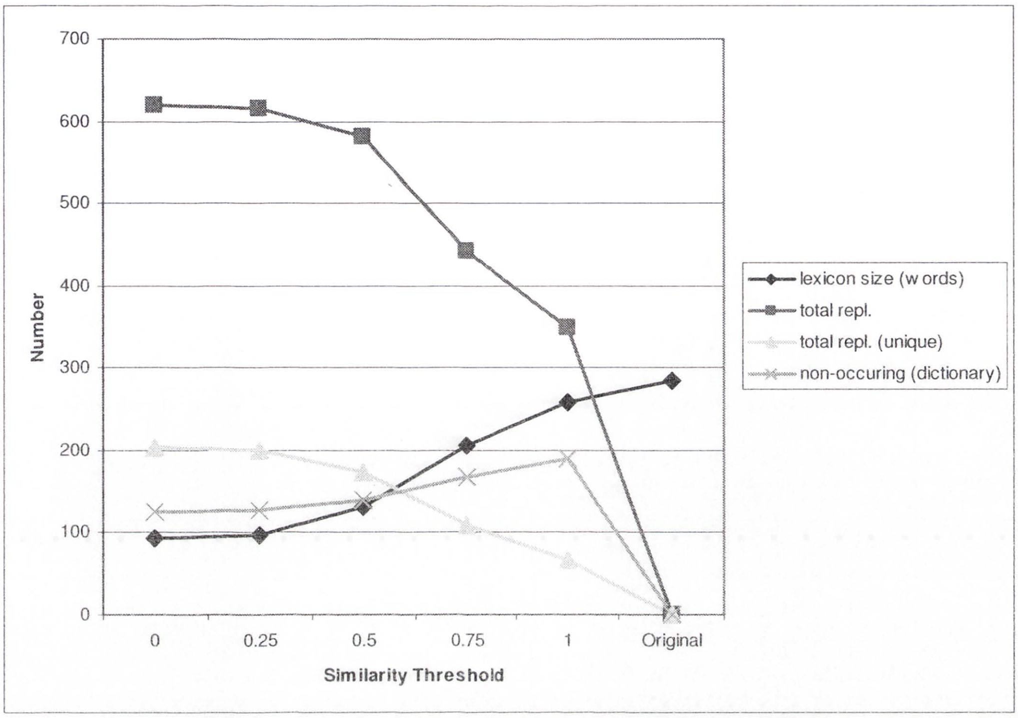

Figure 10 - Optimal Constraint at Similarity Thresholds

Figure 10 confirms the negative relationship between the Similarity Threshold and the number of replacements presented in the Causal Model in Section 4.5.6. I.e. as the Similarity Threshold increases from 0 to 1, the total number of replacements (total repl.) decreases from 620 to 348 (i.e. roughly halves). Figure 10 also confirms that the magnitude of a lexicon can indeed be reduced by reducing the Similarity Threshold. Figure 10 shows that roughly 1/3 of the original words are removed from the lexicon by using a Similarity Threshold of 0.75, and roughly 2/3rds of the original words are removed from the lexicon by using a Similarity Threshold of 0.25. The reduction in lexicon size correlates with an increase in the number of replacements – which is logical, since original words (W_o) are removed from a lexicon because they are re-expressed by replacement (W_r) words.

Figure 10 shows the number of non-occurring dictionary replacements that comprise the total number of replacements (total repl.). Interestingly, as the Similarity Threshold increases from 0 to 1, the percentage of non-occurring dictionary words comprising the total number of replacements increases from 20% (125/620) to 54% (189/348). This is logical since at higher similarity thresholds, there are fewer potential replacements occurring in the original specification (refer Section 4.5.6), and it is more likely that similar words in the dictionary (e.g. synonyms) would maximise the replaceability calculation.

5.4.4. The Sample

Table 6 shows the sample of 40 requirements that we randomly selected from the TPS in accordance with Section 5.3.7. Table 6 shows the "Original" requirement text followed by the "Replacement" requirement text automatically selected by the Software Prototype at each of the Similarity Thresholds (i.e. 0, 0.25, 0.5, 0.75, and 1). The sample formed the basis of our questionnaire to measure the correctness, conventionality, and lexical ambiguity of:

1. the original word (W_o),
2. the replacement word given a similarity threshold of 0 ($W_{r(0)}$),
3. the replacement word given a similarity threshold of 0.25 ($W_{r(0.25)}$),
4. the replacement word given a similarity threshold of 0.5 ($W_{r(0.5)}$),
5. the replacement word given a similarity threshold of 0.75 ($W_{r(0.75)}$),
6. the replacement word given a similarity threshold of 1 ($W_{r(1)}$).

Notice that:

- The same replacement verb often occurs at more than one Similarity Threshold, indicating that the granularity of similarity threshold speculated in Section 5.3.6.1.3 was not too coarse.
- Non-occurring dictionary replacements have been double underlined. Non-occurring dictionary verbs are more prevalent as replacements at higher Similarity Thresholds.

Table 6 - Random Sample of 40 Requirements

#	Reqt ID	Original (W_o)	Replacements				
			$W_{r(0)}$	$W_{r(0.25)}$	$W_{r(0.5)}$	$W_{r(0.75)}$	$W_{r(1)}$
1.	TPS.3.5.6.3.Cabling.1	PPP Co shall **establish#v#6** the cabling for a redundant high-speed data network.	PPP Co shall **use#v#1** the cabling for a redundant high-speed data network.	PPP Co shall **use#v#1** the cabling for a redundant high-speed data network.	PPP Co shall **use#v#1** the cabling for a redundant high-speed data network.	PPP Co shall **install#v#3** the cabling for a redundant high-speed data network.	PPP Co shall **install#v#3** the cabling for a redundant high-speed data network.
2.	TPS.9.7.1.Compressor.Management.1	All compressors shall be controlled to **maximize#v#1** compressor life and optimise the efficiency of the air dryer system.	All compressors shall be controlled to **increase#v#2** compressor life and optimise the efficiency of the air dryer system.	All compressors shall be controlled to **increase#v#2** compressor life and optimise the efficiency of the air dryer system.	All compressors shall be controlled to **increase#v#2** compressor life and optimise the efficiency of the air dryer system.	All compressors shall be controlled to **increase#v#2** compressor life and optimise the efficiency of the air dryer system.	All compressors shall be controlled to **maximize#v#1** compressor life and optimise the efficiency of the air dryer system.
3.	TPS.3.5.6.3.Cabling.2	PPP Co shall **provide#v#1** cabling between CTIP mounting points so that CTIP head and headless units can communicate.	PPP Co shall **supply#v#1** cabling between CTIP mounting points so that CTIP head and headless units can communicate.	PPP Co shall **supply#v#1** cabling between CTIP mounting points so that CTIP head and headless units can communicate.	PPP Co shall **supply#v#1** cabling between CTIP mounting points so that CTIP head and headless units can communicate.	PPP Co shall **supply#v#1** cabling between CTIP mounting points so that CTIP head and headless units can communicate.	PPP Co shall **supply#v#1** cabling between CTIP mounting points so that CTIP head and headless units can communicate.
4.	TPS.4.5.Rolling.Stock.Outline.2	PPP Co shall perform tests to **demonstrate#v#2** that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.	PPP Co shall perform tests to **prove#v#2** that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.	PPP Co shall perform tests to **prove#v#2** that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.	PPP Co shall perform tests to **prove#v#2** that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.	PPP Co shall perform tests to **prove#v#2** that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.	PPP Co shall perform tests to **prove#v#2** that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.
5.	TPS.3.5.6.1.Configuration.1	PPP Co must make provision to **fit#v#8** CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.	PPP Co must make provision to **supply#v#1** CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.	PPP Co must make provision to **supply#v#1** CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.	PPP Co must make provision to **supply#v#1** CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.	PPP Co must make provision to **supply#v#1** CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.	PPP Co must make provision to **equip#v#1** CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.
6.	TPS.5.2.1.Energy.Consumption.Prediction.3	The Actual Energy Consumption of a Set shall be determined and **demonstrated#v#2** by PPP Co using a	The Actual Energy Consumption of a Set shall be determined and **proven#v#2** by PPP Co using a combination of	The Actual Energy Consumption of a Set shall be determined and **proven#v#2** by PPP Co using a combination of	The Actual Energy Consumption of a Set shall be determined and **proven#v#2** by PPP Co using a combination of	The Actual Energy Consumption of a Set shall be determined and **proven#v#2** by PPP Co using a combination of	The Actual Energy Consumption of a Set shall be determined and **proven#v#2** by PPP Co using a combination of

#	Reqt ID	Original (W_o)	Replacements				
			$W_{r(0)}$	$W_{r(0.25)}$	$W_{r(0.5)}$	$W_{r(0.75)}$	$W_{r(1)}$
		combination of simulation and testing.	simulation and testing.	simulation and testing.	simulation and testing.	simulation and testing.	simulation and testing.
7.	TPS.5.3.Security.1	The Set shall **provide#v#5** three levels of locked security access.	The Set shall **supply#v#1** three levels of locked security access.	The Set shall **supply#v#1** three levels of locked security access.	The Set shall **supply#v#1** three levels of locked security access.	The Set shall **supply#v#1** three levels of locked security access.	The Set shall **allow for#v#1** three levels of locked security access.
8.	TPS.6.2.19.Internal.Noise.2	PPP Co shall **undertake#v#1** any reasonable resulting corrective actions on all Sets to rectify the cause of the complaint.	PPP Co shall **attempt#v#2** any reasonable resulting corrective actions on all Sets to rectify the cause of the complaint.	PPP Co shall **attempt#v#2** any reasonable resulting corrective actions on all Sets to rectify the cause of the complaint.	PPP Co shall **attempt#v#2** any reasonable resulting corrective actions on all Sets to rectify the cause of the complaint.	PPP Co shall **attempt#v#2** any reasonable resulting corrective actions on all Sets to rectify the cause of the complaint.	PPP Co shall **attempt#v#2** any reasonable resulting corrective actions on all Sets to rectify the cause of the complaint.
9.	TPS.5.3.Security.2	An Access 1 Security key shall **operate#v#3** Access 1 Security only.	An Access 1 Security key shall **control#v#3** Access 1 Security only.	An Access 1 Security key shall **control#v#3** Access 1 Security only.	An Access 1 Security key shall **control#v#3** Access 1 Security only.	An Access 1 Security key shall **control#v#3** Access 1 Security only.	An Access 1 Security key shall **control#v#3** Access 1 Security only.
10.	TPS.6.2.14.7.Loads.on.Car.Body.Attachments.1	The coupler assembly shall **have#v#2** a built in energy absorbing device.	The coupler assembly shall **feature#v#1** a built in energy absorbing device.	The coupler assembly shall **feature#v#1** a built in energy absorbing device.	The coupler assembly shall **feature#v#1** a built in energy absorbing device.	The coupler assembly shall **feature#v#1** a built in energy absorbing device.	The coupler assembly shall **feature#v#1** a built in energy absorbing device.
11.	TPS.9.1.Electrical.Auxiliary.Power.Supply.1	The EAPS shall be **disconnected#v#1** from the Main Power Supply in Fault or Failure conditions.	The EAPS shall be **switched#v#1** from the Main Power Supply in Fault or Failure conditions.	The EAPS shall be **switched#v#1** from the Main Power Supply in Fault or Failure conditions.	The EAPS shall be **switched#v#1** from the Main Power Supply in Fault or Failure conditions.	The EAPS shall be **unplugged#v#1** from the Main Power Supply in Fault or Failure conditions.	The EAPS shall be **unplugged#v#1** from the Main Power Supply in Fault or Failure conditions.
12.	TPS.9.9.2.Spanner.Size.1	All flexible hoses shall be **fitted#v#8** with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.	All flexible hoses shall be **supplied#v#1** with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.	All flexible hoses shall be **supplied#v#1** with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.	All flexible hoses shall be **supplied#v#1** with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.	All flexible hoses shall be **supplied#v#1** with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.	All flexible hoses shall be **equipped#v#1** with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.
13.	TPS.9.12.Passenger.Exit.Light.1	The light shall not visually **hinder#v#1** Passengers using the doorway.	The light shall not visually **supply#v#1** Passengers using the doorway.	The light shall not visually **supply#v#1** Passengers using the doorway.	The light shall not visually **interfere#v#1** Passengers using the doorway.	The light shall not visually **interfere#v#1** Passengers using the doorway.	The light shall not visually **impede#v#1** Passengers using the doorway.
14.	TPS.9.15.Timetable.and	Details of the task lighting including switching, fitting	Details of the task lighting including switching, fitting	Details of the task lighting including switching, fitting	Details of the task lighting including switching, fitting	Details of the task lighting including switching, fitting	Details of the task lighting including switching, fitting

#	Reqt ID	Original (W_o)	Replacements				
			$W_{r(0)}$	$W_{r(0.25)}$	$W_{r(0.5)}$	$W_{r(0.75)}$	$W_{r(1)}$
	.Task.Lighting.2	type and area coverage shall be **determined#v#1** during the Contract.	type and area coverage shall be **used#v#1** during the Contract.	type and area coverage shall be **used#v#1** during the Contract.	type and area coverage shall be **used#v#1** during the Contract.	type and area coverage shall be **found_out#v#1** during the Contract.	type and area coverage shall be **found_out#v#1** during the Contract.
15.	TPS.9.20.1.Headlight.Low.Beam.1	Each headlight shall **have#v#2** a low beam providing not greater than 20% of the rated light output of the headlight.	Each headlight shall **feature#v#1** a low beam providing not greater than 20% of the rated light output of the headlight.	Each headlight shall **feature#v#1** a low beam providing not greater than 20% of the rated light output of the headlight.	Each headlight shall **feature#v#1** a low beam providing not greater than 20% of the rated light output of the headlight.	Each headlight shall **feature#v#1** a low beam providing not greater than 20% of the rated light output of the headlight.	Each headlight shall **feature#v#1** a low beam providing not greater than 20% of the rated light output of the headlight.
16.	TPS.9.25.6.Code.of.Practice.1	PPP Co shall **follow#v#5** HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.	PPP Co shall **comply_with#v#1** HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.	PPP Co shall **comply_with#v#1** HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.	PPP Co shall **comply_with#v#1** HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.	PPP Co shall **comply_with#v#1** HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.	PPP Co shall **comply_with#v#1** HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.
17.	TPS.9.25.15.Drip.trays.and.drainage.1	The design of drains shall ensure they cannot be **blocked#v#2** by debris.	The design of drains shall ensure they cannot be **prevented#v#1** by debris.	The design of drains shall ensure they cannot be **prevented#v#1** by debris.	The design of drains shall ensure they cannot be **prevented#v#1** by debris.	The design of drains shall ensure they cannot be **prevented#v#1** by debris.	The design of drains shall ensure they cannot be **obstructed#v#1** by debris.
18.	TPS.10.1.1.3.Emergency.(get.home).1	The Set shall be able to **start#v#9** on the steepest gradient.	The Set shall be able to **use#v#1** on the steepest gradient.	The Set shall be able to **use#v#1** on the steepest gradient.	The Set shall be able to **use#v#1** on the steepest gradient.	The Set shall be able to **get_going#v#1** on the steepest gradient.	The Set shall be able to **get_going#v#1** on the steepest gradient.
19.	TPS.10.2.3.Motor.Cut-Out.Operation.4	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall be able to **operate#v#2** an all stations run pattern without overheating the traction equipment when using maximum available	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall be able to **function#v#1** an all stations run pattern without overheating the traction equipment when using maximum available	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall be able to **function#v#1** an all stations run pattern without overheating the traction equipment when using maximum available	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall be able to **function#v#1** an all stations run pattern without overheating the traction equipment when using maximum available	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall be able to **function#v#1** an all stations run pattern without overheating the traction equipment when using maximum available	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall be able to **function#v#1** an all stations run pattern without overheating the traction equipment when using maximum available

#	Reqt ID	Original (W_o)	Replacements				
			$W_{r(0)}$	$W_{r(0.25)}$	$W_{r(0.5)}$	$W_{r(0.75)}$	$W_{r(1)}$
		traction and braking.	traction and braking.	traction and braking.	traction and braking.	traction and braking.	traction and braking.
20.	TPS.10.3.15.2.Emergency.Brake.3	The Emergency Brake shall be applied in the event of the Set **passing#v#1** a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.	The Emergency Brake shall be applied in the event of the Set **supplying#v#1** a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.	The Emergency Brake shall be applied in the event of the Set **supplying#v#1** a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.	The Emergency Brake shall be applied in the event of the Set **crossing#v#1** a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.	The Emergency Brake shall be applied in the event of the Set **crossing#v#1** a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.	The Emergency Brake shall be applied in the event of the Set **going_through#v#3** a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.
21.	TPS.10.3.16.4.Park.Brake.Car.Indication.1	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control **showing#v#4** the Park Brake status for both bogies.	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control **using#v#1** the Park Brake status for both bogies.	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control **using#v#1** the Park Brake status for both bogies.	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control **using#v#1** the Park Brake status for both bogies.	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control **displaying#v#1** the Park Brake status for both bogies.	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control **showing#v#4** the Park Brake status for both bogies.
22.	TPS.6.2.2.Width.1	The Car interior space shall be **maximized#v#1** with a Car interior width measured across the full width of the Car, of at least 2920 mm.	The Car interior space shall be **increased#v#2** with a Car interior width measured across the full width of the Car, of at least 2920 mm.	The Car interior space shall be **increased#v#2** with a Car interior width measured across the full width of the Car, of at least 2920 mm.	The Car interior space shall be **increased#v#2** with a Car interior width measured across the full width of the Car, of at least 2920 mm.	The Car interior space shall be **increased#v#2** with a Car interior width measured across the full width of the Car, of at least 2920 mm.	The Car interior space shall be **maximized#v#1** with a Car interior width measured across the full width of the Car, of at least 2920 mm.
23.	TPS.11.4.3.Incoming.Calls.4	The Crew member taking the PEI call shall be able to **hold#v#2** that call and operate the Intercom to the other Crew member.	The Crew member taking the PEI call shall be able to **maintain#v#1** that call and operate the Intercom to the other Crew member.	The Crew member taking the PEI call shall be able to **maintain#v#1** that call and operate the Intercom to the other Crew member.	The Crew member taking the PEI call shall be able to **maintain#v#1** that call and operate the Intercom to the other Crew member.	The Crew member taking the PEI call shall be able to **maintain#v#1** that call and operate the Intercom to the other Crew member.	The Crew member taking the PEI call shall be able to **maintain#v#1** that call and operate the Intercom to the other Crew member.
24.	TPS.11.5.4.PA.Inputs.and.Priority.Order.3	The minimum volume level shall be agreed in final commissioning or, if not agreed, **determined#v#1** by RailCorp's Representative.	The minimum volume level shall be agreed in final commissioning or, if not agreed, **used#v#1** by RailCorp's Representative.	The minimum volume level shall be agreed in final commissioning or, if not agreed, **used#v#1** by RailCorp's Representative.	The minimum volume level shall be agreed in final commissioning or, if not agreed, **used#v#1** by RailCorp's Representative.	The minimum volume level shall be agreed in final commissioning or, if not agreed, **found_out#v#1** by RailCorp's Representative.	The minimum volume level shall be agreed in final commissioning or, if not agreed, **found_out#v#1** by RailCorp's Representative.
25.	TPS.6.8.4.Offset.Angled.Collisions.2	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall **consider#v#8** the	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall **look_at#v#2** the inclusion	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall **look_at#v#2** the inclusion	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall **look_at#v#2** the inclusion	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall **look_at#v#2** the inclusion	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall **look_at#v#2** the inclusion

#	Reqt ID	Original (W_o)	Replacements				
			$W_{r(0)}$	$W_{r(0.25)}$	$W_{r(0.5)}$	$W_{r(0.75)}$	$W_{r(1)}$
		inclusion of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.	of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.	of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.	of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.	of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.	of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.
26.	TPS.11.5.6. PA.modelling.1	An electro-acoustic virtual model of the carriage shall be **produced#v#2** to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.	An electro-acoustic virtual model of the carriage shall be **used#v#1** to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.	An electro-acoustic virtual model of the carriage shall be **used#v#1** to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.	An electro-acoustic virtual model of the carriage shall be **used#v#1** to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.	An electro-acoustic virtual model of the carriage shall be **generated#v#2** to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.	An electro-acoustic virtual model of the carriage shall be **produced#v#2** to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.
27.	TPS.9.20.2. Headlight.High.Beam.1	Each headlight shall **have#v#2** a high beam providing the full rated light output of the headlight.	Each headlight shall **feature#v#1** a high beam providing the full rated light output of the headlight.	Each headlight shall **feature#v#1** a high beam providing the full rated light output of the headlight.	Each headlight shall **feature#v#1** a high beam providing the full rated light output of the headlight.	Each headlight shall **feature#v#1** a high beam providing the full rated light output of the headlight.	Each headlight shall **feature#v#1** a high beam providing the full rated light output of the headlight.
28.	TOS_Appendix.K.111	Level 1 diagnostic results shall include recommended remedial action required to be **undertaken#v#1** by the Crew	Level 1 diagnostic results shall include recommended remedial action required to be **attempted#v#2** by the Crew	Level 1 diagnostic results shall include recommended remedial action required to be **attempted#v#2** by the Crew	Level 1 diagnostic results shall include recommended remedial action required to be **attempted#v#2** by the Crew	Level 1 diagnostic results shall include recommended remedial action required to be **attempted#v#2** by the Crew	Level 1 diagnostic results shall include recommended remedial action required to be **attempted#v#2** by the Crew
29.	TPS.11.11.Digital.Voice.Annunciator.10	The DVA equipment shall be simple to **operate#v#3** and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.	The DVA equipment shall be simple to **control#v#3** and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.	The DVA equipment shall be simple to **control#v#3** and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.	The DVA equipment shall be simple to **control#v#3** and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.	The DVA equipment shall be simple to **control#v#3** and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.	The DVA equipment shall be simple to **control#v#3** and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.
30.	TPS.6.2.18. Centre.of.Gravity.and.Roll.Centre.1	PPP Co shall nominate and **show#v#2** by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.	PPP Co shall nominate and **prove#v#2** by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.	PPP Co shall nominate and **prove#v#2** by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.	PPP Co shall nominate and **prove#v#2** by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.	PPP Co shall nominate and **prove#v#2** by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.	PPP Co shall nominate and **prove#v#2** by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.
31.	TPS.7.4.Train.Operating.	The TOS shall be **connected#v#10** to TOS	The TOS shall be **inserted#v#1** to TOS	The TOS shall be **inserted#v#1** to TOS	The TOS shall be **inserted#v#1** to TOS	The TOS shall be **inserted#v#1** to TOS	The TOS shall be **plugged_in#v#1** to TOS

#	Reqt ID	Original (W_o)	Replacements				
			$W_{r(0)}$	$W_{r(0.25)}$	$W_{r(0.5)}$	$W_{r(0.75)}$	$W_{r(1)}$
	System.5	screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.	screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.	screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.	screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.	screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.	screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.
32.	TPS.11.10.2.1.Internal.Passenger.area.coverage.4	The camera locations and coverage shall be **demonstrated#v#1** and agreed during the Mock-up review.	The camera locations and coverage shall be **presented#v#1** and agreed during the Mock-up review.	The camera locations and coverage shall be **presented#v#1** and agreed during the Mock-up review.	The camera locations and coverage shall be **presented#v#1** and agreed during the Mock-up review.	The camera locations and coverage shall be **presented#v#1** and agreed during the Mock-up review.	The camera locations and coverage shall be **presented#v#1** and agreed during the Mock-up review.
33.	TPS.11.10.5.1.CCTV.System.Operation.5	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be **demonstrated#v#1** and agreed during the Mock-up review.	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be **presented#v#1** and agreed during the Mock-up review.	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be **presented#v#1** and agreed during the Mock-up review.	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be **presented#v#1** and agreed during the Mock-up review.	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be **presented#v#1** and agreed during the Mock-up review.	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be **presented#v#1** and agreed during the Mock-up review.
34.	TPS.6.7.1.6.Closing.Time.2	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to **show#v#2** that the door timing meets this requirement.	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to **prove#v#2** that the door timing meets this requirement.	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to **prove#v#2** that the door timing meets this requirement.	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to **prove#v#2** that the door timing meets this requirement.	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to **prove#v#2** that the door timing meets this requirement.	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to **prove#v#2** that the door timing meets this requirement.
35.	TPS.3.2.3.Materials.and.End.of.Life.Disposal.6	PPP Co shall **maximize#v#2** the use of recyclable materials throughout the Set.	PPP Co shall **use#v#1** the use of recyclable materials throughout the Set.	PPP Co shall **use#v#1** the use of recyclable materials throughout the Set.	PPP Co shall **use#v#1** the use of recyclable materials throughout the Set.	PPP Co shall **harness#v#2** the use of recyclable materials throughout the Set.	PPP Co shall **maximize#v#2** the use of recyclable materials throughout the Set.
36.	TPS.11.15.2.Operating.Environment.1	PPP Co shall **consider#v#8** mechanisms which overcome the reception and accuracy problem typically encountered by GPS Systems within the Rail Network.	PPP Co shall **look_at#v#2** mechanisms which overcome the reception and accuracy problem typically encountered by GPS Systems within the Rail Network.	PPP Co shall **look_at#v#2** mechanisms which overcome the reception and accuracy problem typically encountered by GPS Systems within the Rail Network.	PPP Co shall **look_at#v#2** mechanisms which overcome the reception and accuracy problem typically encountered by GPS Systems within the Rail Network.	PPP Co shall **look_at#v#2** mechanisms which overcome the reception and accuracy problem typically encountered by GPS Systems within the Rail Network.	PPP Co shall **look_at#v#2** mechanisms which overcome the reception and accuracy problem typically encountered by GPS Systems within the Rail Network.

#	Reqt ID	Original (W_o)	Replacements				
			$W_{r(0)}$	$W_{r(0.25)}$	$W_{r(0.5)}$	$W_{r(0.75)}$	$W_{r(1)}$
37.	TPS.6.2.8.Passenger.Seats.4	The seat design and layout shall **maximize#v#1** the available legroom for seated passengers.	The seat design and layout shall **increase#v#2** the available legroom for seated passengers.	The seat design and layout shall **increase#v#2** the available legroom for seated passengers.	The seat design and layout shall **increase#v#2** the available legroom for seated passengers.	The seat design and layout shall **increase#v#2** the available legroom for seated passengers.	The seat design and layout shall **maximize#v#1** the available legroom for seated passengers.
38.	TPS.6.5.5.2.Gangway.Interior.3	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to **disconnect#v#2** Cars or remove access panels.	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to **supply#v#1** Cars or remove access panels.	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to **supply#v#1** Cars or remove access panels.	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to **push#v#1** Cars or remove access panels.	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to **detach#v#1** Cars or remove access panels.	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to **disconnect#v#2** Cars or remove access panels.
39.	TOS_Appendix.K.142	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be **undertaken#v#1** by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be **attempt#v#2** by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be **attempt#v#2** by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be **attempt#v#2** by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be **attempt#v#2** by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be **attempt#v#2** by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.
40.	TOS_Appendix.K.99	It shall be possible to **call_up#v#1** the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.	It shall be possible to **supply#v#1** the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.	It shall be possible to **supply#v#1** the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.	It shall be possible to **bring_forward#v#2** the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.	It shall be possible to **bring_forward#v#2** the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.	It shall be possible to **bring_forward#v#2** the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.

5.4.5. Questionnaire Results

Section 5.3.6 explained how the questions of the questionnaire were designed to individually collect a complete set of data on each replacement of the sample. The questionnaire was designed to collect data on 1) correctness, 2) conventionality, and 3) lexical ambiguity of both the original word (W_o) and the replacement words ($W_{r(0)}$, $W_{r(0.25)}$, $W_{r(0.5)}$, $W_{r(0.75)}$, $W_{r(1)}$). The questionnaire that we used has been provided in Appendix D: Questionnaire Pilot Test (Sheet 1) and Appendix E: Questionnaire Pilot Test (Sheet 2) and was based entirely on the sample of 40 requirements presented in Table 6.

5.4.5.1. Respondents

We distributed the questionnaire via email to five stakeholders of the RailCorp PPP project. Figure 11 is a copy of the exact email that was sent. Addressees were BCC'd so that recipients of the email could not know who else had received the questionnaire. This was done for two reasons 1) to maintain the privacy of the participants, 2) to prevent collusion on responses.

From: Stephen Boyd
Sent: Thursday, 3 July 2008 10:45 AM
Subject: Questionnaire Support

You are invited to participate in a questionnaire associated with my PhD research on "Optimally-Expressed Requirements".

This questionnaire is how I collect my data to validate my results – so it is very important to me that you participate.

There are three files attached:

1. Information Sheet explaining some background information
2. Instruction Sheet explaining how to complete the questionnaire
3. The Questionnaire (there are two sheets 1 & 2) – both need to be filled in.

The deadline for completing the questionnaire is Friday 11th July.
The questionnaire should take approximately 2 hrs to complete in total (based on a pilot test).
You don't need to finish it in one sitting – just save and continue later.

I hope you can participate – please confirm with me ASAP.

Note: this is not work-related – so please conduct at the appropriate times.

Thanks – in anticipation of your participation!

Steve

Figure 11 - Email asking for Research Participation

Participants were given 8 days to complete the questionnaire and return via email. Most participants asked for extra time to complete the questionnaire, and the last response was received on 16 July (i.e 13 days elapsed from when the questionnaire was issued, to when the last response was received). None of the participants contacted the researcher for assistance or advice whilst completing the questionnaire. The anecdotal feedback received from participants was that the questionnaire was easily understandable (given Appendix C: Questionnaire Instruction Sheet), however the time needed to complete was underestimated. This was unexpected since the length of the questionnaire was assessed as part of the Pilot Test (Appendix D: Questionnaire Pilot Test (Sheet 1) and Appendix E: Questionnaire Pilot Test (Sheet 2)).

In accordance with our Ethics Application and subsequent Ethics Approval (refer Appendix A: Ethics Approval) we have not identified the respondents by name. Table 7 below shows that all respondents were experienced Engineers having:

- Experience writing requirement specifications
- Experience validating the quality of requirement specifications
- Experience developing a product to meet requirement specifications
- Familiarity with the RailCorp PPP project and the TPS

Table 7 - Experience of Respondents

Respondent	Writing Specifications	Validating Specifications	Developing to Specification	RailCorp PPP and TPS
Test Engineer	3 yrs	5 yrs	2 yrs	2 yrs
Design Engineer	10 yrs	12 yrs	8 yrs	2 yrs
Requirements Eng*	5 yrs	5 yrs	2 yrs	2 yrs
Systems Engineer*	7 yrs	3 yrs	2 yrs	2 yrs
Software Engineer	12 yrs	10 yrs	12 yrs	1 yrs

* respondent works for the researcher (this was highlighted in our ethics application).

5.4.5.2. Data Collected

The questionnaire was electronic, so it was possible to automatically check that all required fields had been completed. Fortunately, all five questionnaires were returned in a complete state. Since each questionnaire contained 40 questions, we collected 200 answers in total.

Whilst the same question was responded to by multiple participants – we did not consolidate multiple answers to the same question into a single answer, since each response is subjective and opinion-based and was considered valid in its own right. For instance, if four of the five respondents deemed the original word (W_o) to be lexically unambiguous, but the fifth respondent deemed (W_o) to be lexically ambiguous – then which answer do we take? We took all answers – since the fifth respondent may have been the software developer who may have subsequently developed the wrong solution.

In Section 5.3.7 we nominated a sample size of 40, so there are 40 optimal replacements per similarity threshold (i.e. $\lvert optimal(W_o,context(W_o)) \rvert = 40$). Summing for the five respondents yields 200 unique responses. Each response contains data pertaining to quality factors: correctness, conventionality, and lexical ambiguity.

5.4.5.2.1. Correctness

Test Statistic	Proportion	Value
$p(W_o)_{correct}$	$\sum_{P=1..5} \lvert makes\_sense(P, W_o, context(W_o)) \rvert \,/\, \lvert optimal(W_o,context(W_o)) \rvert$	179/200
$p(W_r)_{correct(0)}$	$\sum_{P=1..5} \lvert makes\_sense(P, W_{r(0)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0)},context(W_o)) \rvert$	129/200
$p(W_r)_{correct(0.25)}$	$\sum_{P=1..5} \lvert makes\_sense(P, W_{r(0.25)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.25)},context(W_o)) \rvert$	129/200
$p(W_r)_{correct(0.5)}$	$\sum_{P=1..5} \lvert makes\_sense(P, W_{r(0.5)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.5)},context(W_o)) \rvert$	139/200
$p(W_r)_{correct(0.75)}$	$\sum_{P=1..5} \lvert makes\_sense(P, W_{r(0.75)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.75)},context(W_o)) \rvert$	166/200
$p(W_r)_{correct(1)}$	$\sum_{P=1..5} \lvert makes\_sense(P, W_{r(1)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(1)},context(W_o)) \rvert$	175/200

5.4.5.2.2. Conventionality

Test Statistic	Proportion	Value
$p(W_o)_{conventional}$	$\sum_{P=1..5} \lvert would\_use(P, W_o, context(W_o)) \rvert \,/\, \lvert optimal(W_o,context(W_o)) \rvert$	168/200
$p(W_r)_{conventional(0)}$	$\sum_{P=1..5} \lvert would\_use(P, W_{r(0)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0)},context(W_o)) \rvert$	132/200
$p(W_r)_{conventional(0.25)}$	$\sum_{P=1..5} \lvert would\_use(P, W_{r(0.25)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.25)},context(W_o)) \rvert$	132/200
$p(W_r)_{conventional(0.5)}$	$\sum_{P=1..5} \lvert would\_use(P, W_{r(0.5)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.5)},context(W_o)) \rvert$	141/200
$p(W_r)_{conventional(0.75)}$	$\sum_{P=1..5} \lvert would\_use(P, W_{r(0.75)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.75)},context(W_o)) \rvert$	150/200
$p(W_r)_{conventional(1)}$	$\sum_{P=1..5} \lvert would\_use(P, W_{r(1)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(1)},context(W_o)) \rvert$	161/200

5.4.5.2.3. Lexical Unambiguity

Test Statistic	Proportion	Value
$p(W_o)_{unambiguous}$	$\sum_{P=1..5} \lvert unambiguous(P, W_o, context(W_o)) \rvert \,/\, \lvert optimal(W_o,context(W_o)) \rvert$	29/200
$p(W_r)_{unambiguous(0)}$	$\sum_{P=1..5} \lvert unambiguous(P, W_{r(0)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0)},context(W_o)) \rvert$	59/200
$p(W_r)_{unambiguous(0.25)}$	$\sum_{P=1..5} \lvert unambiguous(P, W_{r(0.25)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.25)},context(W_o)) \rvert$	59/200
$p(W_r)_{unambiguous(0.5)}$	$\sum_{P=1..5} \lvert unambiguous(P, W_{r(0.5)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.5)},context(W_o)) \rvert$	51/200
$p(W_r)_{unambiguous(0.75)}$	$\sum_{P=1..5} \lvert unambiguous(P, W_{r(0.75)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(0.75)},context(W_o)) \rvert$	55/200
$p(W_r)_{unambiguous(1)}$	$\sum_{P=1..5} \lvert unambiguous(P, W_{r(1)}, context(W_o)) \rvert \,/\, \lvert optimal(W_{r(1)},context(W_o)) \rvert$	67/200

5.4.6. Hypothesis Testing

5.4.6.1. Z-Test Criteria

Before we can use the Z-Test to test each hypothesis presented in Section 5.3.3, we must first confirm that the pre-conditions for using the Z-Test are satisfied:

1) The sample must be random
2) The test statistic distribution should be approximately normal
3) n1.p1 > 5 and n1.(1 – p1) > 5 and n2.p2 > 5 and n2.(1 – p2) > 5

In our case, 1) is considered satisfied, since in Section 5.3.7 we explained that the sample of replacements used to create the questionnaire was randomly taken from the full set of replacements. 2) is considered satisfied, since if we repeatedly took a random sample from the population of replacements – in each case measuring one of the test statistics (say $p(W_o)_{correct}$ or $p(W_o)_{conventional}$ or $p(W_o)_{unambiguous}$) and plotted the test statistic on a graph, we would expect the shape of the curve to approximate a normal distribution. In other words, we expect the *proportions* to be normally distributed. 3) is considered satisfied, since Table 8 shows that each of the four conditions are met (i.e. they are all > 5).

Table 8 - Z-Test Criteria

CORRECTNESS								
hypothesis	n1	p1	n2	p2	n1.p1 > 5	n1.(1 – p1) > 5	n2.p2 > 5	n2.(1 – p2) > 5
$HO_{correct(0)}$	200	0.645	200	0.895	129	71	179	21
$HO_{correct(0.25)}$	200	0.645	200	0.895	129	71	179	21
$HO_{correct(0.5)}$	200	0.695	200	0.895	139	61	179	21
$HO_{correct(0.75)}$	200	0.83	200	0.895	166	34	179	21
$HO_{correct(1)}$	200	0.875	200	0.895	175	25	179	21

CONVENTIONALITY								
hypothesis	n1	p1	n2	p2	n1.p1 > 5	n1.(1 – p1) > 5	n2.p2 > 5	n2.(1 – p2) > 5
$HO_{conventional(0)}$	200	0.66	200	0.84	132	68	168	32
$HO_{conventional(0.25)}$	200	0.66	200	0.84	132	68	168	32
$HO_{conventional(0.5)}$	200	0.705	200	0.84	141	59	168	32
$HO_{conventional(0.75)}$	200	0.75	200	0.84	150	50	168	32
$HO_{conventional(1)}$	200	0.805	200	0.84	161	39	168	32

UNAMBIGUITY								
hypothesis	n1	p1	n2	p2	n1.p1 > 5	n1.(1 – p1) > 5	n2.p2 > 5	n2.(1 – p2) > 5
$HO_{unambiguous(0)}$	200	0.145	200	0.295	29	171	59	141
$HO_{unambiguous(0.25)}$	200	0.145	200	0.295	29	171	59	141
$HO_{unambiguous(0.5)}$	200	0.145	200	0.255	29	171	51	149
$HO_{unambiguous(0.75)}$	200	0.145	200	0.275	29	171	55	145
$HO_{unambiguous(1)}$	200	0.145	200	0.335	29	171	67	133

5.4.6.2. Hypothesis Test Results

Table 9 - Hypothesis Test Results

CORRECTNESS												
hypothesis	**condition**	n1	p1	n2	p2	$\hat{p}$	$\hat{q}$	s_{p1-p2}	Z	α	Z_{crit}	**Result**
$H0_{correct(0)}$	$p(W_r)_{correct(0)} \geq p(W_o)_{correct}$	200	0.645	200	0.895	0.77	0.23	0.04208	-5.9406	0.01	-2.33	REJECT H0
$H0_{correct(0.25)}$	$p(W_r)_{correct(0.25)} \geq p(W_o)_{correct}$	200	0.645	200	0.895	0.77	0.23	0.04208	-5.9406	0.01	-2.33	REJECT H0
$H0_{correct(0.5)}$	$p(W_r)_{correct(0.5)} \geq p(W_o)_{correct}$	200	0.695	200	0.895	0.795	0.205	0.04037	-4.9542	0.01	-2.33	REJECT H0
$H0_{correct(0.75)}$	$p(W_r)_{correct(0.75)} \geq p(W_o)_{correct}$	200	0.83	200	0.895	0.8625	0.1375	0.03444	-1.8875	0.01	-2.33	SUPPORT H0
$H0_{correct(1)}$	$p(W_r)_{correct(1)} \geq p(W_o)_{correct}$	200	0.875	200	0.895	0.885	0.115	0.0319	-0.6269	0.01	-2.33	SUPPORT H0
CONVENTIONALITY												
hypothesis	**condition**	n1	p1	n2	p2	$\hat{p}$	$\hat{q}$	s_{p1-p2}	Z	α	Z_{crit}	**Result**
$H0_{conventional(0)}$	$p(W_r)_{conventional(0)} \geq p(W_o)_{conventional}$	200	0.66	200	0.84	0.75	0.25	0.0433	-4.1569	0.01	-2.33	REJECT H0
$H0_{conventional(0.25)}$	$p(W_r)_{conventional(0.25)} \geq p(W_o)_{conventional}$	200	0.66	200	0.84	0.75	0.25	0.0433	-4.1569	0.01	-2.33	REJECT H0
$H0_{conventional(0.5)}$	$p(W_r)_{conventional(0.5)} \geq p(W_o)_{conventional}$	200	0.705	200	0.84	0.7725	0.2275	0.04192	-3.2203	0.01	-2.33	REJECT H0
$H0_{conventional(0.75)}$	$p(W_r)_{conventional(0.75)} \geq p(W_o)_{conventional}$	200	0.75	200	0.84	0.795	0.205	0.04037	-2.2294	0.01	-2.33	SUPPORT H0
$H0_{conventional(1)}$	$p(W_r)_{conventional(1)} \geq p(W_o)_{conventional}$	200	0.805	200	0.84	0.8225	0.1775	0.03821	-0.916	0.01	-2.33	SUPPORT H0
UNAMBIGUITY												
hypothesis	**condition**	n1	p1	n2	p2	$\hat{p}$	$\hat{q}$	s_{p1-p2}	Z	α	Z_{crit}	**Result**
$H0_{unambiguous(0)}$	$p(W_r)_{unambiguous(0)} \leq p(W_o)_{unambiguous}$	200	0.145	200	0.295	0.22	0.78	0.04142	-3.621	0.01	-2.33	REJECT H0
$H0_{unambiguous(0.25)}$	$p(W_r)_{unambiguous(0.25)} \leq p(W_o)_{unambiguous}$	200	0.145	200	0.295	0.22	0.78	0.04142	-3.621	0.01	-2.33	REJECT H0
$H0_{unambiguous(0.5)}$	$p(W_r)_{unambiguous(0.5)} \leq p(W_o)_{unambiguous}$	200	0.145	200	0.255	0.2	0.8	0.04	-2.75	0.01	-2.33	REJECT H0
$H0_{unambiguous(0.75)}$	$p(W_r)_{unambiguous(0.75)} \leq p(W_o)_{unambiguous}$	200	0.145	200	0.275	0.21	0.79	0.04073	-3.1917	0.01	-2.33	REJECT H0
$H0_{unambiguous(1)}$	$p(W_r)_{unambiguous(1)} \leq p(W_o)_{unambiguous}$	200	0.145	200	0.335	0.24	0.76	0.04271	-4.4488	0.01	-2.33	REJECT H0

5.4.6.3. Summary of Hypothesis Testing

The hypothesis tests presented in Table 9 confirmed that at a 99% significance level, similarity thresholds of 0.75 and 1.0 significantly reduced lexical ambiguity without significantly reducing correctness or conventionality. The hypothesis testing also showed that further reducing the similarity threshold (i.e. to 0.5, 0.25, and 0.0) resulted in significant reductions in correctness and conventionality – thereby no longer solving the research question. The hypothesis test results have thus confirmed that the Similarity Threshold is indeed an independent variable that when varied, affects the dependent variables: correctness, conventionality, and lexical ambiguity.

5.4.7. Reliability

In Section 5.3.10.1 we defined interrater reliability as the extent to which two or more individuals evaluating the same product or performance give identical judgements. We speculated that interrater reliability should be high given that all participants were from the same project, and had a similar engineering background. In other words, we *expect* the five respondents to answer each question in the same way. Figure 12 illustrates the actual result – where *5/5 agree* means all 5 respondents provided the same answer; *4/5 agree* means one respondent disagreed with the other 4 respondents; *3/5 agree* means two respondents disagreed with the other 3 respondents.

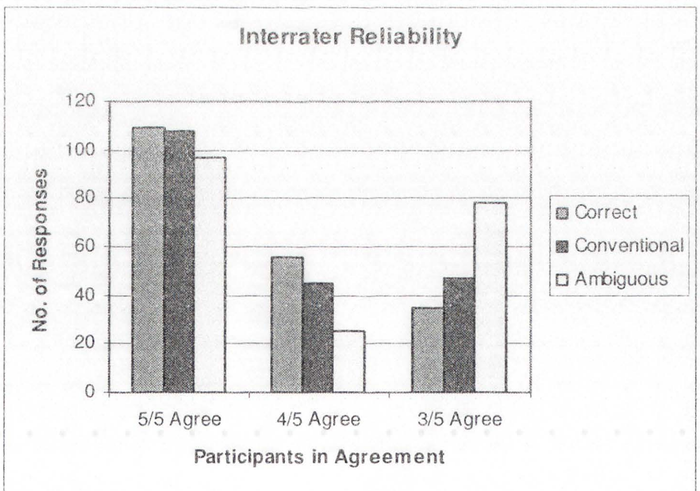

Figure 12 - Interrater Reliability

Figure 12 shows that interrater reliability is highest for correctness, second highest for conventionality, and third highest for lexical ambiguity. In other words, the respondent's answers are much more consistent when asked whether a particular replacement is correct, and much less consistent when asked whether a particular replacement is lexically ambiguous.

If we consider interrater responses to be reliable when *at least* 4 out of the 5 respondents agree, then

- The correctness result is 82.5% reliable (i.e. (109+56)/200)
- The conventionality result is 76.5% reliable (i.e. (108+45)/200
- The lexical ambiguity result is 61.0% reliable (i.e. (97+25)/200)

Chapter 6 Conclusion

6.1. Research Summary

In Chapter 1 we highlighted the problem of ambiguous requirements. We explained that our motivation for solving this problem is driven by its significance, specifically impact to project costs, schedule, and potential safety consequences. In Chapter 2 we evaluated, organised, and synthesised over 200 references relating to the topic of reducing ambiguity in requirements. In Chapter 3 we carefully phrased the research question:

How can a requirements specification be automatically re-expressed in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality?

We then justified March and Smith's (1995) strategic combination of design-science and natural-science as the appropriate research methodology. In Chapter 4 we designed a solution to answer the research question in terms of constructs, models, methods, and instantiations. We developed an understanding of the cause-and-effect relationships between the similarity threshold (independent variable), and the dependent variables: correctness, conventionality, and lexical ambiguity. We defined a new measure called *replaceability* as an effective measure as to the ability of one word to replace another. We also provided a cross-reference from the methods to the relevant function in the software prototype (the source code of which is provided in Appendix F: Prototype Software).

In Chapter 5 we evaluated our design solution to the research question. We defined the required data, justified a quantitative research method, and designed a controlled experiment to collect the data using a custom-built questionnaire. Hypothesis tests were conducted across a range of similarity thresholds in an effort to understand how the independent variable affected the dependent variables. We discovered that there was indeed a causal relationship between the similarity threshold and the dependent variables: correctness, conventionality, and lexical ambiguity. Perhaps most importantly, we discovered that for certain values of similarity threshold, using our design solution, it was possible to solve the research question, in that it was possible to *automatically re-express a requirements specification in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality.*

6.2. Answer to the Research Question

Question: How can a requirements specification be automatically re-expressed in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality?

Answer: A random sample of requirements from a recent rail industry project (RailCorp Contract C01645 2006) that were optimally re-expressed *using the optimal-constraint method defined in Chapter 4* has confirmed that at a 99% significance level, similarity thresholds of 0.75 and 1.0 *significantly* reduced the lexical ambiguity *without significantly* reducing the correctness or conventionality of the requirements specification.

We have explained in Section 5.3.11.1 that this result is expected to be internally valid, in that changes in dependent variables: correctness, conventionality, and lexical ambiguity can be attributed to the change in similarity threshold (independent variable) and not to other confounding factors. We have explained in Section 5.3.11.2 that proximal similarity was used to claim external validity. Specifically, we expect this result can be generalised to other parts of speech (e.g. nouns) and to other large specifications (recall from Chapter 4 that a large number of requirements are needed to adequately profile the characteristics of words (e.g. frequency, polysemy, etc)). We also postulate that a similar result might be achieved for specifications written in languages other than English.

6.2.1. Limitations of Answer

The hypothesis testing showed that further reducing the similarity threshold (i.e. to 0.5, 0.25, or 0.0) resulted in *significant* reductions in correctness and conventionality – to the extent that the research question could no longer be solved by our optimal-constraint method. Thus confirming a cause-and-effect relationship exists between the similarity threshold (independent variable) and dependent variables: correctness, conventionality, and lexical ambiguity.

6.3. Contributions Revisited

In Section 1.5 we claimed that this research would make five important research contributions. Here we recap and confirm the contributions based on the research conducted.

1. The Model: Section 4.3 defined four new propositions: a) Similar b) Correct, c) Conventional, and d) Unambiguous. The novel contribution is the synthesis of new definitions (propositions) from existing literature, e.g. our proposition for *conventional* is

an amalgamation of Fellbaum's (1990) *richly-lexicalised* and Green's (2006) *frequently occurring* concepts.

2. The Method: Section 4.5 defined six new functions: a) Meaning, b) Senses, c) Frequency, d) Width, e) Similarity, and f) Replaceability. The novel contribution is the formal definition we have provided for each function. The method also defines the exact sequence of activities that need to be followed to enable optimal re-expression to occur. We believe the largest single contribution within the Method is the *replaceability* metric which has been published in (Boyd, Zowghi & Gervasi 2007) and aims to provide the optimal trade-off between correctness, conventionality, and lexical ambiguity.
3. The Software Prototype: Appendix F: Prototype Software provides the source code for a re-usable software prototype that is a fully-functional implementation of the method. The purpose of the software prototype was to demonstrate proof-of-concept and to provide a tool to facilitate evaluation. The novel contribution is the source code itself. Other researchers may wish to re-use our software prototype to extend the evaluation to other requirement specifications; or to develop a plug-in validation tool for existing requirements management tools (refer to our suggestions for future research in Section 6.6.2). Following the submission of this thesis the source code will be made freely available in the public domain.
4. The Evaluation Method: Section 5.3 provided a detailed description of the evaluation method that we developed for this thesis. Section 5.2.1 defines the novel metrics we have defined for collecting data from participants and for comparing the correctness, conventionality, and lexical ambiguity of optimally re-expressed specifications to their original NL source specification. The metrics we defined are:
    - makes_sense(P, W_r, context(W_o)), and
    - would_use(P, W_r, context(W_o))
5. Empirical Evidence: Section 5.4 provided valuable empirical evidence in support of the claim that constraining the lexicon of natural language can improve the quality of requirement specifications. In particular, the hypothesis tests presented in Table 9 confirmed that at a 99% significance level, similarity thresholds of 0.75 and 1.0 significantly reduced lexical ambiguity without significantly reducing correctness or conventionality. Such evidence is a contribution since there is a distinct lack of empirical evidence to support the claims made in the CNL literature.

6.4. Implications for Research

The findings and contributions of this research have several implications for future research into the use of CNLs to reduce the ambiguity of requirements. We believe the three largest implications are 1) dynamic-constraint, 2) optimal-constraint, and 3) empirical evidence.

6.4.1. Dynamic Constraint

Throughout the literature on CNLs, we have always found the lexicon to be a static selection of words. By static we mean there was a point in time where the lexicon was *frozen*, and all text written in the CNL needs to be expressed using the *authorised* words. We have explained that this means existing CNLs are difficult to adapt to other domains, or to evolve over time. Furthermore, if a concept cannot be expressed by the authorised words, then the author cannot articulate what needs to be stated.

We have applied a completely different way of thinking in this research. That is, the CNL lexicon is not static, but is dynamically generated based upon the relevant features of words from a selected corpus of text relevant to some nominated application domain. Given the years of development required to generate most existing CNLs, for dynamic constraint to be feasible, we needed to investigate ways to significantly improve the efficiency of the constraint process. We have therefore combined research from the field of Natural Language Processing with existing research from the field of CNLs to develop a new concept of dynamic constraint. We think this is a key step in the direction of future CNLs, and consider this to be a major implication for research on this topic. One key feature of dynamic constraint is the ability to retain the original NL word if no suitable replacement exists – thereby guaranteeing expressiveness.

6.4.2. Optimal Constraint

We found the selection process used to decide which words to retain and which words to exclude from *existing* CNL lexicons to be ad-hoc and preference based rather than based on a balanced trade-off between quality factors. We have demonstrated how important quality factors can be related via a causal model, such that it is possible to develop a quality-conscious *replaceability* metric that can be used to *optimise* the replacements, where *optimise* means achieving some optimal combination of quality factors.

We hope our research sets an expectation in the CNL community that future CNL designers need to be transparent about their lexicon-selection methods. We expect future CNLs will 1) nominate a relevant quality model, 2) develop the equivalent of a *replaceability* metric, and 3) optimally-constrain the lexicon by maximising the equivalent replaceability metric.

6.4.3. Empirical Evidence

This research has generated valuable empirical evidence that supports the claim that optimal re-expression can significantly reduce the lexical ambiguity of a requirements specification, without significantly reducing its correctness or conventionality. Such evidence was not discovered in our review of the literature, and thus generating this evidence was a large focus of our research. We believe that the implication for future research is that our empirical evidence can be cited by other researchers as some justification for conducting further research on the CNL topic. Furthermore, the evaluation method we used to generate the empirical evidence was a novel contribution that we hope sets a benchmark for future CNL evaluations. In other words, an implication for future research is that speculation should be replaced with empirical evaluation and formal (e.g. hypothesis) tests.

6.5. Implications for Industry

One common criticism of research is that it is not practical enough to be adopted by industry. We have demonstrated through this research that optimally-expressed requirements can reduce the lexical ambiguity of requirements specifications without sacrificing their correctness or conventionality. The question we address here is *what does this mean for industry?* In other words, how practical is it for industry to adopt our optimal-constraint solution?

We have already explained in Section 5.3.11.2 that the optimal-constraint method is generalisable to projects comprising many requirements (e.g. 500+). From our experience (refer to Section 1.3), most projects in this category are contractually required to use requirements management tools such as IBM's DOORS (IBM 2008), Vitech's CORE (Vitech Corporation 2008), etc to ensure the correct and efficient management of requirements and requirements traceability. A side effect of using such requirements management tools is the availability of scripting languages for automatically processing the requirements data (such as the DOORS eXtension Language (DXL) (IBM 2008) and COREscript (Vitech Corporation 2008)). In our experience, such scripting languages are typically only used to report on metrics (e.g. requirements traceability) and to export custom traceability reports; however their capabilities are far greater.

Systems engineering standard EIA-632 (2003) states "The developer shall ensure that technical requirement statements and specified requirement statements, individually and as sets, are well formulated". Specifically, "the developer should analyze each requirement statement to ensure 1) ability to preserve competitiveness, 2) clarity, 3) correctness, 4) feasibility, 5) focus, 6) implementability, 7) modifiability, 8) removal of ambiguity, 9) singularity, 10) testability, and 11) verifiability". Of course, as the size of the requirements specification increases, it becomes more onerous to validate that every single requirement is well-formulated (e.g. 500 requirements means 5,500 validation checks). On the other hand, as the size of the requirements specification increases, the number of individual word occurrences increases, so the usage information (e.g. frequency, polysemy, etc) becomes more reliable and the optimal-constraint process is expected to produce better quality replacements.

We therefore suggest the software prototype we have developed for this thesis (refer Appendix F: Prototype Software) be ported to a requirements management tool and executed as part of the requirements validation process. This will enable the automation of 8) removal of ambiguity without adversely affecting 2) clarity, or 3) correctness. The optimal-constraint *plug-in* could be executed with every specification update. We believe this strategy would facilitate rapid industry-adoption since it gives useful results, within the existing tooling environment, and without forcing a management decision to prioritise quality above cost or schedule. For instance, it might be marketed to project management as: *you should optimise the expression of your requirements by executing this plug-in before exporting a new version of the specification. This will result in a less ambiguous specification. This is important because ambiguity is a well known contributor to project failures and to safety-related accidents.*

6.6. Suggestions for Future Research

6.6.1. Automatic Word Sense Disambiguation

In Section 4.5.1.3 we explained that human beings are particularly good at WSD, but that auto-WSD is unreliable and extremely difficult. In Section 5.4.2 we found that manual-WSD is time consuming at approximately 5 minutes per word sense on average. We believe automating WSD remains an important topic for future research. Furthermore, we believe that auto-WSD may need to be coupled with dictionaries that contain proper nouns (such as Wikipedia), since the biggest threat to WSD seems to be unknown words.

6.6.2. Integration with an existing RM Tool

In Section 6.5 we suggested that future practice might be to integrate optimal-constraint within existing requirements management tools (such as DOORS, CORE, Requisite Pro, etc) to facilitate the validation of requirements as described in systems engineering standards such as EIA-632 (2003). We suggest an interesting sequel to this thesis would be the porting of our prototype software into an existing RM tool, followed by a case-study to confirm the theorised benefits.

6.6.3. Quality Model for Requirement Specifications

In Section 2.3.3 we identified that there is no universally accepted quality model for requirement specifications. Furthermore, throughout our literature review we found that various authors often understand the same quality factor differently. We believe there is enormous value in developing a universal quality model for requirement specifications. Defining a quality model that includes cause-and-effect relationships between quality factors would perhaps enable us to define a better measure of *replaceability*.

6.6.4. Query Expansion

Manning and Schütze (1999) explain that semantic similarity is also used for query expansion in information retrieval. A user who describes a request for information in her own words may not be aware of related terms which are used in the documents that the user would be most interested in. If a user describes a request for documents on Russian space missions using the word *astronaut*, then a query expansion system can suggest the term *cosmonaut* based on the semantic similarity between astronaut and cosmonaut. We believe that future research into the use of *replaceability* as a way of improving the recall of search engines would be an extremely interesting topic for future research, and could be considered an extension of our concept to another field, in particular information retrieval. Word properties (such as frequency, sense, polysemy, etc) could be derived from historical search data.

6.7. Concluding Remarks

This thesis is titled "Design and Evaluation of a Method to Reduce the Lexical Ambiguity of Requirement Specifications". We have indeed designed a method to re-express requirements in a way that significantly reduces their lexical ambiguity without significantly reducing their correctness or conventionality. We have optimised the trade-off between lexical ambiguity, correctness, and conventionality through the definition of a new measure called *replaceability*. We have implemented a fully functional software prototype of our method, and have used this prototype to optimally re-express an existing industry specification. We have empirically evaluated the optimal re-expression using actual stakeholders from the industry project, and have statistically proven that our method indeed answers the research question. We have created empirical evidence to help fill a gap in the existing literature about the affects of CNLs on the quality of requirement specifications. We conclude that this thesis is an important contribution to the body of knowledge on reducing ambiguity in requirement specifications.

Appendix A: Ethics Approval

10 March 2008

Associate Professor Didar Zowghi
CB10.04.575
Faculty of Information Technology
UNIVERSITY OF TECHNOLOGY, SYDNEY

Dear Didar,

UTS HREC REF NO 2008-52 – ZOWGHI, Associate Professor Didar, GERVASI, Associate Professor Vincenzo (for BOYD, Mr Stephen PhD student) - "Optimal-Constraint Lexicons for Requirement Specifications"

Thank you for your response to my email dated 19 February 2008. Your response satisfactorily addresses the concerns and questions raised by the Committee, and I am pleased to inform you that ethics clearance is now granted.

Your clearance number is UTS HREC REF NO. 2008-52A

Please note that the ethical conduct of research is an on-going process. The *National Statement on Ethical Conduct in Research Involving Humans* requires us to obtain a report about the progress of the research, and in particular about any changes to the research which may have ethical implications. This report form must be completed at least annually, and at the end of the project (if it takes more than a year). The Ethics Secretariat will contact you when it is time to complete your first report.

I also refer you to the AVCC guidelines relating to the storage of data, which require that data be kept for a minimum of 5 years after publication of research. However, in NSW, longer retention requirements are required for research on human subjects with potential long-term effects, research with long-term environmental effects, or research considered of national or international significance, importance, or controversy. If the data from this research project falls into one of these categories, contact University Records for advice on long-term retention.

If you have any queries about your ethics clearance, or require any amendments to your research in the future, please do not hesitate to contact the Ethics Secretariat at the Research and Innovation Office, on 02 9514 9615.

Yours sincerely,

Professor Jane Stein-Parbury
Chairperson
UTS Human Research Ethics Committee

Appendix B: Research Participant Information Sheet

UNIVERSITY OF TECHNOLOGY, SYDNEY
RESEARCH PARTICIPANT INFORMATION SHEET

This is an information sheet for participants of a research project (titled "Optimally-Expressed Requirements") being conducted by Mr Stephen Boyd for his degree Doctor of Philosophy in Computing Sciences (C02029).

This research aims to answer the question "*How can we significantly reduce the lexical ambiguity of natural English requirements, without significantly reducing the correctness or conventionality?*"

To participate in this research project, you will need to complete an electronic (spreadsheet) questionnaire answering questions pertaining to the correctness, conventionality, and lexical ambiguity of verbs used in a typical requirement specification. Completing the questionnaire should take no more than 60 minutes. Your answers will be collected and used to test a series of hypotheses designed to validate the research problem.

If there are any concerns or questions about this research, please contact Mr Stephen Boyd directly, or alternatively his supervisor A/Prof Dr Didar Zowghi at the Department of Information Technology, University of Technology, Sydney, P.O. Box 123, Broadway, NSW 2007, Australia, phone +61 (02) 9514 1865.

You are free to withdraw your participation from this research project at any time, without consequences, and without giving a reason. Withdrawal from the research will not prejudice the participant's future career opportunities.

Research data gathered from this project may be published in a form that does not identify the participants in any way.

NOTE:
This study has been approved by the University of Technology, Sydney Human Research Ethics Committee. If you have any complaints or reservations about any aspect of your participation in this research which you cannot resolve with the researcher, you may contact the Ethics Committee through the Research Ethics Officer (ph: 02 - 9514 9615, Research.Ethics@uts.edu.au), and quote the UTS HREC reference number (2008-52A). Any complaint you make will be treated in confidence and investigated fully and you will be informed of the outcome.

Appendix C: Questionnaire Instruction Sheet

UNIVERSITY OF TECHNOLOGY, SYDNEY
RESEARCH PARTICIPANT INSTRUCTION SHEET

Instructions:

Step 1	Read the *Research Participant Information Sheet*
Step 2	Open the electronic questionnaire ("questionnaire[1|2].xls" using Microsoft Excel[1][2]
Step 3	**Select the "Sheet 1" tab**
Step 4	Read the "Requirement Text". Notice the missing verb ("________")
Step 5	Think about the "Verb" – is this a verb you would use?
Step 5.a	If you think you might choose this "Verb" to fill in the ("________"), then select "**I might choose this verb**" as your answer.
Step 5.b	If you don't think you would ever choose this "Verb" to fill in the ("________"), then select "**I would not choose this verb**" as your answer.
Step 6	Repeat from Step 5 until each "Verb" has been reviewed for the requirement "No."[3]
Step 7	Repeat from Step 4 until each requirement "No." has been reviewed.
Step 8	**Select the "Sheet 2" tab**
Step 9	Filter the "No." column on the next number (i.e. start with "1", then "2", ... etc)
Step 10	Read the "Requirement Text"[4]. Notice the missing verb ("________")
Step 11	Imagine the missing verb ("________") was defined by the "Synonyms" and "Definition" provided[5].
Step 11.a	If you think defining the missing verb ("________") in this way *would* generate the author's intended meaning of the Requirement, then select "**I think this is the author's intended meaning**" as your answer[6].
Step 11.b	If you think defining the missing verb ("________") in this way *could possibly* generate a valid meaning of the Requirement that is different to the authors intended meaning, then select "**I think this is a possible unintended meaning**" as your answer.
Step 11.c	If you think defining the missing verb ("________") in this way *could not possibly* generate a valid meaning of the Requirement, then select "**I don't think this meaning makes any sense**" as your answer.
Step 12	Repeat from Step 11 until each definition has been reviewed[3]
Step 13	Repeat from Step 9 until each requirement "No." has been reviewed.

Notes:

1. There are two questionnaire-variants randomly assigned to willing participants. You will either receive "questionnaire1.xls" or "questionnaire2.xls".
2. There are two sheets in the spreadsheet. Ensure you complete both Sheet 1 and Sheet 2.
3. There are no right or wrong answers. If you are unsure – then make a forced decision. Do not leave any answer cells blank. The number of remaining entries is shown in the top-right cell.
4. The same requirement text will be repeated (i.e. *"filled down"*) for all possible definitions of the blank ("________").
5. In some cases the "Synonyms" and/or "Definition" cells are blank. In other cases the "Synonyms" and/or "Definition" cells are duplicates. This is not accidental.
6. There may be more than one author's intended meaning per missing verb ("________").

Appendix D: Questionnaire Pilot Test (Sheet 1)

No.	Requirement Text	Verb	Answer
1	PPP Co must make provision to _________ CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.	equip	I would not choose this verb
		fit	I might choose this verb
		supply	I might choose this verb
2	PPP Co shall _________ the cabling for a redundant high-speed data network.	establish	I might choose this verb
		install	I might choose this verb
		use	I might choose this verb
3	PPP Co shall _________ cabling between CTIP mounting points so that CTIP head and headless units can communicate.	provide	I might choose this verb
		supply	I might choose this verb
4	PPP Co shall perform tests to _________ that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.	demonstrate	I might choose this verb
		prove	I might choose this verb
5	The Actual Energy Consumption of a Set shall be determined and _________ by PPP Co using a combination of simulation and testing.	demonstrated	I might choose this verb
		proven	I might choose this verb
6	The Set shall _________ three levels of locked security access.	allow for	I might choose this verb
		provide	I might choose this verb
		supply	I might choose this verb
7	An Access 1 Security key shall _________ Access 1 Security only.	control	I might choose this verb
		operate	I might choose this verb
8	The coupler assembly shall _________ a built in energy absorbing device.	feature	I might choose this verb
		have	I might choose this verb
9	All flexible hoses shall be _________ with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.	equipped	I might choose this verb
		fitted	I might choose this verb
		supplied	I might choose this verb
10	The light shall not visually _________ Passengers using the doorway.	hinder	I might choose this verb
		impede	I might choose this verb
		interfere	I would not choose this verb
		supply	I would not choose this verb
11	Details of the task lighting including switching, fitting type and area coverage shall be _________ during the Contract.	determined	I might choose this verb
		found out	I would not choose this verb
		used	I might choose this verb
12	Each headlight shall _________ a low beam providing not greater than 20% of the rated light output of the headlight.	feature	I might choose this verb
		have	I might choose this verb
13	Each headlight shall _________ a high beam providing the full rated light output of the headlight.	feature	I might choose this verb
		have	I might choose this verb
14	PPP Co shall _________ HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.	comply with	I might choose this verb
		follow	I might choose this verb
15	The design of drains shall ensure they cannot be _________ by debris.	blocked	I might choose this verb
		obstructed	I might choose this verb
		prevented	I would not choose this verb
16	The Set shall be able to _________ on the steepest gradient.	get going	I would not choose this verb
		start	I might choose this verb
		use	I would not choose this verb
17	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall be able to _________ an all stations run pattern without overheating the traction equipment when using maximum available traction and braking.	function	I might choose this verb
		operate	I might choose this verb
18	The Emergency Brake shall be applied in the event of the Set _________ a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.	crossing	I might choose this verb
		going through	I might choose this verb

No.	Requirement Text	Verb	Answer
		passing	I might choose this verb
		supplying	I would not choose this verb
19	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control _________ the Park Brake status for both bogies.	displaying	I might choose this verb
		showing	I might choose this verb
		using	I would not choose this verb
20	The Crew member taking the PEI call shall be able to _________ that call and operate the Intercom to the other Crew member.	hold	I might choose this verb
		maintain	I might choose this verb
21	The minimum volume level shall be agreed in final commissioning or, if not agreed, _________ by RailCorp's Representative.	determined	I might choose this verb
		found out	I would not choose this verb
		used	I would not choose this verb
22	An electro-acoustic virtual model of the carriage shall be _________ to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.	generated	I might choose this verb
		produced	I might choose this verb
		used	I might choose this verb
23	The DVA equipment shall be simple to _________ and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.	control	I might choose this verb
		operate	I might choose this verb
24	PPP Co shall nominate and _________ by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.	prove	I might choose this verb
		show	I might choose this verb
25	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to _________ that the door timing meets this requirement.	prove	I might choose this verb
		show	I might choose this verb
26	It shall be possible to _________ the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.	bring forward	I might choose this verb
		call up	I might choose this verb
		supply	I would not choose this verb
27	The TOS shall be _________ to TOS screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.	connected	I might choose this verb
		inserted	I would not choose this verb
		plugged in	I might choose this verb
28	The camera locations and coverage shall be _________ and agreed during the Mock-up review.	demonstrated	I might choose this verb
		presented	I might choose this verb
29	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be _________ and agreed during the Mock-up review.	demonstrated	I might choose this verb
		presented	I might choose this verb
30	PPP Co shall _________ the use of recyclable materials throughout the Set.	harness	I might choose this verb
		maximize	I might choose this verb
		use	I would not choose this verb
31	The Car interior space shall be _________ with a Car interior width measured across the full width of the Car, of at least 2920 mm.	increased	I might choose this verb
		maximized	I might choose this verb
32	The seat design and layout shall _________ the available legroom for seated passengers.	increase	I might choose this verb
		maximize	I might choose this verb
33	All compressors shall be controlled to _________ compressor life and optimise the efficiency of the air dryer system.	increase	I might choose this verb
		maximize	I might choose this verb
34	The EAPS shall be _________ from the Main Power Supply in Fault or Failure conditions.	disconnected	I might choose this verb
		switched	I might choose this verb
		unplugged	I might choose this verb
35	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to _________ Cars or remove access panels.	detach	I might choose this verb
		disconnect	I might choose this verb
		push	I would not choose this verb
		supply	I would not choose this verb
36	PPP Co shall _________ mechanisms which overcome the reception	consider	I might choose this verb

No.	Requirement Text	Verb	Answer
	and accuracy problem typically encountered by GPS Systems within the Rail Network.		
		look at	I might choose this verb
37	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall _________ the inclusion of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.	consider	I might choose this verb
		look at	I might choose this verb
38	PPP Co shall _________ any reasonable resulting corrective actions on all Sets to rectify the cause of the complaint.	attempt	I might choose this verb
		undertake	I might choose this verb
39	Level 1 diagnostic results shall include recommended remedial action required to be _________ by the Crew	attempt	I might choose this verb
		undertake	I might choose this verb
40	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be _________ by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.	attempt	I might choose this verb
		undertake	I might choose this verb

Appendix E: Questionnaire Pilot Test (Sheet 2)

No.	Requirement Text	Synonyms	Definition	Answer
1	PPP Co must make provision to _________ CTIP modules to every Set as per the following configuration: (i) Two 'headless' units in each Set.	ply, cater	provide what is desired or needed, especially support, food or sustenance	I think this is a possible unintended meaning
1			provide with abilities or understanding	I don't think this meaning makes any sense
1			conform to some shape or size	I don't think this meaning makes any sense
1		fit out, outfit	provide with (something usually for a specific purpose	I think this is the author's intended meaning
1			insert or adjust several objects or people	I think this is a possible unintended meaning
1		match, correspond, check, jibe, gibe, tally, agree	be compatible, similar or consistent; coincide in their characteristics	I think this is a possible unintended meaning
1			make fit	I think this is a possible unintended meaning
1		issue	circulate or distribute or equip with	I don't think this meaning makes any sense
1		go	be the right size or shape; fit correctly or as desired	I think this is a possible unintended meaning
1		fit out, outfit	provide with (something	I think this is the author's intended meaning
1		match	make correspond or harmonize	I don't think this meaning makes any sense
1		render, furnish	provide or furnish with	I think this is the author's intended meaning
1		suit, accommodate	be agreeable or acceptable to	I think this is a possible unintended meaning
1		meet, conform to	satisfy a condition or restriction	I don't think this meaning makes any sense
2	PPP Co shall _________ the cabling for a redundant high-speed data network.		put into an office or a position	I think this is a possible unintended meaning
2		set up	place	I think this is the author's intended meaning
2		lay down, make	institute, enact, or establish	I think this is the author's intended meaning
2		build	build or establish something abstract	I think this is a possible unintended meaning
2			seek or achieve an end by using to one's advantage	I don't think this meaning makes any sense
2		set up, found, launch	set up or found	I don't think this meaning makes any sense
2		practice, apply	avail oneself to	I don't think this meaning makes any sense
2			habitually do something (use only in the past tense)	I don't think this meaning makes any sense
2		base, ground, found	use as a basis for; found on	I think this is a possible unintended meaning
2		give	bring about	I think this is a possible unintended meaning
2		apply, employ	put into service; make work or employ (something for a particular purpose or for its inherent or natural purpose	I think this is a possible unintended meaning
2		expend	use up, consume fully	I think this is a possible unintended meaning
2		found, plant, constitute, institute	set up or lay the groundwork for	I think this is the author's intended meaning
2		prove,	establish the validity of something, as	I think this is a possible

No.	Requirement Text	Synonyms	Definition	Answer
		demonstrate, show, shew	by an example, explanation or experiment	unintended meaning
2		set up	place	I think this is the author's intended meaning
2		habituate	take or consume (regularly or habitually)	I don't think this meaning makes any sense
2		put in, set up	set up for use	I think this is the author's intended meaning
3	PPP Co shall _________ cabling between CTIP mounting points so that CTIP head and headless units can communicate.	put up, offer	mount or put up	I think this is the author's intended meaning
3		render, furnish	provide or furnish with	I think this is the author's intended meaning
3		leave, allow	make a possibility or provide opportunity for; permit to be attainable or cause to remain	I think this is a possible unintended meaning
3			determine (what is to happen in certain contingencies), especially by including a proviso condition or stipulation	I don't think this meaning makes any sense
3		bring home the bacon	supply means of subsistence; earn a living	I don't think this meaning makes any sense
3		issue	circulate or distribute or equip with	I don't think this meaning makes any sense
3		ply, cater	provide what is desired or needed, especially support, food or sustenance	I think this is the author's intended meaning
3		ply, cater	provide what is desired or needed, especially support, food or sustenance	I think this is a possible unintended meaning
3			take measures in preparation for	I think this is a possible unintended meaning
3		render, furnish	provide or furnish with	I think this is the author's intended meaning
4	PPP Co shall perform tests to _________ that the Set remains within the Static Rolling Stock Outline and Kinematic Rolling Stock Outline under all operating conditions of loading, dynamic behaviour and allowable wear, when tested at the Superelevation deficiency requirements nominated in RSU 289.	establish, show, shew	establish the validity of something, as by an example, explanation or experiment	I think this is the author's intended meaning
4		raise, leaven	cause to puff up with a leaven	I don't think this meaning makes any sense
4		testify, bear witness, evidence, show	provide evidence for	I think this is the author's intended meaning
4			obtain probate of	I think this is a possible unintended meaning
4		rise	increase in volume	I don't think this meaning makes any sense
4		show, demo, exhibit, present	show or demonstrate something to an interested audience	I think this is the author's intended meaning
4			prove formally; demonstrate by a mathematical, formal proof)	I think this is the author's intended meaning
4		test, try, try out, examine, essay	put to the test, as for its quality, or give experimental use to	I think this is the author's intended meaning
4		march	march in protest; take part in a demonstration	I don't think this meaning makes any sense
4		certify, manifest, evidence	provide evidence for; stand as proof of; show by one's behavior, attitude, or external attributes	I think this is a possible unintended meaning
4		establish, show, shew	establish the validity of something, as by an example, explanation or experiment	I think this is the author's intended meaning
4		turn out, turn up	be shown or be found to be	I don't think this meaning

No.	Requirement Text	Synonyms	Definition	Answer
				makes any sense
4			take a trial impression of)	I don't think this meaning makes any sense
5	The Actual Energy Consumption of a Set shall be determined and ________ by PPP Co using a combination of simulation and testing.	establish, show, shew	establish the validity of something, as by an example, explanation or experiment	I think this is the author's intended meaning
5			obtain probate of	I think this is a possible unintended meaning
5			take a trial impression of)	I don't think this meaning makes any sense
5			prove formally; demonstrate by a mathematical, formal proof)	I think this is the author's intended meaning
5		rise	increase in volume	I don't think this meaning makes any sense
5		turn out, turn up	be shown or be found to be	I don't think this meaning makes any sense
5		show, demo, exhibit, present	show or demonstrate something to an interested audience	I think this is the author's intended meaning
5		march	march in protest; take part in a demonstration	I don't think this meaning makes any sense
5		establish, show, shew	establish the validity of something, as by an example, explanation or experiment	I think this is the author's intended meaning
5		raise, leaven	cause to puff up with a leaven	I don't think this meaning makes any sense
5		certify, manifest, evidence	provide evidence for; stand as proof of; show by one's behavior, attitude, or external attributes	I think this is a possible unintended meaning
5		test, try, try out, examine, essay	put to the test, as for its quality, or give experimental use to	I think this is the author's intended meaning
5		testify, bear witness, evidence, show	provide evidence for	I think this is a possible unintended meaning
6	The Set shall ________ three levels of locked security access.	leave, allow	make a possibility or provide opportunity for; permit to be attainable or cause to remain	I think this is a possible unintended meaning
6		render, furnish	provide or furnish with	I think this is the author's intended meaning
6		render, furnish	provide or furnish with	I think this is the author's intended meaning
6			determine (what is to happen in certain contingencies), especially by including a proviso condition or stipulation	I think this is a possible unintended meaning
6		ply, cater	provide what is desired or needed, especially support, food or sustenance	I think this is a possible unintended meaning
6		put up, offer	mount or put up	I think this is a possible unintended meaning
6		bring home the bacon	supply means of subsistence; earn a living	I don't think this meaning makes any sense
6		issue	circulate or distribute or equip with	I think this is a possible unintended meaning
6		leave, allow	make a possibility or provide opportunity for; permit to be attainable or cause to remain	I think this is a possible unintended meaning
6		ply, cater	provide what is desired or needed, especially support, food or sustenance	I think this is a possible unintended meaning
6			take measures in preparation for	I think this is a possible unintended meaning
7	An Access 1 Security key shall ________ Access 1 Security only.	manipulate, keep in line	control (others or oneself	I think this is the author's intended meaning
7		command	exercise authoritative control or power over	I think this is the author's intended meaning
7		run	direct or control; projects, businesses, etc.	I don't think this meaning makes any sense
7		master	have a firm understanding or knowledge of; be on top of	I don't think this meaning makes any sense
7		manoeuver,	perform a movement in military or	I don't think this meaning

No.	Requirement Text	Synonyms	Definition	Answer
		maneuver, manoeuvre	naval tactics in order to secure an advantage in attack or defense)	makes any sense
7			handle and cause to function	I think this is the author's intended meaning
7			happen	I don't think this meaning makes any sense
7		work, go, run	perform as expected when applied	I think this is a possible unintended meaning
7			verify by using a duplicate register for comparison	I don't think this meaning makes any sense
7		see, check, insure, see to it, ensure, ascertain, assure	be careful or certain to do something; make certain of something	I don't think this meaning makes any sense
7		verify	verify or regulate by conducting a parallel experiment or comparing with another standard, of scientific experiments	I don't think this meaning makes any sense
7		hold in, hold, contain, check, curb, moderate	lessen the intensity of; temper; hold in restraint; hold or keep within limits	I think this is a possible unintended meaning
7		engage, mesh, lock	keep engaged	I don't think this meaning makes any sense
7		control	handle and cause to function	I think this is the author's intended meaning
7		operate on	perform surgery on	I don't think this meaning makes any sense
8	The coupler assembly shall _________ a built in energy absorbing device.	get, let	cause to move; cause to be in a certain position or condition	I don't think this meaning makes any sense
8		hold, throw, make, give	organize or be responsible for	I think this is a possible unintended meaning
8		give birth, deliver, bear, birth	give birth (to a newborn)	I don't think this meaning makes any sense
8			have left	I think this is a possible unintended meaning
8		have got, hold	have or possess, either in a concrete or an abstract sense	I think this is the author's intended meaning
8		experience	undergo	I don't think this meaning makes any sense
8		own, possess	have ownership or possession of	I think this is the author's intended meaning
8			suffer from; be ill with	I don't think this meaning makes any sense
8		induce, stimulate, cause, get, make	cause to do; cause to act in a specified manner	I think this is a possible unintended meaning
8		suffer, sustain, get	undergo (as of injuries and illnesses)	I don't think this meaning makes any sense
8			be confronted with	I don't think this meaning makes any sense
8		sport, boast	wear or display in an ostentatious or proud manner	I don't think this meaning makes any sense
8		experience, receive, get, undergo	of mental or physical states or experiences	I think this is a possible unintended meaning
8		take	have sex with; archaic use	I don't think this meaning makes any sense
8			have as a feature	I think this is the author's intended meaning
8			have as a feature	I think this is the author's intended meaning
8		accept, take	receive willingly something given or offered	I don't think this meaning makes any sense
8		receive	get something; come into possession of	I don't think this meaning makes any sense
8		consume, ingest, take in, take	serve oneself to, or consume regularly	I don't think this meaning makes any sense
8			have a personal or business relationship with someone	I don't think this meaning makes any sense
8		get, make	achieve a point or goal	I don't think this meaning makes any sense
9	All flexible hoses shall		make fit	I don't think this meaning

No.	Requirement Text	Synonyms	Definition	Answer
	be _________ with metal hexagon fittings, compatible with standard CityRail spanners carried in other rolling stock.			makes any sense
9			insert or adjust several objects or people	I don't think this meaning makes any sense
9		fit out, outfit	provide with (something usually for a specific purpose	I think this is the author's intended meaning
9		issue	circulate or distribute or equip with	I think this is a possible unintended meaning
9			provide with abilities or understanding	I don't think this meaning makes any sense
9		go	be the right size or shape; fit correctly or as desired	I think this is a possible unintended meaning
9		render, furnish	provide or furnish with	I think this is the author's intended meaning
9		match, correspond, check, jibe, gibe, tally, agree	be compatible, similar or consistent; coincide in their characteristics	I think this is a possible unintended meaning
9		meet, conform to	satisfy a condition or restriction	I think this is a possible unintended meaning
9		add, append	state or say further	I don't think this meaning makes any sense
9		suit, accommodate	be agreeable or acceptable to	I think this is a possible unintended meaning
9		ply, cater	provide what is desired or needed, especially support, food or sustenance	I think this is a possible unintended meaning
9		match	make correspond or harmonize	I don't think this meaning makes any sense
9			conform to some shape or size	I don't think this meaning makes any sense
9		fit out, outfit	provide with (something	I think this is the author's intended meaning
10	The light shall not visually _________ Passengers using the doorway.		come between so as to be hindrance or obstacle	I think this is the author's intended meaning
10		intervene, step in, interpose	get involved, so as to alter or hinder an action, or through force or threat of force	I think this is a possible unintended meaning
10		obstruct, obturate, occlude, jam, block, close up	block passage through	I think this is the author's intended meaning
10		handicap, hamper	put at a disadvantage	I think this is the author's intended meaning
10		obstruct, blockade, block,stymie, stymy, embarrass	hinder or prevent the progress or accomplishment of	I think this is a possible unintended meaning
10		add, append	state or say further	I don't think this meaning makes any sense
10			be a hindrance or obstacle to	I think this is the author's intended meaning
10		ply, cater	provide what is desired or needed, especially support, food or sustenance	I don't think this meaning makes any sense
10		issue	circulate or distribute or equip with	I don't think this meaning makes any sense
10			be a hindrance or obstacle to	I think this is the author's intended meaning
10		render, furnish	provide or furnish with	I don't think this meaning makes any sense
11	Details of the task lighting including switching, fitting type and area coverage shall be _________ during the Contract.	check, see, ascertain, watch, learn	find out, learn, or determine with certainty, usually by making an inquiry or other effort	I think this is a possible unintended meaning
11		set	fix conclusively or authoritatively	I think this is the author's intended meaning
11		specify, set, fix, limit	decide upon or fix definitely	I think this is the author's intended meaning
11		decide, make up	reach, make, or come to a decision	I think this is the author's

No.	Requirement Text	Synonyms	Definition	Answer
		one's mind	about something	intended meaning
11		learn, hear, get word, get wind, pick up, get a line, discover, see	get to know or become aware of, usually accidentally	I think this is a possible unintended meaning
11		apply, employ	put into service; make work or employ (something for a particular purpose or for its inherent or natural purpose	I think this is a possible unintended meaning
11			seek or achieve an end by using to one's advantage	I don't think this meaning makes any sense
11			habitually do something (use only in the past tense)	I don't think this meaning makes any sense
11			fix in scope; fix the boundaries of; the tree determines the border of the property)	I think this is a possible unintended meaning
11		habituate	take or consume (regularly or habitually)	I don't think this meaning makes any sense
11		expend	use up, consume fully	I don't think this meaning makes any sense
11		shape, mold, influence, regulate	shape or influence; give direction to	I think this is a possible unintended meaning
11		check, see, ascertain, watch, learn	find out, learn, or determine with certainty, usually by making an inquiry or other effort	I think this is a possible unintended meaning
11		catch out	trap; especially in an error or in a reprehensible act	I don't think this meaning makes any sense
11		practice, apply	avail oneself to	I don't think this meaning makes any sense
11		find, ascertain	after a calculation, investigation, experiment, survey, or study	I think this is a possible unintended meaning
11		find, ascertain	after a calculation, investigation, experiment, survey, or study	I think this is a possible unintended meaning
11		settle, square off, square up	settle conclusively; come to terms	I think this is a possible unintended meaning
12	Each headlight shall _________ a low beam providing not greater than 20% of the rated light output of the headlight.	give birth, deliver, bear, birth	give birth (to a newborn)	I don't think this meaning makes any sense
12		have got, hold	have or possess, either in a concrete or an abstract sense	I think this is the author's intended meaning
12		experience, receive, get, undergo	of mental or physical states or experiences	I don't think this meaning makes any sense
12			have a personal or business relationship with someone	I don't think this meaning makes any sense
12			be confronted with	I don't think this meaning makes any sense
12			suffer from; be ill with	I don't think this meaning makes any sense
12		suffer, sustain, get	undergo (as of injuries and illnesses)	I don't think this meaning makes any sense
12			have left	I don't think this meaning makes any sense
12		experience	undergo	I don't think this meaning makes any sense
12		get, make	achieve a point or goal	I think this is a possible unintended meaning
12		accept, take	receive willingly something given or offered	I don't think this meaning makes any sense
12		sport, boast	wear or display in an ostentatious or proud manner	I don't think this meaning makes any sense
12		get, let	cause to move; cause to be in a certain position or condition	I don't think this meaning makes any sense
12		consume, ingest, take in, take	serve oneself to, or consume regularly	I don't think this meaning makes any sense
12		hold, throw, make, give	organize or be responsible for	I think this is a possible unintended meaning
12		take	have sex with; archaic use	I don't think this meaning makes any sense
12			have as a feature	I think this is the author's intended meaning

No.	Requirement Text	Synonyms	Definition	Answer
12		own, possess	have ownership or possession of	I think this is a possible unintended meaning
12		induce, stimulate, cause, get, make	cause to do; cause to act in a specified manner	I think this is a possible unintended meaning
12			have as a feature	I think this is the author's intended meaning
12		receive	get something; come into possession of	I think this is the author's intended meaning
13	Each headlight shall _________ a high beam providing the full rated light output of the headlight.		suffer from; be ill with	I don't think this meaning makes any sense
13		consume, ingest, take in, take	serve oneself to, or consume regularly	I don't think this meaning makes any sense
13		take	have sex with; archaic use	I don't think this meaning makes any sense
13		experience, receive, get, undergo	of mental or physical states or experiences	I don't think this meaning makes any sense
13		receive	get something; come into possession of	I think this is the author's intended meaning
13			have as a feature	I think this is the author's intended meaning
13			have a personal or business relationship with someone	I don't think this meaning makes any sense
13		get, make	achieve a point or goal	I think this is a possible unintended meaning
13		have got, hold	have or possess, either in a concrete or an abstract sense	I think this is the author's intended meaning
13		hold, throw, make, give	organize or be responsible for	I don't think this meaning makes any sense
13		experience	undergo	I don't think this meaning makes any sense
13			have as a feature	I think this is the author's intended meaning
13			have left	I don't think this meaning makes any sense
13		suffer, sustain, get	undergo (as of injuries and illnesses)	I don't think this meaning makes any sense
13		give birth, deliver, bear, birth	give birth (to a newborn)	I don't think this meaning makes any sense
13			be confronted with	I don't think this meaning makes any sense
13		own, possess	have ownership or possession of	I think this is the author's intended meaning
13		induce, stimulate, cause, get, make	cause to do; cause to act in a specified manner	I think this is a possible unintended meaning
13		sport, boast	wear or display in an ostentatious or proud manner	I don't think this meaning makes any sense
13		accept, take	receive willingly something given or offered	I don't think this meaning makes any sense
13		get, let	cause to move; cause to be in a certain position or condition	I think this is a possible unintended meaning
14	PPP Co shall _________ HB40-1992 Australian Refrigeration and Air Conditioning Code of Good Practice or an approved equivalent International Standard.	come after	come after in time, as a result	I don't think this meaning makes any sense
14		travel along	travel along a certain course	I don't think this meaning makes any sense
14		adopt, espouse	choose and follow; as of theories, ideas, policies, strategies or plans	I think this is the author's intended meaning
14		fall out	come as a logical consequence; follow logically	I think this is a possible unintended meaning
14		be	work in a specific place, with a specific subject, or in a specific function	I don't think this meaning makes any sense
14		trace	follow, discover, or ascertain the course of development of something	I think this is the author's intended meaning

No.	Requirement Text	Synonyms	Definition	Answer
14		play along, accompany	perform an accompaniment to	I don't think this meaning makes any sense
14		pursue	follow in or as if in pursuit	I don't think this meaning makes any sense
14		stick to, stick with	keep to	I think this is a possible unintended meaning
14		keep up, keep abreast	keep informed	I don't think this meaning makes any sense
14		postdate	be later in time	I don't think this meaning makes any sense
14			grasp the meaning	I think this is a possible unintended meaning
14		abide by	act in accordance with someone's rules, commands, or wishes	I think this is a possible unintended meaning
14			adhere to or practice	I think this is the author's intended meaning
14		succeed, come after	be the successor (of)	I don't think this meaning makes any sense
14		come	to be the product or result	I don't think this meaning makes any sense
14			to bring something about at a later time than	I don't think this meaning makes any sense
14		watch, observe, watch over, keep an eye on	follow with the eyes or the mind	I don't think this meaning makes any sense
14			accept and follow the leadership or command or guidance of	I think this is the author's intended meaning
14		surveil, survey	keep under surveillance	I don't think this meaning makes any sense
14			be next	I don't think this meaning makes any sense
14		abide by	act in accordance with someone's rules, commands, or wishes	I think this is a possible unintended meaning
14			to travel behind, go after, come after	I don't think this meaning makes any sense
14		take after	imitate in behavior; take as a model	I don't think this meaning makes any sense
14		conform to	behave in accordance or in agreement with	I think this is the author's intended meaning
15	The design of drains shall ensure they cannot be ________ by debris.	freeze, immobilize, immobilise	prohibit the conversion or use of (assets)	I don't think this meaning makes any sense
15			interrupt the normal function of by means of anesthesia	I don't think this meaning makes any sense
15		blockade, hinder, stymie, stymy, embarrass	hinder or prevent the progress or accomplishment of	I think this is a possible unintended meaning
15		blockade, hinder, stymie, stymy, embarrass	hinder or prevent the progress or accomplishment of	I think this is a possible unintended meaning
15			stamp or emboss a title or design on a book with a block	I don't think this meaning makes any sense
15		jam	interfere with or prevent the reception of signals	I don't think this meaning makes any sense
15		obturate, impede, occlude, jam, close up	block passage through	I think this is the author's intended meaning
15			shape by using a block	I don't think this meaning makes any sense
15		parry, deflect	impede the movement of (an opponent or a ball)	I think this is a possible unintended meaning
15			run on a block system	I don't think this meaning makes any sense
15		prevent, forestall, foreclose, preclude, forbid	keep from happening or arising; have the effect of preventing	I think this is a possible unintended meaning
15		stuff, lug, choke up	obstruct	I think this is the author's intended meaning
15			shut out from view or get in the way so as to hide from sight	I don't think this meaning makes any sense
15		barricadeade, stop off up, bar	render unsuitable for passage	I think this is the author's intended meaning

No.	Requirement Text	Synonyms	Definition	Answer
15			shut out from view or get in the way so as to hide from sight	I don't think this meaning makes any sense
15		forget, blank out, draw a blank	be unable to remember	I don't think this meaning makes any sense
15			support, secure, or raise with a block	I don't think this meaning makes any sense
15		obturate, impede, occlude, jam, close up	block passage through	I think this is the author's intended meaning
15		prevent, keep	prevent from doing something or being in a certain state	I don't think this meaning makes any sense
15			shape into a block or blocks	I don't think this meaning makes any sense
15		stop, halt, kibosh	stop from happening or developing	I don't think this meaning makes any sense
16	The Set shall be able to _________ on the steepest gradient.	begin, lead off, commence	set in motion, cause to start	I think this is the author's intended meaning
16			seek or achieve an end by using to one's advantage	I don't think this meaning makes any sense
16		depart, part, start out, set forth, set off, set out, take off	leave	I think this is the author's intended meaning
16		start up	get going or set in motion	I think this is the author's intended meaning
16			play in the starting line-up)	I don't think this meaning makes any sense
16		go	begin or set in motion	I think this is the author's intended meaning
16		habituate	take or consume (regularly or habitually)	I don't think this meaning makes any sense
16		startle, jump	move or jump suddenly, as if in surprise or alarm	I think this is a possible unintended meaning
16		apply, employ	put into service; make work or employ (something for a particular purpose or for its inherent or natural purpose	I think this is a possible unintended meaning
16		take up	begin work or acting in a certain capacity, office or job	I think this is a possible unintended meaning
16		get down, begin, get, start out, set about, set out, commence	take the first step or steps in carrying out an action	I think this is a possible unintended meaning
16		get cracking, bestir oneself, get moving, get weaving, get started, get rolling	start to be active	I think this is a possible unintended meaning
16		begin	begin an event that is implied and limited by the nature or inherent function of the direct object	I don't think this meaning makes any sense
16		begin	have a beginning, in a temporal, spatial, or evaluative sense	I don't think this meaning makes any sense
16		practice, apply	avail oneself to	I don't think this meaning makes any sense
16		originate, initiate	bring into being	I don't think this meaning makes any sense
16		begin	have a beginning characterized in some specified way	I don't think this meaning makes any sense
16		expend	use up, consume fully	I don't think this meaning makes any sense
16		go	begin or set in motion	I think this is the author's intended meaning
16		start up, embark on, commence	get off the ground	I think this is a possible unintended meaning
16			habitually do something (use only in the past tense)	I don't think this meaning makes any sense
17	With not less than two motor cars (or the equivalent number of motors or traction packages) cut out an AW3 loaded Set shall	work, go, run	perform as expected when applied	I think this is the author's intended meaning

No.	Requirement Text	Synonyms	Definition	Answer
	be able to _________ an all stations run pattern without overheating the traction equipment when using maximum available traction and braking.			
17		manoeuver, maneuver, manoeuvre	perform a movement in military or naval tactics in order to secure an advantage in attack or defense)	I don't think this meaning makes any sense
17		control	handle and cause to function	I think this is a possible unintended meaning
17		work, go, run	perform as expected when applied	I think this is the author's intended meaning
17		officiate	perform duties attached to a particular office or place or function	I think this is a possible unintended meaning
17		run	direct or control; projects, businesses, etc.	I think this is a possible unintended meaning
17			happen	I don't think this meaning makes any sense
17		serve	serve a purpose, role, or function	I think this is a possible unintended meaning
17		operate on	perform surgery on	I don't think this meaning makes any sense
17		engage, mesh, lock	keep engaged	I don't think this meaning makes any sense
18	The Emergency Brake shall be applied in the event of the Set _________ a raised Train Stop by tripping of the trip cock venting the Brake Pipe to atmosphere.	follow through, follow up, follow out, carry out, implement, put through	pursue to a conclusion or bring to a successful issue	I don't think this meaning makes any sense
18		exceed, transcend, overstep, go past, top	go beyond	I don't think this meaning makes any sense
18		communicate, pass on, put across	transmit information	I think this is a possible unintended meaning
18		excrete, egest, eliminate	eliminate from the body	I don't think this meaning makes any sense
18			accept or judge as acceptable	I don't think this meaning makes any sense
18		clear	go unchallenged; be approved	I don't think this meaning makes any sense
18		crossbreed, hybridize, hybridise, interbreed	breed animals or plants using parents of different races and varieties	I don't think this meaning makes any sense
18		go through, go across	go across or through	I think this is the author's intended meaning
18		fall, return, devolve	be inherited by	I don't think this meaning makes any sense
18		work through, run through	apply thoroughly; think through	I don't think this meaning makes any sense
18			allow to go without comment or censure	I don't think this meaning makes any sense
18		excrete, egest, eliminate	eliminate from the body	I don't think this meaning makes any sense
18		traverse, span, sweep	to cover or extend over an area or time period	I don't think this meaning makes any sense
18			transfer to another; of rights or property	I don't think this meaning makes any sense
18		die, decease, perish, go, exit, pass away, expire	pass from physical life and lose all all bodily attributes and functions necessary to sustain life	I don't think this meaning makes any sense
18		elapse, lapse, slip by, glide by, slip away, go by, slide by, go along	pass by	I think this is a possible unintended meaning
18		spend	pass (time	I don't think this meaning makes any sense
18		traverse, track,	travel across or pass over	I think this is the author's

No.	Requirement Text	Synonyms	Definition	Answer
		cover, pass over, get over, get across, cut through, cut across		intended meaning
18			meet and pass	I think this is a possible unintended meaning
18			be identified, regarded, accepted, or mistaken for someone or something else; as by denying one's own ancestry or background	I don't think this meaning makes any sense
18		add, append	state or say further	I don't think this meaning makes any sense
18			fold so as to resemble a cross	I don't think this meaning makes any sense
18		authorize, authorise, clear	grant authorization or clearance for	I don't think this meaning makes any sense
18		go across	go across or through	I think this is the author's intended meaning
18		run, go, lead, extend	stretch out over a distance, space, time, or scope; run or extend between two points or beyond a certain point	I don't think this meaning makes any sense
18		issue	circulate or distribute or equip with	I don't think this meaning makes any sense
18		travel by, pass by, surpass, go past, go by	pass by	I think this is a possible unintended meaning
18		make it	go successfully through a test or a selection process	I don't think this meaning makes any sense
18		evanesce, fade, blow over, pass off, fleet	disappear gradually	I don't think this meaning makes any sense
18		legislate	make laws, bills, etc. or bring into effect by legislation	I don't think this meaning makes any sense
18		thwart, queer, spoil, scotch, foil, frustrate, baffle, bilk	hinder or prevent (the efforts, plans, or desires	I don't think this meaning makes any sense
18		happen, hap, go on, pass off, occur, fall out, come about, take place	come to pass	I don't think this meaning makes any sense
18		hand, reach, pass on, turn over, give	place into the hands or custody of	I don't think this meaning makes any sense
18		render, furnish	provide or furnish with	I don't think this meaning makes any sense
18		devour, down, consume	eat immoderately	I don't think this meaning makes any sense
18		intersect	meet at a point)	I think this is a possible unintended meaning
18			trace a line through or across	I don't think this meaning makes any sense
18		experience, undergo, see	go or live through	I don't think this meaning makes any sense
18		make pass	cause to pass	I don't think this meaning makes any sense
18		overtake, overhaul	travel past	I think this is a possible unintended meaning
18		ply, cater	provide what is desired or needed, especially support, food or sustenance	I don't think this meaning makes any sense
18			throw (a ball	I don't think this meaning makes any sense
18		guide, run, draw	guide or pass over something	I think this is a possible unintended meaning
18		sink, lapse	pass into a specified state or condition	I don't think this meaning makes any sense
19	A local Park Brake off-on status indication shall be provided in each Car adjacent to the Park Brake apply and release control __________ the Park Brake status for both	apply, employ	put into service; make work or employ (something for a particular purpose or for its inherent or natural purpose	I don't think this meaning makes any sense

No.	Requirement Text	Synonyms	Definition	Answer
	bogies.			
19			attract attention by displaying some body part or posing; of animals)	I don't think this meaning makes any sense
19		show up	be or become visible or noticeable	I think this is a possible unintended meaning
19		usher	show (someone to their seats, as in theaters or auditoriums	I don't think this meaning makes any sense
19		expose, exhibit	to show, make visible or apparent	I think this is the author's intended meaning
19		practice, apply	avail oneself to	I don't think this meaning makes any sense
19			habitually do something (use only in the past tense)	I don't think this meaning makes any sense
19			seek or achieve an end by using to one's advantage	I don't think this meaning makes any sense
19		read, register, record	indicate a certain reading; of gauges and instruments	I think this is the author's intended meaning
19		habituate	take or consume (regularly or habitually)	I don't think this meaning makes any sense
19			make visible or noticeable	I think this is a possible unintended meaning
19		demo, exhibit, present, demonstrate	show or demonstrate something to an interested audience	I think this is a possible unintended meaning
19		reveal	make clear and visible	I think this is a possible unintended meaning
19			give evidence of, as of records	I think this is a possible unintended meaning
19		express, evince	give expression to	I think this is a possible unintended meaning
19		testify, bear witness, prove, evidence	provide evidence for	I think this is a possible unintended meaning
19		indicate, point	indicate a place, direction, person, or thing; either spatially or figuratively	I think this is a possible unintended meaning
19			finish third or better in a horse or dog race	I don't think this meaning makes any sense
19		picture, depict, render	show in, or as in, a picture	I think this is a possible unintended meaning
19		expend	use up, consume fully	I don't think this meaning makes any sense
19		prove, demonstrate, establish, shew	establish the validity of something, as by an example, explanation or experiment	I think this is a possible unintended meaning
19		reveal	make clear and visible	I think this is a possible unintended meaning
20	The Crew member taking the PEI call shall be able to _________ that call and operate the Intercom to the other Crew member.	observe, keep	observe correctly or closely	I don't think this meaning makes any sense
20			keep from departing	I think this is a possible unintended meaning
20		defend	state or assert	I don't think this meaning makes any sense
20		deem, view as, take for	keep in mind or convey as a conviction or view	I don't think this meaning makes any sense
20		halt, arrest	cause to stop	I don't think this meaning makes any sense
20		keep	maintain for use and service	I think this is the author's intended meaning
20		control in, contain, check, curb, moderate	lessen the intensity of; temper; hold in restraint; hold or keep within limits	I don't think this meaning makes any sense
20		sustain, keep	supply with necessities and support	I think this is a possible unintended meaning
20			remain committed to	I think this is a possible unintended meaning
20		harbor, harbour, entertain, nurse	maintain (a theory, thoughts, or feelings)	I think this is a possible unintended meaning
20		have, have got	have or possess, either in a concrete or	I think this is a possible

No.	Requirement Text	Synonyms	Definition	Answer
			an abstract sense	unintended meaning
20		agree, concur, concord	be in accord; be in agreement	I don't think this meaning makes any sense
20		defy, withstand up	resist or confront with resistance	I don't think this meaning makes any sense
20		throw, have, make, give	organize or be responsible for	I don't think this meaning makes any sense
20		defend, guard	protect against a challenge or attack	I don't think this meaning makes any sense
20			hold the attention of	I think this is a possible unintended meaning
20		keep	maintain by writing regular records	I don't think this meaning makes any sense
20		carry	drink alcohol without showing ill effects	I don't think this meaning makes any sense
20		declare, adjudge	declare to be	I don't think this meaning makes any sense
20		upmaintain	support against an opponent	I don't think this meaning makes any sense
20		keep	keep in a certain state, position, or activity; e.g., "keep clean"	I think this is a possible unintended meaning
20		carry, bear	support or hold in a certain manner	I think this is a possible unintended meaning
20		conserve, preserve, keep up	keep in safety and protect from harm, decay, loss, or destruction	I don't think this meaning makes any sense
20		retain, keep back back	secure and keep for possible future use or application	I think this is a possible unintended meaning
20		wield, exert	of power or authority)	I don't think this meaning makes any sense
20		reserve, book	arrange for and reserve (something for someone else	I think this is a possible unintended meaning
20		bear	have rightfully; of rights, titles, and offices	I don't think this meaning makes any sense
20			stop dealing with	I don't think this meaning makes any sense
20			aim, point, or direct	I think this is a possible unintended meaning
20			have as a major characteristic	I don't think this meaning makes any sense
20			remain in a certain state, position, or condition	I think this is a possible unintended meaning
20		keep	keep in a certain state, position, or activity; e.g., "keep clean"	I think this is a possible unintended meaning
20		support, sustain up	be the physical support of; carry the weight of	I don't think this meaning makes any sense
20			keep from exhaling or expelling	I don't think this meaning makes any sense
20			take and maintain control over, often by violent means	I don't think this meaning makes any sense
20		assert, asseverate	state categorically)	I don't think this meaning makes any sense
20		contain, take	be capable of holding or containing	I think this is a possible unintended meaning
20		prevail, obtain	be valid, applicable, or true	I don't think this meaning makes any sense
20			assert or affirm	I don't think this meaning makes any sense
20		bear, carry, contain	contain or hold; have within	I think this is a possible unintended meaning
20		restrain, confine	to close within bounds, limit or hold back from movement	I don't think this meaning makes any sense
20		accommodate, admit	have room for; hold without crowding	I don't think this meaning makes any sense
20		oblige, bind, obligate	bind by an obligation; cause to be indebted	I don't think this meaning makes any sense
20			cover as for protection against noise or smell	I don't think this meaning makes any sense
20		take hold	have or hold in one's hands or grip	I think this is a possible unintended meaning
20		apply, go for	be pertinent or relevant or applicable	I don't think this meaning makes any sense
21		find, ascertain	after a calculation, investigation,	I think this is a possible

No.	Requirement Text	Synonyms	Definition	Answer
			experiment, survey, or study	unintended meaning
21		apply, employ	put into service; make work or employ (something for a particular purpose or for its inherent or natural purpose	I think this is a possible unintended meaning
21	The minimum volume level shall be agreed in final commissioning or, if not agreed, _________ by RailCorp's Representative.	settle, square off, square up	settle conclusively; come to terms	I think this is a possible unintended meaning
21		learn, hear, get word, get wind, pick up, get a line, discover, see	get to know or become aware of, usually accidentally	I don't think this meaning makes any sense
21		decide, make up one's mind	reach, make, or come to a decision about something	I think this is the author's intended meaning
21			seek or achieve an end by using to one's advantage	I don't think this meaning makes any sense
21		habituate	take or consume (regularly or habitually)	I don't think this meaning makes any sense
21		specify, set, fix, limit	decide upon or fix definitely	I think this is the author's intended meaning
21		check, see, ascertain, watch, learn	find out, learn, or determine with certainty, usually by making an inquiry or other effort	I think this is a possible unintended meaning
21		expend	use up, consume fully	I don't think this meaning makes any sense
21		shape, mold, influence, regulate	shape or influence; give direction to	I think this is a possible unintended meaning
21		catch out	trap; especially in an error or in a reprehensible act	I don't think this meaning makes any sense
21		practice, apply	avail oneself to	I don't think this meaning makes any sense
21			habitually do something (use only in the past tense)	I don't think this meaning makes any sense
21			fix in scope; fix the boundaries of; the tree determines the border of the property)	I don't think this meaning makes any sense
21		check, see, ascertain, watch, learn	find out, learn, or determine with certainty, usually by making an inquiry or other effort	I think this is a possible unintended meaning
21		find, ascertain	after a calculation, investigation, experiment, survey, or study	I think this is a possible unintended meaning
21		set	fix conclusively or authoritatively	I think this is the author's intended meaning
22	An electro-acoustic virtual model of the carriage shall be _________ to ascertain that the design solution is compliant with this RailCorp Train Performance Specification.	bring forth	bring forth or yield	I think this is a possible unintended meaning
22		render, yield, return, give	give or supply	I think this is the author's intended meaning
22		grow, raise, farm	cultivate by growing, often involving improvements by means of agricultural techniques	I don't think this meaning makes any sense
22		grow, develop, get, acquire	come to have or undergo a change of (physical features and attributes)	I don't think this meaning makes any sense
22		habituate	take or consume (regularly or habitually)	I don't think this meaning makes any sense
22		practice, apply	avail oneself to	I don't think this meaning makes any sense
22			seek or achieve an end by using to one's advantage	I don't think this meaning makes any sense
22		bring forth	bring into existence	I think this is a possible unintended meaning
22		bring about, give rise	cause to occur or exist	I think this is a possible unintended meaning
22		apply, employ	put into service; make work or employ	I think this is a possible

No.	Requirement Text	Synonyms	Definition	Answer
			(something for a particular purpose or for its inherent or natural purpose	unintended meaning
22			habitually do something (use only in the past tense)	I don't think this meaning makes any sense
22		bring on, bring out	bring onto the market or release	I think this is a possible unintended meaning
22		expend	use up, consume fully	I don't think this meaning makes any sense
22			produce (energy)	I don't think this meaning makes any sense
22		bring forth	bring out for display	I think this is a possible unintended meaning
22		make, create	create or manufacture a man-made product	I think this is a possible unintended meaning
22		beget, get, engender, father, mother, sire, bring forth	make children	I don't think this meaning makes any sense
23	The DVA equipment shall be simple to ________ and highly resistant to abuse, Vandalism, cleaning agents and cleaning processes.	operate on	perform surgery on	I don't think this meaning makes any sense
23		work, go, run	perform as expected when applied	I think this is the author's intended meaning
23			verify by using a duplicate register for comparison	I don't think this meaning makes any sense
23		master	have a firm understanding or knowledge of; be on top of	I think this is a possible unintended meaning
23		hold in, hold, contain, check, curb, moderate	lessen the intensity of; temper; hold in restraint; hold or keep within limits	I don't think this meaning makes any sense
23			happen	I don't think this meaning makes any sense
23		run	direct or control; projects, businesses, etc.	I think this is a possible unintended meaning
23			handle and cause to function	I think this is the author's intended meaning
23		engage, mesh, lock	keep engaged	I don't think this meaning makes any sense
23		manoeuver, maneuver, manoeuvre	perform a movement in military or naval tactics in order to secure an advantage in attack or defense)	I don't think this meaning makes any sense
23		see, check, insure, see to it, ensure, ascertain, assure	be careful or certain to do something; make certain of something	I don't think this meaning makes any sense
23		manipulate, keep in line	control (others or oneself	I don't think this meaning makes any sense
23		command	exercise authoritative control or power over	I think this is a possible unintended meaning
23		verify	verify or regulate by conducting a parallel experiment or comparing with another standard, of scientific experiments	I think this is a possible unintended meaning
23			handle and cause to function	I think this is the author's intended meaning
24	PPP Co shall nominate and ________ by full analysis the location of the centre of gravity and roll centre of each Car type under AW0 and AW3 loaded conditions.		obtain probate of	I think this is a possible unintended meaning
24		rise	increase in volume	I don't think this meaning makes any sense
24		turn out, turn up	be shown or be found to be	I don't think this meaning makes any sense
24		show up	be or become visible or noticeable	I think this is a possible unintended meaning
24		demonstrate, establish, shew	establish the validity of something, as by an example, explanation or	I think this is the author's intended meaning

No.	Requirement Text	Synonyms	Definition	Answer
			experiment	
24			make visible or noticeable	I think this is a possible unintended meaning
24		indicate, point	indicate a place, direction, person, or thing; either spatially or figuratively	I think this is a possible unintended meaning
24			take a trial impression of)	I don't think this meaning makes any sense
24		prove, demonstrate, establish, shew	establish the validity of something, as by an example, explanation or experiment	I think this is a possible unintended meaning
24		picture, depict, render	show in, or as in, a picture	I think this is a possible unintended meaning
24		test, try, try out, examine, essay	put to the test, as for its quality, or give experimental use to	I think this is the author's intended meaning
24		raise, leaven	cause to puff up with a leaven	I don't think this meaning makes any sense
24			give evidence of, as of records	I think this is a possible unintended meaning
24		express, evince	give expression to	I don't think this meaning makes any sense
24		read, register, record	indicate a certain reading; of gauges and instruments	I don't think this meaning makes any sense
24		usher	show (someone to their seats, as in theaters or auditoriums	I don't think this meaning makes any sense
24			prove formally; demonstrate by a mathematical, formal proof)	I think this is the author's intended meaning
24		testify, bear witness, evidence	provide evidence for	I think this is a possible unintended meaning
24		reveal	make clear and visible	I think this is a possible unintended meaning
24		testify, bear witness, prove, evidence	provide evidence for	I think this is a possible unintended meaning
24			finish third or better in a horse or dog race	I don't think this meaning makes any sense
25	Doors shall not close in less than 4 seconds in order that Passenger injury, particularly to children, the elderly and infirm is avoided. PPP Co shall provide calculations to ________ that the door timing meets this requirement.	raise, leaven	cause to puff up with a leaven	I don't think this meaning makes any sense
25		reveal	make clear and visible	I think this is a possible unintended meaning
25			prove formally; demonstrate by a mathematical, formal proof)	I think this is the author's intended meaning
25		demonstrate, establish, shew	establish the validity of something, as by an example, explanation or experiment	I think this is the author's intended meaning
25		turn out, turn up	be shown or be found to be	I think this is a possible unintended meaning
25		show up	be or become visible or noticeable	I think this is a possible unintended meaning
25		express, evince	give expression to	I don't think this meaning makes any sense
25		testify, bear witness, prove, evidence	provide evidence for	I think this is a possible unintended meaning
25		test, try, try out, examine, essay	put to the test, as for its quality, or give experimental use to	I think this is a possible unintended meaning
25		usher	show (someone to their seats, as in theaters or auditoriums	I don't think this meaning makes any sense
25			make visible or noticeable	I think this is a possible unintended meaning
25			take a trial impression of)	I don't think this meaning makes any sense
25			finish third or better in a horse or dog race	I don't think this meaning makes any sense
25		picture, depict,	show in, or as in, a picture	I think this is a possible

No.	Requirement Text	Synonyms	Definition	Answer
		render		unintended meaning
25		read, register, record	indicate a certain reading; of gauges and instruments	I think this is a possible unintended meaning
25			give evidence of, as of records	I think this is a possible unintended meaning
25		indicate, point	indicate a place, direction, person, or thing; either spatially or figuratively	I don't think this meaning makes any sense
25		testify, bear witness, evidence	provide evidence for	I think this is a possible unintended meaning
25		prove, demonstrate, establish, shew	establish the validity of something, as by an example, explanation or experiment	I think this is the author's intended meaning
25			obtain probate of	I think this is a possible unintended meaning
25		rise	increase in volume	I don't think this meaning makes any sense
26	It shall be possible to _________ the Driver's Main Operating Screen in any Cab of an unstabled Set to observe air pressures (Main Reservoir, Brake Pipe, Brake Cylinders), tractive effort, speed, and OHW voltage.	mobilize, mobilise, rally	call to arms; of military personnel)	I don't think this meaning makes any sense
26		issue	circulate or distribute or equip with	I don't think this meaning makes any sense
26		ply, cater	provide what is desired or needed, especially support, food or sustenance	I don't think this meaning makes any sense
26			bring forward for consideration	I think this is the author's intended meaning
26		add, append	state or say further	I don't think this meaning makes any sense
26		call, telephone, phone, ring	get or try to get into communication (with someone	I think this is a possible unintended meaning
26			bring forward for consideration	I think this is the author's intended meaning
26		advance	cause to move forward	I think this is a possible unintended meaning
26		remember, retrieve, recall, call back, recollect, think	recall knowledge from memory; have a recollection	I don't think this meaning makes any sense
26		render, furnish	provide or furnish with	I don't think this meaning makes any sense
27	The TOS shall be _________ to TOS screens mounted in the Driver's Workstation and Guard's Workstations for information display, alarm monitoring and data entry.	get in touch, touch base	establish communication with someone	I don't think this meaning makes any sense
27			plug into an outlet	I think this is the author's intended meaning
27		link, tie, link up	connect, fasten, or put together two or more pieces	I think this is the author's intended meaning
27			land on or hit solidly	I don't think this meaning makes any sense
27			establish a rapport or relationship	I don't think this meaning makes any sense
27			hit or play a ball successfully	I don't think this meaning makes any sense
27			be scheduled so as to provide continuing service, as in transportation	I think this is a possible unintended meaning
27		infix, enter	put or introduce into something	I don't think this meaning makes any sense
27		enclose, inclose, stick in, put in	introduce	I don't think this meaning makes any sense
27		link, link up, join, unite	be or become joined or united or linked	I think this is the author's intended meaning

No.	Requirement Text	Synonyms	Definition	Answer
27		associate, tie in, relate, link, colligate, link up	make a logical or causal connection	I don't think this meaning makes any sense
27			join by means of communication equipment	I think this is the author's intended meaning
27		slip in, stick in, sneak in	insert casually	I don't think this meaning makes any sense
27			join for the purpose of communication	I think this is the author's intended meaning
27		tuck	fit snugly into	I don't think this meaning makes any sense
27			plug into an outlet	I think this is the author's intended meaning
28	The camera locations and coverage shall be _________ and agreed during the Mock-up review.	award	give, especially as a reward	I don't think this meaning makes any sense
28		portray	represent in a painting, drawing, sculpture, or verbally	I think this is a possible unintended meaning
28			formally present a debutante, a representative of a country, etc.)	I don't think this meaning makes any sense
28		represent, lay out	bring forward and present to the mind	I don't think this meaning makes any sense
28		show, demo, exhibit	show or demonstrate something to an interested audience	I think this is the author's intended meaning
28		submit	hand over formally)	I think this is a possible unintended meaning
28		give, gift	give as a present; make a gift of	I don't think this meaning makes any sense
28		stage, represent	perform (a play), especially on a stage	I don't think this meaning makes any sense
28		introduce, acquaint	cause to come to know personally	I don't think this meaning makes any sense
28		march	march in protest; take part in a demonstration	I don't think this meaning makes any sense
28		deliver	deliver (a speech, oration, or idea)	I think this is a possible unintended meaning
28		certify, manifest, evidence	provide evidence for; stand as proof of; show by one's behavior, attitude, or external attributes	I think this is a possible unintended meaning
28		salute	recognize with a gesture prescribed by a miltary regulation; assume a prescribed position	I don't think this meaning makes any sense
28		prove, establish, show, shew	establish the validity of something, as by an example, explanation or experiment	I think this is a possible unintended meaning
28		pose	introduce	I think this is a possible unintended meaning
28		confront, face	present somebody with something, usually to accuse or criticize	I don't think this meaning makes any sense
28		show, demo, exhibit, present	show or demonstrate something to an interested audience	I think this is the author's intended meaning
29	The layout, quality, display and user interface of the internal CCTV images on the monitors shall be _________ and agreed during the Mock-up review.	introduce, acquaint	cause to come to know personally	I don't think this meaning makes any sense
29		salute	recognize with a gesture prescribed by a miltary regulation; assume a prescribed position	I don't think this meaning makes any sense
29		award	give, especially as a reward	I don't think this meaning makes any sense
29		prove, establish, show, shew	establish the validity of something, as by an example, explanation or experiment	I think this is a possible unintended meaning
29		confront, face	present somebody with something, usually to accuse or criticize	I don't think this meaning makes any sense
29		march	march in protest; take part in a demonstration	I don't think this meaning makes any sense

No.	Requirement Text	Synonyms	Definition	Answer
29		pose	introduce	I think this is a possible unintended meaning
29		show, demo, exhibit	show or demonstrate something to an interested audience	I think this is the author's intended meaning
29		submit	hand over formally)	I think this is the author's intended meaning
29			formally present a debutante, a representative of a country, etc.)	I don't think this meaning makes any sense
29		represent, lay out	bring forward and present to the mind	I don't think this meaning makes any sense
29		show, demo, exhibit	show or demonstrate something to an interested audience	I think this is the author's intended meaning
29		portray	represent in a painting, drawing, sculpture, or verbally	I don't think this meaning makes any sense
29		give, gift	give as a present; make a gift of	I don't think this meaning makes any sense
29		deliver	deliver (a speech, oration, or idea)	I think this is a possible unintended meaning
29		stage, represent	perform (a play), especially on a stage	I don't think this meaning makes any sense
29		certify, manifest, evidence	provide evidence for; stand as proof of; show by one's behavior, attitude, or external attributes	I think this is a possible unintended meaning
30	PPP Co shall ________ the use of recyclable materials throughout the Set.		seek or achieve an end by using to one's advantage	I don't think this meaning makes any sense
30		habituate	take or consume (regularly or habitually)	I think this is a possible unintended meaning
30			exploit the power of	I think this is a possible unintended meaning
30		apply, employ	put into service; make work or employ (something for a particular purpose or for its inherent or natural purpose	I think this is a possible unintended meaning
30		rein in, draw rein, rein	control and direct with or as if by reins	I don't think this meaning makes any sense
30		practice, apply	avail oneself to	I don't think this meaning makes any sense
30		rule, rein	keep in check	I think this is a possible unintended meaning
30			habitually do something (use only in the past tense)	I don't think this meaning makes any sense
30		tackle	put a harness	I think this is a possible unintended meaning
30		expend	use up, consume fully	I think this is the author's intended meaning
30			make the most of	I think this is the author's intended meaning
30			make as big or large as possible	I think this is a possible unintended meaning
31	The Car interior space shall be ________ with a Car interior width measured across the full width of the Car, of at least 2920 mm.		make the most of	I think this is the author's intended meaning
31			make as big or large as possible	I think this is a possible unintended meaning
31			make bigger or more	I think this is a possible unintended meaning
31			become bigger or greater in amount	I think this is a possible unintended meaning
32	The seat design and layout shall ________ the available legroom for seated passengers.		make the most of	I think this is the author's intended meaning
32			become bigger or greater in amount	I think this is a possible unintended meaning
32			make bigger or more	I think this is a possible unintended meaning
32			make as big or large as possible	I think this is a possible unintended meaning
33	All compressors shall be		become bigger or greater in amount	I think this is a possible

No.	Requirement Text	Synonyms	Definition	Answer
	controlled to _________ compressor life and optimise the efficiency of the air dryer system.			unintended meaning
33			make the most of	I think this is a possible unintended meaning
33			make bigger or more	I think this is a possible unintended meaning
33			make as big or large as possible	I think this is the author's intended meaning
34	The EAPS shall be _________ from the Main Power Supply in Fault or Failure conditions.		make disconnected, disjoin or unfasten)	I think this is the author's intended meaning
34		change over, shift	make a shift in or exchange of	I don't think this meaning makes any sense
34		switch over, exchange	change over, change around, or switch over)	I think this is a possible unintended meaning
34		throw, flip	cause to go on or to be engaged or set in operation	I don't think this meaning makes any sense
34			of electrical appliances)	I don't think this meaning makes any sense
34			unplug, disconnect -- (of electrical appliances)	I think this is the author's intended meaning
34			flog with or as if with a flexible rod)	I don't think this meaning makes any sense
34		trade, swap, swop	exchange or give (something	I don't think this meaning makes any sense
34		interchange, tack, alternate, flip, flip-flop	reverse (a direction, attitude, or course of action))	I don't think this meaning makes any sense
34		switch, shift	lay aside, abandon, or leave for another	I don't think this meaning makes any sense
35	The gangway shall be designed such that any litter dropped by users can be readily removed to eliminate any potential fire or health hazard, without having to _________ Cars or remove access panels.	ply, cater	provide what is desired or needed, especially support, food or sustenance	I don't think this meaning makes any sense
35		come off, come away	come to be detached	I think this is the author's intended meaning
35		issue	circulate or distribute or equip with	I don't think this meaning makes any sense
35		advertise, advertize, promote	make publicity for; try to sell (a product)	I don't think this meaning makes any sense
35		render, furnish	provide or furnish with	I don't think this meaning makes any sense
35		unplug	of electrical appliances)	I don't think this meaning makes any sense
35		bear on	press, drive, or impel (someone	I don't think this meaning makes any sense
35			military use: separate (a small unit	I don't think this meaning makes any sense
35			make disconnected, disjoin or unfasten)	I think this is a possible unintended meaning
35			cause to become detached or separated; take off	I think this is the author's intended meaning
35		crowd	approach a certain age or speed	I don't think this meaning makes any sense
35			press against forcefully without being able to move	I don't think this meaning makes any sense
35		add, append	state or say further	I don't think this meaning makes any sense
35		crusade, fight, press, campaign, agitate	exert oneself continuously, vigorously, or obtrusively to gain an end or engage in a crusade for a certain cause or person; be an advocate for	I don't think this meaning makes any sense
35			sell or promote the sale of (illegal	I don't think this meaning

No.	Requirement Text	Synonyms	Definition	Answer
			goods such as drugs)	makes any sense
35		press	make strenuous pushing movements during birth to expel the baby	I don't think this meaning makes any sense
35		tug, labor, labour, drive	strive and make an effort to reach a goal	I don't think this meaning makes any sense
35		force	move with force, "He pushed the table into a corner")	I think this is a possible unintended meaning
35			move strenuously and with effort	I think this is a possible unintended meaning
36	PPP Co shall _________ mechanisms which overcome the reception and accuracy problem typically encountered by GPS Systems within the Rail Network.	regard	look at attentively)	I think this is a possible unintended meaning
36		think, believe, conceive	judge or regard; look upon; judge	I think this is a possible unintended meaning
36		view	look at carefully; study mentally	I think this is a possible unintended meaning
36		take, deal	take into consideration for exemplifying purposes	I think this is a possible unintended meaning
36		debate, moot, turn over, deliberate	think about carefully; weigh	I think this is a possible unintended meaning
36		count, weigh	show consideration for; take into account	I think this is the author's intended meaning
36			regard or treat with consideration, respect, and esteem	I think this is a possible unintended meaning
36		take, deal	take into consideration for exemplifying purposes	I think this is a possible unintended meaning
36		see, reckon, view, regard	deem to be	I don't think this meaning makes any sense
36		study	give careful consideration to	I think this is the author's intended meaning
36		view	look at carefully; study mentally	I think this is a possible unintended meaning
37	In collisions where two trains are not aligned to enable the couplers and anti-climbers to fully engage PPP Co shall _________ the inclusion of additional Lateral members at Sole Bar and Cant Rail levels to provide improved offset angled collision resistance.	view	look at carefully; study mentally	I think this is a possible unintended meaning
37		think, believe, conceive	judge or regard; look upon; judge	I think this is a possible unintended meaning
37		regard	look at attentively)	I think this is a possible unintended meaning
37		view	look at carefully; study mentally	I think this is a possible unintended meaning
37			regard or treat with consideration, respect, and esteem	I think this is a possible unintended meaning
37		see, reckon, view, regard	deem to be	I don't think this meaning makes any sense
37		study	give careful consideration to	I think this is the author's intended meaning
37		count, weigh	show consideration for; take into account	I think this is the author's intended meaning
37		debate, moot, turn over, deliberate	think about carefully; weigh	I think this is a possible unintended meaning
37		take, deal	take into consideration for exemplifying purposes	I think this is a possible unintended meaning
37		take, deal	take into consideration for exemplifying purposes	I think this is a possible unintended meaning
38	PPP Co shall _________ any reasonable resulting corrective actions on all	guarantee	promise to do or accomplish	I think this is the author's intended meaning

No.	Requirement Text	Synonyms	Definition	Answer
	Sets to rectify the cause of the complaint.			
38		set about	enter upon an activity or enterprise)	I think this is a possible unintended meaning
38		try, seek, essay, assay	make an effort or attempt	I think this is a possible unintended meaning
38		take in charge	accept as a charge)	I don't think this meaning makes any sense
38		contract	enter into a contractual arrangement)	I don't think this meaning makes any sense
38		set about	enter upon an activity or enterprise)	I think this is a possible unintended meaning
38		tackle, take on	accept as a challenge	I think this is a possible unintended meaning
39	Level 1 diagnostic results shall include recommended remedial action required to be _________ by the Crew	try, seek, essay, assay	make an effort or attempt	I think this is a possible unintended meaning
39		contract	enter into a contractual arrangement)	I don't think this meaning makes any sense
39		take in charge	accept as a charge)	I don't think this meaning makes any sense
39		tackle, take on	accept as a challenge	I think this is a possible unintended meaning
39		set about	enter upon an activity or enterprise)	I think this is a possible unintended meaning
39		take in charge	accept as a charge)	I don't think this meaning makes any sense
39		guarantee	promise to do or accomplish	I think this is the author's intended meaning
39		set about	enter upon an activity or enterprise)	I think this is a possible unintended meaning
40	Level 2 diagnostic results shall consist of status, Fault information and recommended remedial action required to be _________ by the authorised personnel to maintain or restore the Set to operational specification in a depot environment.	set about	enter upon an activity or enterprise)	I think this is a possible unintended meaning
40		set about	enter upon an activity or enterprise)	I think this is a possible unintended meaning
40		guarantee	promise to do or accomplish	I think this is a possible unintended meaning
40		tackle, take on	accept as a challenge	I think this is a possible unintended meaning
40		take in charge	accept as a charge)	I don't think this meaning makes any sense
40		try, seek, essay, assay	make an effort or attempt	I think this is a possible unintended meaning
40		take in charge	accept as a charge)	I don't think this meaning makes any sense
40		contract	enter into a contractual arrangement)	I don't think this meaning makes any sense
40		take in charge	accept as a charge)	I don't think this meaning makes any sense

Appendix F: Prototype Software

```
/*=================================================*/
// Header File
/*=================================================*/

using namespace std;

const float SIMILARITY_THRESHOLD[] = {0.0, 0.25, 0.5, 0.75, 1.0};

// config settings
const int WORD_LEN = 60;          //needed for legacy WordNet C code
const int i_SIZE = sizeof(SIMILARITY_THRESHOLD)/sizeof(SIMILARITY_THRESHOLD[0]);

// file settings
const char* STATS_FILE = "c:\\stats.txt";
const char* CONTENT_FILE = "c:\\content.txt";
const char* EXPORT_PAIRS = "c:\\wn_file.txt";
const char* SIMILARITY_FILE = "c:\\file.txt";
const char* WORDNET_SIMILARITY = "c:\\perl\\bin\\perl.exe C:\\similarity.pl -type
WordNet::Similarity::wup -file c:\\wn_file.txt";
const char* OPTIMAL_CONSTRAINT_FILE = "c:\\optimal.txt";


// structures

struct struct_word
{
	string term;
	string pos;
	string sense;
};

struct struct_origin
{
	string RID;
	string position;
	struct_word word;
};

struct struct_feature
{
	struct_word word;
	vector<string> hypernyms;
	int frequency;
	int polysemy;
	int width;
	int tokens;
};

struct struct_relation
{
	struct_feature word1;
	struct_feature word2;
	float similarity;
	float replaceability;
};

struct struct_synset
{
	vector<string> synonyms;
	SynsetPtr synPtr;
	int level;
};

typedef set<string> S_str;
typedef vector<float> V_f;
typedef vector<string> V_str;
typedef set<V_str> S_V_str;
typedef vector<struct_word> V_word;
typedef vector<struct_origin> V_origin;
```

```
typedef vector<struct_feature> V_feature;
typedef vector<struct_relation> V_relation;
typedef vector<struct_synset> V_synset;

--------------------

#include <iostream>
#include <fstream>
#include <sstream>
#include <vector>
#include <set>
#include <string>
#include "wn.h"
#include "optimise.h"
#include "math.h"
#include <algorithm>

//================================================================================
// Function: PrintVector
// Purpose: displays a vector.
//================================================================================

void PrintVector( V_str& TheVector )
{
        for( size_t i = 0; i < TheVector.size(); ++i )
        {
                cout << TheVector[i] << " ";
        }
        cout << endl;
}

//================================================================================
// Function: unpackSynset
// Purpose: unpacks synPtr to term#pos#sense format
//================================================================================

V_str unpackSynset(SynsetPtr synPtr)
{
        V_str V_strSynonyms;

        std::ostringstream os;
        string term_pos_sense;

        for(int i=0; i<synPtr->wcount; i++)
        {
                term_pos_sense = "";
                term_pos_sense += synPtr->words[i];
                term_pos_sense += "#";
                term_pos_sense += synPtr->pos;
                term_pos_sense += "#";
                os.str("");
                os << synPtr->wnsns[i];
                term_pos_sense += os.str();
                V_strSynonyms.push_back(term_pos_sense);
        }
        return V_strSynonyms;
}

//================================================================================
// Function: returnSynset
// Purpose:
//================================================================================

V_str returnSynset(string term, string pos, string sense, bool b_Hypernym)
{
        V_str V_strSynset;
        char c_strTerm[WORD_LEN];
        int i_PosTerm, i_SenseTerm;
        SynsetPtr synPtr, synPtrHyp;

        strcpy(c_strTerm, term.c_str());
        i_SenseTerm = atoi(sense.c_str());
        if(pos == "v") i_PosTerm = VERB;
        if(pos == "n") i_PosTerm = NOUN;

        synPtr = findtheinfo_ds(c_strTerm, i_PosTerm, OVERVIEW, i_SenseTerm);
```

```
	V_strSynset.clear();
	if (b_Hypernym)
	{
		synPtrHyp = NULL;
		synPtrHyp = traceptrs_ds(synPtr, HYPERPTR, i_PosTerm, 1);
		if (synPtrHyp != NULL)
		{
			V_strSynset = unpackSynset(synPtrHyp);
			free_syns(synPtrHyp);
		}
	}
	else
	{
		if (synPtr != NULL)
		{
			V_strSynset = unpackSynset(synPtr);
			free_syns(synPtr);
		}
	}
	return V_strSynset;
}

//=================================================================================
// Function: stripTermPosSense
// Purpose:
// Parses string in format term#pos#sense to strip off term, pos, sense
//=================================================================================

struct_word stripTermPosSense (string str1)
{
	struct_word word1;
	string::size_type pos1, pos2;

	pos1 = str1.find('#',0);
	pos2 = str1.rfind('#',string::npos);

	if((pos1 != string::npos) && (pos2 != string::npos))
	{
		word1.term = str1.substr(0, pos1);
		word1.pos = str1.substr(pos1 + 1, pos2 - pos1 - 1);
		word1.sense = str1.substr(pos2 + 1, string::npos - pos2 - 1);
	}
	else
	{
		cout << "Error: '#' was not found twice in stripTermPosSense: " <<
str1 << endl;
		exit(0);
	}

	return word1;
}

//=================================================================================
// Function: importContentWords
// Purpose: Imports 'content' words from input file into V_contentWords
// Notes: accepts RIDs containing "." or "-".
//=================================================================================

void importContentWords( const char* filename, V_origin& V_origins )
{
	ifstream fin(filename);
	struct_origin originWord;
	struct_word wordSynonym;
	string strRID, strPosition, strTerm, strPos, strSense;
	V_str V_strSynset;
	V_str::iterator it_V_strSynset;

	V_origins.clear();

	while (!fin.eof())
	{
		fin >> strRID; //originWord.RID;
		fin >> strPosition; //originWord.position;
		fin >> strTerm; //originWord.word.term;
		fin >> strPos; //originWord.word.pos;
		fin >> strSense; //originWord.word.sense;
```

```
                V_strSynset = returnSynset(strTerm, strPos, strSense, false);
                for(it_V_strSynset = V_strSynset.begin(); it_V_strSynset !=
V_strSynset.end(); it_V_strSynset++)
                {
                        wordSynonym = stripTermPosSense (*it_V_strSynset);

                        originWord.word.term = wordSynonym.term;
                        originWord.word.pos = wordSynonym.pos;
                        originWord.word.sense = wordSynonym.sense;

                        if(*it_V_strSynset == strTerm + "#" + strPos + "#" + strSense)
// same word in file
                        {
                                originWord.RID = strRID;
                                originWord.position = strPosition;
                                V_origins.push_back(originWord);
                        }
                        originWord.RID = strRID;
                        originWord.position = "-";
                        V_origins.push_back(originWord);
                }

        }
        fin.close();
}

//==========================================================================================
// Function: measureLexicalFeatures
// Purpose: measures frequency, polysemy and width of each word in V_contentWords
// Notes: frequency, polysemy and width are based on usage (V_wordOrigins) not
dictionary.
// Changed to count frequency per sense rather than per term#pos.
//==========================================================================================

void measureLexicalFeatures( V_origin V_origins, V_feature& V_features )
{
        ofstream foutStats(STATS_FILE);
        V_origin::iterator it_V_origin1, it_V_origin2;
        S_str S_strSense, S_strSiblings, S_strExistingWords;
        S_str::iterator it_S_strExistingWords;
        struct_feature featuredWord;
        V_str V_strHypernyms1, V_strHypernyms2;
        V_str::iterator it_V_strHypernyms;
        V_feature::iterator it_V_feature1, it_V_feature2;
        string word;
        bool boolIdentical;

        S_strExistingWords.clear();          // prevents duplicates
        V_features.clear();

        for(it_V_origin1 = V_origins.begin(); it_V_origin1 != V_origins.end();
it_V_origin1++)
        {
                S_strSense.clear();
                S_strSiblings.clear();

                featuredWord.word.term = it_V_origin1->word.term;
                featuredWord.word.pos = it_V_origin1->word.pos;
                featuredWord.word.sense = it_V_origin1->word.sense;
                featuredWord.tokens = count(it_V_origin1->word.term.begin(),
it_V_origin1->word.term.end(), '_') + 1;

                word = featuredWord.word.term + "#" + featuredWord.word.pos + "#" +
featuredWord.word.sense;
                it_S_strExistingWords = S_strExistingWords.find(word);

                if (it_S_strExistingWords == S_strExistingWords.end())
                {
                        S_strExistingWords.insert(word);

                        // width
                        V_strHypernyms1 = returnSynset(featuredWord.word.term,
featuredWord.word.pos, featuredWord.word.sense, true);
                        featuredWord.hypernyms.clear();
                        for(it_V_strHypernyms = V_strHypernyms1.begin();
it_V_strHypernyms != V_strHypernyms1.end(); it_V_strHypernyms++)
```

```
                        {
                                featuredWord.hypernyms.push_back(*it_V_strHypernyms);
                        }

                        featuredWord.frequency = 0;
                        for(it_V_origin2 = V_origins.begin(); it_V_origin2 !=
V_origins.end(); it_V_origin2++)
                        {
                                if ((featuredWord.word.term == it_V_origin2-
>word.term) && (featuredWord.word.pos == it_V_origin2->word.pos))
                                {
                                        S_strSense.insert(it_V_origin2->word.sense);

                                        // MEASURE FREQUENCY PER SENSE - CHANGED:
13APR08
                                        if (featuredWord.word.sense == it_V_origin2-
>word.sense)
                                        {
                                                featuredWord.frequency =
featuredWord.frequency + 1;
                                        }
                                }
                        }
                        featuredWord.polysemy = S_strSense.size();
                        if ((featuredWord.frequency > 0) && (featuredWord.polysemy >
0))  // && (featuredWord.width > 0) && (featuredWord.units > 0))
                        {
                                V_features.push_back(featuredWord);
                        }
                        else
                        {
                                cout << "post-condition of measureLexicalFeatures not
satisfied";
                                exit(0);
                        }
                }
        }

        // width
        for(it_V_feature1 = V_features.begin(); it_V_feature1 != V_features.end();
it_V_feature1++)
        {
                it_V_feature1->width = 0;
                V_strHypernyms1 = it_V_feature1->hypernyms;

                if (V_strHypernyms1.size() > 0)
                {
                        for(it_V_feature2 = V_features.begin(); it_V_feature2 !=
V_features.end(); it_V_feature2++)
                        {
                                V_strHypernyms2 = it_V_feature2->hypernyms;
                                if (V_strHypernyms1 == V_strHypernyms2)
                                {
                                        it_V_feature1->width = it_V_feature1->width +
1;
                                }
                        }
                }
                else
                {
                        it_V_feature1->width = 1;
                }
                foutStats << it_V_feature1->word.term << "#" << it_V_feature1-
>word.pos << "#" << it_V_feature1->word.sense << "        " << it_V_feature1-
>frequency << "        " << it_V_feature1->polysemy << "        " << it_V_feature1-
>width << endl;
        }
        foutStats.close();
        return;
}

//==========================================================================================
// Function: pairWords
// Purpose:
// Returns vector of string pairs. Pairs are determined from V_str1 input string.
// Only returns one order - e.g. if {A,B} then wont also return {B,A} since similarity
// will be the same for both orders. Exports pairs to file based on b_Export
```

```
//==========================================================================================

pairWords( V_word V_words )
{
	V_word::iterator it_V_word1, it_V_word2;
	ofstream foutPairs(EXPORT_PAIRS);

	for(it_V_word1 = V_words.begin(); it_V_word1 != V_words.end(); it_V_word1++)
	{
		for(it_V_word2 = V_words.begin(); it_V_word2 != V_words.end();
it_V_word2++)
		{
			if (it_V_word1->pos == it_V_word2->pos)
			{
				foutPairs << it_V_word1->term << "#" << it_V_word1-
>pos << "#" << it_V_word1->sense << " " << it_V_word2->term << "#" << it_V_word2->pos
<< "#" << it_V_word2->sense << endl;
			}
		}
	}
	foutPairs.close();
	return;
}

//==========================================================================================
// Function: importSimilarityValues
// Purpose:
// Imports the result of WordNet::Similarity. I.e. WN Sim outputs a file, then this
// function reads that file into V_sim WnSimilarity. This is then used later to calc
// similarity between words.
//==========================================================================================

void importSimilarityValues( V_feature V_features, V_relation& V_relations )
{
	ifstream fin(SIMILARITY_FILE);
	V_feature::iterator it_V_feature;
	struct_relation struct_relations;
	string strWord1, strWord2, strFeatureWord, strSimilarity;

	while (!fin.eof())
	{
		fin >> strWord1;
		fin >> strWord2;
		fin >> strSimilarity;

		for(it_V_feature = V_features.begin(); it_V_feature !=
V_features.end(); it_V_feature++)
		{
			strFeatureWord = it_V_feature->word.term + "#" + it_V_feature-
>word.pos + "#" + it_V_feature->word.sense;
			if (strWord1 == strFeatureWord) struct_relations.word1 =
*it_V_feature;
			if (strWord2 == strFeatureWord) struct_relations.word2 =
*it_V_feature;
		}
		struct_relations.similarity = strtod(strSimilarity.c_str(),NULL);
		V_relations.push_back(struct_relations);
	}
	fin.close();
	return;
}

//==========================================================================================
// Function: calculateSimilarity
// Purpose:
//==========================================================================================

void calculateSimilarity( V_feature V_features, V_relation& V_relations )
{
	V_feature::iterator it_V_feature;
	V_word V_words;
	struct_word struct_word1;

	V_words.clear();
	for(it_V_feature = V_features.begin(); it_V_feature != V_features.end();
it_V_feature++)
	{
```

```
                struct_word1.term = it_V_feature->word.term;
                struct_word1.pos = it_V_feature->word.pos;
                struct_word1.sense = it_V_feature->word.sense;
                V_words.push_back(struct_word1);
        }
        pairWords( V_words );
        system(WORDNET_SIMILARITY);

        importSimilarityValues( V_features, V_relations );
        return;
}

//=================================================================================
// Function: calculateReplaceability
// Purpose:
//=================================================================================

void calculateReplaceability( V_relation& V_relations )
{
        V_relation::iterator it_V_relate;
        float f_frequency1, f_frequency2, f_frequency2_1;
        float f_polysemy1, f_polysemy2, f_polysemy1_2;
        float f_tokens1, f_tokens2, f_tokens1_2;
        float f_width1, f_width2, f_width2_1;
        float f_similarity; //, f_replaceability;

        for(it_V_relate = V_relations.begin(); it_V_relate != V_relations.end();
it_V_relate++)
        {

                f_frequency1 = it_V_relate->word1.frequency;
                f_frequency2 = it_V_relate->word2.frequency;
                f_polysemy1 = it_V_relate->word1.polysemy;
                f_polysemy2 = it_V_relate->word2.polysemy;
                f_width1 = it_V_relate->word1.width;
                f_width2 = it_V_relate->word2.width;
                f_tokens1 = it_V_relate->word1.tokens;
                f_tokens2 = it_V_relate->word2.tokens;
                f_similarity = it_V_relate->similarity;


                it_V_relate->replaceability = f_similarity * sqrt((f_frequency2 +
f_width2)/(f_frequency1 + f_width1)) * (f_polysemy1 / f_polysemy2);
                cout << " -> (" << it_V_relate->replaceability << ") ";
                cout << it_V_relate->word2.word.term << "#" << it_V_relate-
>word2.word.pos << "#" << it_V_relate->word2.word.sense;
                cout << endl;
        }
        return;
}

//=================================================================================
// Function: constrainSpecification
// Purpose:
//=================================================================================

void constrainSpecification( V_origin V_origins, V_relation V_relations )
{
        ofstream foutConstraint(OPTIMAL_CONSTRAINT_FILE);
        V_origin::iterator it_V_origin;
        V_relation::iterator it_V_relation;
        V_relation V_relationReplacement(i_SIZE);
        V_f V_fMaxReplaceability(i_SIZE);
        int i;

        string wordOrigin, wordRelation;

        for(it_V_origin = V_origins.begin(); it_V_origin != V_origins.end();
it_V_origin++)
        {
                V_relationReplacement.clear();
                V_fMaxReplaceability.clear();
                for(i = 0; i < i_SIZE; i++)
                {
                        V_fMaxReplaceability.push_back(0);
                }
```

```
                for(it_V_relation = V_relations.begin(); it_V_relation !=
V_relations.end(); it_V_relation++)
                {
                        if ((it_V_relation->word1.word.term == it_V_origin->word.term)
&& (it_V_relation->word1.word.pos == it_V_origin->word.pos) && (it_V_relation-
>word1.word.sense == it_V_origin->word.sense))
                        {
                                for(i = 0; i < i_SIZE; i++)
                                {
                                        if ((it_V_relation->replaceability >
V_fMaxReplaceability[i]) && (it_V_relation->similarity >= SIMILARITY_THRESHOLD[i]))
                                        {
                                                V_fMaxReplaceability[i] =
it_V_relation->replaceability;
                                                V_relationReplacement[i] =
*it_V_relation;
                                        }
                                }
                        }
                }

                foutConstraint << it_V_origin->RID << "         " << it_V_origin-
>position << "         " << it_V_origin->word.term << "#" << it_V_origin->word.pos <<
"#" << it_V_origin->word.sense << "         ";
                for(i = 0; i < i_SIZE; i++)
                {
                        foutConstraint << V_fMaxReplaceability[i] << "         ";
                        if (V_fMaxReplaceability[i] == 1)
                        {
                                foutConstraint << "         " << it_V_origin->word.term
<< "#" << it_V_origin->word.pos << "#" << it_V_origin->word.sense << "         ";
                        }
                        else // > 1
                        {
                                foutConstraint << "(" <<
V_relationReplacement[i].similarity << ")";
                                foutConstraint << "sqrt((" <<
V_relationReplacement[i].word2.frequency << "+" <<
V_relationReplacement[i].word2.width << ")/(" <<
V_relationReplacement[i].word1.frequency << "+" <<
V_relationReplacement[i].word1.width << "))";
                                foutConstraint << "(" <<
V_relationReplacement[i].word1.polysemy << "/" <<
V_relationReplacement[i].word2.polysemy << ")" << "         ";
                                foutConstraint <<
V_relationReplacement[i].word2.word.term << "#" <<
V_relationReplacement[i].word2.word.pos << "#" <<
V_relationReplacement[i].word2.word.sense << "         ";
                        }
                }
                foutConstraint << endl;
        }
        foutConstraint.close();
        return;
}

//==========================================================================================
// Function: optimiseLexicon
// Purpose:
//==========================================================================================

void optimiseLexicon()
{
        V_origin V_origins;
        V_feature V_features;
        V_relation V_relations;

        importContentWords( CONTENT_FILE, V_origins );
        measureLexicalFeatures( V_origins, V_features );
        calculateSimilarity( V_features, V_relations );
        calculateReplaceability( V_relations );
        constrainSpecification( V_origins, V_relations );

        return;
}

//==========================================================================================
```

```
// Function: main
// Purpose:
//-----------------------------------------------------------------------------------------

int main()
{
        if (wninit())
        {
                cout << "Fatal error - cannot open WordNet database" << endl;
                exit (-1);
    }
        optimiseLexicon();

        return 0;
}
```

Bibliography

Adi-wijaya, A., 2003, 'BOYDA: Object Oriented Requirements Engineering', University of Technology, Sydney, Unpublished Capstone Project Report

Adriaens, G., and Schreurs, D. 1992, 'From Cogram to Alcogram: Towards a Controlled English Grammar Checker', *COLING 92*, France, pp. 595-601.

AECMA, 1986, *AECMA/AIA Simplified English: A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language,* Association Europeenne des Constructueurs de Materiel Aerospatial

Agirre, E., and Rigau, G., 1996, 'Word sense disambiguation using conceptual density', *Proceedings of COLING'96*, Copenhagen, Denmark, pp. 16-22.

Alexander, I.F., and Stevens, R., 2002, *Writing Better Requirements*, Addison-Wesley Professional.

Allen, J., 2004, 'How are we responding to industrial and business needs for Controlled Language Machine Translation', *Journées Linguistiques – Langues contrôlées; traduction automatique et langues spécialisées*, Besançon, France, viewed 10 January 2007, <http://www.geocities.com/jeffallenpubs/besanconmay2004.htm>.

Allen, J., 1995, *Natural Language Understanding*, Second Edition, Addison-Wesley, Reading, MA, USA

Alqvist, I., and Sagvall Hein, A., 1996, 'Defining ScaniaSwedish - a Controlled Language for Truck Maintenance', *Proceedings of the First International Workshop on Controlled Language Applications (CLAW96)*, Belgium, pp. 159-167

Ambriola, V., and Gervasi, V., 2006, 'On the Semantic Analysis of Natural Language Requirements with CIRCE', *Automated Software Engineering Journal,* vol 13, no.1, pp.107-167

Bach, K., 2000, *Routledge Encyclopedia of Philosophy – Ambiguity*, London, viewed 20 September 2008, <http://online.sfsu.edu/~kbach/ambguity.html>

Bale, L.S., 1995, 'Cybernetics and the Social Behavioral Sciences', *Cybernetics and Human Knowing: a Journal of Second Order Cybernetics and Cyber-Semiotics,* vol. 3, no. 1, pp. 27-45.

Basili, V., Caldiera, G., Lanubile, F., and Shull, F., 1996, 'Studies in Reading Techniques', *Proceedings of the 21st Annual Software Engineering Workshop, Software Engineering Laboratory Series, SEL-96-002*, Goddard Space Flight Center, Greenbelt, Maryland, USA, pp. 59-65.

Beckwith, R., and Miller, G.A, 1993, 'Introduction to WordNet: an on-line lexical database', *International Journal of Lexicography,* vol. 3, no. 4, pp.235-244.

Bell, T.E., Thayer, R.H., 1976, 'Software Requirements: Are They Really a Problem?' *IEEE International Conference on Software Engineering,* Los Alamitos, California, USA, pp. 61-68.

Ben Achour, C., 1998, 'Guiding Scenario Authoring', *Proceedings of the Eighth European-Japanese Conference on Information Modeling and Knowledge Bases*, IOS Press, Vamala, Finland, pp. 152–171

Berlin, B., 1992, *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*, Princeton: Princeton University Press.

Bernhard, T., 1993, *Systems Engineering: Principles and Practice of Computer-based Systems Engineering*, Chichester: John Wiley & Sons.

Bernth, A., 1997, 'Easyenglish: A Tool for Improving Document Quality', *Fifth Conference on Applied Natural Language Processing*, Washington, USA, pp 159-165.

Berry, D.M., Kamsties, E., and Krieger, M.M. 2003, From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity, A Handbook, *Technical Report, School of Computer Science*, University of Waterloo, Waterloo, Ontario, Canada

Berry, D.M., and Kamsties, E., 2003, 'Ambiguity in Requirements Specification', *Perspectives on Software Requirements Series: The Springer International Series in Engineering and Computer Science, Chapter 2,* Sampaio do Prado Leite, J.C; Doorn, J.H (Eds.), Kluwer Academic Publishers, pp.191-194

Boehm, B., 1975, 'Some experiences with automated aids to the design of largescale reliable software', *IEEE Transactions on Softwar Engineering*, vol.1, no.1, pp.125–133.

Boehm BW., 1984, 'Verifying and Validating Software Requirements and Design Specifications', *IEEE Software*, vol.1, no.1, pp. 75-88.

Boyd, S., Zowghi, D., and Farroukh, A., 2005, 'Measuring the Expressiveness of a Constrained Natural Language: An Empirical Study', *Proceedings of the 13th International Conference on Requirements Engineering (RE'05)*, Paris, France, pp.339-349.

Boyd, S., Zowghi, D., and Gervasi, V., 2007, 'Optimal-Constraint Lexicons for Requirement Specifications', *International Working Conference on Requirements Engineering Foundations for Software Quality (REFSQ '07)*, Trondheim, Norway, pp.203-217.

Brooks, F.P.Jnr, 1987, 'No Silver Bullet: Essence and accidents of software engineering', *IEEE Computer*, vol. 15, no. 1, pp. 10-18.

Brown, R., 1958, 'How shall a thing be called?' *Psychological Review*, vol. 65, pp.14-21.

Brown, P. F., Della-Pietra, S., Della-Pietra, V.J., and Mercer R.L., 1991, 'Word sense disambiguation using statistical methods.' *Proceedings of the 29th Annual Meeting of Association for Computational Linguistics*, Berkeley, California, USA, pp. 264-270.

Buley, E.R., Moore, L.J., and Owess, M.F., 1988, 'B5 (SRS/IRS) Specification Guidelines', Technical *Report M88-57, ESD-TR-88-337, MITRE*, Bedford, MA, USA.

Byrd, R. J. Calzolari, N., Chodorow, M. S., Klavans, J. L., Neff, M. S. and Rizk, O. A., 1987, 'Tools and methods for computational lexicology', *Computational Linguistics*, vol. 13, no 3-4, pp. 219-40.

Cadwell, P., 2008, 'Readability and Controlled Language: does the study of readability have merit in the field of controlled language, and is readability increased by applying controlled-language rules to texts?' *Dissertation-Degree of MA in Translation Studies,* Dublin City University, School of Applied Language and Intercultural Studies.

Campbell, D.T., 1986, 'Relabeling internal and external validity for applied social scientists', W.Trochim (Ed.), *Advances in quasi-experimental design and analysis*, San Francisco, Jossey-Bass, pp. 67-78.

Carbonell, J.G., Mitamura, T., and Nyberg, E.H, 1992, 'The KANT Perspective: A Critique of Pure Transfer (and Pure Interlingua, Pure Statistics, …)', in *Quatrième Colloque international sur les aspects théoriques et méthodologiques de la traduction automatique, Fourth International Conference on Theoretical and Methodological Issues in Machine Translation, TMI-92*, Montréal, Canada, pp. 225–235.

Chantree, F., 2004, 'Ambiguity Management in Natural Language Generation', *7th Annual CLUK Research Colloquium,* Birmingham, England.

Chervak, S., Drury, C., and Ouellette, J., 1996, 'Field Evaluation of Simplified English for Aircraft Workcards', *Proceedings of the 10th FAA/AAM Meeting on Human Factors in Aviation Maintenance and Inspection*, Washington, USA, pp.123-136.

Choueka, Y., and Lusignan, S., 1985, 'Disambiguation by short contexts', *Computers and the Humanities*, vol.19, pp. 147-158.

Cowie, J., Guthrie, J.A., and Guthrie, L., 1992, 'Lexical disambiguation using simulated annealing', *Proceedings of the 14th International Conference on Computational Linguistics, COLING'92,* 23-28 August, Nantes, France, vol. 1, pp. 359-365.

Cruse, D. A., 1986, *Lexical Semantics*, Cambridge, Eng.: Cambridge University Press, pp. 49-68.

Crystal, D., 1992, *An encyclopedic dictionary of language and languages*, Blackwell, Oxford, UK; Cambridge, Mass., USA, p. 326.

Davis, A., 2005, *Just Enough Requirements Management*, Dorset House Publishing, New York, USA.

Davis, A., and T. Rauscher, 1979, 'Formal Techniques and Automated Processing to Ensure Correctness in Requirements Specifications', *IEEE Specifications of Reliable Software Conference*, Los Alamitos, California: IEEE Computer Society Press, USA, pp. 15-35.

Davis, A., Overmyer, S., Jordan, K., Caruso, J., Dandashi, F., Dinh, A., Kincaid, G., Ledeboer, G., Reynolds, P., Sitaram, P.A, Ta, A., and Theofanos, M. 1993, 'Identifying and Measuring Quality in a Software Requirements Specification', *Proceedings of the First International Software Metrics Symposium*, Baltimore, USA, pp. 141-152.

Davis, A., 1993, *Software Requirements: Objects, Functions and States*, Prentice Hall, Englewood Cliffs, N.J, USA.

Dictionary.com, LLC 2008, viewed 3 November 2008, <http://dictionary.reference.com>

Department of Defence MIL-STD-490A, 1985, *Specification Practices*, USA.

Dodd, T., 1990, *Prolog: A Logical Approach*, Oxford University Press, New York, USA.

Douglas, S., and Hurs, M. 1996, 'Controlled Language Support for Perkins Approved Clear English (PACE)', *Proceedings of the 1st International Workshop on Controlled Language Applications (CLAW'96)*, Leuven, Belgium, pp. 93-105.

Easterbrook, S.M. and Callahan, J.R., 1998, 'Formal Methods for Verification and Validation of Partial Specifications: A Case Study', *Journal Systems and Software*, vol.40, no.3, pp. 199–210.

Easterbrook, S. M., Singer, J., Storey, M, and Damian, D., 2007, 'Selecting Empirical Methods for Software Engineering Research', In F. Shull and J. Singer (eds) '*Guide to Advanced Empirical Software Engineering*', Springer.

Electronic Industries Association, EIA-632, 2003, *Processes for Engineering a System,* Information Technology Association of America.

Fabbrini, F., Fusani, M., Gervasi, V., Gnesi, S., and Ruggieri, S. 1998, 'Achieving Quality in Natural Language Requirements', *Proceeding of the 11th International Software Quality Week*, San Francisco, USA.

Fabbrini, F., Fusani, M., Gnesi, G., and Lami, G., 2001, "An Automatic Quality Evaluation for Natural Language Requirements", *Proceedings of the Seventh International Workshop on RE: Foundation for Software Quality (REFSQ'2001)*, Interlaken, Switzerland.

Farroukh, A., 2005, 'To define and use expressiveness to empirically derive the verbs of BOYDA for a particular domain', Faculty of Engineering, University of Technology, Sydney, Unpublished Capstone Project Report.

Faulk, S.R., 1997, *Software Requirements: A Tutorial, in Software Engineering*, IEEE Computer Society Press, Los Alamitos, pp.82-101.

Fellbaum, C., 1990, 'English Verbs as a Semantic Net', *International Journal of Lexicography*, vol. 3, no. 4, pp. 278-301.

Firth, J., 1957, 'A Synopsis of Linguistic Theory 1930-1955' in *Studies in Linguistic Analysis*, Philological Society, Oxford; reprinted in Palmer, F. (ed. 1968) Selected Papers of J. R. Firth, Longman, Harlow.

Fisher, R.A., and Yates, F., 1974, 'Statistical Tables for Biological, Agricultural and Medical Research', Longman Group Ltd, London, UK.

Fowler, M., and Scott, K., 2000, *UML Distilled: A Brief Guide to the Standard Object Modelling Language*, Pearson Education Asia.

Fraser, M.D, Kumar, K., and Vaishnavi, V. 1991, 'Informal and formal requirements specification languages: bridging the gap', *IEEE Transactions on Software Engineering*, vol. 17, no. 5, pp. 454-66.

Frath, P., 2001, 'Polysemy, Homonymy and Reference', *Proceedings of the JASGIL Seminar*, Albert Hamm ed., in RANAM (Recherches Anglaises et Nord-Américaines), no. 34, Strasbourg.

Fuchs, N.E. and Schwitter, R., 1996, 'Attempto Controlled English (ACE)', *Proceedings of the First International Workshop on Controlled Language Applications (CLAW96)*, Belgium, pp. 124-136.

Fuchs, N.E, Schwitter, R. 1995, 'Specifying Logic Programs in Controlled Natural Language', *Workshop on Computational Logic for Natural Language Processing*, Edinburgh, UK.

Fuchs, N.E., Schwertel, U., and Schwitter, R., 1998, 'Attempto Controlled English - Not Just Another Logic Specification Language', *Proceedings of the 8th International Workshop on Logic Programming Synthesis and Transformation,* Manchester, UK.

Gause, D.C., 2000, User *DRIVEN Design—The Luxury that has Become a Necessity, A Workshop in Full Life-Cycle Requirements Management*, ICRE 2000 Tutorial T7, Schaumberg, USA.

Gause, D.C. and Weinberg, G.M., 1989, *Exploring Requirements: Quality Before Design*, Dorset House, New York, USA.

Gay, L., & Airasian, P. 2003, *Educational research:Competencies for analysis and applications*, Englewood, NJ: Prentice Hall, pp.113.

Gazdar, G., Klein, E., Pullum, G., and I. Sag, 1985, *Generalized Phrase Structure Grammar*, Basil Blackwell, London, UK.

Gervasi, V. and Nuseibeh, B., 2000, 'Lightweight Validation of Natural Language Requirements', *Proceedings of Fourth IEEE International Conference on Requirements Engineering (ICRE'2000)*, IEEE Computer Society Press, Schaumburg, IL, pp. 140–148.

Gillon, B., 1990. Ambiguity, generality, and indeterminacy: Tests and definitions, *Synthese*, vol. 85, no.91, pp. 391-416.

Gnesi, S., Lami, G., Trentanni, G., Fabbrini, F., and Fusani, M, et al 2005, 'An Automatic Tool for the Analysis of Natural Language Requirements', *International Journal of Computer Systems Science & Engineering*, vol. 20, no. 1, pp. 53-62.

Goguen, J.A. 1994, 'Requirements Engineering as the Reconciliation of Technical and Social Issues', pp. 165–199 in *Requirements Engineering: Social and Technical Issues*, ed. J.A. Goguen and M. Jirotka, Academic Press, London, UK.

Goyvaerts, P., 1996, 'Controlled English, Curse or Blessing? - A User's Perspective', *Proceedings of the First International Workshop on Controlled Language Applications*, Belgium, pp.139-204.

Gray, J., 2000, *Collection of Ambiguous or Inconsistent/Incomplete Statements*, Vanderbilt University, viewed 2 March 2007, <http://www.gray-area.org/Research/Ambig/>.

Green, R., 2006, 'Vocabulary Alignment via Basic Level Concepts', *OCLC/ALISE research grant report published electronically by OCLC Research*, viewed 6 October 2008, <http://www.oclc.org/research/grants/reports/green/rg2005.pdf>.

Grover, C., Holt, A., Klein, E., and Moens, M., 2000, 'Designing a Controlled Language for Interactive Model Checking', *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW00)*, Washington, USA, pp. 90-104.

Grover, C., Carroll, J., and Briscoe, T., 1993, *The Alvey Natural Language Tools grammar (4th release),* Technical Report 284, Computer Laboratory, University of Cambridge.

Hall, A.D., 1962, *A Methodology for Systems Engineering*, Van Nostrand Reinhold.

Haraldsson, H.V., 2004, 'Introduction to systems thinking and causal loop diagrams', *Technical report, Lund University, Department of Chemical Engineering. Reports in Ecology and Environmental Engineering*, pp. 1-49.

Harwell, R., Aslaksen, E., Hooks, I., Mengot, R., and Ptack, K. 1993, 'What is a Requirement?', *Proceedings of the Third Annual International Symposium*, National Council of Systems Engineers (NCOSE), Arlington, pp. 17–24.

Hevner, A.R., March, S.T., Park, J., Ram, S., 2004, 'Design Science in Information Systems Research', *MIS Quarterly*, vol.28, no.1, pp. 75-105.

Hirst, G., 1987, *Semantic Interpretation and the Resolution of Ambiguity*, Studies in Natural Language processing, Cambridge University Press, Cambridge, UK.

Hladczuk, J., and Eller, W., 1992, *International Handbook of Reading Education*, Greenwood Publishing Group, USA.

Holmback, H., Shubert, S. and Spyridakis, J., 1996, 'Issues in Conducting Empirical Evaluations of Controlled Languages', *Proceedings of the First International Workshop on Controlled Language Applications (CLAW'96)*, University of Leuven, Belgium, pp.166–177.

Hooks, I., 1994, 'Writing good requirements', *Proceedings of the Third International Symposium of the NCOSE*, Vol. 2, pp.197-203.

Hsia, P., Davis, A.M., and Kung, D.C., 1993, 'Status Report: Requirements Engineeering', *IEEE Software*, vol. 10, no. 6, pp.75-79.

Huijsen, W., 1998, 'Controlled Language – An Introduction', *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW'98)*, Pittsburgh, PA, pp. 1–15

Hull, E., Jackson, K., and Dick, J., 2002, *Requirements Engineering*, Springer-Verlag, London, UK.

IBM (previously Telelogic), 2008, *DOORS:Requirements Definition and Management*, viewed 28 June 2008 <http://www.telelogic.com/products/doors/index.cfm>

Ide, N., and Véronis, J., 1998, 'Word Sense Disambiguation: The State of the Art', *Journal of Computational Linguistics*, Vol 24, No.1, pp.1-40.

IEEE Std 610.12-1990, *IEEE Standard Glossary of Software Engineering Terminology*, Institute of Electrical and Electronics Engineering, New York, USA.

IEEE Std 830-1993, *IEEE Recommended Practice for Software Requirements Specifications*, Institute of Electrical and Electronics Engineering, New York, USA.

IEEE Std 1220-1994, *IEEE Trial-Use Standard for Application and Management of the Systems Engineering Process*, Institute of Electrical and Electronics Engineering, New York, USA.

International Council on Systems Engineering 2005, *INCOSE*, Seattle, viewed 10 July 2005, <http://www.incose.org/>

Ishihara, Y., Seki, H., Kasami, T., 1993, 'A Translation Method from Natural Language Specifications into Formal Specifications Using Contextual Dependencies', *Proceedings of IEEE International Symposium on Requirements Engineering*, San Diego, IEEE Computer Society Press, pp.232-239.

Jackson, M., 1995, *Software Requirements & Specifications: A Lexicon of Practice, Principles, and Prejudices*, Addison-Wesley, London, UK.

Jorgensen, J., 1990, 'The psychological reality of word senses', *Journal of Psycholinguistic Research*, vol.19, pp. 167-190.

Kamp, H., and Reyle, U., 1993, *From Discourse to Logic*, Kluwer Academic Publishers, Dordrecht.

Kamprath, C., Adolphson, E., Mitamura, T., and Nyberg, E., 1998, 'Controlled Language for Multilingual Document Production: Experience with Caterpillar Technical English', *Proceedings of the Second International Workshop on Controlled Language Applications (CLAW98)*, Pennsylvania, pp. 51-61.

Kamsties, E., Berry, D.M., and Paech, B. 2001, 'Detecting Ambiguities in Requirements Documents Using Inspections', *Workshop on Inspections in Software Engineering (WISE'01)*, Paris, France, pp. 68-80.

Kamsties, E., von Knethen, A., Philipps, J., and Schaetz, B., 2001, 'An Empirical Investigation of the Defect Detection Capabilities of Requirements Specification Techniques', *Proceedings of the Sixth CAiSE/IFIP8.1 International Workshop on Evaluation of Modelling Methods in Systems Analysis and Design (EMMSAD'01)*, Interlaken, Switzerland.

Kamsties, E. and Paech, B., 2000, 'Taming Ambiguity in Natural Language Requirements', *Proceedings of the Thirteenth International Conference on Software and Systems Engineering and Applications*, Paris, France.

Kamsties, E., 2001, 'Surfacing Ambiguity in Natural Language Requirements', Ph.D. Dissertation, Fachbereich Informatik, Universit¨ at Kaiserslautern, Germany, also Volume 5 of Ph.D. Theses in Experimental Software Engineering, Fraunhofer IRB Verlag, Stuttgart, Germany.

Kincaid, C. D., 1997, 'A Validation Study of Simplified English as a Facilitator in English for Special Purpose Language Learning' *Technical report, Univ. of Central Florida*, Orlando, Florida

Kitchenham, B.A.; Pfleeger, S.L.; Pickard, L.M.; Jones, P.W.; Hoaglin, D.C.; El Emam, K.; Rosenberg, J., 2002, 'Preliminary Guidelines for Empirical Research in Software Engineering', *IEEE Transactions on Software Engineering,* Vol.28, No.8, pp.721-734.

Kiyavitskaya N., Zeni N., Mich L., and Berry D., 2008, 'Requirements for Tools Ambiguity Identification and Measurement in Natural Language Requirements Specification', *Requirements Engineering*, vol. 13, no. 766, pp. 1-1.

Klare, G.R., 1974, 'Assessing readability', *Reading research quarterly*, vol. 10, no. 1, pp.62-102.

Kovitz, B.L., 1998, *Practical Software Requirements: A Manual of Content and Style*, Manning, Greenwich, CT.

Knops, U., 2000, 'Efficient roll-in and roll-out of controlled language applications', *Proceedings of the third international workshop on controlled language applications*, Seattle, USA, pp.134-135.

Krogstie J., Lindland O.I., Sindre G. 1995, 'Towards a deeper understanding of quality in requirements engineering'. *7th International CaiSE Conference*, Jyvaskyla, Finland, pp. 82-95.

Lakoff, G., 1987, *Women, fire, and dangerous things: What categories reveal about the mind*, Chicago: University of Chicago Press, USA, pp,46-47.

Lami, G., 2005, 'QuARS: A Tool for Analyzing Requirements', *Technical Report CMU/SEI-2005-TR-014 ESC-TR-2005-014*, Software Engineering Institute, Carnegie Mellon University, viewed 25 August 2008, <http://www.sei.cmu.edu/publications/documents/05.reports/05tr014.html>

Learn English, 2008, *English Grammar website,* viewed 5 September 2008, <http://www.learnenglish.de/grammar/sentencetext.htm>

Leedy, P.D., and Ormrod, J.E, 2005, *Practical Research: Planning and Design,* 8th Edition, Pearson Education Inc, New Jearsey, USA.

Leite, J.C.S.P. and Franco, A.P.M., 1993, 'A Strategy for Conceptual Model Acquisition', *Proceedings of the IEEE International Symposium on Requirements Engineering*, IEEE Computer Society Press, San Diego, USA, pp. 243–246.

Lesk, M., 1986, 'Automated Sense Disambiguation Using Machine-readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone', *Proceedings of the 1986 SIGDOC Conference*, Toronto, Canada, pp. 24-26.

Levinson, S., 1983, *Pragmatics*, Cambridge University Press, Cambridge, UK.

Lutz, R.R., 1993, 'Analysing Software Requirements Errors in Safety-Critical Embedded Systems', *Proceedings of International Symposium on Requirements Engineering, RE'03*, San Diego, IEEE, pp.126-133.

Lyons, J., 1977, *Semantics I and II*, Cambridge University Press, Cambridge, UK.

Macias, B., and Pulman, S.,1993, *Natural Language Processing for Requirements Specifications, in: F. Redmill, T. Anderson* (eds.), Safety-Critical Systems, Current Issues, Techniques and Standards, Chapman & Hall, pp. 67-89.

Manning, C.D, and Schütze, H., 1999, *Foundations of Statistical Natural Language Processing,* The MIT Press, Cambridge, Massachusetts.

March, S.T., and Smith, G.F., 1995, 'Design and natural science research on information technology', Decision Support Systems, vol.15, no.4, pp.251-266.

Marszal, E., and Scharpf, E., 2002, *Safety Integrity Level (SIL) Selection*, ISA

Martin, J. N., 2008, *Overview of the EIA 632 Standard – Processes for Engineering a System,* Raytheon, viewed 7 September 2008, *<http://sepo.spawar.navy.mil/632_Tutorial.pdf>*

Merriam Webster English Dictionary, 2004, Merriam-Webster; Revised edition.

Mich, L. and Garigliano, R., 2000, 'Ambiguity Measures in Requirements Engineering', *Proceedings of the International Conference on Software Theory and Practice (ICS2000), Sixteenth IFIP World Computer Conference*, Beijing, China.

Mich, L., Franch, M., and Novi Inverardi, P., 2002, 'Market research for requirements analysis using linguistic tools', *Technical Report 66*, Informatica e Studi Aziendali, University of Trento, viewed 10 September 2008, <http://eprints.biblio.unitn.it/view/department/informaticas.html>

Mihalcea, R., and Faruque, E., 2004, 'SenseLearner: Minimally Supervised Word Sense Disambiguation for All Words in Open Text', *Proceedings of ACL/SIGLEX Senseval-3*, Barcelona, Spain.

Miller, G.A, 1990, 'Five papers on WordNet', *Special Issue of International Journal of Lexicography,* vol. 3, no.4., pp.235-244.

Miller, G.A., and Charles, W.G., 1998, 'Contextual Correlates of Semantic Similarity', *Language and Cognitive Processes*, vol.6, no.1., pp.1-28.

Miller, G.A., Leacock, C., Tengi, R., and Bunker, R., 1993, 'A semantic concordance', *Proceedings of the 3rd DARPA Worshop on Human Language Technology*, Plainsboro, New Jersey, pp. 303-308.

Mitamura, T., 1999, 'Controlled Language for Multilingual Machine Translation' *MT Summit VII*, Singapore, pp.46-52.

Mitamura, T., Nyberg, E.H., 1995, *Controlled English for Knowledge- Based MT: Experience with the KANT System*, Center for Machine Translation, Carnegie Mellon University, Pittsburgh.

Mitamura, T., Nyberg, E.H., and Carbonell, J.G, 1991, 'An Efficient Interlingua Translation System for Multi-lingual Document Production', *Proceedings of the Machine Translation Summit III,* Washington, DC, pp.55–61.

Mullery, G., 1996, 'The Perfect Requirements Myth', *Requirements Engineering Journal,* vol. 1, no.2, pp. 132–134.

Neff, M.S., Blaser, B., Lange J.M., Lehmann, H., and Dominguez I.Z., 1993, 'Get it where you can: Acquiring and maintaining bilingual lexicons for machine translation' *In Working Notes of the AAAI Spring Symposium on Building Lexicons for Machine Translation*, Stanford, CA. AAAI Press, pp. 98-98.

Ng, H.T., and Lee, H.B., 1996, 'Integrating multiple knowledge sources to disambiguate word sense: An examplar-based approach', *Proceedings of the 34th Annual Meeting of the*

Association for Computational Linguistics, University of California, Santa Cruz, California, pp. 40-47.

Nurmuliani, N., 2007, 'Investigating Requirements Volatility during Software Development Projects: An Empirical Study', PhD. Dissertation, University of Technology, Sydney, Faculty of Information Technology.

Nuseibeh, B. and Easterbrook, S., 2000, 'Requirements Engineering: a Roadmap', *Proeedings of ICSE: 22nd International Conference on the Future of Software Engineering*, Limerick Ireland, pp. 35-46.

Nyberg, E., Mitamura, T., and Huijsen, W., 2003, *Controlled Language for Authoring and Translation. In Computers and Translation: A Handbook for Translators*, Benjamins Translation Library, 35, Amsterdam.

Nyberg, E.H., and Mitamura, T., 1992, 'The KANT system: fast, accurate, high-quality translation in practical domains', *Proceedings of the fifteenth International Conference on Computational Linguistics, Actes du quinzième colloque international en linguistique informatique, COLING-92*, Nantes, France, pp.1069–1073.

Object Management Group 1997, *Object Constraint Language Specification*, technical report version 1.1, ad/97-08-08, IBM.

O'Brien, S. 2003, 'Controlling controlled English: an analysis of several controlled language rule sets', *European association for machine translation: international workshop*, Dublin, pp.105-114.

Ogden, C.K. 1942, *The General Basic English Dictionary*, W.W. Norton and Co., New York.

Ogden, C.K. 1932, *Basic English, A General Introduction with Rules and Grammar*, Paul Treber and Co., London.

Oriel, J., 1999, 'Guide for Specification Writing for U.S. Government Engineers', *Naval Air Warfare Center Training Systems Division (NAWCTSD)*, Orlando, viewed 17 October 2007 <www.ntsc.navy.mil/Resources/Library/Acqguide/spec.htm>

Oxford English Dictionary 2005, Oxford University Press, New York, viewed 2 December 2005, <http://www.oed.com>

Parnas, D.L., Asmis, G.J.K., and Madey, J., 1991, 'Assessment of Safety-Critical Software in Nuclear Power Plants', *Nuclear Safety,* vol. 32, no.2, pp. 189–198

Patwardhan, S., Banerjee, S., and Pederson, T., 2005, 'SenseRelate::TargetWord - A Generalized Framework for Word Sense Disambiguation', *Twentieth National Conference on Artificial Intelligence (Intelligent Systems Demonstration)*, Pittsburgh, USA.

Pedersen, T, Patwardhan, S and Michelizzi, J. 2004, 'WordNet::Similarity - Measuring the Relatedness of Concepts', *Nineteenth National Conference on Artificial Intelligence*, San Jose, CA, pp.38-41.

Pfahl, D. and Lebsanft, K., 2000, 'Using simulation to analyse the impact of software requirement volatility on project performance', *Information and Software Technology*, vol.42, no.14, pp.1001-1008.

Poesio, M., 1996, 'Semantic Ambiguity and Perceived Ambiguity',' *Ambiguity and Underspecification*, K. van Deemter and S. Peters (ed), CSLI, pp.159-202.

Pulman, S., 1996, 'Controlled Language for knowledge representation' *Proceedings of the first international workshop on controlled language applications*, Katholieke Universiteit Leuven, Belgium, pp. 233-242.

Pulman, S., 1994, 'Natural Language Processing and Requirements Specification', *Presentation at the Prolog Forum*, Department of Computer Science, University of Zurich

Quirk, R., and Greenbaum, S., 1996, *University Grammar of English*, Longman, London

RailCorp Contract C01645 2006, *Exhibit 3 – Train Performance Specification*, viewed 3 December 2008, <http://www.railcorp.info/about_railcorp/rollingstock_public_private_partnership/schedule_of_agreements)>

Randers, J., 1980, 'Elements of the System Dynamics Method', The MIT Press, Cambridge, MA, pp.117-139.

Ravin, Y., and Leacock, C., 2000, *Polysemy: An Overview. Polysemy: Theoretical and Computational Approaches*, eds. Ravin, Y, Leacock, C. Oxford University Press, Oxford.

Maier M. K., and Rechtin E., 2000, *The Art of Systems Architecting*, 2nd Edition, CRC Press.

Resnik, P., 1995, 'Disambiguating Noun Groupings with Respect to WordNet Senses', *Proceedings of the Third Workshop on Very Large Corpora*, Cambridge, Massachusetts, pp. 54-68.

Reuther, U., 1998, 'Controlling language in an industrial application', *Proceedings of the second international workshop on controlled language applications*, Pennsylvania: Carnegie Mellon University, pp.174-184.

Reuther, U., 2003. 'Two in one – can it work? Readability and translatability by means of controlled language', *European association for machine translation: international workshop*, Dublin: Dublin City University, pp.124-132.

Richardson, R., and Smeaton, A.F., 1994, 'Automatic word sense disambiguation in a KBIR application', Working paper CA-0595, School of Computer Applications, Dublin City University, Dublin, Ireland.

Robertson, S. 2005, *The Atlantic Systems Guild Inc*, viewed 13 July 2005, <http://www.systemsguild.com/GuildSite/SQR/SQR.html>

Rolland, C. and Proix, C., 1992, 'A Natural Language Approach for Requirements Engineering', *Proceedings of Conference on Advanced Information Systems Engineering, CAiSE,* Manchester, UK, pp. 257–277.

Rosch, E., Mervis, C., Gray, W., Johnson, D., and Boyes-Braem, P., 1976, 'Basic objects in natural categories', *Cognitive Psychology,* vol.8, pp.382-439.

Rupp, C. and Goetz, R., 2000, 'Linguistic Methods of Requirements-Engineering (NLP)', *Proceedings of the European Software Process Improvement Conference (EuroSPI)*, Denmark.

Saeed, M., 2005, 'To define an unambiguous, disciplined and repeatable process for the coupling of BOYDA with an existing system modelling language', University of Technology, Sydney, Unpublished Capstone Project Report.

Sage, A.P., 1992, *Systems Engineering*, Wiley IEEE.

Saunders S., 2003, 'Reducing Project Re-Work By The Use of Model Based Systems Engineering Processes and Tools', *SETE2003 Conference,* Sydney, Australia.

Schlager, J., 1956, 'Systems engineering: key to modern development', *IRE Transactions*, EM-3, pp. 64–66.

Schneider, G.M., Martin, J., and Tsai, W.T., 1992, 'An Experimental Study of Fault Detection in User Requirements Documents', *ACM Transactions on Software Engineering and Methodology,* vol.1, no.2, pp. 188–204.

Schwitter, R., 2004, 'Representing Knowledge in Controlled Natural Language: A Case Study', *Knowledge-Based Intelligent Information and Engineering Systems, 8th International Conference, KES2004*, Springer LNAI 3213, Wellington, New Zealand, pp.711-717.

Schwitter, R., 2002, 'English as a formal specification language', *Proceedings of the 13th International Workshop on Database and Expert Systems Applications*, Aix-en-Provence, pp. 228-232.

Selvarajan, R., 2005, 'To develop a software tool that implements the BOYDA language and BOYDA modelling rules', University of Technology, Sydney, Unpublished Capstone Project Report.

Shubert, S., Spyridakis, J., Holmback, H., and Coney, M., 1995, 'The Comprehensibility of Simplified English in Procedures', *Journal of Technical Writing and Communication,* vol.25, no.4, pp.347–369.

Simpson, G.B., and Burgess, C., 1989, 'Implications of lexical ambiguity resolution for word recognition and comprehension', *In Small, S., Cottrell, G.W., and Tanenhaus, M.K., (Eds.), 1988, Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence*, Morgan Kaufman, San Mateo, California, pp. 271-288.

Slator, B.M., and Wilks, Y.A., 1987, 'Towards semantic structures from dictionary entries', *Proceedings of the 2nd Annual Rocky Mountain Conference on Artificial Intelligence*, Boulder, Colorado, pp. 85-96.

Software Engineering Institute 2008, *The Capability Maturity Model® Integration*, Carnegie Mellon University, Viewed 16 December 2008, < http://www.sei.cmu.edu/cmmi/>

Somers, H. (ed), 2003, *Computers and Translation: A Translator's Guide*, John Benjamins Publishing Company, Amsterdam.

Sommerville, I. 1992, *Software Engineering*, Fourth Edition, Addison-Wesley Publishing Company, Wokingham, UK.

Sommerville, I., Sawyer, P., 1997, *Requirements engineering, a good practice guide*, 1st Edition, Wiley.

Spaggiari, L., Beaujard, F., Cannesson, E., 2003, 'A Controlled Language at Airbus', *Proceedings of the Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Application Workshop*, Dublin City University, Dublin, pp. 151-159.

Stark, G., Skillicorn, A., and Ameele, R., 1999, 'An Examination of the effects of requirements changes on software maintenance releases', *Journal of Software Maintenance: Research and Practice*, vol.11, pp.293-309.

Stewart, K.M., 1998, 'Effect of AECMA Simplified English on the comprehension of aircraft maintenance procedures by non-native English speakers', PhD, University of British Columbia.

Stokes, D.A, 1991, *Requirements Analysis*, Computer Weekly Software Engineer's Reference Book, pp. 16/3-16/21.

Sussna, M., 1993, 'Word sense disambiguation for free-text indexing using a massive semantic network', *Proceedings of the Second International Conference on Information and Knowledge Base Management, CIKM'93*, Arlington, Virginia, pp. 67-74.

Swaffar, J., Arens, K., and Byrnes, H., 1991, *Reading for meaning: An integrated approach to language learning*,: Prentice-Hall, Englewood Cliffs, NJ.

The Standish Group International, Inc 1994, *The Standish Group International CHAOS Report*, Massachusetts, viewed 23 December 2007, <http://www.standishgroup.com/sample_research/chaos_1994_1.php>

Trochim, W.M.K., 2006, *Research Methods Knowledge Base*, viewed 11 August 2008, <http://www.socialresearchmethods.net/kb/index.php>

Veale, T., 2004, 'Pathways to Creativity in Lexical Ontologies', *Proceedings of The 2nd Global WordNet Conference*, Brno, Czech Republic, pp.220-225.

Vitech Corporation, 2008, *CORE Enterprise*, viewed 7th December 2008, <www.vitechcorp.com/>

Voorhees, E.M., Claudia L., and Geoffrey T., 1995, 'Learning context to disambiguate word senses', *In Petsche, T., Hanson, S.J., and Shavlik, J., eds., Computational Learning Theory and Natural Learning Systems*, MIT Press, Cambridge, Massachusetts.

Walton, D., 1996, 'Fallacies Arising from Ambiguity', *Applied Logic Series*, Kluwer Academic, Dordrecht, NL.

Web WordNet 2.0, 2003, Cognitive Science Laboratory Princeton University, Princeton, viewed 10 May 2008, <http://wordnet.princeton.edu/cgi-bin/webwn>

Wiegers, K., 2006, *More About Software Requirements*, Microsoft Press, Washington, USA.

Wikimedia Foundation Inc, 2008, *Wikipedia*, viewed 10 November 08, <www.wikipedia.org>.

Wilson, W.M., Rosenberg, L.H., and Hyatt, L.E., 1997, 'Automated Analysis of Requirement Specifications', *Proceedings of the 19th international conference on Software engineering,* Massachusetts, USA, pp.161-171.

Wu Z., and Palmer M., 1994, 'Verb Semantics and Lexical Selection', *32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico, pp.133-138.

Yarowsky, D., 1992, 'Word sense disambiguation using statistical models of Roget's categories trained on large corpora', *Proceedings of the 14th International Conference on Computational Linguistics, COLING'92*, Nantes, France, pp. 454-460.

Zave, P., 1997, 'Classification of Research Efforts in Requirements Engineering', *ACM Computing Surveys*, vol.29, no.4, pp.315-321.

Zipf, G.K., 1949, *Human Behavior and the Principle of Least Effort*, MA: Addison-Wesley, Cambridge

Zowghi, D., and Gervasi, V. 2002, 'The Three Cs of Requirements: Consistency, Completeness and Correctness', *Proceedings of 8th International Workshop on Requirements Engineering: Foundation for Software Quality, (REFSQ'02)*, Essen, Germany, pp.155-164.

Zowghi, D., and Gervasi, V., 2004, Erratum to 'On the interplay between consistency, completeness, and correctness in requirements evolution', *Journal of Information and Software Technology,* vol. 46, no.11, Elsevier, pp.763-779.

Zowghi, D., and Nurmuliani, N., 1998, 'Investigating requirements volatility during software development: research in progress', *Proceedings of 3rd Australian Conference on Requirements Engineering*, Geelong, Australia, pp.26-27.