HARNESSING THE POWER OF CONTEXT: A CORPUS-BASED ANALYSIS OF VARIATION IN THE LANGUAGE OF THE REGULATED NUCLEAR INDUSTRY

by

JACQUELINE MARIE HETTEL

(Under the Direction of William A. Kretzschmar, Jr.)

ABSTRACT

This dissertation addresses the importance of extralinguistic contexts on domain-specific language. Using publicly available documents from the regulated nuclear power industry, a representative corpus of over 9 million words was created using the methodology developed for the construction of the Tobacco Documents Corpus at the University of Georgia from tobacco industry documents. The reproducibility of this methodology was confirmed through analysis of the rejection ratios over four sampling iterations by means of the two proportion z-test. Analysis of 20 key terms from this industry and the words with which they co-occur within a span of four words to the right and left of the term, lexical profiles were generated to describe this variety of domain-specific language. The validity of these collocations were measured using their Mutual Information scores. The results indicate that that while there is some shared meaning with regard to key terms in this variety of domain specific language, documents authored by different industry groups, as well as individuals located in different NRC regions, will exhibit differences with regard to their use and connotation of these same key terms. This observation demonstrates that context does matter with regard to domain-specific language in much the same way that has been observed in other language varieties (i.e. speech).

INDEX WORDS: Corpus Linguistics, Domain-Specific Language, Tobacco-Documents Corpus, Government Language, English for Specific Purposes, Mutual Information, Corpus Sampling Methods

HARNESSING THE POWER OF CONTEXT: A CORPUS-BASED ANALYSIS OF VARIATION IN THE LANGUAGE OF THE REGULATED NUCLEAR INDUSTRY

by

JACQUELINE MARIE HETTEL

B.A., Grand Valley State University, 2008

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in

Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

HARNESSING THE POWER OF CONTEXT: A CORPUS-BASED ANALYSIS OF VARIATION IN THE LANGUAGE OF THE REGULATED NUCLEAR INDUSTRY

by

JACQUELINE MARIE HETTEL

Major Professor: William A. Kretzschmar, Jr.

Committee: Christy Desmet
Lewis C. Howe

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

DEDICATION

For my father, W. Grover Hettel.

My contribution to the "family business."

Your love and support throughout this journey means more than you will ever know.

And in memory of Dr. Lisa Lena Opas-Hanninen.

An amazing friend and mentor who believed in me and recognized my potential.

(Even when I did not.)

You will be missed.

ACKNOWLEDGEMENTS

I have been fortunate over the past two years that I have been working on this project to have many different people support my endeavors. It is those individuals that I would now like to acknowledge.

I would like to thank Dr. William Kretzschmar for his support during my time at the University of Georgia. In my first semester at UGA, his Style course introduced me to the world of quantitative analyses of language, and by extension Digital Humanities. Additionally, I would like to thank Dr. Christy Desmet for her encouragement and showing me that corpus linguistic approaches have applications for composition pedagogy. And Dr. Chad Howe, I appreciate him not only being on my committee, but also for providing an alternative perspective on how quantitative approaches can be used to learn about language. I have been extremely fortunate to have all three of these individuals guide me as I have gone from student to academic professional. I will never forget their understanding and support.

There is a group of individuals within the nuclear power industry without whose help this dissertation would never have been possible. First of all, I would like to thank Dwight Mims for not only being a sounding board, but for taking an active interest in the pursuit of my doctorate. It was through him that I was able to meet Ellis Merschoff, who I would also thank for taking time out of his busy schedule to help me better understand the history and infrastructure of the Nuclear Regulatory Commission. And to Jack

Nataro, whose reassurances that my research was of value to the nuclear industry motivated me during a time when I needed it most.

Finally, I would like to thank my family for their love and support through all of the hard work, sweat, and tears. Words can never express how much I appreciate what they have given. Specifically, I would like to acknowledge my parents, Linda and W. Grover Hettel, without whose love and support the pursuit of my doctoral degree would have been much more difficult. While their example has taught me to value the applied as much as the theoretical, more importantly they have shown me that through hard work and perseverance you can accomplish anything.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This dissertation is a description of real language used by the regulated nuclear power industry and how geographic location and industry group affiliation affects the terminology of it. Using corpus linguistic approaches to modeling, key word and collocational patterns were observed through a corpus of approximately 9 million words from publicly available documents internal to the nuclear power industry. This chapter presents the background of this project, including an overview of the regulated nuclear industry in the United States. Then the goals of this study and the significance of this line of inquiry are discussed. Finally, an overview of the methodology, the delimitations of this study, and a summary the chapters to follow are provided.

BACKGROUND OF THE US NUCLEAR POWER INDUSTRY

The regulation of the nuclear industry began as a reaction to the use of atomic bombs on the Japanese cites of Hiroshima and Nagasaki in August of 1945. The United States Congress established the Atomic Energy Commission by passing the Atomic Energy Act of 1946 in order to "maintain strict control over atomic technology and to investigate its military applications," not necessarily to develop it for civilian purposes (Walker and Wellock 1). Following World War II, the primary focus of those individuals involved in nuclear development was directed toward military development. In the early part of 1953, the U.S. Navy began testing nuclear reactors to power their submarine fleet. After the

Atomic Energy Commission observed the success of these reactors in autumn of the same year, it announced the intention to build a power plant. As a result, the first commercial nuclear reactor in the U.S. became operable in Shippingport, Pennsylvania, in 1957 (Bodansky 31). Many more reactors would be built rather quickly in the years that followed.

The Atomic Energy Commission continued to regulate both the commercial use of atomic materials and the development of new technologies using those materials until Congress passed the Energy Reorganization Act of 1974, which divided the AEC into two agencies: the U.S. Energy Research and Development Administration and the U.S. Nuclear Regulatory Commission:

> The U.S. Nuclear Regulatory Commission (NRC) was created as an independent agency by Congress in 1974 to enable the nation to safely use radioactive materials for beneficial civilian purposes while ensuring that people and the environment are protected. The NRC regulates commercial nuclear power plants and other uses of nuclear materials, such as in nuclear medicine, through licensing, inspection and enforcement of its requirements. ("About NRC")

Thus, the NRC came into being in January 1975 to facilitate, and speed up, the licensing of nuclear plants, as well as to develop better regulatory practices for this industry (Walker and Wellock 49). In fact, the issue of reactor safety is thought to be the central one for the NRC in its early years. One event in particular brought the safety of nuclear power plants, as well as the NRC, to the attention of the public, and that was an event known in the industry as the Brown's Ferry Fire:

> The first event was a major fire at the Tennessee Valley Authority's Browns Ferry Nuclear Plant near Decatur, AL, in March 1975. In the process of looking for air leaks in an area containing trays of electrical cables that operated the plant's control room and safety systems, a technician set off a fire. He used a lighted candle to conduct the search, and the open flame ignited the insulation around

> the cables. The fire raged for over 7 hours and nearly disabled the safety equipment of one of the two affected units. (Walker and Wellock 52)

Only four years after this incident, another accident occurred at an American nuclear power generating station:

> On March 28, 1979, an accident at the Three Mile Island Nuclear Station (TMI), Unit 2, near Harrisburg, PA, made the issue [the risks of nuclear power] starkly and alarmingly real. As a result of a series of mechanical failures and human errors, the accident (researchers later determined) uncovered the reactor's core and melted about half of it.…By the time that experts realized that the plant had undergone a loss-of-cool- ant accident and flooded the core, the reactor had suffered irreparable damage. (Walker and Wellock 53-54)

The rapid succession of the Brown's Ferry Fire and Three Mile Island affected the credibility of the nuclear power industry and the NRC, to put it lightly. However, in the years to come, this agency would develop safety requirements and regulatory practices that would help to reduce the risk and likelihood for future accidents.

Although these two events are not the only ones that we have experienced in the U.S. or in the world, the NRC has worked to protect the public and regulate the operating practices of the nuclear power industry. Along the way, the NRC's methods for performing its duties have evolved. One such evolution occurred in the mid-1990s when Dr. Shirley Ann Jackson was appointed Chairman of the NRC. She implemented what Dr. Ellis Merschoff, former Chief Information Officer of the NRC, describes as a radical restructuring of how the NRC now performs its regulatory functions. She developed and introduced risk-informed, performance-based regulation to the NRC that resulted in many changes to the documents being exchanged between licensees and the NRC (Merschoff 2011).

In the light of recent events involving the nuclear Fukushima I Nuclear Power Plant in Japan, the NRC has again been brought into the public spotlight. Criticisms about the actions of the NRC, and even the language of the industry, have been made by individuals for political reasons. One book in particular has been published titled *Nukespeak: The Selling of Nuclear Technology from the Manhattan Project to Fukushima*, which claims that there is a "linguistic filter of the nuclear mindset" that the nuclear power industry and its proponents use for rhetorical means (Hilgartner, et al.. "Introduction"). However, in this extremely politicized, and biased, text the authors do not provide any description for the language of the nuclear power industry in making their claims. In order for any assumption to be made regarding whether or not there is any attempt "to hide the truth about nuclear dangers" (Bell ""No Word for Meltdown: The Return of Nukespeak"), we must first understand the actual language used by the nuclear power industry .

RESEARCH HYPOTHESIS AND OBJECTIVES

In this dissertation using publicly available documents from the regulated nuclear power industry, I intend to test the hypothesis that that geography (NRC region) and industry group contexts (NRC vs. licensee affiliation) are significant as dimensions through the variation of the collocation of nuclear regulatory terms. In order to test this hypothesis, there are three basic goals to this study. The first goal is to establish foundation descriptions of the language being used in this corpus to better understand nuclear regulatory language in use: nuclear regulatory terms. Next, I seek to identify differences in the language used in this industry with regard to author affiliation (i.e. licensees to the

NRC and vice versa, etc.). The final goal is to also identify regional patterns of language use in these documents in order to see if any specific patterns or trends emerge with regard to NRC representatives and the licensees.

WHY ANALYZE THE LANGUAGE OF THE NUCLEAR INDUSTRY?

By answering these questions, this study will extend knowledge of the importance of geographical and industry group contexts on lexical frequencies and collocations. This study also provides an analysis of regulated corporate discourse, more specifically the language of the nuclear power industry as it is used and made available to the public—a variety of language that has yet to be investigated empirically. By comparing this model to a corpus of general American English, it becomes easy to achieve a description of how words co-occurring with key terms in this industry reflect the social dynamic between NRC officials and licensees. Additionally, the geographic/NRC regions where terms are more likely to occur in documents from the NRC's public database can be identified. Academically, these results can be used in the fields of technical writing and lexicography. The regulated nuclear power industry can also reap benefits from the knowledge resulting from this investigation for application in areas where knowledge of the propensities for language use with regard to the NRC/licensee relationship and regional jurisdictions can be of use: i.e. management, communications, or regulatory affairs.

SUMMARY OF THE METHODOLOGY

In order to obtain significant insights I applied the methodology developed to create the Tobacco Documents Corpus (TDC) at the University of Georgia. The reason for adopting this methodology is that the database from which the TDC was sampled is quite similar to the NRC's public document database. The sampling process used for this dissertation was a quota-based framework that randomly selected a set number of documents for each of the NRC's 104 nuclear power licenses. The result was a corpus of over 9 million words complete with metadata provided by the NRC, which was confirmed against each PDF file.

Once the corpus was complete, I used WordSmith Tools to generate frequency lists for this corpus, as well as the Baker Corpus of General American English. I then used this software to perform an analysis to identify the 20 most statistically significant terms that were more likely to occur in the nuclear industry reference corpus than the Baker Corpus. Collocational analyses were then performed that looked for the most frequent words to co-occur with each of the 20 terms within a span of four words to the right and left them. Using these lists, I created lexical profiles to uncover the most salient meanings associated with each term. The terms were then analyzed for frequency of use with respect to industry group and geographic contexts. I then identified any variation in the collocations and lexical profiles was also identified to see how the meaning associated with these industry terms changed. A more detailed description of the methodology for this study is provided in Chapter 4.

DELIMITATIONS OF THE STUDY

Due to the extremely large amount of language-in-use evidence available from the NRC to the public through their online database, there were certain boundaries I had to define and follow in order to make the scope of this dissertation manageable. First of all, this study samples documents from the years 2000 to 2011. The reason for only going back to the year 2000 is because that is the year when all publically available documents from the NRC were provided through the online database. Furthermore, this is a point in time after the NRC was restructured during the tenure of Chairperson Shirley Ann Johnson, which resulted in changes to the communications of this regulated industry. Secondly, all of the documents that are part of this study are available under the governance of the Freedom of Information Act. As such, only documents that have been certified by the NRC as not containing information that would cause detriment to national security or impinge on proprietary rights were included in the corpus. Moreover, the model of regulated nuclear power language that was created for this dissertation is representative of written text, and as a result it is not a model for the spoken variety of this type of language.

Finally, I would like to emphasize that this study is a description of the regulated nuclear power industry's language as it is occurs. This dissertation is not a commentary on the industry's communications with the public, nor does it address why the NRC provides access to the documents they do through their online database. Furthermore, it is not within the scope of this study to provide critique on the improvement of this variety of language. What is provided in this study is a description of what regulated nuclear power industry language is, how it is used, and how it varies by industry group and geographical contexts without judgment of the integrity of the organizations who author the texts.

ORGANIZATION OF THE DISSERTATION

This dissertation is organized into seven chapters. Chapter 2 provides a brief history of corpus linguistics, a discussion of how corpus linguistics is an act of modeling, as well as previous research on creating models of language in use.

The third chapter is a detailed description of the methodology used in this study. I have outlined the procedure used to create the nuclear reference corpus from the NRC ADAMS database in comparison to that used for the TDC. There is great detail provided regarding the sampling framework, as well as an analysis of the reproducibility of this methodology. I also explain the methods and statistical tests employed to identify the industry terms that are analyzed in Chapters 4 through 6.

Chapter 4 is an analysis of the reference corpus to identify 20 regulated nuclear power industry terms. Then, these terms are also analyzed and profiled using collocational patterns to determine the meanings associated with them.

Chapter 5 and Chapter 6 focus on analyzing the frequency of use for the industry terms identified in Chapter 4 with regard to the organizational affiliation of the author being of the NRC or a licensee and geographic context, respectively. The collocational patterns of each term were again analyzed and profiled in order to identify any variation between these two industry groups.

Chapter 7 provides the conclusion for this dissertation. Not only is it a summary discussion of the study and the dissertation, but it also proposes the significance of this line of inquiry and the implications for future research.

CHAPTER 2

CORPUS LINGUISTICS: AN EXERCISE IN MODELING

In the field of language study, corpus linguistics is one research methodology that, while an exercise in modeling, allows the use of "real life language" sampled from the world in which it is used. McEnery and Wilson define corpus linguistics as "the study of language based on examples of real life language use" (McEnery and Wilson 1). However, other scholars in this field like John Sinclair, argue that corpus linguistics is more than just the definition provided by McEnery and Wilson, it is a systematic collection of naturally occurring texts, or "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (2004). Why is it important then to create a model that is composed of samples of real language, texts if you will?

Michael Stubbs in *Words and Phrases: Corpus Studies of Lexical Semantics* reminds us that language, social action, and knowledge all coexist together. In fact, the way in which words are used "can reveal relations between language and culture: not only relations between language and the world, but also between language and speakers with their beliefs, expectations and evaluations" (6). Whether or not we are conscious of it, we have these expectations for the language we use everyday. For example, when we read a newspaper article, we expect the words and phrases it uses to be quite different from those used in a technical manual for putting together a child's toy. That is, we have knowledge of diverse text-types, or documents, and a different set of linguistic expectations for each one.

Kretzschmar also advises us in *The Linguistics of Speech* that the language behavior of those texts that we might be selecting in order to model a language varies by text types:

> Text types are recognizable situations of use for speech, whether written or spoken. Some are large in scale, like 'writing' considered as a text type, as opposed to 'spoken language' as a text type. Others are small such as 'letter,' and even the smaller types can be further delimited as in the sequence 'letter,' 'business letter,' 'job application letter,' and even 'job application letter for which the applicant does not have all of the qualifications.' Each text type can be recognized because it has characteristics that allow it to be distinguished for itself. (Kretzschmar 2009, 159)

Thus, there is a real reason for why our expectations vary from text-type to text-type: we recognize that the language used in different text-types, and thus the documents that are examples of them, are different. However, the language used in these documents is not the only thing that varies: "language at the level of words in texts must be considered as one dimension in the model, and language at the level of word distributions in geographical and social space must be another dimension….And the levels interact with one another" (Kretzschmar 2009, 158). This observation indicates a need for further investigation into the ways in which the interaction of extralinguistic contexts with linguistic elements correlate with differences in the way meaning is created within different text-types.

The perspective that our expectations for language are dependent on our non-linguistic knowledge from the everyday world is also shared by Stubbs: "meanings are not always explicit, but implicit. Speakers can mean more than they say" (20). Thus, language is a "mode of action," or a behavior, as also proposed by twentieth-century researchers, like Malinowski (1923), Firth (1957), and Wittgenstein (1953). The use of corpus linguistics as a model for learning more about language is rooted in the desire to "develop a theory of meaning (Teubert 1999a, 1999b). When people hear or read a text, they are usually interested in its meaning, not in its wording or grammar, and they generally remember its content, not how the content was phrased" (Stubbs 20). These relationships between the words we use and what we mean when we use them are not always clearly visible.

If we look for recurring patterns of words as they are used in different contexts in large collections of textual data, then we can have evidence and quantifiable support for our intuitions as to how meaning is constructed through language. In other words, corpus linguistics possesses the quality of modeling that McCarty describes as computational tractability: the ability to be completely explicit and absolutely consistent so as to be manipulated computationally. Stubbs explains that the way in which we are able to accomplish this task with corpus linguistics is by evaluating the most basic units of meaning in language: words and phrases and how they occur together.

As a result of computational tractability in modeling, we are forced to question how we know what we know, and there may come times when our ideas are problematized by the model. Manipulating our model may enable us to have successes that we cannot explain. We may find something where it should not be. On the other hand, we may have moments of inexplicable failure where we are not able to find something in our model where it is otherwise present. Nonetheless, the act of manipulating a model allows us these temporary states within the "process of coming to know rather than fixed structures of knowledge" (McCarty 4). It is through this process that we are able to make discoveries about abstract concepts, like the creation of meaning through language.

When it comes to meaning, we often associate this notion of being fixed with *denotation,* the "cognitive, conceptual, logical, ideational and propositional meaning….the 'literal meaning' of a word. However, there is another type of meaning words possess called *connotation*, "which is also called affective, associative, attitudinal and emotive meaning" (Stubbs 34). These two types are often contrasted such that denotation is usually assumed to be the meaning associated with a word that is stylistically neutral and not dependent on the relationship a speaker or hearer has with the word—the latter association is relegated to connotation. The difference between these two types of meaning is not always distinct, especially when it comes to which one is primary or secondary in the context of how the word in question is being used.

The meaning we each associate with words and phrases in our language does not refer directly to the world around us. Instead, it indirectly points to our notions of what those words and phrases mean, based on our past experiences. For example, when you read the word *mouse*, its meaning does not come from the combination of the letters *m-o-u-s-e*, but rather your cognitive representation of *mouse* emerges from your past experiences, your reality, where this word was used. The meaning you associate with this word might be most strongly connected with something small and furry that you may only want living in your house if it is in a cage, or it may be a peripheral device for your computer. It is both, and you probably have still more meanings, like a name for a 'black eye" or for a timid person. However, we also associate connotative meanings with a word like *mouse*. We may have negative feelings for our cognitive representation for this word through our past associations of it with disease or filth. On the other hand, we may have a connotation for *mouse* that is positive with it in thinking about a certain cartoon character from our childhood. All of these meanings are individual, based on our expectations and past experience in the use of these words, and it is through the co-occurrence of a word like *mouse* with *dirty* versus *Mickey* that helps us to understand its meaning.

A *phrase,* as defined by Stubbs, is a "unit of meaning in connected language in use [that] is usually not a single word in isolation, but a longer unit of at least a few words in length" (24). When words co-occur with each other within a limited span of a text, like a phrase, we call that *collocation.* Collocations demonstrate that words and context are inseparable, as the meaning of one word is dependent on the environment provided by the words surrounding it (Firth 194): or in his own words, "You shall know a word by the company it keeps!" (11). Collocations are not in and of themselves context, but instead they are "word[s] in habitual company" whose meaning is implied through the context the words provide one another (14). Collocations contribute to lexical cohesion as they help to create and make apparent semantic relationships in a text (Halliday 2004).

When collocations are part of a semantic relationship within a text, they become part of what is often referred to as a *multiword unit.* Bolinger (1976) calls them "the prefabs of language," and later Sinclair (1987, 1991) includes them in his *idiom principle.* This notion is centered on the fact that words do not occur randomly in texts. Instead, a language user will use his large resource of "semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (Sinclair 110). All of these terms denoting specific types of lexical items that are composed of multiple words co-occurring together are considered by many linguists to be the most important linguistic elements for analysis as they are the building blocks for constructing meaning.

Stubbs suggests we should focus on extended lexical units, which are composed of words that co-occur together, because "individual words often do not correspond to units of meaning. Individual forms of a lemma [word] may have quite different uses, and often the unit of meaning is a longer phrase or collocation" (49). The way in which we are able to make observations about these units of meaning is through the use of corpora. We can again go back to Sinclair's definition and find that a corpus is "a collection of pieces of language text [most often] in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair 2004). Baker states that there are several advantages to using computer-based corpora for linguistic investigation:

1. They reduce researcher bias (10-12);
2. A researcher is able to notice the incremental effect of discourse: "how language is employed, often in quite subtle ways, to reveal underlying discourses," lexical priming, and even repeated patterns of evaluations (13);
3. They allow researchers to observe resistant and changing discourses because "discourses are not static" (14);
4. They utilize triangulation, or multiple methods of analysis by going beyond binary oppositions (14).

He concedes that we should "bear in mind that because corpus data does not interpret itself, it is up to the researcher to make sense of the patterns of language which are found within a corpus, postulating reasons for their existence or looking for further evidence to support hypotheses. Our findings from corpus linguistics *are* interpretations…" (Baker 18). The reason for this is that corpora cannot explain why, they can only demonstrate *what* is happening in a language. It is from this perspective that we realize that corpus linguistics as a form of modeling is an advantage because modeling through computing enables us to specify computationally "how we know what we know," and we need to marry an explicit model with intuition because we can know more than we can tell (McCarty 2).

There are those individuals who criticize the corpus-based approach as a means of linguistic inquiry:

> One criticism of corpus-based approaches is that they are too broad--they do not facilitate close readings of texts. However, this is akin to complaining that a telescope only lets us look at far away phenomena, rather than allowing us to look at things close-up like a microscope (Partington 1998:144, quoted in Baker 7).

Most often, this quote is used to demonstrate that while there are those of us who desire to analyze language from this macroscopic perspective in order to quantify what we inherently know about language, there are things corpus linguistics cannot do. First of all, corpora cannot provide negative evidence about a language (i.e. what is correct or possible, or what is incorrect or not possible). Secondly, a corpus cannot tell you why certain patterns occur in language—as we have already discussed, this is where intuition comes into the picture. A corpus can only tell you what happens within it, and with statistics it can help you understand the propensity for those things to happen. Finally, a corpus cannot provide all of the possibilities in language at one time because it must be principled, it must be planned, and it must be systematically constructed (Bennett 3).

This importance on constructing a corpus that is representative of the type of language desired for analysis is based on the need for manipulation of a model that yields accurate interpretations of what we are modeling and not something else entirely.

In *Developing Linguistic Corpora: a Guide to Good Practice*, Sinclair outlines in the first chapter, “Corpus and Text: Basic Principles,” some best practices for sampling a language for a corpus. He begins by explaining that there are three primary considerations that must be addressed in establishing a sampling policy:

1. The orientation to the language or variety to be sampled.
2. The criteria on which we will choose samples.
3. The nature and dimensions of the samples. (Sinclair 2004).

The first question, the notion of the orientation of a corpus, is one he thinks many corpus builders take for granted. In order to make claims about the authenticity of results coming from the analysis of a particular corpus, language researchers must make sure that “only those components of corpora which have been designed to be independently contrastive should be contrasted” (Sinclair 2004). In other words, the type of corpus you choose to create dictates how you sample texts for inclusion in the corpus.

The type of corpus to be created is dependent on the kind of analysis, or manipulation, we want to perform (Anderson and Corbett 7). More specifically, it is dependent on what you want to study (content) and the context of that content (orientation). Table 2.1 lists a few types of corpora as defined by their content and orientation. While this list is by no means comprehensive, we can see that there are many

Table 2.1: Corpora Types by Content and Orientation

Content	Description	Orientation	Description
General	Represents a language or language variety as a whole. Will contain both spoken and written language, or different text-types.	Balanced/ Representative	Texts are selected using pre-defined sampling proportions to reflect a particular language or variety.
Historical	Represents an earlier stage(s) of a language.	Monitor	New texts are added to the corpus continuously to "monitor" language change.
Regional	Represents one regional variety of a language.	Parallel	The same texts are selected in two or more languages or language varieties.
Learner	Represents the language produced by learners of a particular language.	Comparative	Similar texts are selected in two or more languages or language varieties.
Reference	Represents a very specific type of language. For example, domain-specific corpora that are designed to model language for a specific function.	Diachronic	Texts from consecutive time periods are included for comparison purposes.

different kinds of corpora. With regard to content, there are general, historical, regional, learner, and reference corpora. A general corpus seeks to represent a language variety as a whole and includes multiple text-types or modes (American English or British English). A historical corpus represents language during an earlier time period or stage. Regional and learner corpora are exactly what they sound like: a regional corpus represents one regional variety of a language, and a learner corpus represents the language of learners of a particular language. And lastly, a reference corpus represents a very specific type of language, like business language—otherwise known as domain-specific language—which is used for a particular function.

Once the content has been determined for this linguistic model, a corpus builder also needs to decide the context of interest for the inquiry. Depending on the contrastive elements of her inquiry, a corpus builder should orient the sampling of her corpus in different ways. In other words, is she trying to make observations to define a particular language or variety? If so, then she is creating a balanced/representative corpus and should use predefined sampling proportions. Is her real interest in monitoring how a particular language has changed over time? Then her corpus is a monitor one and should continuously add new texts over time, starting with the year corresponding with the beginning of her time period of interest. A corpus builder should make sure to define both content and orientation (Sinclair 2004). For example, she could build a balanced/representative regional corpus in order to make observations about American English in the Southern part of the United States by defining sampling policies that include text-types of both spoken and written English, or she could build a comparative corpus of American English in the South versus American English in the Midwest. Each of these different types of corpora create a situation where researchers can make observations about a language by contrasting particular elements: i.e. time, extralinguistic parameters, text-type, etc.

The second of Sinclair's three questions for developing a sampling policy for a corpus deals with those criteria that help to determine the types of texts to be included in it. These criteria all focus on the nature of a text: its mode, type, domain, language, location, and even date. He suggests that these criteria be "small in number, clearly separate from each other, and efficient as a group for delineating a corpus that is representative of the language or variety under examination" (Sinclair 2004). The decisions made initially regarding these criteria are of great importance to the creation of a corpus, as they directly influence the representativeness of it.

Representativeness is an important aspect of corpus creation. There is no way that one can make accurate generalizations about a language or language variety if the corpus is not sampled validly from the language or language variety in question:

> No one would claim to be able to make valid conclusions about Singaporean English from analysing a corpus of Scottish English, so we cannot draw valid findings about the use of the word *like* in spoken language from a corpus which contains only written language, and vice versa. Similarly, we cannot assume that if a grammatical construction is common in a corpus made up of scientific articles then it is also commonly used in the language as a whole. (Anderson and Corbett 5).

It is suggested by Anderson and Corbett in *Exploring English with Online Corpora* that a corpus that is truly representative will be constructed so that it will reduce the possibility of making invalid claims about the language it is supposed to be modeling. It should be conceded, however, that true representativeness, perfection, is impossible to attain. Nonetheless, we should try to make corpora as representative as is in our power to do.

This issue of representativeness is defined within the sampling parameters defined before the corpus is created. In order to create a corpus that is as representative as possible of a larger population, the corpus builder must define the *sampling unit* and the *sampling frame.* A sampling unit is the individual textual samples to be collected: i.e. a book or a journal article. The population from which the corpus linguist is sampling, then, is all actual sampling units available, while the *sampling frame* is the list of sampling units to be included in the corpus. For example, if a person were interested in making observations about the language used in American newspapers during the 20th century, she would create a sampling frame that was composed of sampling units representing this particular time and place. The amount of documents she would sample using that frame, however, is also something that has to be accounted for in a corpus' sampling design.

The third and final consideration for establishing a sampling policy for a corpus, as outlined by Sinclair, relates to the nature and dimensions of the samples. In other words, how many texts are needed in the corpus in order to create valid estimates: how big does your model need to be? This question is connected directly to the issue of what kind of manipulations you want to do with your model of language in use. There are those corpora that are extremely large in size. For example, the British National Corpus (BNC) is 100 million words in length, the Corpus of Contemporary American English (COCA) is nearing 500 million words, and the Google Books Corpus is in the billions. However, they are large because the type of corpus defined in their sampling policy requires large quantities of sampling units: the BNC is meant to represent the spoken and written language of all of Great Britain in the late 20th century, COCA is a monitor corpus that represents varieties of spoken and written American English since 1990, and Google Books is a monitor corpus of language as it exists in written books up to 2012. Corpora of these sizes are useful for those individuals interested in lexicographical studies, where corpora in the range of tens of millions if not hundreds of millions of words are necessary (Anderson and Corbett 6). On the other hand, if the purpose of your corpus is to represent written language data, then you will want to make sure your corpus is over at least five million words, as this is considered small (Anderson and Corbett 6). Most importantly, the size of the corpus is dependent on the variety and text-types in question. The focus of size in corpus construction is not necessarily to create the largest corpus possible, but rather the importance is to create a model with the understanding of how it will be used.

There are two general types of population sampling that also influence the size of a corpus. Meyer (2004): probability and non-probability sampling. Probability sampling is defined by a researcher carefully pre-selecting the population she wants to study with her corpus by using statistical formulas and demographic information to "ensure that the number and type of people being surveyed are truly representative" (43). This is in contrast to non-probability sampling, in which a pre-selection process is not employed. There are three types of non-probability sampling:

1. "Haphazard, convenience, or accidental sampling" (Kalton 90), where a researcher only samples those individuals who are available to her.
2. "Judgment, or purposive, or expert choice" sampling (Kalton 91), where a researcher decides ahead of time who is best qualified to be sampled (for example, only taking samples from native speakers of a language rather than non-native ones).
3. "Quota sampling" (Kalton 91), where a researcher samples certain percentages of certain populations; for example, she could create a corpus whose samples reflect actual percentages of students from the University of Georgia with regard to gender (60% females, 40% males).

While Meyer suggests that probability sampling is the most reliable type of sampling, because it leads to the least amount of bias, it is often logistically impossible for linguists to do this type of sampling because it often yields extremely large sample sizes, and by extension sizeable resources and funding (44). As a result of these implications associated with probability sampling, it is common for corpus builders to use various combinations of non-probability sampling methods in creating corpora. Moreover, they are able to make valid estimates and observations of their intended populations through the use of documented sampling policies.

Once the population to be investigated with a corpus has been defined, as well as the sampling framework, there are generally two different ways in which the sampling can be performed: *simple* random sampling and *stratified* random sampling. *Simple* random sampling is where all of the sampling units within the sampling frame for a corpus are assigned a number and are then chosen at random using a random number generator or table. This type of sampling can be problematic, as "the chance of an item being chosen correlates positively with its frequency in the population, [and] simple random sampling may generate a sample that does not include relatively rare items in the population, even though they can be of interest to researchers" (McEnery et al. 20). This is in contrast to a *stratified* random sample where the population of sampling units is divided into "relatively homogeneous groups (so-called strata) and samples of each stratum" are taken at random (McEnery et al. 20). For example, if we were interested in sampling according to demographic factors, we could perform a stratified random sample where the population of sampling units are divided on the basis of the age, sex, and/or social class of the writers or speakers. According to Biber (1993), a stratified sample is never less representative than a simple random sample; however, a simple random sample can be less representative than a stratified sample (243).

There is a considerable amount of effort and planning that goes into the design of corpora that enable us to better understand language as it is really used. This methodology is one that is made easier by advances in technology, as evidenced by its growth in popularity. A study performed in 1991 (Johansson) shows that the number of corpus linguistic studies doubled for every five-year period between 1976 and 1991 (Baker 2). One of the major reasons why it was not until the 1980s, as personal computers became widely available (Baker 2), that corpus linguistics became popular is because of this inability to create models that were completely explicit and consistent. However, the perspective that real samples of language could be used as a model for understanding use dates as far back as the mid-eighteenth century, as described in the following section.

CONSTRUCTING MODELS OF LANGUAGE IN USE

In the BC era of corpus linguistics, or before computers, if a person wanted to learn more about language by using samples of it as it was really used, then she had to take slips of paper or cards, hand-write the samples on them, and then file them away in boxes for analysis later. This is exactly what Samuel Johnson described in *Plan of an English Dictionary* his method of creating a corpus of real sentences, or what he called illustrations, from literature to create the *Dictionary of the English Language* (Francis 19). When Johnson created this corpus he ended up with over 150,000 references that yielded nearly 40,000 headwords—all without the use of a computer (Kennedy 14).

In 1747, at about the same time as Johnson, Alexander Cruden created a concordance based on the 1736 King James Version of the Bible. He wanted to identify the major content words of the Bible while also constructing concordances of function words like *he, you, will* and the words that co-occurred with them. His reason for doing this was to check the consistency of factual information in certain parts of the Bible (Kennedy 13-14).

This process of creating models of written and spoken language, as used by actual speakers and authors, continued to be applied throughout the 19th and early 20th centuries, despite its tedious and time consuming nature. It was used in several different areas of linguistic research. Käding (1897) used a corpus of over eleven million words, which was quite large for his time, to establish spelling conventions in German by looking at how letters co-occurred with one another and repeated in significant patterns (McEnery and Wilson 3). Language pedagogy benefited from the use of this methodology through the work of linguists like Thorndike (1921) and Palmer (1933) who created language vocabulary lists for foreign learners using corpora. And even comparative studies of language were performed using this method, like the work of Eaton (1940) who made comparisons of the frequency of word meanings in Dutch, French, German, and Italian (McEnery and Wilson 4).

Thus it is not corpus-based research that is recent, but rather it is the use of technology to compile large corpora with much less effort and to analyze them almost instantaneously that is the current development. This technology helped to alleviate the criticisms that were most often associated with the work of this earlier period: it was done by hand and as such was vulnerable to human error, and there was an inability to observe emerging patterns in the data.

The use of computers to create and analyze corpora began in 1949 with Father Roberto Busa. It was in this year that Busa began work on his *Index Thomisticus*, a machine-readable concordance of the works of St. Thomas Aquinas. Busa started with

10,000 sentences that were each on separate cards that were eventually transferred to punch cards for use on IBM computers. With the assistance of this technology, he was able to perform word-by-word searches on his corpus, which eventually totaled over 15 million words. The result of his work, his legacy if you wish, is the introduction of large-scale computational analysis of the humanities, including language, which is now known as digital humanities or humanities computing (Hockey 2004).

After this point, computer-aided corpus studies grew in popularity, leading to the first computer corpus compiled for linguistic research: the Brown University Standard Corpus of Present-Day English (1963-1964). W. Nelson Francis and Henry Kučera created this corpus at Brown University. It was composed of 500 documents sampled across 15 text types that were authored by native speakers of American English in 1961. Each document consisted of approximately 2,000 words of continuous discourse. Thus, the resulting corpus totaled a little over 1,000,000 words (1,014,312 to be exact). Francis and Kučera designed this corpus to be "a standard of comparison for a variety of studies and analyses of present-day [early 1960s] English," and that is exactly what it became (xvii).

In 1970 a group of linguists from the University of Lancaster, the University of Oslo, and the Norwegian Computing Centre for the Humanities in Bergen came together to create the Lancaster/Oslo-Bergen Corpus of British English (LOB), which duplicated the Brown Corpus. It did so with regard to the date chosen for it to represent (1961) and its sampling framework of 500 documents of approximately 2,000 words of continuous discourse across the same set of 15 text-types (Johansson et al.. 1978). The project was directed by Geoffrey Leech at the University of Lancaster from its inception in 1970 until 1976. However, in 1977 Stig Johansson became the director of LOB and the project was moved to the University of Oslo where it was completed in 1978.

The Brown and LOB Corpora became a source of inspiration for the creation of other corpora using the same model. In the 1990s, researchers at the University of Freiburg took on the task to replicate the Brown and LOB corpora with documents from the early 1990s. The result of their work was the Freiburg-Brown Corpus of American English (FROWN) and the Freiburg-LOB Corpus of British English (FLOB), which represent the language used in the early 1990s for American English and British English. These corpora each contain around 500 texts of approximately 2,000 words each that are distributed again across the same 15 text-types. With the apparent popularity of the Brown and LOB corpora, it is no surprise that in 2006 Paul Baker at the University of Lancaster would again recreate these corpora to represent American English and British English as it was used in documents authored between 2005 and 2007.

In 1980, the Collins Birmingham University International Language Database (COBUILD) project was initiated through a joint venture between the University of Birmingham and the British publishing company Collins. This corpus was designed with the intention of describing the English language based on naturally-occurring texts through lexical analysis, which would eventually lead to the publication of a dictionary of current English for learners of the language. Directed by John Sinclair, the COBUILD corpus contains over 7.3 million words: 6 million words from written texts and 1.3 million from transcribed speech. Data would continue to be added to the COBUILD corpus, as the English Department at the University of Birmingham desired more variety in the corpus, rather than focusing on balance—like the way the Brown style corpora were designed (Kennedy 46). As a result, the COBUILD Corpus, now known as the Bank of English, totals over 2.5 billion words from written and spoken documents of many different text-types and is being added to every month (Harper Collins).

Another corpus of significance to the field of corpus linguistics is the British National Corpus. It was compiled between 1991 and 1994 by a group made up of the Oxford University Press, Addison-Wesley-Longman and Larousse Kingfisher Chambers (publishers), Oxford University Computing Services, the Centre for Computer Research on the English Language at the University of Lancaster, and the Research and Innovation Centre of the British Library. This corpus totals 100,106,008 words resulting from a sampling of 4,124 texts that are intended to be representative of British English in the late 20th century. These documents consist of approximately 90% written texts and 10% transcribed speech. The significant thing about this corpus is that each word within it is automatically tagged for its part of speech and was encoded in compliance with the Text Encoding Initiatives' guidelines using SGML—Standard Generalized Markup Language (University of Oxford). The most recent edition of the BNC, which was released in 2007, has since been converted into XML—Extensible Markup Language. The BNC corpus is considered to be a reference for British English in the last 20th century, due to its finiteness, rather than a historical reference of this variety of English (Kennedy 50).

While the corpora discussed in this brief history of corpus linguistics have mainly covered corpora designed to model content from a general perspective, or a learner perspective as was the case for the COBUILD corpus, we will begin to shift our focus to a type of corpus that has been growing in popularity: those corpora that are more specific with regard to their content.

USING CORPORA FOR MODELING DOMAIN-SPECIFIC LANGUAGE

Domain-specific language corpora are designed to represent language that serves a specific function, like the language of a particular industry (Leech 11). Most of these corpora are corporate in nature. While the study outlined in this dissertation is based on the creation of a domain-specific corpus of regulated nuclear industry discourse, there have been many studies that have been performed prior to this one that help to demonstrate the usefulness of corpus-based analysis of domain-specific language.

The Student Engineering English Corpus (SEEC) is a domain-specific model that contains nearly 2,000,000 words that were sampled from required engineering textbooks of various specializations for students at Walailak University in Thailand (Moudraia 2003). While it was created for use in establishing a list of terms that were central to all engineering disciplines, Olga Mudraya explains in "Engineering English: A Lexical Frequency Instructional Model" how this corpus can be used in the methodology for teaching English to technical students. She advocates for a greater use of lexical-approaches to teaching language in "chunks" (Johns 1991, McEnery and Wilson 1997).

Using the COBUILD Bank of English Corpus and the British National Corpus as her reference for comparison, Mudraya analyzed SEEC to determine those words that were significant for Engineering English. What she discovered was that the most frequently encountered words in the corpus were what she classifies as *sub-technical*, meaning words that have both technical and non-technical senses (i.e. *iron, force, stress, current,* etc.). It was also discovered that the non-technical senses of these words were used more often than the technical ones. She also found that the key verbs used in this corpus corresponded with those most commonly used in academic writing: *act, apply,* *assume, be, become, calculate, consider, correspond to, define, determine,* etc. (Mudraya 242). She determined that these sub-technical terms most often formed prefabricated, formulaic multi-word units with non-technical ones. These clusters, she proposes, are much more effective for learners in that students "can get a much better idea of the use of the word" (Mudraya 237).

The main implication resulting from this work is that sub-technical and academic vocabulary should be given more attention in the English for Special Purposes (ESP) classroom, like identifying those words that are distinctive for the nuclear power industry. Her work indicates that attention should also be placed on the acquisition and use of formulaic multi-word units, i.e. phrases or collocations, that were characteristic of this corpus and that can help make students more effective in their communications within this domain. She finally concludes that the use of domain-specific corpora and corpus-based assignments in the ESP classroom not only help students understand the language of these industries, but will help them to productively use the "language prefabs." Also, her work implies that there is a need for the results from a corpus linguistic analysis of domain-specific language that focuses on extralinguistic contexts like industry group and geography. The outcome of this type of analysis can help to better educate students in ESP classrooms on the importance of the author-audience relationship in the workplace: i.e. the different ways in which these industries use language depending on who the author and audience are, and even where they are located.

While Mudraya's application of the patterns she uncovered in Engineering English using corpus-based methods were for implementation in the ESP classroom, Marlene Kemp-Dynin's dissertation titled "The 'Company' Words Keep: A Corpus-

Based Analysis of Collocational Patterning in Business Terminology" has a different focus. It is one example of a study using corpus linguistic methodologies to create a model for learning about a language variety for use in dictionary creation. She based her research on the fundamental idea that word use "is highly patterned and these patterns vary according to context and text type" (3). In order to discover whether or not most business terms vary according to the specific business practice, industry, or company, she created a domain-specific reference corpus of business language that was balanced and representative (over one million words of text), in addition to using The Tobacco Documents Corpus from the University of Georgia—a corpus of more than 500,000 words from documents internal to the tobacco industry. She then used *WordSmith Tools,* a piece of concordance-making software, to create frequency lists from each corpus and compared the 25 most frequently occurring business terms that both corpora had in common. After this step was completed, she compared the collocates of each term between the corpora.

As a result of her research, Kemp-Dynin discovered that companies do share common business terminology. She demonstrates this fact by comparing the five most frequent collocates of each of her 25 terms between both her reference corpus and the Tobacco Documents Corpus: 15 of the 25 terms (60%) shared at least one collocate—or commonly occurring word within a four-word span in front of a key term or following it—as one of the five most frequent collocates between the two corpora. While we may think that this is a pretty low percentage, if we are to assume that there is such a thing as business language, it demonstrates that there is significant variation between specific business groups and warrants more investigation into domain specific languages. She provides evidence that there is variation in how business terms are used: creating the opportunity and impetus for analysis of domain-specific language with regard to variation resulting from extralinguistic contexts. and the different contexts within which variation is present for them.

Kemp-Dynin's analysis of the collocations of these terms also yielded the conclusion that the language surrounding each of the business terms is affected by the specific industry or the company by comparing both her reference corpus and the Tobacco Documents Corpus to the Brown Corpus (a general reference corpus of American English from the 1960s):

> The meaning of *advertising* is modified according to the context in which it is used. The most frequent collocate of *advertising* in the Brown Corpus, a general purpose corpus, is *magazine*. When *advertising* is used in general business language, as exhibited in the reference corpus, the primary collocate is *sales*. In a more specific context, such as the Tobacco-Documents Corpus, a significant number of collocates will be terms related to the context, in this case the tobacco industry. (Kemp-Dynin 197)

These collocations were used to create dictionary entries for use in the business sector. Furthermore, she confirmed the observation previously made by Kjellmer (1987) and Partington (1998) that collocations and their frequencies vary according to text types. She did this by comparing the frequencies of the three corpora (business reference, Tobacco Documents, and the Brown Corpus) and discovering that "while the reference corpus and the Tobacco-Documents Corpus provide an abundance of collocations per term, the Brown Corpus does not. In fact, many of the terms have only a few collocates which occur more than once" (198). Ultimately, Kemp-Dynin's work demonstrates that further research regarding business language is needed, due to the fact that corporate discourse varies according to industry.

The last example of domain-specific corpus-based modeling I would like to share is the work performed by researchers at the University of Georgia to create the Tobacco Documents Corpus. In "Looking for the Smoking Gun," Kretzschmar explains how this model was created from a database of over four million documents that were produced by tobacco industry defendants for use in legal proceedings (Kretzschmar 2004, 31). The documents that were made part of this database were communications concerning corporate operations. The group of researchers involved in the analysis of the tobacco industry documents (TIDs) made the decision to create a domain-specific reference corpus out of them in order to have a basis of comparison for industry documents that were thought to contain deceptive or manipulative language strategies. Thus, there was a need to create a corpus that would represent the language internal to the tobacco industry, rather than addressing the content of a few documents.

After sampling a reference corpus of over 500,000 words using an innovative, systematic framework (which will be covered in great detail in Chapter 4), they compared it to the Brown Corpus to identify the top 50 content words for it. What they found was that these 50 words could be classified into four groups: trade words like products and components, vocabulary related to selling, disease terms, and terms that applied to research like *market* and *product* (Kretzschmar 41). They were left with one word, *current,* that could not be classified into these four groups.

Using the 50 key terms that were found to occur more than expected in the TID Reference Corpus than in general American English as modeled by the Brown Corpus, they tracked these words with regard to time. More specifically, they identified if each of the 50 key terms were more or less present in the language of the industry in comparison to general

American English from the 1950s through the 1990s using the Brown and FROWN corpora. For example, they determined that the word *cigarette* was used more in the 1950s and less in the 1990s. Meanwhile, *carton, pack,* and *product* were used more in the 1990s. They proposed that these trends possibly indicated a decreased focus on the industry in the 1990s and an increased focus on the product. By using a systematic sampling framework of domain-specific language, Kretzschmar et al.. were able to make a few generalizations about corporate communications, i.e. that they are mainly internal. Most importantly, this research provides a systematic methodology for creating a domain-specific balanced/representative corpus from a database of documents that behaves more like a monitor corpus to talk about domain-specific language in a concrete way.

SUMMARY OF CORPUS LINGUISTICS AS MODELING

Corpus linguistics is a method of modeling language as it is actually used: it makes the implicit explicit. This method allows us to observe that meaning in language is expressed through the frequent use of words with one another in their original context. Although this method of analysis dates back to the eighteenth-century, it has only gained popularity since the invention of the computer, which makes corpus linguistics much more effective and accurate than before. There are different types of corpora, and they serve different purposes. For example, Mudraya suggests there is a need for more corpus-based analysis of the "chunks" of language in different domain-specific fields because it is necessary for effective instruction in English for Specific Purposes classrooms. Moreover, Kemp-Dynin's research indicates that the meaning associated with these chunks and key terms is constructed on a more localized basis, and thus there is a need for more domain-specific research. *The*

Linguistics of Speech also helps us to understand the contextual and local nature of language by suggesting that industry group and geographic contexts interact with the language of different text-types. Analysis of language using corpus linguistics with regard to these dimensions can help us better understand localized meaning, like that in domain-specific language, even more. Based on these ideas, histories, and past accounts of research in domain-specific language and its implications, this study employs the use of a domain-specific balanced/representative corpus of regulated nuclear power discourse. Furthermore, it was created using a systematic sampling methodology originally designed for use with the Tobacco Industry Documents.

CHAPTER 3

CONSTRUCTING A MODEL OF REGULATED NUCLEAR DISCOURSE

For many of us, in order to understand an abstract concept like language, we need a visual representation to see, touch, and even manipulate: a model. Willard McCarty provides an explanation of models and modeling in the humanities in “Modeling.” He defines a model as “either a representation of something for purposes for study, or a design for realizing something new,” and modeling as the process of constructing and manipulating models (1). Moreover, he proposes that it is through modeling that we “make the best and most productive sense through what we observe” (1). Modeling is a relationship between the researcher and her data or theory that is best explained with Minsky’s formula: "To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A" (McCarty 2). In situations where the object of study is abstract, the best method for making explicit the implicit intuition we may have about a particular subject is the use of models.

In order to create an explicit model for investigating our intuition about language, a great deal of care must be put into the planning process for its creation. It would not be wise to construct a house without first putting considerable thought into designing a blueprint that shows where to install the doors, electrical outlets, plumbing, etc. The same is true for building a corpus. We need a sampling framework, or blueprint, to tell us how to put this model together. To make such a blueprint, we need to know what characteristics are important for what we are building. While there are many factors of importance for planning a corpus' sampling procedures, five are most vital: corpus size, text length, text types, text sources, and time frame (Meyer 30-46). The intended use of the corpus affects a researcher's approach to all of these factors. It would not make sense for us to put plumbing for a bathtub in the middle of the living room on our blueprint, as the general function of a living room is not personal hygiene. In exactly the same way, the purpose of a corpus should always be ***the*** deciding factor when planning the sampling frame of a corpus.

To demonstrate this point, I would like to compare the differences in the number and types of texts sampled for corpora with different purposes: general use versus specialized use. A general-use corpus is one that is constructed to investigate a language in its entirety. As a result, this type of corpus requires very careful planning to make sure that text types are balanced. This implies that a corpus linguist makes certain assumptions about the importance of particular text types in a language before sampling. For specialized corpora, those models constructed not necessarily to investigate language as a whole but instead a more specific instance of language in use (Kennedy 20), can be comprised of pieces of texts, as opposed to the full texts suggested by Sinclair for general-use corpora, as long as they make up a coherent unit (Meyer 38). Moreover, these samples should be collected from subcategories like text-type, location, or production in order to create a representative model of specialized language use, like that of the regulated nuclear power industry. In order to do this, I will be following the methodology utilized in the analysis of documents from the tobacco industry at the University of Georgia due to the similar nature of the documents available from the NRC and those that became a part of the Tobacco Documents Corpus.

CREATING THE TOBACCO DOCUMENTS CORPUS

In the fall of 1998, a settlement was reached by the National Association of Attorneys General and seven major United States tobacco industry corporations in order to impose regulatory measures on the tobacco industry. As a result, the seven corporations were required to release all industry documents to the public that were not considered attorney-client privileged nor to have contained proprietary trade information. In "Construction and Analysis of the University of Georgia Tobacco Documents Corpus," Clayton Darwin describes in great detail how rather than performing analysis on the tobacco documents for legal or political purposes, researchers at the University of Georgia sought to "treat the TIDs [Tobacco Industry Documents] as a corpus and to apply accepted methods of corpus and forensic linguistics and rhetorical analysis" to do several things, including "establish baseline descriptions of various linguistic features of this unique set of texts" (Kretzschmar et al.. 33).

To create a representative, specialized reference corpus from such a large database of texts, these researchers employed a systematic approach for their study. Their sampling plan for this corpus was based on the report, "Sampling Plan for Creation of Corpora for the Tobacco Documents Grant" (Kretzschmar 2001). In this report, Kretzschmar proposed that the sampling of the Tobacco Documents Corpus be performed in two stages:

> An initial limited sample from the entire set of documents in order to estimate the prevalence and range of document types; and secondarily, the final representative sample based on quotas derived from the data provided by the initial sample. In other words, first determine what document types exist in significant proportions in the entire document set, and then determine a quota-based sampling procedure that best matches the proposed goal of the Tobacco Documents Project. (Darwin 56)

The first phase, or pilot corpus, was to be drawn in order to "'determine the best classification of text types and estimate their proportions within the overall body of texts'" (Darwin 56). Therefore, special attention needed to be paid to text types for the pilot corpus upon which the reference corpus would be built in order to avoid skewing the data. Before the TDC pilot could even be created to investigate this variety, they had a slight issue from a theoretical standpoint with their sampling population.

In order to deal with large-scale monitor corpora like the Tobacco Documents for comparative corpus-based research, Darwin shares Graeme Kennedy's insight that the dynamic nature of monitor corpora typically renders it unsuitable for comparative studies since one cannot perform descriptive linguistics on it, as it is constantly changing (59). The need arose for the TDC researchers to create a reference corpus from this larger "monitor corpus" that would not be constantly changing. To uncover what kinds of documents were available as part of the Tobacco Industry Documents, as well as the extent of each document, Darwin explains that the UGA researchers sampled the body of documents according to a fixed random sampling frame that would give every document in the collection an equal chance of selection (62-63). The decision was made to take 0.001 of all the documents available, which totaled a little over 300 documents. They then randomly selected a month and a year and queried the Tobacco Documents database to find out how many documents were available for each year. After the random selections were finished, all of the documents in the core corpus were classified using several categories, including:

1. Public Health: Significant for Public Health or not significant for Public Health.

2. Audience: Industry-Internal Audience or Industry-External Audience was established to be exclusive of each other. Documents were classified as internal if they were addressed to persons or groups within or hired by the company from which the document originated, or if they were correspondence between tobacco companies. This was eventually extended to include vendors at all levels of the tobacco industry and all for-profit and for-hire organizations involved in the research, growing, processing, distribution, and sale of tobacco products. Otherwise documents were classified as EX.
3. Addressee: Named or Unnamed.
4. Text Types. (Darwin 73-79)

These criteria were used as the basis for making sure the contents of the corpus matched the intended use of the model. For example, all of the documents that were not designated as being significant for Public Health, being addressed to an industry-internal audience, possessed a named addressee, etc., were rejected from becoming a part of the final quota sample. After creating the core sample for the Tobacco Document Corpus, the researchers used the distributions they observed to develop a protocol for sampling documents that fit their criteria to come a part of the quota sample. What they discovered was that their proportions for document rejection were nearly the same for the final reference corpus as the initial pilot sample: they demonstrated that this particular random selection methodology works.

APPLYING THE TDC METHODOLOGY TO ADAMS

The first obstacle that had to be overcome with creating the TDC was that the tobacco industry documents "as a whole would be classified as a monitor corpus" that "is not fixed in size or content, but dynamic, designed to grow with the release of additional documents in order to monitor change over time" (Darwin 59). As part of the Freedom of Information Act of 1966, the American public has a "right to know" about government records and documents ("The Freedom of Information Act"). Since the tragic events following September 11, 2001, the NRC provides to the public any and all documents about nuclear reactors here in the United States that are not found to contain "sensitive information." Sensitive information is defined by the NRC as being data that has been found to be potentially useful to terrorists, proprietary knowledge for licensees, or "information deemed sensitive because it relates to physical protection or material control and accounting" ("Withholding of Sensitive Information for Nuclear Power Reactors"). All documents that do not possess these characteristics are made available through the NRC's Agency Documents Access and Management System (ADAMS) database.

ADAMS is composed of two secondary collections. First, there is the Publicly Available Records System (PARS) Library that "contains more than 520,000 full-text documents that the NRC has released since November 1999, and several hundred new documents are added each day" ("ADAMS Public Documents") to a web-based archive. The second library is known as the Public Legacy Library and contains over 2 million bibliographic citations for documents earlier than those found in PARS. Starting in January 2000, the Nuclear Regulatory Commission began publishing all of their publicly available documents to ADAMS online ("NRC—Public Document Room").

In order to create a reference corpus of regulated nuclear power language from the ADAMS database, which is essentially a large monitor corpus, I followed the Tobacco Documents Corpus methodology for assembling a pilot corpus. First, I randomly selected a different month for each of the 12 full years available as part of the ADAMS-PARS archive: 2000 through 2011 (Table 3.1).

Table 3.1: ADAMS Random Month Selection

Year	Random Month Selection
2000	November
2001	January
2002	July
2003	September
2004	June
2005	February
2006	May
2007	August
2008	March
2009	October
2010	December
2011	April

I then queried the database for each NRC licensee by using their docket numbers. Docket numbers are unique identification codes assigned to each licensee. Any and all documents being written by the licensee, being written to the licensee, or being sent to the licensee as informed communication for regulatory action or rulemaking are assigned to the licensee's docket. Essentially, the docket is considered a record of communication for the licensee. As such, this identification number proved to be the ideal way for querying the available documents for each nuclear reactor being regulated by the NRC. After I finished my queries, I made an observation quite similar to one Darwin noticed in the Tobacco Documents: the documents varied greatly in count and length for each month/year and each license (Table 3.2). Darwin suggests that this particular behavior

Table 3.2 ADAMS Document Availability by License Excerpt

Sites	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Arkansas Nuclear 1	20	21	21	15	21	19	16	15	11	7	6	17
Arkansas Nuclear 2	19	11	16	94	27	25	20	13	30	6	4	15
Beaver Valley 1	9	13	11	22	15	41	15	150	32	16	3	11
Beaver Valley 2	21	14	13	22	14	40	17	152	38	19	8	13
Braidwood 1	24	27	26	12	10	10	43	13	20	21	13	20
Braidwood 2	25	28	25	12	9	11	40	13	25	19	12	17
Browns Ferry 1	11	10	21	18	32	24	36	33	15	27	15	17
Browns Ferry 2	13	17	22	19	21	19	27	23	14	19	14	15
Browns Ferry 3	12	15	22	19	22	18	29	24	16	18	12	18
Brunswick 1	7	9	12	16	12	33	10	17	22	15	12	17
Brunswick 2	8	9	12	16	11	33	10	16	21	14	12	19
Byron 1	17	28	7	25	10	10	22	15	19	12	14	26

denotes "a considerable change in communicative habit as well as variability in text type" (64). As a result, I made the decision to perform a quota-based sample for each licensee dependent on the number of documents available for each month/year combination. The reason for making this decision is of importance because straight sampling across the years for each licensee without regard for document availability frequencies increases the possibility for a corpus that is not representative of the population of documents available in ADAMS.

I also took 0.001 of all the documents available based on my initial querying, which totaled 30 documents per licensee. These 30 documents were randomly selected over all 12 years based on the number of documents available within each year. An example of the sampling distribution for Indian Point 2, one of the 104 licensees, can be found in Table 3.3. After establishing the number of documents I would take from each year for each licensee, I used Random.org to generate sets of random integers to represent the each result from the query that I would select to be part of the corpus. For example, the random selections for April 2011, for Indian Point 2 were entries 28 and 39. After the random selections were chosen, the appropriate documents were downloaded from ADAMS as .PDF files that had already been converted into a machine-readable format using optical character recognition (OCR) software by NRC librarians.

Before I address the classification system of my randomly selected NRC documents, I want to discuss a feature that is available from ADAMS. When selecting documents for download in this system, you have the option to generate reports about your selections. These reports provide descriptive metadata about each of the documents you have selected (Figure 3.1). One report in particular, the Legacy Custom report,

Table 3.3: Production-Based Document Sample for Indian Point 2

Year	Random Month	Available	Sampled
2000	November	89	4
2001	January	95	4
2002	July	37	2
2003	September	31	1
2004	June	29	1
2005	February	10	1
2006	May	45	2
2007	August	121	5
2008	March	83	4
2009	October	42	2
2010	December	45	2
2011	April	50	2

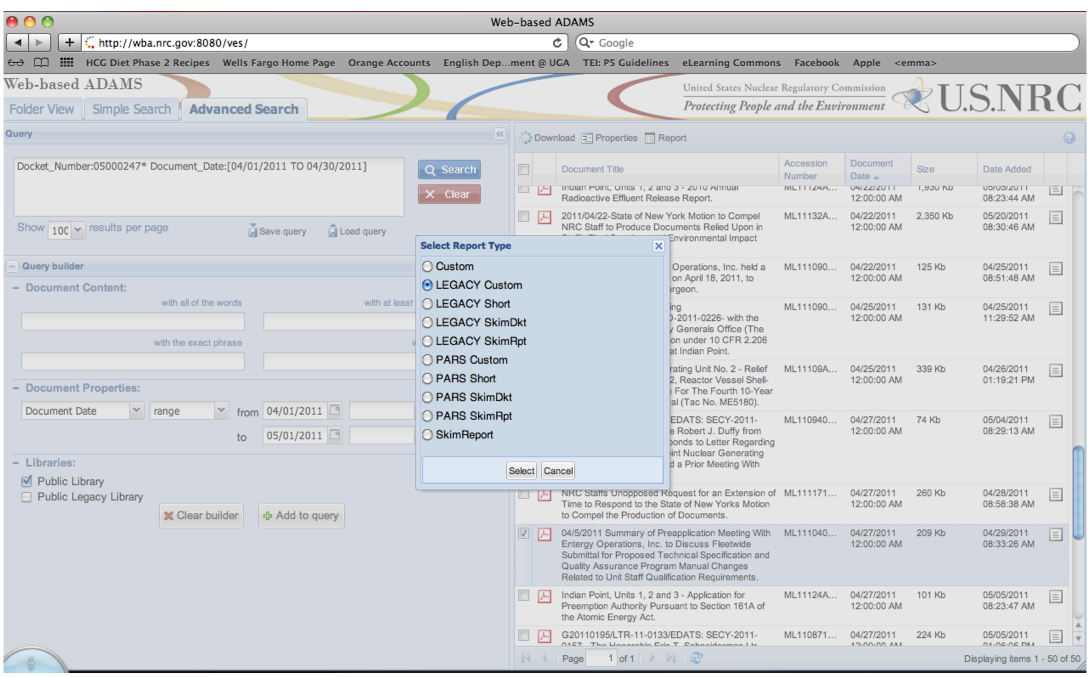

Figure 3.1: ADAMS Report Selection

proved extremely useful for classifying the documents (Figure 3.2). This report provides essential information like Document Type, Author Affiliation, Addressee Affiliation, and even the originating Docket Number of the documents. A report for each Pilot selection was generated and saved as a .CSV file.

In order to take advantage of the metadata in the Custom Legacy reports downloaded from the ADAMS database, I decided to use FileMaker Pro, a suite of database creation software, to create a database for the Pilot selections. First, I imported all of the Pilot Legacy Custom .CSV files into the software and created a database. The resulting file was a searchable series of records from all of the documents. Using all of this metadata, I verified my randomly selected documents against the .PDF files, and classified them according to the following guidelines that were adapted form those used by researchers creating the Tobacco Documents Corpus:

1. Nuclear Power Regulation: No communications involving the regulation of nuclear materials for medical or research uses were included in the pilot corpus, only documents related to the regulation of nuclear power.
2. Industry-Internal Author/Audience or Industry-External Author/Audience: Documents are classified as Audience Industry-Internal if they are addressed to persons or groups within or hired by the licensees or the NRC, or if the document is correspondence between individuals at the NRC or individual licensees. Furthermore, vendors at all levels of the nuclear industry and all consultants (legal, environmental, etc.) and contractors (engineering firms) involved in the production, management, regulation, or business of nuclear power are to be considered internal as well. Otherwise documents are classified as external to the nuclear power industry (Appendix A).
3. Document Types: All documents are assigned document type designations by the NRC librarians. These designations can be found on the Custom Legacy report.
4. Docket Designation: If the docket number assigned to the document is

10/19/11 6:22 PM

Accession Number: ML003769460
Estimated Page Count: 6
Document Date: 11/10/00 12:00 AM
Document Type: License-Monthly Operating Report
Availability: Publicly Available
Document Title: October 2000 Monthly Operating Report for Vermont Yankee Nuclear Power Station.
Author Name: Bronson K H
Author Affiliation: Vermont Yankee Nuclear Power Corp
Author Affiliation Class:
Addressee Name:
Addressee Affiliation: NRC/Document Control Desk
Addressee Affiliation Class:
Docket Number: 05000271
License Number: DPR-028
Case/Reference Number: BVY-00-104
Document/Report Number:
Keyword: DPC converted to PDF
Package Number:
Document Date Received:
Date Docketed:
Related Date:
Comment:
Document Status:
Media Type: Electronic
Physical File Location: ADAMS
FACA Document: No
Date to be Released: 11/28/00 12:00 AM
Distribution List Codes: IE24
Contact Person:
Text Source Flag: OCRed from scanned image - no corrections made
Official Record?: Yes
Document Sensitivity: Non-Sensitive
File Name:
File Size: 184757
****

http://wba.nrc.gov:8080/ves/viewreport.jsp Page 1 of 1

Figure 3.2 ADAMS Legacy Custom Report Sample

the same as the licensee, it was classified as "Own." The designation "Other-Same Site" was used if the docket number was that of a licensed nuclear reactor on the same site. "Other-Same Corporation," designated the situations where the originating docket number assigned to the document represents a licensee owned by the same corporation as the docket number being searched for each document. Finally, the designation "Other-No Affiliation," was used to indicate documents assigned to a licensee's docket that originated from a licensee not possessing any of the aforementioned qualities.

5. Language-Based: All of the documents are marked as being language-based or not in order to identify documents that are image-based like drawings and photographs.
6. Length: Texts shorter than 50 words of continuous discourse were marked so that they can be excluded from the corpus. Likewise, documents longer than 3,000 words are denoted in the metadata so that they can be sampled (1,000 words from the beginning, 1,000 words from the middle, and 1,000 words from the end) to avoid bias.

Once all of the classifications for the pilot corpus had been made, I was able to analyze characteristics of the documents sampled from the population of those available to the public on the ADAMS Database.

THE PILOT

One of the first observations I made when classifying the documents that were selected to be part of the Pilot was that although the sample only allowed for unique document selections of the results from each docket number's database query, duplicate documents (documents being assigned identical accession numbers by the NRC) were sampled because a single document may be assigned to multiple dockets by the NRC. By generating the Custom Legacy Report (Figure 3.2) for each document that was randomly selected to be part of the corpus, I was able to identify exactly which dockets were assigned to a specific document. For the purpose of the reference corpus, this particular occurrence causes distortion in the sampling of the pilot at the docket level. However, I needed to preserve the fact that a single document may be a part of the language in use of multiple licensees. With FileMaker Pro, I was able to denote all of the dockets assigned to each document while only having a single instance of the document in the reference corpus. This way, I am still able to export all of the documents in the corpus that are assigned a unique docket number for comparison to the rest, albeit utilizing sample with replacement statistics. As a result of eliminating all of the duplicate documents from the Pilot, the 3,120 documents downloaded from the ADAMS were reduced to 2,775 unique samples.

Another characteristic documented by the NRC librarians on the Custom Legacy Report is document type. With regard to the types of documents that are part of the Pilot sample, an interesting pattern emerges when you plot their aggregate frequencies. As you can see in Figure 3.3, there is a very distinct, and steep, asymptotic hyperbolic curve, or A-curve. The A-curve follows a pattern known as Pareto's Principle: 80% of the results come from just 20% of the possibilities. In 1906, Vilfredo Pareto, an Italian economist, observed that 80% of Italy's wealth belonged to 20% of the population. He developed a mathematical formula for modeling this type of unequal distribution. After his results were published, individuals in many different disciplines also began noticing that the number of events larger than a specific number in their datasets was an inverse power of that number: i.e. business management and even linguistics (Kirman 2008). Almost 30 years later, an American linguist named George Kingsley Zipf observed this same pattern in corpora composed of natural language utterances. However, he sought a way to calculate the number of times that a word occurred in a particular corpus, given its rank.

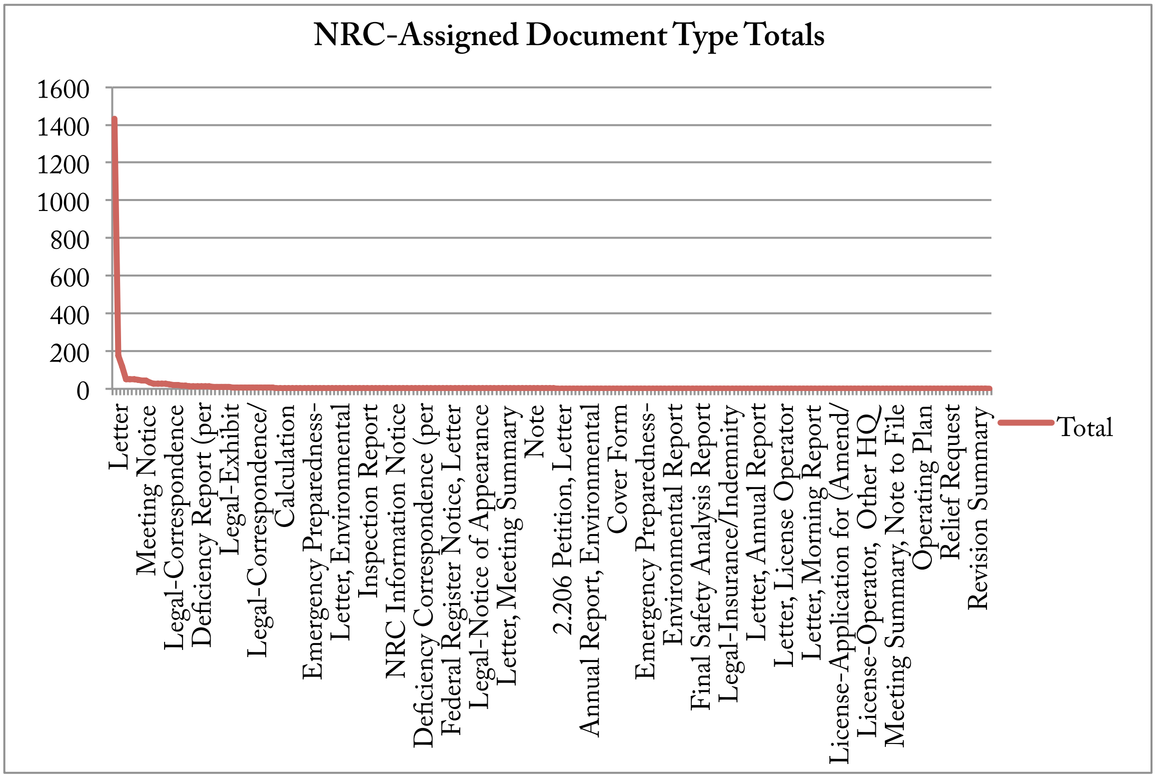

Figure 3.3: Pilot Document Totals Before Splitting Multiples

In other words, calculating the frequency of the word used most often in a corpus, 14th, or even 1,000th. This mathematical model, known as Zipf's law, is a derivative of Pareto's Principle; it is an alternative way of describing the exact same behavior (Adamic and Huberman 143-144). Thus in linguistics, Zipf's Law is used synonymously with the 80/20 Principle, and it is a pattern that is expected also to manifest when the language of the reference corpus is analyzed in the chapters to follow.

In the case of the data in Figure 3.3, we are not plotting word frequencies against their ranks, but rather document types. For the Pilot, we can see that a majority of the documents have been denoted by the NRC as being letters, 1,125 in fact. However, when looking at these documents, many of them appeared to be rather long. So, I went through all of the documents in the pilot and marked them for whether or not they had a unique attachment: 44.45% of them did. Because of this observation, although the NRC librarians have designated a particular file as being a certain document type, when it comes to Letters especially, the potential exists for multiple document types to be present. After splitting these multiple documents, the result was 4,773 individual .PDF files in the sampling. Once all of the files possessing multiple documents were split apart, thereby changing the scale of document types in the Pilot, there still appears to be an A-curve with regard to the relative frequencies of the document types (Figure 3.4).

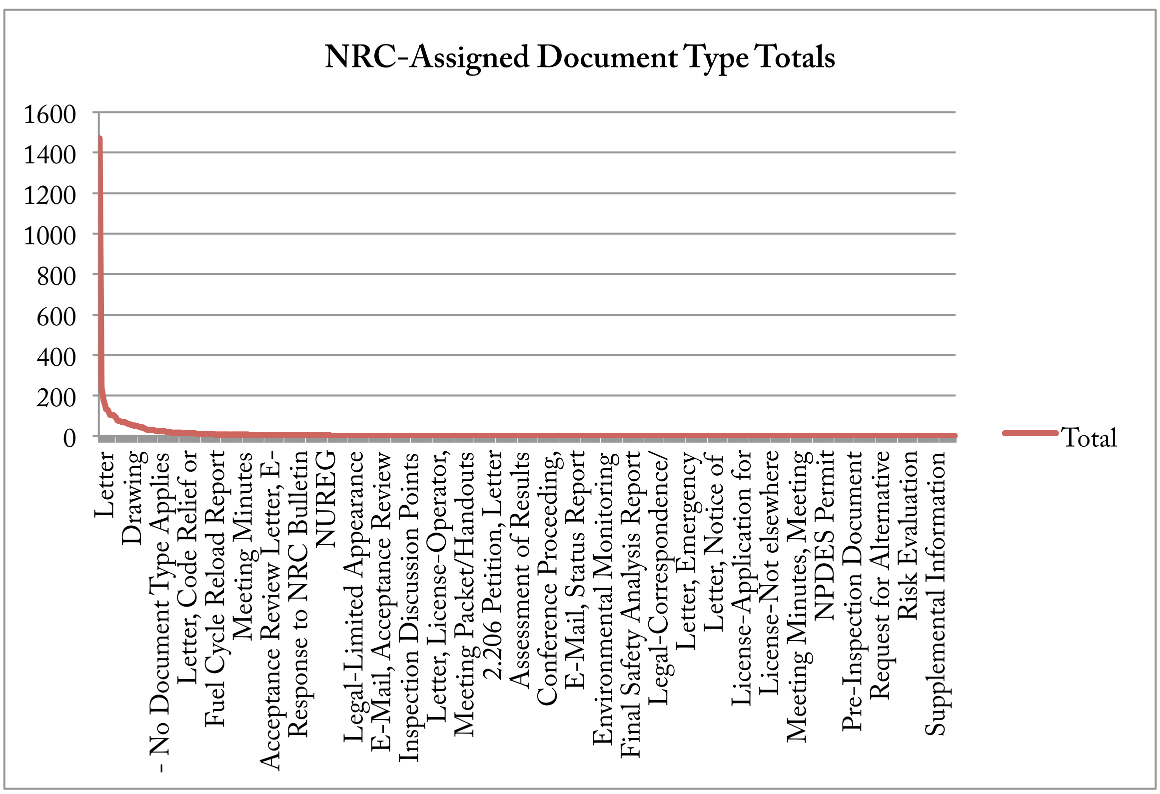

Figure 3.4: Pilot Document Totals After Splitting Multiples

Letters were still the most frequent document after the scale changed, but the frequencies of other documents like Safety Evaluations increased drastically (from 2 to 104). Although the number of document types in the Pilot changed, as well as their relative distribution, the A-curve is still present. This particular behavior is called scaling: the A-curve is present at different aspects, or levels of scale, in the corpus. Scalability of data through A-curve distributions has also been documented extensively in speech data across different linguistic variables, time, and even geographic location (Kretzschmar 2008, Burkette 2009, and Johnson 1992). The frequency of document types is in fact scalable for this particular population of documents. This is an important quality of language in use that should also be documented in the lexical frequencies of the ADAMS documents with regard to proximity.

In order to learn more about the language of the nuclear industry, not only do the documents in the corpus need to be about nuclear power, but also the authors need to be classified as internal. Of the 4,773 documents from the ADAMS-PARS database, 97.76% of them were authored by internal sources. This means that 4,666 documents were kept as part of the reference corpus while 107 documents were not (105 of them were written by externally affiliated authors, and the affiliation of two documents could not be determined). With regard to the internal/external status of the sampled documents' audience affiliations, since the function of the NRC is to ensure "that people and the environment are protected," both internally and externally-directed documents are maintained as part of the corpus.

Of the 4,666 documents remaining in the Pilot, only 2.27% (or 106 of them) were not language-based documents, such as drawings and photographs (Figure 3.5). They

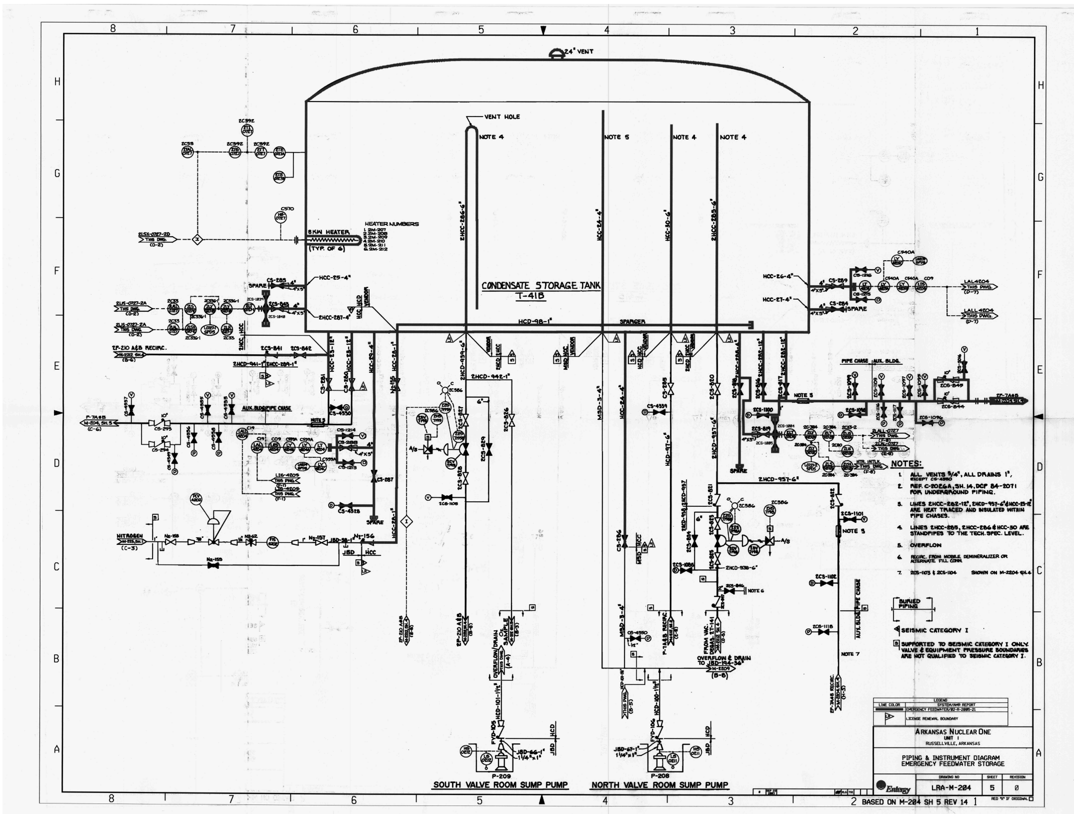

Figure 3.5: Arkansas Nuclear One Condensate Storage Tank Drawing

were not kept as part of the reference corpus. For the 4,560 documents now remaining in the pilot, the average page length was 32.3 pages with a standard deviation of 79.79. This tells us that the length of the documents available from the NRC database is highly variable with documents ranging from one page to 2,996 pages. However, just because a document has numerous pages does not necessarily mean that it contains a great many words. When looking at the sampled documents, 78.79% of them (3,806) contained 50 words or more of continuous discourse. This means that 967 documents could not be used because they were too short. After taking out all of the documents from the pilot sample that were not authored by groups internal to the nuclear power industry, were not language based, and had less than 50 words of continuous discourse, we were left with 3,593 documents. In other words, the Pilot had a rejection rate of 24.72%.

In order to see if this random selection methodology was successful, I performed three iterations of the sampling protocol to look for consistency in the proportions of document rejection to create a sizeable reference corpus from the ADAMS database.

MEASURING REPRODUCIBILITY

One of the most important qualities of a sampling methodology is that it be reproducible. For this reason, I conducted three additional rounds of sampling from the NRC ADAMS database using the previously described protocols. One way to evaluate the reliability of this sampling method is to evaluate the statistical similarities, or rather evaluate if there are any differences statistically in the rates of rejection for documents in the second, third, and fourth iterations of sampling with respect to the Pilot for all of the classification criterion (Darwin 93). Although a quota-derived sampling protocol based on the documents available in the ADAMS database was used, I wanted to see whether or not the ratios of documents rejected due to the qualities of each document were consistent across all of the iterations in comparison to the Pilot.

In order to evaluate my sampling procedures, I performed a two-proportion z-test at a 99% confidence level at each stage where I rejected documents. As was done with the Pilot, I eliminated all of the files that were duplicates with respect to their unique Accession identification numbers for each iteration. There was no statistically significant difference between the rejection ratios of all three iterations in comparison to the pilot (Table 3.4).

Table 3.4: Duplicate Accession ID Rejection Ratios

Iterations	Duplicate Documents	Total Documents	Rejection Ratio
Pilot	345	3120	11.06%
Iteration 2	355	3120	11.38%
Iteration 3	368	3120	11.79%
Iteration 4	371	3120	11.89%

After making sure all of the documents within each iteration were represented only once, I proceeded to verify all files were composed of only one document. The resulting proportions of documents also had no statistical difference from the pilot at a 99% confidence level (Table 3.5).

Table 3.5: Ratio of Original Number to Number After Splitting Multiples

Iterations	Original Number of Documents	Number of Documents After Split	Ratio
Pilot	2775	4773	58.14%
Iteration 2	2765	4625	59.78%
Iteration 3	2752	4618	59.59%
Iteration 4	2749	4581	60%

Beginning with the first classification step after eliminating all duplicates and splitting all files possessing multiple documents, eliminating all of the externally-authored documents, there was still no statistically significant difference between the rejection ratios of all three iterations in comparison to the Pilot (Table 3.6).

Table 3.6: Externally-Authored Document Rejection Ratios

Iterations	Externally-Authored Documents	Total Documents	Rejection Ratio
Pilot	107	4773	2.24%
Iteration 2	111	4625	2.4%
Iteration 3	90	4618	1.95%
Iteration 4	106	4581	2.31%

After all of the externally-authored documents were removed from the samplings for each iteration, all of the remaining documents classified as not being language-based were also filtered out. Yet again, the proportion of internally-authored documents that were not language-based was consistent across all three additional iterations in comparison to the Pilot at a 99% confidence level (Table 3.7).

Table 3.7: Non-Language-Based Document Rejection Ratios

Iterations	Non-Language-Based Documents	Total Documents (Internally-Authored)	Rejection Ratio
Pilot	106	4666	2.27%
Iteration 2	113	4514	2.5%
Iteration 3	104	4528	2.3%
Iteration 4	103	4475	2.3%

The final step for all three of the additional iterations was to identify all of the documents having at least 50 words of continuous discourse. Using FileMaker Pro, I was able to verify the number of documents that were internally-authored and language based, but were too short for inclusion according to the classification criteria. With a 99% confidence level, not only was I able to verify that these proportions also did not have a statistically significant difference with regard to this final classification (Table 3.8), but also with regard to the total rate of rejection for iterations two through four in comparison to the pilot sample (Table 3.9).

Table 3.8: Document Length Rejection Ratios

Iterations	Documents Having Fewer Than 50 Words	Total Documents (Internally-Authored & Language-Based)	Rejection Ratio
Pilot	967	4560	21.21%
Iteration 2	886	4401	20.13%
Iteration 3	865	4424	19.55%
Iteration 4	831	4372	19.01%

Table 3.9: Total Rejection Ratios for All Iterations

Iterations	All Documents Rejected	Total Documents	Rejection Ratio
Pilot	1180	4773	24.72%
Iteration 2	1110	4625	24%
Iteration 3	1059	4618	22.93%
Iteration 4	1040	4581	22.70%

Ultimately, this analysis provides an additional level of confidence to the methodology that the sampling procedure is reliable across multiple iterations.

CREATING THE REFERENCE CORPUS

Knowing that the Pilot, Iteration 2, Iteration 3, and Iteration 4 were all sampled using a quota that was based on production, as well as exhibiting statistically similar proportions throughout the classification process, the decision was made to create the reference corpus for this study by synthesizing all of the iterations together. However, it was discovered that documents had the potential to be sampled in multiple iterations due to the documents being assigned to multiple dockets.

I imported all of the iterations into one database and designated which iteration/s from which each document was sampled. Then, I deleted all duplicate document entries. The resulting reference corpus was composed of 7,110 documents having unique accession numbers, totaling over 9 million words. In order to establish a baseline description of this variety of language, while also focusing on the differences in industry group and geographic proximity, I used WordSmith Tools, a piece of concordancing software, to analyze word frequencies and patterns of co-occurrence. To discover the statistically significant key words in the reference corpus, I compared them to a corpus of general American English created by Paul Baker (2006). Baker's corpus is modeled after the Brown Corpus, a corpus of general American English totaling approximately 1 million words that was created in the 1960s. Baker's corpus has the same distribution of text types as the Brown Corpus, is approximately 1 million words, and is composed to be representative of general American English in the mid 2000s. The Baker Corpus was used to compare with the nuclear reference corpus because they are sampled from populations of documents from similar periods of time.

Using WordSmith Tools, I generated word frequency lists for both Baker's corpus and the nuclear reference corpus. Next, I used the software to statistically compare the two corpora to find out which words were linked to the nuclear industry documents in comparison to general American English: the key terms. Within this software there are two ways that you can measure the strength of the frequency of words associated with one word list to another:

1. Chi Square, a statistical test to determine if there is a statistical difference between the frequency of words used between two corpora;
2. Log-likelihood, a statistical test that, "reveals those words that characterise a set of texts by appearing more frequently (or indeed less frequently) than would be expected" given the norms apparent in the language alone" (Anderson and Corbett 38).

For my analysis, I chose to use log-likelihood because I want to discover those words more likely to occur in regulated nuclear language, rather than just those that are occurring more frequently in order to compile a list of the top 20 key terms in the regulated nuclear power industry variety of language. The log-likelihood statistic is most often associated with model fit analyses, and this makes sense when we are trying to determine which words from the nuclear reference corpus do not "fit" with regard to their frequencies (more or less) in comparison to the Baker Brown Corpus. In the following chapter I will not only identify these 20 key terms, but I will also present my findings for their constructed meanings for the industry as a whole through analysis of their collocates.

CHAPTER 4

LEXICAL PROFILES OF REGULATED NUCLEAR INDUSTRY KEY TERMS

Using WordSmith Tools, I generated word frequency lists for both Baker's corpus and the reference corpus. Next, I used the software to statistically compare the two corpora to find out which words were statistically linked to the reference corpus in comparison to the Baker Corpus of American English: the key terms for regulated nuclear industry language in use. The result was over 2,000 words that were more frequently used (log-likelihood p=0.0001) in the nuclear reference corpus than Baker's corpus of American English. Taking a list of over 2,000 key words and reducing it to 20 key terms was a process informed by a notion of *term* similar to the one used by Kemp-Dynin:

> Not only…words which are clearly used in [for example] the field of business, such as *consumer*, but also to those which take on a 'business sense' in the corpora, e.g., *increase* as in *increase in sales.* This is what Sinclair (1996a, 102) refers to as 'quasi-terms.' (63)

In order to make this determination, I generated a concordance for each word in WordSmith Tools to verify that it possessed a "regulated nuclear sense" until I reached 20 key terms (Table 4.1).

Then, I re-ran the concordances on each of these 20 terms using the "stop list" of function words that was also used to determine keyness. As I performed this task, I limited the frequency for collocations at a minimum of five occurrences and then

Table 4.1: Top 20 Key Terms from the Nuclear Reference Corpus

Key word	Keyness	Freq.
NRC	6554.320801	30454
safety	4217.895508	24259
reactor	4184.859863	19611
plant	4061.947266	22366
inspection	3932.327637	18956
unit	3893.398193	20170
licensee	3592.829834	16660
system	3349.659912	26065
power	3248.645264	24747
nuclear	2641.138184	16731
fuel	2567.255615	13988
license	2450.459473	12035
containment	2348.273926	11139
operating	2334.495361	12775
evaluation	2284.026611	11760
staff	2250.520752	14860
technical	2210.776123	12022
emergency	2090.845947	11814
pressure	2042.388794	13536
inspectors	1979.101563	9369

generated a table of collocates for each term with a span of four words to the right and left of the node, or term being investigated, which is the most commonly used span size that allows room to collect words separated by messages (Sinclair 1972, Clear 1993, Kemp-Dynin 2005). I recorded the 20 most frequent collocates for each term that also had a statistically significant probability of occurring with the node. Self collocates, or the key term co-occurring with itself, were not included as part of the analysis due to time constraints resulting from the process of determining if these instances are true self-collocations or the key term appearing in two separate sentences but within the designated span.

In order to determine which collocates have a statistically significant frequency of co-occurrence with each key term I used the mutual information (MI) score. This particular statistical tool measures the "semantic bond" between a term and its collocate. This is done by comparing the term and its collocates' joint probabilities with the probability of them occurring due to chance. When an MI score is higher than three, it suggests "a strong bond" exists between the words (Anderson and Corbett 34). A statistically strong bond is needed between a node word and its collocates in order to develop a lexical profile that provides context for its meaning. In those instances where collocates have a frequency of co-occurrence less than 10, the MI is discounted and the word is not included in the analysis.

The creation of a lexical profile should be methodical and comprehensive in order to "summarize and present information in a coherent and systematic manner, so as to facilitate comparisons and the discovery of significant patterns" (Stubbs 84). Stubbs describes a process for creating a lexical profile. First, the 20 most frequent collocates of a key word are analyzed according to there grammatical affiliation, which is known as colligation, and are then organized for grammatical category (semantic preference)(Stubbs 87). Then the discourse prosody, or the meta-meaning that results from the most frequent use of a word is established for the lexical item. Stubbs demonstrates this characteristic with his analysis of *undergo* being inherently negative due its association with involuntary, serious, unpleasant events, like surgery and medical tests (Stubbs 89). He also prescribes analyzing the strength of the attraction between the node and its collocate, the position of the collocates in relation to the node, and their distribution in text types to create a more complete profile of a word's meaning by providing probabilities and directions for the descriptions resulting from the first four processes.

For the nuclear reference corpus key terms, I generated a limited lexical profile using Stubbs' methodology. Of the 20 terms identified for analysis, there are 17 nouns, one verb, and two adjectives. The most significant key term, *NRC*, can be grouped with *licensee* and *staff* because they represent those individuals/parties playing major roles in the language modeled by the nuclear reference corpus. Sublists can also be made out of those terms that concern evaluation *(evaluation, inspection, inspectors, license)*, term having to do with things nuclear *(nuclear, fuel, reactor, power)*, as well as plant operations *(plant, unit, system, technical, pressure, operating)*. Next, I identified the words that most frequently co-occurred with each of the 20 terms and the strength of their semantic bonds with the node.

When the collocates for each term are plotted against the total number of co-occurrences, a pattern manifests that has already been observed in this dissertation: an A-curve. Figure 4.1 is a line graph for *NRC* where the collocates in their rank order are plotted on the x-axis and the frequency of their co-occurrence with the term are on the y-axis. The resulting pattern is quite similar to the one noted in Chapter 3 when the frequency of document types for the pilot sample of the reference corpus was graphed in the same way. An A-curve was found for every one of the 20 key terms when the collocates frequencies were plotted. However, one thing that was noticed during this process was that the slope of the curves, or the rank/frequency proportion, was different for each one of the terms (Table 4.2).

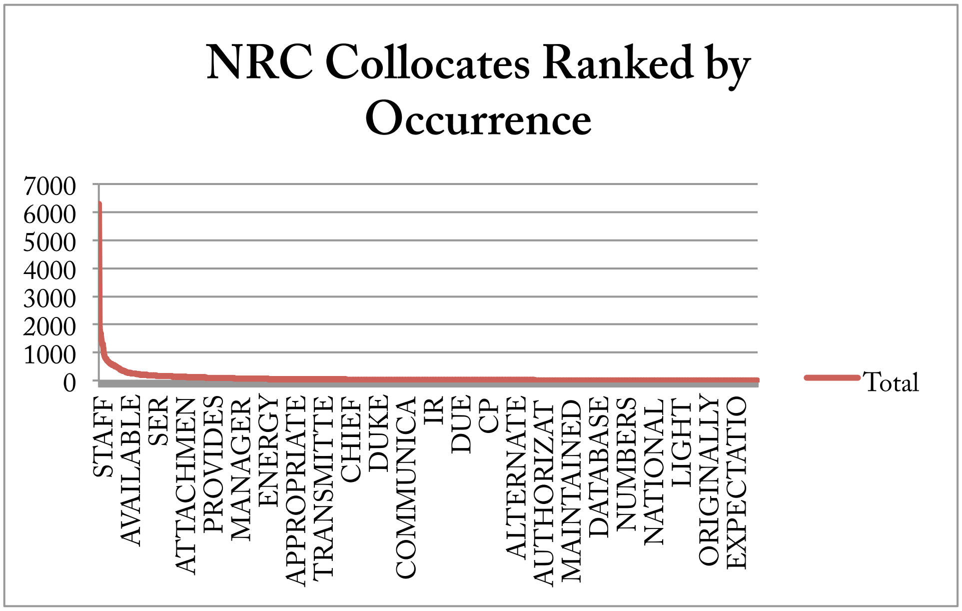

Figure 4.1 Collocates with *NRC* Ranked by Occurrence

When creating lexical profiles for specific words, Stubbs explained that he did so using the 20 most frequently occurring collocates for each one (2002, 219). However, he did not say what percentage of the A-curve his 20 collocates represented in comparison to the total number present for each word, as well as the total number of occurrences they represented. While analyzing 20 collocates is quite manageable for the scope of this dissertation, it should be noted that analysis of them is based only on the top-most level of the A-curve due to the high number of collocates and occurrences. As can be seen in Table 4.3, the percentage of total occurrences demonstrated through an analysis of the 20 most frequently occurring collocates with each term range from 20.02% to 47.02%.

Table 4.2: Rank/Frequency Proportions for 20 Industry Terms

Term	Total Collocates (Types)	Total Occurrences (Tokens)	Token/Type Ratio
NRC	1268	102280	76/24
safety	611	70418	77/23
reactor	698	65969	77/23
plant	811	61399	75/25
inspection	931	66081	75/25
unit	650	44136	72/28
licensee	583	40749	73/27
system	1590	103806	75/25
power	1400	99478	75/25
nuclear	585	62853	77/23
fuel	624	43355	75/25
license	596	39278	76/24
containment	551	33935	73/27
operating	852	46040	73/27
evaluation	492	28330	72/28
staff	491	37168	76/24
technical	397	35335	77/23
emergency	593	38240	74/26
pressure	918	54373	74/26
inspectors	401	28904	75/25

Through analysis of these collocates I generated an Observed Meaning for each term to describe its characteristic usage and discourse function. These industry-level summary statements will serve as the cornerstone for comparison in Chapters 5 and 6 regarding the variation in the Observed Meaning of these key terms due to geography and organizational affiliation. The following lexical profiles of the collocations for each of the 20 key nuclear terms from the reference corpus will proceed in order of keyness, beginning with the word found most likely to occur: *NRC*.

Table 4.3: Representation of Top 20 Collocates and Their Occurrences

Term	Top 20 Collocates' Total Occurrences	Percentage Represented by Top 20 Collocates	Percentage Represented by Top 20 Collocates' Occurrences
NRC	27065	1.58%	26.46%
safety	27003	3.27%	38.35%
reactor	25287	2.87%	38.33%
plant	17755	2.47%	28.92%
inspection	18862	2.15%	28.54%
unit	11689	3.08%	26.48%
licensee	10134	3.43%	24.87%
system	20785	1.26%	20.02%
power	24947	1.43%	25.08%
nuclear	27687	3.42%	44.05%
fuel	16381	3.21%	37.78%
license	16903	3.36%	43.03%
containment	10551	3.63%	31.09%
operating	14394	2.35%	31.26%
evaluation	8669	4.07%	30.60%
staff	17335	4.07%	46.64%
technical	16613	5.04%	47.02%
emergency	12230	3.37%	31.98%
pressure	16658	2.18%	30.63%
inspectors	13307	4.99%	46.04%

NRC

The most statistically significant key word for the nuclear reference corpus is *NRC*, the acronym for the Nuclear Regulatory Commission. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.4. There are five of the 20 nuclear industry key terms from Table 4.1 that are frequent collocates with *NRC*: *nuclear, staff, inspection, evaluation,* and *safety.* When we combine these key terms with the other collocates, we see that most of the words on the list concern things that are

Table 4.4: The 20 Most Frequent Collocates with *NRC*

Word	With	Relation	Total
staff	NRC	20.84883118	6288
regulatory	NRC	19.87118721	1915
approved	NRC	20.72838783	1703
nuclear	NRC	18.76066208	1665
commission	NRC	20.06306267	1630
inspection	NRC	32.57642746	1435
review	NRC	19.02774239	1307
site	NRC	32.43832016	1304
public	NRC	19.16097069	1129
letter	NRC	19.18113327	1104
approval	NRC	20.59997368	933
information	NRC	31.81068993	844
staff's	NRC	21.04694748	800
web	NRC	22.04581451	779
ADAMS	NRC	20.21037292	772
evaluation	NRC	18.10520935	743
will	NRC	16.74832726	717
dated	NRC	18.74554253	686
reviewed	NRC	18.51641464	676
response	NRC	17.96180153	635

information-related (*review, site, letter, information, web, ADAMS, response, request)*. Other words have to do with evaluating *(inspection, approval, approved, evaluation, safety)*.

Most of the information words come together to create clusters indicating resources the *NRC* provides to the public: *NRC PUBLIC DOCUMENT ROOM (ADAMS)* (476) or *NRC WEB SITE* (686). We are also made aware of exactly who serves the public on behalf of the *NRC*: *THE NRC STAFF* (5,294). If we refer back to Table 5.2, we can see that there are two verbs characterizing what the *NRC* does. When we look at the concordance we find rather frequently occurring clusters with these verbs, i.e. *APPROVED BY THE NRC STAFF* (604). While the *NRC* may have *reviewed* things for the licensees, we learn that the *NRC* provides information about the regulation of nuclear power to the *public* for them to evaluate through the recurrence of clusters like *BY THE NRC FOR PUBLIC INSPECTION* (392). Moreover, these clusters of words formed from the 20 most frequent co-occurring words with *NRC* are occurring a significant amount in the corpus, creating clusters of formulaic—or prefabricated—language.

This lexical profile results in the following Observed Meaning for the term *NRC*: **the *NRC* is composed of staff who work with information for evaluation purposes and provide information to the public for their evaluation.**

SAFETY

Although *safety* is a collocate of *NRC*, it is a key term for the nuclear reference corpus as well. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.5. There are four of the industry terms noted as being frequent collocates with *safety*: *evaluation, system, plant, and reactor*. *Plant, reactor,* and *system* join *injection, systems,* and *valves* to create a group of five nouns concerning plant components.

The most frequently occurring clusters indicate that *safety* in the nuclear industry is a social concern: *HEALTH AND SAFETY* (1,936), *HEALTH AND SAFETY OF THE PUBLIC* (1,347). However, *safety* is not solely a government concern, it is of importance to the plant and its equipment: *ON PLANT SAFETY* (101), *REACTOR PRESSURIZER SAFETY VALVES* (95). Also, *safety* is something analyzed, evaluated,

Table 4.5: The 20 Most Frequent Collocates with *Safety*

Word	With	Relation	Total
related	safety	7.119279861	2577
significance	safety	7.061628342	2337
public	safety	6.477663517	2082
analysis	safety	5.792946815	2052
health	safety	7.994673252	2004
evaluation	safety	5.838598251	1829
report	safety	18.15271568	1584
low	safety	6.424854755	1539
function	safety	6.451495647	1351
system	safety	18.85539627	1289
final	safety	6.89980793	996
injection	safety	7.253810883	976
level	safety	5.107435226	915
plant	safety	3.803960323	849
quality	safety	6.469167233	836
systems	safety	15.88266945	821
margin	safety	6.754354	806
updated	safety	7.384805202	749
reactor	safety	3.787545681	736
valves	safety	5.331515789	675

significant to *safety* and can be measured: *MARGIN OF SAFETY (*617), *LEVEL OF QUALITY AND SAFETY* (717), *ACCEPTABLE LEVEL OF SAFETY* (445), *THE SAFETY SIGNIFICANCE* (344), *THE SAFETY RELATED* (320).

This lexical profile results in the following Observed Meaning for the term *safety*: **the focus of *safety* is on the plant, its equipment, and how it is significant to the public. It is evaluated and measured because it is a social concern.**

REACTOR

Reactor is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.6. There are seven of the 20 industry key terms in Table 4.6: *system, pressure, power, safety, nuclear, unit,* and *fuel.*

Table 4.6: The 20 Most Frequent Collocates with *Reactor*

Word	With	Relation	Total
coolant	reactor	20.68819618	3712
vessel	reactor	8.10175705	3436
system	reactor	20.25909996	2757
pressure	reactor	19.79455757	1998
trip	reactor	7.313016891	1729
water	reactor	19.33999062	1458
power	reactor	16.95640373	1397
core	reactor	5.795229435	1105
building	reactor	7.12271452	1019
head	reactor	6.591741562	822
level	reactor	5.247320175	815
safety	reactor	3.783619881	734
nuclear	reactor	4.142054558	649
pump	reactor	5.0218153	591
unit	reactor	3.620404959	545
protection	reactor	5.101470947	528
boundary	reactor	6.420070171	517
operator	reactor	5.75860548	506
RCS	reactor	5.818101406	504
fuel	reactor	3.91940093	465

These key terms combine with the other collocates to create sub-lists for *reactor.* One group concerns resources needed by the *reactor (coolant, water, power, RCS, fuel),* and another group has to do with physical components *(vessel, core, building, head, pump).*

Some of the most frequent clusters of words from the reference corpus for *reactor* often co-occur together to create clusters like *REACTOR COOLANT SYSTEM RCS* (1,555) and *REACTOR COOLANT PRESSURE* (438). However, unlike *NRC* these clusters are not formulaic. They provide semantic information about *reactor*. We can instead classify these collocates as creating cluster terms about components found at many licensee facilities throughout the industry.

There are other clusters that also help us to understand the meanings associated with *reactor*. First of all, the *reactors* being used in this industry are *nuclear*: *OF NUCLEAR REACTOR* (425). Some clusters indicate that *reactor* is composed of many parts: *THE REACTOR VESSEL* (1,317), *THE REACTOR BUILDING* (434), *THE REACTOR COOLANT* (1,313). These components have to be measured—*REACTOR CORE POWER LEVELS* (103), or *THE REACTOR COOLANT PRESSURE* (209)—and are even characterized by those qualities needing measurement, *REACTOR PRESSURE VESSEL* (631), *PRESSURIZED WATER REACTOR* (240), or *BOILING WATER REACTOR* (240). Also, the reactor needs certain resources: *REACTOR VESSEL WATER LEVEL* (205), *THE REACTOR COOLANT* (1313), or *FUEL IN THE REACTOR* (53). There are many systems involving the *reactor* and its components: *REACTOR COOLANT SYSTEM* (1,555). Moreover, some of these systems have to do with safety: *REACTOR PROTECTION SYSTEM* (380), *REACTOR COOLANT PRESSURE BOUNDARY* (414). We also learn that people manage the *reactor* and its *systems*: i.e. *SENIOR REACTOR OPERATOR* (220).

This lexical profile results in the following Observed Meaning for the term *reactor*: **a reactor operated in this industry is nuclear and is characterized by its resources**

and qualities needing measurement. It is composed of many parts that work systematically to maintain safety, i.e. at certain conditions initiating a reactor *trip* (shutdown).

PLANT

Plant is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.7. Six of the collocates in this table are industry key terms from Table 4.1: *nuclear, power, unit, safety, operating,* and *system.*

Table 4.7: The 20 Most Frequent Collocates with *Plant*

Word	With	Relation	Total
nuclear	plant	6.299460411	3302
power	plant	17.45947456	2258
specific	plant	6.592871666	1651
operation	plant	5.125110149	994
unit	plant	4.214092731	938
conditions	plant	5.062491894	893
safety	plant	3.803960323	849
operating	plant	17.15627098	732
operations	plant	5.635857105	611
procedures	plant	5.111435413	584
systems	plant	15.49356937	578
design	plant	4.178104401	561
shutdown	plant	17.7227211	542
performance	plant	4.420265675	519
units	plant	14.84723759	517
equipment	plant	17.53235626	475
personnel	plant	4.878104687	457
modifications	plant	6.590603352	454
risk	plant	4.754329205	441
system	plant	17.28081703	399

These key terms come together with the remaining 14 collocates to show a concern for daily operations *(operation, conditions, operations, procedures, shutdown, performance, equipment, personnel)*, as well as a things that work together or that form larger entities *(unit, units, system, systems)*. I would like to take a moment to ponder the presence of *personnel*. This word is the only collocate having to do with people, and we see it present in this list rather than *staff* as we saw with *NRC*. This indicates the possibility that *staff* is a term used mainly to refer to *NRC* employees and *personnel* is a designation for plant-site employees.

The most frequent clusters of words for *plant* are related to plant identification: *NUCLEAR POWER PLANT* (1,583), *NUCLEAR GENERATING PLANT* (201). A *plant* is made up *units*, and discourse centers on an individual *unit*: *NUCLEAR PLANT UNIT* (195), *PLANT PERFORMANCE FOR UNIT* (125). Also, a *plant* is operated—*OF PLANT OPERATION* (119) or *DURING PLANT OPERATION* (102)—and it can even be shutdown: *SHUTDOWN OF THE PLANT* (73). Issues related to the *plant* are said to be *plant specific: THE PLANT SPECIFIC* (308), *PLANT SPECIFIC OPERATING EXPERIENCE* (73), *PLANT SPECIFIC RISK* (53).

This lexical profile results in the following Observed Meaning for the term *plant*: **a nuclear *plant* is operated by personnel to generate power and is concerned with performance, modifications, safety, and risk. Issues resulting from the plant are specific to it.**

INSPECTION

Inspection is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.8, and you will notice that there is only one industry key term in this table: *NRC*.

Table 4.8: The 20 Most Frequent Collocates with *Inspection*

Word	With	Relation	Total
scope	inspection	21.911726	2416
inspectors	inspection	31.33751869	1824
inservice	inspection	22.40297508	1452
NRC	inspection	32.57239914	1431
program	inspection	33.22658157	1126
reviewed	inspection	20.23188019	1110
public	inspection	20.13127708	1106
results	inspection	33.06973267	1010
report	inspection	31.96764183	941
interval	inspection	20.89180756	718
period	inspection	19.92710495	713
performed	inspection	18.92003822	646
available	inspection	31.35168648	614
visual	inspection	21.10973549	595
One	inspection	18.80801582	571
findings	inspection	32.2240181	562
will	inspection	17.38920212	559
conducted	inspection	20.01841545	524
activities	inspection	19.278862	488
team	inspection	20.41716576	456

Many of the words, all of them nouns, have to do with predetermined methods *(scope, program)*. Another group has to do with results *(results, findings, report)*, and how they are uncovered *(reviewed, performed conducted)*.

While *inservice, available,* and *visual* did not fit into the sub-lists for this term, two of them formed clusters with *inspection* that indicate specific types of inspections: *THE INSERVICE INSPECTION* (229), *VISUAL INSPECTION OF* (220). *Inspection* is also frequently shown to be an action performed by the public on information provided by the NRC: *AVAILABLE ELECTRONICALLY FOR PUBLIC INSPECTION IN THE NRC PUBLIC DOCUMENT ROOM* (401). The most frequent clusters from the reference corpus indicate that *inspection* is a quality of its methods, and results: *A INSPECTION SCOPE* (2,091), *INSERVICE INSPECTION PROGRAM* (213). It also should be noted that the clusters mentioned thus far for *inspection*, as well as others like *THE INSPECTORS REVIEWED* (633), *MATERIALS EXAMINED DURING THE INSPECTION SHOULD BE* (87), and *PRESENTED THE INSPECTION RESULTS TO* (210) all appear to be prefabricated phrases.

This lexical profile results in the following Observed Meaning for the term *inspection*: **there are different types of *inspection*, and each one has predetermined methods and plans that inform actions taken to review information and generate evaluations. For the NRC and licensees, these happen at regular intervals of time.**

UNIT

Unit is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.9. There are six industry key terms listed in this table *power, nuclear, plant, reactor, operating,* and *system*. Most of the words in this list relate to nuclear power plants and their operation

Table 4.9: The 20 Most Frequent Collocates with *Unit*

Word	With	Relation	Total
station	unit	6.153031826	1266
power	unit	16.63515854	1150
nuclear	unit	4.836247444	1080
plant	unit	4.214092731	938
outage	unit	5.737971306	627
refueling	unit	5.842178345	626
cycle	unit	16.95457458	574
steam	unit	4.798697472	571
Lucie	unit	7.761995792	568
reactor	unit	3.620404959	545
Point	unit	5.59649992	431
operating	unit	16.52098656	425
mode	unit	5.362751961	412
One	unit	4.016082287	406
Millstone	unit	7.714501858	402
shutdown	unit	17.38585854	387
system	unit	17.20752144	342
operation	unit	3.66140008	325
B	unit	14.0842886	314
containment	unit	3.575154781	300

(station, power, nuclear plant, outage, refueling, cycle, steam, reactor, operating, mode, shutdown, system, operation, containment). It should also be noted that there are three proper nouns in this table that have to do with the names of specific plants: *Lucie, Point,* and *Millstone*. The co-occurrence of these three facility names with *unit* indicates that these licensees are in some way significant to the ongoing dialogue of this industry, as they are frequently mentioned to a point of statistical significance.

The most frequent clusters from the reference corpus indicate that *unit* is a way of identifying separate reactors that are licensed for operation at a facility: *POWER STATION UNIT* (304), *NUCLEAR PLANT UNIT* (201), *NUCLEAR STATION UNIT*

(195), and *UNIT # REACTOR* (341). We also learn that the operation of the individual units is done systematically: *DURING THE UNIT OUTAGE* (120), or *OPERATING CYCLE FOR UNIT* (21).

This lexical profile results in the following Observed Meaning for the term *unit*: **a *unit* is a way of identifying individual nuclear reactors at a power generating facility. A *unit* operates in cycles, shutting down for refueling or other outages.**

LICENSEE

Licensee is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.10.

Table 4.10: The 20 Most Frequent Collocates with *Licensee*

Word	With	Relation	Total
identified	licensee	5.460438728	1040
stated	licensee	6.669305325	674
staff	licensee	4.59004879	668
proposed	licensee	4.36514616	584
inspectors	licensee	15.65794754	579
NRC	licensee	17.21269608	567
event	licensee	4.729462624	504
shall	licensee	4.628351212	501
provided	licensee	4.873188019	494
requested	licensee	5.870744705	461
company	licensee	6.077727318	444
failed	licensee	6.746170521	443
submitted	licensee	5.975144386	441
report	licensee	16.79000473	423
also	licensee	4.442747593	413
determined	licensee	4.831207275	397
reviewed	licensee	4.717143536	395
information	licensee	17.64669228	383
performed	licensee	4.091928959	370
changes	licensee	4.075873375	353

There are three industry key terms that are collocates for *licensee*: *staff, inspectors,* and *NRC.* Most of the words on this list relate to the exchange of information between *licensees* and the *NRC.* We can divide this group into three sets: one concerns **analyzing** information *(identified, determined, reviewed),* the second has to do with the **dissemination** of information *(stated, proposed, provided, requested, submitted)*, and the third centers on the **evaluation** of information in action, or synthesis, by the *NRC (failed, performed)*.

The most frequent clusters from the reference corpus indicate that the *licensee* is the source of *information*: *LICENSEE EVENT REPORT* (283), *LICENSEE IDENTIFIED VIOLATIONS* (251), *THE LICENSEE PROPOSED* (231), *THE LICENSEE STATED* (522), *THE LICENSEE DETERMINED THAT* (197), *INFORMATION PROVIDED BY LICENSEE* (61). When distributing this information, the *licensee* is not always successful in the way it has synthesized previous information: *THE LICENSEE FAILED TO* (402). This information is part of a circuit of communication with the *NRC,* its *staff,* including *inspectors: INSPECTORS ASKED THE LICENSEE* (98), or *STAFF REQUESTED THAT THE LICENSEE* (33). A *licensee* also receives instruction from the *NRC* in the form of imperatives through the word *shall: THE LICENSEE SHALL* (417), *LICENSEE SHALL OPERATE* (157).

This lexical profile results in the following Observed Meaning for the term *licensee*: **a *licensee* reports to the *NRC*. Its role in this relationship is to analyze information, provide it to the *NRC*, and receive evaluations and assessments.**

SYSTEM

The collocational data for the 20 most frequently co-occurring words with *system* are illustrated in Table 4.11.

Table 4.11: The 20 Most Frequent Collocates with *System*

Word	With	Relation	Total
reactor	system	20.25700569	2753
coolant	system	33.89210892	1786
cooling	system	21.1265564	1433
water	system	33.56633759	1425
safety	system	18.84641457	1281
control	system	19.29047585	1179
pressure	system	33.22273254	1123
component	system	20.94678688	1062
emergency	system	19.59342766	1047
protection	system	20.1398201	906
ADAMS	system	21.29210854	817
containment	system	19.31159592	812
core	system	19.55579948	782
power	system	30.23323059	707
ventilation	system	21.90181541	705
management	system	19.94033432	664
service	system	19.73131561	631
leakage	system	32.22144699	561
RCS	system	20.22690201	559
test	system	32.1981163	552

Six of the collocates in this table are industry key terms: *reactor, safety, pressure, emergency, containment,* and *power*. Most of the collocates on this list are nouns that concern the reactor and its physical components *(reactor, core, component, coolant, water, pressure, feedwater, ventilation, leakage, RCS)*, four come together regarding safety *(safety, emergency, protection, containment)*, and four nouns have to do with management and control *(control, management, test, service)*. The most frequent clusters from the reference corpus indicate that many of the nouns help us to understand the types of *systems* in place in the nuclear industry: *REACTOR COOLANT SYSTEM* (1,556), *AGENCYWIDE DOCUMENTS ACCESS AND MANAGEMENT SYSTEM* (436), or *CONTAINMENT SPRAY SYSTEM* (219). The collocates having to do with safety indicate the protective purposes of some of the systems: *EMERGENCY CORE COOLING SYSTEM* (384), *REACTOR PROTECTION SYSTEM* (384), *FIRE PROTECTION SYSTEM* (124). We also learn that some of these systems function to manage and control water and the release of gases: *SYSTEM LEAKAGE TEST* (147), *LEAKAGE DETECTION SYSTEM* (30). There is also concern for when these *systems* are not performing properly: i.e. *COOLANT SYSTEM LEAKAGE* (66).

This lexical profile results in the following Observed Meaning for the term *system*: **There are many different kinds of *systems*: some involve controlling and managing plant components, others safety, and all of these systems are managed to maintain control over plant operations.**

POWER

Power is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are provided in Table 4.12. There are five of the key terms being analyzed in this chapter that are also collocates with *power: nuclear*, *plant*, *reactor*, *unit,* and *system.* All four of these key terms concern the physical plant *(plant, station, unit, plants)*. The most frequent clusters from the reference corpus

Table 4.12: The 20 Most Frequent Collocates with *Power*

Word	With	Relation	Texts	Total
nuclear	power	18.76222038	1891	4167
plant	power	17.45883369	1211	2257
station	power	18.46882248	870	1600
reactor	power	16.95640373	679	1397
offsite	power	19.29302406	407	1304
thermal	power	19.19017601	431	1290
unit	power	16.63264656	646	1148
plants	power	25.25384712	682	1143
loss	power	18.12469482	436	1105
level	power	17.61082649	468	1095
operation	power	17.35038185	657	1064
uprate	power	20.57789803	293	1003
full	power	19.0809536	475	999
core	power	17.50854301	367	946
percent	power	18.49900818	320	822
company	power	18.66496468	591	820
rated	power	20.13314438	303	744
system	power	30.23323059	314	707
electric	power	18.74476814	384	672
Light	power	20.08606529	465	664

result from these words grouping together to describe what type of *power* fuels the plants and reactors in this industry: *NUCLEAR POWER PLANT(S)* (2,630), *NUCLEAR POWER REACTOR(S)* (413). *Power* is also used as a branding tool to describe what some of these companies provide: *POWER AND LIGHT COMPANY* (213).

Some of the other most frequently used clusters show us that *power* is something that is measured: *RATED THERMAL POWER* (431), *AT 100 PERCENT POWER* (219), *EXTENDED POWER UPRATE* (241), *POWER LEVEL OF* (160). Furthermore, *power* is measured both within the plant and outside of it: *LOSS OF OFFSITE POWER* (486), *CORE THERMAL POWER* (142).

This lexical profile results in the following Observed Meaning for the term *power*: ***power* comes in different types, is measured in loss from its fullest state, is both a resource and a product at the plant and offsite. *Power* is also linked to identity in this industry: used by corporations as a branding tool to identify themselves and what they provide to their customers.**

NUCLEAR

The collocational data for the 20 most frequently co-occurring words with *nuclear* are listed in Table 4.13, and seven of these are key terms: *power*, *plant*, *unit*, *operating*, *reactor*, *safety*, and *staff*.

Table 4.13: The 20 Most Frequent Collocates with *Nuclear*

Word	With	Relation	Total
power	nuclear	18.76256752	4168
commission	nuclear	8.077274323	3362
plant	nuclear	6.299023151	3301
regulatory	nuclear	7.597817898	3236
station	nuclear	7.17802906	2137
NRC	nuclear	18.75632286	1660
plants	nuclear	13.69981289	1272
unit	nuclear	4.836247444	1080
units	nuclear	15.97309971	844
operating	nuclear	17.5926857	741
company	nuclear	6.802697182	737
generating	nuclear	7.77021265	721
reactor	nuclear	4.139830112	648
LLC	nuclear	7.644014835	595
safety	nuclear	3.685415983	585
Point	nuclear	6.16449213	530
energy	nuclear	5.90221405	529
staff	nuclear	4.236382484	525
Entergy	nuclear	7.137629986	516
office	nuclear	7.14862299	500

Seven nouns and two verbs have to do with identifying power and the production of energy *(power, energy, plant, station, plants, unit, units, operating, reactor, safety)*. The rest of the collocates have to do either with the NRC *(commission, regulatory, NRC, staff, office)* or the licensees *(company, LLC)*. There are also two proper nouns present, *Point* and *Entergy*, which directly relate to specific licensees whose clusters of identification involve *nuclear,* and who are referenced a statistically significant amount in the nuclear reference corpus: i.e. *Entergy Nuclear Operations* (203).

The most frequent cluster from the reference corpus is identifying the NRC as the *NUCLEAR REGULATORY COMMISSION* (3,021). Furthermore, the NRC frequently uses *nuclear* to provide specificity to itself and what it seeks to do: *NUCLEAR REACTOR REGULATION* (404), *OFFICE OF NUCLEAR REACTOR REGULATION* (383). However, the strongest semantic bond with *nuclear* is *power*, indicating that *nuclear* will most likely co-occur with *power* in the discourse from the nuclear reference corpus. This observation makes sense when we encounter the same clusters we saw for power: like *NUCLEAR POWER PLANT(S)* (2,630), and *NUCLEAR POWER REACTOR(S)* (413). The word *nuclear* is also used frequently as a branding tool for companies to identify themselves as an owner and/or operator of *nuclear power plants*: *NINE MILE POINT NUCLEAR* (146), or *ENTERGY NUCLEAR OPERATIONS* (203).

This lexical profile results in the following Observed Meaning for the term *nuclear*: ***nuclear* is a type of power that is regulated by the government for safety in its use, which in this industry is the commercial generation of electricity. *Nuclear* is also used by**

companies in their branding to identify themselves as owners and/or operators within this industry.

FUEL

Fuel is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are illustrated in Table 4.14.

Table 4.14: The 20 Most Frequent Collocates with *Fuel*

Word	With	Relation	Total
spent	fuel	9.345326424	2846
pool	fuel	8.490381241	1631
storage	fuel	8.156466484	1552
assemblies	fuel	8.720793724	1204
oil	fuel	8.233034134	1059
handling	fuel	8.722023964	761
irradiated	fuel	9.111623764	749
assembly	fuel	7.871488094	727
design	fuel	5.119969845	674
movement	fuel	8.825162888	666
cycle	fuel	17.40780258	545
rods	fuel	7.507568359	521
core	fuel	5.18130064	515
rod	fuel	6.285531044	488
cladding	fuel	7.891842365	482
reactor	fuel	3.91940093	465
accident	fuel	5.171433926	437
system	fuel	17.8175354	362
containment	fuel	4.337881565	353
nuclear	fuel	3.713718176	344

There are four industry key terms listed in Table 4.14: *reactor*, *system*, *containment*, and *nuclear*. All of these terms have to do with the *nuclear reactor*, and they are joined by *pool*, *core, storage, assemblies, assembly, rods, rod, cladding, spent,* and *irradiated*. In this instance, *spent* is included with the other *nuclear* collocates due to its definition of being a depletion to the point that a nuclear reaction can no longer be sustained ("Spent (depleted or used) nuclear fuel").

We are able to learn from the clusters that while the adjectives are describing types of *fuel*, *fuel* itself helps to explain the places within the system, as well as providing specificity to components: *SPENT FUEL POOL* (1,239), *SPENT FUEL STORAGE* (615), *FUEL STORAGE RACKS* (153). Moreover, *fuel* is not a static resource in this industry. It is moved and handled as part of a cycle: *MOVEMENT OF IRRADIATED FUEL* (194), *THE FUEL HANDLING* (149), *URANIUM FUEL CYCLE* (92), *IMPACTS FROM THE FUEL CYCLE* (23). The clusters also help us to understand that while movement words do frequently co-occur with *fuel,* it is that movement which brings the relatively frequent use of the word *accident* (437 instances) into context: *FUEL HANDLING ACCIDENT* (366).

This lexical profile results in the following Observed Meaning for the term *fuel*: ***fuel*** **mainly refers to that which is nuclear: it is dangerous, focus is placed on the storage and containment of it, and moving it can lead to accidents. Types of** ***fuel*** **help to identify specific components and places within the reactor system.**

LICENSE

The collocational data for the 20 most frequently co-occurring words with *license* are illustrated in Table 4.15.

Table 4.15: The 20 Most Frequent Collocates with *License*

Word	Relation	Total
renewal	22.87748718	3277
amendment	21.202425	2499
operating	33.14080048	2122
facility	20.69419289	1091
renewed	22.78870392	815
application	20.18762016	803
request	19.89316177	791
condition	19.37022781	769
scope	20.0932827	685
proposed	18.40643692	591
within	18.26177216	415
nuclear	17.72469139	406
incorporated	21.04856491	390
conditions	18.27504921	379
hereby	21.38882065	335
changes	18.0071373	331
effective	20.26077271	305
attachment	19.02955627	303
amended	20.65359688	299
environmental	18.56560707	297

There are two industry key terms that are also collocates with *license*: *operating* and *nuclear*. Most of the remaining collocates are nouns. Four of the nouns concern documents *(amendment, application, request, attachment)*, four more words—two nouns and two verbs—relate to change *(renewal, changes, renewed, amended)*.

We are able to see that this term is also formulaic like *NRC*, *inspection*, and *power,* through the clusters resulting from *facility* and *operating: FACILITY OPERATINGLICENSE* (978), *RENEWED FACILITY OPERATING LICENSE* (243). From these set phrases, *operating* lets us know that the type of license being described in these documents is not a credential for driving a car, but rather for operating a facility that uses nuclear materials. There are several other formulaic clusters that provide clarification as to the extent of documents surrounding the maintenance of this license: *LICENSE AMENDMENT REQUEST* (690), or *PROPOSED LICENSE RENEWAL APPLICATION* (304). We also learn from other clusters that some of the words that did not fit into sub-lists, namely *hereby,* is evidence of the formality associated with a *license* and provides us more proof of its formulaic status: *AMENDMENTS TO THIS LICENSE ARE HEREBY INCORPORATED* (283), *PROPOSED LICENSE IS EFFECTIVE AS OF* (295).

This lexical profile results in the following Observed Meaning for the term *license*: **a *license* directly relates to the operation of a facility that uses nuclear materials. It is dynamic in nature, and documents are created and used to communicate these changes. It is rooted in formality.**

CONTAINMENT

Containment is the next key term for the nuclear reference corpus. The collocational data for the 20 most frequently co-occurring words with this term are provided in Table 4.16. There are two key terms that are collocates with *containment: system* and *reactor*. These collocates can be organized into the following sub-lists. There are eight nouns concerning physical constructs and components in the plant and how they work together *(spray, sump, valves, fuel, building, reactor, system, program),* and there are six nouns having to do with control and separation *(isolation, leakage, rate, integrity, inside, outside). Pressure* was the one word that did not fit into any of the sub-lists.

Table 4.16: The 20 Most Frequent Collocates with *Containment*

Word	Relation	Total
primary	7.748493195	994
isolation	7.723342896	937
pressure	19.47439575	909
system	19.31159592	812
spray	8.085451126	783
sump	7.840876102	523
integrity	7.066363335	523
secondary	7.576007366	501
leakage	18.61204338	500
inside	7.949046612	482
outside	7.207916737	426
valves	5.769917965	420
atmosphere	8.714278221	380
cooling	5.682217121	358
fuel	4.337881565	353
air	5.867667675	351
rate	5.808701038	341
building	6.316507339	331
reactor	3.704296589	319
program	17.91304588	308

From the clusters formed by the collocates with *containment* we find that *pressure* is used to provide context to the meaning of *containment*: *THE CONTAINMENT PRESSURE* (97), or *THE CONTAINMENT PRESSURE BOUNDARY* (23). We now know that *containment* is an entity that has its own components, like pressure, and several others: *THE CONTAINMENT ATMOSPHERE* (209). Furthermore, these components require measurement: *CONTAINMENT LEAKAGE RATE* (190).

There are several other frequently used clusters that also allow us to understand that *containment* is an entity within the system. It has two ordinal divisions, *primary* and *secondary*: *PRIMARY CONTAINMENT ISOLATION* (148). Whether or not *containment* is exerting the control it should, its *integrity*, is of importance: *INTEGRITY OF THE CONTAINMENT* (32), *STRUCTURAL INTEGRITY OF THE CONTAINMENT* (24), *CONTAINMENT INTEGRITY DURING* (24).

This lexical profile results in the following Observed Meaning for the term *containment*: **c*ontainment* is an entity within the plant that is composed of components and systems that require measurement. This entity functions to control the division between inside and outside to promote isolation, and any breaches of this control are also measured.**

OPERATING

The collocational data for the 20 most frequently co-occurring words with *operating* are listed in Table 4.17, seven of which are key terms from Table 4.1: *license, nuclear, plant, power, unit, system,* and *pressure*. Most of the words in this list relate to the licensees and their business, including the physical components at their facilities *(facility, plant, core, company, industry, system, nuclear, power, pressure)*, while five other lexical items can be made into a group concerning licenses and other documents *(license, licenses, procedures, report, renewed)*.

From the most frequent clusters, we learn which words are the primary nouns being modified by *operating*, even when *operating* is modified by one of the other nouns or adjectives: *FACILITY OPERATING LICENSE* (981), *OPERATING LIMITS REPORT* (324), *MONTH OPERATING CYCLE* (62). We can also see from these clusters that *operating* is a quality that is of importance not only to regulation,

Table 4.17: The 20 Most Frequent Collocates with *Operating*

Word	With	Relation	Total
license	operating	33.14215851	2124
experience	operating	21.31646919	1450
facility	operating	19.88236809	1243
cycle	operating	30.75492668	812
limits	operating	18.96371269	776
nuclear	operating	17.59073639	740
plant	operating	17.15034676	729
licenses	operating	21.31914902	656
core	operating	18.26444626	639
conditions	operating	17.93781471	600
power	operating	28.97955132	593
procedures	operating	18.5380497	575
report	operating	30.2214489	561
renewed	operating	21.2291832	553
normal	operating	19.23706627	539
unit	operating	16.51758766	424
system	operating	30.6820488	386
company	operating	18.80725861	362
industry	operating	19.18876266	318
pressure	operating	30.38421249	314

RENEWED FACILITY OPERATING LICENSE (286), but also to daily functions at the plant, *NORMAL OPERATING CONDITIONS* (125), and even the guidelines in place for protection, *EMERGENCY OPERATING PROCEDURES* (168) and *CORE OPERATING LIMITS* (504). *Operating* is also used by companies, in a similar way to nuclear and power, as a branding tool to identify themselves by what they do: i.e. *SOUTHERN NUCLEAR OPERATING COMPANY* (346).

This lexical profile results in the following Observed Meaning for the term *operating*: ***operating* is a measured quality that extends to the physical components of the**

plant, regulatory licenses and documents, and the limits and conditions the industry follows. It is also used as a branding tool by companies in this industry for identification.

EVALUATION

The collocational data for the 20 most frequently co-occurring words with *evaluation* are illustrated in Table 4.18, and four of its collocates are industry key terms: s*afety*, *NRC*, *technical*, and *staff*.

Table 4.18: The 20 Most Frequent Collocates with *Evaluation*

Word	With	Relation	Total
safety	evaluation	5.842536449	1834
NRC	evaluation	18.09937286	740
report	evaluation	17.91893005	653
performed	evaluation	4.933412552	468
technical	evaluation	4.866350651	462
model	evaluation	6.495110035	427
staff	evaluation	4.323554516	392
cause	evaluation	5.911351204	391
based	evaluation	4.572329998	333
engineering	evaluation	6.481850624	321
staff's	evaluation	7.185271263	316
results	evaluation	17.83477783	308
licensee's	evaluation	5.437344074	302
regulatory	evaluation	4.660776615	297
root	evaluation	7.058977604	260
proposed	evaluation	3.63782382	249
flaw	evaluation	6.410139084	240
review	evaluation	4.048916817	238
dated	evaluation	4.583236217	220
risk	evaluation	4.665297985	218

Eight of the collocates on this list concern different parties involved in nuclear power regulation, as well as their interests *(NRC, staff, staff's, licensee's, regulatory, technical, engineering, safety)*, and all but one of the remaining words deal with information and how it is ascertained and handled *(cause, results, flaw, report, model, review, proposed, based, performed)*. The one word that did not fit into these sub-lists was *dated*.

We learn that ascertained information in this industry often comes from *evaluation* through the cluster *THE RESULTS OF THE EVALUATION* (139). But the types of *evaluation* and their informative functions can be gleaned from *SAFETY EVALUATION REPORT* (374), *A SAFETY EVALUATION* (346), and *TECHNICAL EVALUATION REPORT* (18). We can see from the clusters that oftentimes the focus of an *evaluation* is to uncover the source of a problem: *ROOT CAUSE EVALUATION* (230), or *THE FLAW EVALUATION* (60).

This lexical profile results in the following Observed Meaning for the term *evaluation*: **an *evaluation* is performed to ascertain information involving safety, regulatory, and technical interests of the NRC and licensees to often seek the source of a problem. It is more strongly associated with the NRC than other parties and is communicated formally.**

STAFF

The collocational data for the 20 most frequently co-occurring words with *staff* are provided in Table 4.19. There are three key terms that are recognized in this table: *NRC, licensee,* and *nuclear*. When they are combined with the remaining 17 words, these

Table 4.19: The 20 Most Frequent Collocates with *Staff*

Word	With	Relation	Total
NRC	staff	20.85957336	6335
concludes	staff	8.676509857	1034
reviewed	staff	6.154234409	954
finds	staff	8.998927116	833
review	staff	5.371141434	752
licensee	staff	4.526016235	639
determined	staff	5.596786022	602
proposed	staff	4.542399883	589
commission	staff	5.665117264	561
information	staff	18.3363533	551
nuclear	staff	4.239128113	526
licensee's	staff	5.875413418	517
regulatory	staff	5.094753742	507
therefore	staff	18.14930916	484
will	staff	3.233073235	455
evaluation	staff	4.46038866	431
applicant	staff	6.397253036	419
members	staff	16.92741203	415
requested	staff	5.775790691	385
also	staff	4.352332115	346

collocates can be organized into the following sub-lists. First, there are those parties involved in nuclear power production and regulation *(NRC, commission, licensee, licensee's, applicant, members)*, and we also have a large group of words related to information and what is done with it *(review, information, reviewed, requested, concludes, finds, determined)*.

We can see from Table 4.17 that *staff,* when used, most likely is referring to the *NRC*—which echoes the observation made earlier about this term being associated with individuals working for this regulatory agency, versus *personnel* being more strongly associated with individuals working at a *plant*. Moreover, we can expect that those clusters formed with *staff* and *NRC* will be formulaic phrases. From the most frequently occurring clusters, we learn that the *staff* is responsible for evaluating *information* as a consequence of its interactions with the data from the licensees: *NRC STAFF CONCLUDES THAT* (855), *THEREFORE THE STAFF CONCLUDES* (85), *NRC STAFF REVIEWED THE LICENSEE'S* (145).

This lexical profile results in the following Observed Meaning for the term *staff*: **staff is a term used mostly to refer to those individuals at the NRC who handle information obtained from the licensees, evaluate it, and provide any consequences that result from that information.**

TECHNICAL

The collocational data for the 20 most frequently co-occurring words with *technical* are listed in Table 4.20. There are two key terms that frequently co-occur with *technical* enough to be in the top 20: *evaluation* and *unit*. When put into context with the other collocates for *technical*, we see that many of them form a group about standards and conformity *(specification, specifications, requirements, accordance, standard)*. Another sub-list concerns change and its evaluation *(change, changes, evaluation, review)*, and there is a third group that has to do with information *(support, bases, basis, information)*.

From the most frequent clusters from the nuclear reference corpus, we learn that *technical specifications* are a type of document, and this cluster occurs most often as a semantic unit in the reference corpus. For example, we can find clusters that reveal the fact that *technical specifications* are informative in nature, *TECHNICAL SPECIFICATION BASES* (132), they are meant to provide information about

Table 4.20: The 20 Most Frequent Collocates with *Technical*

Word	With	Relation	Total
specifications	technical	21.88437653	5214
specification	technical	9.016460419	3650
changes	technical	5.706941605	789
requirements	technical	19.0907917	752
accordance	technical	5.321621895	539
proposed	technical	4.579723835	489
evaluation	technical	4.866350651	462
change	technical	4.949527264	456
required	technical	4.166353226	416
environmental	technical	5.466784	407
appendix	technical	5.346675873	398
support	technical	5.645680904	390
bases	technical	6.438448906	378
contained	technical	6.483231544	371
basis	technical	4.987867355	364
standard	technical	18.01193619	356
information	technical	17.80766487	309
review	technical	4.312134743	292
unit	technical	3.421154261	291
section	technical	3.599686146	290

conformity, *STANDARD TECHNICAL SPECIFICATIONS* (261), and they have power over those topics they concern, *TECHNICAL SPECIFICATION REQUIREMENTS* (153), *REQUIRED TECHNICAL SPECIFICATIONS* (76) or *IN ACCORDANCE WITH TECHNICAL SPECIFICATION* (68). Experiences on the unit level may require changes to be proposed: *PROPOSED TECHNICAL SPECIFICATION CHANGE(S)* (80).

This lexical profile results in the following Observed Meaning for the term *technical*: ***technical* most commonly used as part of a formulaic sequence with *specifications*, a document of accepted standards for individual units that can be changed.**

EMERGENCY

The collocational data for the 20 most frequently co-occurring words with *emergency* are illustrated in Table 4.21, two of which are key terms: *system* and *power*.

Table 4.21: The 20 Most Frequent Collocates with *Emergency*

Word	With	Relation	Total
response	emergency	6.249684334	1118
system	emergency	19.59066963	1045
preparedness	emergency	9.541996002	961
diesel	emergency	8.000451088	908
plan	emergency	7.008765697	872
cooling	emergency	6.812228203	831
core	emergency	5.806733131	671
generator	emergency	6.483263493	656
power	emergency	16.25347328	517
site	emergency	18.57260895	516
radiological	emergency	6.557211399	475
general	emergency	6.86485815	454
room	emergency	5.479088783	444
planning	emergency	8.166430473	443
control	emergency	4.262821674	417
procedures	emergency	5.438610554	387
classification	emergency	8.302089691	386
director	emergency	8.008155823	383
area	emergency	18.11597061	376
action	emergency	16.50780106	370

Four of the words on this list have to do with power *(diesel, power, generator, radiological)*, four collocates have to do with locations *(core, site, room, area)*, seven of them concern management and planning *(control, classification, director, preparedness, plan, procedures, planning)*. Looking at the most common clusters involving *emergency* from the nuclear reference corpus we see that this term is also a quality used to modify or describe many of its collocates, which range from mechanical systems, to states of being, and even position titles: *EMERGENCY CORE COOLING* (641), *EMERGENCY DIESEL GENERATOR* (627), *THE EMERGENCY DIRECTOR* (181), *EMERGENCY OPERATING PROCEDURES* (162), *EMERGENCY OPERATIONS FACILITY* (157), and *RADIOLOGICAL EMERGENCY PREPAREDNESS* (75).

This lexical profile results in the following Observed Meaning for the term *emergency*: **e*mergency* is most commonly associated with systems and of a plant that are needed in a radiological *emergency*, i.e. diesel generators. Those entities having been designated as *emergency* are a result of preparedness, they are in place to control, and they result in actions.**

PRESSURE

The collocational data for the 20 most frequently co-occurring words with *pressure* are provided in Table 4.22. There are two key terms on this list that are collocates with *pressure: reactor* and *system.* The remaining collocates can be grouped together in order to uncover the underlying meaning of this term.

Six of the words are related to measuring *(temperature, differential, psig, test, high, low)*, and five nouns concern control *(boundary, containment, code, control, leakage)*. When looking at the most common clusters with *pressure,* we learn that it is a force intrinsic to several different mechanisms in the plant: i.e. *STEAM DOME PRESSURE* (87) or *REACTOR PRESSURE VESSEL* (645). This force and its characteristics are often measured: *REACTOR COOLANT PRESSURE BOUNDARY* (399), *PRESSURE AND*

Table 4.22: The 20 Most Frequent Collocates with *Pressure*

Word	With	Relation	Total
reactor	pressure	19.79599953	2000
vessel	pressure	21.06583023	1400
boundary	pressure	21.98170662	1275
temperature	pressure	20.49952698	1270
system	pressure	33.22529984	1125
high	pressure	20.64486313	1040
coolant	pressure	33.0568161	1001
containment	pressure	19.47598267	910
low	pressure	20.18138695	878
RCS	pressure	20.670	760
code	pressure	31.44273758	654
injection	pressure	21.13009644	605
ASME	pressure	19.60862732	532
differential	pressure	22.62859344	502
boiler	pressure	22.75588608	488
steam	pressure	18.85715294	483
test	pressure	32.00547028	483
psig	pressure	21.73595428	477
control	pressure	17.69075394	389
leakage	pressure	31.6820488	386

TEMPERATURE LIMITS (75), *DIFFERENTIAL PRESSURE ACROSS* (50). These quantified characteristics are then used to describe certain mechanisms and functions: *HIGH PRESSURE COOLANT* (151) or *LOW PRESSURE INJECTION* (71). Context for control of these forces is provided by authorities like the Association for Mechanical Engineers: *ASME PRESSURE VESSEL CODE* (432) or *ASME CODE CLASS* (10).

This lexical profile results in the following Observed Meaning for the term *pressure*: ***pressure* is a force intrinsic to the machines, systems, and their components, at a plant. It is measured, quantified, and evaluated in order to maintain control with regard to certain prescribed standards.**

INSPECTORS

The collocational data for the 20 most frequently co-occurring words with *inspectors* are listed in Table 4.23.

Table 4.23: The 20 Most Frequent Collocates with *Inspectors*

Word	With	Relation	Total
reviewed	inspectors	18.97610474	2789
inspection	inspectors	31.33751869	1824
scope	inspectors	18.5135746	1375
also	inspectors	16.77684975	768
identified	inspectors	16.15565872	621
determined	inspectors	16.90413666	616
observed	inspectors	18.64836502	580
verified	inspectors	18.7377739	578
licensee	inspectors	15.59427071	554
licensee's	inspectors	17.12141991	507
evaluated	inspectors	16.82661247	409
performed	inspectors	15.59214306	386
resident	inspectors	19.21742821	370
selected	inspectors	17.61271667	355
concluded	inspectors	17.38836861	283
conducted	inspectors	16.48217964	271
following	inspectors	15.02200031	265
finding	inspectors	16.64840508	255
results	inspectors	28.4761734	251
plant	inspectors	14.02139378	250

There are three key terms that collocate with *inspectors: inspection, licensee,* and *plant*. It makes sense that these three would co-occur with *inspectors*, as they correspond with what these people do (*inspection),* who they do it to *(licensee)*, and what they are inspecting *(plant)*. Most of the collocates for this term are verbs that can be easily organized into a sub-list that deals with the application, analysis, synthesis, and evaluation of information:

what the inspectors do *(reviewed, observed, identified, selected, verified, determined, performed conducted, concluded, evaluated)*. By analyzing the concordance lines for *inspectors,* we can also learn that how *resident* fits into this picture and that it is regularly used to specify a type of NRC inspector that is assigned in residence at each plant: RESIDENT INSPECTOR (308).

We can see in Table 4.21 that *inspectors* is another key term whose collocates all have high MI scores, and by looking at the clusters for this term we can see that this is another formulaic term: i.e. *INSPECTORS PRESENTED THE INSPECTION RESULTS* (166). Looking at more of the clusters for this term, we see that all of the verbs that collocate with *inspectors* are the actions they perform during an inspection: *THE INSPECTORS REVIEWED* (2,136), *THE INSPECTORS DETERMINED* (568), *THE INSPECTORS OBSERVED* (478), and *THE INSPECTORS IDENTIFIED* (469). Moreover, when they performed these actions they did this for more than one object of analysis: *INSPECTORS ALSO REVIEWED* (320), or *THE INSPECTORS REVIEWED THE FOLLOWING* (131). For *scope*, by looking back to the analysis of this key term we are reminded that it is a planning or procedural aspect of an inspection, and as such makes sense to be a collocate of *inspectors.*

This lexical profile results in the following Observed Meaning for the term *inspectors*: **i*nspectors* are representatives of the NRC, some of whom are in residence at individual plants, who perform inspections to analyze and evaluate the synthesis and application of information at each plant. They produce results that affect the licensee.**

SUMMARY OF KEY TERM PROFILING

This chapter has identified the 20 most statistically significant key terms from the nuclear reference corpus. For each of these terms, the 20 most frequent collocates were identified and used to develop a lexical profile and Observed Meaning for each key term. The strength of the semantic bonds between the terms and their collocates was determined using mutual information scores.

Using the MI scores and cluster frequency patterns, it was determined that five of the key terms have formulaic status: *NRC, inspection, power, license,* and *inspectors*. All of these words concern the NRC and its main interests in regulation. Meanwhile, *system* and *pressure* also exhibited extremely high semantic bonds for all of its collocates, but they were determined to be behaving semantically, as the prevalence of formulaic-esque clauses is a result of components that are used across the industry on the plant level. It should be conceded, however, that although a term was not labeled as formulaic in its status does not mean that it is void of pre-fabricated phrases—just that it is not characterized by mainly being a part of prefabricated language.

The same procedure will be performed in the following chapter to identify any variation between the Observed Meanings that were synthesized from the lexical profiles and the patterns of co-occurrence for the key terms with regard to group membership of the author. This inquiry will allow us to see if domain-specific language, like that of the regulated nuclear power industry, is influenced by this particular dimension. In other words, we seek to discover if the meaning associated with the industry key terms that we observed in this chapter are in fact the same meanings that emerge for the different industry groups or if there are differences in the ways that these terms are used.

CHAPTER 5

VARIATION IN KEY TERM COLLOCATIONS WITH REGARD TO ORGANIZATIONAL AFFILIATION

Determining those terms that are fundamental and distinctive for the language of the nuclear industry is merely the first step in learning about the influence some dimensions may have on domain-specific language. One of these dimensions is that of group membership, or what I will call industry groups. The results in this chapter indicate that there are differences in the interaction of groups in the nuclear power industry with regard to the way meaning is constructed for the industry terms identified in the previous chapter. These differences are underlying the industry perspective and may or may not have emerged in the industry-level. The reason for this behavior is that both groups share aspects of the meanings observed in Chapter 4, which emerge in the aggregate. However, the data in this chapter will demonstrate that industry groups also have their own unique lexical profiles for industry key terms, which reflect their respective roles in this industry.

GROUP REPRESENTATION IN THE NUCLEAR REFERENCE CORPUS

The 7,110 documents in the corpus were written by individuals classified as being internal to the regulated nuclear power industry, authors who can be organized into four general groups: the Nuclear Regulatory Commission, the corporations/utilities holding licenses to operate nuclear power reactors or licensees, vendors providing services to the licensees, and government agencies other than the NRC that regulate activities concerning nuclear power production (i.e. emergency management or environmental protection).

Using metadata from the nuclear reference corpus, we are able to calculate the distribution of documents written by individuals from each of these four groups (Table 5.1).

Table 5.1: Document Distribution by Regulated Nuclear Industry Organization Group

	NRC	Licensees	Vendors	Government Agencies (Non-NRC)
# Of Documents	3,354	3,473	199	84
% Of Total	**47.17%**	**48.85%**	**2.8%**	**1.18%**

Although the lexical profiles in Chapter 4 result from texts authored by all four author classifications, the primary industry groups represented in this corpus are the NRC and the licensees. As a result, these are the ones that will be analyzed in this chapter for differences with regard to the industry group designation.

One way to see if membership with regard to organizational affiliation matters in the language of regulated corporate discourse is to compare the ways in which meaning is created for the industry key terms used by each of these groups. First, I performed a key words analysis for the NRC and the licensees in comparison to one another. These lists were used to see if any of the 20 industry key terms were used more (or less) often by one group in comparison to the other (Table 5.2). The NRC-authored texts used the key terms *NRC, staff, licensee, license, operating, evaluation, safety, inspection, inspectors, plant,*

Table 5.2: Key Term Analysis by Industry Group

Key word	NRC Freq.	NRC %	Licensee Freq.	LIC. %	Group with Positive Keyness	|log-likelihood|
NRC	20643	0.52	9132	0.20	NRC	6083.30
safety	13369	0.34	10192	0.23	NRC	900.13
reactor	8566	0.22	10273	0.23	Licensee	17.62
plant	10528	0.27	10940	0.24	NRC	36.16
inspection	13144	0.33	5628	0.13	NRC	4077.28
unit	7450	0.19	12186	0.27	Licensee	655.98
licensee	15211	0.38	1337	0.03	NRC	15407.05
system	11317	0.29	13921	0.31	Licensee	47.17
power	10429	0.26	12973	0.29	Licensee	55.70
nuclear	9409	0.24	6473	0.14	NRC	958.77
fuel	5151	0.13	8153	0.18	Licensee	370.55
license	8088	0.20	3603	0.08	NRC	2351.09
containment	3955	0.10	6911	0.15	Licensee	498.43
operating	6134	0.15	6181	0.14	NRC	39.19
evaluation	5713	0.14	5390	0.12	NRC	88.63
staff	12368	0.31	2107	0.05	NRC	9357.11
technical	5432	0.13	6421	0.14	None	2.84
emergency	4289	0.11	6590	0.15	Licensee	253.04
pressure	4657	0.12	8086	0.18	Licensee	568.29
inspectors	9301	0.23	68	-	NRC	13339.96

and *nuclear* more than what was found in documents authored by licensee representatives. It can also be noted from Table 5.2 that eight of the remaining nine key terms were found to be used more often by the licensee group in comparison to the NRC-authored documents *(reactor, unit, system, power, fuel, containment emergency, pressure)*. *Technical* was the only term not used at significantly different rates by the two groups (p-value<0.0001). When we juxtapose those terms found to be used more frequently than expected by the authors of NRC documents to their licensee counterparts, some interesting semantic sets become apparent (Table 5.3).

Table 5.3: Key Terms by Industry Group

NRC Key Terms	Licensee Key Terms
NRC	reactor
safety	unit
plant	system
inspection	power
licensee	fuel
nuclear	containment
license	emergency
operating	pressure
evaluation	
staff	
inspectors	

The first thing we are able to notice is that the NRC uses those terms representing larger entities more than expected *(NRC, staff, licensee, inspectors,* and *plant*). Conversely, we can see from the licensee key term column in Table 5.3 that there are more terms relating to smaller entities that are local as well as plant-based items characterized by a mechanical nature *(reactor, unit, system, fuel, power)*. We are also able to observe a distinction in the preference for the abstract in the NRC-authored documents *(safety, nuclear, operating)* versus the applied in documents written by licensee representatives *(emergency, containment, pressure)*. Lastly, we see that the NRC most frequently uses those industry terms related to their supervisory role *(inspection, evaluation, license, licensee, inspectors)*

When the collocates for each term are plotted against the total number of co-occurrences for each group, we are able to confirm again that these distributions all form A-curves. Figure 5.1 is a line graph for *NRC* from the documents authored by NRC representatives, while Figure 5.2 represents data from licensee-authored texts.

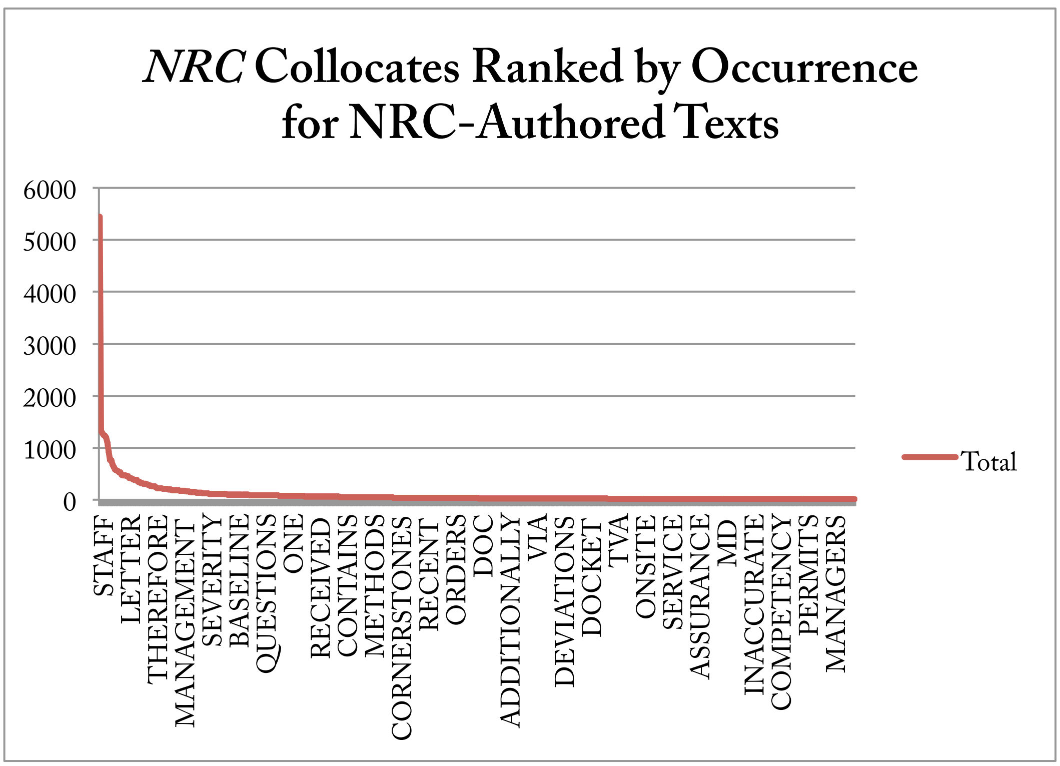

Figure 5.1: Collocates with *NRC* in NRC-Authored Texts

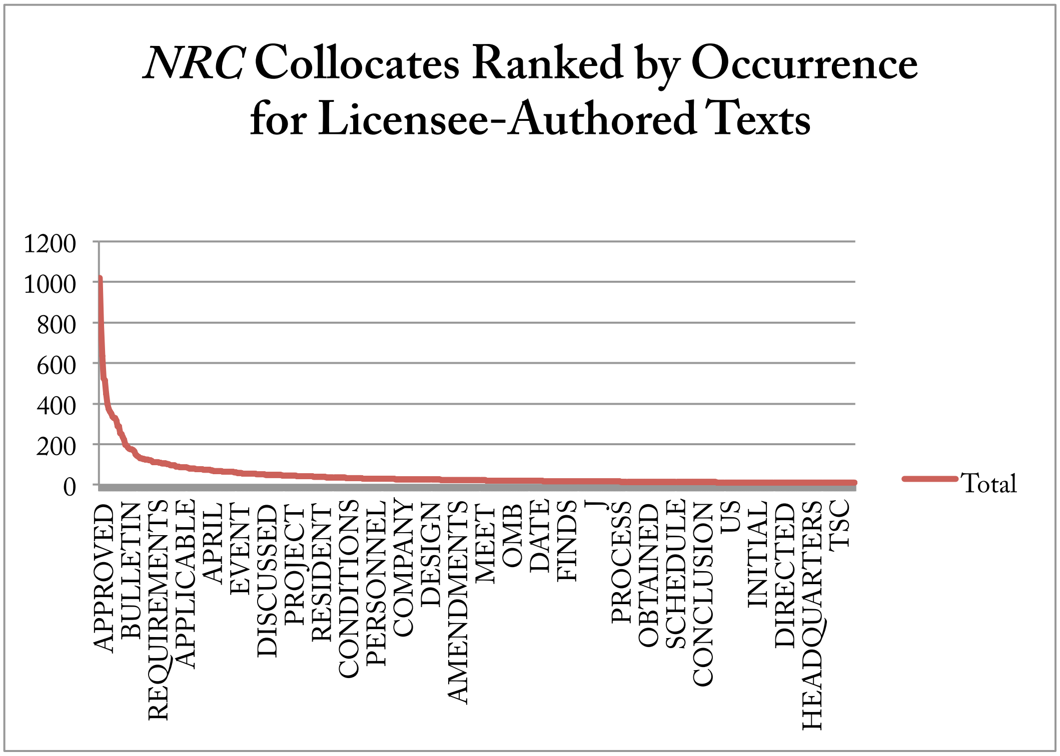

Figure 5.2: Collocates with *NRC* in Licensee-Authored Texts

The resulting patterns demonstrate the same behavior noted in Chapter 4 that the rankings of each collocate are inversely proportionate to their frequency of co-occurrence. Furthermore, the slope of the curves for the collocates for each term and each industry group are different (Table 5.4). The percentage of total occurrences demonstrated through an analysis of the 20 most frequently occurring collocates with each term for each group also had different distributions (Table 5.5).

When we look at the difference between each group for the percentage of occurrences represented by the top 20 collocates, we find that for authors affiliated with the NRC it was 24.42% to 57.76%, while for the licensees it was 24.10% to 53.47%. On

Table 5.4: Rank/Frequency Proportions for 20 Terms by Industry Group

Term	NRC Total Colls. (Types)	NRC Total Occurs. (Tokens)	NRC Token/Type Ratio	Licensee Total Colls. (Types)	Licensee Total Occurs. (Tokens)	Licensee Token/Type Ratio
NRC	586	54190	76/24	583	27563	72/28
safety	722	47522	75/25	652	34785	73/27
reactor	423	28595	74/26	462	32199	75/25
plant	499	28242	71/29	853	36126	70/30
inspection	649	43577	75/25	326	15798	70/30
unit	394	16312	69/31	471	23875	70/30
licensee	518	34670	72/28	98	2648	66/34
system	836	42414	72/28	651	40506	73/27
power	518	33510	73/27	558	37948	74/26
nuclear	356	35002	76/24	420	23177	72/28
fuel	432	17623	71/29	413	23180	73/27
license	261	20262	77/23	251	11346	73/27
containment	285	10728	70/30	395	20018	71/29
operating	282	16677	74/26	485	19862	71/29
evaluation	280	12795	70/30	306	11783	69/31
staff	376	29658	77/23	130	4219	67/33
technical	200	14267	75/25	283	18257	75/25
emergency	276	13481	71/29	408	19464	72/28
pressure	310	15086	73/27	636	30742	73/27
inspectors	254	25015	75/25	1	10	-

the other hand, comparing these ranges with what we observed in Chapter 4—20.02% to 47.02%—allows us to see that the number of collocates and their frequencies for each group are less than those for the entire industry. Thus, the Observed Meanings for each term based off of the 20 most frequently co-occurring words for each group are based off of a greater number of occurrences, or we are able to move a little further down the top-most part of the A-curve.

The analysis in this chapter will be of the 10 most frequently occurring industry key terms from the previous chapter due to the scope of this dissertation (the remaining 10 profiles can be found in Appendix B). Even analyzing only half of the industry key

Table 5.5: Representation of Top 20 Collocates and Their Occurrences by Industry Group

Term	NRC Top 20 Total Occurs.	NRC Top 20 Percentage	NRC Top 20 Occurs. Percentage	Lic. Top 20 Total Occurs.	Lic. Top 20 Percentage	Lic. Top 20 Occurs. Percentage
NRC	20778	3.41%	38.34%	8271	3.43%	30.01%
safety	16624	2.77%	34.98%	10925	3.07%	31.41%
reactor	11168	4.73%	39.06%	14504	4.33%	45.04%
plant	8908	4.01%	31.54%	8705	2.34%	24.10%
inspection	15027	3.08%	34.48%	5317	6.14%	33.65%
unit	4855	5.08%	29.76%	7077	4.25%	29.64%
licensee	9127	3.86%	26.33%	1416	20.41%	53.47%
system	10358	2.39%	24.42%	11251	3.07%	27.78%
power	11928	3.86%	35.60%	12229	3.58%	32.23%
nuclear	17498	5.62%	49.99%	9379	4.76%	40.47%
fuel	6549	4.63%	37.16%	9395	4.84%	40.53%
license	11398	7.66%	56.25%	5678	7.97%	50.04%
containment	3745	7.02%	34.91%	6795	5.06%	33.94%
operating	7625	7.09%	45.72%	6780	4.12%	34.14%
evaluation	5226	7.14%	40.84%	3876	6.54%	32.89%
staff	15447	5.32%	52.08%	2031	15.38%	48.14%
technical	20	10.00%	57.76%	9172	7.07%	50.24%
emergency	5471	7.25%	40.59%	6464	4.92%	33.21%
pressure	6274	6.45%	41.59%	9898	3.14%	32.20%
inspectors	13279	7.87%	53.08%	-	-	-

terms will allow us to see any differences that exist between the industry groups' lexical profiles. The following results focus on the variation between the industry groups with regard to the use of nuclear industry key terms measured through collocate analysis and the creation of lexical profiles. The resulting meanings both reflect the roles of the respective industry groups, as well as any underlying meaning that did or did not emerge in the industry level observations made in Chapter 4.

NRC

When we look at the 20 most frequent collocates with *NRC* for both of these groups (MI > 3.0), more differences become apparent in the use of this term (Table 5.6).

Table 5.6: Top 20 Collocates with *NRC* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
staff	6.400703	5445	approved	19.87102	1021
regulatory	5.683362	1330	*staff*	19.67430	789
site	6.259449	1277	letter	18.62722	635
inspection	17.86508	1244	approval	19.71141	526
commission	5.549767	1238	regulatory	18.26500	516
nuclear	4.615767	1202	request	18.04827	449
public	5.106264	1099	*nuclear*	17.07503	400
review	4.678869	930	dated	18.61484	373
web	7.601296	769	information	16.94588	361
ADAMS	5.779656	764	review	17.74697	350
staff's	6.512188	673	commission	19.17993	336
approved	6.057064	629	response	16.95887	330
reviewed	4.037202	576	question	18.87935	330
concludes	6.454955	565	provided	17.21340	316
document	5.507801	560	reference	17.00246	288
room	16.63959	532	submitted	18.59899	287
determined	4.656599	532	*evaluation*	16.68401	254
evaluation	4.011169	480	requested	18.23957	251
enforcement	6.104842	468	additional	16.96877	236
oversight	6.294382	465	*safety*	26.30733	223

Both groups do share some of the same collocates for *NRC*: i.e. *staff, approved,* and *evaluation*. However, these collocates have very different frequencies and rankings. This is a result of the scaling nature of language that we also observed in the text-type distributions during corpus creation as discussed in Chapter 3. In the subsequent tables, we can see that the scaling behavior observed in Table 5.6 is present for all of the collocate distributions for the 20 key terms with respect to industry group. By isolating the 20 most frequent collocates for each industry group that are not found in both columns and juxtaposing them to one another, we can see the differences in the way this term is used by each one (Table 5.7).

Table 5.7: Collocates with *NRC* by Industry Group (Not Repeated in Table 5.6)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
site	6.259449	1277	letter	18.62722	635
inspection	17.86508	1244	approval	19.71141	526
public	5.106264	1099	request	18.04827	449
web	7.601296	769	dated	18.61484	373
ADAMS	5.779656	764	information	16.94588	361
staff's	6.512188	673	response	16.95887	330
reviewed	4.037202	576	question	18.87935	330
concludes	6.454955	565	provided	17.21340	316
document	5.507801	560	reference	17.00246	288
room	16.63959	532	submitted	18.59899	287
determined	4.656599	532	requested	18.23957	251
enforcement	6.104842	468	additional	16.96877	236
oversight	6.294382	465	*safety*	26.30733	223

.

The collocates for each group in Table 5.7 can be formed into semantic sets. For the regulatory group, we can form a group related to providing information to the public *(public, site, web, ADAMS, public, document, room)*. The reason why *document* and *room* are included in this set is mainly due to the collocate cluster *NRC PUBLIC DOCUMENT ROOM* (474). A second group involves those words related to the NRC's role *(enforcement, oversight)*, as well as how it is enacted *(inspection, reviewed, concludes, determined, identified)*. From the perspective of NRC-authored documents, it is the *NRC* and its staff that has the ability to make determinations, to make policy, to provide oversight, and to provide approval in their relationships with licensees and vendors: *NRC STAFF CONCLUDES* (453), *NRC STAFF HAS DETERMINED* (264), *NRC STAFF HAS REVIEWED* (229), and *APPROVED BY THE NRC* (223) (the use of all capitalized, italicized letters in this chapter is reserved for clusters from the corpus using the collocates and terms). Thus we can see that when the *NRC* refers to itself, it is often demonstrating its supervisory capacity in this industry. We can also see that it does also possess some power, which is demonstrated through the presence of enforcement: *NRC ENFORCEMENT POLICY* (357).

When we look at the groups that can be made of the collocates in Table 5.7 for the industry group of licensees, their perception of the NRC's role in the industry becomes quite evident. The first group we can form from these collocates relates to specific documents *(reference, form, bulletin, report)*. It seems that evidentiary support is something that characterizes the licensee relationship with *NRC*. Evidence from *reference*s is provided to the *NRC*, sometimes even from documents authored by the *NRC*: *THE NRC IN REFERENCE* (41), *IN LETTER TO NRC THE NRC REFERENCE* (28), *BY THE NRC IN REFERENCE* (23), *TO THE NRC IN REFERENCE* (12). In other words, we find that the perceptions of what the *NRC* is differs with regard to industry group. For those documents authored by NRC-affiliates it is power and control, while for licensee-authored documents it is information and documents.

The second semantic set we can make from these collocates relates to an *information* exchange circuit. One sub-group of this set relates to requests *(request,* *question, requested*), and the second sub-group involves responses *(response, provided, submitted)*. From the clusters associated with these words, we find out that the first sub-group for this set is the role of the NRC. Moreover, the licensee is the one who responds to *request*s made by the *NRC*: *RESPONSE TO NRC BULLETIN* (59), *SUBMITTED TO THE NRC* (205), *RESPONSE TO NRC QUESTIONS* (24). Although the word *dated* does not appear to belong to this semantic set, we learn from its most frequent clusters that when these exchanges occur is a defining element for identifying them in communications: *LETTER TO NRC DATED* (62), *NRC LETTER DATED* (68), *LETTER TO THE NRC DATED* (13). There are two more collocates that also belong to this set: *approval* and *additional*. What these collocates contribute to our understanding of the information exchange aspect of *NRC* is the nature of this circuit. *Additional* allows us to understand that this is a repetitive process. There are not only *requests,* but also *RESPONSE TO NRC REQUEST FOR ADDITIONAL INFORMATION* (51), which suggests that this circuit repeats itself. The inclusion of *approval* to this group also lets us know that there is finality to this circuit as well when the NRC provides a judgment that completes the information exchange circuit regarding a particular issue: *NRC APPROVAL OF* (90), *NRC REVIEW AND APPROVAL* (74). Lastly, while *safety* does not fit into any of these semantic sets, it does provide a connection for the collocates and *NRC* on a topical level regarding what the *NRC* supervises: *NRC SAFETY EVALUATION* (64).

Despite all of the differences between these two groups with regard to the 20 most frequently occurring collocates for *NRC*, there is also some shared meaning between them in the use of this term, which can be noticed in the collocates listed in Table 5.8.

We see that within the 20 most frequently co-occurring words with *NRC* that both groups have seven words that are most salient. What we can see from this combination of words is that there is a shared understanding of who the *NRC* is, what it does, and who does it for the

Table 5.8: Collocates Present for Both Industry Groups in Top 20

Repeated from Table 5.6 for Both Groups
regulatory
commission
nuclear
staff
review
approved
evaluation

NRC. This is not to say that some of the observations that have already been made are not also shared on some level, just not in the most salient level through the 20 most frequently co-occurring words.

This lexical profile results in the following Observed Meanings for the term *NRC* for the NRC and licensees as separate industry groups:

1. **Authors of NRC documents perform a supervisory role in this industry, which is exemplified through the use of this term, while also serving the public by providing them access to documents and information from the regulated nuclear industry.**

2. **The authors of licensee documents view the *NRC* as a regulatory figure from which they seek validation and approval, and to whom they also provide cogent proof and information.**

SAFETY

When we look at the 20 most frequent collocates with *safety* for both of these groups we are able to see that there are similarities and differences between their use for this term as well (Table 5.9).

Table 5.9: Top 20 Collocates with *Safety* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
significance	19.34161	1952	related	18.37604	1221
public	18.91414	1527	analysis	16.73006	1096
health	20.58847	1424	*evaluation*	16.76142	804
low	19.65174	1408	function	17.28317	711
related	19.29486	1279	*system*	15.15720	683
evaluation	18.40343	1000	report	16.18005	647
analysis	18.37313	917	margin	17.74714	545
report	18.26066	908	injection	18.11862	535
function	19.12739	632	health	18.63035	486
radiation	18.49796	616	public	17.50489	466
final	19.34669	588	*plant*	28.29107	441
system	30.07249	570	analyses	16.65046	396
reactor	16.89767	528	final	17.98638	394
green	29.95110	524	valves	16.19649	387
level	17.94057	524	level	15.75459	386
quality	19.17449	509	significance	18.85608	381
performance	17.14280	462	systems	16.08461	378
updated	20.15489	432	limit	16.17847	334
systems	17.32729	420	quality	17.23678	326
injection	19.73354	404	equipment	15.89287	308

From the perspective of those words co-occurring most with *safety* for each group, we notice that there are 14 words that are present for both groups *(significance, public, health, related, evaluation, analysis, report, function, final, system, level quality, systems, injection)*. These words demonstrate that the notion of *safety* being significant to the public *(significance, public, health, related)* and systems *(function, system, systems, injection)* that need to be evaluated *(evaluation, analysis)*, measured *(level, quality)*, and reported *(report, final)* is shared on the highest level in this industry.

Looking at the 20 most frequent collocates for each industry group that are not found in both columns in Table 5.9 and juxtaposing them to one another, we can see that there are differences in the way this term is used by each group (Table 5.10).

Table 5.10: Collocates with *Safety* by Industry Group (Not Repeated in Table 5.9)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
low	19.65174	1408	margin	17.74714	545
radiation	18.49796	616	*plant*	28.29107	441
reactor	16.89767	528	analyses	16.65046	396
green	29.95110	524	valves	16.19649	387
performance	17.14280	462	limit	16.17847	334
updated	20.15489	432	equipment	15.89287	308

The collocates for the NRC-authored texts in Table 5.10 do not appear to form any semantic sets, but we can see that *safety* is most often also associated with the nuclear reactors used by the licensees *(radiation, reactor)*. This focus on *safety* has to do with protecting people from *radiation* resulting from the *reactor*: *OCCUPATIONAL*

RADIATION SAFETY (235), *REACTOR SAFETY RADIATION SAFETY SAFEGUARDS* (84), *PUBLIC RADIATION SAFETY* (126). We can also get a better understanding of how *safety* is measured in documents authored by NRC representatives through the presence of *low* and *green*: *VERY LOW SAFETY SIGNIFICANCE* (1,211), *VERY LOW SAFETY SIGNIFICANCE GREEN* (422). Moreover, we learn that *safety* is measured and evaluated because it is a attribute that is performed by licensees: *SAFETY PERFORMANCE DURING* (65), *ASSESSMENT OF YOUR SAFETY PERFORMANCE* (59). Finally, there is one collocate that is frequently used in NRC-authored texts with *safety* to communicate that those *reports* made concerning this issue can in fact be revised and *updated*: *UPDATED FINAL SAFETY ANALYSIS REPORT* (324).

For the licensee-authored documents, Table 5.10. shows that *safety* is related to the *plant* and its *equipment (plant, valves, equipment): SAFETY RELATED EQUIPMENT* (144), *SAFETY RELIEF VALVES* (91). We can also see that on this highest level of occurrence for licensee-authored documents that while they communicate the characteristic of *safety* being something measured, there are implications for boundaries of these *safety* measurements that cannot be crossed by the frequent co-occurrence of *margin* and *limit*: *MARGIN OF SAFETY* (514), *THE SAFETY LIMIT* (40), *THE SAFETY LIMIT MINIMUM* (8), *SAFETY LIMIT IS* (37). The sixth collocate in Table 5.10 for the licensee industry group is *analyses*. The significant use of *analyses* in addition to *analysis* implies that safety requires multiple critical inquiries: *THE SAFETY ANALYSES* (131), *ASSUMED IN THE SAFETY ANALYSES* (20), *APPLICABLE SAFETY ANALYSES* (62), *CONTINUED APPLICABLE SAFETY*

ANALYSES (14), *ASSUMPTIONS OF THE SAFETY ANALYSES* (17). This perspective of *safety* shows us that while there is a shared understanding of this term for both industry groups, there is also a difference in the meaning observed from the 20 most frequently co-occurring words with it for each group.

This lexical profile results in the following Observed Meanings for the term *safety* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, *safety* directly relates to protecting the public from radiation due to the use of reactors by the licensees. They are also responsible for evaluating *safety*, and reports on this topic can be updated.**
2. **Authors of documents for the licensees view *safety* as a quality for equipment at the plant that is in place to protect the public. *Safety* is also measured on an ongoing basis.**

REACTOR

An analysis of the term *reactor* shows us that there are also differences in the 20 most frequent words that co-occur with it between the two industry groups (Table 5.11). The first thing to notice is that there are 12 collocates found in the lists for both groups in Table 5.11 *(coolant, vessel, system, pressure, trip, water, power, core, building, head, pump, unit)*. The presence of these collocates indicates that the shared meaning associated with this term for both groups directly relates to the components of the *reactor (system, core, head, pump, coolant, water, pressure, power)*. We can also see that the idea that a *reactor* is contained within other structures is also shared between the industry groups *(vessel,*

Table 5.11: Top 20 Collocates with *Reactor* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
coolant	8.340731	1491	vessel	8.108613	2153
vessel	8.089266	1110	coolant	8.267525	2126
system	18.88256	1045	*system*	5.668736	1628
pressure	6.145844	713	*pressure*	19.04444	1243
trip	7.386607	597	trip	7.125992	990
power	4.634240	560	water	5.519383	887
water	5.437552	530	*power*	4.712626	782
safety	16.89767	528	building	6.985383	589
core	6.245060	513	level	5.554957	562
nuclear	4.630724	504	core	5.483571	536
building	7.334835	429	head	6.214962	423
office	7.096395	403	protection	5.842977	339
process	14.47166	393	boundary	6.317590	324
head	7.137866	391	temperature	4.792819	322
oversight	7.305880	389	RCS	5.668444	313
operator	6.006049	362	pump	4.744263	283
regulation	8.146052	343	scram	6.940775	270
senior	7.734932	310	*unit*	3.247172	266
pump	5.291820	299	shutdown	5.078326	240
unit	4.001473	258	low	5.084976	228

building, unit), as well as the notion that the *reactor* can *trip*. As a result, these are the meanings that emerged in the industry-level analysis.

When we analyze the 20 most frequent collocates with *reactor* for each industry group that are not shared by both columns in Table 5.11, we can see that there are differences in the way this term is used (Table 5.12). From the perspective of NRC-authored documents, we see that *reactor* emerges most often with words related to the NRC being the entity responsible for supervising the use of *nuclear reactors* due to *safety*

Table 5.12: Collocates with *Reactor* by Industry Group (Not Repeated in Table 5.11)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
safety	16.89767	528	level	5.554957	562
nuclear	4.630724	504	protection	5.842977	339
office	7.096395	403	boundary	6.317590	324
process	14.47166	393	temperature	4.792819	322
oversight	7.305880	389	RCS	5.668444	313
operator	6.006049	362	scram	6.940775	270
regulation	8.146052	343	shutdown	5.078326	240
senior	7.734932	310	low	5.084976	228

implications*: NUCLEAR REACTOR REGULATION* (341), *OFFICE OF NUCLEAR REACTOR REGULATION* (337), *REACTOR SAFETY CORNERSTONES* (125). Furthermore, role the NRC possesses with regard to *reactors* is one that is governed by an established *process: REACTOR OVERSIGHT PROCESS* (349).

We also discover through the presence of *senior* and *operator* that the NRC has the authority to govern who can operator *reactors: SENIOR REACTOR OPERATOR LICENSE* (35). This indicates that a *reactor* requires oversight, such that a person must obtain their endorsement to become a *reactor operator* through a credentialing process: *SENIOR REACTOR OPERATOR LICENSE APPLICATIONS* (20). The NRC connotation of a reactor being something requiring regulation and licensure again implies that this government agency is in a position of authority over the licensees. It should also be noted that the presence of *senior* and *operator* in the 20 most frequently co-occurring words with this term also communicates the characterization of those individuals whose jobs directly involve a *reactor*, as well as there being different levels to the experience of these individuals*: SENIOR REACTOR OPERATOR* (158), *REACTOR OPERATOR AND SENIOR REACTOR OPERATOR* (42), *A SENIOR REACTOR INSPECTOR* (15), *A SENIOR REACTOR ANALYST* (10).

When looking at the collocates listed in Table 5.12 for the licensee industry group, we find a different perspective in the 20 most frequently co-occurring words. First of all, we can create one semantic set from these collocates relating to the measurement of entities in the *reactor (level, temperature, RCS, low, fuel)*. We are also reminded that in addition to the measurement of the *reactor* there is also a pervasive focus in these documents on safety and limits that should not be passed *(protection, boundary): REACTOR PROTECTION SYSTEM* (264), *REACTOR COOLANT PRESSURE BOUNDARY* (257).

One of the biggest differences between the NRC and licensees' use of *reactor* is the presence of two terms on the licensee collocate list for the shutdown of the nuclear reactor in addition to *trip: scram* and *shutdown*. The NRC website states that a nuclear reactor scram is "the sudden shutting down of a nuclear reactor, usually by rapid insertion of control rods, either automatically or manually by the reactor operator—also known as a 'reactor trip'" ("Reactor Scram"). The presence of these terms related to the *shutdown* of a *reactor* in emergency situations in Table 5.12 for the licensee industry group helps us to understand the importance of *safety* when they operate a nuclear reactor, and what actions become necessary for them to take when their measurements indicate a certain *boundary* has been breached.

This lexical profile results in the following Observed Meanings for the term *reactor* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, a *reactor* is a machine over which they have regulatory authority, and any individual wanting to operate a *reactor* must gain the NRC's approval and endorsement**
2. **For the authors of documents for licensees, a *reactor* is a machine that they operate and sometimes have to shutdown. This machine is measured, as it has limits and boundaries for its operation.**

PLANT

An analysis of the 20 most frequently co-occurring words with the term *plant* shows us that there are also differences in the meaning associated with this term between the two industry groups (Table 5.13). There are 13 words shared by both industry groups in Table 5.13 *(nuclear, power, specific, unit safety, operation, operating, conditions, operations, procedures, design, systems, shutdown)*. The presence of these words in the list of 20 most frequently used collocates with *plant* for both groups indicates that there are several shared meanings for *plant.* First of all, the notion of *plant* identification is pervasive for both groups *(nuclear, power)*: *NUCLEAR POWER PLANT*—as well as the fact that a *plant* is composed of *unit*(s) and *systems* that work together through a specific *design*. This notion is embedded in the language for the entire industry—confirming that the industry-wide denotation for *plant* being a place generating nuclear power. We can also see that the semantic set relating to *plant operations* is shared *(operation, operating, operations, shutdown, conditions, procedures),* as well as *safety* being a concern in relation to *plant*.

Table 5.13: Top 20 Collocates with *Plant* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	6.057213	1665	*nuclear*	19.90082	1418
power	5.469514	1228	*power*	18.17636	860
specific	6.544244	813	specific	31.68198	771
performance	4.799476	468	operation	18.71448	628
unit	4.347344	403	conditions	18.71068	553
safety	16.15565	388	*unit*	17.47846	498
modifications	6.892541	335	*safety*	28.28779	440
operation	4.809669	328	*operating*	29.65364	378
units	14.07031	320	systems	18.47802	331
operating	4.239906	308	shutdown	18.81089	318
conditions	4.744020	307	operations	19.02951	308
risk	4.848057	306	design	17.58904	286
operations	5.705699	295	procedures	18.98588	285
procedures	4.812157	294	personnel	18.48267	276
areas	5.251546	270	equipment	18.19404	253
design	4.151516	253	will	16.23371	251
inspectors	3.332564	249	*system*	16.19061	233
systems	4.115282	233	changes	17.17681	212
status	6.146927	232	data	29.77790	206
inspection	16.29044	213	normal	18.53007	200

When we move past this shared perspective to look at the seven words for each group that are not repeated in Table 5.13, we find that there is some variation to the meaning associated with this word at such a high level (Table 5.14). The NRC's use of *plant* indicates that their interests are of a broad nature *(units, areas)*. For example, they refer to *AREAS OF THE PLANT* (80) and *LARGE AREAS OF THE PLANT* (73), and *POWER PLANT UNITS* (49). Furthermore, *risk* is a word that is commonly used with *plant* by this group of authors, which communicates that there is a certain element of danger related to *plant* operation, and that this situation is specific to each *plant: PLANT*

Table 5.14: Collocates with *Plant* by Industry Group (Not Repeated in Table 5.13)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
performance	4.799476	468	personnel	18.48267	276
modifications	6.892541	335	equipment	18.19404	253
units	14.07031	320	will	16.23371	251
risk	4.848057	306	*system*	16.19061	233
areas	5.251546	270	changes	17.17681	212
inspectors	3.332564	249	data	29.77790	206
inspection	16.29044	232	normal	18.53007	200

SPECIFIC RISK (43), *IMPACT ON PLANT RISK* (29), *PLANT SPECIFIC RISK CONSEQUENCES* (18).

The NRC also relates the *plant* with *inspection* and the individuals who have been given the job of performing them: *INSPECTION OF NUCLEAR PLANT* (48), *INSERVICE INSPECTION OF PLANT* (48), *THE INSPECTORS REVIEWED* (81), *RESIDENT INSPECTORS ASSIGNED TO THE PLANT* (27). This emphasis on evaluation is made even more apparent with the presence of *performance* in combination with the frequent use of *status: DETAILS SUMMARY OF PLANT STATUS* (65), *PLANT STATUS AND INSPECTOR INSIGHTS* (13), *ASSESSING PLANT PERFORMANCE* (8). By analyzing the most frequent clusters of *plant* with *modifications,* we learn that *TEMPORARY PLANT MODIFICATIONS* (63) and *PERMANENT PLANT MODIFICATIONS* (26) are more aspects of the *plant* that are evaluated in NRC-authored documents.

When we take and compare the collocates from the NRC-authored documents to those from the licensee industry group, we notice some interesting variation. First of all, notice in Table 5.14 that the noun *modifications* is present rather than *changes.* The word *modifications* connotes the act of something changing so as to be a variety, maintaining some notion of its origin. When we contrast *modifications* with *changes,* we become privy to a difference in attitudes between the NRC and licensees on this level. Where the NRC refers to *TEMPORARY PLANT MODIFICATIONS* (63), the licensee-authored texts provide evidence of *CHANGES TO THE PLANT* (30) and *CHANGES IN PLANT OPERATION* (7). This is not to say that the licensees do not also refer to modifications, or that the NRC does not talk about changes as well. However, it does not happen on this highest level of frequency analysis.

We also find from the licensee group perspective that *plant* is directly related to local entities *(personnel, equipment, system)*: *PLANT SPECIFIC PERSONNEL* (17), *ANY PLANT EQUIPMENT* (24), *ANY PLANT SYSTEM* (16). Moreover, the collection of facts and statistics for analysis of a *plant* also appears to be frequently discussed in documents authored by this group: *PLANT SPECIFIC DATA* (19), *PLANT DESIGN DATA* (11). In addition to *data*, it is also the perspective of licensee authors to relate *plant operations* to the noun *normal* in order to communicate standard or typical *plant operations: METHODS GOVERNING NORMAL PLANT OPERATIONS* (37), *DURING NORMAL PLANT OPERATIONS* (19), *NORMAL PLANT OPERATING PARAMETERS* (5). Finally, we learn from the licensee perspective that *plant* is quite frequently related to the word *will.* This frequent collocational pair indicates that a *plant* makes commitments regarding its actions in the future: *THE PLANT WILL BE* (15), *THE PLANT WILL BE PERFORMING* (5), *PLANT WILL SATISFY APPLICABLE REGULATORY REQUIREMENTS* (22).

This lexical profile results in the following Observed Meanings for the term *reactor* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, a *plant* is a place for generating nuclear power whose performance is evaluated by inspectors, often with regard to safety.**
2. **For the authors of documents for licensees, a *plant* is a place for generating nuclear power that is dynamic, and any changes that are made are specific to it, including commitments regarding its future actions. *Plant* operation is often classified as normal, and data is collected about it.**

INSPECTION

There are also differences in the meaning associated with *inspection* between NRC-authored documents and those written by licensee representatives, which we can see through an analysis of the 20 most frequently co-occurring words with the term (Table 5.15). On this level of frequency analysis, we find that there are nine collocates that are shared by both groups in their twenty most frequently co-occurring words with *inspection* *(scope, NRC, results, report, inservice, period, program, performed, one)*. The presence of these terms in both columns of Table 5.15 indicates that there is a shared meaning for *inspection* being predetermined with regard to a planned series of events *(program),* what is relevant *(scope),* and time *(period)*. There is also consensus between the two groups at this level that *inspection* is directly related to the *NRC*, which is the entity that is responsible for performing *(performed)* them and generating the outcome of this type of assessment *(results, report)*. We also find that *inservice* is a type of *inspection* that is

Table 5.15: Top 20 Collocates with *Inspection* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
scope	21.1064	2288	inservice	8.97774	868
inspectors	19.5650	1821	program	5.93891	630
NRC	17.8604	1240	interval	7.5567	465
reviewed	19.3052	1101	visual	7.66108	389
public	19.4383	1098	will	3.61754	225
results	19.6290	790	methods	6.81819	221
report	18.9260	720	report	4.71718	215
available	19.3548	592	performed	4.64420	214
inservice	29.5201	583	results	4.95963	211
period	19.6338	546	tube	17.3334	208
findings	31.0051	544	requirements	4.17670	188
program	29.2605	487	*NRC*	16.1644	185
conducted	19.3153	481	year	6.08529	183
team	19.6534	434	one	15.5558	182
performed	18.5515	427	steam	4.69555	176
electronically	21.3941	400	period	5.26515	167
one	30.5213	389	time	4.12363	162
activities	18.3106	373	generator	5.26785	148
procedures	18.8284	368	SG	6.35490	147
Mr.	19.5570	345	plan	5.67216	133

frequently discussed in the writing of both groups, the frequency with which aspects of the inspection re-occur *(one)*—as in *ONE INSPECTION SAMPLE*—as well as the future nature of inspections (will).

On the other hand, looking at those words that most frequently co-occurred with *inspection* that were not shared by both groups in Table 5.15, we notice that there is a difference in the notion of this term for them on this level of saliency (Table 5.16). By putting these collocates into small groups, we can get a better sense of the notions related to *inspection* being used by each group most frequently. First of all, we find that the people responsible for conducting *inspections* are frequently used by the *NRC*

Table 5.16: Collocates with *Inspection* by Industry Group (Not Repeated in Table 5.15)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
inspectors	19.5650	1821	interval	7.5567	465
reviewed	19.3052	1101	visual	7.66108	389
public	19.4383	1098	methods	6.81819	221
available	19.3548	592	tube	17.3334	208
findings	31.0051	544	requirements	4.17670	188
conducted	19.3153	481	year	6.08529	183
team	19.6534	434	steam	4.69555	176
electronically	21.3941	400	time	4.12363	162
activities	18.3106	373	generator	5.26785	148
procedures	18.8284	368	SG	6.35490	147
Mr.	19.5570	345	plan	5.67216	133

representatives *(inspectors, team)*. So, not only are there multiple people who perform an *inspection* of a licensee, but they work as a team: *INSPECTORS PRESENTED THE INSPECTION RESULTS* (167), *THE TEAM REVIEWED THE INSPECTION* (93). For documents authored by this group, there are words related the types of evaluative actions the NRC performs during an *inspection (reviewed, conducted): THE INSPECTORS REVIEWED* (681), *INSPECTION EXAMINED ACTIVITIES CONDUCTED UNDER* (82). It also becomes apparent through the frequent collocation of *inspection* with *findings* that the conclusion or outcome of an inspection is also of importance in NRC-authored documents: *THE INSPECTION FINDINGS* (83), *REACTOR INSPECTION FINDINGS* (54), *INSPECTION FINDINGS BEING CLASSIFIED AS* (49).

The connotation also exists for the NRC-authored texts that the evaluative nature of an *inspection* is standardized by the use of *procedure*: *WITH NRC INSPECTION PROCEDURE* (40), *THE INSPECTION PROCEDURE* (67). Although it is clear that evaluation is the primary focus of *inspection* from the NRC perspective, it is also related to sharing: sharing with both plant officials, *INSPECTION RESULTS DISCUSSED WITH MR.* (12), *INSPECTION RESULTS PRESENTED TO MR.* (27); and the public, *AVAILABLE ELECTRONICALLY FOR PUBLIC INSPECTION* (400).

From the licensee group's perspective, those words most frequently co-occurring with *inspection* provide more specific details regarding what type of inspections are being performed and of what equipment *(visual, tube, steam, generator, SG)*. There are not just *inservice inspections;* there are *STEAM GENERATOR TUBE INSERVICE INSPECTION[s]* (60) and *VISUAL INSPECTION OF* (160). The licensee-authored texts also share a frequent discussion of the means of measurement employed by the NRC during inspections, as well as those criteria used for evaluation during the process: *INSERVICE INSPECTION PLAN* (48), *METHODS OF INSPECTION* (25), *THE INSPECTION METHODS* (22), *OF ASME CODE* (28), *METHODS NEED TO BE USED* (19), *INSPECTION REQUIREMENTS FOR REACTOR* (11), and *INTERIM INSPECTION REQUIREMENTS* (12).

We are also provided clarification that *time* is an important element of *inspection* from the perspective of this industry group. Not only are specific inspections performed at set intervals, like the *10 YEAR INTERVAL MASTER INSERVICE INSPECTION* (6), but also the results of an inspection happen at a specific point in time, providing context or support for licensee-authored documents: *AT THE TIME OF INSPECTION* (7).

This lexical profile results in the following Observed Meanings for the term *inspection* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, an *inspection* is an evaluation they perform that is guided by manuals and procedures, and the results of this process are shared with licensee officials. *Inspection* is also an action performed by the public on NRC documents that are available electronically.**
2. **For the authors of documents for licensees, an *inspection* may come in different forms, i.e. some focus on specific machinery. There are requirements for the way in which an inspection is performed, and it may serve as a point of reference with respect to time for the licensees.**

UNIT

An analysis of the 20 most frequently co-occurring words with the term *unit* shows us that there are also differences in the meaning associated with this term between the two industry groups (Table 5.17). In Table 5.17 we can observe that there are 12 words from the 20 collocates that most frequently co-occur with *unit* that are shared by both industry groups *(nuclear, station, power, plant, reactor, outage, refueling, steam, Point, Lucie, shutdown, one).* The use of this word as a way of pointing to a specific reactor *(reactor)*, its components *(steam)*, its identity and ownership *(nuclear, station, power, plant, Point, Lucie,*

Table 5.17: Top 20 Collocates with *Unit* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	4.989252	562	station	5.833503	699
station	6.656549	557	*power*	4.159016	632
power	4.697860	509	*plant*	17.47846	498
plant	4.347344	403	*nuclear*	4.788978	488
reactor	4.001473	258	cycle	17.39804	471
outage	5.823473	228	refueling	5.637607	410
refueling	5.986037	213	outage	5.587662	395
steam	5.258627	210	Lucie	7.746060	370
Point	6.110164	202	mode	5.425053	342
system	16.64711	193	steam	4.539440	342
Lucie	7.580050	172	*operating*	15.78525	308
pump	4.661517	168	Millstone	7.746505	277
Nos.	7.273769	165	*reactor*	3.247172	266
generating	6.705067	155	one	14.88769	248
shutdown	4.882126	154	shutdown	4.783049	232
B	14.28347	150	*technical*	3.715367	230
one	16.22458	144	HBRSEP	8.412469	223
inspection	16.19420	141	Salem	7.564682	222
Mile	7.705600	137	operation	3.595111	215
generator	4.971629	134	Point	5.258592	209

one), and even the processes it undergoes *(outage, refueling, shutdown)* is shared by both groups at this level of analysis. However, those words from Table 5.17 that are not shared by both industry groups help us to see that there is a difference in the meanings they associate with *unit* at this level of analysis reflect the roles of each group (Table 5.18).

When comparing the collocates from the NRC-authored texts to those from the licensees, there is some variation with regard to specific *units* belonging to licensees. The NRC list in Table 5.18 contains references to Nine Mile Point and/or Three Mile Island through the presence of *Mile,* while for the licensees we see frequent references to HB Robinson, Salem, and Millstone *(Millstone, HBRSEP, B, Salem)*. There are also other

Table 5.18: Collocates with *Unit* by Industry Group (Not Repeated in Table 5.17)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
system	16.64711	193	cycle	17.39804	471
pump	4.661517	168	mode	5.425053	342
Nos.	7.273769	165	*operating*	15.78525	308
generating	6.705067	155	Millstone	7.746505	277
B	14.28347	150	*technical*	3.715367	230
inspection	16.19420	141	HBRSEP	8.412469	223
Mile	7.705600	137	Salem	7.564682	222
generator	4.971629	134	operation	3.595111	215

differences in the meaning associated with *unit* for these two groups. From the NRC-authored documents we find several references to specific systems and components found in a *unit (system, pump, generator),* as well as a frequent use of *generating* as part of a unit's identity: *NUCLEAR GENERATING UNIT* (56), *GENERATING STATION UNIT* (62). It also becomes apparent that the NRC is regularly referring to more than one unit operated at a single licensee's facility *(Nos.): UNIT NOS. 1 AND 2* (114), *STATION UNIT NOS.* (55). We are also able to see the NRC's perspective that a *unit* requires their oversight with the presence of *inspection* and *amendment*: *INSPECTION OF THE UNIT* (13), *UNIT 2 LICENSE AMENDMENT* (8).

For licensee-authored documents, we find that from the frequent use of *cycle* and *mode* that *units* possess a dynamic quality: *SUBSEQUENT OPERATING CYCLE* (18), *CYCLE 23 REFUELING OUTAGE* (11), *WAS IN MODE* (77). There is also the frequent reference in licensee-authored documents of a *unit* undergoing *operation: OPERATION OF THE UNIT* (25), *SAFE OPERATION OF THE UNIT* (7), *THE OPERATING UNIT* (30), *UNIT WAS OPERATING IN MODE* (19). Finally, the presence of *technical* in Table 5.18 indicates that there are frequent references to the fact that *technical specifications* are specific to each *unit: UNIT NO.# TECHNICAL SPECIFICATIONS* (78), *UNIT TECHNICAL SPECIFICATIONS* (10).

This lexical profile results in the following Observed Meanings for the term unit for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, *unit* is a way of identifying reactors and components that work as part of a system at specific licensee facilities and requires their oversight through inspections and licensure.**
2. **For the authors of documents for licensees, a *unit* is also a way of identifying individual reactors that work as part of a system at their facilities. However, these units are dynamic, i.e. they operate in cycles, and technical specifications are specific to each unit.**

LICENSEE

There are also differences in the meaning associated with *licensee* between NRC-authored documents and those written by licensee representatives, which we can see through an analysis of the 20 most frequently co-occurring words with the term (Table 5.19). We can see from this table that there are five collocates shared by both groups in their 20 most frequently co-occurring words with *licensee (staff, company, stated, shall, changes)*. From these words we are able to see that at this high level of saliency that there is some consensus of the notion of *licensee* that was presented in Chapter 4. First of all, a *licensee is* not only a *company* that operates a nuclear reactor, but it also provides information *(stated)* to the NRC. Also, there is a shared knowledge that a *licensee* receives instruction

Table 5.19: Top 20 Collocates with *Licensee* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
identified	4.894190	1018	event	6.882845	231
staff	3.741137	635	report	6.784077	214
stated	6.079838	615	LER	9.323378	129
inspectors	4.016604	578	shall	5.796326	99
proposed	4.359791	562	controlled	8.281208	78
provided	4.941274	459	commitments	7.565016	60
failed	6.413832	439	*NRC*	16.58926	59
requested	5.632055	433	stated	7.482038	59
submitted	5.738823	415	selected	7.787959	57
also	4.334080	403	states	6.899258	49
shall	4.520546	400	provide	5.094725	47
company	5.870389	396	company	6.589930	46
reviewed	3.926201	393	commission	6.829097	43
determined	4.641768	388	regulatory	5.190913	40
information	3.598963	354	will	3.199313	40
performed	4.367661	349	reports	6.870306	37
violations	6.746733	341	document	6.280348	35
performance	3.746955	326	C	15.12165	32
changes	4.187256	317	*staff*	5.619834	31
facility	4.395778	306	changes	3.971000	30

from the NRC through the use of *shall*, as well as the fact that a *licensee* is associated with *changes*.

When we look at those words not shared by both groups in the list of the 20 most frequently occurring collocates with *licensee,* we find that there is variation between them (Table 5.20). We see that most of the words frequently co-occurring with *licensee* in documents authored by NRC representatives concern information and assessment. Four of these words are nouns: one of them concerns information *(information)*, one of them is a person who performs assessments *(inspector),* and two of them relate to assessment *(performance, violations)*. There are also eight verbs that frequently co-occur with *licensee*

Table 5.20: Collocates with *Licensee* by Industry Group (Not Repeated in Table 5.19)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
identified	4.894190	1018	event	6.882845	231
inspectors	4.016604	578	report	6.784077	214
proposed	4.359791	562	LER	9.323378	129
provided	4.941274	459	controlled	8.281208	78
failed	6.413832	439	commitments	7.565016	60
requested	5.632055	433	*NRC*	16.58926	59
submitted	5.738823	415	selected	7.787959	57
also	4.334080	403	states	6.899258	49
reviewed	3.926201	393	provide	5.094725	47
determined	4.641768	388	commission	6.829097	43
information	3.598963	354	regulatory	5.190913	40
performed	4.367661	349	will	3.199313	40
violations	6.746733	341	reports	6.870306	37
performance	3.746955	326	document	6.280348	35
facility	4.395778	306	C	15.12165	32

for this group: five of them concern assessment *(identified, reviewed, determined, performed, failed),* and three of them concern the information circuit *(proposed, requested, submitted).* There were also two words that did not fit into any of these groups *(facility, also).*

For the licensee-authored texts, we see a similar pattern of groupings. First, there are four nouns that relate to information, or more specifically documents *(event, report, LER, reports, document).* The reason for including *event* in this group is that we learned in Chapter 4 that one of the most common clusters was *LICENSEE EVENT REPORT LER* (283), and we find this also to be the case for licensee-authored texts: *LICENSEE EVENT REPORT LER* (176). As a result, *event* appears to belong in this semantic set, and we can also infer that the document type *Licensee Event Report* is most frequently within the domain of the *licensee*.

There are also two verbs found in Table 5.20 for the licensee-authored documents that relate to the information circuit *(states, provide)*. These words are the present tense forms of *provided* and *stated*, which were shared by both groups on this level of frequency. There are also several words that relate to a *licensee's* control *(controlled, selected, implement, commitments, will, may)*. Finally, there is a group related to the *NRC (NRC, regulatory, commission)*. The collocation of *licensee* with these words for the licensee group demonstrates how their own identity is semantically linked to those who judge them. This is interesting when put in the context that without the *NRC*, they would not be *licensees*. There was one collocate, *C*, that did not fit into any of these groups. When we look at the clusters for it with *licensee*, we find that most often it is in reference to a specific licensee: *LICENSEE FOR DONALD C COOK NUCLEAR* (20).

In order to better understand the perceived role of the *licensee* by both industry groups at this level, we can organize these collocates into a Venn diagram so that we can observe the differences in perspectives of the NRC and licensee authors (Figure 5.3). The collocates we find residing only in the left circle relates explicitly to the actions of the *licensee*, both those that are a part of the information circuit and those requiring evaluation, as perceived by the NRC. Likewise, all of the words listed only in the circle on the right are associated with the licensee group and relate to control in actions taken by them. Within the shared region of the two circles are those words that were listed for both groups in Table 5.19, as well as the present-tense forms of those verbs that were in Table 5.20 for the licensee industry group. The conclusion we can make from this visual

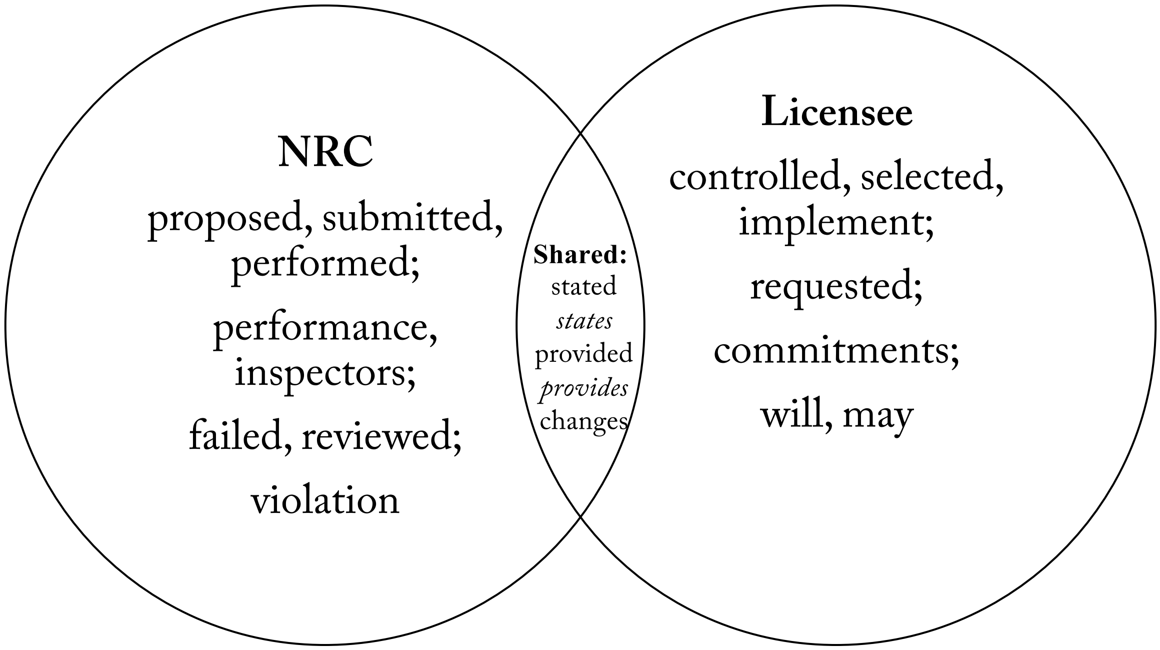

Figure 5.3 Venn Diagram of *Licensee*

representation is that there appears to be an assertion of power with regard to *licensee* for both industry groups.

The licensees provide a picture of the themselves as an entity that asserts itself in an applied sense by way of its choices: *LICENSEE MAY MAKE CHANGES* (7), *LICENSEE INITIATED CHANGES* (6), *LICENSEE SHALL FULLY IMPLEMENT* (12), *SELECTED LICENSEE COMMITMENTS* (50), *LICENSEE COMMITMENTS IS PERFORMED* (6), *CONTROLLED BY THE LICENSEE* (20), *LICENSEE CONTROLLED DOCUMENT* (34). However, we do still see the role of the *licensee* as being the one reporting to the NRC through the presence of *requested: LICENSEE IS REQUESTED TO PROVIDE* (20). The NRC-authored texts provide the complementary perspectives of *licensees* performing actions requiring evaluation, which as we recall has been commonly associated with the NRC for other terms. The verbs in the NRC domain of the Venn Diagram further clarify this evaluative power to be over those choices made, and actions taken, by the *licensee*: *LICENSEE IDENTIFIED VIOLATIONS* (251), *LICENSEE PERFORMANCE DEFICIENCY* (9). As a result of this comparison, we find that the NRC-authored documents associate *licensee* with assessment and **following** the rules/guidelines/standards established by the NRC, while the licensee authors perceive a *licensee* as an entity in relation to the NRC that has control in the **application** of those rules/guidelines/standards.

This lexical profile results in the following Observed Meanings for the term *licensee* for the NRC and licensees as separate industry groups:

1. **For the NRC, a *licensee* is an entity that provides them information, makes changes, and requires evaluation. A *licensee* is assessed according to success and failure, which may result in violations.**
2. **For licensees, a *licensee* provides information and makes requests to the *NRC*. They perceive a *licensee* as also having power with regard to taking action: i.e. making changes and commitments.**

SYSTEM

There are also differences in the meaning associated with *system* between NRC-authored documents and those written by licensee representatives, which we can see through an analysis of the 20 most frequently co-occurring words with the term (Table 5.21). There are eleven collocates shared by both groups in this table *(reactor, coolant, water, cooling,*

Table 5.21: Top 20 Collocates with *System* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	18.87702	1041	*reactor*	5.668736	1628
ADAMS	20.15947	789	coolant	6.714773	982
coolant	20.42340	755	cooling	6.267767	723
water	18.94797	722	control	4.529317	686
cooling	20.04341	659	water	4.701816	682
component	19.98540	567	*safety*	15.15085	680
safety	30.05978	565	*emergency*	4.883533	606
management	19.22709	520	*pressure*	17.55764	601
pressure	18.70116	501	protection	6.041089	527
documents	19.29989	478	*containment*	4.582970	516
access	20.18510	461	component	6.067149	480
document	19.47343	434	core	4.805625	454
control	18.04674	428	ventilation	7.122888	429
emergency	18.54822	415	*power*	3.288905	395
NRC's	19.42673	391	monitoring	4.986623	345
protection	18.35352	355	RCS	5.353648	341
service	30.34813	345	leakage	4.469429	302
accession	20.14330	335	feedwater	5.686651	297
test	30.14168	299	steam	4.104475	289
core	18.52582	298	air	4.911853	288

component, safety, pressure, control, emergency, protection, core). Two semantic sets can be formed from these words. One group concerns the *reactor* and physical *components* on the plant level *(reactor, component, coolant, water, cooling, pressure, core)*, while the other relates to *safety* and *control (safety, emergency, protection, control)*. As a result we can infer that at this highest level of frequency there is a shared understanding of a *system* that most often concerns the physical machinery and components on the plant level in addition to *safety* and maintaining a level of *control* over it: the notion of *system* that emerged in Chapter 4.

When we look at the words not shared by the two groups in Table 5.21 we become aware that although there is shared meaning between them for *system*, there is also variation between them with regard to the context of the frequent use of this term (Table 5.22).

Table 5.22: Collocates with *System* by Industry Group (Not Repeated in Table 5.21)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
ADAMS	20.15947	789	*containment*	4.582970	516
management	19.22709	520	ventilation	7.122888	429
documents	19.29989	478	*power*	3.288905	395
access	20.18510	461	monitoring	4.986623	345
document	19.47343	434	RCS	5.353648	341
NRC's	19.42673	391	leakage	4.469429	302
service	30.34813	345	feedwater	5.686651	297
accession	20.14330	335	steam	4.104475	289
test	30.14168	299	air	4.911853	288

The NRC list in Table 5.22 appears to primarily concern the NRC's own *system*, which exists to facilitate their role of providing the public *access* to *documents* through the *NRC's ADAMS* database *(ADAMS, documents, document, access, accession, NRC's, management): ADAMS IS ACCESSIBLE FROM* (287), *NRC'S AGENCYWIDE DOCUMENT ACCESS AND MANAGEMENT SYSTEM ADAMS* (361). There were two words that did not fit into this group *(service, test)*. By analyzing the clusters formed with *service* and *system,* we find that this collocate provides the added meaning to this term of a specific type of *system* on the plant level, as well as the importance of *systems* on the plant level being functional: *SERVICE WATER SYSTEM* (200), *THE SERVICE WATER SYSTEM* (64), *ESSENTIAL SERVICE WATER SYSTEM* (21), *SYSTEM WAS*

OUT OF SERVICE (8), *SYSTEM RETURNED TO SERVICE* (6). The presence of *test* in this list also re-emphasizes the importance of assessment from the perspective of the NRC, even with regard to a *system*: *SYSTEM LEAKAGE TEST* (90), *SYSTEM FUNCTIONAL TEST* (42).

On the other hand, the point of view provided by licensee-authored documents in Table 5.22 provides a more specific meaning to a *system* on the plant level. We are made aware that licensee authors frequently write about specific systems used for components and resources like *RCS, feedwater, steam,* and *air: AUXILIARY FEEDWATER SYSTEM* (75), *MAIN FEEDWATER SYSTEM* (27), *THE MAIN STEAM SYSTEM* (49), *STEAM SUPPLY SYSTEM* (40), *BOTTLED AIR SYSTEM* (35), *AIR CONDITIONING SYSTEM* (27). The presence of the word *power* allows us to understand licensee-authored documents frequently express that some systems are in place to distribute *power* on the plant level for both onsite and offsite use*: THE OFFSITE POWER SYSTEM* (27), *ELECTRIC POWER SYSTEM* (30), *POWER DISTRIBUTION SYSTEM* (27), *ONSITE POWER SYSTEM* (17).

From the licensee perspective, the importance of standards and protection are related to *system* quite often—both in the function of a *system* for this purpose, as well as the monitoring of it *(containment, leakage, monitoring, ventilation): CONTAINMENT SPRAY SYSTEM* (121), *CONTAINMENT ISOLATION SYSTEM* (26), *SYSTEM LEAKAGE TEST* (57), *COOLANT SYSTEM LEAKAGE* (28), *RADIATION MONITORING SYSTEM* (44), *FLOW MONITORING SYSTEM* (28), *MONITORING SYSTEM TO OPERABLE STATUS* (10), *EMERGENCY VENTILATION SYSTEM* (63), *BUILDING VENTILATION SYSTEM* (49). There is another word in Table 5.22

for the licensee group that indirectly indicates the frequent relationship between a *system* and protection: *room*. While this relationship might not be clear at first, we can look at the clusters for this collocate with *system* to realize that there are frequent references to specific systems in place to protect and service the *control room* where operators make decisions about the operation of the nuclear reactor: *CONTROL ROOM EMERGENCY AIR SYSTEM* (36), *CONTROL ROOM EMERGENCY VENTILATION SYSTEM* (43), *CONTROL ROOM POST ACCIDENT RECIRCULATION SYSTEM* (27), *CONTROL ROOM MAKEUP AND CLEANUP FILTRATION SYSTEM* (15). All of these collocates help us to understand that the most frequently used words with *system* by licensee authors indicate that many different types of systems are in place at the plant level to perform many different jobs.

This lexical profile results in the following Observed Meanings for the term *system* for the NRC and licensees as separate industry groups:

1. **For the NRC, a *system* is most often related to the NRC's Agencywide Document Access and Management System (ADAMS). Otherwise, this term is linked to the notion of *safety*, especially with regard to those components and elements connected with a reactor.**
2. **For licensees, a *system* most directly relates to machines (i.e. reactor), components, and elements that are connected together to serve many different functions. *System* is also often linked with safety, especially the application of safety through emergency and protection.**

POWER

In Table 5.23, we can see that there are both shared meanings for *power* between the two industry groups as well as different ones.

Table 5.23: Top 20 Collocates with *Power* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	6.647987	2484	*nuclear*	6.226360	1407
plant	5.467163	1226	thermal	7.064047	876
station	6.583056	741	*plant*	18.17804	861
plants	12.30633	680	station	5.978815	823
offsite	7.476901	627	*reactor*	4.712626	782
reactor	4.634240	560	loss	5.885093	664
unit	4.692180	507	offsite	17.79231	659
operation	5.384611	484	*unit*	4.159016	632
reactors	17.48225	482	level	5.315097	601
percent	6.250718	440	full	6.706079	585
loss	5.813355	429	operation	4.844094	544
level	5.262738	417	core	5.122487	527
uprate	8.495543	400	uprate	8.193281	509
company	6.374261	385	rated	7.848259	477
full	7.020959	373	company	6.484894	415
thermal	6.752172	355	system	3.288905	395
light	8.044551	351	distribution	7.032396	377
electric	6.871244	335	percent	6.338442	374
electrical	6.758879	329	supply	6.659842	364
core	5.293728	323	plants	6.953756	357

There are 16 collocates shared by the two groups, which indicates that on the highest level of frequency there is significant consensus between the two groups as to how they write about *power (nuclear, plant, station, plants, offsite, reactor, unit, operation, percent, loss, level, uprate, company, full, thermal, core)*. First of all, we see that there is consensus in the frequent discussion of different types of power in this industry *(nuclear, thermal)*. Also, many of the nouns from this list come together to create a semantic set connoting where *power* is made in this industry *(plant, station, plants, unit)* through the *operation* of certain equipment *(reactor, core)*. Another group can be made with those words related to the characteristic also observed in Chapter 4 about *power* being something that is measured *(percent, loss, level, uprate, full)*. Finally, there appears to be frequent use, in the language of both groups, of *power* as a branding tool through its collocation with *company*.

Despite all of these shared meanings for *power* by both industry groups, we can see that there are also some different relationships being made for the meaning of this term between them (Table 5.24).

Table 5.24: Collocates with *Power* by Industry Group (Not Repeated in Table 5.23)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactors	17.48225	482	rated	7.848259	477
light	8.044551	351	*system*	3.288905	395
electric	6.871244	335	distribution	7.032396	377
electrical	6.758879	329	supply	6.659842	364

From the perspective of NRC-authored documents, we see an increased use of words with *power* that indicate this term is used most often when referring to licensees and other organizations who use *power* as a branding tool for identification through the product they make, sell, or are involved with in some capacity *(light, electric): POWER LIGHT COMPANY* (134), *POWER AND LIGHT COMPANY* (90), *ELECTRIC*

POWER RESEARCH INSTITUTE (60), *ELECTRIC AND POWER COMPANY* (45). However, *electric* and *electrical* are also used by this group to describe the kind of power used on the plant level: *DC ELECTRICAL POWER* (84), *AC ELECTRICAL POWER* (46), *ELECTRICAL POWER DISTRIBUTION* (56), *ELECTRIC POWER SYSTEMS* (56), *REQUIRES THAT ELECTRIC POWER* (17). We find that the last of the 20 most frequently co-occurring words with *power* for NRC-authored documents that is not also present in this same list for the licensee group is *reactors.* While the singular version of this term was shared by both groups in this same list, we see that the NRC frequently refers to more than one *reactor* when talking about power: *NUCLEAR POWER REACTORS* (291), *COMMERCIAL NUCLEAR POWER REACTORS* (185), *SAFE OPERATION OF COMMERCIAL NUCLEAR POWER REACTORS* (180), *STANDARDS FOR POWER REACTORS* (116).

There is a definite theme underlying three of the words that most frequently co-occur with *power* for licensee-authored texts, as listed in Table 5.20: business. When we combine the collocates *distribution, supply,* and *system* we find that *power* is often treated as a product, a commodity even, from the licensee perspective. It is a resource to be distributed and allocated throughout the *plant*: *THE POWER SUPPLY* (61), *CORE POWER DISTRIBUTION* (51), *POWER. DISTRIBUTION SYSTEM* (26), *POWER DISTRIBUTION MEASUREMENTS* (17), *STANDBY POWER SUPPLY* (8). Moreover, it is the reason why the licensees are in business: *TO SUPPLY POWER* (15), *SUPPLY THE POWER* (11). Finally, power is also related to the notion of measurement with the presence of *rated: OF RATED THERMAL POWER* (77), *AT RATED THERMAL POWER* (31), *PERCENT RATED THERMAL* (30).

This lexical profile results in the following Observed Meanings for the term *POWER* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, *power* is a part of the identities of the parties they work with. However, they also associate this term with resources needed for operating *reactors*.**
2. **For the authors of documents for licensees, *power* is not only who they are, it is a product they manufacture as part of their business. It is also a resource, or commodity, that is distributed throughout the plant and measured.**

NUCLEAR

When the 20 most frequently co-occurring words with the term *nuclear* for each industry group are juxtaposed, we notice some interesting similarities and differences (Table 5.25). We can see from this table that there are 14 words that are shared between both industry groups when we analyze collocate frequencies at this level *(commission, power, regulatory, plant, NRC, station, plants, unit, units, operating, LLC, generating, company, Point).* First and foremost, both groups frequently pair *nuclear* with *regulatory, commission,* and *NRC* to refer to the *NRC* in their writing, and similarly there is the shared use of this term as a branding tool for companies to identify themselves *(LLC, company, Point).* However, the biggest group that we can make from these words demonstrates the connection between *nuclear* and the production of power *(power, plant, station, plants, unit, units, generating, operating);* this is not surprising as both groups exist because of the use of *nuclear power* in this way.

Table 5.25: Top 20 Collocates with *Nuclear* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
commission	7.690862	2489	*plant*	19.90082	1418
power	6.647987	2484	*power*	6.226360	1407
regulatory	7.610242	2305	station	7.209961	964
plant	6.055479	1663	regulatory	7.264206	815
NRC	4.609693	1197	commission	8.712359	768
station	7.281693	1085	*unit*	4.788978	488
plants	12.70542	809	*NRC*	17.07503	400
unit	4.989252	562	plants	8.046837	380
units	14.96578	532	company	7.225561	346
reactor	4.627859	503	*safety*	15.09888	305
staff	3.953249	455	generating	7.871219	291
operating	4.876705	428	*operating*	16.59608	287
office	7.017146	419	units	6.055472	287
Entergy	7.182915	401	energy	6.155868	217
LLC	7.371359	396	*fuel*	4.037975	194
generating	7.703166	391	LLC	7.933991	192
company	6.461511	369	Point	5.873342	170
regulation	8.047999	352	operations	5.442353	162
reactors	17.10587	335	management	5.519775	149
Point	6.450542	323	Ferry	8.578575	139

Despite the many similarities in the 20 words that most frequently co-occur with *nuclear* for each of these industry groups, there are also those words that they do not share in Table 5.25 (Table 5.26). When looking at the words in this table for the NRC-authored documents, *nuclear* and *reactor* frequently co-occur with *staff* and *office* to identify who they are and what they do: *NUCLEAR REACTOR REGULATION* (353), *OFFICE OF NUCLEAR REACTOR REGULATION* (349). Meanwhile, *reactors* often co-occurs with *nuclear* in the language of this group to provide specificity to what it is they are helping to supervise: *SAFE OPERATION OF COMMERCIAL NUCLEAR*

Table 5.26: Collocates with *Nuclear* by Industry Group (Not Repeated in Table 5.25)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	4.627859	503	*safety*	15.09888	305
staff	3.953249	455	energy	6.155868	217
office	7.017146	419	*fuel*	4.037975	194
Entergy	7.182915	401	operations	5.442353	162
regulation	8.047999	352	management	5.519775	149
reactors	17.10587	335	Ferry	8.578575	139

POWER REACTORS (180), *IMPROVEMENTS FOR NUCLEAR POWER REACTORS* (18). Along the same semantic lines, this term also collocates with two words that specifically refer to the fact that these reactors they regulate are owned and operated by specific businesses *(Entergy)*.

Both groups are also talking about specific licensees in these documents, but they are talking about them at different rates. From Tables 5.25 and 5.26 we can see that the NRC-authored documents make the most references to those licenses held by *ENTERGY NUCLEAR OPERATIONS INC* (130), in addition to the shared references to Indian *Point*, Nine Mile *Point*, *Point* Beach, and/or Turkey *Point* by both groups—it should be noted that Indian Point is also owned and operated by Entergy Nuclear Operations, Inc. Whether or not these numerous mentions of *Entergy Nuclear Operations Inc.* are utilized in a positive or negative manner is not part of this study, however analyzing the prosody of this cluster is would be an interesting future study.

We can also see in Table 5.26 that, in addition to *Point*, the licensee-authored documents also have a tendency to relate *nuclear* to a specific licensee *(Browns, energy,* *operations, management, fuel)*. In this case, it is *BROWNS FERRY NUCLEAR PLANT* (116), *NUCLEAR MANAGEMENT COMPANY* (89), *GLOBAL NUCLEAR FUEL* (35), *THE NUCLEAR ENERGY INSTITUTE* (31), *INSTITUTE OF NUCLEAR POWER OPERATIONS* (15), *NORTHEAST NUCLEAR ENERGY* (26), *NUCLEAR ENERGY COMPANY* (25), and/or *GE NUCLEAR ENERGY* (16). Licensee authors also frequently relate *nuclear* to *operations* and *management* in reference to specific employees: *PRESIDENT NUCLEAR POWER OPERATIONS* (49), *NUCLEAR SECURITY MANAGEMENT* (7).

Despite the inclination of both groups to use *nuclear* to identify people, places, and organizations, the licensee group also frequently uses this term to describe a specific type of *safety* that is important to them: *NUCLEAR SAFETY RELATED* (32), *THE NUCLEAR SAFETY ASSESSMENT BOARD* (36), *NUCLEAR SAFETY REVIEW* (23), *AFFECT NUCLEAR SAFETY* (19). *Fuel* is another collocate that, in addition to its use with *nuclear* for branding and identification, is used by licensees to describe specific types of *fuel* on the plant level: *SPENT NUCLEAR FUEL* (23), *OF NUCLEAR FUEL* (15), *NUCLEAR FUEL CYCLE* (13). It becomes clear through this analysis that *nuclear* is most often used for branding and identification by licensee authors, but it also ascribes specificity to employee job titles, *safety,* and even *fuel.*

This lexical profile results in the following Observed Meanings for the term *nuclear* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, *nuclear* is the defining quality of what they regulate, and it also is part of their identity with regard to**

their departments or offices as well as the identities of those entities they regulate.

2. **For the authors of documents for licensees, *nuclear* is not only the type of fuel they use to generate electricity, but it is a type of safety they must be concerned with, it provides specificity to the job titles for certain employees, and it is part of many organizations' identities within this industry.**

SUMMARY OF INDUSTRY GROUP PROFILING

This chapter has analyzed the lexical profiles of the top 10 most frequently occurring industry key terms from Chapter 4. For each of these terms, the 20 most frequent collocates were identified and the resulting lexical profiles and Observed Meaning for each key term were examined with regard to industry group. We were able to conclude that the differences between each group's lexical profile were directly related to the roles of each group in this industry. In other words, domain specific language is affected by group membership. This particular observation is of use to both the general public, who needs to know and understand that industry group membership reflects its role in the industry when reading documents and correspondence from this industry, and those individuals in the nuclear industry whose job is directly related to communicating with/for another industry group.

This perspective also afforded us the opportunity to observe that those meanings we observed in Chapter 4 for the industry key terms emerged as an aggregate of the collocates most frequently used by the industry groups. In order to see if this behavior also applies to the geographical dimension, the same procedure will be performed in the following chapter. Therefore, the results from Chapter 5 will serve as the baseline for the regional implications uncovered in Chapter 6.

CHAPTER 6

REGIONAL VARIATION IN KEY TERM COLLOCATIONS

For the variety of regulated corporate discourse that is modeled in the nuclear reference corpus, it is quite clear that language, and by extension meaning, varies with regard to the author of each document being affiliated either with the NRC or a licensee. However, organizational membership is not the only way in which the nuclear industry is divided. Where a licensee is located geographically is another dimension through which we might observe variation in the ways language is used, i.e. to conclude that the emergence of meaning that was observed in Chapter 5 is actually a characteristic of any sub-grouping of domain-specific language, or to uncover patterns that might be of use within the nuclear industry with regard to differences in the interaction of licensees and NRC representatives across the industry. The patterns we observe that emerge throughout the analysis in this chapter will not only help us to understand the different concerns of licensee and NRC staff with regard to their location, but also will help them to identify potential linguistic best practices utilized in different regions or within their own.

THE FOUR NRC REGIONS OF THE U.S. NUCLEAR POWER INDUSTRY

The NRC has divided all 104 reactors in the United States into four separate geographic regions:

1. Region I offices are located in King of Prussia, Pennsylvania, and it oversees all nuclear facilities in Connecticut, Delaware, Maine, Maryland, Massachusetts,

New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and Washington D.C.;

2. Region II offices are located in Atlanta, Georgia, and it oversees Alabama, Florida, Georgia, Kentucky, Mississippi (except for Grand Gulf, which is overseen by Region IV), North Carolina, Puerto Rico, South Carolina, Tennessee, Virginia, the Virgin Islands, and West Virginia;
3. Region III offices are located in Lisle, Illinois, and it oversees Illinois, Indiana, Iowa, Michigan, Ohio, and Wisconsin; and
4. Region IV offices are located in Arlington, Texas, and it oversees all nuclear facilities in Alaska, Arizona, Arkansas California, Colorado, Hawaii, Idaho, Kansas, Louisiana, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, South Dakota, Texas, the U.S. Pacific Territories, Utah, Washington, and Wyoming.

Although each of these regions is responsible for overseeing a vastly different number of states—i.e. Region 3 is responsible for six states while Region 4 regulates facilities in 22 states in addition to the U.S. Pacific Territories—the land area each office regulates is not indicative of the number of nuclear reactors within its jurisdiction (Table 6.1)

Table 6.1: Number of Reactors by NRC Region

	Region I	Region II	Region III	Region IV
Number of Reactors	26	33	24	21

All of the regions have approximately the same number of reactors (mean = 26, standard deviation = 5.099). In the same way we analyzed the metadata of the documents in the reference corpus for author organizational affiliation we can also calculate the number of documents in the corpus that are assigned to the docket numbers, or unique identifying numbers for each reactor, by region (Table 6.2).

Table 6.2: Number of Documents in Nuclear Reference Corpus by NRC Region

	Region I	Region II	Region III	Region IV
Number of Documents	1,763	2,039	1,650	1,598
Number of Documents (Normalized by # of Reactors)	67.808	61.788	68.750	76.095

It was explained in Chapter 3 that sampling for the corpus was performed by docket number, so each reactor had the same number of documents sampled from the ADAMS database. However, documents sampled for one docket number were found to also be assigned to other dockets. Looking at Table 6.2 we might think initially that the varying number of documents for each region falls in line with this sampling process and that Region II having the greatest number of documents is to be expected since it has the greatest number of reactors. While this is the case, when we normalize the number of documents by the number of reactors (mean = 68.610, standard deviation = 5.866) we see that more documents were assigned to the dockets of reactors in Region 4 in comparison to the rest of the regions, and fewer documents were assigned to reactors in Region 2 than to the rest of the regions. This particular observation is important because the fact that Region 4 has more documents assigned to dockets across the industry could possibly indicate that the language used by these licensees has more influence across the industry through increased exposure, and likewise Region 2 could possibly have less influence.

REGIONAL KEY TERM ANALYSES

Using the same process that was described in Chapter 4, I performed a key word analysis for each region in comparison to the rest to identify those industry key terms that are used more or less than expected to a point of statistical significance. I looked for those key terms that were statistically significant at p-value 0.01, or with 99.0% confidence (Table 6.3).

Table 6.3: Key Industry Terms for All Regions p<0.01

Region I		Region II		Region III		Region IV	
Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
license	inspection	fuel	inspectors	inspection	emergency	emergency	fuel
nuclear	inspectors	unit	license	inspectors	staff	safety	nuclear
staff	licensee		nuclear	licensee	unit	technical	unit
NRC	plant		emergency	plant		licensee	
	fuel		reactor	reactor			
	operating		safety	system			
	emergency		technical				

The variation in the use of the industry key terms with regard to region has practical value for industry executives, plant personnel, and NRC officials who regularly move regions—and as a result have to change their focus.

For Region I (Figure 6.1), we find that the documents assigned to reactors in this part of the country use *license, nuclear, staff, and NRC* more than expected in comparison to all of the other regions.

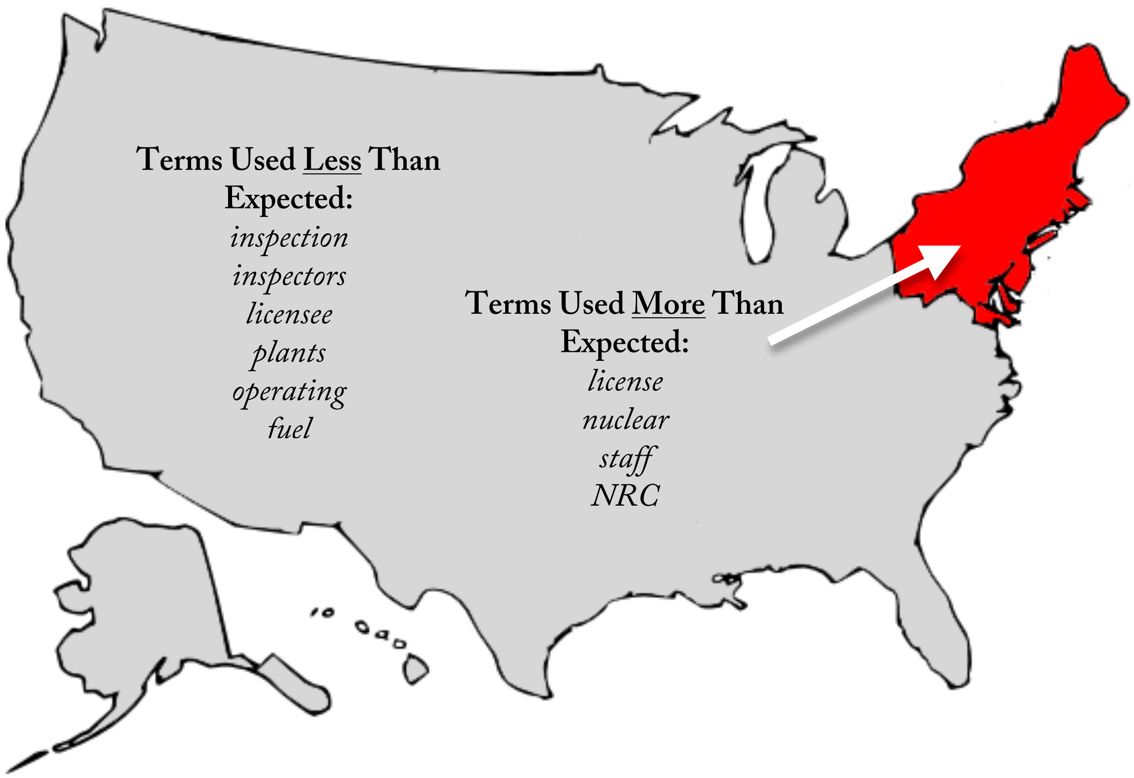

Figure 6.1: Key Terms for Region I (p<0.01)

On the other hand they used *inspection, inspectors, licensee, plants, operation, fuel,* and *emergency* less than expected based off of those documents assigned to dockets in the other regions. The focus of communication in this region is primarily on the NRC and its identity as regulator and administrator over the use of nuclear materials. The frequent use of these terms is unique to Region I. We find that the language used in documents designated as belonging to dockets in Region II has a different focus (Figure 6.2).

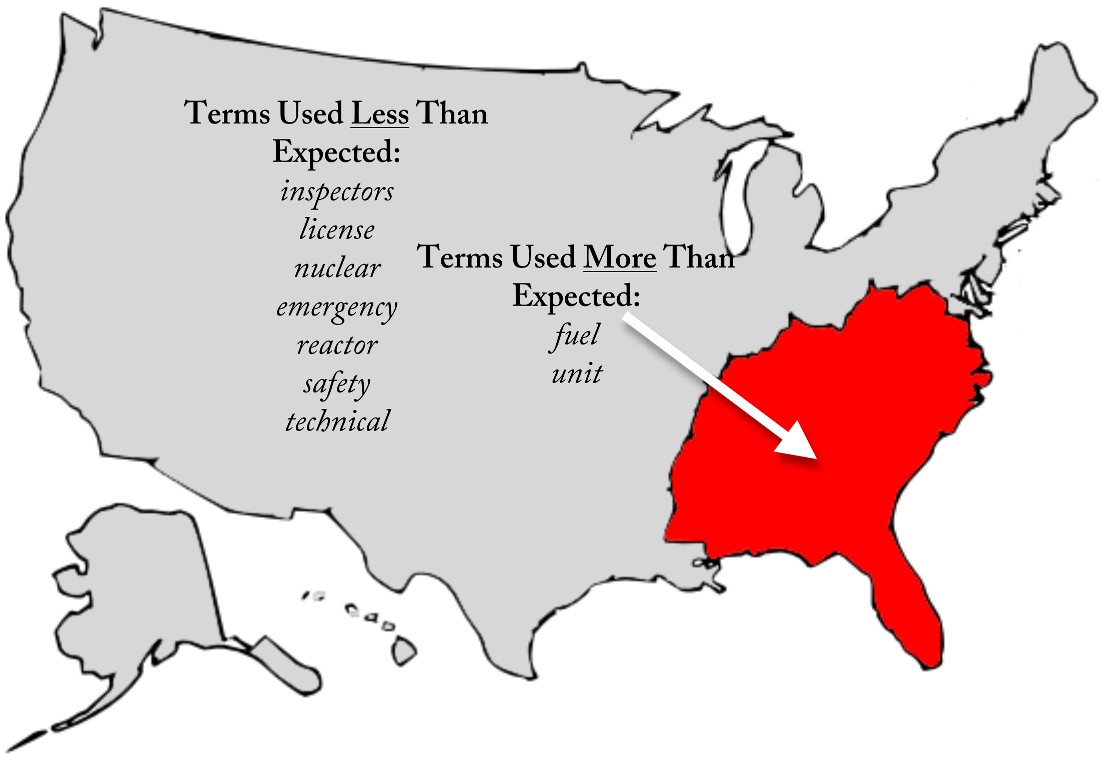

Figure 6.2: Key Terms for Region II (p<0.01)

For this region, *fuel* and *unit* are those words used more often than the documents assigned to the dockets of Regions I, III, and IV. This frequency could possibly be accidental, except when we find that *inspectors, license, nuclear, emergency, reactor, safety,* and *technical* were used far less often. What this helps us to see is that there is a shift in the focus of this region, in comparison to the other three, from those terms related to the regulation of the industry to that of the physical plant site.

In Region III (Figure 6.3), we find that they do not share Region II's frequent use of *unit* and *inspectors*, but they do share Region I and II's infrequent use of *emergency*.

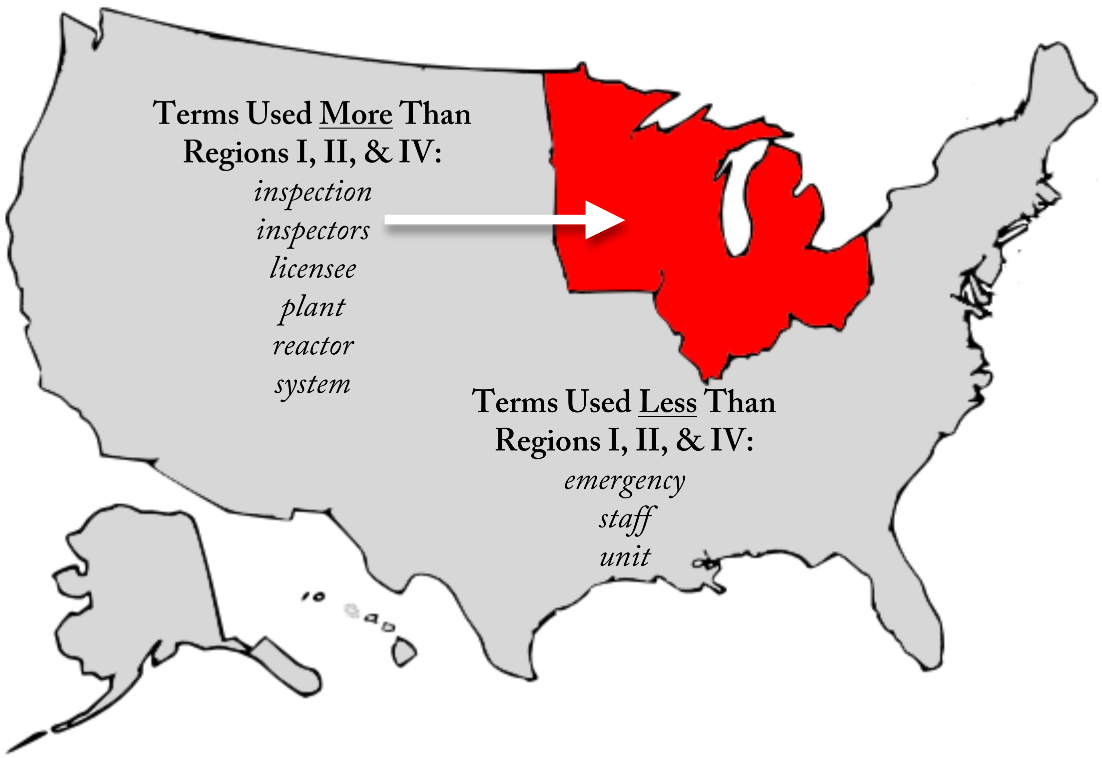

Figure 6.3: Key Terms for Region III (p<0.01)

Region III also seems to be involved in more communication involving the *plant, reactor,* and *system* than the other regions, as well as the assessment of these things by the NRC *(inspection, inspectors).*

Documents assigned to dockets in Region IV demonstrate a propensity for using *emergency, safety, technical,* and *licensee*, while *fuel, nuclear,* and *unit* are used far less (Figure 6.4).

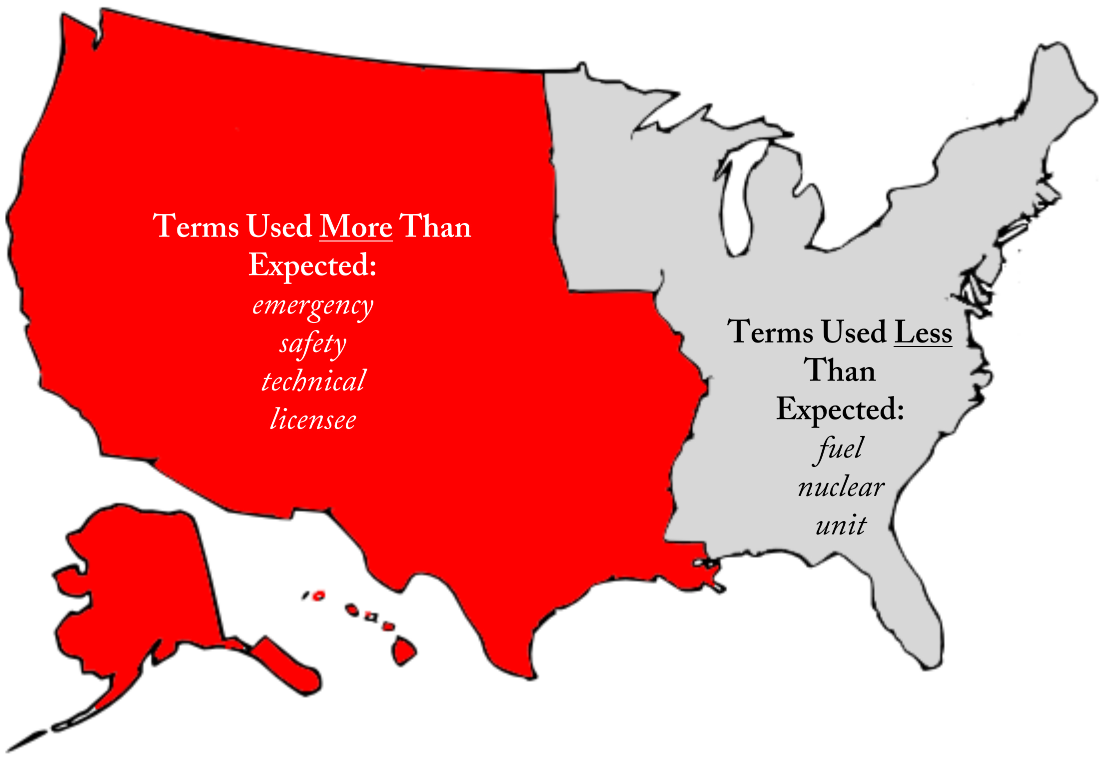

Figure 6.4: Key Terms for Region IV (p<0.01)

The communication associated with this region is focused more on *safety* and *emergency* than is to be expected in comparison to Regions I, II, and III. There also seems to be an increased focus on plant standards (or *technical specifications*), as well as the role of plants as *licensees*. The frequent use of these terms by plants in this region is a characteristic that could be of use to public relations departments in licensee organizations, as well as the NRC.

From Figures 5.1, 5.2, 5.3, and 5.4 we are able to see that documents assigned to docket numbers within each of the four NRC regions have different propensities for using the 20 industry key terms identified in Chapter 4. However, when we juxtapose the terms that occur more and less than expected in each region along with those that were used within expected ranges, we gain a better perspective (Table 6.4). Despite Region I's focus on the regulatory aspect of the industry, Region II's frequent use of *unit* and *fuel,* a greater amount of discussion about things related to *inspections* in Region III, and an increased focus on *safety* and *emergency* issues in Region IV, we can see in this table that *power, containment, evaluation,* and *pressure* are the four industry key terms that are used within expected ranges by all four regions. Or in other words, there is no difference in the frequency of occurrence of these terms with regard to the regional affiliation of the docket number to which the document is assigned.

KEY TERM ANALYSIS FOR NRC REGIONS BY INDUSTRY GROUP

We can see in Figure 6.5 that the use of the key terms in documents assigned to Region I that were NRC-authored had a statistically significant overuse of *NRC*, *nuclear,* and *staff* in comparison to the rest of the NRC-authored documents in the remaining regions.

Table 6.4: Key Term Distribution by Region

	Region I	Region 2	Region 3	Region 4
Occurred More than Expected	NRC staff nuclear license	unit fuel	inspection inspectors licensee plant reactor system	emergency technical licensee safety
Occurred Less than Expected	fuel inspection inspectors plant operating	emergency nuclear inspectors technical license reactor safety	emergency staff unit	fuel nuclear unit
Occurred within Expected Ranges	safety reactor unit licensee system **power** **containment** evaluation technical **pressure**	NRC plant inspection licensee system **power** **containment** operating **evaluation** staff **pressure**	NRC safety **power** nuclear fuel license **containment** operating **evaluation** **pressure**	NRC reactor plant inspection system **power** nuclear license **containment** operating **evaluation** staff **pressure** inspectors

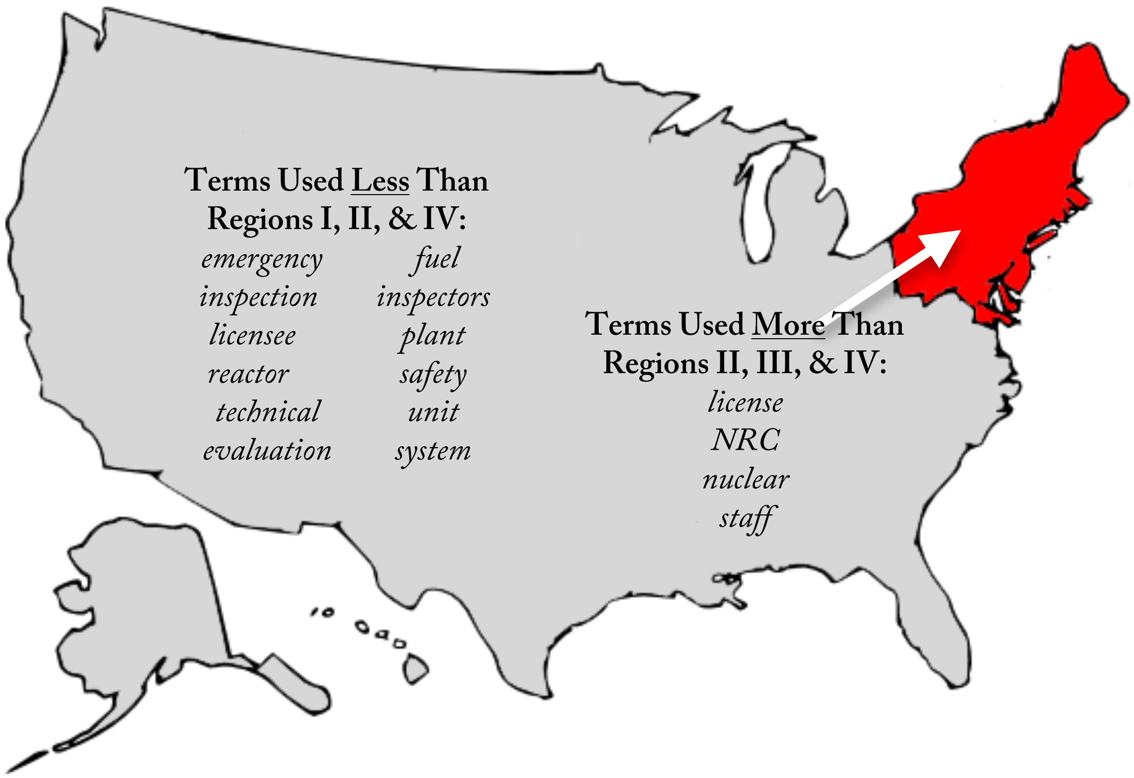

Figure 6.5: Key Terms for NRC-Authored Documents in Region I (p<0.01)

Furthermore, *emergency, fuel, inspection, inspectors, licensee, plant, reactor, safety, technical, unit, evaluation,* and *system* were used far less than expected in comparison to the other regions. As a result, we can see that a majority of the usage of these terms comes from NRC representatives in this region, and we now know that this industry group in this region is the one writing most about the *NRC* as a regulatory entity and least about those terms directly related to the plant site.

Those documents designated as Region II that were also authored by NRC representatives exhibit a different pattern from Region I (Figure 6.6). Namely, for documents authored by the NRC for this Region, we observe a change in the focus to

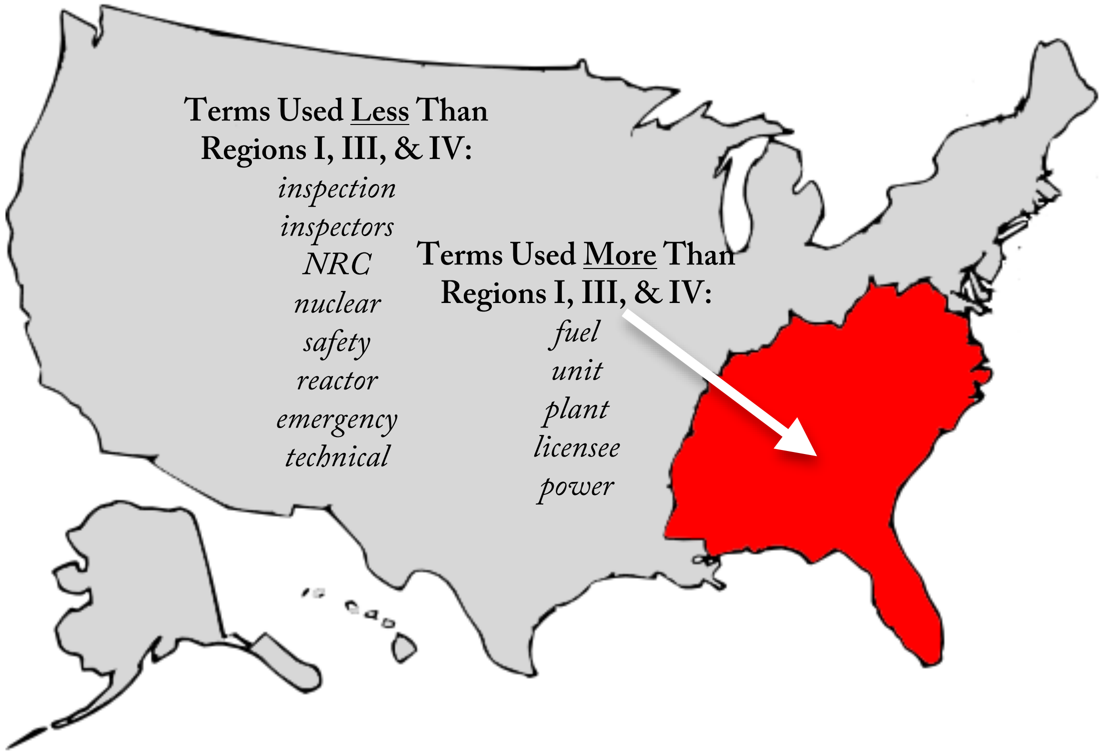

Figure 6.6: Key Terms for NRC-Authored Documents in Region II (p<0.01)

fuel, unit, plant, licensee, and *power* more than the NRC language assigned to reactors in the other regions: those terms directly relating to plant operations. They also are represented as having less communication about *inspection, inspectors, NRC, nuclear, safety, reactor, emergency,* and *technical* in comparison to Regions I, III, and IV.

In Figure 6.7 we notice that Region III also has an idiosyncratic distribution for statistically significant usage of key terms. We find that those terms related to the regulation of nuclear power on the plant level, or in practice—*inspection, inspectors, licensee, plant, reactor, evaluation, power, pressure, safety,* and *system*—are all used more than

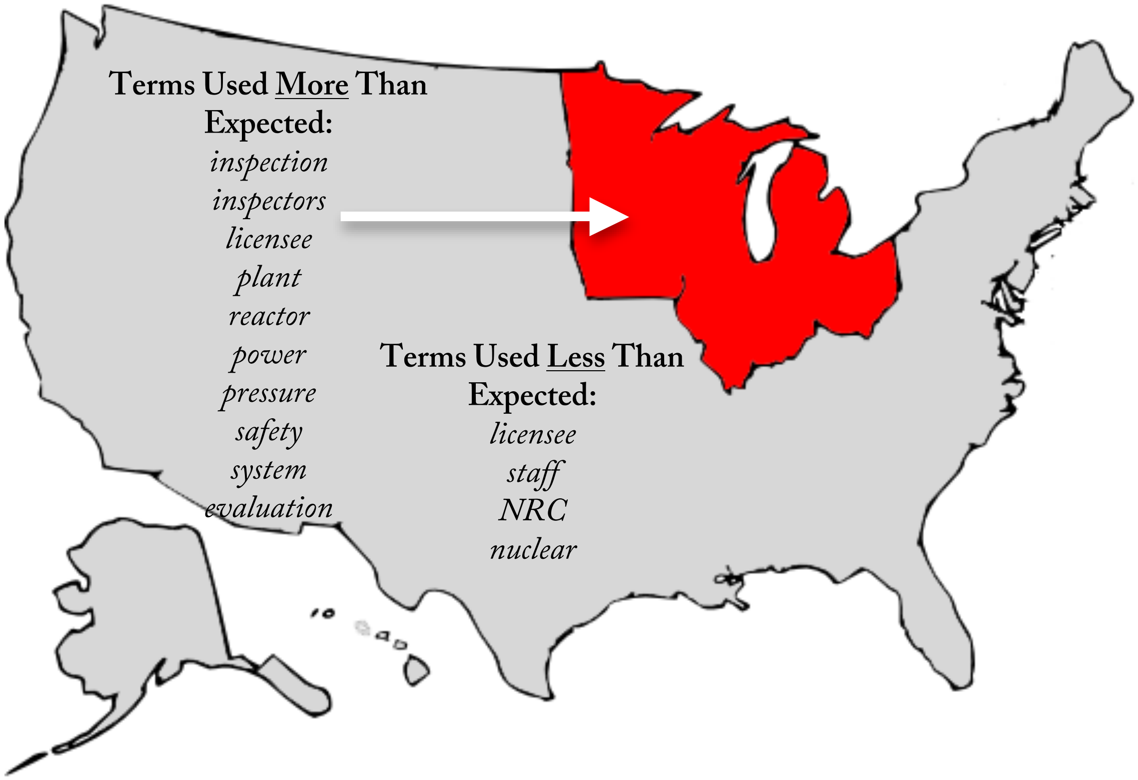

Figure 6.7: Key Terms for NRC-Authored Documents in Region III (p<0.01)

expected in these documents. On the other hand, all of those terms used more in Region I *(license, staff, NRC,* and *nuclear)* that correspond to the regulatory nature of the *NRC* conceptually are used less than expected based off of the documents assigned to the other regions.

What we can see in the documents assigned to the dockets of this region is an increased communicative focus on the evaluation and assessment of the plant and its components, and less focus on the NRC and its administration over licenses. Region IV (Figure 6.8) has a similar focus on assessment, but more so with plant-level standards and

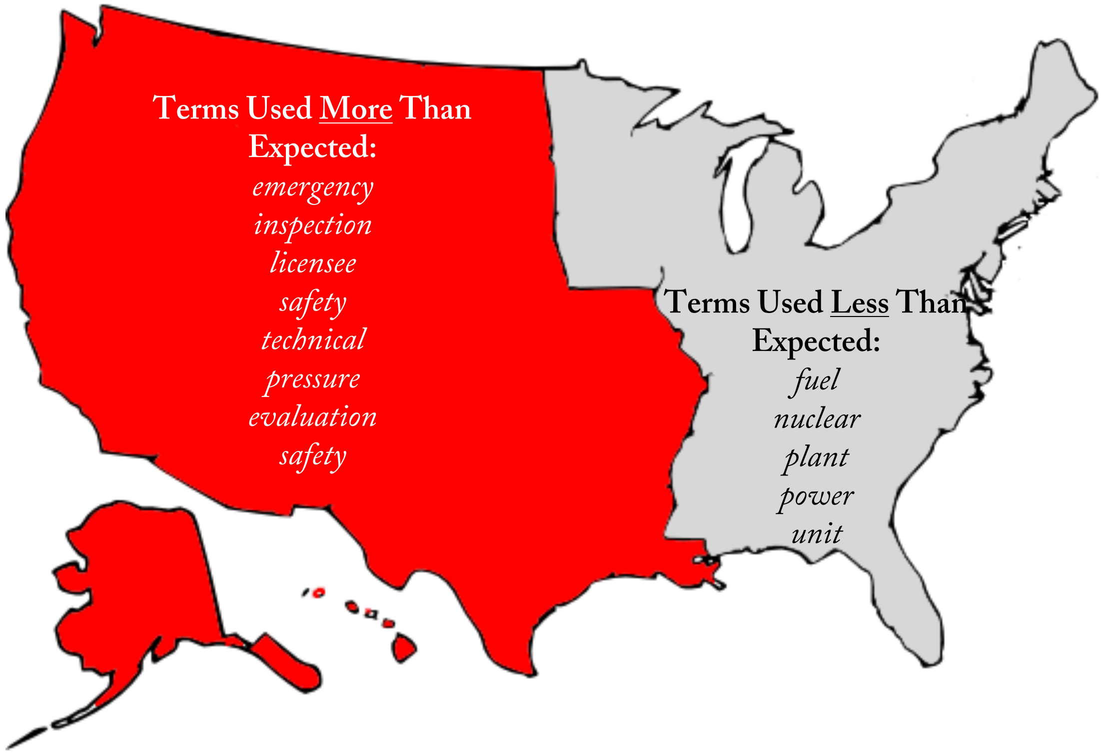

Figure 6.8: Key Terms for NRC-Authored Documents in Region IV (p<0.01)

the implementation of safety *(emergency, inspection, licensee, safety, technical,* and *pressure),* and a decreased focus on *fuel, unit, plant,* and *power.*

There are a few observations to be made about these key term distributions between the four regions for NRC-authored documents (Figure 6.9). First of all, there appears to be a divide between Regions I and II and Regions III and IV with regard to the use of *inspection* and *safety* in those documents assigned to them. The NRC-authored documents assigned to the dockets in Regions III and IV appear to use these words significantly more than Regions I and II. We can also see that documents written by the

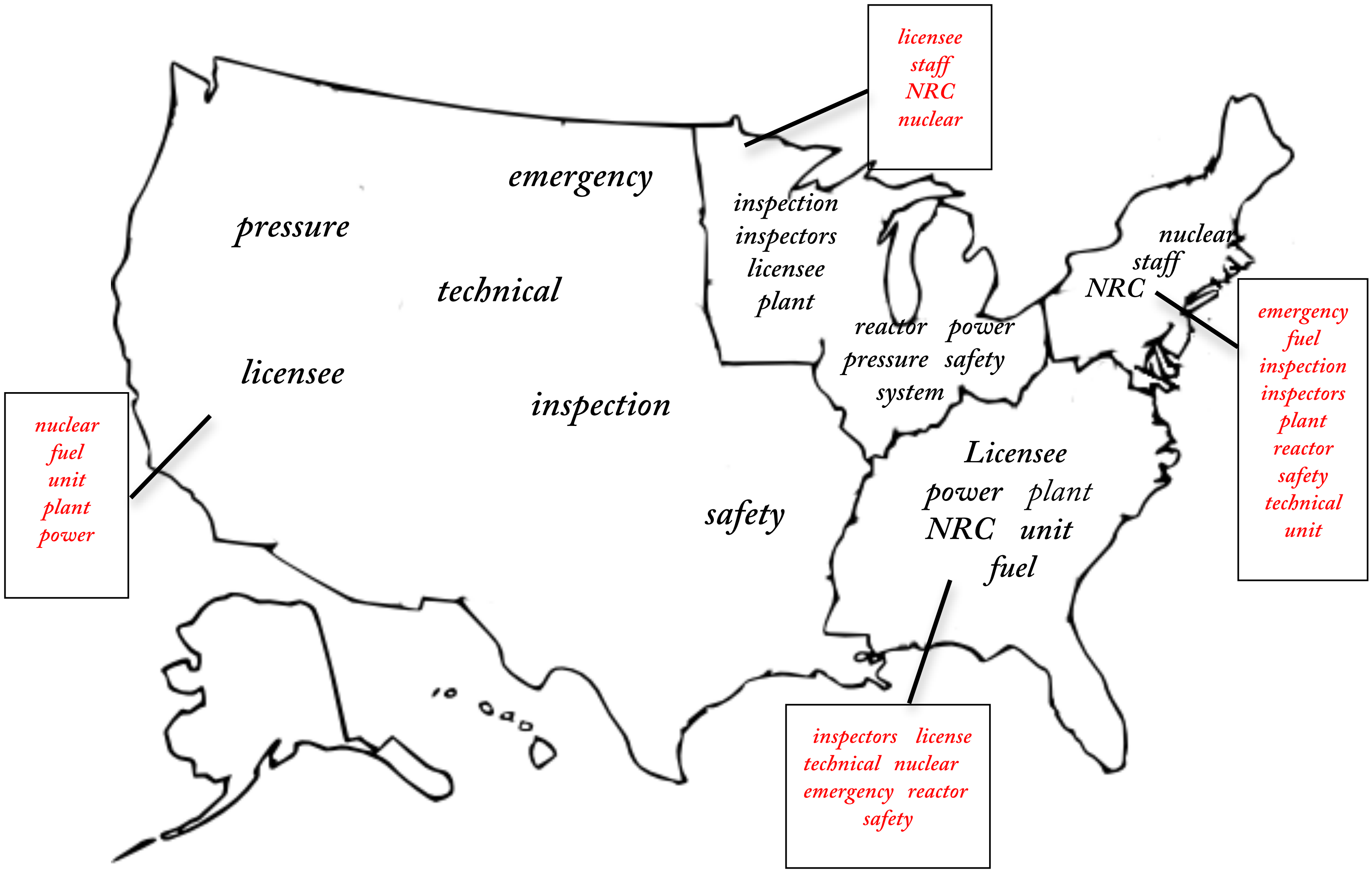

Figure 6.9: Key Terms for NRC-Authored Documents in All Regions ($p<0.01$)

NRC and assigned to reactors belonging to licensees regulated by Regions II and III use *plant* and *power* to the greatest degree of all four regions.

We also see that the NRC documents assigned to Region IV are most frequently referring to *emergency* and *technical* in comparison to documents authored by the same industry group for the other regions, while the same can be said about Region III regarding *inspectors* and *reactor*. We also find that the documents affiliated with NRC authorship and assigned to Region II are most concerned about fuel, while Region I is characterized by a similar abundance of the use of *NRC* and *nuclear*, which makes sense since the *NRC* headquarters are in Region I. With regard to Regions I and III, they are characterized by an opposition in the significant over-use (Region I) and under-use (Region III) of *staff*, while the reverse is true for *evaluation*: positive key term designation for Region III and negative for Region I. If we now go over and analyze the regional use of the key terms with regard to licensee authorship, we are able to observe a different pattern in the frequency of key terms used than what was noted for NRC-authored documents.

First of all, we find that licensee-authored documents assigned to dockets in Region I are characterized by a greater use of *license, technical, emergency,* and *unit* (Figure 6.10). There is also a statistically significant under use of *inspection*, *licensee*, *NRC*, *operating*, and *plant*. Licensee-authored documents designated as being part of regulatory communication for reactors in Region II demonstrate a greater use of *inspection, NRC*, *unit*, *fuel*, and *pressure*, while also having a lesser degree of use of *emergency*, *license*, *technical*, *unit*, *containment*, and *safety* than the documents authored by the same industry

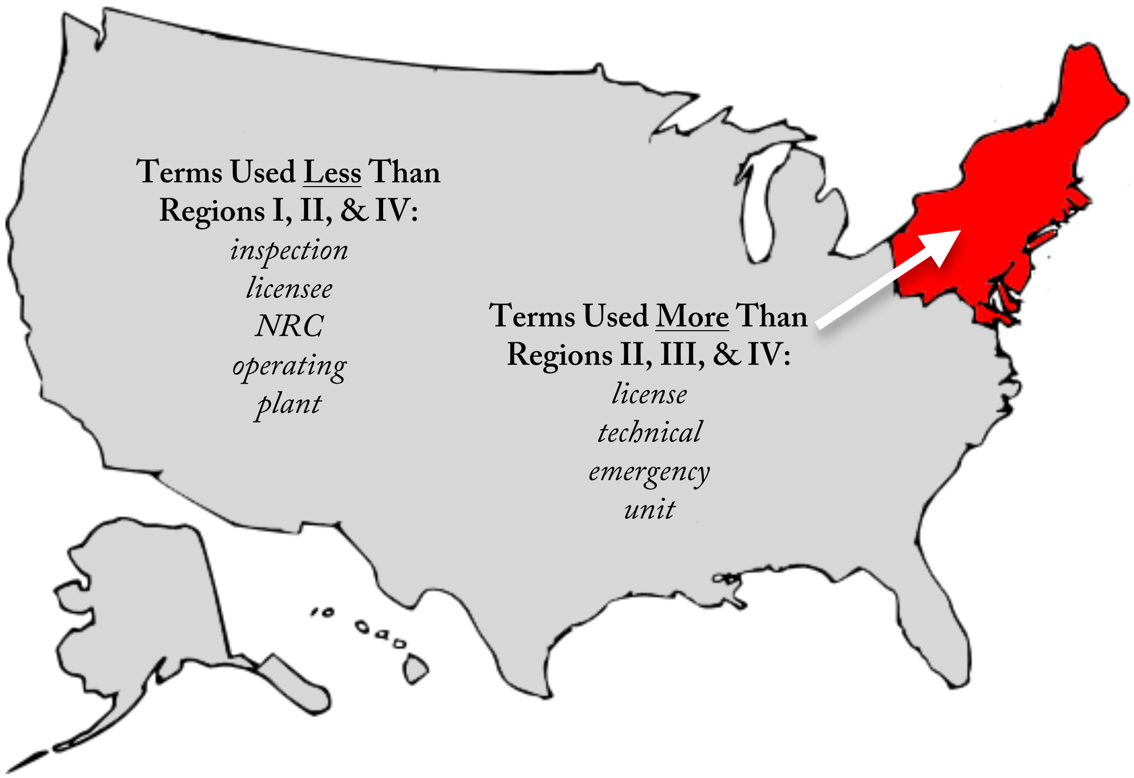

Figure 6.10: Key Terms for Licensee-Authored Documents in Region I ($p<0.01$)

group in the other regions (Figure 6.11). On the other hand, Figure 6.12 demonstrates that Region III is characterized by a different set of key terms being used more than expected in comparison to the other regions in licensee-authored documents: *nuclear, reactor*, and *plant*. These same documents use *emergency, NRC, technical, unit, containment,* and *safety* far less in comparison to the other regions.

Similarly to the patterns observed in Region III, Region IV is also found to have more instances of *emergency, safety,* and *technical* than is expected based off of the rest of the industry, in addition to *inspection, licensee, NRC, operating,* and *staff* (Figure 6.13).

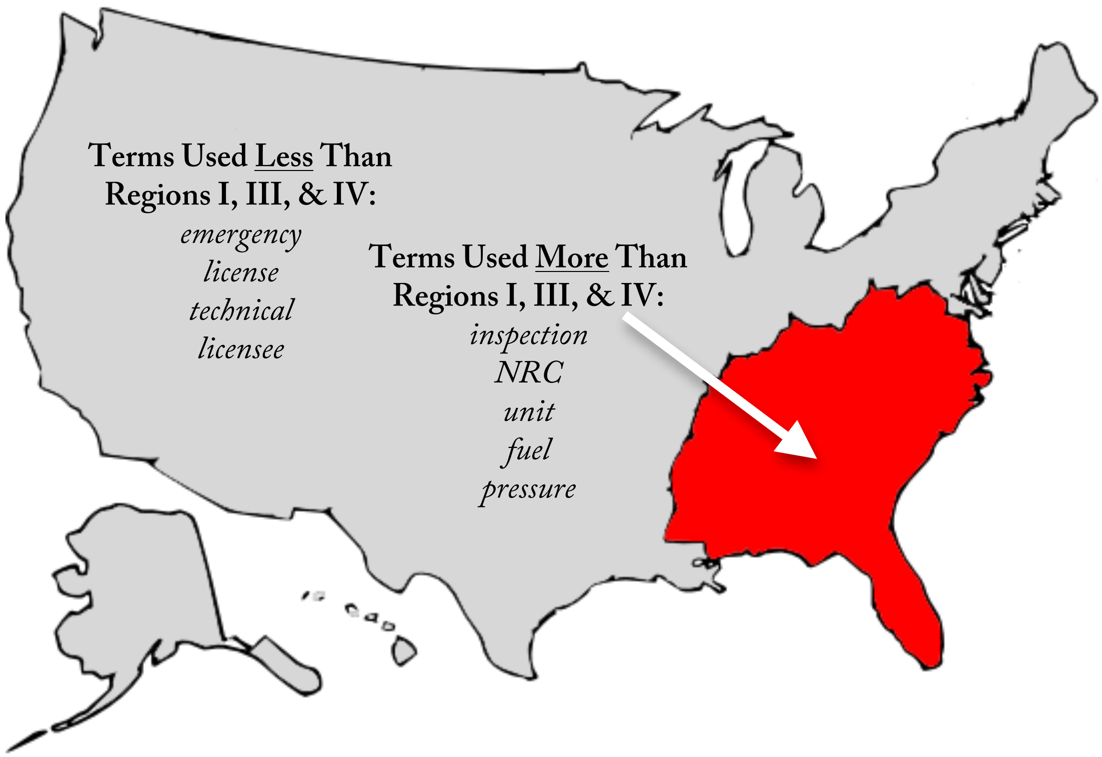

Figure 6.11: Key Terms for Licensee-Authored Documents in Region II (p<0.01)

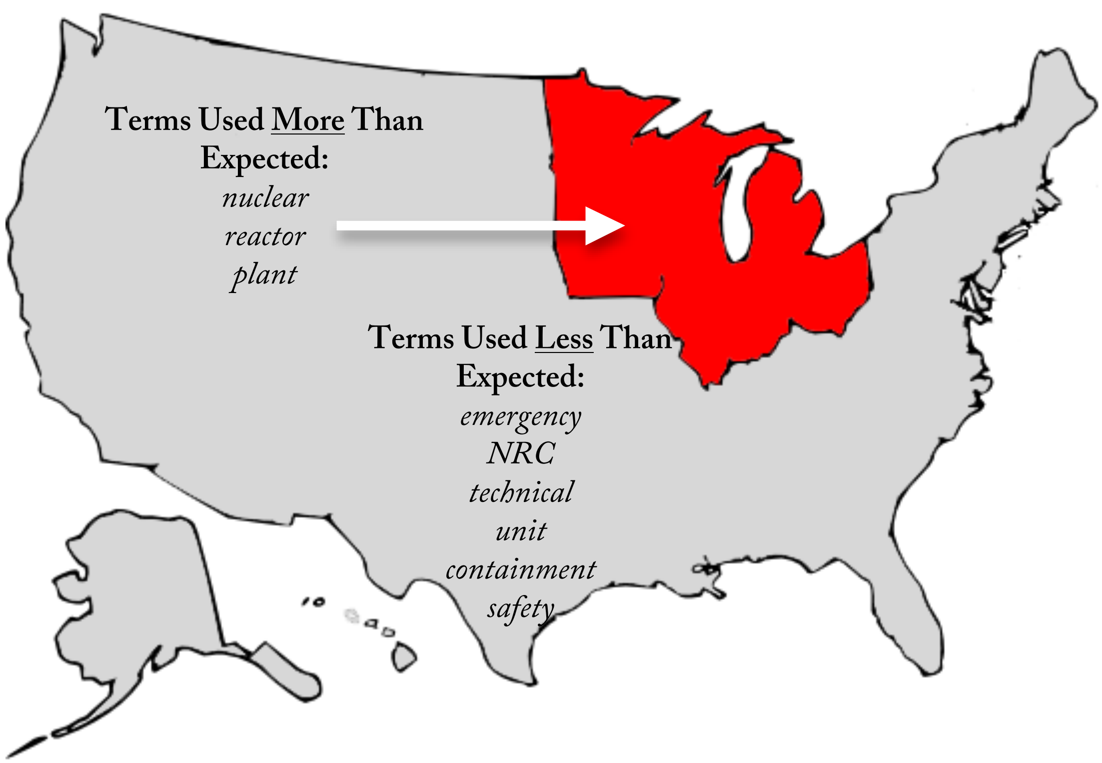

Figure 6.12: Key Terms for Licensee-Authored Documents in Region III (p<0.01)

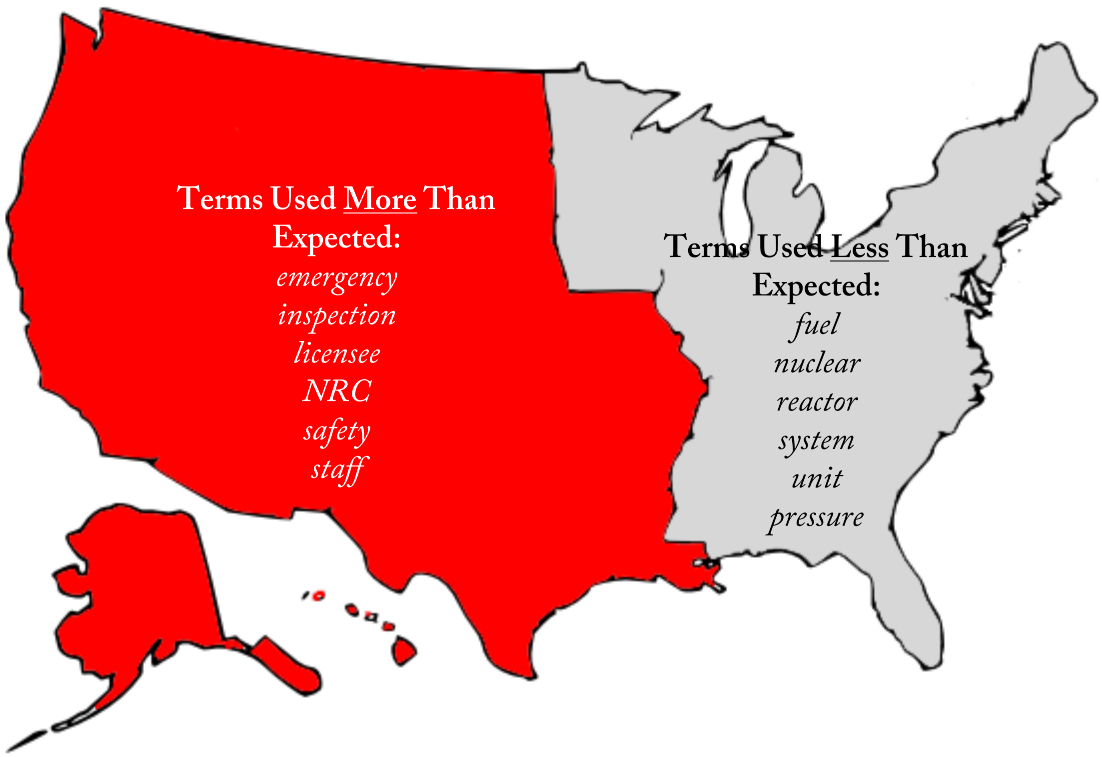

Figure 6.13: Key Terms for Licensee-Authored Documents in Region IV (p<0.01)

In addition, licensee-authored documents in Region IV also have a tendency for *fuel, nuclear, reactor, system, unit,* and *pressure* to be used less frequently in comparison to documents authored by licensee representatives in the other regions. When we compare these distributions of key terms used by licensees between the regions (Figure 6.14), we find an interesting pattern of oppositional use for pairs of regions. First, the documents in Regions I and IV communicate most about *emergency*, while this term is used less by Regions II and III. The licensee-authored documents for dockets in Regions II and IV employ *NRC* more than is expected by the other regions, while Regions I and III are characterized as using it less than expected. This same type of pattern can be observed

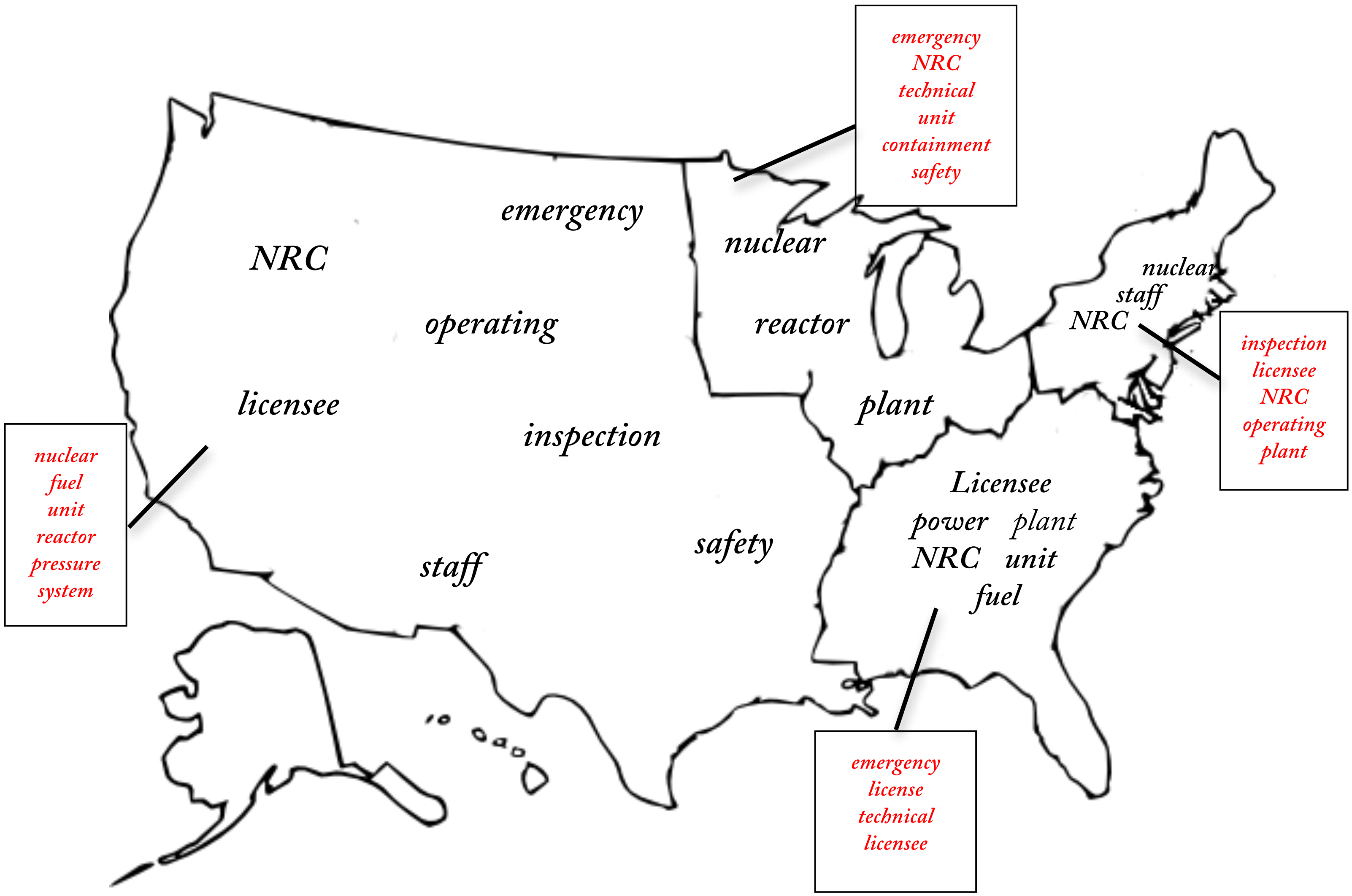

Figure 6.14: Key Terms for Licensee-Authored Documents in All Regions (p<0.01)

through the use of *unit* more often than Regions I and II, and far less by Regions III and IV. Also, it appears as though Regions I and IV are opposites beyond their use of *emergency*, specifically with regard to many other terms: *inspection, licensee, NRC,* and *operating*. We can also see that there is also opposition in the use of terms by Regions II and IV. Region II uses *unit, fuel,* and *pressure* more than is to be expected while Region IV uses these three terms far less than is to be expected. *Reactor* is another term that is used in opposition by two regions: it is used more by Region II and less by Region IV.

There are also some interesting patterns that emerge when we compare the results of our regional industry group analysis to one another: i.e. comparing the data for Region I NRC-authored documents (Figure 6.5) to Region I licensee-authored documents (Figure 6.10). As we can see, the documents authored by both groups in Region I used *license* more than their compatriots in all of the other regions, indicating that this focus is pervasive in the culture for this region. However, the licensee-authored documents for Region I used *emergency* and *unit* more than those in Regions II, III, and IV, while the NRC-authored texts were characterized as using these key terms less than the documents authored by this same group in the other regions. This suggests that although the licensees are talking about *emergency* more than is to be expected, the NRC is talking about it less than NRC authors in other regions. On the other hand, the NRC-authored texts in this region used *NRC* more than NRC-authored texts in the other regions, while the documents written by licensees for Region I used *NRC* less than the documents written by licensees for the other regions. Ultimately, what we are seeing is that there is more disparity in the use of the industry key terms in Region I than in the other Regions.

Performing these same comparisons for Region II, We find that the documents written by both industry groups in Region II used *unit* and *fuel* more than the documents written by both NRC and licensee authors in the other Regions. However, if we look at the patterns of oppositional terms between the two industry groups in Region II, we notice that the licensee-authored texts for Region II used *NRC* and *inspection* more than all of the other regions, the NRC-authored documents used both of these terms less than those documents also authored by the NRC for the other regions. On the other hand, documents authored by the NRC for communication in this region used *licensee* far more than the texts written by this group for communication in the other regions, and licensee-authored documents used this term less than Regions I, III, and IV. What we can tell from this opposition is that both groups tend to refer to one another more than any other region.

Comparing the significant use, or lack of use, of key terms for each industry group in the context of regional patterns for Region III demonstrates that both groups in this region have a tendency to use *reactor* and *plant* more, while they also use *NRC* less than the rest of the regions. Thus, we see an increased focus on the plant-level and decreased focus on the *NRC* is pervasive throughout the region for both groups. However, this tendency does not extend to *nuclear*, where the licensee-authored documents used this term more than their counterparts in the other regions, and the documents written by the NRC for this region were characterized by a lack of occurrences for this term in comparison to Regions I, II, and IV combined. This same behavior was true for the propensity for *safety* to be used more in NRC-authored texts in Region III than Regions

I, II, and IV, while it occurred rather infrequently in licensee-authored documents for Region III.

Region IV's focus on *emergency, inspection, licensee,* and *safety* appears to be shared by both industry groups, indicating that these issues are pervasive in the practice of this region. Furthermore, both NRC- and licensee-authored documents in Region IV are characterized by an infrequent use of *fuel, nuclear,* and *unit* in comparison to documents authored by individuals having the same affiliations in Regions I, II, and III. The only key term that these two groups possess an oppositional use of in comparison to the other regions is *pressure*, which is used more in the NRC-authored documents in Region IV than the other regions and less by the licensees in Region IV than the rest of their counterparts in the industry. In opposition to the observation made about Region I having the most disparity with regard to the use of key terms between industry groups, Region IV has the most agreement between the key terms used more often in comparison to the other regions by both industry groups. These types of differences were not the only ones observed for the language of the regulated nuclear power industry with regard to regional variation.

DIFFERENCES IN OBSERVED MEANING

Upon analyzing the key term frequencies for each region with respect to industry groups, I generated lists of the 20 most frequently used collocates for each term (Appendix C). The result of this process was 160 collocate lists, one list for Regions I through IV for both licensee-authored texts and those written by the NRC. Analysis of these tables for each key term yields the observation that, in general, the meanings constructed for each of these key terms becomes more specific for each industry group on a regional basis, but they are still in the context of the Observed Meanings constructed on the industry group level. The terms that behaved this way are *safety*, *reactor*, *unit* (NRC-authored documents), *system, fuel, evaluation, staff, operating, technical, emergency, inspectors* (NRC-authored documents), *power, nuclear,* and *pressure*. For example, we learn in the distributions of *safety* that the specific aspects of this term differ from region-to-region for both groups.

In Table 6.5, for the licensee-authored documents in Regions I, II, and III, there are five new collocates: *relief* (RI), *nuclear* (RI and RII), *updated* (RII and RIII), and *non* (RIII).

Table 6.5: 20 Most Frequent Collocates with *Safety* by Region (Licensee-Authored)

Region I	Region II	Region III	Region IV
related	related	analysis	related
analysis	analysis	related	analysis
evaluation	***evaluation***	***system***	function
system	function	***evaluation***	***evaluation***
plant	***system***	report	report
function	injection	function	***system***
report	report	margin	health
public	health	***plant***	significance
injection	margin	health	margin
level	analyses	significance	public
margin	public	equipment	injection
nuclear	final	limits	***plant***
systems	valves	public	systems
limit	***nuclear***	relief	analyses
valves	systems	updated	final
health	level	non	level
quality	quality	limit	quality
final	updated	final	limits
relief	***plant***	systems	valve
equipment	limit	Valves	valves

We see that these collocates of specificity work to provide context through *NUCLEAR SAFETY ASSESSMENT* (RI-23), *BEEN DETERMINED TO BE NUCLEAR SAFETY RELATED* (RII-12), *SAFETY RELIEF VALVES* (RII-52), *UPDATED FINAL SAFETY ANALYSIS REPORT* (RIII-63), *NON-SAFETY RELATED* (RIII-70), *NON-SAFETY RELATED SNUBBERS* (RIII-8). This same type of behavior can be observed as well in the other terms like *fuel, evaluation,* or *pressure* in the tables provided in Appendix C. Moreover, we find for *NRC, plant, inspection, unit, licensee, license, containment,* and *pressure* all have regional variation in the Observed Meaning, beyond increased specificity, with regard to industry groups.

NRC

For the NRC-authored documents in each region, we find that *NRC* co-occurs with words in such a way that the Observed Meaning for this term is quite similar across all of the regions (Table 6.6). The only difference that really can be pointed out is that Region IV is the one that has more frequent co-occurrences of *policy* with *NRC*, as this word is not found in the 20 most frequent collocates for any other industry group in any of the regions. This same observation can be made about the presence of *safety* in Region III. It should be conceded that this is a small difference, as *policy* and *safety* both co-occur with *NRC* for all of the regions for this industry group—just not in the top 20 collocates listed in Table 6.6.

Conversely, we can see a much larger variation in the words co-occurring with *NRC* for the licensee-authored texts across Regions I through IV. We can see in Regions I and III that there the new collocates *consider, recommend,* and *AMERGEN* (Table 6.7).

Table 6.6: 20 Most Frequent Collocates with *NRC* by Region (NRC-Authored)

Region I	Region II	Region III	Region IV
staff	*staff*	*staff*	*staff*
site	site	site	public
public	*inspection*	*inspection*	*inspection*
inspection	public	public	site
review	web	review	regulatory
web	approved	ADAMS	commission
staff's	review	web	*nuclear*
ADAMS	ADAMS	staff's	ADAMS
will	staff's	document	review
reviewed	determined	room	concludes
issued	finds	*evaluation*	web
information	will	oversight	*licensee*
approved	document	approved	document
document	*licensee*	*licensee*	reviewed
room	room	enforcement	room
identified	reviewed	concludes	staff's
evaluation	information	*safety*	enforcement
letter	letter	reviewed	policy
requirements	concludes	determined	identified
concludes	approval	reading	approved

When we look at the most frequent clusters involving these words, we find that there is a specific corporation in Regions I and III that frequently makes recommendations to the *NRC* regarding this agency's language use: *EXELON AMERGEN RECOMMEND THAT THE NRC CONSIDER* (RI-36, RIII-60), *THE NRC CONSIDER REVISING* (RI-16, RIII-24), *THE NRC CONSIDER REVISING THE FIRST SENTENCE* (RI-8, RIII-12), *THE NRC CONSIDER REMOVING* (RI-8, RIII-12), *THE NRC CONSIDER REMOVING ALL* (RII-6), *THE NRC CONSIDER DELETING* (RI-8, RII-12), *THE NRC CONSIDER DELETING ALL* (RIII-6), *THE NRC CONSIDER REWORDING THE PARAGRAPH* (RIII-6), *THE NRC CONSIDER ELIMINATING*

Table 6.7: 20 Most Frequent Collocates with *NRC* by Region (Licensee-Authored)

Region I	Region II	Region III	Region IV
approved	***staff***	approved	approved
staff	letter	request	***staff***
review	approved	reference	approval
approval	approval	***staff***	letter
letter	response	letter	request
information	request	approval	information
request	TVA	information	review
submitted	question	additional	provided
provided	Bulletin	requested	question
evaluation	information	response	response
reference	review	submitted	***evaluation***
requested	provided	provided	***safety***
safety	submitted	issued	***inspection***
dated	SER	***evaluation***	will
notification	requested	consider	prior
consider	***inspection***	***safety***	reference
question	additional	recommend	order
response	report	AMERGEN	submitted
issued	issued	dated	requested
AMERGEN	reference	review	made

THIS OBJECTIVE (RIII-6), *THE NRC CONSIDER REMOVING ALL REFERENCES* (RIII-6).

It should be noted that no other region, and for that fact no other licensee, is found in the corpus to have *recommend* or *consider* as a collocate with *NRC* at all. This is not to say that licensees in Regions I, II, III, and IV are not also making these kinds of recommendations to the NRC. For example, when a concordance is created for all instances of the word *sentence* in the documents authored by licensees in Region IV, we find that there is a frequent use of the cluster *REVISE SENTENCE TO READ* (RIV-16). This cluster has the same essential function as those noted in the collocation of *recommend* with *NRC* of suggesting NRC revision of a document. However, the examples resulting from the collocation of *recommend* with *NRC* indicates that EXELON is using a more direct mode of communication regarding these revisions that has pragmatic implications beyond the semantic meaning of licensees suggesting revisions to NRC language. These observations cause us to question whether Exelon's relationship with the NRC. Do they in fact have a better repoire with NRC officials, and as a result better performance? Did this linguistic approach to discussing revisions aid in establishing such a repoire? And does this behavior reflect a linguistic attitude of politeness toward the NRC that other licensees should adopt? While analysis of the nuclear reference corpus cannot answer these questions, they are ones that do warrant future study.

PLANT

For *plant,* the variation between the regions for the licensee-authored texts does exhibit an added degree of specificity to the meaning of this term, but we are able to uncover additional dimensions of its meaning through the different collocates (Table 6.8). In terms of specificity, we see more terms referring to specific plants in the licensee-authored documents for each region in addition to those authored by the NRC, and even titles of personnel: i.e. *THE PLANT MANAGER* (RIV-20).

Beyond specificity, the first observation we can make from Table 6.8 is that Regions I and II exhibit a frequent occurrence of *will,* which indicates the commitments made by a *plant* to do things in the future: *THE PLANT WILL SATISFY APPLICABLE REGULATORY REQUIREMENTS* (RI-10, RII-21), *THE PLANT WILL SATISFY*

Table 6.8: 20 Most Frequent Collocates with *Plant* by Region (Licensee-Authored)

Region I	Region II	Region III	Region IV
power	*nuclear*	*nuclear*	*power*
nuclear	*unit*	*power*	operation
specific	specific	specific	specific
conditions	operation	operation	conditions
safety	*power*	conditions	*nuclear*
operation	electric	*safety*	*safety*
operating	steam	generating	operations
equipment	conditions	Cook	*operating*
shutdown	will	shutdown	Callaway
systems	Ferry	*unit*	systems
operations	units	*operating*	manager
based	Browns	design	personnel
Fitzpatrick	Hatch	equipment	shutdown
current	systems	data	design
design	procedures	Prairie	Canyon
personnel	*operating*	Island	Diablo
system	*safety*	procedures	*system*
Ginna	design	systems	normal
changes	operations	personnel	required
will	generating	Monticello	status

(RI-10), *WILL COORDINATE WITH THE PLANT ASSESSMENT* (RII-9), *DIRECTOR WILL COORDINATE WITH PLANT ASSESSMENT* (RII-9), *PLANT ASSESSMENT TEAM WILL DETERMINE* (RII-6).

For the licensee-authored documents in Region III, *plant* is associated with *data*: *IN PLANT DATA* (RIII-9), *PLANT DATA COLLECTED* (RIII-7), *PLANT RESPONSE DATA* (RIII-8). And for the documents in Region IV that are authored by licensee representatives, when *plant* is a quality of *condition* or *evaluations*, it is at times *required*: *REQUIRED PLANT CONDITIONS* (RIV-8), *REACH THE REQUIRED PLANT CONDITION FROM FULL SHUTDOWN* (RIV-5), *WHICH REQUIRE A*

PLANT SPECIFIC EVALUATION (RIV-10). In addition, we learn from this region that *plant operations* are regularly classified as normal, and this characterization is a quality of *plant components* and *methods: NORMAL PLANT OPERATION* (RIV-27), *NORMAL PLANT COMPONENTS* (RIV-8), and *NORMAL PLANT METHODS* (RIV-8). Similarly in Region I, we find that *plant conditions* can also be *current: BASED ON CURRENT PLANT CONDITIONS* (RI-46).

Some additional meanings for *plant* become apparent from the words co-occurring with this key term in NRC-authored texts across all four regions. In Table 6.9 we see that *plant* often occurs with *activities* in documents authored by this industry group in Region I, and that the NRC concedes that the topics they are writing about *MAY CONFLICT WITH YOUR PLANT ACTIVITIES* (RI-23). From Region II, the additional meaning of *plant* being synonymous with *site* filters into the 20 most frequent collocates*: THE PLANT SITE* (RII-32), *PLANT OR SITE CHARACTERISTICS* (RII-24). And in Region IV, the collocates demonstrate that although licenses are credentials given to *licensees*, they directly govern the *plant*: *PLANT CONDITIONS BEYOND LICENSES BASIS* (RIV-28), *LICENSES BASIS OF THE PLANT* (RIV-17). Furthermore, changes made at the *plant* precipitate a *response* to/from the NRC: *DOESN'T CLEARLY ADDRESS RESPONSE TO PLANT CONDITIONS BEYOND LICENSES* (RIV-14), *OPERATORS IN RESPONSE TO CHANGING PLANT* (RIV-6).

Table 6.9: 20 Most Frequent Collocates with *Plant* by Region (NRC-Authored)

Region I	Region II	Region III	Region IV
nuclear	*nuclear*	*nuclear*	*power*
power	*power*	*power*	*nuclear*
specific	specific	specific	specific
Calvert	*unit*	performance	performance
Cliffs	units	*safety*	conditions
performance	*safety*	modifications	*safety*
Ginna	electric	risk	procedures
operating	performance	operations	licensing
operation	Hatch	status	modifications
design	operation	operation	operation
safety	steam	areas	Callaway
unit	*operating*	design	operating
systems	Harris	*unit*	*inspectors*
risk	Lucie	units	equipment
operations	conditions	*inspectors*	risk
Fitzpatrick	modifications	conditions	reviewed
procedures	risk	procedures	areas
areas	site	*operating*	Diablo
LLC	procedures	equipment	response
activities	areas	Cook	Canyon

One last observation to be made about the data in Table 6.9 is the variation in the proper nouns listed for each region. All of the regions show different proper name distributions than those listed for the licensees in Table 6.8. For example, we see in the Region I column on Table 6.9 the addition of *Calvert Cliffs*, which has a higher frequency of occurrence in this context than both Fitzpatrick and Ginna that were the only two proper nouns listed for Region I in Table 6.8.

INSPECTION

Inspection also shows variation in the specificity provided in the words co-occurring with it for both industry groups across all four regions. For example, licensee-authored documents in Region I provide evidence of more types of inspections (Table 6.10): *RISK INFORMED INSERVICE INSPECTION* (RI-33), *REFUELING OUTAGE INSPECTION* (RI-7).

Table 6.10: 20 Most Frequent Collocates for *Inspection* by Region (Licensee-Authored)

Region I	Region II	Region III	Region IV
inservice	inservice	inservice	inservice
program	program	program	program
interval	visual	interval	interval
visual	interval	visual	scope
methods	will	requirements	visual
requirements	report	methods	tube
year	methods	period	*NRC*
will	one	results	results
performed	performed	used	performed
informed	time	year	steam
scope	*NRC*	performed	will
risk	tube	*NRC*	*inspectors*
reactor	results	penetration	requirements
outage	penetration	*nuclear*	generator
ASME	metal	third	report
results	steam	*plant*	methods
section	period	second	period
evaluation	next	steam	public
penetration	year	guidelines	plan
third	activities	SG	ASME

Similarly, the documents written by licensee representatives in Regions I, II, and III show a tendency for communicating about other specific inspections like *BARE METAL*

VISUAL INSPECTION OF EACH PENETRATION, or *LOWER HEAD PENETRATION INSPECTION*. We can also observe the degree of specificity related to *inspection* in Region IV's frequent co-occurrences with *manual* and *chapters*: *NRC INSPECTION MANUAL* (RIV-11), *NRC INSPECTION MANUAL PART # TECHNICAL GUIDANCE* (RIV-7), *NRC INSPECTION MANUAL CHAPTER* (RIV-6). However, there is more to the collocates with *inspection* than just specificity.

For the NRC-authored documents in Region I, inspection is something used (Table 6.11): *USING INSPECTION MANUAL CHAPTER* (RI-20), *MANAGED USING INSERVICE INSPECTION* (RI-8), *USING NRC INSPECTION* (RI-8).

Table 6.11: 20 Most Frequent Collocates with *Inspection* by Region (NRC-Authored)

Region I	Region II	Region III	Region IV
scope	scope	scope	scope
NRC	*NRC*	*inspectors*	*inspectors*
public	public	reviewed	*NRC*
inspectors	reviewed	*NRC*	public
reviewed	results	public	program
team	available	results	reviewed
results	report	report	results
available	inservice	period	report
findings	conducted	available	available
report	period	findings	inservice
inservice	plan	conducted	findings
program	findings	electronically	reports
conducted	program	inservice	period
period	team	inspector	performed
procedure	activities	constituted	electronically
electronically	electronically	activities	procedure
performed	procedure	Mr.	manual
identified	performed	program	Mr.
using	will	performed	team
visual	last	sample	chapter

On the other hand, for Region II an *inspection* is a point of reference that provides context and is used differently by the NRC authors and the licensees. The Region II NRC-authored documents frequently refer to the *last* inspection: *SINCE THE LAST INSPECTION* (RII-62), *REPORTS GENERATED SINCE THE LAST INSPECTION* (RII-16), *LAST SIX MONTHS OF THE INSPECTION* (RII-14), and licensees point to the *next* inspection (Table 6.10): *THE NEXT INSPECTION* (RII-41), *UNTIL THE NEXT INSPECTION* (RII-18), *PLANT RESTART FOLLOWING THE NEXT INSPECTION* (RII-12), *MAINTAINED UNTIL THE NEXT INSPECTION* (RII-9).

From the NRC-authored texts in Region III we find that they often designate an *inspection* as being *constituted* of *samples: THE INSPECTION CONSTITUTED ONE SAMPLE* (RIII-42), *THE INSPECTION CONSTITUTED TWO, THE INSPECTION CONSTITUTED ONE READINESS* (RIII-6), *THE INSPECTION CONSTITUTED TWO UNPLANNED* (RIII-5). There is even further evidence these *samples* are defined by the NRC: *SAMPLE AS DEFINED* (RIII-56), *SAMPLED AS DEFINED IN INSPECTION PROCEDURE* (RIII-7). One last observation we can make from Table 6.11 about the meaning created for *inspection* in the documents authored by the NRC in Region III is that evaluations are made about these samples during inspections: *INSPECTION SAMPLE B FINDINGS* (RIII-33).

UNIT

As we have already determined, increased specificity is an inherent characteristic of the collocational patterns of each key term when we zoom in to look at the differences between the documents authored by each industry group with regard to regional variation, and this can be seen in Table 6.12.

Table 6.12: 20 Most Frequent Collocates with *Unit* by Region (NRC-Authored)

Region I	Region II	Region III	Region IV
station	*plant*	*power*	station
nuclear	*nuclear*	*plant*	*nuclear*
power	Lucie	station	steam
Point	*power*	*reactor*	Verde
Mile	station	Byron	Arkansas
Milestone	outage	pump	*power*
generating	*reactor*	amendment	generator
Salem	refueling	*nuclear*	*plant*
Indian	WBN	*containment*	refueling
Cliffs	steam	*system*	electric
plant	shutdown	Braidwood	outage
system	Bar	shutdown	model
Island	room	train	Waterford
TMI	*license*	refueling	generating
Seabrook	facility	outage	Palo
reactor	building	opposite	CPSES
inspection	Watts	generator	*reactor*
Nine	electric	water	diesel
water	Point	trip	STP
RAI	*operating*	diesel	*emergency*

We see in the NRC group that there is variation in the specific items used at these facilities that co-occur with *unit:* like *water* and *reactor* for Region I; *room, facility, building*, and *operating* for Region II; *containment, train, water, trip,* and *diesel* in Region IV, and *model* and *diesel* for Region IV. *Unit* also exhibits this characteristic as well through a greater amount of proper names present for each of the regions in both the NRC- and licensee-authored documents (Tables 6.12 and 6.13).

Table 6.13: 20 Most Frequent Collocates with *Unit* by Region (Licensee-Authored)

Region I	Region II	Region III	Region IV
Millstone	Lucie	station	steam
Salem	*plant*	Byron	outage
station	*nuclear*	Braidwood	station
power	HBRSEP	*power*	*power*
BVPS	cycle	cycle	refueling
TMI	*power*	*nuclear*	*nuclear*
Point	refueling	refueling	mode
Mile	outage	outage	CPSES
technical	steam	*reactor*	model
nuclear	North	*plant*	Arkansas
fuel	Anna	DNPS	Creek
TS	BFN	mode	*operating*
Nine	mode	event	electric
reactor	electric	section	Wolf
amendment	*operating*	FOL	*plant*
cycle	shutdown	opposite	Waterford
operating	operation	*operating*	operated
specifications	*technical*	*fuel*	cycle
operation	TS	TMI	peak
month	restart	LaSalle	month

We learn that there are different meanings associated with *unit* that are depicted in these collocational frequencies as well. First of all, *unit* is inextricably linked to licensure and *technical specifications* from the perspective of licensees in Region I *(technical, TS, amendment, operating, specifications)*. For Region II in Table 6.13, we find that *restart* is an added focus of *unit* in Region II for the licensee-authored documents, in opposition to *shutdown*.

A theme of otherness becomes present for Region III as *THE OPPOSITE UNIT* (RIII-39) is often used to refer to the other licensed reactor unit at a plant that is not the one having just been discussed. Also, we see the presence of *event* as a collocate that designates this key term is related to something important that happens. The last observation that is to be made about the variation in the collocates of *unit* with regard to variation in Region III is the presence of *TMI,* or Three Mile Island, in the collocate list of a region other than Region I where it is located. We see that *TMI* is the 19th most frequently co-occurring word with *unit* for this region, and it becomes clear that this particular *unit* is frequently referred to in the communications of Region III: *THE TMI UNIT 1* (RIII-8), *AT TMI UNIT 1* (RIII-6).

LICENSEE

For this term, the roles are reversed from those observations made earlier about the term *NRC*: the licensees exhibited essentially the same meanings associated with *licensee*, while the NRC has some variation between the regions. The best way for us to view this variation is by taking a deeper look at the clusters associated with the co-occurrence of *failed* with *licensee*, which as we can see in Table 6.14 appears in all of the 20 most frequent collocate lists except for Region I.

When we zoom in to look at *failed*, we see that it co-occurs with *licensee* only 15 times for Region 1, 109 times for Region 2 NRC-authored documents, 150 times for Region 3, and 177 times for Region 4. As a result of the infrequent use of *failed* in the NRC-authored documents assigned to dockets in Region I, no clusters with *failed* occur at least 5 times. This is not to say that *failed* is never used in NRC communications with Region I. However, it is just that this word is used much less in a four word span with *licensee* in the documents assigned to docket that were sampled as part of the reference

Table 6.14: 20 Most Frequent Collocates with *Licensee* by Region (NRC-Authored)

Region I	Region II	Region III	Region IV
identified	identified	identified	identified
staff	stated	*inspectors*	failed
provided	*staff*	*staff*	*NRC*
proposed	proposed	proposed	stated
stated	*inspectors*	shall	*staff*
submitted	*NRC*	company	proposed
shall	determined	requested	*inspectors*
information	company	entered	submitted
requested	also	performance	reviewed
facility	provided	failed	determined
nuclear	requested	also	performance
LLC	performed	management	changes
response	reviewed	stated	provided
request	*power*	facility	requested
violations	shall	provided	also
dated	submitted	response	management
also	will	reviewed	performed
INC	failed	violations	personnel
company	verify	information	violations
may	dated	submitted	findings

corpus. What we can say, however, is that Region I is supervised and communicated with differently regarding "failure." Nonetheless, for the remaining three regions we can see in Table 6.15 there are different frequencies and rankings of clusters for each region. While there are many similar constructions across all three regions, there are also distinct clusters involving *failed* and *licensee* for each one, most of which begin with *licensee failed to*. For Region I, we see that there are the distinct clusters *LICENSEE FAILED TO PROTECT*, *LICENSEE FAILED TO MEET*, and *LICENSEE FAILED TO DETERMINE*. In Region III we find the specification of *licensee personnel* being the ones who *failed*, as well as the patterns of *SPECIFICALLY THE LICENSEE FAILED*, and

Table 6.15: Clusters by Region for *Licensee* Co-Occurring with *Failed*

Region II	Region III	Region IV
tracking number requirement the licensee failed to meet (8)	***specifically the licensee failed* (17)**	*licensee failed to identify* (16)
***licensee failed to protect* (8)**	*failed to ensure* (15)	*licensee failed to perform* (14)
***licensee failed to meet* (8)**	*failed to implement* (14)	***licensee failed to evaluate* (13)**
licensee failed to adequately (7)	*failed to identify* (14)	***licensee failed to establish* (12)**
licensee failed to implement (7)	***licensee personnel failed* (9)**	*licensee failed to ensure* (10)
licensee failed to ensure (7)	*license failed to adequately* (9)	***licensee failed to properly* (8)**
***licensee failed to determine* (6)**	*licensee failed to perform* (8)	*licensee failed to adequately* (7)
licensee failed to identify (5)	***licensee failed to take appropriate* (7)**	***licensee failed to provide* (6)**
	tracking number requirement the licensee failed to meet (7)	***licensee failed to submit* (6)**
		***licensee failed to promptly identify* (6)**

LICENSEE FAILED TO TAKE APPROPRIATE. We see the greatest amount of unique clusters in the Region IV column, which is to be expected since *failed* co-occurred the most with *licensee* in this region when compared to the others: *LICENSEE FAILED TO EVALUATE, LICENSEE FAILED TO ESTABLISH, LICENSEE FAILED TO PROPERLY, LICENSEE FAILED TO PROVIDE, LICENSEE FAILED TO SUBMIT*, and *LICENSEE FAILED TO PROMPTLY IDENTIFY.*

LICENSE

With regard to the regional variation of the collocates with *license* identified for NRC-authored documents, we can see that the Observed Meanings they create are essentially the same as what we observed in Chapter 5. However, we do find that there are a few additional meanings for documents written by NRC representatives of the four regions for the key term *license*.

In Table 6.16 there is a new collocate present in the lists for Regions II, III, and IV: *paragraphs*.

Table 6.16: 20 Most Frequent Collocates with *License* by Region (NRC-Authored)

Region I	Region II	Region III	Region IV
renewal	renewal	*operating*	renewal
operating	amendment	amendment	*operating*
amendment	*operating*	renewal	amendment
facility	renewed	facility	renewed
scope	facility	condition	facility
application	condition	renewed	condition
transfer	scope	conditions	conditions
within	application	incorporated	application
proposed	hereby	hereby	incorporated
renewed	incorporated	effective	request
nuclear	request	proposed	hereby
subject	effective	amended	scope
transfers	attachment	request	changes
condition	amended	changes	effective
review	proposed	accordingly	amended
effective	accordingly	application	review
environmental	within	paragraph	attachment
applications	paragraph	will	will
accordingly	environmental	conducted	accordingly
amended	*nuclear*	*nuclear*	paragraph

Taking a close look at the concordance for each of these groups we find that in addition to everything else the licenses are, they are in fact documents that have *paragraphs*. Additionally, we learn from Region I that these documents that are administered by the

NRC and have authority to permit an entity, or individual, to operate a nuclear facility can be transferred from one owner to another, specifically a transfer of control of the license: *TRANSFER OF THE LICENSE* (RI-40), *THE LICENSE TRANSFER* (RI-23), *TRANSFER OF CONTROL OF THE LICENSE* (RI-26). It should also be noted that there is also meaning created for *license* in the words co-occurring with this key term for the licensee-authored documents in Region III (Table 6.17).

Table 6.17: 20 Most Frequent Collocates with *License* by Region (Licensee-Authored)

Region I	Region II	Region III	Region IV
renewal	renewal	amendment	amendment
amendment	amendment	renewal	*operating*
operating	*operating*	*operating*	request
request	request	request	renewal
application	proposed	application	application
scope	scope	facility	proposed
proposed	facility	scope	condition
unit	condition	condition	facility
nuclear	amendments	within	station
facility	application	TS	conditions
condition	renewed	proposed	change
within	within	*nuclear*	scope
BVPS	conditions	permit	within
NRC	*unit*	construction	*technical*
station	will	following	will
power	*NRC*	amendments	*unit*
environmental	*nuclear*	approved	*NRC*
amendments	change	manner	changes
technical	incorporated	*NRC*	incorporated
review	*plant*	*plant*	requested

Switching over to analyze the collocates of this particular group, we find that the *license* is also related to *construction permits*: *AMENDMENT OF LICENSE OR*

CONSTRUCTION PERMIT (34). The perspective that a *license* has a relationship with another document if something that was also observed with the co-occurrence of *technical specifications* with this term in Chapter 5. We also notice for Region I that the acronym for one of the licensees in this region filters up into this list of collocates: *BVPS,* which standards for Beaver Valley Power Station.

CONTAINMENT

For *containment*, we find that there is a greater frequency of *accident* co-occurring with this term in Region I for both licensee and NRC-authored documents (Table 6.18).

Table 6.18: Top 20 Collocates with *Containment* for Region I by Industry group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
primary	8.716741	116	primary	8.010806	221
pressure	6.460024	105	secondary	8.511730	187
spray	8.926990	96	*system*	5.014587	157
integrity	8.013622	91	isolation	7.435794	156
isolation	8.237426	91	*pressure*	5.757265	147
system	4.568516	64	sump	7.869611	94
accident	6.354671	60	integrity	7.304549	91
secondary	8.351171	58	atmosphere	8.571278	84
leakage	6.370459	44	*fuel*	4.935608	82
outside	7.706462	44	spray	7.138478	67
systems	5.191535	42	inside	7.922352	65
cooling	5.953671	40	accident	5.200184	63
liner	8.453137	38	air	5.804593	59
valve	5.810931	36	building	6.334270	55
valves	15.04874	35	outside	7.028241	54
water	4.376313	35	within	4.352497	52
heat	5.784296	35	monitor	6.628994	51
credit	7.299559	34	loss	4.815322	48
program	14.00692	34	valve	5.100034	46
removal	6.735031	33	leakage	5.161579	44

An analysis of the concordance reveals that accidents can occur *inside containment*, and this occurrence is connected with *pressure*, thus is measured, and has become a part of the licensees' programs: *CREDIT FOR CONTAINMENT ACCIDENT PRESSURE* (RI/NRC-24), *POST-ACCIDENT CONTAINMENT* (RI/NRC-6), *CONTAINMENT POST-ACCIDENT* (RI/LIC-19), *ACCIDENT INSIDE CONTAINMENT* (RI/LIC-23). Moreover, we become aware that these accidents are typically a result of *fuel* handling: *FUEL HANDLING ACCIDENT INSIDE CONTAINMENT* (RI/LIC-23). Looking at the words co-occurring with *containment* in licensee-authored documents in Region IV, we notice a similar pattern of co-occurrence to that just described in Region I involving *accident* (Table 6.19).

These documents often have frequent occurrences of *FUEL HANDLING ACCIDENT* (RIV = 18) within a four-word span of *containment*. However, *accident* does not appear in the 20 most frequent collocates for the Region IV licensees. Instead, we find they have a strong propensity to use *failure* with containment in a similar way: *CONTAINMENT FAILURE PRESSURE* (15). It is interesting to compare this frequent collocate cluster with *CONTAINMENT ACCIDENT PRESSURE,* which was noticed in Region I and posit as to why these two regions would choose to use the words they do to communicate the same message. There is the possibility that the choice to use *failure* over *accident* is a rhetorical move on the part of the licensee authors in Region I to appear to an audience with a different political climate than Region IV. This proves especially profound when we combine it with our earlier observation regarding the frequent co-occurrence of *failure* with *licensee* for Regions II, III, and IV, and its absence for Region I. Regardless of whether or not this variation is the result of rhetorical strategy or not, it

Table 6.19: Top 20 Collocates with *Containment* for Region IV by Industry group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
system	4.906576	68	*pressure*	6.362336	214
spray	8.660822	67	primary	7.535203	147
isolation	7.529017	55	spray	8.205293	136
inside	8.733522	49	isolation	7.762296	131
fuel	15.62264	48	failure	6.224213	102
cooling	6.475256	43	sump	7.863910	100
reactor	4.570639	43	purge	9.379925	99
sump	8.572012	42	*system*	4.504462	99
primary	7.000201	38	valves	6.222248	92
outside	7.763314	37	atmosphere	8.475502	78
pressure	5.146467	37	building	6.699385	73
building	7.218592	35	air	6.128700	69
systems	4.929785	33	*fuel*	4.639601	59
integrity	6.557244	31	leakage	5.094522	56
leakage	5.893840	31	temperature	5.256542	56
atmosphere	9.403804	29	outside	7.006466	55
heat	6.186820	28	inside	7.458313	52
removal	6.768215	28	pump	4.942037	46
within	5.047369	9	valve	5.033122	46
air	6.485450	10	hydrogen	7.482684	40

further demonstrates that even for industry key terms there is variation across both industry group and geographic contexts.

SUMMARY OF REGIONAL PROFILING

From those terms analyzed in this chapter, we can deduce that regulated nuclear discourse, as well as the meanings associated with the industry key terms in Chapter 4, varies with regard to NRC regional designation. This behavior appears to be a characteristic of sub-groupings of domain-specific language. We also observed in the analysis from this chapter that there are certain ways in which language is used in different regions that has pragmatic implications, and might possibly influence the ways in which certain licensees interact with NRC representatives. The observations made in this chapter, like this one, are not only of use to licensee affiliates and NRC representatives, but particularly to those individuals moving from one region to another. Ultimately, these analyses provided us the opportunity to learn that the behaviors of domain-specific language are influenced by other dimensions like industry group and regional affiliation, as well as results that are of use to individuals both within and without the nuclear power industry.

CHAPTER 7

CONCLUSION

In this dissertation, I examined the importance of extralinguistic contexts, like industry group and regional affiliation, on the terminology of domain-specific language, as well as the ways in which meaning is communicated through collocations. My hypothesis was that these contexts are significant as dimensions through the variation of the collocation of regulated nuclear power industry terms. I sought to test this hypothesis through three basic goals:

1. Establish foundation descriptions of this variety of domain-specific language through actual language used in this industry to identify key terms;
2. Identify differences in the language used in this industry with regard to author affiliation;
3. Identify regional patterns of language use in these documents in order to see if any specific patterns or trends emerge with regard to affiliation with either the NRC or licensees.

BUILDING THE FOUNDATION

Using the same methodology that was used in creating the Tobacco Documents Corpus at the University of Georgia, I produced a corpus of the language used in the regulated nuclear power industry with documents sampled from the NRC ADAMS public database. The reason why this methodology was used is because it was developed for use on a database that behaved like a monitor corpus, having documents added to it systematically, making it a similar database to ADAMS. Following the same protocols as the TDC, I began the creation of a Pilot Corpus by randomly selecting 0.001 of all of the documents available from the ADAMS database. This turned out to be 30 documents from each of the docket numbers that uniquely identify the licenses held for all 104 nuclear power reactors in the United States. These documents were sampled for 12 independent months that were randomly selected—one for each of the twelve years from 2000 to 2011. The documents were randomly selected using a percentage ratios established by all of the documents available in ADAMS, at the time of sampling, for the 12 months being sampled: i.e. if Palo Verde Unit 1 had .4 of its total documents available in April 2011, then 12 documents were randomly selected for that month.

Using metadata obtained from the Legacy Custom Report from the ADAMS database that was verified against each PDF downloaded, the documents were classified using a specific set of criteria: they must be related to the regulation of nuclear power (e.g. not nuclear isotopes used for cancer research), the author must be industry-internal but the audience could be either internal or external, document type, docket assignment, whether or not the document was language-based, and the length of the document had to be a minimum of 50 words of continuous discourse.

After all of the documents were classified and made into a relational database, it was discovered that their distribution reflected Zipf's Law: Most of the documents were of a few document types, while a minority of the documents were of many different document types. Because each of the documents were downloaded as PDFs that were assigned unique NRC Accession Numbers, I was able to filter out all of the documents that were duplicated because of their assignment to multiple dockets. However, I was able to maintain in the metadata which documents were assigned to multiple dockets. There were also many documents in the corpus samples that were found to have more than one document in a single PDF file. Each of these files were split into their individual documents, and their original Accession number was preserved in the file names used. For example, ML11069A001 was found to have both a Safety Evaluation and a Cover Letter. Thus, these two documents because ML11069A001_1 and ML11069A001_2 respectively and were classified individually using the same criteria. Using the corpus database created from the samples taken from ADAMS, I was finally able to filter out all of the documents that did not fit the criteria of the project. The original 3,120 documents resulted in 4,560 individual files for the Pilot Corpus.

I then performed three more iterations of the sampling, classification, and rejection procedures. At this point I was interested in discovering whether or not my sampling method was reproducible, so I analyzed the splitting and rejection ratios for all four iterations by comparing them against one another using a two proportion z-test at 99% confidence. What I discovered was that there was no statistically significant difference between the iterations with regard to the sampling procedures. Therefore, I was able to determine that the methodology was reproducible.

After I had confirmed that all three had no statistically significant difference in their rejection ratios, I combined the documents from all four iterations. When I performed this process, the decision was made to eliminate the duplicate PDF files that were sampled in different iterations due to the document being assigned to multiple dockets. Yet again, I preserved the original iteration identification in the reference corpus database metadata so that I could filter the documents according to which iterations they were sampled in if needed. After this process was completed, I determined that there were 7,110 unique NRC Accession Number identified documents in the nuclear reference corpus.

Using WordSmith Tools, I generated a frequency list for this corpus and compared it to the frequency list of the Baker Brown Corpus of General American English, which is representative of American English used in 2006 (a similar time period to that of the nuclear reference corpus). I evaluated these two lists using the Keyness function of WordSmith Tools and discovered which content words in the nuclear reference corpus were used more than expected based on the Baker Brown Corpus through the log-likelihood statistic. My threshold for statistical significance was a p-value of less than 0.0001. A stop list of function words was used to make sure only content words were part of the analysis. Over 2,000 key words were identified, so I went through this list and took the 20 terms with the highest frequencies and lowest p-values (Table 7.1). A key word was selected to be a term if it either clearly was used in the regulated nuclear power industry, like *nuclear*, as well as terms that took on a sense of this industry like *staff*, i.e. *NRC staff*. This was performed by generating a concordance for each term and looking at the word in question in the context of its use.

The next step was to identify the 20 most frequently used content words to co-occur with each key term. The collocates were limited to a four word span to either side of the node word. Mutual Information scores were calculated for each word that co-occurred with the key term in order to identify whether or not the 20 most frequent collocates had a

Table 7.1: Top 20 Key Terms from the Nuclear Reference Corpus

Key word	Keyness	Freq.
NRC	6554.320801	30454
safety	4217.895508	24259
reactor	4184.859863	19611
plant	4061.947266	22366
inspection	3932.327637	18956
unit	3893.398193	20170
licensee	3592.829834	16660
system	3349.659912	26065
power	3248.645264	24747
nuclear	2641.138184	16731
fuel	2567.255615	13988
license	2450.459473	12035
containment	2348.273926	11139
operating	2334.495361	12775
evaluation	2284.026611	11760
staff	2250.520752	14860
technical	2210.776123	12022
emergency	2090.845947	11814
pressure	2042.388794	13536
inspectors	1979.101563	9369

significant semantic bond between them and the node. The MI score had to be greater than 3.0 and the collocate had to have a minimum of 10 co-occurrences with the node in order to be deemed statistically significant. These collocates were used to create lexical profiles for the terms so as to identify the meaning of each term as characterized by its use in the corpus from the perspective of the entire industry: Observed Meaning.

During this phase an abundance of prefabricated lexical "chunks" or clusters that involving each term was noted, often in co-occurrence with other terms. In many of the terms that were characterized by formulaic language, extremely high Mutual Information scores for all of the 20 most frequently used collocates were extremely high (over 10).

Moreover, when analyzed with regard to industry groups these chunks appeared to collocate more with sub-technical or non-technical language for the NRC and sub-technical or technical language for the licensees.

INFLUENCE OF INDUSTRY GROUP MEMBERSHIP

It was found that the documents in the nuclear reference corpus were distributed such that over 95% of them were written by either NRC or licensee-affiliated authors (Table 7.2).

Table 7.2: Document Distribution by Regulated Nuclear Industry Group

	NRC	Licensees	Vendors	Government Agencies (Non-NRC)
# Of Documents	3,354	3,473	199	84
% Of Total	**47.17%**	**48.85%**	**2.8%**	**1.18%**

Due to the significant presence of these two industry groups in the corpus, they were the ones that were compared to see if organizational affiliation is a dimension of domain-specific language that leads to variation in the use of industry terms. The first step that was taken was to analyze the frequency of nuclear industry term use through keyness analysis in WordSmith Tools.

All of the documents from the reference corpus that were designated as having been written by a representative of the NRC were compared to those authored by a licensee-affiliated individual. It was found that NRC-authored texts used *NRC, staff,* *licensee, license, evaluation, safety, inspection, inspectors, plant,* and *nuclear* more than expected in comparison to the licensee-authored documents. On the other hand, licensees used *reactor, unit, system, power, fuel, containment, operating, emergency,* and *pressure* more than was expected. It was discovered that *technical* was used no differently in comparison of one group to the other. Also, all of the terms found to have positive keyness for one of the groups were also found to have negative keyness for the other.

Three of the terms having positive keyness for the NRC indicate the power and authority of this organization *(inspection, evaluation, license)*. There were also terms used by the NRC that indicate larger entities *(NRC, staff, licensee, inspectors, plant)*, versus those used more by the licensees that represent a local or plant-based perspective *(unit, pressure, fuel, systems)*. There also seemed to be a difference in the positive keyness of industry terms for these two groups with regard to abstract notions for the NRC *(safety, nuclear)* in comparison to more applied concepts for the licensees *(emergency, containment, power)*.

When the 20 most frequent collocates were calculated for the 20 industry terms from the documents authored by each of these two groups, it was found that the collocates were idiosyncratic: neither group had the exact same collocates in the same ranking order, nor did they have the same MI scores quantifying the semantic bond between those collocates and each respective term. Furthermore, these collocates were scaled differently in their rankings in comparison to both the industry and one another. Lexical profiles were also generated for each term with these collocates in order to determine if variation existed between the Observed Meaning of each term between both industry groups and in comparison to the industry-level analysis in Chapter 4.

INFLUENCE OF REGIONAL DESIGNATION

The NRC has divided the nuclear power industry in the United State into four geographic regions. These four jurisdictions have their own NRC officials who represent the headquarters in Rockville, Maryland, and who work to provide oversight and regulation on a local level to the licensees in the states for which they are responsible. Although all of the dockets were sampled equally, it was found that there is one region that has more documents per reactor/license than the others: Region IV (Table 7.3).

Table 7.3: Number of Documents in Nuclear Reference Corpus by NRC Region

	Region I	Region II	Region III	Region IV
Number of Documents	1,763	2,039	1,650	1,598
Number of Documents (Normalized by # of Reactors)	67.808	61.788	68.750	76.095

When the patterns of term use were evaluated for each region in comparison to the others, it became evident that there is a connection between the use of the 20 industry terms with respect to geography in addition to industry group, as well as the interaction of geography and organizational affiliation.

It was discovered that Region I uses terms relating to the NRC and its identity as administrator and regulator more than expected in comparison to the rest of the industry *(license, nuclear, staff, NRC)*. Conversely, there was less of a focus on using those key terms that relate to plant operations and inspection than Regions II, III, and IV. When these tendencies were analyzed with regard to industry group affiliation of the author as well, it was found that the NRC-authored documents for this region had more use of *license, nuclear, staff,* and *NRC* than was expected. They also had less of the same terms from the regional level in addition to *safety, emergency, technical* and *unit* in comparison to the NRC-authored documents in the other three regions. The licensee-authored documents in Region I also shared the same focus on *license* more than the licensees of the other regions, and they also used terms related to standards *(technical, emergency, unit)* more than expected as well. With regard to those terms used less frequently by Region I licensees than their counterparts in Regions II, III, and IV, we learned that they were *inspection, licensee, NRC, operating,* and *plant.*

In Region II, all of the documents associated with dockets in these states used *fuel* and *unit* more than expected in comparison to the rest of the country. They also used terms related to NRC oversight less than expected *(inspectors, license, nuclear, emergency, reactor, safety, technical)*. For those documents in this region that were authored by NRC representatives, we found more of a focus on the *licensee* and its *plant* than was anticipated *(fuel, unit, plant, licensee, power*), and there was much less use of NRC oversight terms *(inspection, inspectors, NRC, nuclear, safety, reactor, emergency, technical)*. Alternatively, the licensees in this region focused more on the *NRC* and its *inspection* of the *plant (NRC, inspection, unit, fuel, pressure)* than was expected based on the frequency of these terms in Regions I, III, and IV. They also communicated those terms associated with being regulated *(license, licensee, technical, emergency)* far less than was expected in comparison to their industry counterparts.

Region III is characterized by more application of terms related to the oversight of licensees as *reactor* operators than the other three regions *(inspection, inspectors, licensee, plant, reactor, system*). While on the other hand, they use *emergency, staff,* and *unit* far less than is expected. We can see where these tendencies come from by looking at each industry group in this region in comparison to their complements in the industry. The NRC-authored documents for Region III use those terms related to oversight of the licensee on the plant-level far more than was expected *(inspection, inspectors, licensee, plant, reactor, power, pressure, safety, system*), while they were less focused on the NRC as an administrator of licenses *(license, staff, NRC, nuclear).* The terms used more than expected in documents authored by licensee representatives for this region form a cluster related to *reactors* at the plants *(nuclear, reactor, plant),* while having a negative keyness for the use of terms relating to the *NRC* and the application of safety *(emergency, NRC, technical, unit, containment, safety).*

In Region IV's documents we find more talk of *licensee* and the application of *safety* than was expected based off of Regions I, II, and III *(emergency, safety, technical, licensee),* but there was less use of *nuclear, fuel,* and *unit.* The NRC authors in this region had more of a focus on oversight of standards and *safety* than was expected *(emergency, inspection, licensee, safety, technical, pressure),* while also demonstrating less use of *plant* terms *(fuel, nuclear, plant, power, unit).* Similarly, the licensee-authored documents for this region demonstrated more lexical focus on the NRC oversight of *safety* and the application of *safety* than expected based off of licensees in the other regions through the use of *emergency, inspection, licensee, NRC, safety,* and *staff.* Then again, they also used terms related to the *reactor* less than those same groups *(fuel, nuclear, reactor, system, unit, pressure)*.

It is important to note that these results indicate frequency of use in the comparison of one group of documents to another. These results are not meant to indicate that one region has a lack of concern in a certain area. What they are indicating is that some regions communicate using these terms more than the others. This being said, we can see that there is some disparity between the keyness of terms between industry groups belonging to the same region.

It might be expected, or even assumed, that these groups would have the same terms be significant with regard to keyness since they are communicating with one another. However, there are some instances where there were differences present. For example, in Region I, we find that there is a distinctive difference with regard to terms relating to the NRC and its role and identity as the authority figure in this industry. While the NRC-authored documents for this Region use these terms more than is expected in comparison to all of the other regions' NRC-authored texts, the licensees in Region I are using these terms far less than is expected in comparison to the other licensees. This same type of disparity is present also in Region I with the licensees using *safety* and *technical,* words related to standards, more than is expected while the NRC-authored documents in this region use it less in comparison to their own equivalents in the other regions.

We can see these same patterns present in the comparison of industry group use of key terms within Regions II and III. While industry groups in each of these regions respectively demonstrate agreement in the significant use of terms related to plant topics like *fuel* (Region II) or *reactor* (Region III), they have a discrepancy in the focus on the NRC in Region III and NRC assessment and oversight in Region II. The region that demonstrated the most agreement across both industry groups with regard to nuclear industry term use was Region IV. In this region both groups used terms relating to the application of *safety* and *technical* standards more than was expected in comparison to the industry groups of Regions I-III. They also seemed to share the characteristic under use of terms related to *plant* and *reactor*, their *nuclear* identity. Instead, it seems as though this region has some consensus in their communications on an interest in *safety,* assessment, and standards on the plant level.

VARIATION IN OBSERVED MEANING

When the lexical profiles and Observed Meanings were recognized with regard to industry group and geographic context, it was discovered that the meanings for each key term did vary and change in comparison to those documented on the industry level. Industry group affiliation indicated that most of the terms reflected the perceptions of each group with regard to their roles in the industry. Most of the terms began co-occurring with words related to evaluation and authority for the NRC-authored documents, and many of the collocations for the terms in the licensee-authored texts were related to plant-level operations. There were also some variation in the types of collocates noticed for these two groups when the influence of geography was also factored in by comparing the collocates of the NRC and licensee groups across the four regions.

The evolution of the variation exhibited in the collocates and Observed Meaning of the 20 industry terms through the interaction of the industry group and geographic dimensions will be demonstrated with a review of *NRC* and *licensee.*

For the term *NRC* we found that the collocates clustered into two significant groups: those that are information-related *(review, site, letter, information, web, ADAMS, response, request)* and those that have to do with *evaluation (evaluation, inspection, approval, safety*). The most frequent clusters involving this term and its 20 most frequent collocates also help to communicate the meaning of *NRC* and its role in the industry (Table 7.4).

Table 7.4: Industry Level *NRC* Collocate Clusters

Cluster	Frequency
THE NRC STAFF	5,294
NRC WEB SITE	686
APPROVED BY THE NRC STAFF	604
NRC PUBLIC DOCUMENT ROOM (ADAMS)	476
BY THE NRC FOR PUBLIC INSPECTION	392

Once all of these factors were taken into consideration, it was determined that the Observed Meaning for the term *NRC* was that **the *NRC* is composed of staff who work with information for evaluation purposes and provide information to the public for their evaluation.**

When we look at how this term was used in NRC-authored texts in comparison to those written by licensees, we find that there are some considerable differences in the collocates for these two groups. Both of them do share some of the same collocates for

NRC like *staff, approved,* and *evaluation.* But it should be noted that these words have very different frequencies and rankings between the two groups. Additionally, the groups had very different MI scores related to their idiosyncratic collocate lists. The MI scores for the licensees were significantly higher than those for the NRC. We saw that *NRC* was attracted strongest to *inspection* for NRC-authored documents and *safety* for the licensees. Another trend that was noted between these two groups was the different words used by each of them to make clear the *NRC's* role in the industry: how the NRC perceives of itself versus how the licensees perceive the *NRC.* We found that the NRC authors often used verbs like *concludes, determined, reviewed,* and *approved* in clusters like those found in Table 7.5, while the licensees primarily used *approved*.

Table 7.5: Clusters of *NRC* Indicating Authority by Industry Group

NRC-Cluster	Frequency	Licensee-Cluster	Frequency
NRC STAFF CONCLUDES	453	*THE NRC APPROVED*	201
NRC ENFORCEMENT POLICY	357	*REVIEWED AND APPROVED BY THE NRC*	90
NRC STAFF HAS DETERMINED	264	*PRIOR NRC APPROVAL*	86
NRC STAFF HAS REVIEWED	229	*NRC APPROVED METHODOLOGY*	76
APPROVED BY THE NRC	223		

We also learned through the collocates the perspective of the licensee that they are to respond to the *NRC*: *SUBMITTED TO THE NRC* (205), *RESPONSE TO NRC REQUEST FOR ADDITIONAL INFORMATION* (51), *RESPONSE TO NRC*

QUESTIONS (24). The result of the variations in the collocations found for *NRC* is a change in the Observed Meaning depicted in Figure 7.1.

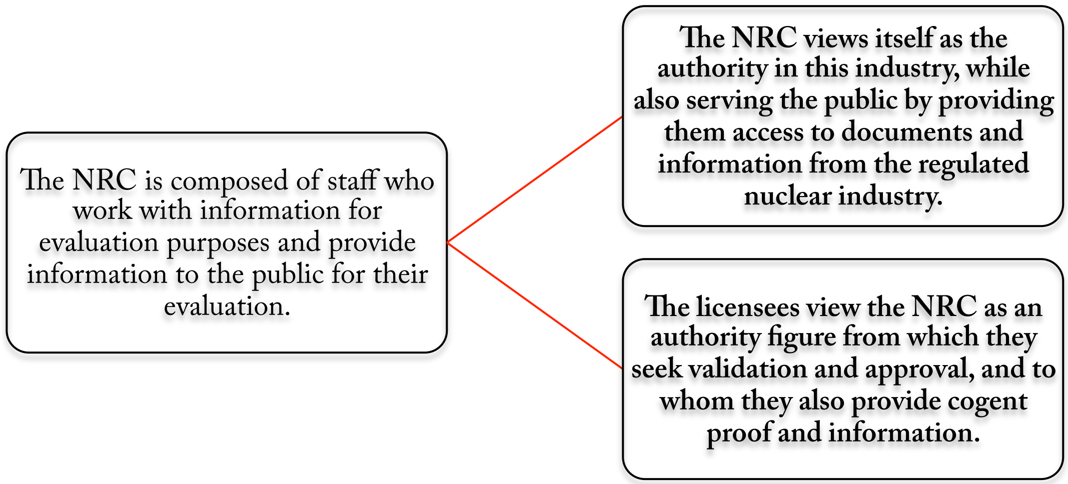

Figure 7.1: *NRC* Observed Meaning Map—Industry Group

When we go even further to compare the ways in which *NRC* is used by the NRC across the four regions, we find that the view of the NRC being the authority on *policy* is more frequently used by Region IV, as this collocate was not found in the top 20 lists for the other regions. This same type of observation was made for *safety* in Region III. Switching over to the licensee perspective across the four regions, we noticed the presence of *consider, recommend,* and *AMERGEN* (the name of one of the licensee corporations) in the collocate lists for Regions I and III. Putting these words into the context of their clusters, we learned that for the licensees in these regions, they associate the *NRC* with an entity to whom *AMERGEN* can make recommendations for the NRC to consider regarding the language used by the NRC in its regulation of licenses (Table 7.6).

Table 7.6: NRC Region I and II Clusters for *NRC* with *Consider* or *Recommend*

NRC Region I and III Clusters
EXELON AMERGEN RECOMMEND THAT THE NRC CONSIDER
THE NRC CONSIDER REVISING
THE NRC CONSIDER REVISING THE FIRST SENTENCE
THE NRC CONSIDER REMOVING
THE NRC CONSIDER REMOVING ALL
THE NRC CONSIDER DELETING
THE NRC CONSIDER DELETING ALL
THE NRC CONSIDER REWORDING THE PARAGRAPH
THE NRC CONSIDER ELIMINATING THIS OBJECTIVE
THE NRC CONSIDER REMOVING ALL REFERENCES

When we compare the language found in Table 7.6 to the cluster *REVISE SENTENCE TO READ* that was found in a concordance for *sentence* in all documents authored by licensees in Region IV, it becomes clear that the Exelon licensees in Regions I and III are quite direct with their recommendations to the NRC about language revisions, yet at the same time using politeness strategies through the use of *recommend* or *consider*. This is made apparent when juxtaposed with the common syntax in Region IV that involves an imperative with an implied "you" for the *NRC.* All of these observations help us to see that there is a great degree of variation in the meaning associated with *NRC* when you look at it from the different industry group and geographic contexts, especially with respect to the *NRC's* authority (Figure 7.2)

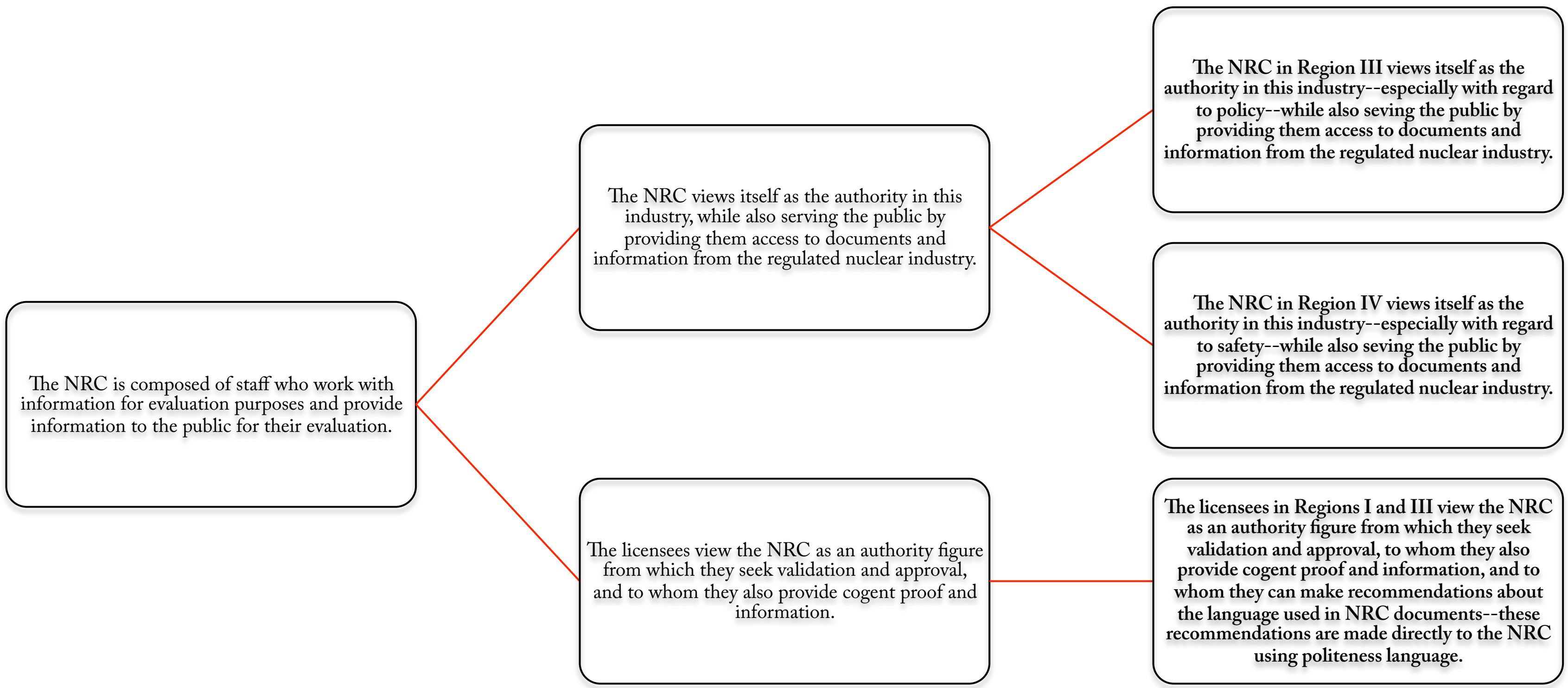

Figure 7.2: ***NRC*** **Observed Meaning Map—Industry Group and Region**

If we look at the variation of the collocates and Observed Meaning of the other main group involved in this industry, *licensee*, we find some interesting patterns with regard to the influence of industry group and geography on the perceptions of this term's role in the regulated nuclear power industry. On the industry level, the collocates for *licensee* can be divided into three semantic sets: one has to do with **analyzing** information *(identified, determined, reviewed)*, the second concerns the **dissemination** of information *(stated, proposed, provided, requested, submitted)*, and the third involves the **evaluation** of information by the *NRC (NRC, inspectors, failed, performed*). We learned even more about the role of information from the most frequent clusters associated with this term: the *licensee* is the source of it; and it is part of the communication circuit with the *NRC*, its *staff*, which includes *inspectors* (Table 7.7).

Table 7.7: Industry Level *Licensee* Collocate Clusters

Cluster	Frequency
THE LICENSEE STATED THAT	522
THE LICENSEE FAILED TO	402
LICENSEE IDENTIFIED VIOLATIONS	251
THE LICENSEE PROPOSED	231
THE LICENSEE DETERMINED THAT	197
INSPECTORS ASKED THE LICENSEE	98

It was also determined that the *licensee* receives instruction from through the presence of *shall* as a most frequent collocate: *THE LICENSEE SHALL* (417), *LICENSEE SHALL OPERATE* (157). When these patterns were all put into context with one another, it became clear that the Observed Meaning for the term *licensee* on the industry level was that **the *licensee* reports to the *NRC*. Its role in this relationship is to analyze information, provide it to the *NRC*, and receive evaluations and assessments.**

When we look at the differences in how this term was used in NRC-authored texts in comparison to those written by licensees, we saw again that both collocate lists were idiosyncratic. The second major pattern to be noticed was the presence of a significant amount of different verbs present in both lists. These verbs were organized into a Venn diagram in order to see the difference in the perception of what a *licensee* does, as well as what it has power over from the perspectives of both groups Figure 7.3.

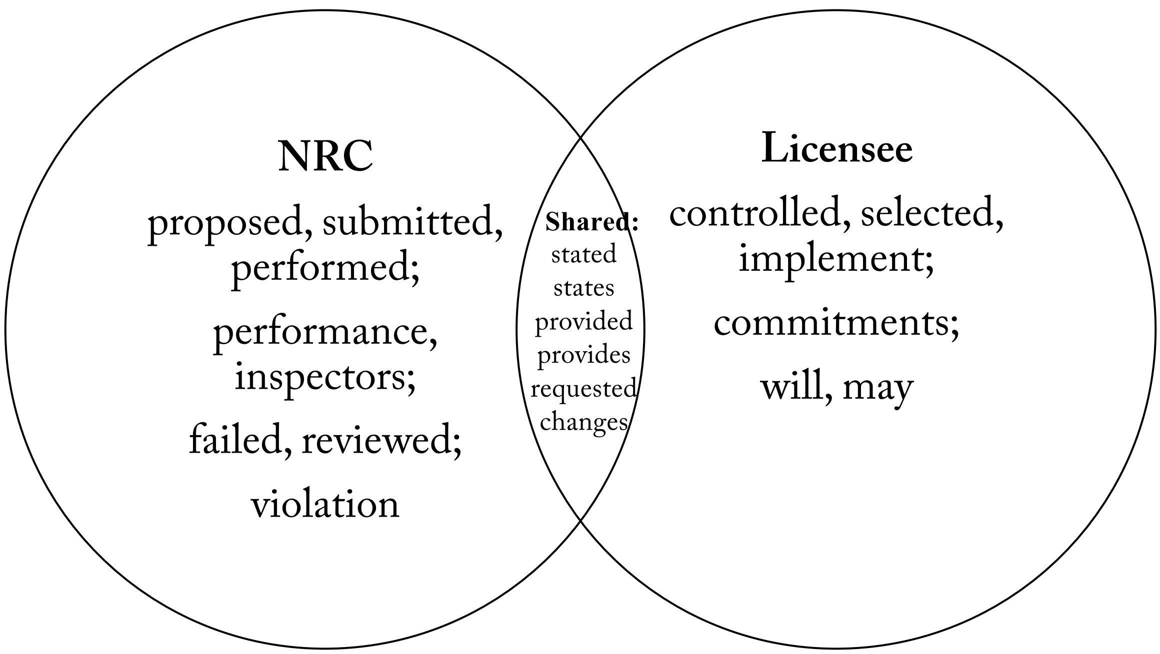

Figure 7.3 Venn Diagram of *Licensee*

Many of the NRC collocates for *licensee* appeared to be associated with evaluating *licensees (proposed, submitted, performed, performance, information, inspectors, failed, reviewed, violation)*. The licensee perspective let us know what a *licensee* does in this industry *(stated, provided, requested, states, provide, changes)*. What became most apparent from looking at Figure 7.3 is that from the NRC perspective a *licensee* is something that requires evaluation, while the licensees perceive themselves as having control and power over the choices they make and the actions they take (which are also evaluated by the NRC).

When we bring into context all of the collocates associated with licensee for these two industry groups, we find that they have very different yet complementary perceptions of what a *licensee* is (Figure 7.4).

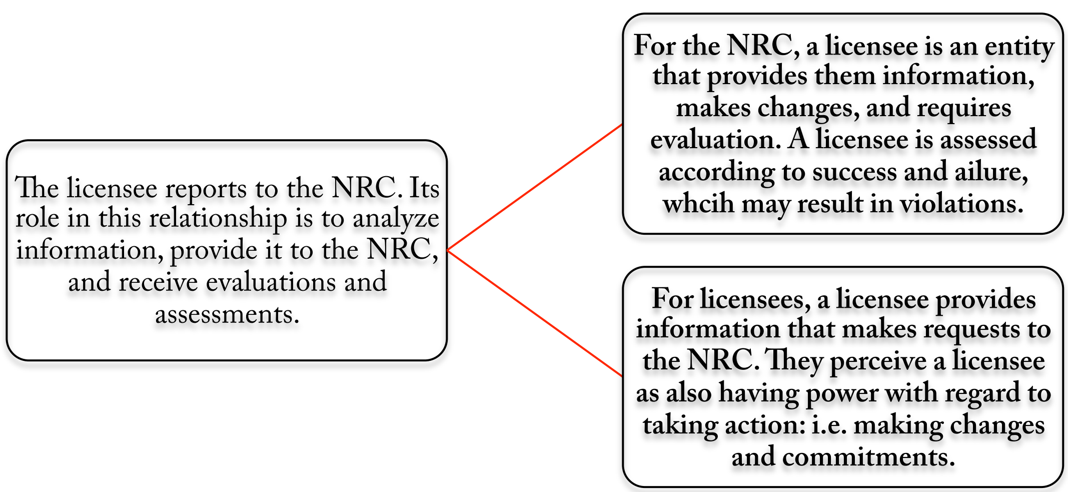

Figure 7.4: *Licensee* Observed Meaning Map—Industry group

By looking at the variation between these Observed Meanings again with regard to the interaction of industry group and geography, we found that there is essentially the same sense for *licensee* from the NRC-authored texts, albeit with more specificity. The biggest difference we noticed in the Observed Meaning for licensee was found in the collocate clusters involving *licensee* and *failed* between Regions II, III, and IV. For Region II, we found that the types of failure most frequently associated with *licensee* was communicated with the infinitival verbs *to meet* (8), *to protect* (8), *to implement* (7), *to ensure* (7), *to determine* (6), and *to identify* (5), while one cluster communicated the quality associated with the failure: *adequately* (7). The failures involving protection and determination were unique to this region in comparison to the clusters from Regions III and IV. In Region III, the clusters involving *licensee* and *failed* also included the verbs *to ensure* (15), *to implement* (14), *to identify* (14), *to perform* (8), *to take* (7), and *to meet* (7). We learned that in this region as well the failure could come in different qualities *specifically* (17), *adequately* (9), and *appropriate* (7)—while also involving *personnel* (9). The unique attributes for this region were *specifically*, *personnel*, and *to take appropriate*. In Region IV, we find that there are also a great many verbs associated with *failed: to identify* (16), *to perform* (14), *to evaluate* (13), *to establish* (12), *to ensure* (10), *to provide* (6), *to submit* (6). There were also failures related to being prompt (6), properness (8), as well as adequacy (7). The aspects of failing that were unique to Region IV were *to evaluate* (13), *to establish* (12), *to properly* (8), *to provide* (6), *to submit* (6), and *to promptly identify* (6). It should also be noted that Region IV had the greatest amount of clusters involving *failed* and *licensee*. These clusters help us to further develop the notion of a *licensee*.

SIGNIFICANT CONTRIBUTIONS FROM THIS STUDY

By creating a domain-specific corpus of regulated nuclear power industry language I have demonstrated that the Tobacco Documents Corpus methodology is reproducible for other varieties of domain-specific language, especially when the population of documents to be sampled comes from a database that acts like a monitor corpus. This confirms the assumption made by Darwin about the effectiveness of this methodology, as he was unable to verify reproducibility for the TDC due to a lack of information during the rejection phase of sampling.

The results of this study not only also provide evidence to support Kretzschmar's assertion that extralinguistic factors like industry group membership and geography are important dimensions for linguistic analysis through corpus linguistics, but it does so specifically for domain-specific language. The meanings for each of the industry terms identified from the nuclear reference corpus varied in their collocational frequencies when these contexts were factored into the analysis. For example, the analysis of these terms from an industry group perspective revealed that there is a relationship between the situation of power in this industry—who has the authority and over whom—and the ways in which the key terms are used. The geographic perspective helped us to better understand the influence of a local perspective on the industry terms with regard to situational use and specificity: a lesson that could be passed on to nuclear engineering students who can begin developing a more local mindset with regard to the language they use, as this will be expected in the regulated nuclear power industry. Moreover, we also were able to uncover that there were definite differences in the ways in which the different industry groups communicated with one another in different regions: an observation that could be of use to the nuclear power industry. Ultimately, we now know that these contexts, as well as their interaction, play a definite role in the language of the nuclear industry.

Analysis of the nuclear reference corpus also revealed that this variety of domain-specific language is characterized by a large amount of "chunks" of formulaic language. Many of the terms were involved in these prefabricated lexical units, even with regard to industry group and regional contexts. Some of these an industry terms were classified as formulaic, due to their propensity to make these clusters. Moreover, many of these terms exhibited the characteristic of having all of their 20 most frequently co-occurring words possessing an MI score of over 10. Another characteristic of this prefabricated language was that they often involve sub-technical and non-technical terms when used by the NRC, and the clusters formed by the licensees involved more sub-technical and technical terms.

The results from this study also demonstrate that the Engineering English described and analyzed by Mudraya that was sampled from textbooks as part of the SEEC corpus is quite similar in nature to that used in practice within the regulated nuclear power industry: a business based on engineering. This is especially clear with regard to the use of sub-technical and non-technical writing, as well as the use of academic verbs that co-occurred with *NRC, licensees, staff, inspectors,* etc.

The final significant observation resulting from this study relates to the frequent use of *safety* in this industry at both the industry group and geographic dimensions, as well as its frequent presence as a collocate for many of the other terms. In Chapter 1, I shared the perspective of nuclear industry language that has been proposed by the author of *Nukespeak* that there is an attempt "to hide the truth about nuclear dangers" by the parties involved in this industry (Bell 2011). The data, as it has been observed in this corpus, implies that there is a significant amount of discussion on *safety, emergency,* even *violations* on the part of licensees, and *accidents* and *failures* that happen with regard to *fuel* and *containment.* Whether or not people in this industry are trying to "hide" information about the dangers of the nuclear industry, I cannot attest to. But what I can say, with proof, is that there is significant discussion in this industry regarding these issues.

IMPLICATIONS FOR FUTURE STUDY

There is a great deal of opportunities for future research that have resulted from this initial foray into examining the language of the regulated nuclear power industry. The first opportunity for future research is to take this analysis one step further and to analyze key term use with regard to corporation. As the analysis was concentrated on more of a local level with regard to geography, we noticed more instances of key terms co-occurring with specific licensees and corporations—like AMERGEN with *NRC.* The influence of corporate affiliation with language use also needs to be explored in domain-specific language with regard to the changes in industry language and term use when entities change corporate hands. In the last 12 years in the nuclear industry there have been several licenses that been bought, sold, transferred and acquired by different corporations. If industry group and geographic dimensions have such a powerful influence on language use, I would hypothesize that corporate affiliation would as well through differences in corporate culture.

Another implication for future study would be an investigation of the importance and influence of certain documents on the rest of the industry. From the licensee perspective, providing references and cogent information is an integral part of the nuclear power industry, as well as its language. Thus, it would be an interesting exercise to see which industry documents are cited more or less than others, as well as to see how clusters from those documents become used across the industry.

Through the results in this dissertation, it was determined that there is a difference in the domain-specific language with regard to the interaction of extralinguistic context. It would be interesting to take this analysis one step further to see if there are differences in the language of evaluation on the part of the NRC with regard to region and/or licensee. There also exists the possibility of further analysis to be performed on the language used in this industry with regard to text-types and registers. Further analysis is also needed with regard to evaluation language in this industry, as well as the differences in language used by the NRC and licensees for internal audiences versus the general public.

Obviously, there is a great deal of research still to be done with regard to domain-specific language in general, and more specifically the language of the regulated nuclear power industry. The interpretations made as a result of this study, while informative to the field of linguistics and ESP pedagogy, also have a place in the industry that was modeled. The characteristics of this variety of domain-specific language outlined in this study, as well as those to be uncovered in future research, can be applied to the areas of communications, regulatory affairs, and even management. Ultimately, the era we are in where technology allows the general public access to vast amounts of documents that were previously much more cumbersome to analyze in paper form, especially from government-regulated industries, provides us ample resources and opportunity for learning even more about domain-specific language and how it is affected by different contexts. My hope is that those who read this dissertation will become inspired to start their own journeys into these new frontiers of domain-specific language research.

Bibliography

"About NRC." *The United States Nuclear Regulatory Commission.* The United States Nuclear Regulatory Commission, 15 Apr. 2011. Web. 4 Nov. 2011.

"About the Collins Corpus and the Bank of English." *myCOBUILD.com.* Harper Collins Publishers. Web. 1 Mar. 2013.

Adamic, Lada A., and Bernardo A. Huberman. "Zipf's Law and the Internet." *Gleometrics* 3 (2002): 143-150. Print.

"ADAMS Public Documents." *The United States Nuclear Regulatory Commission.* The United States Nuclear Regulatory Commission, 11 Apr. 2011. Web. 4 Nov. 2011.

Baker, Paul. *Using Corpora in Discourse Analysis.* London: Continuum, 2006. Print.

———. *AmE06 Corpus*. 2006. Web. 4 Nov. 2011. http://www.ling.lancs.ac.uk/staff/paulb/AmE06zipped.zip.

Bell, Richard C. "No Word for Meltdown: The Return of Nukespeak." *Nukespeak*. Nukespeak, 16 Mar. 2011. Web. 4 Nov. 2011.

Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8:243-57.

———. *Dimensions of Register Variation: A Cross-Linguistic Comparison.* Cambridge: Cambridge UP, 1995.

Bodansky, David. *Nuclear Energy: Principles, Practices, and Prospects.* New York: Springer Publishers, 2004. Print.

Bolinger, D., 1976. Meaning and memory. Forum Linguisticum 1, 1–14.

Burkette, Allison. "The Lion, the Witch, and the Armoire: Lexical Variation in Case Furniture Items." *American Speech* 84.3 (2009): 315-339. Print.

Burnard, Lou. "Metadata for Corpus Work." *Developing Linguistic Corpora: A Guide to Good Practice.* Digital. 2004. Web. 4 Nov. 2011.

Darwin, Clayton M. "Construction and Analysis of the University of Georgia Tobacco Documents Corpus." Diss. The University of Georgia, 2008. Digital.

"The Freedom of Information Act." *The United States Department of Justice.* The United States Department of Justice, 1996. Web. 4 Nov. 2011.

Firth, J.R. "A Synopsis of Linguistic Theory, 1930-1955." *Studies in Linguistic Analysis*. Oxford: Blackwell, 1957. 1-32. Print.

Francis, W. Nelson and Henry Kučera. *Manual of Information to Accompany a Standard Sample of Present-Day Edited American English, For Use with Digital Computers*. Providence, RI: Houghton Miflinm 1964. Print.

Halliday, M.A.K. "Lexicology." *Lexicology and Corpus Linguistics.* Eds. M.A.K. Halliday, Wolfgang Teubert, Colin Yallop, and Anna Čermáková. London: Continuum, 2004. 1-22. Print.

Hiltgartner, Stephen, Richard C. Bell, and Rory O'Connor. *Nukespeak: The Selling of Nuclear Technology from the Manhattan Project to Fukushima.*Washington, D.C.: Sierra Club, 2011. Digital.

Hockey, Susan. "The History of Humanities Computing." A Companion to Digital Humanities. Eds. Susan Schreibman and Ray Siemens. Malden, MA: Blackwell, 2004. 3-19. Print.

Johansson, Stig. “Computer Corpora in English Language Research.” *English Computer Corpora: Selected Papers and Research Guide.* Ed. Stig Johansson and Anna-Brita Stenström. New York: Mouton de Gruyter, 1991. 3-32. Print.

Johansson, Stig, Geoffrey N. Leech, and Helen Goodluck. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Oslo: Department of English, University of Oslo, 1978. Print.

Johnson, Ellen. “Lexical Change and Variation in the Southeastern United States in the Twentieth Century.” Diss. The University of Georgia, 1992. Print.

Kalton, Graham. “Introduction to Survey Sampling.” *SAGE University Paper 35 (1983).* Thousand Oaks, CA: SAGE Publications. Print.

Kemp-Dynin, Marlene. “The ‘Company’ Words Keep: A Corpus-Based Analysis of Collocational Patterning in Business Terminology.” Diss. The University of Georgia, 2005. Digital.

Kennedy, Graeme. *An Introduction to Corpus Linguistics.* London: Addison Wesley Longman, 1998. Print.

Kirman, Alan, "Pareto, Vilfredo (1848–1923)." *The New Palgrave Dictionary of Economics Online*. Eds. Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan, 2008. Web. 28 Aug. 2012.

Kjellmer, Göran. “Aspects of English Collocations.” *Corpus Linguistics and Beyond.* Ed.William Meijs. Amsterdam: Rodopi B.V., 1987. 133-140. Print.

Kretzschmar, Jr., William A. *The Linguistics of Speech.* Cambridge: Cambridge UP, 2009. Print.

———. "Sampling Plan for Creation of Corpora for the Tobacco Documents Grant." 2001. In "Construction and Analysis of the University of Georgia Tobacco Documents Corpus." Clayton M. Darwin. Diss. The University of Georgia, 2008. Digital.

Kretzschmar, William A., Jr., Clayton Darwin, Cati Brown, Donald L. Rubin, and Douglas Biber. 2004. Looking for the smoking gun: Principled sampling in creating the Tobacco Industry Documents Corpus. *Journal of English Linguistics* 32 (1): 31-47.

Kučera, Henry and W. Nelson Francis. *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press, 1967. Print.

Larsen-Freeman, Diane and Lynne Cameron. *Complex Systems and Applied Linguistics.* Oxford: Oxford UP, 2008. Print.

Malinowski, Branislaw. "The Problem of Meaning in Primitive Languages." *The Meaning of Meaning: A Study of Influence of Language Upon Thought and of the Science of Symbolism.* Eds. C. K. Ogden and I. A. Richards. New York: Harcourt, Brace and World, 1923. 296-336. Print.

McCarty, Willard. "Modeling: A Study in Words and Meanings." *A Companion to Digital Humanities.* Eds. Susan Schreibman and Ray Siemens. Malden, MA: Blackwell, 2004. 254-272. Print.

McEnery, Tony and Andrew Wilson. *Corpus Linguistics: An Introduction.* Edinburgh: Edinburgh UP, 2001. Print.

McEnery, Tony, Richard Xiao, and Yukio Tono. *Corpus-Based Language Studies: An Advanced Resource Book.* New York: Routledge, 2006. Print.

Meyer, Charles F. *English Corpus Linguistics: An Introduction.* Cambridge: Cambridge UP, 2004. Print.

Merschoff, Ellis. Face-to-face interview, June 2011.

Moudraia, Olga. "The Student Engineering English Corpus." *ICAME Journal* 28 *(2004):*139-43.

"NRC—Public Document Room." *The United States Nuclear Regulatory Commission.* The

Mudraya, Olga. "Engineering English: A lexical frequency instructional model." *English for Specific Purposes 25 (2006)*: 235-256. Print.

United States Nuclear Regulatory Commission, 03 May 2011. Web. 4 Nov. 2011.

Palmer, Harold and A. S. Hornby. *Second interim report on English collocations*. Tokyo: The Institute for Research in English Teaching, Department of Education, 1933. Print.

Partington, Alan. *Patterns and Meanings: Using Corpora for English Language Research and Teaching.* Philadelphia: John Benjamins, 1998. Print.

Sinclair, John. Collocation: A progress report. In *Language topics: Essays in Honour of Michael Halliday*, ed. Ross Steele and Terry Threadgold. Philadelphia: J. Benjamins, 1987. Print.

———.*Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991. Print.

———. "Corpus and Text: Basic Principles." *Developing Linguistic Corpora: A Guide to Good Practice.* Ed. Martin Wynne, 2004. Web. 4 Nov. 2011.

Stubbs, Michael. *Words and Phrases: Corpus Studies of Lexical Semantics.* Malden, MA: Blackwell Publishing, 2002. Print.

Thorndike, Edward L. *Teacher's Word Book.* New York: Teacher's College, Columbia University, 1921. Print.

Walker, Samuel J. and Thomas R. Wellock. *A Short History of Nuclear Regulation, 1946–2009*. U.S. Nuclear Regulatory Commission, 2010. Digital.

Wittgenstein, L. Philosophical Investigations. Oxford: Blackwell, 1953.

"Withholding of Sensitive Information for Nuclear Power Reactors." *The United States Nuclear Regulatory Commission.* The United States Nuclear Regulatory Commission, 18 Apr. 2011. Web. 4 Nov. 2011.

APPENDIX A

INDUSTRY EXTERNAL DESIGNATION LIST

EXTERNAL AUTHORS
American Coal Ash Association
Ann Riley & Associates, Ltd
Ashtabula County, OH
Avila Valley Advisory Council
Barnegat Bay National Estuary Program
Beyond Nuclear
Blue Ridge Environmental Defense League
Brattleboro Reformer
Brunswick County, NC
C-10 Research & Education Foundation, Inc.
Cambria Legal Foundation
Canadian Coalition for Nuclear Responsibility
Center for a Sustainable Coast
Central Coast Peace & Environmental Council
Citizens Allied for Safe Energy, Inc. (CASE)
Citizens Awareness Network
Citizens for Alternatives to Chemical Contamination
Citizens for Renewable Energy
Citizens Resistance at Fermi Two
Citizens' Utility Board, Clean Wisconsin
City of Harrisburg, PA
City of Manitowoc, WI
Coalition for a Nuclear Free Great Lakes
Connecticut Coalition Against Millstone
Don't Waste Michigan, Home for Peace & Justice
Ecology Party of Florida
Environmental Ctr of San Luis Obispo County
Everglades Law Center, Inc.
Friends of the Coast
Friends of the Earth
Georgia Women's Action for New Directions
Georgians for Clean Energy
Giblin & Nickerson, PA
Grandmothers, Mothers & More for Energy Safety
Great Rivers Environmental Law Ctr

Green Party of Ohio
Harmon, Curran, Spielberg & Eisenberg, LLP
Hudson River Sloop Clearwater, Inc.
Huron Environmental Activist League
Indigenous Environmental Network (IEN)
Institute for Energy & Environmental Research
International Institute of Concern for Public Health
Inwater Research Group, Inc.
J/R/A Associates
Jacobson, Buffalo, Magnuson, Anderson & Hogen, P C
Jersey Shore Nuclear Watch
Kalamazoo River Protection Association
Kauffman & Eye
Kickapoo Traditional Tribe of Texas
Lac du Flambeau Band of Lake Superior Chippewa
Lone Tree Council
Long Island Coalition Against Millstone
McGraw-Hill
Michigan Citizens for Water Conservation
Michigan Environmental Council
Michigan Interfaith Climate & Energy Campaign
Michigan Land Trustees
Michigan Representative National Environmental Trust
Miller & Chevalier
Montana State Univ
Nabors
National Environmental Protection Ctr
New England Coalition, Inc.
New Jersey Energy Coalition
New Jersey Environmental Federation
New Jersey Public Interest Research Group (NJPIRG)
North Carolina Waste Awareness & Reduction Network (NC WARN)
Northwest Environmental Advocates
Nuclear Age Peace Foundation
Nuclear Energy Information Service
Nuclear Free Great Lakes Action Campaign
Nuclear Information & Resource Service (NIRS)
Nuclear Policy Research Institute
Ocean County, NJ, Board of Chosen Freeholders
Orange County, NC, Board of Commissioners
Peg Pinard
Pilgrim Watch
Prairie Island Indian Community
Project on Government Oversight
Protect All Children's Environment

Public Commenter
Public Health & Sustainable Energy (PHASE)
Public Policy Advocates, LLC
Riverkeeper, Inc.
Riverkeeper, Inc.
Rockland County Conservation Association, Inc.
Rutgers Environmental Law Clinic
San Luis Obispo Cancer Action Now
San Luis Obispo Chapter of Grandmothers for Peace International
San Luis Obispo County, CA
San Luis Obispo Mothers for Peace
Santa Lucia Chapter
Santa Margarita Area Residents Together
Savannah Riverkeeper
Seacoast Anti-Pollution League
Shems, Dunkiel, Kassel, & Saunders, PLLC
Sierra Club
Sierra Club, Atlantic Chapter
Sierra Club, Miami Group
Sierra Club, New Hampshire Chapter
Sierra Club, New Jersey Chapter
Sierra Club, Ohio Chapter
Sierra Club, Santa Lucia Chapter
Sierra Club, South Carolina Chapter
Sierra Club, Van Buren County Greens
Southern Alliance for Clean Energy
Spiegel & McDiarmid
Surfrider Foundation, Ventura County Chapter
Susquehanna River Basin Commission
Sustainable Energy & Economic Development Coalition
Tap Pilam Coahuiltecan Nation
The New York Times
Three Mile Island Alert, Inc.
Town of Kingston, MA
Town of Oak Island, NC
Town of Plymouth, MA
Town of Yorktown, NY
Turner Environmental Law Clinic
Union of Concerned Scientists
Univ of Connecticut
Univ of Toledo
UNPLUG Salem Campaign
Utility Workers Union of America
Village of Pinecrest, FL
Voices for Earth Justice

Waste Awareness & Reduction Network NC (WARN)
Westchester Citizens Awarenesss Network (WestCAN)
Windham Regional Commission
Yell County Wildlife Federation

APPENDIX B

ADDITIONAL LEXICAL PROFILES FOR INDUSTRY GROUP ANALYSIS

FUEL

For this industry key term, the similarities and differences in the way the two industry groups uses *fuel* in the language represented in the corpus can be seen in Table 5.27.

Table 5.27: Top 20 Collocates with *Fuel* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
spent	21.94497	1494	spent	9.138858	1262
pool	21.31053	839	storage	8.109026	876
storage	20.58771	644	pool	8.072441	747
assemblies	21.37418	400	assemblies	8.491449	742
oil	20.29094	329	oil	8.408904	715
handling	21.03665	240	irradiated	9.027769	559
design	17.40698	235	handling	8.671224	508
core	18.17084	233	movement	8.652185	494
reactor	16.69876	230	assembly	7.675590	470
assembly	20.59858	204	design	5.054443	393
cycle	18.8757	199	cycle	17.38368	312
cooling	18.30863	198	rods	7.230575	311
irradiated	21.59106	185	accident	5.069199	310
cladding	20.76359	173	cladding	7.595219	302
movement	21.36666	172	rod	5.839547	294
rods	20.23239	162	core	4.762164	258
system	29.23056	159	*reactor*	3.559614	221
rod	19.27272	157	recently	8.284105	211
damage	19.04787	155	diesel	6.427927	205
containment	17.10777	141	*containment*	4.023082	205

Most of the collocates, 17 in fact, in this table are shared between both industry groups *(spent, pool, storage, assemblies, oil, handling, design, core, reactor, assembly, cycle, irradiated, cladding, movement, rods, rod, containment)*. When we think back to the Observed Meaning of *fuel* in Chapter 4, we see that the notion of *fuel* being related to the *reactor* is shared between these two groups for our analysis of the highest peak of collocate frequencies *(pool, storage, assemblies, core, reactor, assembly, cladding)*, as well as specific types of *fuel (spent, oil, irradiated, rods, rod)*. Both groups also have a shared understanding that this term relates to the ideas of *movement* and *handling*, indicating that *fuel* is not a static object in this industry. Its use is planned through use of a *design* and is replenished cyclically *(cycle)*. Moreover, both groups at this highest level of frequency share the idea that *fuel* in this industry needs to be stored and contained *(storage, containment)*.

Despite all of this agreement, there are three remaining collocates for each group at this level that are not shared between them (Table 5.28).

Table 5.28: Collocates with *Fuel* by Industry Group (Not Repeated in Table 5.27)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
cooling	18.30863	198	accident	5.069199	310
system	29.23056	159	recently	8.284105	211
damage	19.04787	155	diesel	6.427927	205

For those documents authored by representatives of the NRC, we see that *fuel* is imbued with the idea that it is part of something bigger on the plant level through its collocation with *system: THE FUEL SYSTEM* (16), *SPENT FUEL POOL COOLING SYSTEM FUEL OIL SYSTEM* (14). This term is also frequently paired with *cooling* by NRC authors: *SPENT FUEL POOL COOLING CAPABILITIES* (42), *COOLING CONTAINMENT AND SPENT FUEL* (44), *SPENT FUEL COOLING AND CONTAINMENT INTEGRITY* (20). When we combine these clusters with the frequent collocation of *fuel* with *damage*, we become aware that much focus on this term is related to avoiding impairing its usefulness: *OPERATIONS TO MITIGATE FUEL DAMAGE* (65), *RESULT IN FUEL DAMAGE* (8).

The perspective of the licensees with regard to the three words that collocate with *fuel* from Table 5.28 is slightly different. First of all, rather than most often relating *fuel* to *damage,* like the NRC-authored documents, we find an increased focus on the result of such damage: an *accident* like *FUEL HANDLING ACCIDENT* (277), *POSTULATED FUEL HANDLING ACCIDENT* (38), *FUEL HANDLING ACCIDENT INSIDE CONTAINMENT* (38), *FUEL HANDLING ACCIDENT ANALYSIS* (17). We also find that the licensee group frequently refers to more specific types of *fuel (recently, diesel)*: *RECENTLY IRRADIATED FUEL* (192), *NON RECENTLY IRRADIATED FUEL* (27), *DIESEL FUEL OIL* (86), *DIESEL GENERATOR FUEL* (35).

This lexical profile results in the following Observed Meanings for the term *fuel* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, *fuel* is something needed in the operation of nuclear reactors that is part of a system, requires cooling, and can be damaged.**

2. **For the authors of documents for licensees, *fuel* is a resource needed for the operation of different machines a plant that generates electricity, not just nuclear reactors but also things like diesel generators, and it must be replenished regularly. Mishandling of *fuel* can lead to accidents.**

LICENSE

When we look at the 20 most frequent collocates with *license* for both of these groups we are able to see that there are similarities and differences between their use for this term as well (Table 5.29).

Table 5.29: Top 20 Collocates with *License* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
renewal	8.694685	2141	amendment	20.38938	1057
operating	6.849351	1444	renewal	21.96552	1031
amendment	6.887042	1373	*operating*	30.42231	644
facility	6.799520	861	request	19.27387	525
renewed	8.746032	703	application	19.38704	311
condition	5.875205	545	proposed	17.24642	269
application	6.022089	471	facility	19.13350	226
scope	5.800975	457	scope	19.47949	221
incorporated	7.554680	342	condition	17.48140	219
hereby	7.608781	320	within	16.58593	143
proposed	4.422027	312	*unit*	15.56285	132
effective	7.050679	298	amendments	20.37417	125
conditions	4.991585	280	*nuclear*	16.21254	110
amended	6.723724	279	LAR	19.65646	108
nuclear	3.839571	275	*NRC*	27.67636	96
attachment	6.169483	272	renewed	21.22813	94
within	4.764961	262	conditions	16.13870	93
request	5.208614	256	*technical*	15.96638	92
changes	4.778862	254	NPF	20.23287	91
accordingly	7.341674	253	station	16.46509	91

There are 13 collocates in this table that are shared by both groups *(renewal, operating, amendment, facility, renewed, condition, application, scope, proposed, conditions, nuclear, within, request).* Two semantic sets can be formed from this group of words. The first set relates to the shared meaning of these two groups of *license amendment* and *renewal (renewal, renewed, amendment, application, proposed, request).* The second set concerns what kind of credentials are most frequently talked about by both groups, and the extent of them *(operating, facility, nuclear, condition, scope, conditions, within).*

In addition to all of these shared meanings derived from the 20 most frequently co-occurring words with *license,* we see that there are seven words from each of the two industry groups in Table 5.29 that are not shared (Table 5.30).

Table 5.30: Collocates with *License* by Industry Group (Not Repeated in Table 5.29)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
incorporated	7.554680	342	***unit***	15.56285	132
hereby	7.608781	320	amendments	20.37417	125
effective	7.050679	298	LAR	19.65646	108
amended	6.723724	279	*NRC*	27.67636	96
attachment	6.169483	272	*technical*	15.96638	92
changes	4.778862	254	NPF	20.23287	91
accordingly	7.341674	253	station	16.46509	91

We see that much of the formulaic language noted in Chapter 4 involving *license* is most frequently used in the NRC-authored documents *(incorporated, hereby, effective, accordingly): AMENDMENTS ARE HEREBY INCORPORATED IN THE LICENSE*

(167), *ARE HEREBY INCORPORATED INTO THE LICENSE* (105), *AMENDMENT IS EFFECTIVE AS OF* (250), *LICENSE IS EFFECTIVE AS OF* (39), *RENEWED LICENSE IS EFFECTIVE AS OF* (13), *ACCORDINGLY THE LICENSE IS* (221), *ACCORDINGLY THE LICENSE IS AMENDED* (190). We also can see that this group more frequently refers to a *license* going through the amendment process in the past tense through more than one change*: LICENSE IS AMENDED BY CHANGES* (192). Moreover, the frequent collocation of *license* with *attachment* indicates to us that quite often the authors of NRC documents are referring to supporting documents involved with *license amendment* communications: *IN THE ATTACHMENT TO THIS LICENSE AMENDMENT* (243).

From Table 5.30, we find that the words most frequently used with *license* provide a different meaning for the licensees. First of all, there are more frequent references to those parties involved with the *license (unit, station, NRC)*. We see from the clusters of *NRC* with *license* that they possess the role of administrator over this credential: *THE NRC ISSUED LICENSE* (5), *NRC LICENSE RENEWAL*(7), *NRC APPROVAL OF LICENSE* (7). Meanwhile, a *license* is assigned to a specific *unit* or *station: STATION UNIT NO.# LICENSE* (19), *UNIT NO.# OPERATING LICENSE* (31), *UNIT # LICENSE* (27), *STATION LICENSE RENEWAL* (11), *GENERATING STATION LICENSE* (10).

We also learn that for the licensees a license is a dynamic entity that moves forward into the future. *Amendments* is a clue from the licensees that not only is change is inherent in *license*, but that it happens more than once: *THE LICENSE APPLICATION AND AMENDMENTS* (6), *PROPOSED LICENSE AMENDMENTS* (157), *PREVIOUS LICENSE AMENDMENTS* (7). We learn from these most frequently occurring collocates the specific acronyms used to describe requests for license amendments *(LAR)* and licenses themselves *(NPF)*.

When we see the presence of *technical* in Table 5.30, we get the first connotation of change occurring—since we know from Chapter 4 that *technical* most often means *technical specifications*, which are documents dictating standards that must be followed on the plant level but can be changed with NRC approval. It appears that these *technical specifications* also have a close relationship with the *license: OPERATING LICENSE AND TECHNICAL SPECIFICATIONS* (14), *LICENSE CONDITIONS AND TECHNICAL SPECIFICATIONS* (6), *TECHNICAL SPECIFICATIONS FOR LICENSE* (5).

This lexical profile results in the following Observed Meanings for the term *license* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, a *license* is and endorsement for operating a facility that uses nuclear materials. The NRC holds the power to issue, renew, and approve *licenses*, which often involve supporting documentation. They often use formulaic language involving this term.**
2. **For the authors of documents for licensees, a *license* is an endorsement for operating a power generating station using nuclear fuel administrated by the NRC. It is closely connected to Technical Specifications, and it can undergo revision.**

CONTAINMENT

For this industry key term, the similarities and differences in the way the two industry groups use it in the language represented in the corpus can be seen in Table 5.31.

Table 5.31: Top 20 Collocates with *Containment* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
primary	8.179400	358	isolation	7.523578	640
spray	8.565068	335	primary	7.507436	625
isolation	7.849668	289	*pressure*	18.62212	624
system	17.99612	261	*system*	4.582970	516
pressure	5.783029	256	spray	7.693799	431
integrity	7.271960	239	sump	7.548851	362
leakage	17.51511	187	secondary	7.492630	349
reactor	4.252536	163	inside	7.776442	334
cooling	6.069660	162	leakage	5.474940	301
sump	8.153152	160	atmosphere	8.397002	281
secondary	7.659725	151	valves	5.565005	281
systems	4.843487	145	outside	7.055847	278
inside	8.253776	142	integrity	6.832290	264
fuel	17.10777	141	air	5.746578	255
outside	7.422102	138	rate	5.524790	225
valves	5.900669	133	purge	8.479193	218
unit	4.149584	132	building	6.055590	208
core	5.299704	123	*fuel*	4.023082	205
building	6.587588	118	temperature	4.699129	203
heat	5.487857	112	program	3.950756	195

A majority of the collocates listed in this table are shared by both groups—14 to be exact *(primary, spray, isolation, system, pressure, integrity, leakage, sump, secondary, inside, fuel, outside, valves, building)*. Two semantic sets can be formed from these words to help us understand the shared meaning of *containment* for both groups at this frequency level.

The first one concerns physical components and entities monitored on the plant level *(pressure, spray, system, sump, fuel, valves, building)*. The second group relates *containment* to words having to do with a degree of separation *(isolation, integrity, leakage, inside, outside)*, as well as the extent of that separation *(primary, secondary)*. As a result of the presence of these words in Table 5.31, we are able to see that the resulting notions of *containment* formed from them are not only shared by both groups, but they are shared on the most salient levels of frequency.

It can also be noted in Table 5.31 that seven of the 20 most frequently co-occurring words with *containment* are not mutual to both groups (Table 5.32).

Table 5.32: Collocates with *Containment* by Industry Group (Not Repeated in Table 5.31)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	4.252536	163	atmosphere	8.397002	281
cooling	6.069660	162	air	5.746578	255
systems	4.843487	145	rate	5.524790	225
unit	4.149584	132	purge	8.479193	218
core	5.299704	123	temperature	4.699129	203
heat	5.487857	112	program	3.950756	195

For the documents authored by NRC representatives, *containment* is frequently used with words concerning the *reactor (reactor, unit, core)*. From this association we find that *containment* is specific to each *reactor* or *unit*: *THE REACTOR CONTAINMENT* (31), *THE UNIT # CONTAINMENT* (31). *Containment* also frequently co-occurs with *heat* and *cooling*: *CONTAINMENT HEAT REMOVAL* (46), *CONTAINMENT HEAT*

REMOVAL PUMPS (9), *CORE COOLING CONTAINMENT* (47), *MAINTAIN OR RESTORE CORE COOLING CONTAINMENT* (44), *COOLING AND CONTAINMENT INTEGRITY* (22). From these clusters we see that when *containment* co-occurs with *heat* that there is a negative connotation for this word, while *cooling* has a positive association with the term. In addition to *containment* being a discrete element in relation to each *reactor* or *unit* at a plant that needs to be kept cool, it appears also that it is frequently discussed in NRC-authored documents as being part of a *system* in order to accomplish these tasks: *COOLING DECAY HEAT REMOVAL AND CONTAINMENT SPRAY SYSTEMS* (29), *CONTAINMENT HEAT REMOVAL SYSTEM* (11). As the NRC-affiliated authors are regularly writing about these aspects of *containment,* we find that on the same level of the 20 most frequently use collocates with this term for licensee-authored documents that there is a more local focus to the Observed Meaning.

While the NRC-authored documents had more frequent references to *containment* in relation to the *reactor* and the removal of *heat* through *containment systems*—broad concerns that can be addressed with many licensees—we find that the licensee group frequently uses collocates with *containment* that have a more local application. In Table 5.32, The licensee group frequently relates *containment* with *purge*. Through this association we discover that *containment* is composed of elements that require removal, which is similar to the use of *heat* by the NRC except in this case it is hydrogen that might be radioactive: *THE CONTAINMENT PURGE* (79), *CONTAINMENT PURGE HYDROGEN PURGE* (21), *CONTAINMENT PURGE EXHAUST* (46), *CONTAINMENT PURGE EXHAUST RADIATION* (19), *CONTAINMENT PURGE EXHAUST FILTERS* (16). The licensee group also frequently uses *air*, *atmosphere,* and *temperature* with *containment,* further demonstrating a regular use of this term as having its own components that require monitoring and measurement: *THE CONTAINMENT ATMOSPHERE* (121), *THE CONTAINMENT AIR* (46), *PRIMARY CONTAINMENT ATMOSPHERE* (30), *CONTAINMENT ATMOSPHERE RADIOACTIVITY* (21), *CONTAINMENT ATMOSPHERE CONTROL* (20), *CONTAINMENT AIR WEIGHT* (17), *CONTAINMENT AIR TEMPERATURE* (16).

The presence of *rate* as a frequent collocate further establishes the recurring relationship of *containment* with measurement, and most often as part of a recurring series of measurements *(program)*: *CONTAINMENT LEAKAGE RATE* (121), *CONTAINMENT LEAKAGE RATE TESTING* (104), *CONTAINMENT LEAKAGE RATE TESTING PROGRAM* (80). It should be noted that there are also clusters involving *containment* and *program* that concern the monitoring of other aspects of *containment:* *CONTAINMENT TENDON SURVEILLANCE PROGRAM* (8), *CONTAINMENT INSERVICE INSPECTION PROGRAM* (7). As a result of these relationships on the level of the 20 most frequently used collocates with this term, we see an Observed Meaning for the licensee group that is centered on the treatment of *containment* during plant operations.

This lexical profile results in the following Observed Meanings for the term *license* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, *containment* is part of the reactor's system that requires cooling and heat removal.**

2. **For the authors of documents for licensees, *containment* has its own components that require monitoring and measurement on a regular basis. Some of these components and elements are often removed, or purged, and this action also requires monitoring and measurement.**

OPERATING

When we look at the 20 most frequent collocates with *operating* for both of these groups we are able to see that with this term too there are similarities and differences between the way it is used (Table 5.33).

Table 5.33: Top 20 Collocates with *Operating* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
license	6.849351	1444	experience	20.34031	773
facility	7.323586	939	*license*	30.42679	646
experience	8.636154	653	limits	17.99860	540
renewed	8.475508	442	core	17.50228	433
nuclear	4.876705	428	cycle	29.81591	423
licenses	8.400818	404	report	17.11739	413
cycle	7.207211	375	*plant*	29.64981	377
plant	4.230508	306	conditions	16.93939	324
procedures	5.610985	298	*unit*	15.78525	308
power	4.161676	289	facility	18.55172	302
test	17.22381	237	*nuclear*	16.59608	287
tests	6.618439	225	*power*	15.54191	277
conditions	5.075038	225	normal	17.98425	274
renewal	5.723059	207	procedures	17.79685	250
limits	6.142529	204	licenses	20.66850	229
normal	6.619323	202	based	16.38275	209
core	5.316496	193	*system*	14.91138	192
system	16.88196	187	*pressure*	28.61496	184
company	6.074809	184	temperature	16.20632	171
written	6.339011	183	company	17.84347	168

There are 15 collocates in Table 5.33 that are shared by both industry groups *(license, facility, experience, nuclear, licenses, cycle, plant, procedures, power, conditions, limits, normal, core, system, company)*. From these words we can deduce that both groups most frequently share several meanings that were observed in Chapter 4 for this term. First of all, we notice they both frequently contribute to the use of *operating* in relation to a credential *(license, licenses)* needed by licensees to operate facilities and equipment *(facility, plant, system, core)* in the generation of electricity using *nuclear power*. Both NRC and licensee authors also frequently relate *operating* to established *conditions, limits*, and ways of doing things throughout the industry in order to achieve a certain standard *(normal)*. They also both have a shared understanding that *operating* often relates to events and happenings that may occur regularly *(experience, cycle)*. Finally, there is consensus at this high level of frequency for this term to be used as a branding tool by some companies in this industry to identify themselves by what they do *(company)*.

Despite the many words both groups do share in Table 5.34, there are five collocates that are not mutual for them (Table 5.34). When we look at those words for the NRC group in this table, we find that *operating* is directly related to the supervisory and assessment role of this government agency in the nuclear power industry. As *operating* regularly co-occurs with *identified* in the documents authored by NRC representatives, we find that this term is both something that is overseen: *RECENTLY IDENTIFIED OPERATING EXPERIENCE* (7), *OPERATING EXPERIENCE*

Table 5.34: Collocates with *Operating* by Industry Group (Not Repeated in Table 5.33)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
renewed	8.475508	442	report	17.11739	413
test	17.22381	237	*unit*	15.78525	308
tests	6.618439	225	based	16.38275	209
renewal	5.723059	207	*pressure*	28.61496	184
written	6.339011	183	temperature	16.20632	171

IDENTIFIED (6). This supervisory role of the NRC also becomes noticeable through the frequent use of *renewed* and *renewal* with operating.

Although the co-occurrence of *operating* with *license(s)* was shared by both group in their top 20 collocates, we find that the *renewal* of these documents is most frequently discussed in NRC-authored documents: *RENEWED FACILITY OPERATING LICENSE* (225), *RENEWED OPERATING LICENSE* (173), *RENEWAL OF THE OPERATING LICENSES* (36), *RENEWAL OF OPERATING LICENSES* (34), *RENEWAL OF THE OPERATING LICENSE* (23). The position of authority the NRC holds with regard to *operating* is also present through the collocates *test, testing,* and *written*: *ANNUAL OPERATING TEST* (19), *REQUALIFICATION WRITTEN EXAMINATIONS* (12), *OPERATING TEST RESULTS* (9), *OPERATING TESTS REQUIRED* (6), *OPERATING TESTS WERE ADMINISTERED* (5). As a result of these relationships with *operating,* we can deduce that the use of this term by the NRC group most frequently relates to having passed a test, undergone evaluation, and thereby receiving the commission's approval.

Through the contextual information provided in Table 3.34, we can see that *operating* frequently reflects a different set of roles for the licensee group. First of all, we find that this term is directly related to another industry term, *unit*. Moreover, it is a quality of a *unit: THE OPERATING UNIT* (31), *UNIT # OPERATING* (48). We can also see that *operating* frequently co-occurs with *report,* indicating that the events of operating a nuclear power plant are documented and shared with others, which fits with the earlier observation of the *licensee* serving the role of information provider in the communication circuit with the *NRC*: *CORE OPERATING LIMITS REPORT* (216), *ENVIRONMENTAL OPERATING REPORT* (96), *MONTHLY OPERATING REPORT* (23). This role is also echoed in their use of the collocate *based,* as they provide support and information to regulators as a rationale for a *license* to be allowed continued operation*: OPERATING FREQUENCY IS BASED* (15), *ADEQUATE BASED ON OPERATING EXPERIENCE* (9)*, ARE REASONABLE BASED ON OPERATING EXPERIENCE* (49).

Operating is also frequently used by licensee-affiliated authors in demonstrating their role and perspective in the industry in the collocates *pressure* and *temperature.* However, the relationship of the term with these words is not indicative of the licensee group's role in the communication circuit, but rather their local focus on the plant through measurement: *NORMAL OPERATING TEMPERATURE* (34), *NORMAL OPERATING PRESSURE* (28), *NOMINAL OPERATING PRESSURE* (23), *OPERATING TEMPERATURE AND PRESSURE* (21), *MAXIMUM OPERATING TEMPERATURE* (18), *AT NORMAL OPERATING TEMPERATURE* (17), *AT NORMAL OPERATING PRESSURE* (15). As a result of these collocates, we can see that the role licensees play in this industry is also present in the Observed Meaning of *operating*, albeit a different one than that detected for the NRC-authored documents.

This lexical profile results in the following Observed Meanings for the term *license* for the NRC and licensees as separate industry groups:

1. **For the authors of documents for the NRC, *operating* is a characteristic of licenses and licensees that have successfully undergone evaluation and examination, thereby receiving the commission's approval to continue on into the future.**
2. **For licensees, *operating* is the action of running the plant, which requires measurement and documentation. These daily happenings are used as a basis for various aspects of decision-making.**

EVALUATION

For *evaluation*, the similarities and differences in the way the two industry groups use this term in the language represented in the corpus can be seen in Table 5.35. There are eight collocates in this table that are shared by both industry groups *(safety, NRC, technical report, performed, based, proposed, engineering)*. From these collocates we see that the notion of an evaluation having different types *(engineering, safety, technical)* is shared by both groups at this level of analysis, as well as *evaluations* being *performed* and *proposed*. Furthermore, both groups frequently relate *evaluation* with creating documents *(report)*, as well as the notion it needs to have evidence for its conclusions *(based)*. Finally, it is clear that both groups frequently refer to the *NRC* as being the entity in this industry that most often performs an *evaluation.*

Table 5.35: Top 20 Collocates with *Evaluation* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
safety	18.41062	1005	*safety*	16.76142	804
NRC	4.002124	477	model	7.096483	362
staff	4.343164	362	report	5.206250	289
technical	5.472332	342	*NRC*	16.68401	254
report	5.289470	323	performed	4.936613	251
cause	6.367016	289	engineering	6.882010	207
licensee's	5.307502	288	results	4.558252	153
staff's	7.037146	268	based	4.527248	152
regulatory	5.029849	234	ECCS	6.219473	135
root	7.281340	201	summary	6.029833	135
performed	4.955424	197	flaw	6.439488	134
licensee	3.076197	185	proposed	3.697751	121
based	4.706384	171	models	7.522641	116
review	4.010964	162	*technical*	3.892235	115
proposed	3.603944	125	Westinghouse	6.263083	112
identified	3.269659	124	used	3.798562	112
reviewed	3.663124	123	risk	5.298022	110
operability	5.412853	119	determined	4.783002	106
dated	4.117838	119	criteria	4.586205	105
engineering	6.060123	112	change	3.553691	105

Although both groups do share several words in Table 5.35, there are 12 collocates that are not mutual for them (Table 5.36). Those words present for the NRC-authored documents in this table help us to understand that the most frequent use of *evaluation* relates to who is performing the evaluations *(staff, staff's)*, who is getting evaluated *(licensee, licensees)*, the types of *evaluation* most often discussed by this industry group *(cause, root, regulatory, identified, operability)*, as well as the importance of when *evaluations* are performed *(dated)*. We learn from the analysis of the clusters formed from the NRC's collocates with *evaluation* that *NRC staff* performs evaluations and evaluative

Table 5.36: Collocates with *Evaluation* by Industry Group (Not Repeated in Table 5.35)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
staff	4.343164	362	model	7.096483	362
cause	6.367016	289	results	4.558252	153
licensee's	5.307502	288	ECCS	6.219473	135
staff's	7.037146	268	summary	6.029833	135
regulatory	5.029849	234	flaw	6.439488	134
root	7.281340	201	models	7.522641	116
licensee	3.076197	185	used	3.798562	112
review	4.010964	162	Westinghouse	6.263083	112
identified	3.269659	124	risk	5.298022	110
reviewed	3.663124	123	determined	4.783002	106
operability	5.412853	119	criteria	4.586205	105
dated	4.117838	119	change	3.553691	105

actions to regulate the licensees and what they do: *EVALUATION OF LICENSEE IDENTIFIED CHANGES* (21), *EVALUATION OF OPERABILITY REPORTABILITY ISSUES* (16), *REGULATORY EVALUATION IN* (18). We also find out in Table 5.36 that *evaluation* is synonymous with *review: REVIEW AND EVALUATION* (53), *EVALUATION BASED ON ITS REVIEW* (37). Not only are reviews part of the *evaluation* process, when an *evaluation* is performed by a licensee it is often *reviewed* by NRC *staff: REVIEWED THE LICENSEE'S EVALUATION* (30), *NRC STAFF REVIEWED THE LICENSEE'S EVALUATION* (17). The NRC is not the only entity that they associate with performing evaluations of the licensees, *STAFF HAS REVIEWED THE EVALUATION* (18), and the licensee perspective confirms this.

The licensee-authored documents frequently refer to different types of *evaluation (ECCS, flaw, risk, change)*: *EVALUATION OF PROPOSED CHANGES* (14), *THE*

FLAW EVALUATION (28), *THE RISK EVALUATION* (21), *ECCS PERFORMANCE EVALUATION* (24). They also relate *evaluations* to the fact that this term requires resources *(used, criteria): USED IN THE EVALUATION* (48), *CRITERIA FOR PREPARATION AND EVALUATION* (13). The parts of an *evaluation,* however, that are most frequently referred to are the *results* and *summary: SUMMARY OF THE EVALUATION* (27), *SUMMARY OF AGING MANAGEMENT EVALUATION* (25), *SUMMARY OF EVALUATION* (22), *THE RESULTS OF THE EVALUATION* (35), *THE RESULTS OF THIS EVALUATION* (20).

From the different types of *evaluations, models* are often constructed and discussed: *ECCS EVALUATION MODEL* (63), *LOCA EVALUATION MODEL* (45), *THE MODEL SAFETY EVALUATION* (28), *ECCS PERFORMANCE EVALUATION MODELS* (21). Finally, we are made aware of certain vendors who are synonymous with *evaluation* and *evaluation models* for the licensees: i.e. *Criteria* Analytics and *Westinghouse* in *WESTINGHOUSE SMALL BREAK LOCA EVALUATION* (14), *WESTINGHOUSE ECCS (EMERGENCY CORE COOLING SYSTEM) EVALUATION MODEL* (9).

This lexical profile results in the following Observed Meanings for the term *evaluation* for the NRC and licensees as separate industry groups:

1. **For the NRC, *evaluation* is primarily an assessment of safety issues that they perform at the plant level at a specific moment in time, but sometimes it relates to assessments of licensees that are made by other parties.**
2. **For licensees, an *evaluation* is the identification and assessment of issues relating to a variety of topics (i.e. safety, aging management, emergency**

core cooling system, etc.). These analyses are performed by the NRC and other vendors like Westinghouse.

STAFF

One observation that was made in Chapter 4 was that the word *staff* was strongly connected with *NRC.* without regard for social perspective. By comparing the collocates of *staff* found in texts authored by the NRC versus those affiliated with licensees, we are able to see that *staff* is not only used by both industry groups to refer to the *NRC*, but that it is also used to refer to individuals working for entities associated with the licensees (Table 5.37). Including *NRC,* there are 12 words shared by both groups in this table *(NRC, review, reviewed, determined, commission, nuclear, regulatory, information, will, members, applicant, requested).* Both groups frequently communicate the notion that *staff* most often refers to those people working on behalf of the *NRC,* it is composed of *members*, and that these people are most often performing actions *(review, reviewed, determined, requested)* that relate to *information.* Both groups frequently use this word along with *applicant,* indicating a shared understanding of whom *staff* interacts with, as well as the commitment of actions to the future by *staff* through a frequent collocation with *will.*

Although both groups do share many words in Table 5.37, there are eight collocates that are not mutual for them (Table 5.38). The words used in NRC-authored documents that are listed this table help us to understand that the most frequent uses of *staff* by this group is to communicate what *NRC staff* does *(evaluation, concludes, finds)*—*STAFF CONCLUDES THAT* (850), *NRC STAFF CONCLUDES* (455), *THE NRC*

Table 5.37: Top 20 Collocates with *Staff* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
NRC	6.413103	5492	*NRC*	19.67430	789
concludes	8.041992	1017	review	6.149606	119
reviewed	5.428092	905	information	5.008960	97
finds	8.143957	793	requests	7.544472	86
review	4.823609	616	requested	6.636446	85
licensee	3.673698	606	*plant*	17.45981	85
determined	5.446267	551	additional	5.384998	81
proposed	4.613980	545	approved	5.942807	69
licensee's	4.986181	499	*nuclear*	4.434597	66
commission	4.963380	494	regulatory	5.095437	59
nuclear	3.956416	456	will	3.103837	59
therefore	5.690572	455	commission	6.528001	55
regulatory	4.839608	444	applicant	7.895699	53
information	4.214495	441	control	3.531693	52
evaluation	4.490777	401	determined	5.024887	49
will	3.189567	384	EOF	6.966575	47
members	6.792227	364	manager	6.067180	46
applicant	5.565670	363	applicable	4.672362	46
also	4.304375	321	members	7.449157	44
requested	5.401154	300	reviewed	6.324978	44

Table 5.38: Collocates with *Staff* by Industry Group (Not Repeated in Table 5.37)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
concludes	8.041992	1017	requests	7.544472	86
finds	8.143957	793	*plant*	17.45981	85
licensee	3.673698	606	additional	5.384998	81
determined	5.446267	551	approved	5.942807	69
licensee's	4.986181	499	control	3.531693	52
therefore	5.690572	455	EOF	6.966575	47
evaluation	4.490777	401	manager	6.067180	46
also	4.304375	321	applicable	4.672362	46

STAFF FINDS (347), *NRC STAFF EVALUATION* (38)—as well as what they are making evaluations about *(proposed): STAFF CONCLUDES THAT THE PROPOSED* (146). The presence of *licensee* and *licensee's* in Table 5.38 also indicates for which group of people in this industry the *NRC staff* is making these conclusions: *STAFF CONCLUDES THAT THE LICENSEE* (50), *STAFF CONCLUDES THAT THE LICENSEE'S* (47). The frequent collocation of *staff* with *also* helps us to understand that *staff* performs evaluations on multiple aspects: *STAFF ALSO REVIEWED* (44), *STAFF ALSO CONCLUDES* (40). Finally, we learn from the frequent co-occurrence of *staff* with both *therefore* and *concludes* that this term is directly related to consequences, and more specifically the fact that conclusions are the consequence of *staff's* evaluative role in this industry: *THEREFORE THE NRC STAFF CONCLUDES* (46), *THEREFORE THE STAFF CONCLUDES* (91), *THEREFORE THE STAFF FINDS* (43), *THEREFORE THE NRC STAFF FINDS* (51). The NRC perspective of *staff* having a supervisory role in this industry is also present in the most frequent collocates for the licensee-authored documents.

The characteristic of *staff* having a supervisory role in the nuclear industry is a shared perspective of the licensees, as exemplified in some of the words in for this group in Table 5.38. First of all, we find that licensee-authored texts frequently relate *staff* with *requests, additional,* and *approved* to describe actions taken by the *NRC staff: NRC STAFF REQUESTS FOR ADDITIONAL INFORMATION* (6), *NRC STAFF REQUESTS THE LICENSEE* (7), *APPROVED BY THE NRC STAFF* (98), *NRC STAFF APPROVED CODES AND METHODS* (11). However, we can see that the licensees often evaluate the quality of *NRC staff's* supervisory actions, and even its *members*, as being *applicable* or not*: BY THE NRC STAFF ARE APPLICABLE* (32), *APPLICABLE NRC STAFF MEMBERS* (13), *ANALYZED WITH APPLICABLE NRC STAFF* (12). It should be mentioned, however, that the licensee-authored documents did frequently collocate *staff* with other words in order to talk about individuals who work for them rather than the NRC. When the licensee texts employ *staff* with the words *plant, control, EOF*, or *manager* most likely the industry term will be modified by one of these words in the L1 position, or one slot to the left of it: *PLANT OPERATING STAFF* (9), *THE PLANT STAFF* (25), *PLANT OPERATIONS STAFF* (5), *THE EOF (EMERGENCY OPERATIONS FACILITY) STAFF* (7), *CONTROL ROOM STAFF* (28), *THE STAFF MANAGER* (14).

This lexical profile results in the following Observed Meanings for the term *staff* for the NRC and licensees as separate industry groups:

1. **For the NRC, *staff* typically refers to their own employees, unless explicitly specified otherwise. These individuals have the authority to evaluate licensees on behalf of the NRC, and their conclusions are the consequence of these evaluations.**
2. **For licensees, *staff* typically refers to those individuals working for the NRC—who may or may not be applicable to what they are discussing—unless explicitly specified otherwise as having plant affiliations. *Staff* typically request information from licensees and provide approval.**

TECHNICAL

Although *technical* was not found to have a statistically significant difference in its frequency of use between the NRC and licensees, there are patterns in the words that co-occur with this term for each industry group. There are nine words that most frequently co-occur with *technical* that are present for both industry groups in Table 5.39 *(specification, specifications, accordance, requirements, basis, support, required, proposed, changes)*.

Table 5.39: Top 20 Collocates with *Technical* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
specifications	9.632138	2858	specification	8.957869	2568
specification	8.947187	1017	specifications	9.273857	2318
changes	6.189480	446	requirements	5.311520	471
accordance	5.772685	361	change	5.142155	369
environmental	6.256611	356	proposed	4.977586	350
evaluation	5.472332	342	changes	5.179689	333
contained	7.297473	329	bases	17.52496	271
appendix	6.099467	320	required	4.237989	270
requirements	17.61666	271	standard	6.886421	238
review	4.588162	226	***unit***	3.715367	230
indicated	6.723345	216	support	5.585738	220
protection	5.219781	211	section	4.061544	206
basis	5.189609	190	accordance	4.643179	174
follows	7.820341	185	force	8.009854	170
information	4.124836	179	task	8.215493	167
read	8.534512	169	analysis	3.952113	167
support	5.742698	160	***plant***	16.80037	164
required	3.940412	141	basis	4.738995	164
proposed	3.811844	135	manual	5.989061	162
provide	4.180814	128	requirement	5.670664	160

The presence of these words indicates that the notion of *technical* being directly linked to requirements and conformity *(specification, specifications, requirements, accordance)*, as well as an informational or evidentiary foundation *(basis, support)*, is shared by both groups at this high level of frequency. Furthermore, both industry groups frequently connect this term to *changes* and the act of proposition *(proposed).*

It can also be noted in Table 5.39 that 11 of the 20 most frequently co-occurring words with *technical* are not shared by both groups (Table 5.40).

Table 5.40: Collocates with *Technical* by Industry Group (Not Repeated in Table 5.39)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
environmental	6.256611	356	change	5.142155	369
evaluation	5.472332	342	bases	17.52496	271
contained	7.297473	329	standard	6.886421	238
appendix	6.099467	320	***unit***	3.715367	230
review	4.588162	226	section	4.061544	206
indicated	6.723345	216	force	8.009854	170
protection	5.219781	211	task	8.215493	167
follows	7.820341	185	analysis	3.952113	167
information	4.124836	179	***plant***	16.80037	164
read	8.534512	169	manual	5.989061	162
provide	4.180814	128	requirement	5.670664	160

First of all, we can see that the NRC directly relates *technical* to *protection*, and even more frequently *environmental protection*: *TECHNICAL SPECIFICATIONS AND ENVIRONMENTAL PROTECTION PLAN* (194) *FIRE PROTECTION PLAN AND TECHNICAL SPECIFICATIONS* (7). However, despite the connotation of *environmental* and *protection* with *technical*, this term is attributed to something that is not originally produced by people associated with the NRC. Instead, it is *provided* to the NRC in the form of documents and *information*: *TECHNICAL SPECIFICATIONS CONTAINED IN APPENDIX* (240), *TECHNICAL SPECIFICATIONS INDICATED IN THE* (214), *TECHNICAL INFORMATION IN SUFFICIENT DETAIL* (36). There is clearly a strong relationship between language and *technical* through its frequent collocation with both *read* and *follows* to create the cluster *TO READ AS FOLLOWS* (204). Even more importantly, it is the authors of documents for the NRC who possess the ability to make the decision regarding exactly how something *technical* is to *read*. We also learn from the perspective of the NRC-authored documents that it is the NRC staff who is responsible for performing evaluations of the *technical: THE NRC STAFF'S TECHNICAL EVALUATION* (40), *DETAILED TECHNICAL REVIEW* (109), *NRC STAFF TO COMPLETE ITS DETAILED TECHNICAL REVIEW* (41).

The notion that *technical* is associated with oversight is also present in the collocations from the licensee model. However, rather than the NRC or one singular authority, collaborative entities and resources are connoted with this term: *TECHNICAL SPECIFICATION TASK FORCE (TSTF)* (103), *THE TECHNICAL SUPPORT CENTER* (64). The rest of the collocates in Table 5.40 indicate different relationships for the underlying meaning of *technical* than those we observed for the NRC-authored documents. First, there is the notion that *technical* is explicitly related to a document, however this document is not most often a supplement to be found in an appendix (like for the NRC-authored documents). Instead something *technical* is often found in a *manual* and/or a document divided into *sections* that are specific to each *unit* at each *plant: UNIT # TECHNICAL SPECIFICATIONS* (128), *THE PLANT'S TECHNICAL*

SPECIFICATIONS (54), *TECHNICAL REQUIREMENTS MANUAL* (134), *TECHNICAL SPECIFICATION SECTION* (46). These informational documents are often analyzed and changed, and even the *bases* upon which they are designed are even controlled *TECHNICAL SPECIFICATION CHANGE* (118), *CHANGE TO TECHNICAL SPECIFICATION* (22), *BASES CONTROL PROGRAM* (55), *STANDARD TECHNICAL SPECIFICATION CHANGE* (38), *TECHNICAL AND REGULATORY ANALYSIS* (18), *TECHNICAL ANALYSIS AND REGULATORY* (17).

We can deduce from these two observations that although the NRC perceives *technical* as a form of compliance to their prescribed standards for all plants that needs to be evaluated, while the licensees view *technical* as local standards. This lexical profile results in the following Observed Meanings for the term *evaluation* for the NRC and licensees as separate industry groups:

1. **For the NRC, *technical* most often forms the cluster *technical specifications*, a document of accepted standards that are equated with other guidelines for protection that are assessed for compliance. These documents can be changed, but only with adequate evidence and justification.**
2. **For licensees, *technical* most often forms the cluster *technical specifications*, a document of accepted standards. While these documents are in place across the industry, they represent guidelines that are specific and local to each plant.**

EMERGENCY

There are also similarities and differences in the meaning associated with *emergency* between NRC-authored documents and those written by licensee representatives, which we can see through an analysis of the 20 most frequently co-occurring words with the term (Table 5.41).

Table 5.41: Top 20 Collocates with *Emergency* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
preparedness	9.870913	591	*system*	4.881150	605
diesel	8.618698	566	response	6.135303	600
plan	7.438118	521	cooling	6.697451	461
generator	7.457026	432	site	6.307378	411
system	18.54474	414	general	7.256737	371
cooling	7.104699	360	director	8.518343	365
response	6.099707	352	core	5.549842	360
core	6.483400	303	plan	6.807074	342
action	5.007302	228	diesel	7.430539	332
power	4.080366	191	area	5.798535	306
systems	5.085773	186	*power*	3.999497	306
procedures	5.247853	162	room	17.48215	270
room	17.16413	159	classification	8.435592	265
planning	7.975090	155	control	4.128680	246
control	4.514967	155	generator	5.612100	220
procedure	5.408307	144	personnel	5.456270	218
level	5.010647	144	preparedness	9.411997	217
operating	4.428509	143	operations	5.683989	195
water	4.483716	137	planning	8.239655	189
generators	7.920513	128	procedures	5.676383	185

There are 13 collocates shared by both industry groups in Table 5.41 *(preparedness, plan, procedures, planning, diesel, power, generator, system, core, cooling, response, room, control).* These words can be divided into two major semantic groups. One group concerns *preparedness* and *planning (preparedness, plan, procedures, planning, response),* and the other one has to do with plant-level words *(control, room, diesel, generator, power, core, cooling).* Only one word, *system,* did not fit these two semantic sets.

When we look at those words not shared by both groups in the list of the 20 most frequently occurring collocates with *emergency,* we find that there is variation between them with regard to the meaning associated with this term at such a high level of frequency: there are eight words for each group in Table 5.41 that are not shared (Table 5.42).

Table 5.42: Collocates with *Emergency* by Industry Group (Not Repeated in Table 5.41)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
action	5.007302	228	site	6.307378	411
systems	5.085773	186	general	7.256737	371
level	5.010647	144	director	8.518343	365
procedure	5.408307	144	area	5.798535	306
operating	4.428509	143	classification	8.435592	265
water	4.483716	137	personnel	5.456270	218
generators	7.920513	128	operations	5.683989	195

The first thing that can be observed from Table 5.42 for the NRC-authored documents is that there are three words that are different numbers of word forms that are shared by both groups *(systems, generators, procedure).* This demonstrates the perspective of the NRC having a perspective of the entire industry in that they would be concerned with multiple *systems* and *generators* at different licensee facilities, while also being concerned with an individual *procedure: EMERGENCY CORE COOLING SYSTEMS* (72), *EMERGENCY*

DIESEL GENERATORS (118), *EMERGENCY PLAN IMPLEMENTING PROCEDURE* (31). We also see in Table 5.42 that *emergency* is a construct falling under the authority of authors associated with the NRC when this term is associated with *action*: *THE EMERGENCY ACTION LEVEL* (122), *EMERGENCY ACTION LEVELS SHALL BE* (14), *EMERGENCY CLASSIFICATION AND ACTION LEVEL SCHEME* (33). Finally, we learn of individual aspects of *operating* a nuclear power plant that when they posses the quality of *emergency* NRC authors frequently discuss *(water, operating): EMERGENCY SERVICE WATER* (61), *EMERGENCY SERVICE WATER PUMPS* (13), *EMERGENCY OPERATING PROCEDURE* (31). This more microscopic perspective is also shared by the licensee-authored documents.

From Table 5.42 we find that *emergency* is yet again more localized to each licensee. Firstly, *emergency* is something that requires *classification*, and this action is performed at the local (plant) level using appropriate evidence for rationale: *THE EMERGENCY CLASSIFICATION* (33), *DECLARATION OF AN EMERGENCY CLASSIFICATION* (8), *EMERGENCY CLASSIFICATION BASED ON* (19). We also find that this term is a characteristic that can be assigned to a licensee's physical *site* or *area: SITE AREA EMERGENCY* (272), *SITE AREA EMERGENCY OR GENERAL EMERGENCY* (82). Finally, we see that the notion of *emergency* being related to operating a nuclear plant is also shared by licensee-authored documents when it is used to characterize organizational operations for licensees—from facilities to personnel: *EMERGENCY OPERATIONS FACILITY* (109), *THE EMERGENCY DIRECTOR* (180), *THE SITE EMERGENCY DIRECTOR* (18), *EMERGENCY OPERATIONS FACILITY DIRECTOR* (10), *EMERGENCY RESPONSE PERSONNEL* (57),

EMERGENCY RESPONSE ORGANIZATION PERSONNEL (13), *THE EMERGENCY OPERATIONS CENTER* (18).

This lexical profile results in the following Observed Meanings for the term *emergency* for the NRC and licensees as separate industry groups:

1. **For the NRC, *emergency* is generally associated with systems in place to control a situation when the NRC evaluates it to be at a certain level that requires action.**
2. **For licensees, *emergency* is a local concern that has been integrated into their culture and structure such that facilities and personnel are allocated for use specifically in these situations.**

PRESSURE

An analysis of the 20 most frequent collocates with *pressure* for both groups results in similarities differences in the use of this term between the NRC and licensees (Table 5.43). Almost all of the collocates listed in Table 5.43 are shared by both groups—18 to be exact. These words can be organized into semantic sets in order for us to see which notions of this term are mutual between NRC- and licensee-authored documents. The first set concerns plant equipment and resources *(reactor, boiler, vessel, system, RCS, injection, coolant)*. Another group can be formed from *high, low, temperature, test, differential,* and *psig* that has to do with measurement. The third group formed from the collocates shared by both groups in Table 5.43 concerns *control (boundary, containment, control, code, ASME).* This only leave two words that co-occur with *pressure* in the 20 most frequently occurring collocates for each group (Table 5.44).

Table 5.43: Top 20 Collocates with *PRESSURE* by Social Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	6.149885	715	*reactor*	19.04444	1243
vessel	7.899266	529	temperature	19.51128	845
system	18.70691	503	boundary	20.99431	826
high	7.359033	451	vessel	19.98144	786
boundary	8.641746	439	*containment*	18.62212	624
coolant	7.353844	409	*system*	17.55764	601
low	6.561714	376	coolant	19.71759	579
temperature	7.131162	336	high	19.65777	564
injection	8.273845	334	RCS	19.69368	508
code	17.99928	308	low	19.47347	476
containment	5.788653	257	code	30.41332	320
RCS	7.270800	232	steam	17.99373	315
test	17.39070	202	ASME	18.77270	313
boiler	9.528957	198	differential	21.58890	312
ASME	5.963960	191	psig	20.62177	295
retaining	9.606274	175	leakage	18.12528	280
differential	9.341409	167	boiler	21.75436	266
psig	8.600847	157	injection	19.67910	263
components	5.222888	149	test	17.49166	251
control	4.309908	146	control	16.72398	231

Table 5.44: Collocates with *Pressure* by Industry Group (Not Repeated in Table 5.43)

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
retaining	9.606274	175	steam	17.99373	315
components	5.222888	149	leakage	18.12528	280

For the NRC-authored documents, we see in Table 5.44 that *pressure* is a defining quality of *components,* and more specifically how they *retain pressure*: *PRESSURE RETAINING COMPONENTS* (44), *PRESSURE RETAINING WELDS* (28),

PRESSURE RETAINING BOLTING (22), *COMPONENTS AND SYSTEM PRESSURE TESTS* (61). We know from NRC-authored documents for other terms like *technical, safety,* etc., that this group functions in a supervisory and evaluative capacity for the industry. Thus, these frequent clusters make sense as *PRESSURE RETAINING COMPONENTS* are something that could be evaluated. The licensee-authored texts also reflect the focus of this industry group on local issues through the frequent use of *steam* and *leakage* with *pressure: STEAM DOME PRESSURE* (57), *STEAM GENERATOR PRESSURE* (34), *MAIN STEAM PRESSURE* (16), *PRESSURE BOUNDARY LEAKAGE* (131), *LEAKAGE FROM REACTOR PRESSURE* (24). We can see from the frequent clusters that both of these collocates are used to describe specific places where *pressure* is measured on the plant level.

This lexical profile results in the following Observed Meanings for the term *pressure* for the NRC and licensees as separate industry groups:

1. **For the NRC, *pressure* is a force intrinsic to systems that can be measured, and it is a defining quality for some components used on the plant level.**
2. **For licensees, the Observed Meaning of *pressure* is similar to that of the industry level, that it is a "force intrinsic to the machines, systems, and their components, at a plant. It is measured, quantified, and evaluated" in specific locations at the plant in order to monitor issues like *leakage*.**

INSPECTORS

While all of the industry terms to this point have had similarities and differences between the words that co-occur with them, *inspectors* is a term that is used in much different frequencies in NRC-authored documents in comparison to those written by representatives of licensees (Table 5.45).

Table 5.45: Top 20 Collocates with *Inspectors* by Industry Group

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reviewed	7.460934	2785	NRC	18.32586	10
inspection	19.56502	1821			
scope	7.188538	1375			
also	5.974063	768			
identified	4.883788	618			
determined	6.015955	615			
observed	7.597003	580			
verified	7.621719	578			
licensee	3.952814	553			
licensee's	5.420281	507			
evaluated	6.256549	408			
performed	5.218950	385			
resident	7.658813	362			
selected	6.482520	355			
concluded	6.299640	280			
conducted	5.299075	270			
following	4.457254	264			
finding	4.966893	255			
results	4.791665	251			
plant	3.332564	249			

There are a few aspects of Table 5.45 that might not make sense upon first glance. As you might have noticed, the number of collocates for *inspectors* from the licensee-authored texts is far fewer than 20—one to be exact. This is due to the infrequency of *inspectors* in those texts designated as licensee-authored in the reference corpus that was presented in Table 5.2. It should be noted that there were seven other collocates with *inspectors* for the licensees, however their raw frequencies were less than 10, and thus their MI scores were discounted. As a result, the context surrounding *inspectors* for the NRC is exactly the same as that which was presented in Chapter 4 and that of the licensees is much more limited. The reason for this limitation is the infrequent use of *inspectors* (68). Essentially, the only significant patterns we can find using this methodology for licensee-authored documents is that they associate *inspectors* as being representatives of the *NRC*.

APPENDIX C

COLLOCATE LISTS OF TOP 20 WORDS CO-OCCURRING WITH INDUSTRY KEY TERMS FOR EACH NRC REGION

REGION I COLLOCATION LISTS

Table C.1: Top 20 Collocates with *Containment* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
primary	8.716740608	116	primary	8.010806084	221
pressure	6.46002388	105	secondary	8.511730194	187
spray	8.926989555	96	*system*	5.014586926	157
isolation	8.237425804	91	isolation	7.435793877	156
integrity	8.013622284	91	*pressure*	5.757264614	147
system	4.568516254	64	sump	7.869610786	94
accident	6.354671001	60	integrity	7.304549217	91
secondary	8.351171494	58	atmosphere	8.571277618	84
outside	7.706461906	44	*fuel*	4.935608387	82
leakage	6.370458603	44	spray	7.138478279	67
systems	5.191534996	42	inside	7.922352314	65
cooling	5.953671455	40	accident	5.200183868	63
liner	8.453137398	38	air	5.804593086	59
valve	5.810931206	36	building	6.334270477	55
valves	15.04873848	35	outside	7.028240681	54
heat	5.784296036	35	within	4.352497101	52
water	4.37631321	35	monitor	6.628993511	51
program	14.00691891	34	loss	4.815322399	48
credit	7.299559116	34	valve	5.100034237	46
removal	6.735031128	33	leakage	5.161578655	44

Table C.2: Top 20 Collocates with *Emergency* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
diesel	9.490834236	109	general	7.462481976	154
system	5.813617229	104	response	6.053694248	151
generator	8.316775322	93	site	6.586158276	149
preparedness	10.41575336	88	*system*	4.76870966	149
plan	7.703895569	80	area	6.500388622	123
cooling	7.326151371	71	cooling	6.566418648	114
core	6.830480576	51	director	8.386055946	97
planning	9.031157494	50	plan	6.481164455	82
response	5.633639336	42	none	7.738000393	81
radiological	7.311164856	38	planning	8.255501747	80
room	5.945515156	38	personnel	5.98942709	78
water	5.039618969	38	room	5.728340626	78
power	4.355592251	37	preparedness	9.31431675	77
EDG	7.182830811	36	classification	8.208402634	74
plans	6.966617107	34	diesel	7.424439907	74
procedures	5.840235233	32	core	5.368046761	70
feedwater	7.475798607	31	exposure	6.429999352	66
action	4.923457623	31	control	4.021417141	66
pump	5.900032997	29	generator	6.159351826	63
operating	4.751293182	29	air	5.705721378	62

Table C.3: Top 20 Collocates with *Evaluation* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
safety	6.058279514	263	*safety*	16.51710129	210
NRC	3.998762131	128	report	5.351531506	73
technical	6.202083111	112	*NRC*	4.679269314	58
staff	4.405491352	102	model	7.295895576	57
report	15.90400791	80	flaw	7.334085941	56
staff's	7.043984413	73	engineering	7.280434608	55
regulatory	5.267596722	68	performed	4.775849819	43
dated	4.82196331	51	results	4.897167683	40
cause	5.93454504	47	models	8.368453026	37
SE	7.51649189	43	criteria	5.461562157	37
reviewed	4.589437962	43	BWRVIP	7.266802788	35
review	4.033213615	43	based	4.716605663	35
licensee's	5.498866558	40	ECCS	7.079065323	33
SER	7.463685513	38	risk	5.819343567	33
engineering	6.7639184	38	aging	5.453271389	31
based	4.574787617	37	*inspection*	4.444090843	29
root	7.236305237	36	management	4.980314255	27
performed	4.865411758	34	used	4.023515224	26
licensee	3.335339546	32	reviewed	6.441087246	25
section	3.428575039	27	performance	5.0803895	25

Table C.4: Top 20 Collocates with *Fuel* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
spent	10.01121712	385	spent	9.270540237	383
pool	9.429102898	182	irradiated	9.324151039	270
storage	8.719905853	169	assemblies	9.02889061	234
assemblies	9.162283897	57	pool	8.412519455	217
oil	7.924863815	53	movement	8.97307682	210
reactor	4.75316143	51	storage	7.994202137	205
nuclear	3.897417307	46	recently	9.281571388	155
handling	9.095421791	45	handling	8.821550369	121
cycle	7.136748314	42	accident	5.531592846	90
core	5.971698761	38	cladding	8.19846344	82
design	5.223834991	38	*containment*	4.935608387	82
pools	9.842206001	36	design	5.142400742	81
independent	7.30873394	32	involving	8.765183449	80
enrichment	9.626477242	31	*unit*	3.917832613	79
burnup	9.399066925	31	assembly	7.996587276	73
racks	9.181836128	30	*nuclear*	4.453076363	67
cladding	9.054924011	29	new	4.992484093	58
rods	8.187374115	29	oil	7.692120075	56
stored	9.00429821	28	building	6.124634266	54
assembly	8.313626289	28	cycle	6.08165741	54

Table C.5: Top 20 Collocates with *Inspection* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
scope	7.125285149	332	inservice	9.184839249	194
NRC	4.182201385	278	program	17.37675095	186
public	5.765705109	235	interval	7.702795506	97
inspectors	6.117922783	209	visual	7.889478683	92
reviewed	5.820128441	193	methods	7.609564304	73
team	6.545930862	173	requirements	4.114992619	41
results	6.034515381	167	year	5.832854271	38
available	5.91012764	140	will	3.428647995	38
findings	5.574962616	140	performed	4.593774796	37
report	15.73404121	136	informed	8.137152672	36
inservice	7.767432213	112	scope	6.439398289	35
program	14.34701824	104	risk	5.938969135	35
One	4.900334835	93	*reactor*	3.644383669	35
conducted	5.67257309	91	ISI	6.72048378	34
period	5.670974731	82	outage	5.667667389	33
procedures	5.79242897	78	ASME	5.344768047	33
electronically	7.765222549	76	results	4.609976292	32
performed	5.051900387	74	BWRVIP	7.126452446	31
identified	3.896213055	71	section	3.57625556	31
using	5.352917671	67	penetration	6.561146736	29

Table C.6: Top 20 Collocates with *Inspectors* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reviewed	7.718163013	339			
inspection	6.117922783	209			
scope	7.083523273	152			
identified	5.626232147	111			
determined	6.417948723	91			
also	5.69217062	85			
verified	8.356060982	84			
resident	8.809930801	76			
observed	8.078901291	69			
performed	5.760687351	57			
selected	7.308716774	55			
conducted	5.709560394	44			
evaluated	6.138195515	41			
system	3.590442419	37			
presented	7.16077137	35			
results	4.736161232	32			
actions	4.309863567	32			
addition	5.516199589	31			
finding	5.369692326	30			
concluded	6.082114697	29			

Table C.7: Top 20 Collocates with *License* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
renewal	8.386268616	815	renewal	9.921943665	409
operating	6.584297657	375	amendment	8.219831467	214
amendment	6.537953854	345	*operating*	6.950743198	149
facility	6.360707283	231	request	7.298665524	109
scope	6.401509762	207	application	8.003517151	99
application	6.007538319	202	scope	7.610103607	68
transfer	7.066698551	186	proposed	4.88581419	50
within	5.520384312	146	***unit***	4.024516582	48
proposed	4.713107586	136	***nuclear***	4.671921253	44
renewed	8.335420609	122	facility	6.724758625	42
nuclear	3.855268002	120	appendix	5.709419727	42
subject	5.954687119	109	B	3.915816545	41
transfers	7.577883244	108	condition	4.942638874	40
condition	5.345162868	94	within	4.616323471	40
effective	6.458677769	80	BVPS	5.846834183	38
review	3.951151848	80	*NRC*	4.197833061	35
environmental	4.857711792	78	station	4.576292515	34
applications	7.07308054	73	*power*	3.428554058	30
accordingly	6.894434452	70	amendments	8.264242172	29
amended	6.155047894	65	environmental	5.178477764	29

Table C.8: Top 20 Collocates with *Licensee* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
identified	4.970837116	142	event	7.028291702	50
staff	3.814702272	123	report	6.818640709	50
provided	5.464139462	116	shall	6.467923164	37
stated	6.178565025	98	LER	9.284213066	23
proposed	4.357201099	98	controlled	8.578893661	21
submitted	6.064089775	91	states	7.519691944	17
shall	4.600995541	82	provide	5.842700005	17
information	3.932268858	76	reports	7.799125195	16
requested	5.770783901	74	*NRC*	4.834365845	16
facility	4.775551796	71	violations	10.19702339	14
nuclear	3.109618187	66	stated	7.584369183	14
LLC	5.399273872	64	provided	5.512152672	14
resonse	4.45260334	62	*staff*	6.389668465	12
request	4.859148979	61	protection	6.094833851	12
violations	7.419311523	58	document	6.526836395	10
dated	4.096014977	56	*unit*	3.527297974	10
also	4.026940823	54			
INC	5.91384697	53			
company	5.597373486	53			
may	3.511748552	50			

Table C.9: Top 20 Collocates with *NRC* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
staff	6.205381393	1629	approved	7.815679073	223
regulatory	5.630136967	401	*staff*	7.698168278	199
commission	5.300837517	393	regulatory	6.293271065	127
nuclear	4.279393673	375	review	5.796740532	99
site	6.090112209	340	*nuclear*	4.804998398	95
public	4.792141438	287	approval	7.493283272	88
inspection	4.19254303	280	commission	7.085407734	86
review	4.420720577	258	letter	6.012519836	85
web	7.397921562	218	information	4.73179388	79
staff's	6.384624481	212	request	5.702881813	71
ADAMS	5.463012695	208	submitted	6.469706535	66
will	3.277211666	207	provided	4.868525505	60
reviewed	4.33204031	165	*evaluation*	4.679269314	58
issued	5.534586906	157	reference	4.835539818	57
approved	5.89643383	149	requested	6.352648735	56
information	3.566932678	149	*safety*	13.71662712	50
document	5.389574051	146	dated	5.837043762	50
room	4.716127872	137	consider	7.341957569	47
identified	3.550601721	134	notification	6.467488289	47
evaluation	3.998762131	128	question	6.35123539	47

Table C.10: Top 20 Collocates with *Nuclear* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
power	6.818945408	907	*power*	6.630770683	492
commission	7.075358391	707	station	6.732237816	270
regulatory	7.280705929	662	*plant*	6.011859894	263
plant	5.955388069	469	regulatory	7.300274849	231
Entergy	7.229018688	414	commission	8.577869415	219
NRC	4.267805576	372	LLC	7.979351997	137
station	6.905488968	338	*safety*	15.07469082	116
LLC	7.386641502	337	INC	8.260910034	116
plants	12.34016132	310	plants	8.125562668	115
INC	7.671506882	238	PSEG	7.930658817	114
PSEG	7.298967361	221	Entergy	7.966732025	111
Point	6.459218025	219	*NRC*	4.804998398	95
unit	5.559057713	216	operations	6.389546871	93
reactor	4.846683025	179	Mile	7.73950386	88
office	6.941126823	174	Point	6.436600208	84
Fitzpatrick	7.500726223	159	*unit*	3.928074598	80
generating	7.487644196	157	Fitzpatrick	8.345696449	76
operations	6.36374712	156	company	6.986097336	75
regulation	7.777628422	145	*fuel*	4.453076363	67
Calvert	7.531606197	131	generating	6.936602116	63

Table C.11: Top 20 Collocates with *Operating* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
license	6.584297657	375	experience	8.8244524	151
facility	7.367830276	268	*license*	6.950743198	149
experience	8.7495718	233	*plant*	5.135882378	100
licenses	8.206707001	163	limits	6.254668713	90
nuclear	4.427660942	103	report	5.304100037	74
plant	4.610061169	87	procedures	6.451514244	66
renewal	5.848244667	81	*power*	4.25173521	66
procedures	6.002075195	75	*nuclear*	4.828904629	61
power	4.288613796	74	core	5.619847775	59
industry	6.858514309	58	facility	6.851008892	57
renewed	7.978825569	55	*unit*	3.93258667	56
tests	6.589783192	49	normal	6.298389912	54
limits	6.164477348	45	conditions	5.199512005	54
current	5.669217587	44	licenses	9.042456627	47
report	14.8234024	43	*pressure*	4.409170151	46
reactors	6.906812191	42	cycle	6.047403336	37
agreement	6.864643097	42	temperature	4.679307938	36
Nos	6.778228283	42	pump	4.691401005	35
core	5.413004398	40	current	5.194698811	32
written	5.87518549	39	environmental	5.006174564	32

Table C.12: Top 20 Collocates with *Plant* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	5.955388069	469	*power*	5.50611639	299
power	6.259148121	458	*nuclear*	6.011859894	263
specific	6.720000267	220	specific	6.52164793	168
Calvert	8.000956535	135	conditions	5.823179245	158
Cliffs	8.009594917	134	*safety*	15.09597111	156
performance	5.011136055	96	operation	4.993719578	119
Ginna	7.438868523	89	*operating*	5.135882378	100
operating	4.610061169	87	equipement	5.653656006	99
operation	4.655387402	69	shutdown	5.787236691	92
design	4.683996201	64	systems	5.091020584	75
unit	4.184323788	62	operations	5.552771091	69
safety	3.129398823	62	based	4.639430046	66
systems	4.50032711	59	procedures	5.436090469	62
risk	4.722458839	58	Fitzpatrick	7.598518372	60
Fitzpatrick	6.446690559	57	current	5.127453327	58
operations	5.337192535	57	design	4.221394062	57
procedures	4.895400047	55	personnel	5.032126904	54
areas	5.278490543	53	Ginna	6.482764244	49
LLC	5.088419437	51	changes	3.719421148	46
activities	4.500664234	50	*technical*	3.500713825	45

Table C.13: Top 20 Collocates with *Power* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	6.818945408	907	*nuclear*	6.630770683	492
plant	6.255994797	457	station	6.330918312	330
station	7.10470438	307	*plant*	5.50611639	299
plants	12.43546009	262	thermal	7.511056423	242
unit	5.455052853	159	*reactor*	4.803305149	194
level	5.933122158	148	*unit*	4.492749214	191
uprate	8.477497101	126	loss	5.732168674	167
Calvert	7.801967621	125	offsite	6.61179781	160
Cliffs	7.762439251	120	rated	7.955568314	132
reactor	4.571232319	117	operation	4.780001163	125
operation	5.186825752	106	full	6.540239334	101
reactors	7.481566906	105	plants	7.20395565	98
thermal	6.757947445	104	Valley	8.127334595	88
Ginna	7.38297224	91	Beaver	8.127334595	88
Bottom	6.675055504	87	Millstone	6.13356638	88
Peach	6.786210537	86	supply	6.666061401	81
units	13.47439289	85	*system*	3.066889524	75
Millstone	6.882410526	84	core	4.717118263	73
offsite	7.1390028	80	flow	4.303491592	72
LLC	5.575963974	76	reduced	6.940921307	71

Table C.14: Top 20 Collocates with *Pressure* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	6.044840336	147	*reactor*	6.167061806	312
coolant	7.744416714	128	vessel	7.102749348	206
high	7.715038776	124	temperature	6.628790855	201
system	5.325969219	118	*system*	4.908655643	168
vessel	7.893345356	113	boundary	7.749585152	162
containment	6.473699093	106	coolant	6.922709465	159
low	6.792368412	95	*containment*	5.757264614	147
boundary	8.545444489	85	high	6.912596703	144
injection	8.630146027	83	low	6.619188786	118
code	6.215108395	83	code	5.814823627	92
temperature	7.06953001	77	ASME	6.093197346	86
differential	9.446342468	55	PSIG	7.804403782	83
boiler	9.544905663	53	boiler	8.938076019	79
retaining	9.734144211	47	steam	5.388683796	76
accident	5.846152306	46	RCS	6.177941322	75
ASME	6.255451202	45	leakage	5.66844368	72
PSIG	8.537107468	41	limits	5.253539085	65
RCS	7.472337246	40	core	5.227928638	65
oil	7.209302902	38	injection	7.05505991	62
engineers	8.037934303	37	testing	4.870445728	61

Table C.15: Top 20 Collocates with *Reactor* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
coolant	8.452812195	358	coolant	8.274513245	563
vessel	8.216609001	242	vessel	8.060287476	555
system	5.429457188	217	*system*	5.687889576	400
nuclear	4.854720592	180	*pressure*	6.167061806	312
office	7.490466595	156	water	5.365097046	225
water	5.594347477	152	*power*	4.803305149	194
pressure	6.044840336	147	trip	6.68846941	172
regulation	8.474524498	144	level	5.653705597	137
core	6.66697979	124	building	6.899345875	130
building	7.495810509	119	temperature	5.177859783	102
power	4.571232319	117	head	5.784578323	101
oversight	7.716282368	105	core	5.303111553	95
process	5.900450706	95	protection	5.800824165	83
safety	4.025981426	95	boundary	6.22283268	78
trip	7.110823631	85	shutdown	5.334699154	71
operator	6.312867641	80	hours	4.839934349	71
level	5.34047699	76	leakage	5.134738445	69
leakage	6.032784462	65	control	3.558977842	68
director	7.441660404	62	mode	5.539895058	65
senior	7.215110779	57	pump	4.558169365	64

Table C.16: Top 20 Collocates with *Safety* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
health	8.035139084	383	related	17.99338531	336
public	6.22148037	381	analysis	16.43766594	240
related	16.82307625	342	*evaluation*	16.51710129	210
significance	6.784281254	330	*system*	14.75022888	167
evaluation	6.047266483	261	*plant*	15.09597111	156
low	6.722213745	237	function	16.6535759	146
report	16.01255417	195	report	15.59309864	145
analysis	5.606135845	175	public	17.17025757	136
system	4.263737202	148	injection	17.99544525	132
final	6.537470818	137	level	15.8893919	129
function	6.421710968	132	margin	17.28592873	118
systems	5.27366972	127	*nuclear*	15.07469082	116
green	6.657739639	123	systems	15.88733006	110
performance	4.678437233	96	limit	15.97159767	108
reactor	4.025981426	95	valves	16.03049088	105
level	5.03351593	94	health	18.04741859	104
quality	6.346329212	90	quality	16.66067505	102
will	3.096176147	90	final	17.59880638	101
endangering	8.405222893	86	relief	16.58278847	97
assessment	5.016344547	80	equipment	15.53485489	77

Table C.17: Top 20 Collocates with *Staff* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
NRC	6.215090752	1640	*NRC*	7.698168278	199
reviewed	6.035411835	323	review	6.225186825	34
finds	7.915836334	258	regulatory	6.319320679	33
concludes	7.741365433	234	*nuclear*	5.159678936	31
review	4.809133053	203	commission	7.536373615	30
determined	5.573278904	155	additional	5.910533905	25
therefore	5.904062748	151	information	5.042209148	25
commission	4.586680889	144	*plant*	4.384304047	24
will	3.47789979	143	member	8.281270027	23
licensee's	5.769069195	133	assessment	6.303571224	21
proposed	4.195500374	133	requests	7.721182823	19
applicant	5.425551891	126	EOF	7.147580147	17
licensee	3.779078484	120	attachment	5.258668423	17
evaluation	4.603430748	117	RAD	9.23575592	16
nuclear	3.333315372	117	requested	6.422524452	15
regulatory	4.549872398	114	reviewed	6.844841003	14
also	4.411246777	107	finds	10.49562263	13
information	3.768568516	103	approved	5.685558796	13
noted	5.902833462	82	support	5.302681446	13
applicant's	6.533966064	78	reference	4.673429489	13

Table C.18: Top 20 Collocates with *System* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	5.416099072	215	*reactor*	5.687889576	400
ADAMS	6.646003246	201	coolant	6.794568062	260
water	5.522006989	191	cooling	6.528035164	203
coolant	6.985007286	171	control	4.660782337	188
cooling	6.721608162	168	water	4.725152969	186
management	5.920147896	155	*pressure*	4.908655643	168
safety	4.2441082	146	*safety*	14.75022888	167
access	7.059785843	140	*containment*	5.014586926	157
documents	6.196671486	136	*emergency*	4.76870966	149
pressure	5.313690662	117	core	5.41232872	132
component	6.067446232	110	protection	6.037743092	126
emergency	5.813617229	104	monitoring	5.255322456	117
accession	6.828898907	97	component	5.999467373	107
control	4.807503223	96	components	4.882530212	90
document	6.001983643	95	spray	6.388686657	82
service	5.476703644	82	isolation	5.412918091	79
fire	4.627445221	78	ventilation	6.618504047	78
test	5.025726795	76	RCIC	7.43363905	77
protection	5.205891609	69	fire	5.207046509	77
NRC's	5.310227394	67	design	4.21124649	77

Table C.19: Top 20 Collocates with *Technical* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
specifications	9.953770638	610	specification	8.630458832	740
specification	8.86858654	169	specifications	9.065667152	636
evaluation	6.202083111	112	requirements	5.183705807	129
changes	6.078691006	76	*unit*	4.211921692	95
review	4.805459023	64	required	4.582011223	94
accordance	5.537491798	63	change	5.041919231	93
contained	7.152906895	61	changes	5.113653183	89
requirements	4.55325222	52	section	4.463361263	86
appendix	5.803310394	48	proposed	4.80202198	82
information	4.298318863	47	non	13.93445396	77
support	6.143160343	45	description	6.515664101	75
provide	4.563729286	42	brief	8.28217411	74
indicated	6.233102798	39	standard	7.223678112	74
justification	7.644813538	38	bases	6.65359211	68
basis	5.150466919	38	limits	5.186540604	60
staff	3.019039154	34	support	5.512200356	52
follows	7.486859322	31	improved	8.048271179	46
read	8.371785164	29	*plant*	3.500713825	45
detailed	7.001946926	29	analysis	4.050995827	40
environmental	4.55589962	28	requirement	5.268905163	39

Table C.20: Top 20 Collocates with *Unit* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
station	7.472971916	226	Millstone	7.658114433	268
nuclear	5.559057713	216	Salem	7.376135826	250
power	5.464097977	160	station	5.670389175	221
Point	6.982361317	142	*power*	4.492749214	191
Nos	8.471288681	130	BVPS	6.484416485	180
Mile	8.255328178	118	TMI	7.333350658	132
Millstone	8.120455742	113	Point	6.225543499	124
generating	8.013631821	102	Mile	7.275321007	109
Salem	7.077144146	74	*technical*	4.211921692	95
Indian	7.061257839	73	*nuclear*	3.928074598	80
Cliffs	7.688099861	65	TS	4.721031666	79
plant	4.184323788	62	*fuel*	3.917832613	79
system	4.016099453	60	Nine	7.088509083	77
Island	8.05134201	58	*reactor*	3.098562956	63
TMI	6.782075882	50	amendment	4.826218128	62
Seabrook	6.81506443	47	cycle	5.353283882	56
reactor	3.796452999	39	specifications	4.751283646	56
inspection	3.308708429	36	*operating*	3.93258667	56
Nine	7.047764778	34	Two	4.108299732	53
water	3.832116604	33	Valley	7.286233425	52

Table C.21: Top 20 Collocates with *Containment* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
primary	6.536973	98	*pressure*	5.512087822	199
spray	6.762039661	84	spray	7.83732748	177
leakage	4.676865578	66	*system*	4.592923164	174
pressure	3.875463247	64	inside	8.035918236	169
sump	6.085928917	56	isolation	7.182790756	166
isolation	5.56263876	54	sump	7.478246689	147
reactor	3.420665026	54	leakage	5.598255634	125
secondary	6.380730152	52	primary	6.767753124	115
integrity	5.948325157	51	outside	7.079697132	102
failure	4.560322762	45	atmosphere	8.624520302	99
building	5.285124779	43	rate	5.931252003	91
core	4.077034473	42	purge	8.405826569	83
inside	5.992591381	41	program	4.234399796	83
rate	4.781009197	41	cooling	5.66083765	78
outside	5.729199886	40	testing	4.678115845	78
system	4.064541817	40	air	5.504177094	74
concrete	6.580126286	36	temperature	4.882544041	73
cooling	4.504718781	35	building	5.903254032	66
fuel	3.031302929	35	exhaust	7.474216461	65
atmosphere	7.082626343	34	valves	5.12360096	65

Table C.22: Top 20 Collocates with *Emergency* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
diesel	7.628774643	146	*system*	5.254566669	250
plan	7.303643703	142	*power*	5.001301765	196
preparedness	10.60517502	132	cooling	6.781024933	154
cooling	6.468434811	124	site	6.431470871	150
system	4.070141792	110	core	5.983391285	143
core	5.578407288	108	response	5.642961025	143
generator	5.851748466	99	general	7.333370686	123
power	3.671994209	78	room	6.129281998	120
response	4.749876022	77	control	4.836362362	117
action	4.891605377	64	area	5.954145432	113
systems	4.466378689	48	diesel	7.082806587	100
operating	13.98337173	45	plan	6.783253193	99
generators	6.440339565	45	notification	7.088480473	83
procedure	5.179460526	44	director	7.91893959	78
water	3.584627628	44	path	7.344676971	77
ECCS	5.434068203	40	ECCS	6.340958595	75
feedwater	5.196897507	40	required	3.951940298	74
procedures	5.241645336	39	bus	7.68416357	68
level	4.00460577	39	generator	5.288480759	67
AC	7.161812782	37	center	7.287984848	66

Table C.23: Top 20 Collocates with *Evaluation* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
safety	5.907670021	258	*safety*	5.890796185	255
NRC	4.711368561	121	model	7.134250164	138
staff	6.655613899	106	report	5.136826515	104
report	5.06574297	99	Westinghouse	7.551794529	102
technical	5.053908348	84	performed	4.858100414	93
licensee's	9.732678413	74	engineering	7.014959335	89
staff's	9.415431976	69	break	6.272208691	78
performed	4.427466393	69	ECCS	6.460128307	75
regulatory	5.166237831	66	*NRC*	3.982330084	73
licensee	6.77098465	63	based	4.768656731	69
cause	6.05313158	62	models	7.734262943	67
based	4.151985168	45	*technical*	4.661591053	64
flaw	6.88524437	44	summary	6.425576687	63
information	3.872903109	44	LOCA	6.113065243	61
process	5.357347012	42	will	3.187175989	61
results	4.252298355	42	results	4.692871094	57
root	7.181844711	39	small	6.394868374	55
identified	4.585985661	38	change	3.989067078	51
SE	8.460941315	32	*plant*	3.239432335	46
engineering	5.539225578	32	large	6.215251446	44

Table C.24: Top 20 Collocates with *Fuel* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
spent	9.858260155	528	oil	16.67119026	383
pool	8.408031464	278	spent	9.131278992	319
storage	7.733652115	232	storage	8.123599052	304
assemblies	7.978959084	175	assemblies	8.398112297	234
oil	14.93548012	115	pool	8.003335953	210
assembly	7.345451832	114	handling	8.682447433	193
design	4.686065197	112	movement	8.595518112	159
rod	6.200694084	110	irradiated	8.94189167	148
rods	7.562743187	102	assembly	7.631502628	139
cladding	7.319161415	80	accident	5.260891914	127
system	3.209182501	80	design	4.711600304	114
core	4.591917038	72	cycle	6.141459465	107
reactor	3.552799463	71	tank	6.818030834	94
cycle	5.529275417	70	diesel	6.513155937	89
cooling	5.241993427	70	cladding	7.282635689	78
irradiated	7.639328957	60	mark	8.448444366	75
use	4.672700405	60	core	4.650810719	75
pools	10.2329483	53	*system*	3.116073132	75
handling	6.817910671	53	rod	5.628787518	74
damage	6.417031288	53	BW	8.972950935	71

Table C.25: Top 20 Collocates with *Inspection* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
scope	9.442183495	580	inservice	17.52257729	287
inspectors	12.88630962	427	program	5.858669281	255
NRC	5.865918159	321	visual	7.439559937	222
public	8.385505676	314	interval	7.711508751	200
reviewed	8.66476059	266	will	4.222590446	149
results	6.130526066	184	report	5.125884056	123
available	6.947746754	159	methods	7.270933628	117
report	5.47799015	157	One	5.2892766	112
inservice	16.63379478	155	performed	4.80739212	107
conducted	7.88013792	144	time	4.855170727	105
period	6.382178783	141	*NRC*	4.168845177	99
plan	7.11808157	137	tube	5.575602531	74
findings	9.214306831	134	results	4.796788692	73
program	4.853000641	127	metal	6.864902496	70
team	8.666998863	114	penetration	6.492600918	70
activities	6.610437393	112	next	7.144267559	69
electronically	13.5408926	109	plan	6.128573895	69
procedure	6.314102173	106	year	6.027403831	69
performed	4.70978117	100	activities	5.911606789	69
will	3.58838439	96	period	5.351151943	69

Table C.26: Top 20 Collocates with *Inspectors* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reviewed	16.00720596	687			
inspection	12.88630962	427			
scope	14.57122612	323			
also	12.25312042	173			
licensee	13.84516621	161			
observed	14.9892292	157			
determined	12.20323849	128			
identified	12.00111771	123			
verified	14.57001972	118			
evaluated	12.40028954	115			
licensee's	16.02512932	110			
selected	13.581954	105			
performed	10.4318037	84			
following	10.70152092	82			
conducted	13.00570202	80			
resident	16.59062004	79			
plant	9.502504349	67			
results	10.53472233	62			
concluded	12.954669	57			
review	10.88799858	57			

Table C.27: Top 20 Collocates with *License* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
renewal	11.419981	763	renewal	10.21948147	332
amendment	9.226385117	449	amendment	8.644632339	300
operating	18.02183151	387	*operating*	16.76549339	162
renewed	11.56099415	268	request	7.255717754	135
facility	9.350541115	262	proposed	5.974571228	118
condition	7.112365246	184	scope	7.930158138	97
scope	8.781456947	175	facility	7.743382931	86
application	8.291215897	128	condition	5.874205589	78
hereby	10.79910088	114	amendments	8.960350037	67
incorporated	9.393509865	111	application	7.33561039	66
request	6.778814793	97	renewed	9.327795029	57
effective	8.397859573	90	within	4.983037472	49
attachment	7.136881351	88	conditions	4.797743797	40
amended	10.49354649	87	*unit*	3.406051874	40
proposed	5.449480057	82	will	3.393321276	40
accordingly	9.902080536	80	LAR	8.05151844	38
within	5.672108173	79	*NRC*	3.777384758	36
paragraph	8.114999771	77	*nuclear*	4.20343399	35
environmental	6.55117321	72	change	3.989577293	29
nuclear	5.20343399	70	dated	5.3795681	26

Table C.28: Top 20 Collocates with *Licensee* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
identified	9.615980148	287	event	6.767392635	66
stated	11.15826225	248	report	6.5937953	66
staff	9.822067261	220	selected	9.067533493	56
proposed	7.929146767	186	commitments	9.22280407	53
inspectors	13.91512871	169	LER	9.351734161	34
NRC	7.19092226	156	shall	5.62528038	27
determined	8.755579948	143	will	4.124340534	27
company	9.668497086	141	controlled	8.366150856	24
also	8.329903603	139	provide	5.303321362	20
provided	7.977748394	132	commission	7.331385136	19
requested	9.558718681	130	states	7.054423809	18
performed	7.362693787	122	actions	5.427041054	18
reviewed	9.882400513	120	document	6.698563099	16
shall	7.728374004	116	required	3.974671125	16
power	6.47675705	116	*NRC*	3.905519962	16
failed	10.57899761	109	regulatory	5.141748905	15
submitted	9.31446743	109	manual	6.695007324	14
will	6.137637615	109	may	4.666097641	14
verify	9.627943993	102	concludes	8.723702431	13
dated	8.531777382	94	changes	4.443288803	13

Table C.29: Top 20 Collocates with *NRC* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
staff	9.735607147	1659	*staff*	7.444650173	339
site	7.134077072	416	letter	6.675555706	325
regulatory	6.863015652	396	approved	7.444038868	310
commission	8.648312569	379	dated	6.827738285	231
nuclear	5.85885191	359	approval	7.532319546	220
inspection	5.874878883	323	responses	5.357931137	200
public	7.717811584	307	request	6.09781456	197
web	13.09505749	290	regulatory	5.818621635	192
approved	7.286855698	278	TVA	5.976649761	178
review	6.463187218	259	question	6.909134865	171
ADAMS	10.98684406	221	bulletin	7.518489361	158
staff's	10.003685	192	*nuclear*	4.489953041	139
determined	6.078324795	179	information	4.591375828	134
finds	12.02831173	178	review	5.423658848	126
will	3.699171066	161	provided	4.839045525	120
document	7.001144886	158	submitted	6.326443672	110
licensee	7.172306538	154	SER	7.688556194	105
room	5.691781521	151	requested	6.235590935	104
reviewed	7.183763981	148	*inspection*	4.168845177	99
information	4.623318672	137	commission	6.682171345	97

Table C.30: Top 20 Collocates with *Nuclear* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
commission	10.02670765	713	*plant*	6.532781601	604
regulatory	8.073215485	663	*power*	5.642689228	377
plant	6.52799654	602	station	7.604232311	351
power	6.205801964	557	regulatory	6.860221863	286
NRC	5.85885191	359	commission	8.598798752	265
station	7.461788177	318	*unit*	4.635578632	221
units	6.636398792	257	plants	7.970487595	168
plants	8.285529137	209	generation	8.060408592	166
unit	4.371237755	184	company	7.154309273	143
staff	6.809980869	158	*NRC*	4.489953041	139
operating	15.24447727	133	Ferry	8.316656113	136
Southern	8.65182209	113	Browns	8.300226212	135
reactor	4.202696323	104	units	5.664196968	131
McGuire	7.526967049	103	Energy	6.429934025	122
reactors	8.635941505	92	*operating*	15.10805798	121
company	6.502232552	91	*safety*	4.357592106	118
Catawba	7.302576542	86	Southern	8.545887947	105
Energy	5.925461769	86	chief	9.263195038	100
office	8.408761024	85	McGuire	7.46982336	99
Harris	8.735162735	73	office	8.26615715	77

Table C.31: Top 20 Collocates with *Operating* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
license	18.02183151	387	experience	19.03281403	410
facility	18.80284882	331	*license*	16.76549339	162
experience	17.98269081	198	limits	16.20870399	143
renewed	20.13351059	184	cycle	16.92750359	135
licenses	19.87586975	164	*unit*	14.22153473	130
nuclear	15.24447727	133	core	15.88391495	129
cycle	16.60557556	108	facility	17.3979454	125
limits	15.73533344	103	*nuclear*	15.10805798	121
procedures	16.63476753	99	licenses	19.3127346	111
Southern	18.69092941	91	report	15.16080093	111
plant	14.12160206	89	normal	16.7018261	110
company	16.7553196	85	based	15.35831451	109
test	14.79525948	85	conditions	15.33232689	107
core	15.26500511	84	*plant*	14.28972435	100
tests	16.5915451	83	Southern	18.72229385	93
power	13.64521313	74	company	16.85372353	91
system	13.4876976	71	renewed	18.98498917	83
renewal	17.08876801	70	industry	16.93582344	72
written	18.18160439	69	*power*	13.58550739	71
prior	15.70741844	67	*pressure*	14.06343365	64

Table C.32: Top 20 Collocates with *Plant* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	6.530391216	603	*nuclear*	6.532781601	604
power	4.735699177	293	*unit*	4.827911377	368
specific	6.565774918	254	specific	6.801091671	299
unit	4.025448799	211	operation	5.231930733	205
units	5.626606464	186	*power*	4.213367462	204
safety	3.862360239	122	electric	7.215807438	203
electric	6.408452511	116	steam	4.824369431	145
performance	5.224404812	105	conditions	4.825324535	140
Hatch	7.171382427	104	will	3.291619062	128
steam	4.168022156	92	Ferry	7.568509579	118
operation	4.076012611	92	Browns	7.562726974	118
operating	14.12160206	89	units	4.970090866	118
Harris	8.461443901	88	Hatch	7.291121483	113
Lucie	5.78960371	87	procedures	5.824838638	105
conditions	4.10543251	85	systems	4.750011921	105
modifications	6.502910614	84	*operating*	14.28972435	100
risk	5.733195305	83	*safety*	3.560979605	99
site	4.604929924	76	design	3.926441669	90
procedures	5.280518055	72	operations	5.43871069	85
areas	5.540073872	69	generating	7.62879467	84

Table C.33: Top 20 Collocates with *Power* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	6.205801964	557	*nuclear*	5.642689228	377
plant	4.735699177	293	company	7.498390675	306
Light	8.27585125	273	thermal	7.073481083	275
company	7.30709219	268	offsite	7.004857063	260
electric	6.697915077	164	station	6.355472088	249
plants	7.110137463	156	light	8.077912331	238
reactor	4.034224033	156	uprate	8.098725319	209
station	5.624289036	150	*reactor*	4.449261665	208
offsite	6.191942692	148	*plant*	4.213367462	204
operation	4.512267113	144	*unit*	3.738097191	200
unit	3.264166117	144	*emergency*	5.001301765	196
reactors	8.488226891	140	level	5.263358116	194
loss	5.229652405	140	loss	5.60816431	182
Carolina	7.382907867	134	operation	4.834195614	180
uprate	7.446648121	133	core	5.168640614	169
Florida	7.250884533	128	full	6.594053268	166
core	4.74504137	126	Florida	7.572812557	160
percent	6.273740292	122	plants	7.100859642	155
thermal	5.889056683	121	rated	7.89702034	148
licensee	6.47675705	116	conditions	4.685605526	147

Table C. 34: Top 20 Collocates with *Pressure* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	5.020351887	200	*reactor*	6.164398193	442
system	4.320129395	176	boundary	8.206480026	339
vessel	5.940051556	140	temperature	6.674901009	309
boundary	6.888802052	136	vessel	6.892917633	271
RCS	6.357881069	130	coolant	16.81832123	216
high	5.965313435	119	RCS	7.022014141	206
temperature	5.171622276	109	*containment*	5.512087822	199
coolant	15.66334629	97	high	6.632738113	189
injection	6.685532093	91	*system*	4.407592297	187
low	5.654042721	88	low	6.489232063	157
code	4.782361984	77	steam	5.201577187	141
containment	3.875463247	64	differential	8.72943306	119
retaining	8.684922218	60	leakage	5.163368702	113
components	4.489534378	60	code	5.27046299	108
boiler	7.9603405	54	ASME	5.75416708	103
ASME	4.822554111	54	test	4.432177544	92
test	3.66350317	54	peak	7.301593781	92
differential	7.535054684	52	boiler	8.664884567	88
pressurizer	5.752646446	52	design	4.278418541	86
PSIG	6.519113541	48	PSIG	7.237931728	79

Table C.35: Top 20 Collocates with *Reactor* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
coolant	17.59184074	427	vessel	8.210261345	781
system	4.945632935	314	coolant	18.34158897	718
vessel	6.561084747	249	*system*	5.730496883	541
trip	6.714819431	230	*pressure*	6.164398193	442
pressure	5.020351887	200	trip	7.159348488	313
power	4.034224033	156	water	5.777444839	313
building	6.617280483	153	building	7.242535591	236
core	5.414567947	150	level	5.849617958	218
water	4.564241409	135	*power*	4.449261665	208
safety	4.16187048	130	core	5.653355122	177
head	5.572684765	111	head	6.170586109	168
process	5.989449501	110	boundary	6.940381527	163
oversight	9.774313927	106	protection	6.270807743	119
nuclear	4.202696323	104	RCS	5.740537167	98
unit	3.141585827	99	pump	5.194662094	98
senior	9.116782188	78	lower	5.735750675	88
office	7.930709839	77	leakage	4.592935562	88
fuel	3.552799463	71	low	5.427868843	87
pump	4.667415142	68	cooling	5.302732944	86
operator	6.122175694	67	scram	7.44742775	83

Table C.36: Top 20 Collocates with *Safety* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
significance	10.35695553	442	related	7.512046337	384
public	8.249651909	413	analysis	6.111651897	354
health	9.068122864	389	*evaluation*	5.890796185	255
related	7.344936371	342	function	6.509432316	235
low	7.301914215	325	*system*	4.411607265	221
evaluation	5.902067184	257	injection	7.6618433	211
analysis	5.627039909	253	report	5.303190708	201
report	5.400331974	215	health	8.079206467	196
level	5.553971767	181	margin	6.977171898	189
quality	6.83864975	178	analyses	6.244413376	164
function	5.945798397	159	public	6.899504185	162
green	11.40175343	149	final	7.465693474	132
final	7.62097168	147	valves	5.529154778	124
system	3.742645741	139	*nuclear*	4.357592106	118
reactor	4.16187048	130	systems	5.061828613	115
injection	6.894940376	124	level	4.887016296	114
plant	3.862360239	122	quality	6.131100655	109
systems	5.098981857	118	updated	7.227403164	102
valves	5.432939053	116	limit	5.479949474	99
performance	5.498086452	112	*plant*	3.560979605	99

Table C.37: Top 20 Collocates with *Staff* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
NRC	9.745141029	1670	*NRC*	7.444650173	339
finds	14.91658497	300	CECC	8.262604713	71
concludes	12.31206989	285	information	5.332900047	51
licensee	9.741147995	208	requested	7.162192822	45
reviewed	9.781924248	204	proposed	4.653260231	35
review	8.097267151	183	question	6.713933468	34
therefore	7.8684659	168	response	4.849263191	32
proposed	6.89029932	165	review	5.535763741	31
nuclear	6.809980869	158	additional	5.212840557	29
licensee's	12.05563068	156	position	5.838742733	26
determined	7.996483803	154	will	3.147298336	25
information	6.927261353	154	determined	5.253259182	23
commission	9.465373993	152	assessment	5.737397194	23
regulatory	7.508398533	141	requests	6.972827435	22
applicant	10.24561691	123	letter	4.858765602	21
requested	8.528320313	116	operations	5.633460045	21
evaluation	6.773183823	115	control	3.502095699	18
will	5.203881741	104	reviewed	6.196961403	17
members	10.29426193	90	*plant*	3.050336599	17
concluded	8.988088608	81	finds	10.68776703	16

Table C.38: Top 20 Collocates with *System* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	4.941030979	313	*reactor*	5.730496883	541
coolant	16.11273766	219	coolant	16.60941315	309
ADAMS	10.56927586	216	*emergency*	5.254566669	250
water	4.692727089	211	cooling	6.243313313	236
cooling	6.040150166	205	water	4.841992855	234
component	6.307253361	186	*safety*	4.411607265	221
pressure	4.320129395	176	control	4.41975975	195
control	3.920953751	138	*pressure*	4.407592297	187
safety	3.732228994	138	*containment*	4.592923164	174
protect	5.871418953	129	ventilation	7.316513062	167
document	6.312810421	128	protection	6.200041771	162
management	5.64218092	119	core	4.849860668	145
documents	7.091310024	118	component	5.907645702	141
access	6.72375679	114	*power*	3.33105588	137
emergency	4.083198071	111	test	4.238414288	133
NRC's	7.852523327	110	leakage	4.571902275	124
fire	5.485291958	103	spray	6.261847019	120
service	4.891963959	98	air	5.186849117	120
test	3.783044815	97	RCS	5.442919731	114
components	5.310227394	67	inoperable	5.175465584	108

Table C.39: Top 20 Collocates with *Technical* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
specifications	9.709494591	892	specification	9.238760948	776
specification	7.325348377	206	specifications	9.426263809	733
changes	5.829904079	149	change	5.601323605	158
accordance	5.715247631	105	requirements	5.24934864	134
contained	7.493454933	103	proposed	5.212070942	124
appendix	5.730815887	89	changes	5.430914402	113
evaluation	5.053908348	84	***unit***	3.879426718	99
indicated	7.586912632	77	bases	16.09938431	91
basis	5.14069891	77	required	4.217597008	83
follows	7.261668205	72	support	5.618283749	75
environmental	5.696941853	71	basis	4.940001965	67
read	9.409750938	67	*evaluation*	4.661591053	64
requirements	4.137455463	62	force	8.205935478	62
information	4.122292995	53	report	4.371483803	62
TSS	9.044100761	52	standard	6.967737675	61
NRC	3.473845959	52	section	4.018003941	61
protection	5.786256313	51	task	8.834859848	58
provide	4.341132641	45	information	4.175732136	55
review	4.707782269	42	prohibited	8.795947075	53
proposed	3.61542654	41	accordance	4.701441765	52

Table C.40: Top 20 Collocates with *Unit* by Industry Group for Region II

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
plant	4.025448799	211	Lucie	7.446853638	374
nuclear	4.371237755	184	*plant*	4.827911377	368
Lucie	6.326224327	172	*nuclear*	4.635578632	221
power	3.264166117	144	HBRSEP	8.090834618	218
station	5.041520596	118	cycle	6.195670128	206
outage	4.788952827	103	*power*	3.738097191	200
reactor	3.141585827	99	refueling	5.527836323	181
refueling	4.454255104	86	outage	5.553663254	175
WBN	6.614204884	71	steam	4.590076447	168
steam	3.175039053	63	North	5.991664886	160
shutdown	4.156281471	58	Anna	6.106908798	159
Bar	6.500675678	57	BFN	6.317151546	156
room	3.762930632	57	mode	5.481079578	150
license	3.891478539	56	electric	6.191233158	136
facility	4.84553051	54	*operating*	14.22153473	130
Watts	7.404486179	49	shutdown	5.275580406	126
electric	4.71848011	49	operation	3.749605894	100
building	4.320021629	49	*technical*	3.879426718	99
B	12.7841301	48	TS	3.871177435	95
operating	12.75375652	47	restart	7.061663628	90

Table C.41: Top 20 Collocates with *Containment* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
primary	8.785447121	137	isolation	8.121419907	209
isolation	7.978878021	96	primary	7.992083073	201
spray	8.413210869	91	*system*	4.663704395	131
system	15.34874821	79	secondary	8.324370384	130
integrity	7.380081177	79	spray	7.83024931	90
unit	4.969876289	62	leakage	5.564303398	84
valaves	6.644521713	59	*pressure*	15.81146812	83
leakage	6.350379944	57	integrity	7.385202885	83
pressure	5.679894447	56	valves	5.773463249	81
inspection	3.609194756	51	radiation	5.897062778	75
cooling	6.375554562	47	atmosphere	8.528059006	65
vessel	5.92048502	43	rate	5.634413719	58
interval	7.075340748	42	*reactor*	3.70710063	58
fuel	4.771342278	41	monitor	6.430002213	56
reactor	3.933831453	40	cooling	5.714352608	56
secondary	8.399632454	39	temperature	4.693111897	51
pool	6.466183186	39	air	5.673017025	50
water	4.840653419	38	outside	6.703978539	48
core	5.391783237	37	sump	7.337938786	46
program	13.08936119	33	design	4.319745541	46

Table C.42: Top 20 Collocates with *Emergency* by Industry Group for Region I

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
preparedness	9.978368759	198	*system*	5.095695019	146
plan	7.418410301	153	response	6.060334206	125
diesel	8.539961815	119	cooling	6.883026123	104
system	15.78181458	114	diesel	8.193946838	103
response	6.157762051	96	general	7.443238735	94
cooling	7.248485088	92	site	6.32275629	90
generator	7.077780247	79	director	8.663107872	87
core	6.353074074	77	plan	7.09871006	84
room	5.859004498	69	core	5.386616707	83
control	4.980258465	64	area	5.907741547	67
systems	15.92620468	63	procedure	5.864627838	66
procedures	5.142715454	45	operations	6.64217329	63
water	4.956115246	44	implementing	8.747937202	61
power	13.01931477	42	notification	7.613283157	60
action	4.054418087	39	classification	8.855752945	58
generators	8.376673698	38	center	7.352474689	53
drill	8.049854279	31	procedures	6.129324913	53
exercise	7.269635201	31	*plant*	3.891115665	53
equipment	4.668530464	29	generators	8.247348785	50
cornerstone	6.428852081	29	generator	6.390089035	50

Table C.43: Top 20 Collocates with *Evaluation* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
safety	5.72859478	313	*safety*	6.044932842	177
NRC	4.303498268	170	report	5.022488117	64
cause	6.854895115	121	*NRC*	14.81435013	63
licensee's	5.278187752	116	performed	4.895691395	62
technical	5.289436817	92	model	6.324000835	48
staff	4.332853317	91	proposed	4.319526672	47
staff's	7.185507774	85	changes	4.374519348	42
report	15.72612572	82	cause	5.934879303	41
root	7.716496944	77	flaw	7.26418066	40
regulatory	4.923558712	64	criteria	5.288001537	39
risk	4.984717846	57	attachment	5.466612339	38
performed	4.778056145	54	results	14.96635342	35
operability	5.635015011	53	summary	5.949545383	35
review	4.147172451	52	*inspection*	4.474500656	35
based	4.754441261	51	determined	5.04916811	32
proposed	3.858315229	45	based	4.420272827	32
dated	4.397163868	43	used	3.971087933	31
reviewed	3.562417507	43	change	3.87293601	28
inspectors	14.72612572	41	concluded	6.759068012	27
plant	3.024386168	41	using	4.673094749	27

Table C.44: Top 20 Collocates with *Fuel* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
spent	9.507479668	432	spent	9.253917694	391
pool	8.568092346	247	storage	8.418066978	266
storage	8.168585777	219	oil	8.763503075	215
assemblies	8.906853676	122	pool	7.891112328	183
handling	8.634327888	101	assembly	7.695273876	138
oil	8.214878082	100	cycle	16.27122498	134
core	6.265130043	100	rack	7.624673843	107
design	5.414618492	94	handling	8.343518257	82
reactor	4.240671635	73	assemblies	8.062841415	82
racks	7.91654253	55	rods	7.337238789	80
cycle	6.382627487	55	rod	5.642490864	76
damage	6.997512341	53	design	4.754714489	73
cooling	5.987829685	53	***nuclear***	4.301733494	69
capacity	7.413521767	50	***reactor***	3.726350784	69
diesel	6.82398653	50	racks	8.564677238	68
loading	7.737956524	48	new	5.376671314	63
Westinghouse	7.320680618	48	will	3.313206673	63
new	5.901455402	48	clad	8.157829285	60
stored	8.65383625	45	***unit***	3.547165632	59
assembly	8.251964569	43	cells	8.493260384	58

Table C.45: Top 20 Collocates with *Inspection* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
scope	7.621675968	892	inservice	8.944189072	247
inspectors	17.6704483	811	program	5.817509174	139
reviewed	5.588250637	450	interval	7.418962479	107
NRC	4.14344883	391	visual	7.591618061	96
public	5.785609245	371	requirements	4.625608921	71
results	5.904139519	254	methods	6.703540325	55
report	15.87725544	234	period	5.620090485	54
period	6.30697155	213	results	15.28713608	50
one	5.216813564	204	year	6.321352005	50
available	5.563681126	187	used	4.466957092	50
findings	4.179538727	174	performed	4.332667351	48
conducted	5.605253696	169	*NRC*	14.1978693	47
electronically	7.642126083	146	penetration	6.603751659	45
inservice	7.385402203	143	third	7.690277576	39
inspector	5.534402847	140	*nuclear*	3.916756153	39
constituted	7.586558819	136	second	6.386940479	36
activities	4.452528954	135	steam	4.533594608	36
Mr.	5.763572216	132	*plant*	3.264383078	36
program	12.99557495	127	guidelines	6.5827384	35
sample	6.041656017	126	SG	6.401875496	35

Table C.46: Top 20 Collocates with *Inspectors* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reviewed	19.19214439	1269			
inspection	17.6704483	811			
scope	19.21437645	624			
also	18.12847519	406			
licensee's	17.64650154	357			
identified	16.81599045	293			
determined	18.12251854	292			
verified	19.27054214	279			
licensee	15.85364532	274			
observed	19.06899261	231			
evaluated	18.4698658	211			
finding	17.50408363	183			
performed	17.12723541	164			
selected	18.13646126	153			
concluded	18.25463867	141			
resident	19.0853405	120			
conducted	17.15806198	115			
addition	17.53012466	111			
following	16.27030754	106			
results	16.68186378	101			

Table C.47: Top 20 Collocates with *License* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
operating	7.086558819	421	amendment	8.740703583	320
amendment	6.954901695	352	renewal	10.07718849	228
renewal	8.799072266	285	*operating*	7.089099884	192
facility	6.848906994	234	request	7.328952789	120
condition	6.151546955	196	application	8.062797546	109
renewed	8.889392853	131	facility	7.522930145	72
conditions	5.451569557	101	scope	7.818071365	58
incorporated	7.672662735	96	condition	5.608175278	56
hereby	7.967240334	95	within	5.394892216	55
attachment	6.584606647	91	TS	5.212084293	50
effective	7.471750259	87	proposed	4.585793972	44
proposed	4.487535	82	Nos.	8.591337204	43
amended	6.845693111	74	*nuclear*	4.612833977	43
request	5.542131424	72	permit	7.935460567	40
accordingly	7.639325619	70	construction	7.804417133	39
changes	4.891780376	70	following	4.641419411	37
application	5.486490726	66	amendments	8.379735947	36
B	3.223574638	59	C	4.581110954	33
paragraph	7.04777813	54	manner	7.55849123	30
conducted	5.00200367	51	approved	5.747833729	30

Table C.48: Top 20 Collocates with *Licensee* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
identified	4.717084885	351	report	7.213420868	120
inspectors	15.94048882	291	event	6.827141762	99
staff	3.719892502	182	shall	6.714771271	78
proposed	4.245354176	180	LER	8.743954659	41
shall	5.108174801	177	violations	9.928467751	35
company	5.842179775	176	company	6.840651035	30
requested	5.554366589	162	C	5.366225243	30
entered	6.073551655	158	states	7.303136349	29
performance	3.831738234	152	requested	6.562844753	25
failed	6.46873188	150	employee	9.198280334	24
also	4.273589611	144	stated	7.566523552	24
management	5.077541351	140	Donald	9.121112823	22
stated	5.694644928	139	*unit*	4.039931774	22
facility	4.700093269	137	reports	7.308858871	21
provided	4.761088371	132	environmental	5.857005119	21
response	4.390715122	129	commission	7.125865936	19
reviewed	3.534419298	129	protection	5.94907999	19
violations	6.505238056	124	controlled	7.661681175	18
submitted	5.698127747	123	one	4.795182705	18
information	3.657152891	123	*nuclear*	4.279111385	18

Table C.49: Top 20 Collocates with *NRC* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
staff	6.519171238	1372	approved	18.10400963	321
site	16.25533104	392	request	16.58751869	150
inspection	4.14344883	391	reference	15.67570114	140
public	5.343496799	352	regulatory	16.27715302	136
regulatory	5.650934219	351	*staff*	17.31108093	135
commission	5.605751038	327	letter	16.20597267	120
nuclear	4.622005939	308	*nuclear*	14.89139175	109
review	4.848063946	280	commission	17.56438446	102
ADAMS	6.055659294	254	approval	17.56953621	100
web	7.424303532	204	information	14.96071243	89
staff's	6.595058918	187	additional	15.29151917	87
document	5.441513538	174	requested	16.29056168	82
room	4.876804829	173	response	14.83959389	80
oversight	6.335767269	170	submitted	16.7794342	76
evaluation	4.303498268	170	provided	15.14357376	75
approved	5.800816059	157	issued	17.31857872	67
concludes	6.557632446	152	*evaluation*	14.81435013	63
enforcement	6.015105247	152	consider	17.62989807	62
reviewed	3.578125715	144	*safety*	13.80612755	61
determined	4.577168465	139	recommend	18.64970589	60

Table C.50: Top 20 Collocates with *Nuclear* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
power	16.06999969	756	*power*	6.66566515	625
commission	7.919426918	713	*plant*	6.819937706	607
regulatory	7.72452116	648	station	6.23096323	264
plant	6.046809673	484	regulatory	6.821619511	200
NRC	4.612607479	306	commission	8.472490311	193
station	7.011921406	217	generating	8.545230865	138
plants	7.678072929	185	company	7.023317814	136
company	6.916252613	176	*unit*	4.648684025	134
units	13.95070171	174	management	6.941003799	117
reactor	4.797501564	169	units	6.436223507	116
office	7.167267323	139	security	6.308638573	114
regulation	8.206453323	124	*NRC*	14.89139175	109
management	5.92922163	120	plants	7.716950417	103
reactors	7.786058903	114	Cook	8.558422089	96
operation	5.164421558	99	Quad	7.889893532	94
operating	4.695830345	99	Cities	7.960268497	92
commercial	8.065919876	96	Dresden	7.79804039	89
LLC	6.880448818	94	Prairie	8.254851341	86
special	7.482882977	92	LLC	7.304122925	70
staff	3.809732676	92	Energy	6.202641487	70

Table C.51: Top 20 Collocates with *Operating* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
license	7.086558819	421	*license*	7.096594334	193
facility	7.192677021	255	limits	6.642217159	188
test	5.596988201	116	report	6.165351391	186
cycle	7.250592232	110	cycle	16.72617722	156
renewed	8.817218781	107	core	5.969154358	150
experience	8.337661743	101	experience	8.180908203	114
power	13.65930462	99	facility	7.336751938	107
nuclear	4.695830345	99	*plant*	4.446542263	94
licenses	8.427426338	98	*power*	4.017449379	80
plant	4.125869751	89	procedures	6.396948814	77
system	14.6372776	78	normal	6.454296112	67
procedures	5.320599556	77	annual	6.959317207	65
tests	6.711418152	68	environmental	5.784350395	64
amendment	4.573232174	58	radiological	6.437869549	62
written	6.248127937	57	procedure	5.503187656	62
procedure	5.249064922	55	conditions	4.795058727	62
limits	5.978923798	50	*unit*	3.78280592	59
current	5.450767517	50	*system*	3.492609501	58
reactor	3.562668085	50	*pressure*	15.27366543	57
conditions	4.627849102	49	temperature	4.806393147	55

Table C.52: Top 20 Collocates with *Plant* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	6.049787045	485	***nuclear***	6.819937706	607
power	14.58727551	361	***power***	4.846681595	257
specific	6.510455608	265	specific	16.23736382	201
performance	4.607868195	165	operation	5.277305603	157
safety	3.673174858	146	conditions	5.073861122	136
modifications	6.998311043	138	***safety***	4.30384779	126
risk	5.230025768	131	generating	7.782453537	118
operations	6.209304333	130	Cook	8.0947752	101
status	6.419803619	112	shutdown	5.460467339	96
operation	4.900087357	110	***unit***	3.615554333	95
areas	5.387321949	103	***operating***	4.446542263	94
design	4.475967407	103	design	4.469628334	92
unit	4.070561886	103	equipment	4.999967575	87
units	12.73527622	100	data	4.434630394	84
inspectors	15.04270554	99	Prairie	7.666729927	83
procedures	4.638847828	92	Island	7.499896526	81
conditions	4.598174095	92	procedures	5.579479218	79
operating	4.157930851	91	systems	4.903738022	75
equipment	4.227281094	79	personnel	4.918608189	72
Cook	7.175188541	78	Monticello	7.326667309	71

Table C.53: Top 20 Collocates with *Power* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	16.06999969	756	*nuclear*	6.66566515	625
plant	14.58327389	360	station	5.887869358	346
station	16.46177101	304	*plant*	4.852284431	258
offsite	24.02837563	226	thermal	6.663380146	249
reactor	14.14145565	220	level	5.74577713	235
unit	14.33372879	190	*reactor*	4.583930016	220
percent	15.87764454	189	*unit*	4.528746605	205
electrical	16.51593208	177	uprate	8.333571434	201
reactors	17.28296471	165	full	6.944468021	188
plants	16.42304802	159	core	5.020356178	161
operation	14.81142235	159	operation	5.043640614	153
full	16.44593811	147	rated	7.884884834	136
uprate	17.74030876	134	loss	5.91267252	132
AC	16.81130791	120	offsite	6.100655079	116
inspection	12.42642975	107	company	6.009881973	112
loss	15.06164932	106	flow	4.489066601	111
system	23.45137215	101	AC	7.699362278	101
units	22.12944412	101	Quad	7.216686249	98
thermal	16.04623032	101	range	6.311101913	98
company	15.06415081	100	Cities	7.288341045	96

Table C.54: Top 20 Collocates with *Pressure* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	6.41046381	239	*reactor*	17.25518417	418
vessel	7.923295498	185	temperature	17.66190529	246
high	7.291100502	127	vessel	17.97056198	240
system	15.81302547	117	boundary	18.9669323	186
low	6.659206867	114	coolant	18.03264999	163
boundary	8.434048653	97	high	17.73734665	163
coolant	7.735218525	97	low	17.40241623	136
injection	8.445473671	96	*system*	15.35135365	130
code	15.09240818	71	RCS	17.5080452	90
test	5.479539394	71	code	16.71302032	89
temperature	6.959667206	64	PSIG	18.99762917	85
containment	5.679894447	56	*containment*	15.81146812	83
head	6.769656181	52	injection	17.90397072	82
control	4.588703632	49	flow	15.79745674	82
ASME	5.977925777	44	leakage	16.21054077	81
boiler	9.641004562	43	ASME	16.82807922	78
PSIG	8.909493446	43	control	15.15845585	76
RPV	6.526575089	43	differential	19.31909561	71
retaining	9.704501152	39	RPV	17.06873131	69
valve	5.505255222	39	boiler	19.72548866	67

Table C.55: Top 20 Collocates with *Reactor* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
vessel	7.821275711	413	vessel	7.923445225	698
coolant	8.054578781	290	coolant	7.881247997	441
system	15.58289146	239	*pressure*	17.25518417	418
pressure	6.404414654	238	*system*	5.445580959	417
power	14.14145565	220	water	14.51591492	313
safety	4.403078079	201	*power*	4.583930016	220
trip	7.077508926	171	level	5.692541599	203
nuclear	4.797501564	169	trip	6.597609043	199
water	5.596105099	165	core	5.476741314	198
head	7.157308578	163	head	6.460947037	145
office	7.040215969	141	building	6.890453339	136
core	5.927911282	138	scram	6.596155643	129
operator	5.909434795	138	temperature	5.062819958	122
regulation	7.987238407	118	internals	7.627792835	102
building	7.195669651	111	*unit*	3.651068211	100
process	14.46338367	110	control	3.92313242	97
oversight	6.709245682	107	protection	5.697719574	95
senior	7.591170788	100	shutdown	5.229335308	84
pump	5.182178497	96	RCS	5.768733501	81
lecel	4.860022545	84	boundary	6.107251644	77

Table C.56: Top 20 Collocates with *Safety* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
significance	6.54826355	683	analysis	6.146037579	342
low	6.981635571	482	related	7.748925686	322
public	6.286514759	464	*system*	4.639098644	190
health	7.974523067	420	*evaluation*	6.044932842	177
related	6.543272018	340	report	5.512018681	175
evaluation	5.72859478	313	function	6.46807003	164
analysis	5.86210537	273	margin	7.218277454	143
report	16.25655174	269	*plant*	4.326568127	128
radiation	6.185090065	241	health	7.950677872	118
reactor	4.403078079	201	significance	8.397342682	112
performance	4.489021301	178	equipment	5.6142869	109
function	6.415937424	164	limits	5.222682953	104
system	14.49797344	159	public	6.741711617	103
final	6.739749908	159	relief	6.547815323	102
updated	7.498948574	156	updated	7.671856403	99
plant	3.663259506	145	non	6.239801407	98
green	5.941387177	144	limit	5.714108944	93
significant	5.196485043	143	final	6.960159779	92
cornerstone	6.472372055	126	systems	5.471792698	91
level	4.913285255	123	valve	5.347458363	86

Table C.57: Top 20 Collocates with *Staff* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
NRC	6.544189453	1396	*NRC*	17.31108093	135
concludes	8.263023376	260	requests	8.256292343	33
reviewed	5.165700436	227	review	6.583432198	33
licensee	3.703950882	180	applicant	9.001605988	30
finds	8.346785545	172	*plant*	4.54342556	29
proposed	4.970340729	169	information	4.976429462	21
determined	5.579003334	146	*nuclear*	4.544268131	20
review	4.694126129	132	TSC	8.312269211	19
licensee's	4.645573616	130	requested	6.280069828	19
commission	5.183555603	128	security	5.834735394	19
information	4.437489033	120	personnel	5.566563129	18
members	6.986610413	108	approved	5.964177132	17
evaluation	4.566525459	107	additional	5.035171509	17
regulatory	4.813058376	103	members	8.10498333	16
therefore	5.597568035	98	support	5.603089333	16
also	4.534334183	98	will	3.365057945	16
nuclear	3.809732676	92	EOF	7.874370575	15
concluded	5.946612358	83	determined	5.353254318	15
requested	5.239929676	74	issued	7.158977985	14
approved	5.503235817	67	control	3.817016602	14

Table C.58: Top 20 Collocates with *System* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	15.57684231	238	*reactor*	5.445580959	417
ADAMS	17.41795731	235	coolant	6.723265171	230
component	17.65909386	191	control	4.74823761	200
water	16.04655075	167	*safety*	4.623831749	188
cooling	17.25740814	165	water	13.407938	169
safety	14.47971058	157	protection	6.166413784	153
NRC's	16.88664627	153	ventilation	7.309620857	148
coolant	17.47765732	144	*emergency*	5.115323544	148
document	16.63290787	143	cooling	5.939106941	141
management	16.43701553	140	*containment*	4.663704395	131
documents	16.21792221	129	*pressure*	15.35135365	130
access	17.20435524	125	component	6.021318913	122
core	16.21823502	125	monitoring	5.47191	110
control	15.08893585	123	will	3.227962255	109
pressure	15.80064201	116	core	4.383434296	108
RHR	17.88728142	114	service	5.050559998	101
emergency	15.78181458	114	instrumentation	5.623729706	92
protection	15.44844151	103	RCS	5.669476986	88
test	15.17464542	102	steam	4.492319107	87
power	23.45137215	101	feedwater	5.921393394	85

Table C.59: Top 20 Collocates with *Technical* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
specifications	19.20063972	825	specifications	9.408466339	581
specification	8.95724678	277	specification	9.175656319	502
changes	6.272074699	140	requirements	5.636200905	150
environmental	6.568846226	114	analysis	5.085806847	101
accordancde	5.715667248	107	proposed	5.072523117	95
contained	7.276761055	93	changes	5.112249374	84
evaluation	5.289436817	92	manual	6.684029102	72
appendix	5.855476856	88	standard	6.920795918	68
requirements	4.756288052	88	changes	4.780590534	63
protection	5.225331783	67	accordance	4.991877079	59
justification	7.587796688	54	*plant*	3.807337999	55
indicated	6.708856106	54	bases	6.571401596	53
review	4.290987968	52	required	4.093453884	49
support	5.990057945	51	improved	8.651968002	48
follows	7.609686375	46	surveillance	5.191095829	48
standard	15.00424194	45	regulatory	5.055621147	43
information	3.793342829	40	appendix	4.981275558	42
basis	4.665934563	39	station	4.030479908	42
branch	7.532018185	37	basis	4.864917755	41
read	8.27292347	36	support	5.229025364	39

Table C.60: Top 20 Collocates with *Unit* by Industry Group for Region III

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
power	14.34130192	191	station	6.426066875	436
plant	4.070561886	103	Byron	6.963387966	287
station	6.106397629	97	Braidwood	6.955502987	227
reactor	4.027776718	83	*power*	4.528746605	205
Byron	7.224184513	82	cycle	15.87064552	155
pump	5.322920799	80	*nuclear*	4.648684025	134
amendment	4.658629417	74	refueling	5.824388027	129
nuclear	3.822766542	65	outage	5.706311703	114
system	14.08580112	64	*reactor*	3.651068211	100
containment	5.015679836	64	*plant*	3.615554333	95
Braidwood	7.198894501	63	DNPS	6.003559589	80
train	6.991164207	59	mode	5.631546497	74
shutdown	5.093462467	59	one	4.248412609	71
refueling	5.684038639	58	event	3.737282276	67
outage	5.544427872	56	section	3.298910618	65
opposite	8.227819443	48	opposite	7.457016945	60
DG	6.922570705	48	FOL	6.706995487	60
generator	5.434437275	46	*operating*	3.78280592	59
water	3.991111755	41	*fuel*	3.547165632	59
trip	5.385361195	40	two	4.198654175	58

Table C.61: Top 20 Collocates with *Containment* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
system	4.90657568	68	***pressure***	6.362335682	214
spray	8.660821915	67	primary	7.53520298	147
isolation	7.529016972	55	spray	8.205292702	136
inside	8.733522415	49	isolation	7.7622962	131
fuel	15.6226387	48	failure	6.2242136	102
cooling	6.475255966	43	sump	7.863909721	100
reactor	4.570639133	43	purge	9.379924774	99
sump	8.572011948	42	***system***	4.504462242	99
primary	7.000201225	38	valves	6.222248077	92
outside	7.76331377	37	atmosphere	8.475502014	78
pressure	5.146467209	37	building	6.699384689	73
building	7.21859169	35	air	6.12869978	69
systems	4.929785252	33	***fuel***	4.639600754	59
integrity	6.557243824	31	temperature	5.256542206	56
leakage	5.893840313	31	leakage	5.094522476	56
atmosphere	9.403803825	29	outside	7.006466389	55
removal	6.768215179	28	inside	7.458312511	52
heat	6.186819553	28	valve	5.033122063	46
within	5.04736948	28	pump	4.942037106	46
air	6.485450268	26	hydrogen	7.482684135	40

Table C.62: Top 20 Collocates with *Emergency* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
diesel	8.048662186	216	reponse	6.69875288	258
preparedness	9.309370995	186	director	8.398600578	138
generator	7.269251823	175	*system*	4.724468708	136
plan	7.170099735	162	plan	6.445501328	108
response	6.060055256	145	cooling	6.475666523	95
system	4.856570721	110	site	6.1736269	86
action	15.89368725	97	personnel	5.7116642	86
cooling	6.662641525	82	preparedness	9.141824722	85
core	6.275852203	70	classification	8.321804047	81
planning	8.041387558	68	operations	5.767886162	79
organization	8.330300331	60	diesel	7.544095039	75
procedure	5.273704529	58	facility	6.340127945	73
level	5.194020271	55	core	5.1442976	73
procedures	5.069280148	49	organization	8.205955505	69
systems	4.75619936	49	manager	5.847744942	68
plans	6.352508068	47	procedures	5.982113838	63
levels	6.895225048	44	general	6.872364521	56
room	5.004853249	42	area	5.190169334	55
B	14.65132618	41	procedure	5.383916378	53
power	3.668332815	41	action	4.564396381	48

Table C.63: Top 20 Collocates with *Evaluation* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
safety	5.572608471	232	*safety*	16.23316383	180
NRC	3.677660465	86	model	7.352457047	139
report	5.346886635	76	*NRC*	4.550887585	66
staff	4.267739296	74	performed	5.029236317	57
licensee’s	5.403157711	64	report	5.044075966	52
technical	4.974030018	63	engineering	6.648441315	47
cause	6.360477924	60	risk	5.502450943	42
problem	6.779468536	52	results	4.600930214	37
staff’s	7.021543503	51	proposed	3.893313885	35
root	7.187104702	49	based	4.443248272	32
regulatory	4.79392767	48	environmental	5.433412075	31
licensee	3.165507078	48	ECCS	6.303905487	30
based	5.041213512	45	*plant*	3.274792433	30
performed	4.881991863	44	described	5.156348228	27
review	4.158987522	38	specific	4.922551155	27
request	4.85161829	36	*technical*	3.893499374	27
section	3.542109489	35	reviewed	6.237179279	26
operability	5.698488712	32	determined	5.084198475	26
proposed	3.899926424	31	required	3.231688023	26
complete	6.168545723	30	models	6.917561531	25

Table C.64: Top 20 Collocates with *Fuel* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
spent	19.78253746	245	spent	9.333757401	158
pool	19.30579948	156	assemblies	8.550075531	149
storage	18.25132179	98	design	5.738741875	118
oil	18.4729023	93	handling	9.145610809	113
irradiated	19.6482296	81	pool	8.105718613	112
handling	18.81481361	74	rod	7.266396523	111
movement	19.50017738	68	rods	7.948185921	109
assemblies	19.48800278	59	cladding	7.887291431	107
diesel	16.61728668	59	building	6.958893299	88
reactor	14.59622955	54	storage	8.119018555	79
cooling	16.47387886	53	irradiated	9.172380447	75
containment	15.6226387	48	assembly	7.326031208	73
accident	16.03684807	42	core	5.269937038	68
core	15.90936661	40	movement	8.900341034	65
cycle	25.02004623	39	accident	5.210114956	63
cladding	19.08930779	39	***reactor***	4.34459734	62
material	15.35121822	31	oil	8.358049393	61
nucleaer	13.7642498	30	***containment***	4.639600754	59
damage	16.81783676	29	clad	8.321463585	44
one	14.53734207	29	optimized	7.889200211	43

Table C.65: Top 20 Collocates with *Inspection* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
scope	7.318282604	546	inservice	8.848594666	169
inspectors	5.762030125	432	program	5.744998932	107
NRC	4.135407448	320	interval	7.141860008	84
public	5.489840508	308	scope	7.64238739	72
program	5.067448616	247	visual	7.766969681	66
reviewed	5.495406628	224	tube	5.958766937	66
results	5.867640972	213	results	5.274833679	63
report	5.297667027	199	*NRC*	4.389850616	63
available	5.680175304	168	steam	5.315711975	59
inservice	7.36498642	165	performed	4.985066414	59
findings	15.13821602	141	will	3.643769503	56
reports	6.38933754	137	*inspectors*	7.95106411	54
period	5.759970188	132	generator	5.788431644	51
performed	5.018066883	131	requirements	4.259792805	51
electronically	7.701946735	127	report	4.804302216	47
procedure	4.940792561	113	methods	6.187956333	41
manual	5.97609663	111	period	5.263683796	40
Mr.	6.400645733	110	public	5.666003704	39
team	5.530678272	107	plan	5.325459957	35
chapter	6.25450325	100	ISI	14.6473875	32

Table C.66: Top 20 Collocates with *Inspectors* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reviewed	7.564365864	561	*inspection*	7.95106411	54
inspection	5.762030125	432	reviewed	10.06493092	53
scope	7.309832573	324	scope	9.652058601	39
also	5.776288509	132	observed	9.768992424	24
observed	7.743121624	131	performed	5.903508186	15
determined	5.831798553	130	determined	6.884469986	13
completed	6.839239597	128	issue	8.149264336	12
verified	7.481064796	123	also	6.188435078	12
licensee	3.694440126	112	resident	9.75308609	11
identified	4.525186062	110	verified	8.634691238	11
resident	7.722262859	103			
performed	5.236895561	91			
B	15.1181612	83			
presented	6.925498486	80			
documents	5.807137966	72			
plant	3.827198982	69			
one	4.732233524	66			
results	4.877388477	64			
noted	6.88751173	63			
licensee's	4.616589069	60			

Table C.67: Top 20 Collocates with *License* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
renewal	8.95443058	354	amendment	8.401260376	279
operating	7.116682529	329	*operating*	7.270231247	200
amendment	6.999708176	315	request	7.426169872	153
renewed	8.960062027	192	renewal	10.07667446	131
facility	6.86833334	191	application	6.922973156	73
condition	5.661869049	99	proposed	5.21984005	62
conditions	5.580774784	99	condition	5.805135727	58
application	6.008241177	89	facility	7.006289482	54
incorporated	7.756931782	87	station	5.659543991	38
request	5.490577221	72	NPF	8.710909843	37
hereby	7.704976559	70	page	5.777882099	35
scope	5.346794605	66	LAR	7.787471771	33
changes	5.044872284	64	condition	4.666713715	31
effective	7.063762188	61	change	4.197749138	28
amended	6.70706892	61	scope	6.822885513	27
review	4.407999516	58	within	4.426678658	26
attachment	6.221116543	56	*technical*	4.284080982	25
will	3.342766285	56	*unit*	4.088308334	25
accordingly	7.528796196	51	will	3.075806618	25
paragraph	6.989809513	49	incorporated	7.481083393	23

Table C.68: Top 20 Collocates with *Licensee* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
identified	5.03918314	286	event	6.170680523	50
failed	6.555049896	177	report	6.041243553	43
NRC	3.042150021	163	shall	5.718922138	31
stated	5.878607273	148	LER	8.575248718	24
staff	3.639863491	141	states	7.035373688	24
proposed	4.49628067	138	*NRC*	4.170153618	21
inspectors	3.732575178	115	proposed	4.283301353	19
submitted	5.719617367	103	identified	5.562898159	17
reviewed	4.254494667	103	will	3.108571053	15
determined	4.617341995	102	submit	8.119568825	14
performance	3.858218431	101	controlled	6.96202755	14
changes	4.477296352	99	submittal	6.781089306	14
provided	4.707030296	97	use	4.562076569	13
requested	5.663827896	95	may	4.244469166	13
management	4.555002213	86	*staff*	5.393044949	12
also	4.293644428	86	performance	4.74899292	12
performed	4.273977757	85	changes	3.974649191	12
personnel	4.836914063	81	SCE	8.384622574	11
violations	6.234750748	80	findings	6.917496204	11
findings	14.18238735	79	stated	6.405995846	11

Table C.69: Top 20 Collocates with *NRC* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
staff	6.427371979	1223	approved	7.479149342	265
public	5.109238148	323	***staff***	7.314955711	217
site	6.590615749	320	approval	7.311460018	153
inspection	4.135407448	320	letter	5.997541904	125
regulatory	5.531295776	296	regulatory	5.866267204	121
commission	5.720342636	282	***nuclear***	5.205749989	104
nuclear	4.770206451	274	request	5.327690601	100
ADAMS	5.755630493	199	information	4.822319508	96
review	4.61639452	193	commission	6.640139103	94
concludes	6.865642548	188	review	5.584327698	86
web	7.40093565	166	provide	5.052132607	82
licensee	3.024338722	161	question	6.417900085	79
document	5.695516586	148	response	4.473601341	72
room	5.007771015	141	***evaluation***	4.550887585	66
reviewed	4.378396988	141	***safety***	13.75797558	64
staff's	6.570519924	138	***inspection***	4.389850616	63
enforcement	6.118118763	132	prior	5.154477119	58
policy	6.821984291	122	reference	4.382337093	58
identified	3.444980621	119	submitted	6.135538578	57
approved	5.956083775	114	order	5.651286602	57

Table C.70: Top 20 Collocates with *Nuclear* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
commission	7.989355087	549	*power*	6.214261532	272
regulatory	7.706559181	540	station	7.670371532	217
power	6.560866356	412	regulatory	7.536886215	195
station	7.460842133	294	commission	8.637557983	190
NRC	4.764931202	273	*plant*	5.55654335	146
plants	7.898231506	179	*NRC*	5.205749989	104
plant	5.292343616	176	*unit*	5.571012497	99
generating	7.709260464	148	generating	7.996902943	92
unit	5.434827328	126	plants	8.20407486	90
staff	4.456344604	126	*operating*	5.532969952	85
operating	5.419201851	118	Cooper	9.51185894	70
Palo	7.876638412	113	one	5.425431252	70
Arkansas	8.556189537	112	Arkansas	9.356799126	69
Verde	7.916521549	112	*safety*	14.49501514	54
one	5.596492767	111	Onofre	9.231114388	52
Energy	6.780737877	110	San	7.885978699	52
Institute	8.465672493	109	units	6.525115013	49
units	6.573566437	92	*reactor*	4.220134258	44
special	7.889780998	90	special	7.478635311	39
reactor	4.423489571	88	ANO	6.127684593	39

Table C.71: Top 20 Collocates with *Operating* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
license	7.116682529	329	*license*	7.27742672	201
facility	7.23653841	195	experience	8.552274704	148
experience	8.787059784	136	cycle	7.016033649	119
cycle	16.47993279	134	limits	6.212324142	107
renewed	8.779492378	134	*plant*	4.744302273	107
nuclear	5.419201851	118	conditions	5.546277523	104
conditions	5.765455246	89	*nuclear*	5.532969952	85
normal	7.570568562	85	core	5.528742313	81
corporation	8.988536835	70	report	5.263971806	78
licensee	3.623895645	67	*power*	4.0108428	76
subsequent	7.401459217	58	normal	6.398270607	66
power	4.212209225	55	facility	6.362002373	63
licenses	8.34855175	54	*system*	3.848705292	63
plant	4.117316246	53	procedures	6.146143913	60
procedures	5.218495369	50	licenses	8.841907501	58
Wolf	7.584727287	47	*unit*	4.410662651	57
Creek	7.550579071	47	based	4.663144112	48
C	13.84010792	43	*emergency*	4.048838139	46
industry	6.596933842	43	mode	5.17211771	44
one	4.784825325	43	NPF	8.061041832	43

Table C.72: Top 20 Collocates with *Plant* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
power	5.334533215	195	*power*	4.851545334	230
nuclear	5.292343616	176	operation	5.634160995	184
specific	6.582210064	159	specific	6.529552937	179
performance	5.326217175	157	conditions	5.564711094	178
conditions	5.244286537	101	*nuclear*	5.55654335	146
procedures	5.297257423	86	*safety*	14.53824711	121
safety	3.414507389	86	operations	5.848555088	120
licensing	6.212880135	82	*operating*	4.757722855	108
modifications	7.328947067	81	Callaway	7.490486622	107
operation	5.059610844	77	systems	5.364127159	93
Callaway	7.63619566	76	manager	5.647232056	85
operations	5.735114098	73	personnel	5.137998104	83
inspectors	3.827198982	69	shutdown	14.90478611	78
equipment	5.021398067	67	design	4.193119526	68
risk	4.809425354	63	Canyon	7.734861374	65
reviewed	4.282226563	59	Diablo	7.68755579	65
Diablo	7.358724117	58	*system*	3.068761587	62
areas	5.325003147	58	normal	5.503907204	60
response	4.154549599	58	required	3.267302752	58
Canyon	7.353261948	57	status	5.748370647	57

Table C.73: Top 20 Collocates with *Power* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
nuclear	6.560866356	412	*nuclear*	6.214261532	272
offsite	7.934194565	244	*plant*	4.851545334	230
plant	5.334533215	195	loss	6.545637131	216
plants	7.666258335	164	*reactor*	4.954627991	188
loss	6.559293747	156	offsite	15.90302563	184
public	5.063502312	136	distribution	7.370902061	168
reactor	4.665703297	112	full	6.779551029	160
operation	5.548523426	105	thermal	7.164782524	134
reactors	15.39550591	100	level	5.340842724	129
district	8.751649857	94	range	6.454651833	128
percent	6.426249504	83	core	5.157896519	125
electric	6.712121487	81	*system*	3.817159653	123
electrical	7.131063461	80	percent	6.208146095	115
full	7.107063293	79	supply	7.127400875	110
systems	4.90314579	79	operation	4.571118832	104
level	5.155938148	78	*unit*	4.267492771	103
commercial	7.895014286	74	plants	6.980914593	99
Southern	7.994506359	71	uprate	8.290927887	79
supply	7.588513851	71	electric	5.778768539	76
California	7.500517368	71	*operating*	4.0108428	76

Table C.74: Top 20 Collocates with *Pressure* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	6.026082039	156	boundary	8.467146873	241
boundary	8.897015572	135	*containment*	6.362335682	214
vessel	7.958415985	108	*reactor*	5.876363754	197
high	7.182285309	102	RCS	6.816329002	152
coolant	7.12153101	102	vessel	7.446073055	137
system	5.030127525	98	temperature	6.379815578	135
low	6.450032234	91	*system*	4.762341022	131
temperature	7.097928047	80	coolant	6.407385349	97
injection	8.139216423	77	low	6.541498184	95
crevice	9.113724709	76	differential	8.845811844	94
code	5.806429386	76	high	6.406414032	82
contact	7.186522007	55	pressurizer	6.868740559	66
safety	3.466616392	47	leakage	5.163347721	65
boiler	9.542833328	46	testing	5.063812256	61
ASME	5.897799969	46	ASME	5.506755829	58
control	4.65869379	46	injection	6.696796417	56
RCS	6.877906799	45	design	4.501369476	55
test	5.173092842	45	relief	5.719857693	52
head	6.761285305	37	test	4.562400341	52
containment	5.146467209	37	*safety*	13.90616226	51

Table C.75: Top 20 Collocates with *Reactor* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
coolant	8.43454361	439	coolant	8.567242622	505
system	5.832231998	296	*system*	5.93900156	345
vessel	8.299633026	237	vessel	8.351673126	299
pressure	6.016804218	155	trip	7.728710651	298
core	6.716505051	130	*pressure*	5.876363754	197
safety	4.050125599	122	*power*	4.954627991	188
trip	7.387431622	119	core	5.671505928	115
process	5.636072159	112	building	6.897890568	108
power	4.665703297	112	head	6.820723057	98
pump	5.95980072	111	water	5.227198601	98
water	5.370920658	107	RCS	5.902620792	94
oversight	7.511193275	105	pump	5.096312523	66
head	7.431618214	102	protection	5.388876915	63
nuclear	4.423489571	88	*fuel*	4.34459734	62
operator	5.851989269	85	boundary	6.264625072	61
senior	7.274526596	80	flow	4.746581554	60
boundary	7.027560711	64	instrumentation	5.714576244	44
building	6.799865246	60	*nuclear*	4.220134258	44
sources	7.091992855	58	RPS	7.88212347	42
described	5.957856178	58	temperature	4.474947929	42

Table C.76: Top 20 Collocates with *Safety* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
significance	6.500029564	566	related	17.76001358	257
low	6.971998215	402	analysis	16.17639732	250
public	5.728932858	360	function	16.94548798	180
health	7.721116066	317	*evaluation*	16.23316383	180
related	6.506940365	287	report	15.97882271	167
report	5.741157055	268	*system*	14.77337646	156
analysis	5.605906963	241	health	18.16398048	143
radiation	5.788058758	230	significance	17.97232819	140
evaluation	5.560117245	230	margin	17.52216339	140
function	6.41753149	188	public	16.78128815	133
final	6.818454266	170	injection	17.62577629	126
system	4.003580093	148	*plant*	14.53824711	121
injection	7.411009789	143	systems	15.95326519	108
level	5.142803192	129	final	17.42822266	95
quality	6.322691917	128	analyses	16.23173141	95
reactor	4.050125599	122	level	15.41867828	89
updated	7.331511974	116	quality	16.90109253	80
green	5.690761566	116	limits	15.27723026	73
occupational	7.406226158	111	valve	15.23095322	69
significant	5.096920967	108	valves	15.39901733	68

Table C.77: Top 20 Collocates with *Staff* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
NRC	6.437950134	1232	*NRC*	7.314955711	217
concludes	8.151367188	262	approved	6.413743496	31
reviewed	5.488679409	174	review	5.943196774	27
review	5.030499935	147	applicable	5.404106617	21
commission	5.442964077	133	commission	6.363724232	19
licensee	3.522680044	130	support	5.392093182	18
nuclear	4.456344604	126	*unit*	4.159675121	18
regulatory	5.02355051	119	*plant*	3.584738016	18
finds	8.150764465	111	will	3.147173882	18
proposed	4.65991354	111	finds	10.12079716	17
members	6.762255192	105	members	7.432741165	17
determined	5.109158516	103	manager	5.494142056	17
will	3.339782476	92	EOF	6.711406231	16
licensee’s	4.732316494	85	TSC	6.368683338	15
evaluation	4.325072765	77	provide	4.631717682	15
information	3.944280148	76	proposed	3.717832565	15
applicant	5.195482731	71	prepared	7.584349632	14
also	4.3895154	66	requested	5.820789814	14
requests	6.270192146	65	*nuclear*	4.34291172	14
complete	5.909343243	53	control	3.596986532	14

Table C.78: Top 20 Collocates with *System* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
reactor	5.832231998	296	*reactor*	5.93900156	345
coolant	7.195391178	233	coolant	6.816596985	216
ADAMS	6.905806541	226	*safety*	14.77337646	156
water	5.607972622	158	cooling	6.461413383	148
safety	4.003580093	148	control	4.618666649	140
component	6.704408169	143	component	6.568592548	138
cooling	6.625427723	137	*emergency*	4.724468708	136
management	5.853673458	136	*pressure*	4.762341022	131
document	6.475052834	130	*power*	3.817159653	123
documents	6.352686405	123	water	4.945075512	116
NRC's	6.489715576	122	protection	5.731693268	115
access	6.952554703	113	RCS	5.495063782	102
accession	6.935456753	111	*containment*	4.504462242	99
emergency	4.856570721	110	core	4.604403019	79
pressure	5.030127525	98	operation	4.188642502	74
control	4.509084225	90	instrumentation	5.879498005	71
protection	4.692545414	76	monitoring	4.977384567	68
power	3.662369728	70	steam	4.296022892	65
containment	4.90657568	68	*operating*	3.848705292	63
service	14.5604744	66	feedwater	5.982738972	62

Table C.79: Top 20 Collocates with *Technical* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
specifications	9.29033947	644	specification	9.016384125	223
specification	9.023083687	386	specifications	9.282249451	217
environmental	7.55279398	171	requirements	5.284270287	79
changes	6.071500301	112	support	6.363199234	25
appendix	6.307002068	104	changes	5.274921894	63
protection	5.881561756	101	bases	6.502027512	41
accordance	5.810997009	96	proposed	4.500277042	45
contained	7.195870876	81	center	7.272910595	14
requirements	4.667758942	76	change	4.491761208	46
review	4.978744984	74	requirement	6.191076279	34
violation	5.890915871	65	required	4.086922169	39
evaluation	4.974030018	63	manual	6.469024658	24
required	4.616922379	62	task	8.20687294	30
plan	5.55920887	53	force	7.900289059	30
information	4.307241917	51	standard	6.259020329	36
indicated	6.666348457	49	accordance	4.746238232	37
surveillance	5.689950466	48	surveillance	4.803591251	27
TSS	6.885798931	45	analysis	3.85584259	33
basis	4.964891911	45	revise	6.611651897	32
bases	14.72107697	43	condition	4.223608494	26

Table C.80: Top 20 Collocates with *Unit* by Industry Group for Region IV

NRC-Word	NRC-Relation	NRC-Freq	Licensee-Word	Licensee-Relation	Licensee-Freq
station	6.938313961	147	steam	5.8988657	121
nuclear	5.434827328	126	outage	6.641635418	113
steam	6.212512493	89	station	6.157430172	111
Verde	7.609006405	65	*power*	4.267492771	103
one	5.301917076	65	refueling	6.641283512	99
Arkansas	8.180502892	62	*nuclear*	5.571012497	99
power	4.159193039	56	mode	5.905802727	83
generator	5.64045763	55	one	5.12538147	83
plant	4.038304806	53	CPSES	7.987870693	78
refueling	6.242385387	51	B	4.14867878	70
electric	6.599309444	50	model	5.686526775	64
outage	5.938313961	49	Arkansas	8.585111618	59
model	6.591607094	46	Creek	7.282548428	59
Waterford	7.666813374	45	electric	6.17858839	57
generating	6.469132423	45	*operating*	4.410662651	57
Palo	6.960196495	43	Wolf	7.372319698	56
April	5.841452599	42	*plant*	3.602272034	55
B	14.69238377	41	Waterford	6.250905037	52
CPSES	7.825818062	39	operated	7.001676083	49
reactor	3.689457417	38	cycle	5.462854862	46