Thesis for the Degree of Doctor of Philosophy

A Corpus-based Language Network Analysis of Near-synonyms in a Specialized Corpus

by

Wenyu Lu

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

English Language and Literature

Korea Maritime and Ocean University

July 2017

Approval

Name: Wenyu Lu
Degree: Doctor of Philosophy
Title of thesis: A Corpus-based Language Network Analysis of Near-synonyms in a Specialized Corpus

Examining Committee:

Chair:

Dr. Kim, Jae-Hoon
Chair, Professor of Computer and Information Engineering
Korea Maritime and Ocean University

Dr. Jhang, Se-Eun
Senior Supervisor, Professor of English Language and Literature
Korea Maritime and Ocean University

Dr. Kong, Gil-Young
Internal Examiner, Professor of Navigation Science
Korea Maritime and Ocean University

Dr. Lee, Soo-Sang
External Examiner, Professor of Library, Archives Information Studies
Pusan National University

Dr. Deng, Yaochen
External Examiner, Professor of English
Dalian Maritime University

Date Approved: May 25, 2017

A Corpus-based Language Network Analysis of Near-synonyms in a Specialized Corpus

Wenyu Lu

Department of English Language and Literature

Graduate School of Korea Maritime and Ocean University

Abstract

As the international medium of communication for seafarers throughout the world, the importance of English has long been recognized in the maritime industry. Many studies have been conducted on Maritime English teaching and learning, nevertheless, although there are many near-synonyms existing in the language, few studies have been conducted on near-synonyms used in the maritime industry.

The objective of this study is to answer the following three questions. First, what are the differences and similarities between different near-synonyms in English? Second, can collocation network analysis provide a new perspective to explain the distinctions of near-synonyms from a micro-scopic level? Third, is semantic domain network analysis useful to distinguish one near-synonym from the other at the macro-scopic level? In pursuit of these research questions, I first illustrated how the idea of incorporating collocates in corpus linguistics, Maritime English,

near-synonyms, semantic domains and language network was studied. Then important concepts such as Maritime English, English for Specific Purposes, corpus linguistics, synonymy, collocation, semantic domains and language network analysis were introduced. Third, I compiled a 2.5 million word specialized Maritime English Corpus and proposed a new method of tagging English multi-word compounds, discussing the comparison of with and without multi-word compounds with regard to tokens, types, STTR and mean word length. Fourth, I examined collocates of five groups of near-synonyms, i.e., *ship* vs. *vessel*, *maritime* vs. *marine*, *ocean* vs. *sea*, *safety* vs. *security*, and *harbor* vs. *port*, drawing data through WordSmith 6.0, tagging semantic domains in Wmatrix 3.0, and conducting network analyses using NetMiner 4.0. In the final stage, from the results and discussions, I was able to answer the research questions. First, maritime near-synonyms generally show clear preference to specific collocates. Due to the specialty of Maritime English, general definitions are not helpful for the distinction between near-synonyms, therefore a new perspective is needed to view the behaviors of maritime words. Second, as a special visualization method, collocation network analysis can provide learners with a direct vision of the relationships between words. Compared with traditional collocation tables, learners are able to more quickly identify the collocates and find the relationship between several node words. In addition, it is much easier for learners to find the collocates exclusive to a specific word, thereby helping them to understand the meaning specific to that word. Third, if the collocation network shows learners relationships

of words, the semantic domain network is able to offer guidance cognitively: when a person has a specific word, how he can process it in his mind and therefore find the more appropriate synonym to collocate with. Main semantic domain network analysis shows us the exclusive domains to a certain near-synonym, and therefore defines the concepts exclusive to that near-synonym: furthermore, main semantic domain network analysis and sub-semantic domain network analysis together are able to tell us how near-synonyms show preference or tendency for one synonym rather than another, even when they have shared semantic domains.

The options in identifying relationships of near-synonyms can be presented through the classic metaphor of "the forest and the trees." Generally speaking, we see only the vein of a tree leaf through the traditional way of sentence-level analysis. We see the full leaf through collocation network analysis. We see the tree, even the whole forest, through semantic domain network analysis.

Acknowledgment

My deepest gratitude goes first and foremost to Prof. Se-Eun Jhang, my supervisor, for his constant support of my Ph.D. study and academic research, and also for his patience, motivation, encouragement, and professional knowledge. His guidance has walked me through all the stages of the research and writing of this thesis. His wisdom has led me through the difficulties not only in the academic life but also in my personal life. Without his illuminating instruction and valuable suggestions, the completion of this thesis would be impossible. I could not have imagined having a better supervisor and mentor for my Ph.D. study.

Additionally, I would like to thank the rest of my thesis committee: Prof Jae-Hoon Kim, Prof. Gil-Young Kong, Prof. Soo-Sang Lee and Prof. Yaochen Deng, not only for their insightful comments and encouragement, but also for the enlightening questions which incentivized me to broaden my research through various perspectives. I am also grateful to Prof. Kevin Parent and Prof. Mi-lim Ryoo, who have inspired me with enormous knowledge in linguistics. In addition, I deeply appreciate the contribution to this thesis made in various ways by my fellow students and friends, especially Dr. Sung-min Lee who has given precious support to conduct this research.

Moreover, my sincere gratitude goes to leaders and colleagues at School of Foreign Languages, Dalian Maritime University. They have provided me with this precious opportunity to study at Korea Maritime and Ocean University and kindly

shared my heavy teaching tasks. Any progress that I have made is the result of their profound concern and selfless devotion. Without their help, I would be unable to finish my Ph.D. in such a short period.

Last but not the least, I would like to thank my family: my father Tangyao Lu, my mother Lina Han and my husband Lianbo Li. Each time when I meet difficulties, I will think about my father's words. It is he who taught me to be a kind, responsible and honest person. My mother is the one who pushes me to go forward, who comforts me when I feel frustrated and depressed, and who is always there to support me. In those years, it is she who walked me through all the laughs and tears. My husband not only supports me spiritually and financially, but also encourages me to pursue my dreams. He understands and trusts me as a soulmate, cherishes and honors me as a friend, comforts and protects me as a guardian angel, just as he said in our wedding oath. I dedicate all my success to my beloved family.

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Focus of Inquiry

English is recognized as the international medium of communication for seafarers throughout the world (Ziarati, Ziarati, Calbas, B., & Moussly, 2008). In fact, “a ship which would formerly have been manned by a crew of around 40 sharing the same language could now be run by 17, with 17 different mother tongues” (Johnson, 1995, p. 127). An issue of increasing importance to the international maritime community over the last two decades has been the quality of Maritime English training for non-English-speaking maritime personnel and the lack of unified standards of English testing (Shen, 2010) in maritime training establishments. As pointed out by Bocanegra-Valle (2012, pp. 3580-3582), Maritime English terminology and phraseology pose a real challenge due to their specialization and unfamiliarity: Passengers sleep in *cabins* and meals are cooked in the *galley*; a ship does not have *walls* but *bulkheads*; smoke comes out of a *funnel* not a *chimney*; objects are not positioned *left* or *right* but on the *port* or *starboard* sides. In the particular case of multilingual crews, a proficient command of the English language in general, and of Maritime English in particular, is vitally important to overcome language barriers which can cause accidents. It has been estimated that this is the case in approximately 30-40% (Trenkner, 2000) or even

42% (Hetherington, Flin, & Mearns, 2006) of shipping accidents reported, where an insufficient command of English led to misunderstanding. Therefore, studies on the Maritime English field are very important.

According to several studies, English tends to have a larger number of words, if not the largest, than many other languages (Crystal, 2007). Some of the words have been borrowed from other languages (Finegan, 2007). This leads to the fact that many near-synonyms exist in the language. Many researchers have shown their interest in the distinction between near-synonyms such as *forest* and *woods* (Room, 1981), *foe* and *enemy* (Gove, 1984), *task* and *job* (Hirst, 1995). Taylor (2002) examined the near-synonyms *tall* and *high* by using an acceptability rating task and argued that the differences between these two adjectives were able to be captured using Vantage Theory. Inkpen and Hirst (2006) used an unsupervised decision-list algorithm to distinguish near-synonyms listed in the dictionary of *Choose the Right Word* (Hayakawa, 1994). Xiao and McEnery (2006) explored the collocational behavior and semantic prosody of near-synonyms from a cross-linguistic perspective. More recently, Webb and Kagimoto (2011) studied the interactions affected by the number of collocates, the position of the node word, and synonymy in learning collocations. That study indicated that learning through collocations with a node word might be a better way, but they also mentioned that synonymy had a negative effect on learning. Many studies have been conducted on the near-synonyms in general English texts; however, few studies of these were on the near-synonyms used in the maritime industry. Therefore, it would be interesting to see

how they behave in a maritime specialized context.

Language choices matter. Each time we choose a word in speaking or writing, we are influenced by the words we have just uttered or written, and also by the words we are planning to speak or write next. There is certainly some truth in the expression “one word led to the next.” Language is subject to conditions imposed by other cognitive subsystems or computational principles as well as memory limitations (particularly during language acquisition). This kind of “one word led to the next” phenomenon is known as the “collocation of a node” in corpus linguistics. Collocation of the node, as the fundamental interacting unit, is very important partly because it is very common in linguistic theorizing, but also because it is relatively straightforward to obtain sufficient corpus data to be statistically significant. In this study I will focus on the collocations of near-synonyms in the maritime industry.

With regard to language network analysis, collocation networks deal with linguistic features because the notion of collocation contributes to shaping the meaning of words (Firth, 1957) and also forming lexico-grammar (Halliday, 1961). Network analysis has been used in recent research in various fields, particularly biology and physics, which provides tools for characterizing statistical properties or complex structures. The insights from this thesis work as a piece of evidence for recent data-driven statistical approaches to natural language. It will also be helpful for studying fundamental unsolved puzzles in cognitive science.

It is very difficult to deal with all the collocates of node words because the number of collocates could be several hundred or thousand words. However, visualization can help to some extent. Williams (2002) advocated the advantages of visualization on collocational studies, claiming that the widened network could provide the semantic environment of the word and also helped to differentiate between potential synonyms. Therefore, in this study visualization of collocation network is conducted. In addition, semantic domains of collocates have a stronger tendency to connect pairs of near-synonyms than simple collocates. It will be helpful for beginning learners of L1 and L2 and lexicographers to understand the differences and similarities between near-synonyms through visualization of semantic domain networks.

This thesis examines collocates of five groups of near-synonyms, i.e., *ship* vs. *vessel*, *maritime* vs. *marine*, *sea* vs. *ocean*, *safety* vs. *security*, and *port* vs. *harbor*, drawing data from the self-built specialized Maritime English Corpus (MEC) through WordSmith 6.0, tagging semantic domains in Wmatrix 3.0 and conducting network analyses using NetMiner 4.0. The objective of the work done in this study is to answer the following three questions. First, what are the differences and similarities between different near-synonyms in English? Second, can collocation network analysis provide a new perspective to explain the distinctions of near-synonyms from a microscopic level? Third, is semantic domain network analysis useful to distinguish one near-synonym from the other at the macroscopic level?

1.2 Outline of the Thesis

In pursuit of the research questions, this thesis is organized as follows. Chapter 1 introduces the focus of inquiry, and outlines the thesis in terms of how the idea of incorporating collocate in corpus linguistics, Maritime English, near-synonyms semantic domains, and language network is studied.

With a brief synopsis in Section 2.1, Chapter 2 explores five important ideas involved in this thesis. Firstly, the relationship among Maritime English, English for Specific Purposes (ESP) and corpus linguistics is illustrated in Section 2.2. Then the main concepts of synonymy, collocation, language network and semantic domain analysis are addressed in detail. Synonymy in Section 2.3 is illustrated from the perspectives of lexicography, philosophy and linguistics. Then, five ways of categorizing synonym proposed by other scholars are brought about, followed by criteria for synonymy differentiation. At last, the near-synonyms mentioned in this thesis are defined in the field of corpus linguistics. As one of several main discussions in this thesis, collocation is defined under theoretical linguistics, lexicography, psychology and corpus linguistics in Section 2.4. Specifically, explanation of the differences between collocation and colligation is also addressed in this section. In Section 2.5, I provide the definition of language network analysis along with some basic concepts for network analysis. An overall review of the previous language network analysis papers is also given in order to provide explanatory power for the descriptions of Maritime English to follow. Studying

semantic domains will provide more insights into the concept of language network analysis. Semantic domain network analysis is also the highlight of this thesis. Therefore, in Section 2.6, concepts of semantic domains are brought forward, followed by previous studies on semantic domain analysis.

Chapter 3 discusses the definition and characteristics of a corpus from the aspects of corpus-driven vs. corpus-based research, specialized corpora for specialized discourse, and how to compile a corpus focusing on representativeness, balance, and size. I propose a new method of tagging English multi-word compounds, discussing the comparison of with and without multi-word compounds with regard to tokens, types, STTR and mean word length. The result shows that a corpus which goes through the process of "compounding" is more suitable for analysis. In addition, Chapter 3 presents the research tools and methodology for collocates extraction, network visualization, semantic tagging and process of data analysis. With the guidance of this chapter, the network visualization will be conducted in subsequent Chapter 4 and Chapter 5.

Chapter 4 develops my convergence idea incorporating corpus linguistics and language networks in order to provide answers to my three research questions. It proposes collocation network analysis in order to give further explanatory power to the corpus descriptions for the better understanding of Maritime English. With brief introduction of how traditionally people distinguish near-synonyms in dictionary, such as *ship* vs. *vessel*, *maritime* vs. *marine*, *sea* vs. *ocean*, *safety* vs. *security*, and

port vs. *harbor*, I will show the results of collocation network visualization of these near-synonyms.

In Chapter 5, I propose a new way of viewing the collocates of near-synonyms. That is the semantic domain network analysis, which will be proved to be an effective way to show the differences among near-synonyms.

Chapter 6 summarizes my findings in this study, and points out limitations and implications.

Chapter 2. Literature Review

2.1 A Brief Synopsis

This chapter explores five important ideas involved in this thesis. Firstly, the relationship among Maritime English, English for Specific Purposes (ESP) and corpus linguistics is illustrated in Section 2.2. Then, main concepts of synonymy, collocation, language network analysis and semantic domain analysis are addressed in detail.

Synonymy in Section 2.3 is firstly illustrated from the perspectives of lexicography, philosophy and linguistics. Then, five ways of categorizing synonyms by degree are presented, followed by criteria for synonymy differentiation. At last, the near-synonym is defined from the perspective of corpus linguistics. In this thesis, near-synonym means pairs that have very similar cognitive or denotational meanings, but which may differ in collocational or prosodic behaviors (Xiao & McEnery, 2006, p. 108).

As a way to show the context where near-synonyms exist, collocation is another important concept in this thesis. Collocation is defined from the standpoints of theoretical linguistics, lexicography, corpus linguistics and psychology in Section 2.4. Specifically, an explanation of the differences between collocation and colligation is addressed in this section.

In Section 2.5, the definition and classification, along with some basic concepts

of language network analysis, is provided. An overall review of previous language network analysis papers is also given in this section in order to provide explanatory power for descriptions of near-synonyms in Maritime English.

Studying semantic domains will provide more insights into the concept of language network analysis. Semantic domain network analysis is also the highlight of this thesis. Therefore, in Section 2.6, concepts of semantic domains are presented, followed by previous studies on semantic domain analysis.

2.2 Maritime English as an English for Specific Purposes (ESP)

2.2.1 What is ESP?

ESP has gradually developed into a new field of study since the 1960s. According to Hutchinson and Waters (1987, p. 19), three factors were included in the emergence of ESP, i.e., the demands of a *Brave New World* (Huxley, 1932), a revolution in linguistics, and the focus on the learner. They believed that ESP is an independent subject, hence it does not belong to a particular kind of methodology, teaching materials or language. Strevens (1988) regarded ESP as a special case, which is contained in a general scope of specific-purpose language teaching. Later, Robinson (1991) indicated that ESP is based on two criteria: one is “goal-directed” and another is “need analysis.”

2.2.2 Maritime English as ESP

From ancient times, until the emergence of modern national states, maritime commerce had been largely restrained to coastal areas. In the 15th to the 17th centuries, with the advent of the great maritime era, western countries began developing their navigation technology. The corresponding shipping trade, a global phenomenon that evolved with a globalized commerce across countries, started to boom. Entering the 18th and 19th centuries, an internationally-accepted language was badly needed to solve language barriers in the maritime field across various countries. In response, English, considered a lingua franca for mariners, has started to gain worldwide recognition as the international language of seaborne trade in the maritime industry.

The importance of Maritime English has been well recognized globally not only because it is the only designated common language for maritime operations, but also because over 70% of maritime accidents have been attributed to communication problems, as reported by the International Maritime Organization (IMO). Maritime English is defined as an official language within the international maritime community, contributing to the safety of navigation and the facilitation of seaborne trade (Trenker, 2009). At present, there exists undisputed international agreement on English as the required common language of seafaring and port operations.

Maritime English terminology and phraseology pose a real challenge due to its

specialization and unfamiliarity. There are at least three reasons for its appearance: (i) the need for clear communication between ships and shores, between crew members, and between crew and passengers; (ii) the development of a set of terms that refer to the parts of ships, and the procedures involved in sailing; and (iii) the globalization of the shipping industry. Particularly important is Resolution A.380 (X) adopted by the IMO General Assembly in 1977, in virtue of which English was recognized as "a common language for international communications between ships and between ships and shore services."

According to Bocanegra-Valle (2012, pp. 3579-3580), as an umbrella term, Maritime English refers to the English language used by seafarers both at sea and in port, and by individuals working in the shipping and shipbuilding industry. It subsumes five different sub-varieties according to the specific purpose they serve within the maritime context: (i) English for navigation and maritime communications, (ii) English for maritime commerce, (iii) English for maritime law, (iv) English for marine engineering, and (v) English for shipbuilding. To summarize, Maritime English refers to the specialized English language used at sea and in port by mariners. It is one of the branches of ESP, and is used as an official language within the international maritime community.

2.2.3 ESP and Corpus Linguistics

In corpus linguistics, specialized English can assist in recognition of language

used in specific areas, which is very different from general English (Lu, Lee, & Jhang, 2017). According to Johns (2013, p. 5) the origin of ESP corpus-based research can be dated back to the 1960s when the central focus of ESP research was English for science and technology (EST) in academic contexts. The research at that time was mostly descriptive, involving few statistical grammatical counts within written discourse. In the 1990s, along with the development of corpus linguistics, researchers started to pay special attention to different subfields, particularly specific written academic registers. Many corpora have been built for ESP research with the booming of computer science since the 20th century (Lu, Lee, & Jhang, 2017).

2.3 Synonymy

As an international language, English has absorbed many loans from other countries through constant invasion and cultural integration; therefore it developed into a language filled with synonyms. Synonyms are used to reflect the nuance between different objects, express sophisticated emotions, or merely to avoid redundancy in text (spoken or written). Synonymy in a language can be viewed as a basic concept in lexicology. When meaning relations of words are studied, most researchers are inclined to prioritize the concept of synonyms in their investigation (Harley, 2006). Etymologically speaking, the term *synonymy* originates from the Greek word *sunonumon*, meaning “having the same name” (Jackson & Amvela,

2000, p. 92).

English itself has many different synonyms. Two concepts need to be clarified here before entering further discussion: (i) If the meaning of word A is similar to word B, word A is called the synonym of word B, and vice versa; (ii) Synonymy refers to the semantic paradigmatic relationship of synonyms. The lexical relation which parallels identity in the membership of two classes is synonymy (Cruse, 1986, p. 88). In this section, the definition of synonymy (synonym) from different perspectives is offered. Then the classification of synonymy on the basis of granularity is provided. At last, the topic of this thesis, near-synonym, will be discussed from the standpoint of corpus linguistics.

2.3.1 Definition of Synonymy

Defining “synonymy” has proven to be very difficult. If one takes the view that synonyms are words that have the exact same meaning, it can be argued that synonyms do not and even cannot exist in any language since words tend to exist with somewhat “contrastiveness” which eliminates “exactness.” But if one relaxes this view to include words that are merely similar in meaning, then almost any two words can be considered synonyms at some level of granularity. As a fairly complicated language phenomenon, many scholars in different fields have tried to give “synonymy” a widely acceptable definition.

Lexicographers clearly take the view that synonymy is matter of likeness,

disagreeing only in how broad the definition ought to be. The breadth of the definition depends on the purpose of the dictionary. A broad definition is better for word finding, for example, in a hierarchically structured dictionary like *Roget's Thesaurus* (Roget, 1911). A narrow definition is better for discrimination, for example, in *Merriam-Webster's New Dictionary of Synonyms* (Gove, 1984), "a synonym, in this dictionary, will always mean one of two or more words in the English language which have the same or very nearly the same essential meaning. Two or more words which are synonyms can be defined in the same terms up to a certain point" (Gove, 1984, p. 24).

Philosophers treat synonymy as one important aspect of "truth value" and people are skeptical about the existence of synonyms. Goodman (1952) proposed to substitute "sameness of meaning" with "likeness of meaning" in the definition of synonym since "sameness" is an unqualified notion. Quine (1961, p. 22) defined synonym as the primary business of the theory of meaning in which "intersubstitutability" is an important concept. Generally speaking, limited by the "truth value," the concept of synonym is too wide for philosophers to disclose any intrinsic meaning.

Theoretical linguists consider synonymy in terms of necessary resemblance and permissible differences (significant degree of semantic overlap) and contextually, by means of diagnostic frames. Cruse (1986, p. 266) believed that synonyms must not only manifest a high degree of semantic overlap, but also have a low degree of

implicit contractiveness. He proposed that "a word is said to be a synonym of another word in the same language if one or more of its senses bears a sufficiently close similarity to one or more of the senses of the other word" (Cruse, 2006, p. 174). Clark (1992) believed that the meanings of any two different words are in an opposite position. Hence synonyms will either disappear or develop a new meaning to show the difference.

2.3.2 Synonymy as a Matter of Degree

Within the class of synonyms, some pairs of items are more synonymous than others. Palmer believed there are two kinds of synonyms. "Total synonyms are fully exchangeable in all situations, while loose synonyms are words that are close in meaning or have their meanings overlap" (Palmer, 1981, p. 91). Jackson agreed and stated that "words are strict synonyms if they can be used interchangeably in all sentence contexts, while loose synonyms are pairs of words that can be substituted for each other in a wide range of contexts but not necessarily" (Jackson 1988, p. 65). Lyons (1981, p. 50) proposed a different classification. He defined absolute synonyms as "expressions that are *fully*[1], *totally*[2] and *completely*[3] synonymous", partial synonyms as "expressions which satisfy at least one, but not all three, of the criteria" and near-synonyms are "more or less similar, but not

[1] Synonyms are fully synonymous if, and only if, all their meanings are identical.
[2] Synonyms are totally synonymous if, and only if, they are identical in all contexts.
[3] Synonyms are completely synonymous if, and only if they are identical on all (relevant) dimensions of meaning.

identical in meaning". Lyons (1981) strongly insisted that there be a clear distinction between near-synonyms and partial synonyms.

As an influential linguist in semantics, Cruse (1986, 2000) has offered two different forms of synonymy classification. According to his three degrees of synonymy, Cruse (1986, p. 266) classified synonymy into absolute synonymy, cognitive synonym and plesionyms. Later, in 2000, he classified synonymy into absolute synonymy, propositional synonymy, and near-synonymy (Cruse, 2000, p. 156). The definition of each sub-category is as follows:

(i) Absolute synonymy refers to complete identity of meaning. (e.g., *sofa*: *settee*; *pullover*: *sweater*).

(Cruse, 2000, p. 157)

(ii) Cognitive synonymy may be defined as follows: X is a cognitive synonym of Y if (i) X and Y are syntactically identical, and (ii) any grammatical declarative sentence S containing X has equivalent truth conditions to another sentence S', which is identical to S except that X is replaced by Y. (e.g., *fiddle*: *violin*).

(Cruse, 1986, p. 88)

Propositional synonymy can be defined in terms of entailment. If two lexical items are propositional synonyms, they can be substituted in any expression with truth-conditional properties without effect on those properties. Differences in the meanings of propositional synonyms, by definition, necessarily involve one or more aspects of non-propositional meaning, the most important being (i)

differences in expressive meaning, (ii) differences of stylistic level (on the colloquial-formal dimension), and (iii) differences in the presupposed field of discourse. Most usually, more than one of these comes into play at any one time. (e.g., *fiddle*: *violin*; *shin*: *fibular*).

(Cruse, 2000, p. 158)

(iii) Near-synonyms must share the same core meaning and must not have the primary function of contrasting with one another in their most typical contexts.

(Cruse, 2006, p. 174)

From the definition and examples provided by Cruse (above), it can be seen that he believed there are three types of synonyms, which are absolute synonyms, cognitive synonyms (i.e., propositional synonyms), and near-synonyms (i.e., plesionyms). Cognitive synonyms must be identical in respect of propositional traits, but they may differ in respect of expressive traits (Cruse, 1986, p. 273). He also mentioned that "complete identity of meaning (absolute synonymy) is very rarely, if ever, encountered" (Cruse, 2006, p. 174) since "natural languages abhor absolute synonyms just as nature abhors a vacuum" (Cruse, 1986, p. 270).

Table 2.1 is a generalization of various scholars' ideas of synonymy as a matter of degree. However, although the borderline between different categories is, at least in principle, clear, the boundaries between near-synonyms and non-synonyms are less straightforward.

Table 2.1
Classification of synonymy

Scholars	Classification		
Palmer (1981)	total synonyms	loose synonyms	
Jackson (1988)	strict synonyms	loose synonyms	
Lyons (1981)	absolute synonyms	partial synonyms	near-synonyms
Cruse (1986)	absolute synonymy	cognitive synonymy	plesionyms
Cruse (2000)	absolute synonymy	propositional synonymy	near-synonymy

2.3.3 Criteria for Synonymy Differentiation

Practically, synonyms can be differentiated in many ways. First is place of origin. Since English is widely used in many countries, one concept may be expressed differently across countries, such as *football* (BrE) and *soccer* (AmE), *lift* (BrE) and *elevator* (AmE). Even within any one country, different words can be used to refer to the same object, such as *armpit* in London dialect and *armhole* in Bradford dialect. Second is formality. In different contexts, conversation situations, or genres, slang, informal English and formal English can be used to refer to the same concept, such as *kick the bucket*, *die*, *pass away*, etc. Third is connotation. This third type of synonyms is the focus of this thesis. Examples of this type are numerous, such as *look*, *see*, *watch*, *read*, *view*, *glare*, *stare*, *glance*, *glimpse*, *observe*, *notice*, *spot*, *scan*, *skim*, *gaze*, etc.

2.3.4 Near-synonyms in Corpus Linguistics

Inkpen and Hirst (2006) defined synonyms as words that have almost the same meaning but are not fully intersubstitutable. They argued that synonyms can differ in the frequency of use, attitudes (pejorative, neutral, or favorable), and stylistics (formality, concreteness, force, floridity, or familiarity). Although it is a debatable topic in theoretical semantics and it appears that it might not be possible to precisely and rigorously define near-synonyms since the definition might depend to a large extent on how the near-synonyms are to be used, corpus linguistics probably does not need a fully rigorous definition.

Computational linguistics has developed tools and methods to distinguish near-synonyms. Taylor (2003) examined near-synonyms *tall* and *high* by using an acceptability rating task. Test subjects were required to find whether the use of *tall* and *high* was acceptable under different contexts. Taylor later argued that the differences between these two verbs can be captured using Vantage Theory.[4] Inkpen and Hirst (2006) presented a method to automatically acquire a new type of lexical resource: a knowledge-base of near-synonym differences was based on an unsupervised decision-list algorithm so that they were able to distinguish near-synonyms listed on the dictionary of *Choose the Right Word* (Hayakawa, 1994). Among all the computational tools, WordNet (Miller, 1995; Princeton University, 2010) is considered important and the most widely-used. WordNet is an online

[4] For more on Vantage Theory, see, for example, MacLaury, 2002.

lexical database of near-synonyms. It contains English nouns, verbs, adjectives, and adverbs organized into sets of synonyms, which represent a lexicalized concept.

Collocation, the way how words company each other, is one useful way to distinguish near-synonyms. Halliday and Hasan (1976, p. 73) observed that the adjective that often occurs with *tea* is *strong* but not *powerful,* whereas a *car* is more likely to be described as *powerful* than *strong*. Similarly, while *weak* and *feeble* have similar cognitive meanings, native speakers of English prefer to say *weak tea* rather than *feeble tea* (Mackin, 1978, p. 150). It is also noticed that the adjective *daunting* often collocates with *task* but not *job*.

In the 21st century, the introduction of large machine readable corpora and advanced searching engines like WordSmith has further helped the study in the distinction of near-synonyms. As mentioned by Xiao and McEnery (2006, p. 108), near-synonyms can also differ in semantic prosodies, e.g., *fickle* is negative, whereas *flexible* is positive. Moreover, in addition to the lexical level, near-synonyms at the morphological level also demonstrate different collocational behaviors. Webb and Kagimoto (2011) studied the effect of the number of collocates, the position of the node word, and the synonymy on learning collocations. The study showed that the number of collocates had a positive effect to the learning process of collocation, whereas the position of the node word did not show an obvious effect.

In this thesis, near-synonyms are defined as “lexical pairs that have very similar cognitive or denotational meanings, but which may differ in collocational or prosodic behaviors” (Xiao & McEnery, 2006, p. 108). As such, synonymous words are not collocationally interchangeable (Conzett, 1997, pp. 70-87; Tognini-Bonelli, 2001, p. 34). For example, Greenbaum (1974, as cited in Xiao & McEnery, 2006, p. 108) noted that synonyms may “be separated collocationally because of restrictions to a language variety or style”, as shown in his examples *to cashier an army officer* vs. *to expel a school child*. In this thesis, near-synonyms are considered at the word and phrase levels; the collocational behavior of several sets of near-synonyms in Maritime English are discussed.

2.4 Collocation

Collocation is, put crudely, the property of language whereby two or more words seem to appear frequently in each other’s company. The term has been used and explained in different ways.

2.4.1 Definition of Collocation

The term “collocation” was first brought forth by John Firth, widely accepted as the founder of the concept. Firth (1957, p. 181) observed that “collocations of a given word are statements of the habitual or customary places of that word.” Since

then, many scholars have defined collocation in many ways. Halliday and Hasan (1976, p. 287) referred to collocation as a cohesive device and described it as "a cover term for the kind of cohesion that results from the co-occurrence of lexical items that are in some way or other typically associated with one another, because they tend to occur in similar environments." Leech (1974, p. 20) believed that "collocative meaning consists of the associations a word acquires on account of the meanings of words which tend to occur in its environment." His opinion was supported by Cruse (1986, p. 40) who saw collocation as sequences of lexical items which occur together habitually.

In 1933, Harold Palmer's *Second Interim Report on English Collocations* highlighted the importance of collocation as a key to producing natural-sounding language. Thus from the 1940s onwards, many lexicographers began to place collocation usage in their dictionaries, and with the booming of the large readable corpora in the 21st century, many dictionaries such as the *Macmillan English Dictionary for Advanced Learners* (Macmillan Publishers Ltd., 2002) and the *Longman Dictionary of Contemporary English (5th Edition)* (Pearson Longman, 2008) included boxes or panels with lists of frequent collocations.

2.4.2 Collocation in Corpus Linguistics

Generally speaking, functional linguistics concentrates on general abstraction about the properties of phrases and sentences. In contrast, corpus linguistics

emphasizes the importance of context. In this study, under the corpus linguistics orientation, an analysis based on collocation is conducted.

2.4.2.1 Definition of Collocation in Corpus Linguistics

There is general agreement that individual words and their co-occurrences contribute to shaping the meaning of words (Lewis, 1997; Nation, 2001; Schmitt, 2000; Sinclair, 1991). Sinclair, the founder of modern corpus linguistics, defined collocation in much more technical words. Collocation is regarded as "the co-occurrence of two items (node words) within a specified environment (a selected span)" (Sinclair, Jones, & Daley, 2004, p. 10). "The collocational pattern of an item consists of a list of all words appearing significantly often in its environment with information about them" (Sinclair et al., 2004, p. 73). Since then, using corpus linguistics in collocation research has become a trend. Lindquist (2009, p. 57) believed that collocation is the relation between a word and individual word-forms which co-occur frequently with it.

Thus as one of the key types of data that corpora can reveal, through collocation, the meaning of words and their behaviors can be derived. Collocation is concerned with repeated, statistically significant patterns, where words occur together more often than chance alone might dictate. The statistical relationships are calculated on measures of frequency and probability, and various calculations can be used, depending on the type of relationships being explored.

2.4.2.2 Collocation vs. Colligation

The difference between “collocation” and “colligation” was first coined by Firth (1957) who believed that “you shall know a word by the company it keeps” (p. 11). Firth (1957, p. 183) saw colligation as “the interrelation of grammatical categories in syntactical structure” and collocation as “actual words in habitual company”. Thus, two terms interrelated with each other were introduced. Later, theoretical linguists proposed many different theories to distinguish collocation with colligation. For example, Nattinger and DeCarrico (1992) considered colligation as general classes of collocations, for which at least one construction is specified by category rather than as a distinct lexical item. Tognini-Bonelli (1996, p. 74) defined colligation as interrelations of grammatical categories, which concern categories such as word classes and sentence classes. And with the introduction of corpus linguistics, Sinclair (1996) saw form and meaning as complementary: different senses of a word will characteristically be realized in different structural configurations. According to Sinclair (1996, pp. 80-88), “collocation is a frequent co-occurrence of words; it does not have any profound effect on the individual meanings of the words, but there is usually at least a slight effect on the meaning, if only to select or confirm the meaning appropriate to the collocation, while colligation is the occurrence of grammatical choices to account for the greater variation.” Stubb (2009) further generalized that collocation suggests predictability of word combinations and colligation is one step more abstract than collocation, dealing with the predication of the grammatical classes in the syntactical level.

2.4.3 Lexical Priming of Collocation in Psychology

The notion of collocation means not only statistically significant co-occurrence with node words (Sinclair et al., 2004), but also psychological reality (Hoey, 2005). Hoey (2005, p. 7) brought about the idea of "the pervasiveness of collocation" and borrowed the concept of "priming" as discussed in the literature of psychology (e.g., Anderson, 1983; Neely, 1977; Neely, 1991) where the notion of semantic priming was used to discuss the way a "priming" word may provoke a particular "target" word. Chomsky (1986) distinguished the study of linguistic data, which he termed "E-Language" (externalized language) from "I-Language" (internalized language), the language found in the brains of speakers. Lexical priming was intended as a bridge between the two categories.

As for the synonym, Hoey (2005, p.13) specifically discussed a hypothesis that "co-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations." He also pointed out that corpus linguistics can provide a way of analyzing language. A listener will recognize a word more quickly when a related word is given (i.e., *body* and *heart*). Therefore, psychologically speaking, collocation can work as a way to distinguish near-synonyms.

2.5 Language Network Analysis

2.5.1 Definition

The development of human beings is the result of information exchange on both intellectual and emotional levels. All people, objects or relationships are intertwined in different types of networks; there can be no single absolutely independent item. Network analysis can contribute powerfully in evaluating relationships among abstract elements, people, and knowledge (Barabasi, 2002; Christakis & Fowler, 2010; Newman, 2001; Scott, 2000; Watts & Strogatz, 1998). Therefore, many studies have been conducted based on networks, such as ecological webs (Montoya & Sole, 2002), genre identification (Stevanak, Larue, & Carr, 2010), Web query analysis (Saha Roy, Ganguly, Choudhury, & Kumar Singh, 2011), semantic analysis (Biemann, Roos, & Weihe, 2012), opinion mining (Amancio, Fabbri, Oliveira, Nunes, & da Fontoura Costa, 2011) software maps (Valverde, Ferrer-Cancho, & Sole, 2002), genomes (Sole & Pastor-Satorras, 2002), brain networks (Sporns, Chialvo, Kaiser, & Hilgetag, 2004; Eguiluz, Cecchi, Chialvo, Baliki, & Apkarian, 2005) or Internet architectures (Albert, Jeong, & Barabasi, 1999). Language is clearly an example of a complex dynamic system. It exhibits highly complicated network structures at all levels (phonetic, lexical, syntactic, semantic, discourse) and the network shaped and reshaped by millions of language users over long periods of time, as they adapt and change them to their needs as part of ongoing local interactions. How people conceive and use words

highly depends on the language acquisition process which happens in the brain and the cognitive process to be produced, parsed, and interpreted by highly complex cognitive networks as well.

2.5.2 Classification

Sole, Corominas, Valverde, & Steels (2005, p. 4) presented three kinds of language networks in their analysis. The first type of network constructions is co-occurrence networks in which two words are linked if they appear together within at least one sentence. Such graphs can be undirected or directed. The second type of network constructions is syntactic networks, which are built up based upon constituent structures, where units form higher level structures which in turn then behave as units in other structures. The third type of network constructions is semantic networks, which can be built starting from individual words that lexicalize concepts and then mapping out basic semantic relations such as part-whole or binary opposition.

Figure 2.1 is a summary of language networks proposed by Sole, Corominas, Valverde, & Steels (2005, p. 8).

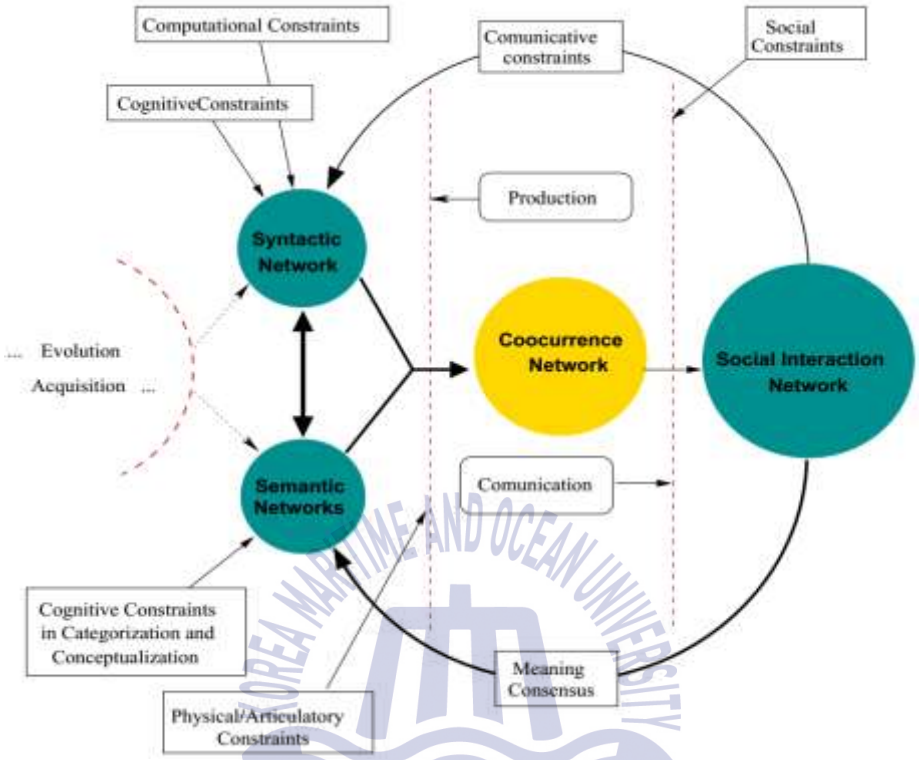
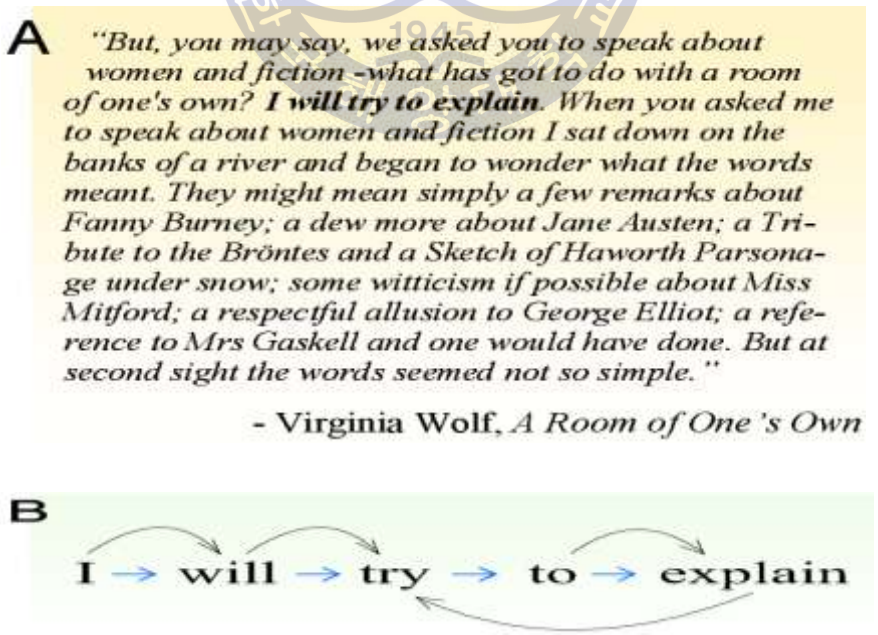


Figure 2.1 Relationships of language networks.
From Sole, Corominas, Valverde, & Steels, 2005, p. 8.

Among all the three types of language networks (excluding the social interaction network), co-occurrence network, i.e., collocation network is my focus for this study. Collocation networks can be considered as the fundamental type of network not only because it is common in language, but also because it is based on a meaning expression phase. Word collocation networks (Ke, 2007), also known as collocation graphs (Choudhury & Mukherjee, 2009; Heyer, Lauter, Quasthoff, Wittig, & Wolff, 2001), are networks of words found in a document or a document

collection, where each node corresponds to a unique word type, and edges correspond to word collocations (Ke & Yao, 2008). As a method originating in corpus linguistics, a tool for visualizing sequences of texts, a collocation network can provide a two-dimensional image of the most central words in a text and the connections between them (Magnusson & Vanharanta, 2003, p. 276). It offers clear visualization to quickly discover the most significant relationships between words. In this thesis, word co-occurrence and collocation network are considered since it is, as Sole et al. (2005, p. 3) put it, "relatively straightforward to obtain sufficient corpus data to be statistically significant, and because several large scale projects are under way for manual annotation of text based on lexical entries, such as WordNet (Fellbaum, 1987) and Framenet (Baker, Fillmore, & Lowe, 1998)." A good example of a word collocation network is shown in Figure 2.2.



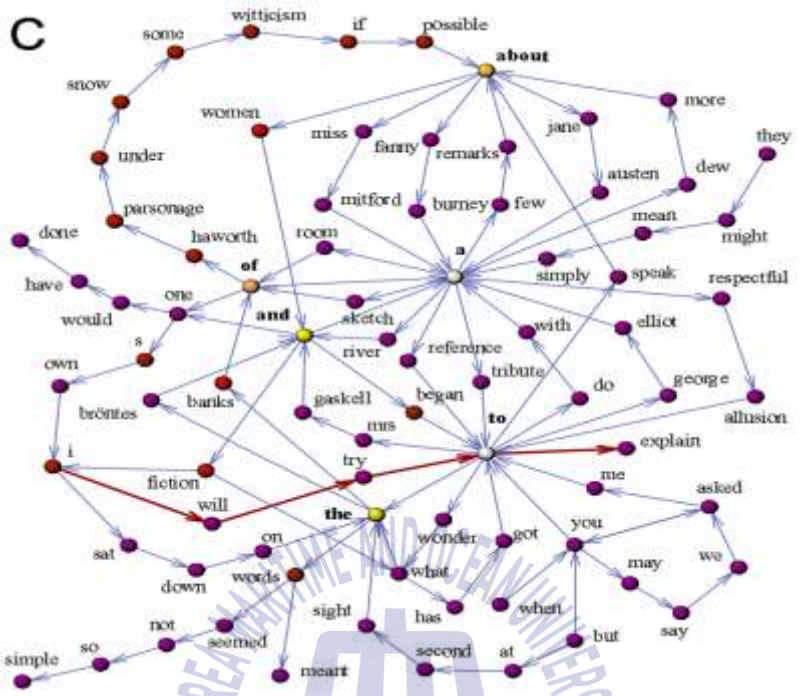


Figure 2.2 A collocation network.
From Sole, Corominas, Valverde, & Steels, 2005, p. 2.

From Figure 2.2, it can be seen that in A, a sentence exists in a paragraph. Its meaning is understandable. However there is no easy way for readers to know how these words are constructed to show the meaning. In B, syntactic analysis is given to show how words are connected in a sentence, but it is still a liner structure. However, in C, it can be easily seen how words are connected in the whole paragraph within a network.

2.5.3 Basic Concepts

The idea of converging corpus linguistics and language networks can be traced back to the study of Jones (1971). In her keyword retrieval study Jones (1971, p. 56) discussed four types of links between keyword nodes, “strings”, “stars”, “cliques”, and “clumps”, as shown in Figure 2.3.

Figure 2.3 Four types of link between keywords.
From Jones, 1971, p. 56.

Based on Jones’ work, Scott and Tribble (2006) hypothesized that words could be redrawn as a network of connections by showing Jones’ formal pattern of keyword linkage, as shown in Figure 2.4.

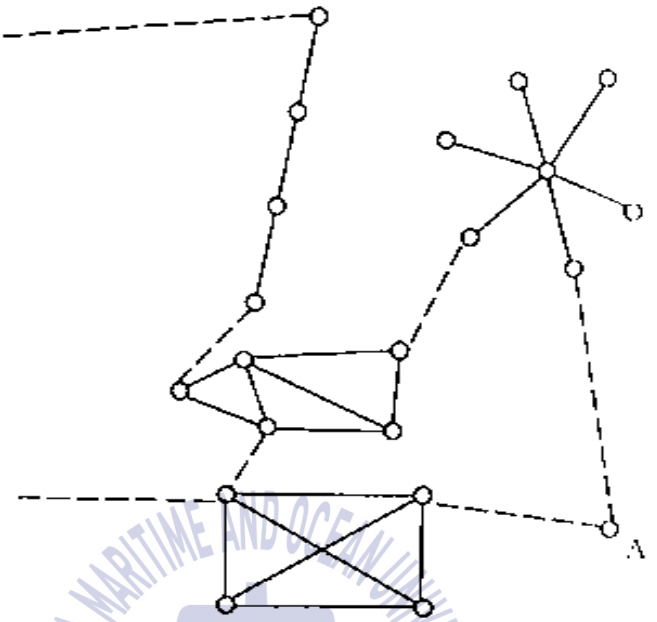


Figure 2.4 Keyword linkage network.
From Scott and Tribble, 2006.

To help understand network analysis, the following definitions of network concepts are necessary, as outlined in Nooy, Mrvar, and Batagelj (2011). The following definitions are borrowed in their book *Exploratory Social Network Analysis with Pajek*. These definitions will be used in Chapter 4 and Chapter 5 to describe the framework of network analysis. Notice that "network", "node", and "link" are in accordance with computer science terminology. These terms are called "graph", "vertex", and "edge" in mathematics.

a. A "graph" is a set of vertices and a set of lines between pairs of vertices.

b. A "simple undirected graph" contains neither multiple edges nor loops.

c. The “degree of a node” is the number of lines incident with it.

d. In a “one-mode network”, each vertex can be related to every other vertex.

e. In a “two-mode network”, vertices are divided into two sets and vertices can only be related to vertices in the other set.

f. “Geodesic” is the shortest path between two vertices.

g. The “betweenness” centrality of a vertex is the proportion of all geodesics between pairs of other vertices that include this vertex.

h. The “eigenvector centrality” of a vertex is the extent to which it is linked to vertices with high eigenvector centrality.

i. A “strong component” is a maximal strongly connected sub-network.

j. A “clique” is a maximal complete sub-network containing three vertices or more.

2.5.4 Previous Studies

In corpus linguistics, many researchers have analyzed corpus data relying on visualization techniques. These studies have demonstrated that collocation network is a useful method of exploring complex relationships between lexical items. The concept of collocation networks originated in an article by Williams (1998). In order to create specialized dictionaries, Williams (1998) explored collocation

networks to visualize a web of interlocking patterns of biology terms in a corpus of plant biology articles. In addition, Williams (2002) visualized DNA with the immediate statistical collocates which revealed the thematic environment of DNA as a technical word within a 560,000-word biology corpus. He explained this type of visualization in accordance with ESP vocabulary learning.

Following Williams, many scholars have applied collocation networks to various studies. McEnery (2006) employed a visualization technique that draws collocation networks automatically. Masucci and Rodgers (2006) studied network properties of Orwell's *1984*, as shown in Figure 2.5.

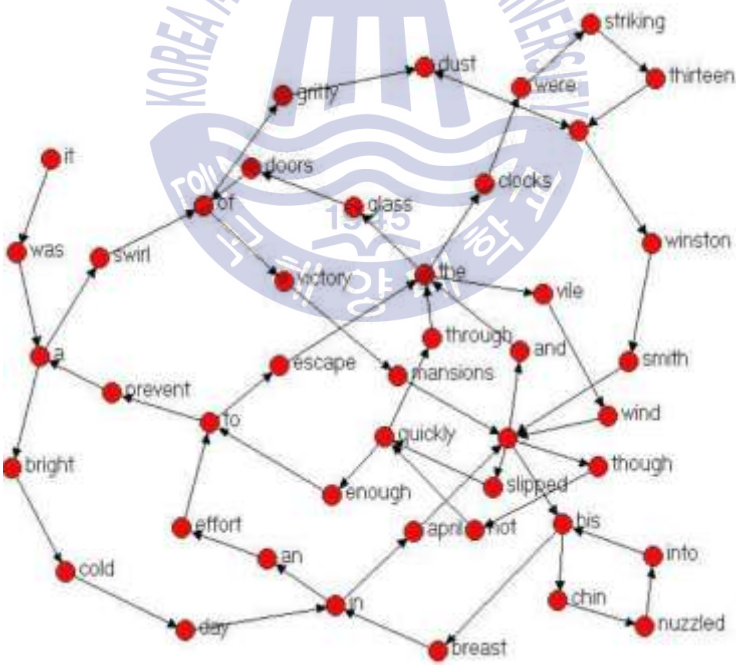


Figure 2.5 Language network for the first 60 words of Orwell's *1984*.
From Masucci and Rodgers, 2006.

Treating each word in the text as a node, they analyzed the properties of the nearest neighbors and clustering coefficients and found that they followed the characteristics of power law (Zipf's law, Pareto distribution).

Language network analysis was also applied to non-English languages. Zhou, Hu, Zhang, & Guan (2008) studied Chinese language networks from *The People's Daily* corpus by building two different networks based on different criteria to define link relations, as shown in Figure 2.6. Liang, Shi, Tse, Liu, Wang, & Cui (2009) compared word collocation networks of Chinese and English text, and pointed out their similarities and differences. They further constructed character collocation networks in Chinese, showed their small world structure, and used these networks in a follow-up study to accurately segregate Chinese essays from different literary periods.

Figure 2.6 Network analysis of Chinese sentences.
From Zhou et al., 2008.

Language network is useful on lexical level, but also proved to be helpful in discourse analysis. Stuart and Botella (2009) analyzed knowledge networks of specific science discourse communities. They analyzed keywords and clusters in terms of their distributions across text plots and discipline levels. Their results indicated that in science discourse university communities, people tended to share some keywords and clusters in writings. Stevanak et al. (2010) used collocation networks to distinguish between novels and news articles. Alonso, Millon, and Williams (2011) used the "E-Advanced Learner's Dictionary of Verbs in Science (DicSci)" for a vocabulary visualization. They adopted a corpus-driven approach to a new dictionary by using collocation networks, as shown in Figure 2.7.

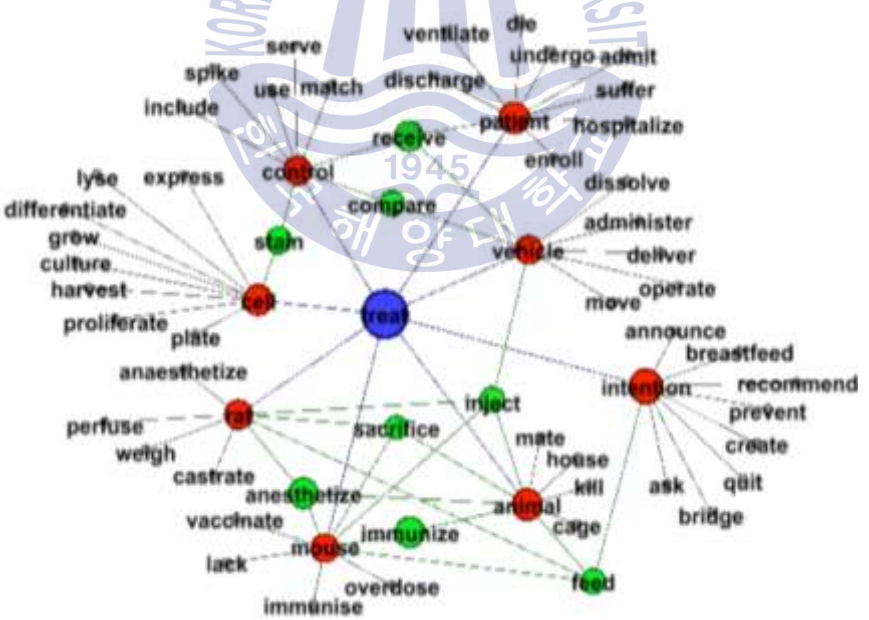


Figure 2.7 Collocation network of *treat*.
From Alonso, Millon, and Williams, 2011.

Biemann et al. (2012) used graph motifs on collocation networks to further distinguish real natural language text from generated natural language text, and to point out the shortcomings of n-gram language models. Moretti (2013) adopted a computational approach to deal with huge literary corpora by using z-score and social network analysis to investigate Shakespeare's plays and other novels. In his *Distant Reading*, he used network visualization among characters and argued for his findings from the character networks of *King Lear*, as shown in Figure 2.8.

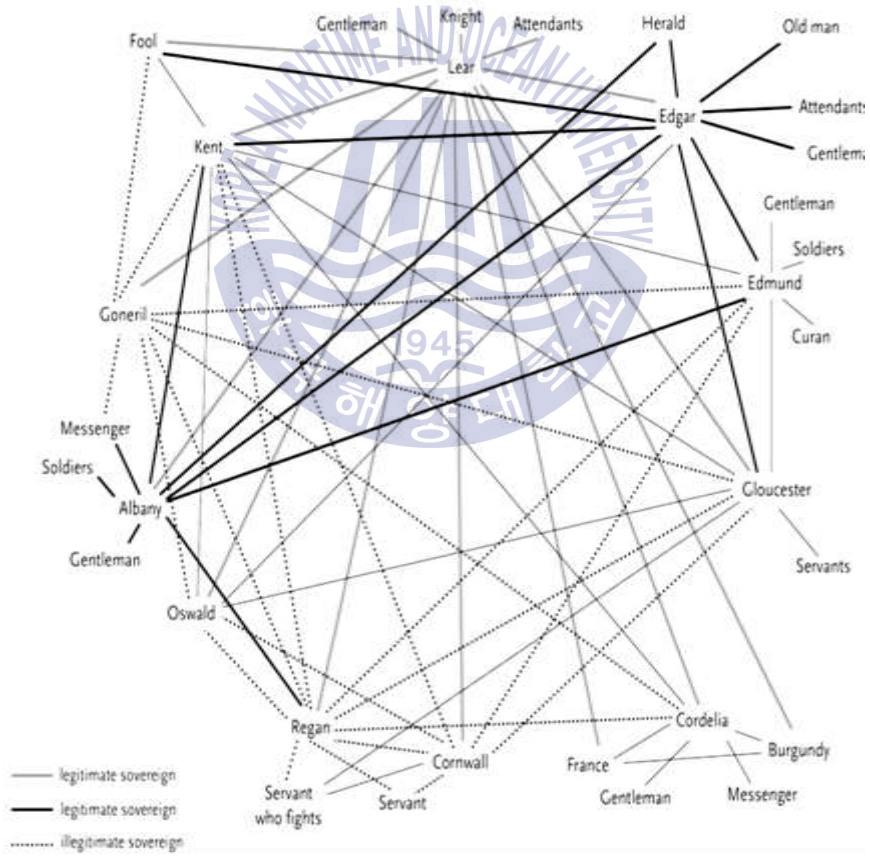


Figure 2.8 Character Network of *King Lear*.
From Moretti, 2013.

The most important contribution to visualization of collocation networks seems to be a study conducted by Brezina, McEnery, and Wattam (2015), including various options for operationalization of the network construction by developing GraphColl 1.0 software, which made a systematic analysis of collocation networks available, as shown in Figure 2.9. The study also offered different statistical measures for identifying collocations.

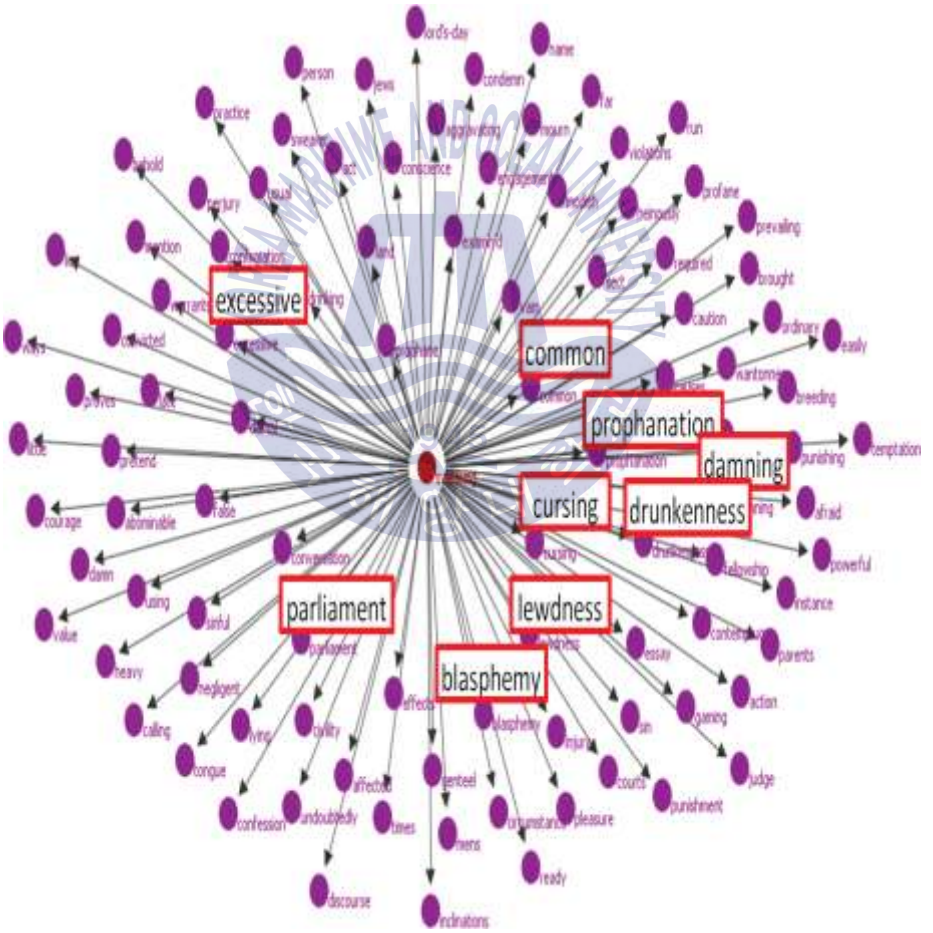


Figure 2.9 Collocates of *swearing*.
From Brezina, McEnery, and Wattam, 2015.

2.6 Semantic Domain Analysis

2.6.1 Concepts of Semantic Domains

Semantic domain has been used in many fields. Ottenheimer (2006, p. 18), a writer in Linguistic Anthropology, defined a semantic domain as a "specific area of cultural emphasis." In lexicography, a semantic domain or semantic field is defined as an area of meaning and the words used to talk about it. For instance, English has a semantic domain WIND, which includes words such as *storm*, *typhoon*, *hurricane*, *tornado*, *breeze*, etc. In the social sciences, the concept of semantic domain stemmed from the ideas of cognitive anthropology. The quest was originally to see how the words that groups of humans use to describe certain things are relative to the underlying perceptions and meanings that those groups share (Ottenheimer, 2006, p. 18).

2.6.2 Previous Studies on Semantic Domain Analysis

Phillips (1983) suggested that the empirical patterns of association discovered in the data should be viewed theoretically as "lexical networks." He argued that traditional linguistics could not adequately account for the concept of subject matter, as subject matter relied on regularities in the lexical organization of text. Phillips suggested what he called a knowledge-free analysis of the terms in a text. An analysis of this kind could presumably reveal systematic textual patterning, which in turn contributed to the semantic structure of the text and functions as a

basis for the emergence of the notion of content (Magnusson & Vanharanta, 2003, p. 277). Gries, and Dagmar (2009) and Gries and Otani (2010) proposed viewing synonyms from a cognitive semantic analysis. Recently, Jhang and Lee (2013) visualized collocation networks of near-synonyms of *maritime* vs. *marine* and *ship* vs. *vessel* in the Maritime English Corpus and explained their semantic relations. They focused on the strength of MI scores by presenting diverse thickness and marking numbers along with node words. Besides, Faber, León-Araúz, and Reimerink (2014) drew conceptual networks of *groyne* which showed semantic relations among words, as shown in Figure 2.10.

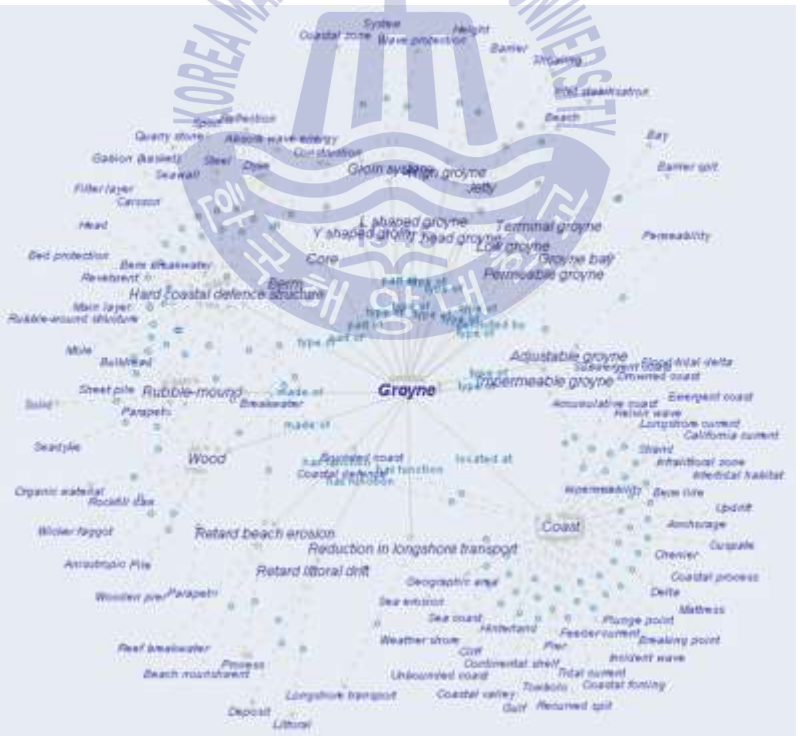


Figure 2.10 Conceptual map of *groyne*.
From Faber, León-Araúz, and Reimerink, 2014.

Chapter 3. Data and Methodology

3.1 Maritime English Corpus

3.1.1 What is a Corpus?

The origins of corpus construction can be dated back centuries, when biblical scholars and their teams indexed words in the Bible, along with citations of where and in which passages they occurred. Later, lexicographers such as Dr. Samuel Johnson, who first published a comprehensive dictionary of English in 1755, started to build a collection of word usages in a rented room. Without support from any academic institution, Johnson stored endless slips of paper logging samples of usage from the period 1560 to 1660 in the room. His years of efforts became a "paper corpus." Before it found its way into the linguistic terminology, the term "corpus" had long been in use to refer to a collection or binding together of written works of a similar nature. The modern idea of a linguistic corpus was developed mainly in the 1950s when theoretical linguists had the thought of compiling a collection of language usages so that both scholars and learners were able to easily extract data or samples from the corpus. And when computational work became popular in the 1960s, the modern written and spoken corpora started to emerge.

Biber, Conrad, and Reppen (1998) described corpus linguistics as having four main features: (i) it is an empirical (experiment-based) approach in which patterns of language use that are observed in real language texts (spoken and written) are

analyzed, (ii) it uses a representative sample of the target language stored as an electronic database (a corpus) as the basis for the analysis, (iii) it relies on computer software to count linguistics patterns as part of the analysis, and (iv) it depends on both quantitative and qualitative analytical techniques to interpret the findings. This definition is accepted by most corpus linguists, with few disagreements.

3.1.2 Characteristics of a Corpus

3.1.2.1 Corpus-driven vs. Corpus-based research

The term of “corpus-driven” and “corpus-based” was originally introduced by Tognini-Bonelli (2001). Corpus-based studies build or use a corpus to test a theory or hypothesis, while corpus-driven studies look at the data characteristics provided by the corpus and form a theory solely based on the corpus itself. Therefore, those who believe that corpus is just a method may conduct a corpus-based research. And those who claim that the corpus itself embodies its own theory of language usually prefer a corpus-driven research. This idea of corpus-driven is also closely associated with the work of scholars usually referred to as “neo-Firthians” (Tognini-Bonelli, 2001, pp. 84-85).

In this thesis, both corpus-driven and corpus-based analysis are conducted. First, based on collocates extracted for the near-synonyms, I am able to distinguish the

near-synonyms by observing significant collocates within a window span of 1-left and 1-right (1–1). In this corpus-driven research, then considering semantic domains, the data is put into network analysis to test whether semantic domain network analysis is helpful to distinguish near-synonyms, hence the corpus-based analysis.

3.1.2.2 Specialized Corpora for Specialized Discourse

Currently, there are two opposite trends in building a corpus. One is advocated by John Sinclair, the father of corpus linguistics, who claimed that “small is not beautiful; it is simply a limitation” (Sinclair, 2004, p. 189). One reason for Sinclair’s words is the fact that some linguistic features are so rare that a small-sized corpus is not able to provide enough instances that a linguistic feature will be a frequency of occurrence approaching zero. Therefore, many large, even mega corpora, began to be compiled, such as the 100 million word British National Corpus (BNC), the 400 million word Corpus of Historical American English (COHA), and the 520 million word Corpus of Contemporary American English (COCA), let alone the 1.9 billion word Corpus of Global Web-Based English (GloWbe) and the still increasingly-added corpus News on the Web (NOW), which currently includes more than 4.2 billion words. However, others have argued that a smaller corpus may be perfectly adequate for some purposes. As Koester (2010, p. 67) mentioned, the smaller, more specialized corpora have a distinct advantage;

they allow a much closer link between the corpus and the contexts in which the texts occur in the corpus. Whereas very large corpora, through their de-contextualization, give insights into lexico-grammatical patterns in the language as a whole, smaller specialized corpora give insights into patterns of language use in particular settings.

In ESP, this kind of specialized corpus is greatly in need since a large general corpus cannot "represent" the lexicon or phrase usage in a specific industry and cannot yield insights that are directly relevant for teaching and learning for specific purposes (Flowerdew, 2002). Researchers find it practical or necessary to create specialized corpora for the particular research questions they have, hence many specialized corpora were built in recent years such as the 1.8 million word *The Michigan Corpus of Academic Spoken English (MICASE)*. Therefore, I prefered to build a relatively smaller but specialized corpus, so that when collocates are extracted for near-synonyms, data is available for the analysis.

3.1.3 Maritime English Corpus (MEC)

Generally speaking, the Maritime English Corpus (MEC) is an opportunistic corpus (McEnery & Hardie, 2011, p. 11). Relevant materials have been included because of their availability. Given the time and resource constraints of the study, text collection has been limited to materials that are available in electronic digital form either from the internet or from publishers, a number of whom have allowed

their materials to be included. For the MEC, the key text types were first identified, after which texts were selected to reflect those areas, in order to “represent” the maritime discourse. The text selections were based on what Biber (1993, p. 243) described as “situational” or Sinclair (2004, p. 1) called “external” factors; that is, they were chosen for their communicative function in the discourse, not for “linguistic” or “internal” features.

3.1.3.1 Sampling of the MEC

In order to build a specialized MEC, I carefully selected the data involved, cleansed the data manually and ended up with an approximately 2.5 million word corpus. For the sampling of the corpus I divided it into two sub-corpora, arts, and sciences, each of which further stemmed from two sub-subcorpora, as shown in Figure 3.1.

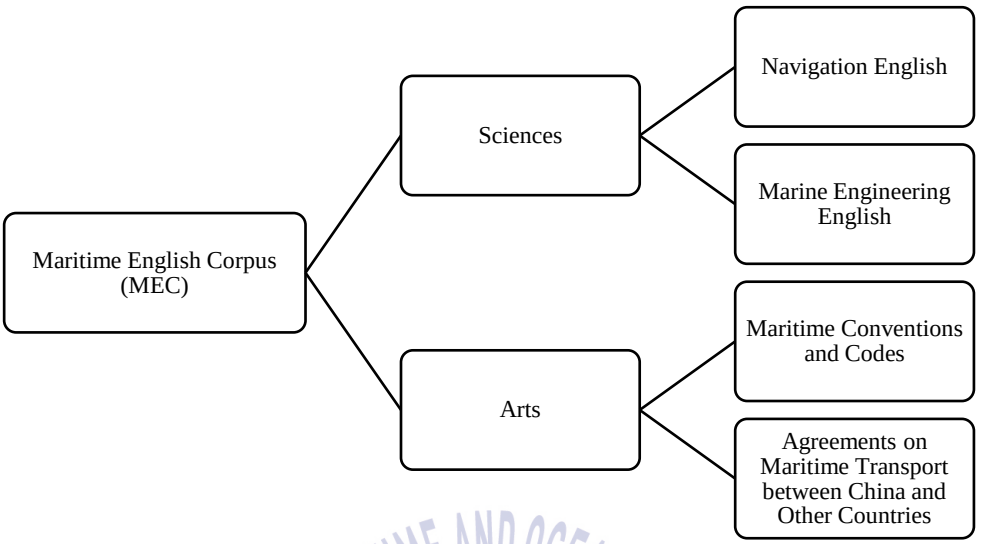


Figure 3.1 Sampling of Maritime English Corpus

In the sub-corpus "sciences", I included two genres of text. One is Marine Engineering English. Another is Navigation English. This way of classifying Maritime English is based on function, an approach accepted and used worldwide. Generally speaking, Navigation English is used "above" the deck, while Marine Engineering English is used "below" the deck. The data were extracted from major international maritime journals, as listed in Table 3.1 and English-written maritime textbooks, as listed in Table 3.2.

Table 3.1
List of maritime journals

TEXT ID	TITLES
A1	JOURNAL FOR MARITIME RESEARCH (2010-2015)
A2	MARITIME STUDIES (2012-2014)
A3	GYROSCOPY AND NAVIGATION (2010-2014)
A4	WMU JOURNAL OF MARINE AFFAIRS (2010-2015)
A5	THE JOURNAL OF NAVIGATION (2010-2012-2013)

Table 3.2
List of maritime textbooks

TEXT ID	TITLES
B1	COMMAND OF THE SEA THE HISTORY AND STRATEGY OF MARITIME EMPIRES, CLARK G. REYNOLDS, 1974
B2	USE OF ENGLISH FOR MARITIME STUDENTS, MERCEDES HERRERA ARNAIZ, 2014
B3	ENGLISH FOR THE MARITIME INDUSTRY: AN ENGLISH LANGUAGE COURSE BOOK FOR SEA FARERS, TONY GRICE, 2012
B4	THE ARTS OF THE SAILOR, HG SMITH, 2012

In the sub-corpus “arts”, I similarly included two genres of texts: one is maritime conventions and codes, the other is the agreements on maritime transport between China and other countries.

As a specialized agency of the United Nations, the International Maritime Organization (IMO) is the global standard-setting authority for the safety, security,

and environmental performance of international shipping. Its main role is to create a regulatory framework for the shipping industry that is fair and effective, universally adopted, and universally implemented (http://www.imo.org/en/Pages/Default.aspx). As an internationally important authority, IMO has promoted the adoption of many conventions and protocols, nearly all of which are now in force (Lu, Lee, & Jhang, 2017). Conventions and protocols are binding legal instruments, and upon entry into force their requirements must be implemented by all countries which are party to them. Therefore, in the MEC, conventions and protocols take important roles in the sampling frame. The list below shows the conventions and codes included in the MEC.

Table 3.3
List of maritime conventions and codes

TEXT ID	TITLES
C1	CONVENTION FOR THE UNIFICATION OF CERTAIN RULES OF LAW RELATION TO ASSISTANCE AND SALVAGE AT SEA, 1910
C2	INTERNATIONAL CONVENTION FOR THE UNIFICATION OF CERTAIN RULES OF LAW WITH RESPECT TO COLLISION BETWEEN VESSELS, 1910
C3	INTERNATIONAL CONVENTION AND STATUTE CONCERNING THE REGIME OF NAVIGABLE WATERWAYS OF INTERNATIONAL CONCERN, 1921
C4	CONVENTION AND STATUTE ON THE INTERNATIONAL REGIME OF MARITIME PORTS, 1923
C5	INTERNATIONAL CONVENTION FOR THE UNIFICATION OF CERTAIN RULES OF LAW RELATING TO BILLS OF LADING, 1924
C6	INTERNATIONAL CONVENTION FOR THE UNIFICATION OF CERTAIN RULES OF LAW RELATING TO MARITIME LIENS AND MORTGAGES, 1926
C7	INTERNATIONAL CONVENTION FOR THE UNIFICATION OF CERTAIN RULES CONCERNING THE IMMUNITY OF STATE-

	OWNED SHIPS, 1926
C8	CONVENTION ON THE INTERNATIONAL MARITIME ORGANIZATION, 1948
C9	INTERNATIONAL CONVENTION FOR THE PREVENTION OF POLLUTION OF THE SEA BY OIL, 1954
C10	INTERNATIONAL CONVENTION RELATING TO STOWAWAYS, 1957
C11	INTERNATIONAL CONVENTION RELATING TO THE LIMITATION OF THE LIABILITY OF OWNERS OF SEA-GOING SHIPS, 1957
C12	GENEVA CONVENTION ON TERRITORIAL SEA AND CONTIGUOUS ZONE, 1958
C13	GENEVA CONVENTION ON THE HIGH SEAS, 1958
C14	CONVENTION ON THE LIABILITY OF OPERATORS OF NUCLEAR SHIPS, 1962
C15	CONVENTION ON FACILITATION OF INTERNATIONAL MARITIME TRAFFIC, 1965
C16	CONVENTION RELATING TO REGISTRATION OF RIGHTS IN RESPECT OF VESSELS UNDER CONSTRUCTION, 1967
C17	INTERNATIONAL CONVENTION ON CIVIL LIABILITY FOR OIL POLLUTION DAMAGE, 1969
C18	INTERNATIONAL CONVENTION ON TONNAGE MEASUREMENT OF SHIPS, 1969
C19	INTERNATIONAL CONVENTION FOR PREVENTING COLLISIONS AT SEA, 1972
C20	UNIFORM RULES FOR A COMBINED TRANSPORT DOCUMENT, 1973
C21	INTERNATIONAL CONVENTION FOR THE PREVENTION OF POLLUTION FROM SHIPS, 1973
C22	PROTOCOL RELATING TO INTERVENTION ON THE HIGH SEAS IN CASES OF MARINE POLLUTION BY SUBSTANCES OTHER THAN OIL, 1973
C23	1996 PROTOCOL TO THE CONVENTION ON THE PREVENTION OF MARINE POLLUTION BY DUMPING OF WASTES AND OTHER MATTER, 1972
C24	INTERNATIONAL CONVENTION FOR THE SAFETY OF LIFE AT SEA, 1974
C25	YORK ANTWERP RULES, 1974
C26	PROTOCOL OF 2002 TO THE ATHENS CONVENTION RELATING TO THE CARRIAGE OF PASSENGERS AND THEIR LUGGAGE BY SEA, 1974
C27	TORREMOLINOS INTERNATIONAL CONVENTION FOR THE

	SAFETY OF FISHING VESSELS, 1977
C28	PROTOCOL OF 1996 TO AMEND THE CONVENTION ON LIMITATION OF LIABILITY FOR MARITIME CLAIMS, 1976
C29	INTERNATIONAL CONVENTION ON STANDARDS OF TRAINING, CERTIFICATION AND WATCHKEEPING FOR SEAFARERS, 1978
C30	UNITED NATIONS CONVENTION ON THE CARRIAGE OF GOODS BY SEA, 1978
C31	INTERNATIONAL CONVENTION ON STANDARDS OF TRAINING, CERTIFICATION AND WATCHKEEPING FOR SEAFARERS, 1978
C32	INTERNATIONAL CONVENTION ON MARITIME SEARCH AND RES CUE, 1979
C33	UNITED NATIONS CONVENTION ON INTERNATIONAL MULTIMODAL TRANSPORT OF GOODS, 1980
C34	COMMITTEE MARITIME INTERNATIONAL CONSTITUTION, 1981
C35	UNITED NATIONS CONVENTION ON THE LAW OF THE SEA, 1982
C36	UNITED NATIONS CONVENTION ON CONDITIONS FOR REGISTRATION OF SHIPS, 1986
C37	CONVENTION FOR THE SUPPRESSION OF UNLAWFUL ACTS AGAINST THE SAFETY OF MARITIME NAVIGATION, 1988
C38	UNITED NATIONS CONVENTION ON THE LIABILITY OF OPERATORS OF TRANSPORT TERMINALS IN INTERNATIONAL TRADE, 1991
C39	SEAFARERS' HOURS OF WORK AND THE MANNING OF SHIPS CONVENTION, 1996
C40	PROTOCOL ON PREPAREDNESS, RESPONSE AND CO-OPERATION TO POLLUTION INCIDENTS BY HAZARDOUS AND NOXIOUS SUBSTANCES, 2000
C41	INTERNATIONAL CONVENTION ON THE CONTROL OF HARMFUL ANTI-FOULING SYSTEMS ON SHIPS, 2001
C42	INTERNATIONAL CONVENTION ON CIVIL LIABILITY FOR BUNKER OIL POLLUTION DAMAGE, 2001
C43	INTERNATIONAL CONVENTION FOR THE CONTROL AND MANAGEMENT OF SHIP'S BALLAST WATER AND SEDIMENTS, 2004

In addition, 19 agreements on maritime transport between China and other countries were included as another source of sub-corpus arts in MEC, including those with France, Greece, Indonesia, Iran, Israel, South Korea, Laos, Lebanon,

Mongolia, Morocco, Norway, Poland, South Africa, Southeast Asian Nations, Sweden, Thailand, Tunisia, United States, and Bulgaria. Neither China nor most of the countries involved in the agreements is a “native-speaker of English” country; however the agreements were all signed in English, which further proves that English is used an official language in maritime industry.

All the four genres mentioned above were carefully cleansed and edited manually so that as a UTF-8 text version, it can be freely imported into Wordsmith and NetMiner for further analysis.

3.1.3.2 Size, Balance, and Representativeness

A corpus is composed of many selected texts which is intended to provide a somewhat representative sample of language from a particular genre or specific discourse community. As mentioned in Section 3.1.2.2, there is no need to make a large corpus so as to provide the maximum amount of data to the analyst. For a specialized corpus, as long as it is selected from the range of text types in the target population (Biber, 1993, p. 243), then if the data is sufficient enough to generate the results needed, it is not necessary to deliberate the size of the corpus. Especially considering the fact that there are many government documents involved in the Maritime English Corpus (MEC), the data collected will be limited to the materials that are available.

The issue of balance and representativeness may, on the other hand, limit the number of texts in particular sub-corpora. This has been particularly pertinent in the construction of the MEC, where the texts have been selected to be representative of various text types within the maritime discourse. Atkins, Clear, and Ostler (1992, p. 6) posited the idea of the balanced corpus as "a corpus so finely tuned that it offers a manageably small scale model of the linguistic material which corpus users wish to study." However, they also pointed out that such balance "relies heavily on intuition." Since the MEC does not include spoken word texts, it cannot be regarded as truly representative of the total discourse, for just as Sinclair (2005, p. 9) pointed out, "for a corpus to be truly balanced, that is, representative of actual language in use, it would have to be composed of up to 90% spoken language to reflect the relative amount of speech, compared to written language, that most people will encounter." In this study I have based decisions on what to exclude and include in the corpus as "maritime texts" and built the corpus based on the availability and suitability of the maritime texts. Therefore, I consider the MEC as a balanced corpus, in which each sub-corpus has about 1.25 million words.

As for the factor of "representativeness," at the beginning of the construction of the corpus, I considered adding maritime fiction such as Melville's (1851) *Moby-Dick* or Joseph Conrad's maritime fictions. However, I discarded this thought considering the fact that literature writings will polish or exaggerate the words used so as to deliver another meaning. This is not helpful to the near-synonyms analysis

in this thesis. Moreover, I did not include the genre of maritime news as Lee (2016) did because news language is often brief and simple, which is not the actual language use in the maritime industry. Therefore, in the MEC, I only included journals, textbooks and government documents, all of which can be considered as representative, standard and formal maritime language.

3.1.3.3 Multi-word Compounds in the MEC

Generally speaking, few researchers will consider multi-word compounds when building a corpus. However, in Maritime English, many terms are used as a compound to designate a specific meaning. Plag, Braun, Lappe, & Schramm (2009) has classified compounds in General English and Engineering English, as shown in Table 3.4.

Table 3.4
Classification of multi-word compounds
From Plag et. al., 2009.

	General English	Engineering English
a.	*ashtray, windmill, hotline*	*hypsometer, azeotrope, nuclide*
b.	*fast-food, icy-cold, call-girl*	*Sabatier–Senderens process, two-phase multiplier*
c.	*ice cream, income tax increase*	*absolute density, butex process, safety audit*

Since only adjacent collocates (i.e., window span 1-1) will be considered in this thesis, it will be best if multi-word compounds are marked so as not to disturb the collocation extraction result. Lee (2016) considered the situation and compiled his

study based on a corpus in which the multi-word compounds were treated as single items. In this thesis, his process of “compounding”, as shown in the following Figure 3.2 was adopted.

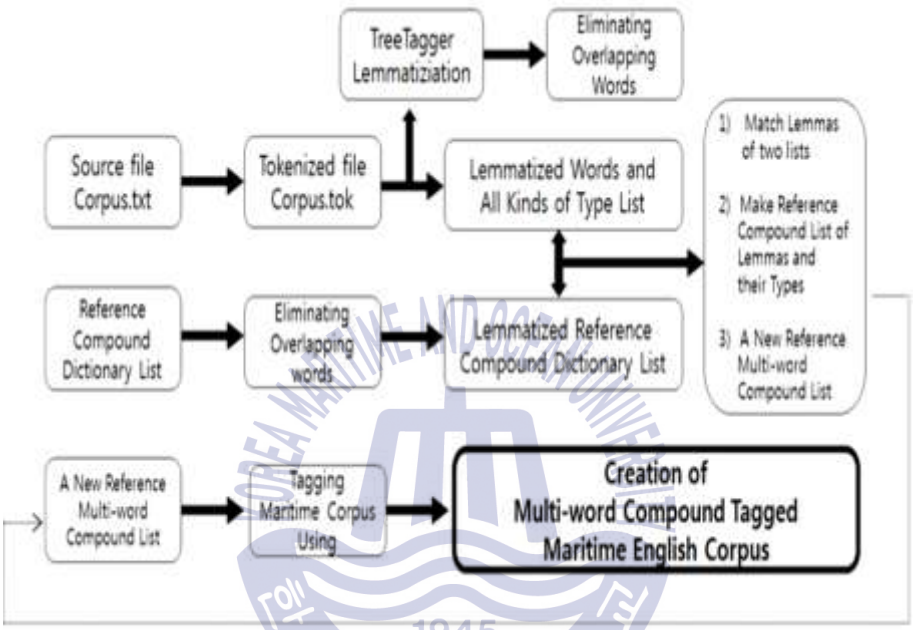


Figure 3.2. Creation of multi-word compound tagged MEC. From Lee, 2016.

Microsoft Visual FoxPro 6.0 (hereafter referred to as FoxPro) (Microsoft, 2011) is a powerful and widely-used computer database management system. It is particularly suitable for linguistic computing. A work platform of Foxpro is displayed below in Figure 3.3.

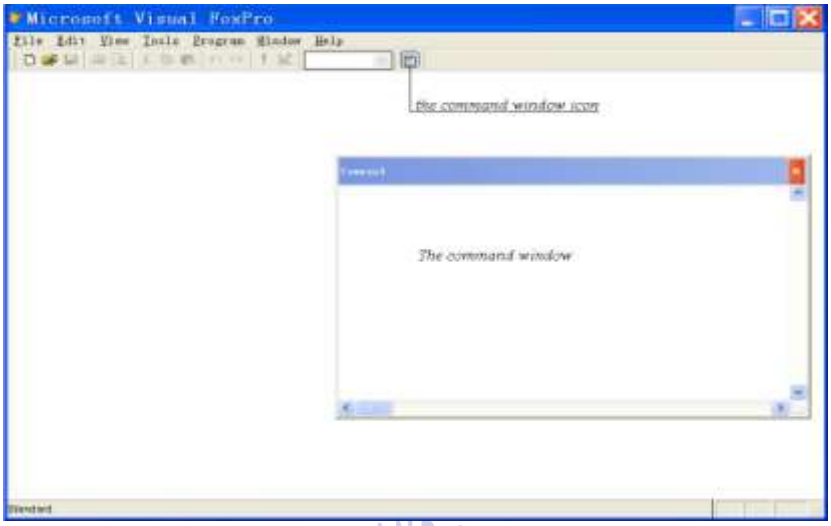


Figure 3.3. A work platform of Visual FoxPro 6.0.

At first, a lemmatization process was performed through TreeTagger (Schmidt, n.d.), although one could use any type of lemmatizer available. Then the following command instructions were conducted on the corpus so as to get the final MEC with multi-words compounds in it.

```
SET DEFA to D:\DATA-PROCESSING20160403\Lu Wenyu\Maritime-English-Corpus
SET SAFE OFF
SET TALK OFF
CLOSE ALL DATA
CLEAR

creat table tempt (stng c (40), source m (4))
APPEND FROM MARITIME-COMPOUND
appen memo source from NEC.txt
DELETE ALL FOR ALLTRIM(SOURCE)=="
PACK
REPLACE ALL SOURCE WITH LOWER(ALLTRIM(SOURCE))

SELECT 2
USE MARITIME-COMPOUND
```

```
REPLACE ALL COMPOUND WITH
STRTRAN(ALLTRIM(COMPOUND),"","""")
REPLACE ALL STNG WITH STRTRAN(ALLTRIM(STNG),"","""")
GO TOP
DO WHILE! EOF ()
AA=ALLTRIM(STNG)
? AA
BB=ALLTRIM(COMPOUND)

SELECT 1
REPLACE ALL SOURCE WITH
STRTRAN(ALLTRIM(SOURCE),"&AA","&BB")
SELECT 2
SKIP
ENDDO

SELECT 1
GO TOP
COPY MEMO SOURCE TO NEC-COMPOUND.TXT
```

3.1.3.4 Basic Information of the MEC

As mentioned above, the MEC went through a process of "compounding" so as to better display the characteristics of Maritime English. Instead of viewing each word as a separate single item, each maritime technique term is considered as a single multi-word compound. Therefore, besides the words like *bow*, *aft*, *bridge*, *deck*, *captain*, *cadet*, etc., in this study word combinations such as *engine room*, *deep draught*, *fog signal*, *passing light*, *cathedral hull*, *on the water*, *fore and aft*, *estimated time of arrival*, *International Convention for the Safety of Life at Sea*, *global maritime distress and safety system*, *left-hand draft in this set of marks*, *between the devil and the deep blue sea*, *dry bulk self-unloaded ship* are also considered as one single collocate of the node word.

After this compounding, the statistics of the MEC changed. Table 3.5 displays the statistics of the original MEC and compounding MEC, Table 3.6 shows the information of the sub-corpora of the original MEC, while Table 3.7 presents the information of the sub-corpora in MEC after the compounding process.

Table 3.5
General information of original and compounding MEC

Corpora	**Tokens**		**Types**	**Standardized Type-Token Ratio (STTR)**	**Mean Word Length**
Original MEC	All: 2,610,521	Sciences:1,287,173 Arts: 1,323,348	49,166	36.56	4.94
Compounding MEC	All: 2,589,669	Sciences:1,280,696 Arts: 1,308,973	50,764	36.90	4.99

Table 3.6
General information of subcorpora in original MEC

Corpora	**Tokens**	**Types**	**Standardized Type-Token Ratio (STTR)**	**Mean Word Length**
Maritime Engineering English	402,541	13,081	35.70	4.84
Navigation English	884,632	33,039	40.68	4.91
Maritime Conventions and Codes	814,197	21,013	36.84	4.89
Agreements on Maritime Transport	509,151	11,354	29.11	5.14

Table 3.7
General information of subcorpora in compounding MEC

Corpora	Tokens	Types	Standardized Type-Token Ratio (STTR)	Mean Word Length
Maritime Engineering English	401,524	13,350	35.73	4.86
Navigation English	879,172	33,973	40.89	4.95
Maritime Conventions and Codes	804,087	21,997	37.68	4.96
Agreements on Maritime Transport	504,886	11,762	29.63	5.20

Before moving into observation, four concepts should be clarified. First, "token" is the number of words a corpus has. If the word such as *sea* exists in the corpus 27 times, then it is counted 27 times. However, types refer to the number of different words in a corpus. Therefore, although for example, *sea* exists in the corpus 27 times, it is only counted as 1 type. In addition, standardized type-token ratio (STTR) is computed per 1,000 words as a word list goes through each text file. It is an effective way to show a variety of vocabulary in the corpus. If two corpora are compared, Corpus A has a lower type-token ratio than Corpus B; Corpus A could be said to have more repetition and thus it is less varied in vocabulary than Corpus B (Olohan, 2004). At last, word length is an important factor to determine the style of a text. The shorter and simpler the words are, the less difficult and complicated the text will be in general. The common core such as Anglo-Saxon vocabulary (e.g., *cow*, *lord*, *grave*, etc.) is usually used in informal style for smooth and easy

communication, while Latinate and Greek vocabulary (e.g., *impromptu*, *amphitheater*, etc.) is usually used in formal style to express conventions and standards.

When comparing Table 3.5 with Tables 3.6 and 3.7, we can observe that after the process of compounding, the token number has decreased 0.8% and the number of types has increased about 3.2%. This is because many 1-gram words have been linked together into 2 or more grams' compounds, hence fewer tokens and more types. Most noticeably, the subcorpus "Arts" (Conventions and codes, Agreements) changed most, especially the maritime conventions and codes genre. This indicates that in these types of legal documents, there are many multi-word compounds. Therefore, it is important for maritime law teachers to specifically teach these compounds to learners.

For the value of STTR, with the decreasing of tokens and the increasing of types, it is understandable that generally speaking, the value of STTR is greater after compounding. What attracts attention is that the STTR in subcorpus "Arts," especially the agreements on maritime transport, is much smaller than other genres. This can be explained by the fact that as a specialized genre, many of the word types used in maritime legal documents are used repeatedly, whereas in academic books and journals a variety of terms are used to avoid redundancy. Moreover, in the subcorpus "Science," it seems that Navigation English is more varied than Maritime Engineering English. But here I do not wish to offer a too-hasty

conclusion, since the data collected might not be sufficient to generate any definite answer.

For mean word length, because of the process of compounding, many words become longer, hence bigger mean word length (4.99) in Compounding MEC when compared with the Original MEC (4.94). In addition, although the STTR of Agreement on Maritime Transport is the lowest among the four, its mean word length (5.20) is much higher than others, followed by Maritime Conventions and Codes (4.96). Northcott (2013, p. 215) mentioned that historical development of languages and desire of power are the two main factors that make the image of the law as something inaccessible, mysterious and frightening. For one thing, legal English has borrowed many words from Romans (e.g., *incrimination*), Anglo-Saxons (e.g., *shareholder*), and Norman French (e.g., *immunity*), which are quite long compared with contemporary English. For another, increasing word length will help to maintain the image of the law as something inaccessible, complex and difficult to be grasped by general populace, therefore enabling authority to maintain power.

3.2 Methodology for Collocates Extraction

Wordsmith 6.0 (Scott, 2015) was used to extract significant collocates of node words in the MEC. Significant collocation takes place when two or more words occur together more frequently than would be expected by coincidence. The most

common words in any corpus are likely to be grammatical words and these will usually appear at the top of collocate lists as being the words most commonly found in proximity to lexical items. However, unless we are looking specifically for grammatical patterns (colligation), what we are interested in is the patterning of the lexical items within the text that gives it meaning. Statistical tools are able to sort the data by strength or significance of collocation. In this thesis, Mutual Information Value 3 (MI3) (Oakes, 1998) is used to select significant collocates of near-synonyms.

Traditionally, most quantitative corpus studies have used statistical hypothesis tests (chi-squared, z-test, t-test, Fisher's exact test, etc.) to measure the strength of collocation (Oakes, 1998). Evert (2004) extensively explored background theories and tested various statistical measures for finding collocations. His study suggested that both old and new hypotheses have to be tested under a range of conditions. More recently, Gries (2013) pointed out that there are three criteria for finding collocations: (i) directionality, (ii) dispersion, (iii) type-token distribution among collocates. Directionality means that the degree of strength between collocates. For example, a second word, *course*, in *of course* shows a more intense relationship with *of*, but *of* is not a good cue for *course*. Dispersion indicates the degree of distribution among given texts in corpora. For example, the term *of course* appeared in a number of different texts. Third, type-token distribution measures that how many different types can be collocates which compete with *course* for a position around *of*. Jhang and Lee (2014) tested several statistical hypothesis tests

with a one-million-word corpus of Maritime English to find out which test produces better results. They concluded that z-score was a better measure than the others for their corpus.

However, I chose statistical measures based on information theory instead since natural language does not satisfy assumptions required under hypothesis tests. In corpus linguistics, there are two widely used measures based on information theory. They are Mutual Information (MI) (Church & Hanks, 1990) and Mutual Information 3 (MI3) (Daille, 1995; Oakes, 1998). MI measurement calculates the strength of collocation based on how strongly words are connected to the target word (node) within a given span, where the higher the score, the stronger the link and the lower the score, the more tenuous the relationship between the words. Just as Hunston (2002 p. 71) put it "the MI-score measures the amount of non-randomness present when two words co-occur."

There is a problem, however, with MI scores, in that rare words in the corpus can gain unwarranted prominence in collocation lists as MI scores will tend to highlight rarer words that appear in limited collocations (Oakes, 1998, p. 171). Recognizing the tendency for MI to over-emphasize rare items, Daille (1995) experimented with the MI equation and developed MI3, which gives greater weight to more frequent items in the corpus and thus reduces the tendency for unusual lexical pairings to dominate the collocate list. MI3 is calculated by dividing observed frequencies of the co-occurring words by expected frequencies of the co-occurring words within

specific spans, taking the logarithm to the base 2 of the outcome: $I(x,y) = \log^2 \frac{P(x,y)^3}{P(x)P(y)}$. By adding a "cubing" of observed frequencies, MI3 made it possible to give more weight to high frequencies than to low frequencies (Oakes 1998, pp. 171-172). A minimum MI3 value of 3, according to Hunston (2002, p. 71) , can be regarded as good evidence of meaningful collocation, hence MI3≥3 in this study.

Given the size of the MEC, I set the minimum co-occurrence frequency to 2. As for window spans, I decided to use adjacent collocates, i.e., a window span of one word to the left and one to the right, so as to reduce irrelevant words and show strong semantic relationships between nodes and collocates for the purpose of investigating semantic influences.

Therefore, within a 1-1 window span, items which have a minimum co-occurrence frequency of 2 as a collocate of a given node word and a minimum MI3 score above 3 are considered to be collocates of a node word.

3.3 Methodology for Networks Visualization

Many tools are designed to show the networks between different nodes, such as LaNCoA toolki, UNICET, WORDNET, and GRAPHCOLL. In this thesis, in order to visualize collocations, NetMiner Version 4.0 (Cyram, 2016) from social network analysis was used to discover nodes and links. NetMiner is a software tool to present a visualization of the network of large quantities of data. It is useful to

explain relationships between words. It was originally designed to detect underlying patterns and structures of the network and has been used for general research and teaching in social networks in various fields such as information science, biology, geography, economics, and political science.

The significant collocates were chosen and loaded into NetMiner. In order to effect network analysis, extraction of bigrams was also done through WordSmith. Only the bigrams that ad frequencies not less than 2, and those collocates which have MI3 scores no less than 3 will be considered to for analysis in NetMiner. In addition, in order to avoid unnecessary redundancy, the bigrams which contain the following kinds of words were also removed before the bigrams were imported into NetMiner, these words were (1) articles: *the*, *a*, *an*; (2) possessive pronouns: *its*, *their*, *his*, *her*; (3) pronouns: *I*, *they*, *he*, *she*; (4) prepositions: *about*, *on*, *in*, *of*, *after*, *between*, *against*, *under*, etc.; (5) conjunctions: *and*, *or*, *however*, *as*, etc.; (6) auxiliary verbs: *is*, *are*, *have*, *do*, etc.; (7) modal verbs: *can*, *may*, *must*, etc.; (8) demonstrative pronouns: *this*, *that*, *these*, *those*; (9) numbers: *one*, *1000*, etc.; (10) units: *tons*, *tonnage*, etc.; (11) interrogative pronouns: *what*, *which*, *when*, etc.; (12) countries: *UK*, *US*, *New Zealand*, etc.; (13) negations: *not*, *non*, *no*; (14) other functional adjectives and adverbs: *than*, *then*, *also*, *already*, *to* , *all*, *any*, *more*, *most*, etc.

Using NetMiner, community groups were automatically calculated through Analyze >> Cohesion >> Betweenness, as suggested by Girvan and Newman

(2002), who proposed a property of community structure, in which network nodes were joined together in tightly knit groups, between which there were only looser connections. I then compiled all bidirectional networks linked with the node and used 2D "spring embedding" algorithms to visualize the data. This is a straightforward implementation of Kamada and Kawai's (1989) spring embedding algorithm, $k_{i,j} = K/d_{i,j}^2$, which is one of the force-directed graph layout algorithms. The strength of the spring between vertices *i* and *j* is defined as where *K* is a constant. This algorithm computed by force-based graph layout algorithms has the advantage of drawing networks.

3.4 Methodology for Semantic Tagging

Most linguistic analyses of frequently-used words have been based either on parts of speech (word class) or on syntactic categories. With the application of more advanced searching techniques such as the Wmatrix system utilized in the present study, semantic domains can be accounted (Lu, Lee, and Jhang, 2017). The Wmatrix web interface program for the UCREL semantic analysis system (USAS) (Rayson, Archer, Piao, & McEnery, 2004; Rayson & Mariani, 2009) was initially developed in the REVERE project by Rayson (2008) and has been used across a number of research projects. Semantic domains are analyzed based on McArthur's (1981) *Longman Lexicon of Contemporary English*. The semantic categories used by USAS have a multi-tier structure with 21 major discourse fields. Each major category is further fine-grained into several subcategories. There are total of 113

subcategories, as shown in Table 3.8. These subcategories can be further fine-grained into subgroups.

Table 3.8
USAS semantic tagset
From UCREL semantic analysis system (USAS), Rayson, 2008.

A GENERAL & ABSTRACT TERMS	**I MONEY & COMMERCE**	**S SOCIAL ACTIONS, STATES & PROCESSES**
A1 General	I1 Money generally	S1 Social actions, states & processes
A1.1.1 General actions, making etc.	I1.1 Money: Affluence	S1.1 Social actions, states & processes
A1.1.2 Damaging and destroying	I1.2 Money: Debts	S1.1.1 General
A1.2 Suitability	I1.3 Money: Price	S1.1.2 Reciprocity
A1.3 Caution	I2 Business	S1.1.3 Participation
A1.4 Chance, luck	I2.1 Business: Generally	S1.1.4 Deserve etc.
A1.5 Use	I2.2 Business: Selling	S1.2 Personality traits
A1.5.1 Using	I3 Work and employment	S1.2.1 Approachability and Friendliness
A1.5.2 Usefulness	I3.1 Work and employment: Generally	S1.2.2 Avarice
A1.6 Physical/mental	I3.2 Work and employment: Professionalism	S1.2.3 Egoism
A1.7 Constraint	I4 Industry	S1.2.4 Politeness
A1.8 Inclusion/Exclusion	**K ENTERTAINMENT, SPORTS & GAMES**	S1.2.5 Toughness; strong/weak
A1.9 Avoiding	K1 Entertainment generally	S1.2.6 Sensible
A2 Affect	K2 Music and related activities	S2 People
A2.1 Affect: Modify, change	K3 Recorded sound etc.	S2.1 People: Female
A2.2 Affect: Cause/Connected	K4 Drama, the theatre & show business	S2.2 People: Male
A3 Being	K5 Sports and games generally	S3 Relationship
A4 Classification	K5.1 Sports	S3.1 Relationship: General
A4.1 Generally kinds, groups, examples	K5.2 Games	S3.2 Relationship: Intimate/sexual
A4.2 Particular/general; detail	K6 Children's games and toys	S4 Kin
A5 Evaluation	**L LIFE & LIVING THINGS**	S5 Groups and affiliation
A5.1 Evaluation: Good/bad	L1 Life and living things	S6 Obligation and necessity
A5.2 Evaluation: True/false	L2 Living creatures generally	S7 Power relationship
A5.3 Evaluation: Accuracy	L3 Plants	S7.1 Power, organizing
A5.4 Evaluation: Authenticity	**M MOVEMENT,**	S7.2 Respect
A6 Comparing		S7.3 Competition
A6.1 Comparing:		S7.4 Permission
		S8 Helping/hindering
		S9 Religion and the

Similar/different	**LOCATION, TRAVEL & TRANSPORT**	supernatural
A6.2 Comparing: Usual/unusual	M1 Moving, coming and going	**T TIME**
A6.3 Comparing: Variety	M2 Putting, taking, pulling, pushing, transporting	T1 Time
A7 Definite (+ modals)	M3 Movement/transportation: land	T1.1 Time: General
A8 Seem	M4 Movement/transportation: water	T1.1.1 Time: General: Past
A9 Getting and giving; possession	M5 Movement/transportation: air	T1.1.2 Time: General: Present; simultaneous
A10 Open/closed; Hiding/Hidden; Finding; Showing	M6 Location and direction	T1.1.3 Time: General: Future
A11 Importance	M7 Places	T1.2 Time: Momentary
A11.1 Importance: Important	M8 Remaining/stationary	T1.3 Time: Period
A11.2 Importance: Noticeability	**N NUMBERS & MEASUREMENT**	T2 Time: Beginning and ending
A12 Easy/difficult	N1 Numbers	T3 Time: Old, new and young; age
A13 Degree	N2 Mathematics	T4 Time: Early/late
A13.1 Degree: Non-specific	N3 Measurement	**W THE WORLD & OUR ENVIRONMENT**
A13.2 Degree: Maximizers	N3.1 Measurement: General	W1 The universe
A13.3 Degree: Boosters	N3.2 Measurement: Size	W2 Light
A13.4 Degree: Approximators	N3.3 Measurement: Distance	W3 Geographical terms
A13.5 Degree: Compromisers	N3.4 Measurement: Volume	W4 Weather
A13.6 Degree: Diminishers	N3.5 Measurement: Weight	W5 Green issues
A13.7 Degree: Minimizers	N3.6 Measurement: Area	**X PSYCHOLOGICAL ACTIONS, STATES & PROCESSES**
A14 Exclusivizers/particularizers	N3.7 Measurement: Length & height	X1 General
A15 Safety/Danger	N3.8 Measurement: Speed	X2 Mental actions and processes
B THE BODY & THE INDIVIDUAL	N4 Linear order	X2.1 Thought, belief
B1 Anatomy and physiology	N5 Quantities	X2.2 Knowledge
B2 Health and disease	N5.1 Entirety; maximum	X2.3 Learn
B3 Medicines and medical treatment	N5.2 Exceeding; waste	X2.4 Investigate, examine, test, search
B4 Cleaning and personal care	N6 Frequency etc.	X2.5 Understand
B5 Clothes and personal belongings	**O SUBSTANCES, MATERIALS, OBJECTS & EQUIPMENT**	X2.6 Expect
C ARTS & CRAFTS	O1 Substances and materials generally	X3 Sensory
C1 Arts and crafts	O1.1 Substances and	X3.1 Sensory: Taste
E EMOTIONAL ACTIONS, STATES &		X3.2 Sensory: Sound
		X3.3 Sensory: Touch
		X3.4 Sensory: Sight
		X3.5 Sensory: Smell
		X4 Mental object
		X4.1 Mental object:

PROCESSES E1 General E2 Liking E3 Calm/Violent/Angry E4 Happy/sad E4.1 Happy/sad: Happy E4.2 Happy/sad: Contentment E5 Fear/bravery/shock **E6 Worry, concern, confident** **F FOOD & FARMING** F1 Food F2 Drinks F3 Cigarettes and drugs F4 Farming & Horticulture **G GOVT. & THE PUBLIC DOMAIN** G1 Government, Politics & elections G1.1 Government etc. G1.2 Politics G2 Crime, law and order G2.1 Crime, law and order: Law & order G2.2 General ethics G3 Warfare, defence and the army; Weapons **H ARCHITECTURE, BUILDINGS, HOUSES & THE HOME** H1 Architecture, kinds of houses & buildings H2 Parts of buildings H3 Areas around or near houses H4 Residence H5 Furniture and household fittings	materials generally: Solid O1.2 Substances and materials generally: Liquid O1.3 Substances and materials generally: Gas O2 Objects generally O3 Electricity and electrical equipment O4 Physical attributes O4.1 General appearance and physical properties O4.2 Judgement of appearance (pretty etc.) O4.3 Colour and colour patterns O4.4 Shape O4.5 Texture O4.6 Temperature **P EDUCATION** P1 Education in general **Q LINGUISTIC ACTIONS, STATES & PROCESSES** Q1 Communication Q1.1 Communication in general Q1.2 Paper documents and writing Q1.3 Telecommunications Q2 Speech acts Q2.1 Speech etc: Communicative Q2.2 Speech acts Q3 Language, speech and grammar Q4 The Media Q4.1 The Media: Books Q4.2 The Media: Newspapers etc. Q4.3 The Media: TV, Radio & Cinema	Conceptual object X4.2 Mental object: Means, method X5 Attention X5.1 Attention X5.2 Interest/boredom/excited/energetic X6 Deciding X7 Wanting; planning; choosing X8 Trying X9 Ability X9.1 Ability: Ability, intelligence X9.2 Ability: Success and failure **Y SCIENCE & TECHNOLOGY** Y1 Science and technology in general Y2 Information technology and computing **Z NAMES & GRAMMATICAL WORDS** Z0 Unmatched proper noun Z1 Personal names Z2 Geographical names Z3 Other proper names Z4 Discourse Bin Z5 Grammatical bin Z6 Negative Z7 If Z8 Pronouns etc. Z9 Trash can Z99 Unmatched

As pointed out by Lu, Lee, and Jhang (2017), the understanding of macroscopic analysis (the study of the characteristics of whole texts or varieties of language) is able to inform the microscopic level (focusing on the use of a particular linguistic feature). Therefore, in this thesis, I assign semantic tags to the collocations of near-synonyms, hoping to see the intricate connections between near-synonyms from a higher vantage point. The accuracy rate for the semantic tagger is 91%-92% (Rayson et al., 2004). After the completion of analysis by the software, manual proof-checking was conducted before entering into serious discussions of the tagging results.

3.5 Process of Data Analysis

The raw data were processed step by step by WordSmith, Wmatrix, and NetMiner. Firstly, WordSmith was used to extract collocates of the target near-synonyms. Secondly, the collocates were tagged by Wmatrix for their semantic domains. Thirdly, both the collocates bigrams and semantic domains bigrams were imported into NetMiner for a 2-D language network analysis.

Chapter 4. Collocation Network Analysis of Near-synonyms

In this chapter I examined the collocations of five groups of typical maritime near-synonyms from the perspective of collocation network analysis. Firstly, a review of the differences among these near-synonyms were given from the aspect of dictionary definition, etymology origin and studies in other areas. These near-synonyms are *ship* vs. *vessel*, *maritime* vs. *marine*, *sea* vs. *ocean*, *safety* vs. *security*, and *port* vs. *harbor*. The meanings of these near-synonyms vary considerably from one context to another, leading to potential ambiguities for related practitioners. The maritime field is particularly prone to such difficulties. As a specialized agency of the United Nations, International Maritime Organization (IMO) — the global standard-setting authority for the safety, security, and environmental performance of international shipping — has published many documents including the concepts of these near-synonyms, both at the policy and technical levels; however, there is no policy established for maritime field to distinguish between these terms. Many misunderstandings have occurred where rights and duty were not clearly addressed. Misfortunes and disputes arouse from natural disasters or human operation errors; therefore, this chapter focused on these frequently-used but rarely-defined near-synonyms in the maritime field. Secondly, a cosine-similarity calculation was applied to the pairs of words. Judging from the result, it can be seen that some pairs of words have less or no similarity. The degree of similarity also affected the behavior of near-synonyms. Lastly, when analyzing

the collocates of near-synonyms, it is very important to distinguish collocates by word types rather than lemmas since different types of words may prefer to collocate with different words.

4.1 Meaning Differences

4.1.1 *Ship* vs. *Vessel*

Vessel usually refers to a container with a cavity. The main task of *vessel* is usually carrying goods or people. It can be preceded by a noun to indicate its function, as in *merchant vessel*, *passenger vessel*, or *whaling vessel*. In modern English, it can also refer to spacecraft or seaplane. People usually use the feminine words like *she* or *her* to refer to the *ship*. In literature, *ship* is poetic and full of emotions, while *vessel* does not entail this colorization. In the poem *O Captain, My Captain* written by Walt Whitman, the first two lines are "O Captain! My Captain! Our fearful trip is done. This *ship* has weather'd every rack, the prize we sought is won..." Obviously, the poet took the United States as the *ship*, with President Lincoln as the captain to lead the USA through fierce currents and horrible waves. Of course, there are also other types of water transportation carriers such as *boat*, *canoe*, *yacht*, *cruiser*, *raft*, *freighter*, *liner*, *tanker*, *container*, etc. They can usually be distinguished by their functions; therefore, I do not consider these terms as near-synonyms in this thesis.

In the *Oxford Advanced Learner's English-Chinese Dictionary* (Hornby, 2002, p. 1389, 1687), although originating from different words, *ship* and *vessel* are used exchangeably to define each other. *Ship* is defined as "a large vessel for carrying people or goods by sea", originating in the 1300s, from Old English *scipian*. Whereas originating in the 1300s, from Modern French *vaisseau* and from Late Latin *vascellum*, *vessel* is defined as "a *ship* or boat, esp. a large one". Therefore, it is very hard for language learners to understand the differences of the two words from the dictionary.

4.1.2 *Maritime* vs. *Marine*

The near-synonyms *maritime* and *marine* were selected because these words are ambiguous in terms of pedagogical aspects for English learners of L2 to study their appropriate usages. *Marine* is usually used to talk about something that is directly connected to the sea and sea transport or is encountered at sea. For example, *marine plants*, *marine animals*, *marine life*, *marine traffic*, *marine engineer*, *marine product*, *marine pollution*, etc. *Maritime* is chosen when we talk about human activity at sea, such as *maritime insurance* (for ships and cargo transported by sea), *maritime town* (built on the coast of a sea), *maritime museum* (contains exhibits connected with the sea), etc.

In addition, most dictionaries define *maritime* and *marine* almost in the similar words. In the *Oxford Advanced Learner's English-Chinese Dictionary* (Hornby,

2002, p. 906), although originating from different words, *maritime* and *marine* are used exchangeably to define each other. *Maritime* is defined as "of the sea, sailing or shipping", originating in the mid-16th Century from Latin *maritimus*. Whereas originating in the mid-1660s, from Old French *marin* / *marine*, and from Latin *marinus*, *marine* is defined as "of, near, found in or produced by the sea." Therefore, it is very hard for language learners to understand the differences of the two words from the dictionary.

4.1.3 *Sea* vs. *Ocean*

In the *Oxford Advanced Learner's English-Chinese Dictionary* (Hornby, 2002, p. 1013, 1351), *sea* is used to define the *ocean*. *Sea* is defined as "the salt water that covers most of the earth's surface and encloses its continents and islands", originating in the early 1440, from Proto-Germanic *saiwaz*. Whereas originating in the late 13th Century, from Old French *occean*, *ocean* is defined as "mass of salt water that covers most of the earth's surface" Generally speaking, *ocean* is bigger while *sea* is smaller. There are four *ocean*s in the world, namely, the *Pacific Ocean*, *Atlantic Ocean*, *Indian Ocean*, and *Arctic Ocean*. But those areas close to the coast are usually called *sea* such as *Yellow Sea*, *Sea* of *Japan*, etc. However, when it comes to the collocates of these two near-synonyms — whether it is the *sea* culture or *ocean* culture or whether it is *sea* exploration or *ocean* exploration — people are not clear about it. Therefore, it is very important to see how these two near-

synonyms work with their general collocations.

4.1.4 *Safety* vs. *Security*

From the linguistic aspect, *safety* and *security* are not clearly distinguished internationally. Some languages have one single word for both *safety* and *security*. This is the case in German (*sicherheit*), Norwegian (*sikkerhet*) Spanish (*seguridad*), Portuguese (*segu-rança*), Swedish (*säkerhet*) and Danish (*sikkerhed*). Some other languages, however, have corresponding different words for *safety* and *security*. Such cases are English (*safety* vs. *security*), French (*sûreté* vs. *sécurité*), Chinese (安全 vs. 安保/保安) and Korean (**안전** vs.**안보**/**보안**).

In the *Oxford Advanced Learner's English-Chinese Dictionary* (Hornby, 2002, p. 1324, 1357), although originating from different words, *safety* and *security* are used exchangeably to define each other. *Safety* is defined as "being safe, not being dangerous or in danger", originating in the early 14th Century, from Old French *sauvete*, and from Medieval Latin *salvitatem* (nominative *salvitas*). To the contrary, originating in the mid-15th Century, from Latin *securitas*, *security* is defined as "freedom or protection from danger or worry". Therefore, it is very hard for language learners to understand the differences of the two words from the dictionary.

Because of the importance of the *safety* and *security* issue, many researchers have developed theories and methods to distinguish them, mainly in industry and

the information field. Burns, McDermid, and Dobson (1992) considered the distinction in terms of the differences in causal structure and in the degree of harm caused. According to them, a *safety*-critical system is one whose failure could be sufficient to cause absolute harm, while a *security*-critical system is one whose failure could not be sufficient to cause harm, but could increase the number of possibilities, or likelihood of existing possibilities, for others to intentionally cause relative harm. Dutertre and Stavridou (1998) suggested promising lines of research in the intersection of *safety* and *security*, in the application of security concepts and models to different classes of safety or fault-tolerance properties, and in the theory and practice of fault tolerant system applied to intrusion tolerance. Albrechtsen (2002) compared the differences between industrial *safety* and information *security*, concluding that the basic ideas of industrial *safety* and *security* are the same, both of which are protecting assets from danger, creating safe / secure conditions with only some nuances from the aspects of traditions, causes, threats / hazards, loss, surroundings and uncertainty. Pièvre-Cambacédès and Chaudet (2010) proposed a SEMA referential framework that makes the latent differences underlying the use of the terms *security* and *safety* explicit. They argued that on one hand, *security* is concerned with the risks originating from the environment and potentially impacting the system, whereas *safety* deals with the risks arising from the system and potentially impacting the environment (i.e., a system vs. environment distinction). On the other hand, *security* typically addresses malicious risks while *safety* addresses purely accidental risks (a malicious vs. accidental

distinction). Boholm, Möller, & Hansson (2015) analyzed the concepts of risk, *safety*, and *security*, as well as their relation, based on empirical observation of their actual everyday use. However, these analyses are based on the information and industry field and whether these theories are applicable to the maritime field is worth further discussion.

4.1.5 *Port* vs. *Harbor*

Many words are used to designate the meaning of the place where ships or vessels rest when they are close to the coast. In the *Oxford Advanced Learner's English-Chinese Dictionary* (Hornby, 2002, p. 677, 1142), *port* originates from Latin *portus*, and is defined as "(i) a town or city with a *harbor*, (ii) a *harbor*." In the premises of a modern *port* you will usually find a lot of stacked shipping containers as well as lined up cranes. Originating in the Early 12th Century, from Middle English *herberwe*, *harbor* (*harbour* in BrE) is defined as "a place of shelter for ships." Some *harbors* are considered natural because the "structures" that protect the coast are parts of their natural topography.

4.2 Similarity Degree of Groups of Near-synonyms

As mentioned in Section 2.3.2, within the class of near-synonyms, some pairs of items are more synonymous than others. This can be proven furthermore, if

statistical methods are used to measure how similar the five groups of words analyzed in this thesis are. Two methods were used to measure similarity as a matter of degree: one is based on number of shared collocates, while the other is based on MI3 cosine similarity.

4.2.1 Similarity Degree Based on Number of Shared Collocates

The number of shared collocates has long been used as an important measure factor to the similarity between two words. With the different number of collocates for each node word, since all the words are from the same MEC corpus there is no need to normalize the results. Therefore, Table 4.1 shows the overlapped rate of five groups of words based on shared collocates number.

Table 4.1
Similarity degree based on number of shared collocates

Near-synonyms	Total	Shared Collocates	Overlapped Rate
ship vs. *vessel*	671	140	20.9%
maritime vs. *marine*	312	40	12.8%
sea vs. *ocean*	163	19	11.7%
safety vs. *security*	274	15	5.5%
port vs. *harbor*	256	6	2.3%

From the table above, it can be seen that although the five pairs of words are considered near-synonyms in the maritime field, and can be easily confused by related practitioners, *safety* vs. *security* or *harbor* vs. *port* are not that exchangeablely-used, as might be assumed, compared to *vessel* vs. *ship*. This is

also one of the reasons why it is necessary to conduct this study to distinguish them in the maritime field.

4.2.2 Similarity Degree Based on MI3 Cosine Similarity

Cosine similarity is a very useful way to measure the similarity between two words. Based on the following formula, the cosine similarity based on MI3 scores of collocation strength was calculated.

$$\frac{\sum_{i=1}^{i=n} a_i \times b_i}{\sqrt{\sum_{i=1}^{i=n} a^2} \times \sqrt{\sum_{i=1}^{i=n} b^2}}$$ (a, b is the number of collocational strength: MI3)

FoxPro commands were used to automatically calculate all the collocates of the five pairs of near-synonyms as follows:

```
SET DEFAULT TO D:\DATA-PROCESSING20160403\LvWenyu\cosine-
similarity
SET TALK OFF
SET SAFETY OFF
CLOSE ALL data
CLEAR

CREATE TABLE tempt(word c(30),pmi n(10,4), pmi1 n(10,4), pmi2
n(10,4),d1d2 n(16,4),d1 n(10,4), d2 n(10,4),sym n(5))

synm1='port.txt'
synm2='wharf.txt'

APPEND FROM &synm1 DELIMITED WITH tab
REPLACE ALL pmi1 WITH pmi
DELETE ALL FOR pmi<3
DELETE ALL FOR ALLTRIM(word)=="
pack
COPY TO tempt1 FIELDS word, pmi1
```

```
ZAP
APPEND FROM &synm2 DELIMITED WITH tab
DELETE ALL FOR ALLTRIM(word)==''
REPLACE ALL pmi2 WITH pmi
DELETE ALL FOR pmi<3
pack
COPY TO tempt2 FIELDS word, pmi2

ZAP
APPEND FROM tempt1
APPEND FROM tempt2

REPLACE ALL word WITH ALLTRIM(PROPER(word))
INDEX on word TAG word
TOTAL ON word TO tempt8

SELECT 2
USE tempt8
COUNT TO overlap FOR pmi1>0 AND pmi2>0

COUNT TO colw1 ALL FOR pmi1>0
COUNT TO colw2 ALL FOR pmi2>0
cp1=overlap/colw1
cp2=overlap/colw2

REPLACE ALL d1d2 WITH pmi1*pmi2
REPLACE ALL d1 WITH pmi1^2
REPLACE ALL d2 WITH pmi2^2

SUM d1d2 TO fenzi

SUM d1 TO dd
sqrtd1=sqrt(dd)

SUM d2 TO ee
sqrtd2=sqrt(ee)

*cossim=(d1¡¤d2) / ||d1||*||d2||
cossim=fenzi/(sqrtd1*sqrtd2)
*?cossim
```

The results are shown in Table 4.2, which shows not only the cosine similarity, but also the shared collocates' proportion in each near synonym.

Table 4.2
Similarity degree based on MI3 cosine similarity

Near-synonyms	Cosine similarity	Proportion in the 1st word	Proportion in the 2nd word
vessel vs. *ship*	0.4485	0.3955	0.4416
maritime vs. *marine*	0.2820	0.3200	0.2130
sea vs. *ocean*	0.2200	0.1418	0.6552
safety vs. *security*	0.1520	0.0649	0.3571
port vs. *harbor*	0.0736	0.2220	0.0260

The ranking in Table 4.2 is the same as the Table 4.1. Both measures have proven that items commonly assumed to be near-synonyms do not show the same degree of similarity. In the following network analysis we are able to see how different the collocational behaviors are depending on different degree of similarity.

4.3 Collocation Network Analysis

4.3.1 *Ship* vs. *Vessel*

Figure 4.1 shows the collocation network of *ship* and *vessel*. We can see that the words *ship* and *vessel* are used a lot in Maritime English--they are the main subjects in the maritime industry. Both of them have many exclusive collocates to themselves. However, it is very hard to distinguish the differences between these

terms based only from the image presented through this collocational analysis. This is a shortcoming of this kind of network analysis, and is why later in Chapter 5, semantic domain network will be introduced.

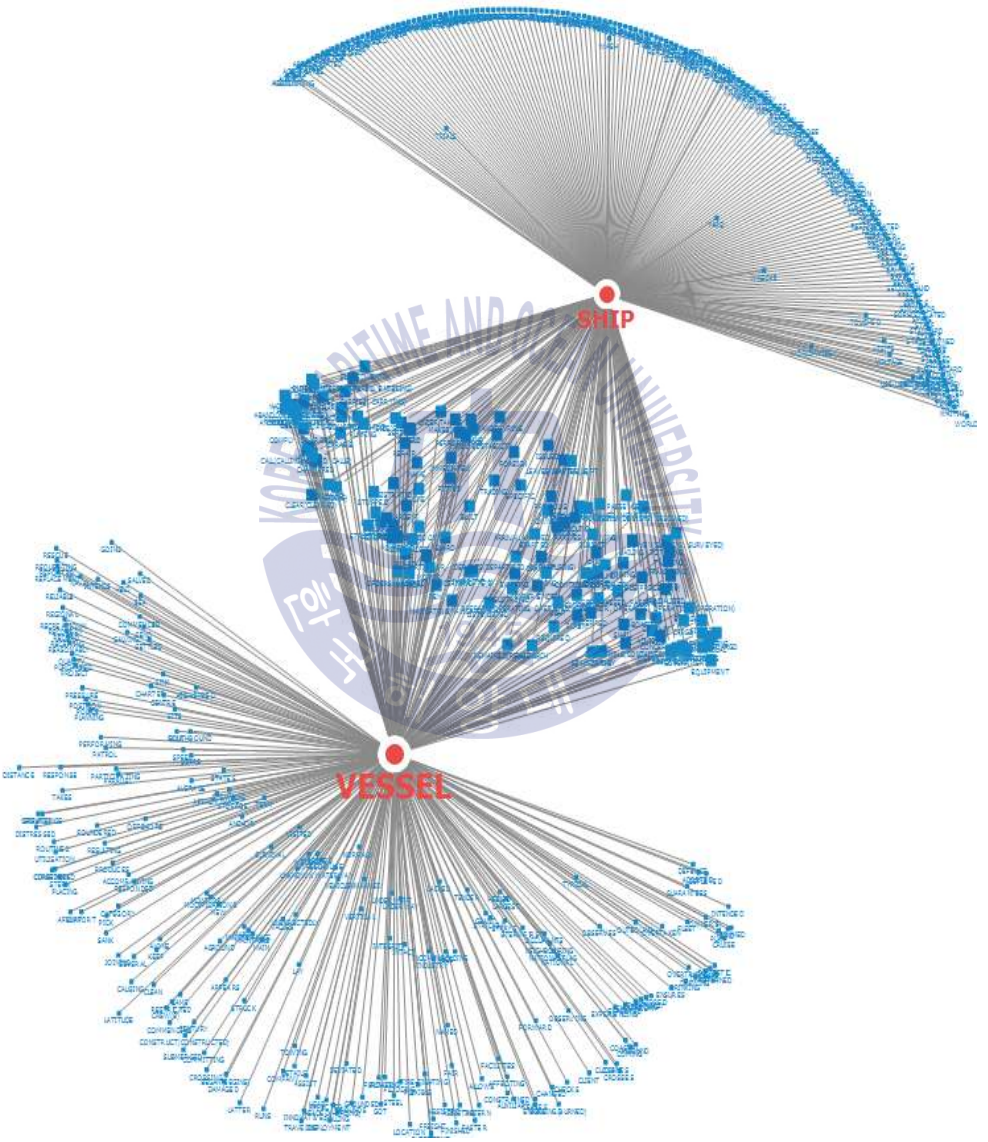


Figure 4.1. Collocation network of *ship* and *vessel*.

4.3.2 *Maritime* vs. *Marine*

Figure 4.2 shows the collocation network of *maritime* and *marine*. From this Figure, despite many collocates, we can still see easily the fact that *maritime* and *marine* have their own typical exclusive collocation groups in the specialized MEC. It can be observed that *maritime* has more types of collocates than *marine*. Therefore, by looking at the tendency of collocations in the corpus, we are able to see the differences between the two words. In addition, compared with the higher similarity degree pair *ship* vs. *vessel*, the collocates shared by *maritime* and *marine* can be seen clearly in Figure 4.2. However, since many collocates are involved in the network, it is not always easy to distinguish them in complex images of word nets. Moreover, the learners are not able to identify whether the words such as *accident(s)* are more preferable to *maritime* or *marine*.

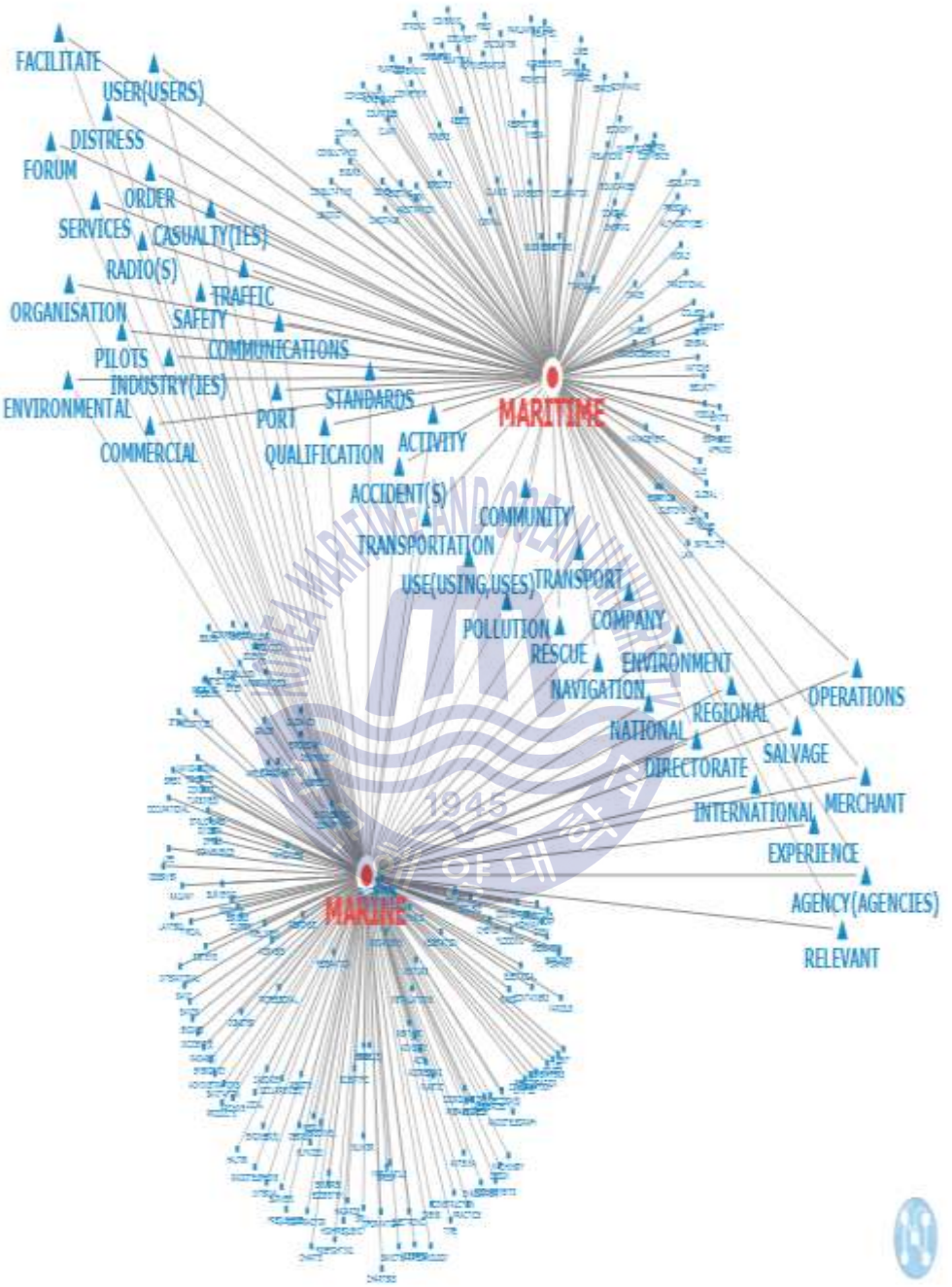


Figure 4.2. Collocation network of *maritime* and *marine*.

4.3.3 *Sea* vs. *Ocean*

Figure 4.3 shows the collocation network of *sea* and *ocean*. It can be clearly seen that people usually prefer to say *ocean environment*, *ocean policy*, *ocean carries*, *ocean cargo*, *national ocean*, *ocean voyage*, etc., while most of the other words tend to collocate with the more specific *sea*. In addition, they share some collocates on physical structures or parts in general. Words such as *bed, passages, inlets, open, areas*, *floor* or *bottom* all refer to the physical part of *ocean* or *sea*.

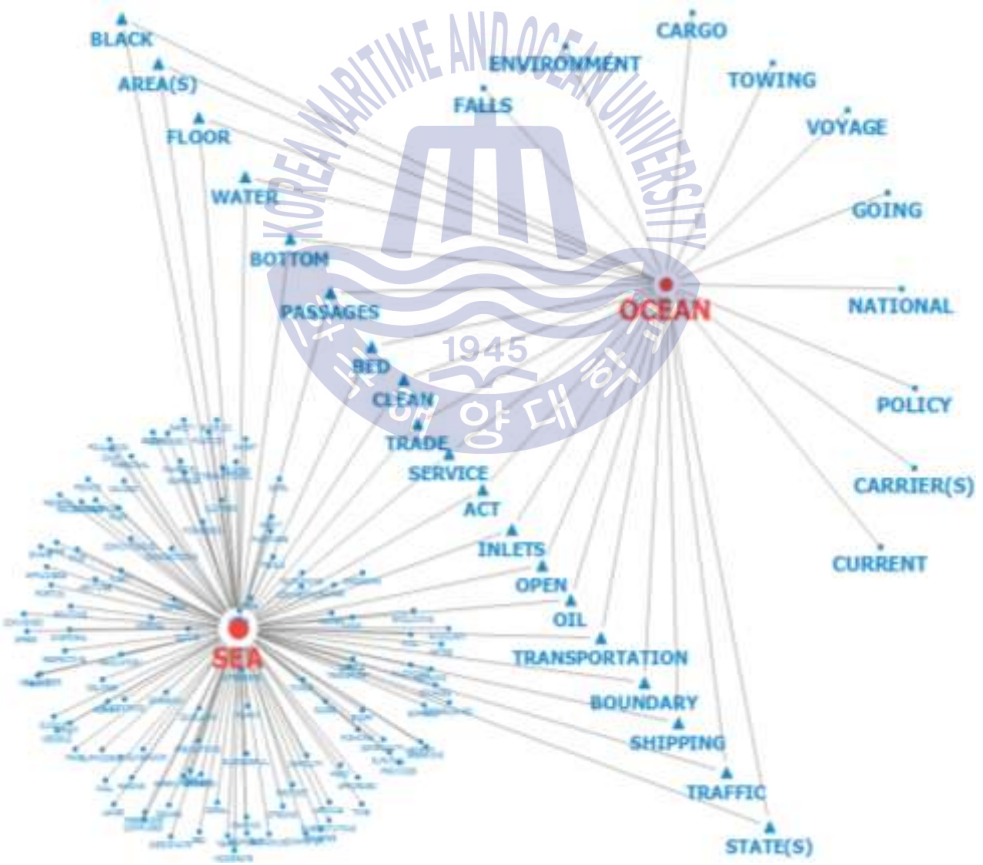


Figure 4.3.Collocation network of *sea* and *ocean.*

4.3.4 *Safety* vs. *Security*

Figure 4.4 shows the collocation network of *safety* and *security*. From this figure, we can see that firstly, *safety* has many more collocates than *security*, suggesting that in Maritime English, *safety* has more types of collocates than *security*. Secondly, in the specialized MEC, generally speaking, collocates tend to show very clear preference for either *safety* or *security*. Therefore, by looking at the tendency of collocations in the corpus, we are able to see the differences between the two words. Thirdly, the words *maritime* and *port* are among those collocates shared with both *safety* and *security*. Interestingly, *maritime* is the shared collocates of both *safety* and *security*, however *marine* is exclusive for *safety* only. This supports the aim of this study, that is, to determine usage differences among near-synonyms. At last, it can be observed that *system* is shared by both nodes, countering the System - Environment distinction proposed by Pietre-Cambacédès and Chaudet (2010).

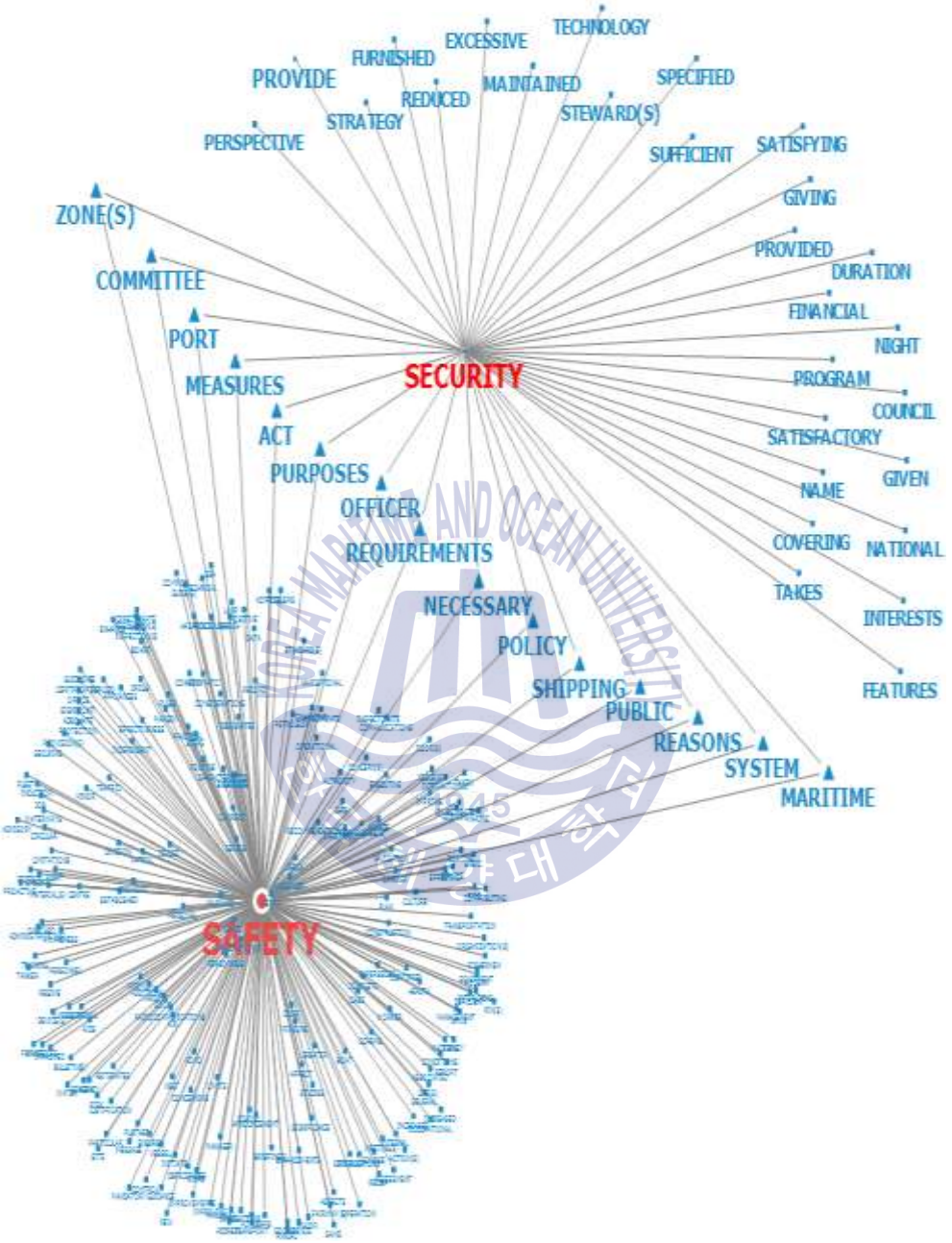


Figure 4.4. Collocation network of *safety* and *security*.

4.3.5 *Port* vs. *Harbor*

Figure 4.5 shows the collocation network of *port* and *harbor*. Firstly, it is clearly *port* has far more types of collocates, as compared with *harbor.* This supports what is explained in the dictionary explanation: sometimes *port* is so that it might be a city containing several *harbors*. In addition, with low degree of similarity, *port* and *harbor* have fewer shared collocates compared with the other four groups of near-synonyms.

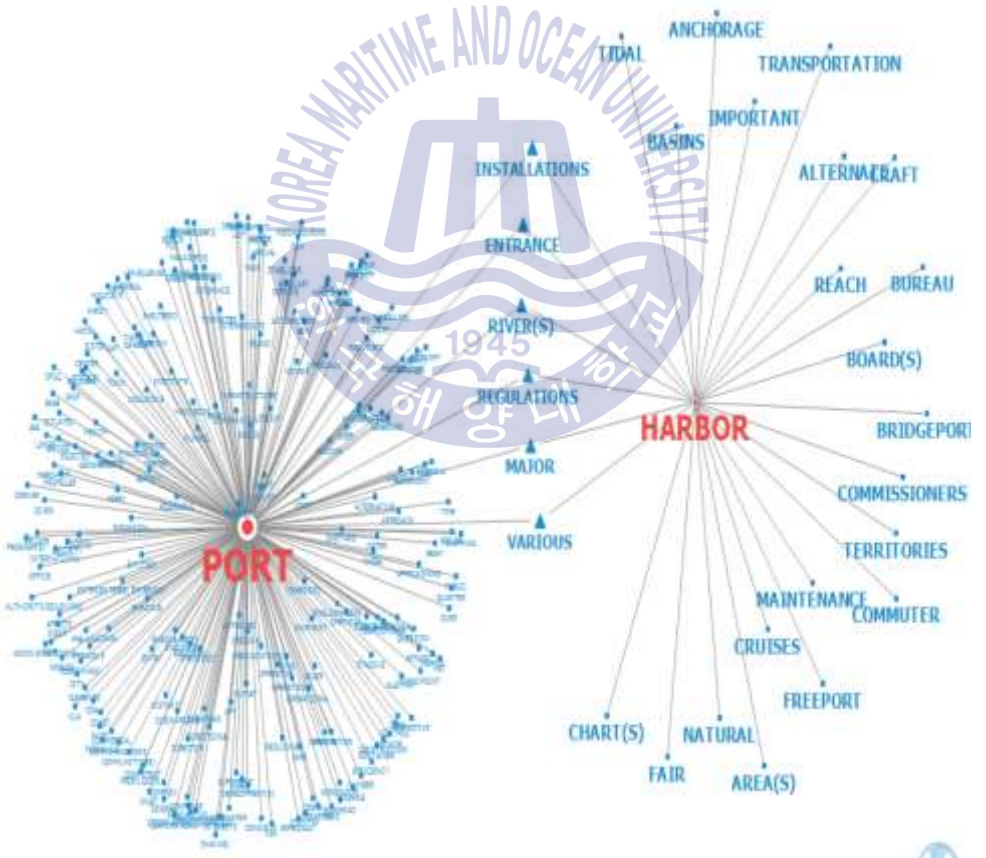


Figure 4.5. Collocation network of *port* and *harbor*.

4.4 Advantages and Limitations of Collocation Network Analysis

As a special visualization method, collocation network analysis can provide learners with a direct vision of the relationships between words. Compared with tradition collocation tables, learners are able to quickly identify the collocates and find the relationship between several node words through shared collocates. In addition, it is much easier for us to find the collocates exclusive to a specific word, hence helping us to understand the meaning specific to that word.

However, collocation network analysis has its limitations. First, for those which have many collocates in the corpus (*ship* vs. *vessel*), the network will be too complicated to analyze. In addition, for those frequently-used words such as *sea*, *safety*, *maritime*, etc. there are so many collocates existing in the corpus so that we can barely see which collocates are exclusive to them. At last, it cannot provide a general picture for students to cognitively identify and distinguish near-synonyms, especially when they shared the same collocates.

Chapter 5. Semantic Domain Network Analysis of Near-synonyms

In this section, I combined the network analysis with semantic domain analysis so as to provide a new viewpoint for the relationship between words and analyze the connection from a more general perspective, semantic domains, instead of only focusing on the lexis perspective collocation.

5.1 Comparison between Collocation and Semantic Domain Analysis

Compared with collocation analysis, semantic domain analysis clearly has its advantages, especially when involving a large quantity of collocates of one or several synonyms. Table 5.1 shows the total number of collocates, sub-semantic domains, and main semantic domains of the five pairs of near-synonyms.

Table 5.1
Total number of collocates, sub-semantic domains, main semantic domains of five groups of near-synonyms

Near-synonyms	Collocates	Sub-semantic domains	Main semantic domains
vessel	354	107	20
ship	317	211	19
maritime	125	81	16
marine	187	78	19
sea	134	63	18
ocean	29	20	10
safety	232	81	17
security	42	29	14
port	229	85	17
harbor	27	17	8

It can be seen that the number of collocates of each near-synonym is usually too large to observe. However, with semantic domain analysis, the number of nodes involved in the network decreases dramatically. And when it is scaled-down to the main semantic domains, the data quantity is within a controllable size to be analyzed or recognized by language learners. In addition, when the learners want to distinguish near-synonyms through words that only collocate with them, semantic domain analysis is also a preferred option.

Table 5.2 shows the total number of collocates, sub-semantic domains, and main semantic domains exclusive to each near-synonym.

Table 5.2
Total number of collocates, sub-semantic domains, main semantic domains exclusive to each near-synonym

Near-synonyms	Collocates	Sub-semantic domains	Main semantic domains
vessel	214	20	1
ship	177	124	0
maritime	85	37	1
marine	147	34	4
sea	115	45	8
ocean	10	2	0
safety	217	59	3
security	27	7	0
port	223	69	9
harbor	21	1	0

It can be seen there are many collocates exclusive to each near-synonym: it will be too difficult to generalize the cognitive concept for learners to understand. Therefore, semantic domain analysis is recommended since it offers a relatively controllable amount of data for analysis. However, what should be noticed here is that some words such as *sea*, *harbor*, *port*, etc. have no exclusive main semantic domains. In this case, it is highly recommended that learners refer to the sub-semantic domains for generalization of an exclusive concept.

An additional interesting observation from the total number of collocates, sub-semantic domains, and main semantic domains shared by each pair of near-synonym, as shown in Table 5.3: with the decreasing of cosine similarity, the number of shared collocates and semantic domains goes down accordingly, with

the exception of *safety* vs. *security* and *harbo*r vs. *port*. It is believed since these two pairs of words show a very low degree of similarity (below 0.2), they tend to have fewer shared collocates at the word-level. However, when it goes to a higher semantic level, the words find more common features, hence the 22>15 for *safety* vs. *security* and 16>6, 8>6 for *harbo*r vs. *port*.

Table 5.3
Total number of collocates, sub-semantic domains, main semantic domains shared by each pair of near-synonym

Near-synonyms	Cosine similarity	Collocates	Sub-semantic domains	Main semantic domains
vessel vs. *ship*	0.4485	140	87	19
maritime vs. *marine*	0.2820	40	44	15
sea vs. *ocean*	0.2200	19	18	10
safety vs. *security*	0.1520	15	22	14
harbor vs. *port*	0.0736	6	16	8

From the analysis above, it can be seen that semantic domain analysis has clear advantages over merely collocate analysis.

5.2 Semantic Domain Network Analysis of Exclusiveness

In this section, semantic domain network analysis will be offered on the five pairs of near-synonyms. Special attention will be paid to which semantic domains are exclusive to which near-synonym. Through this, learners will be able to

cognitively define the concepts usually collocated with each near-synonym, hence helping them to distinguish them.

5.2.1 *Ship* vs. *Vessel*

Figure 4.1 has shown a chaotic relationship between *ship* and *vessel* based on the collocation network analysis. However, the semantic domain network can overcome this weakness and display a much clearer connection between words. Figure 5.1 shows the semantic domain networks of *ship* and *vessel*. One is a network of more general main categories; the other is a network of more detailed subcategories. It can be seen that *vessel* has one F (Food and Farming) main category that is exclusive to it. Further information of exclusiveness can be seen from the sub-semantic domain network too. Therefore, through semantic domain network analysis, we are able to see clearly which semantic domain is exclusive for the near-synonyms. Then if we are able to know the correct semantic domain of a word, we may easily know whether it is usually collocated with *ship* or *vessel*.

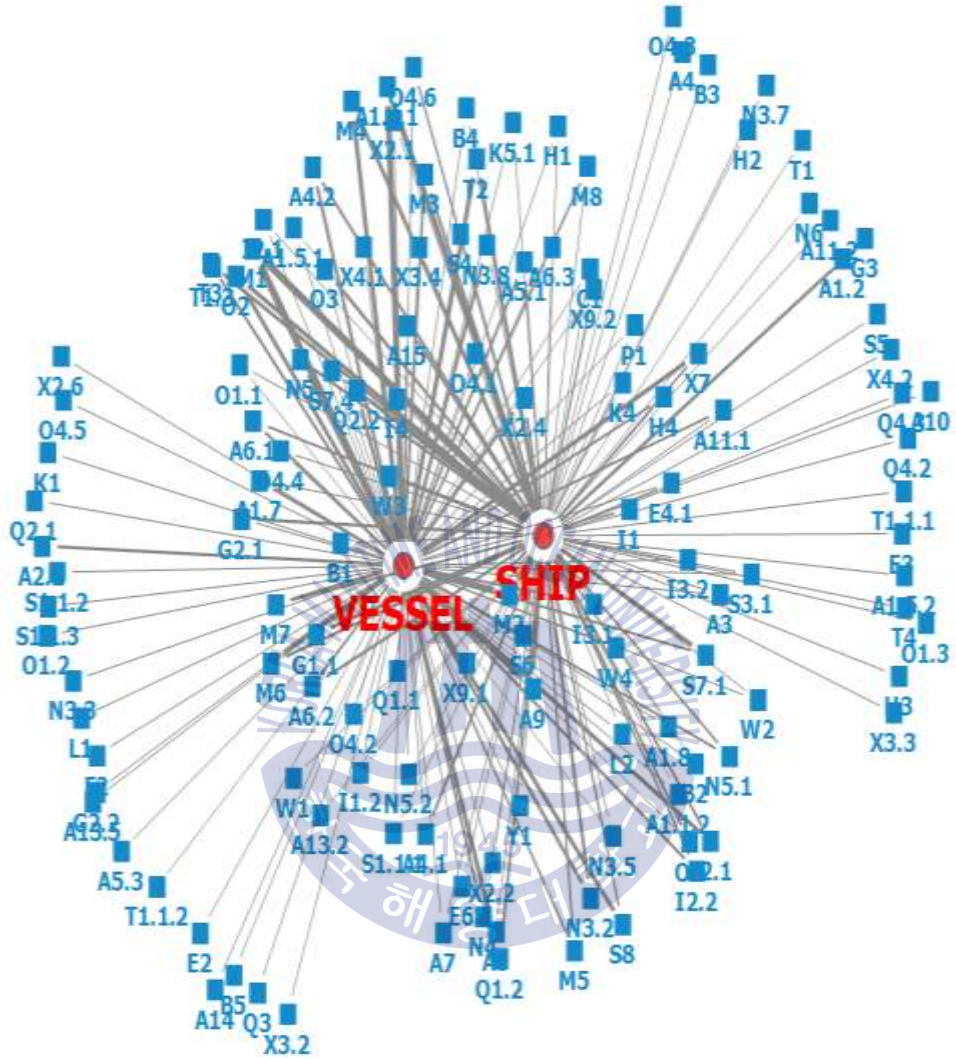


Figure 5.1a. Sub-semantic domain network of *ship* and *vessel*.

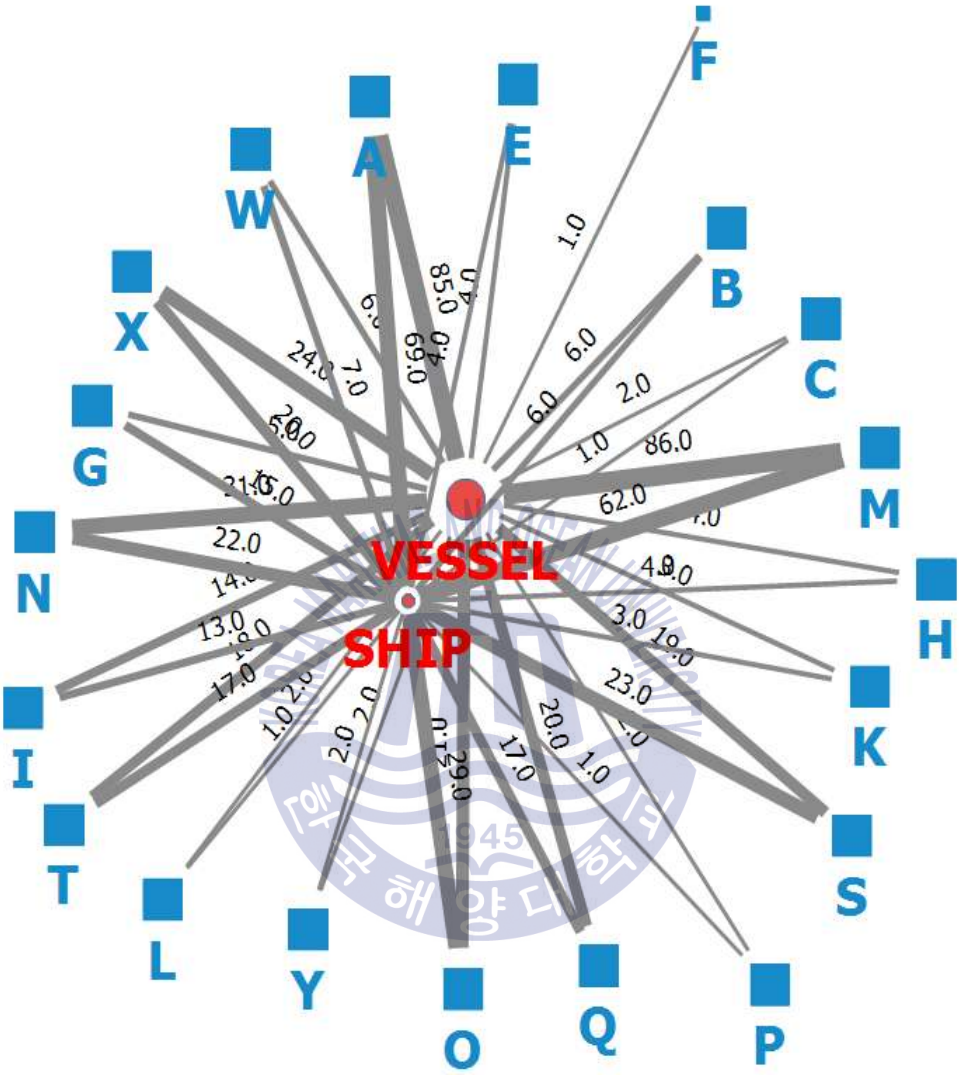


Figure 5.1b. Main Semantic domain network of *ship* and *vessel*.

5.2.2 *Maritime* vs. *Marine*

Figure 5.2 shows the semantic domain networks of *maritime* and *marine*. One is a network of more detailed subcategories; the other is a network of more general main categories. Comparing and combining the results from the two semantic domain networks, it can be seen that *maritime* has the following 19 exclusive semantic domains: G1.2 (Politics); E4.1 (Happy / Sad: Happy); S3.1 (Relationship: General); W1(The universe); X9.1 (Ability: Ability; intelligence); X4.1(Mental object: Conceptual object); T1.1.2 (Time: General: Present; simultaneous); O4.4 (Shape); A7 (Definite); W3 (Geographical terms); A2.2 (Affect: Cause / Connected); A5.3 (Evaluation: Accuracy); G2.1 (Law and order); I1.1 (Money: Affluence); S1.2.5 (Toughness; strong /week); T1.1.1 (Time: General: Past); K5.1 (S*port*s); Q4 (The Media); and A1.7 (Constraint). However, since *marine* has so many exclusive subcategories, it will be much better if we move to the main semantic domain, hence *marine* has L (Life and living things), H (Architecture, buildings, houses and the home), C (Arts and *Craft*s), Y (Science & Technology), i.e., four types of exclusive categories. Through semantic domain network analysis we are able to see clearly which semantic domain is exclusive for *maritime* or *marine*.

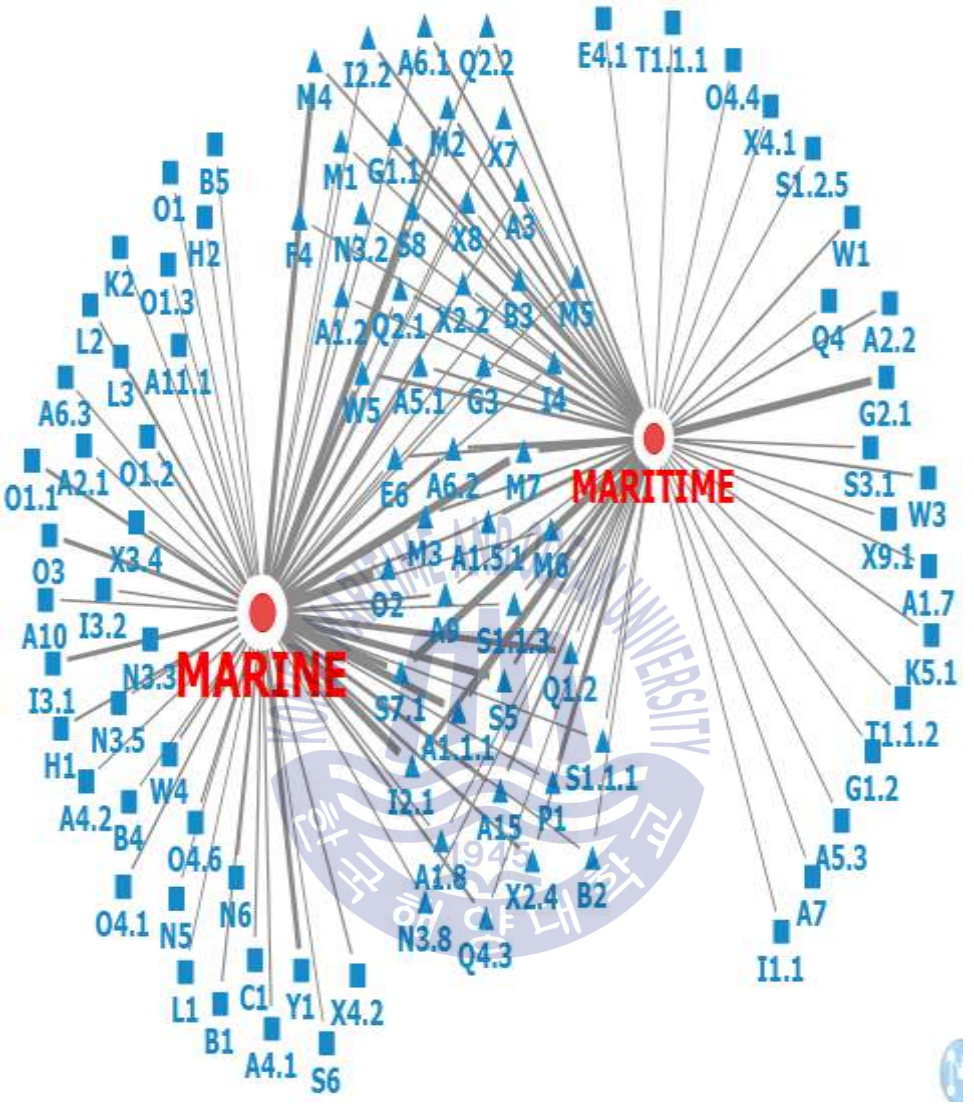


Figure 5.2a. Sub-semantic domain network of *maritime* and *marine*.

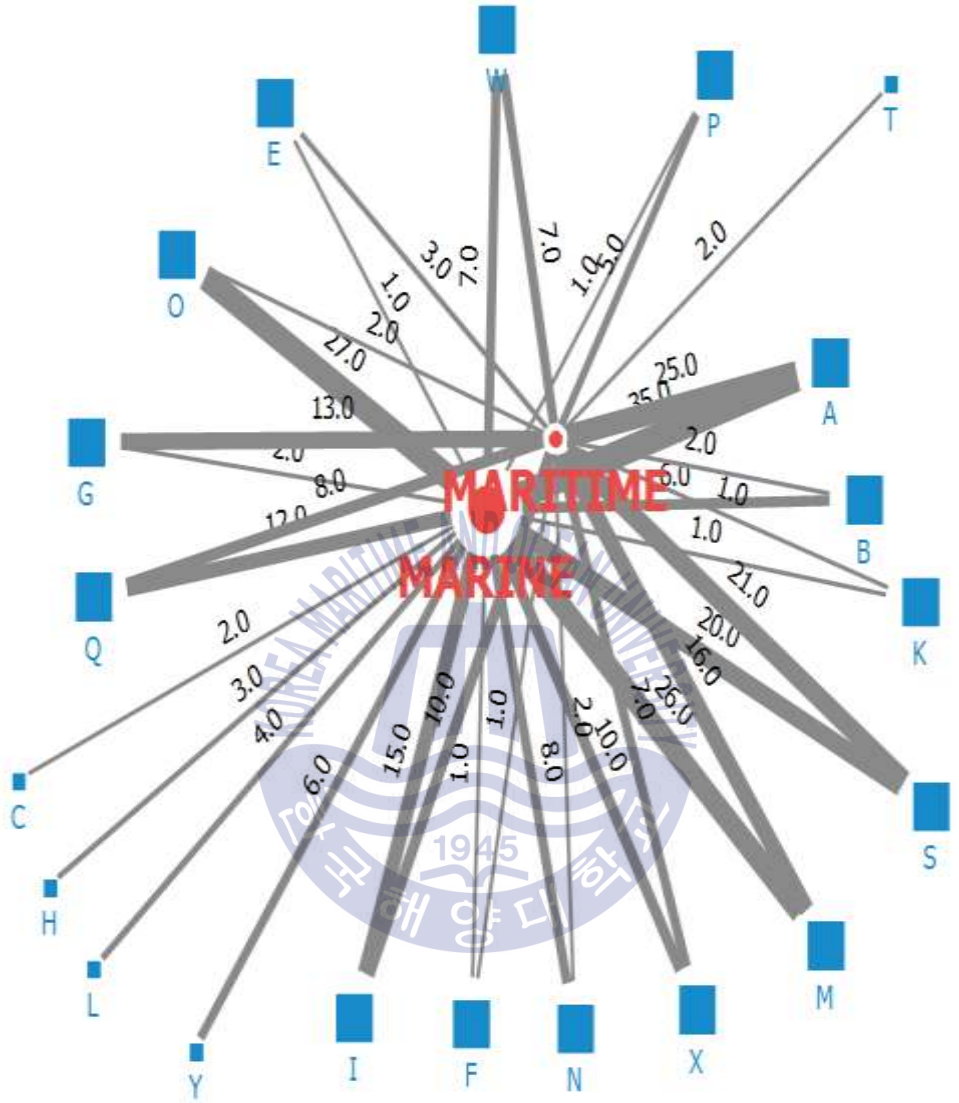


Figure 5.2b. Main Semantic domain network of *maritime* and *marine*.

5.2.3 *Sea* vs. *Ocean*

Figure 5.3 shows the semantic domain networks of *ocean* and *sea*. One is a network of more detailed subcategories, the other is a network of more general main categories. Comparing and combining the results from two semantic domain networks, it can be seen that *ocean* has the following 2 exclusive semantic domains, i.e., O1 (Substances and materials generally) and T1.1.2 (Time: General: Present; simultaneous). However, since *sea* has so many exclusive subcategories, it will be much better if we move to the main semantic domains. *Sea* has L (Life and living things); Y (Science & Technology); B (The body & the individual); E (Emotional actions, states and Processes); F (Food and farming); K (Entertainment Sports & Games); N (Numbers & Measurement); Q (Linguistic Actions, States, and Processes): eight types of exclusive categories. Through semantic domain network analysis, we are able to see clearly which semantic domain is exclusive for *ocean* or *sea*.

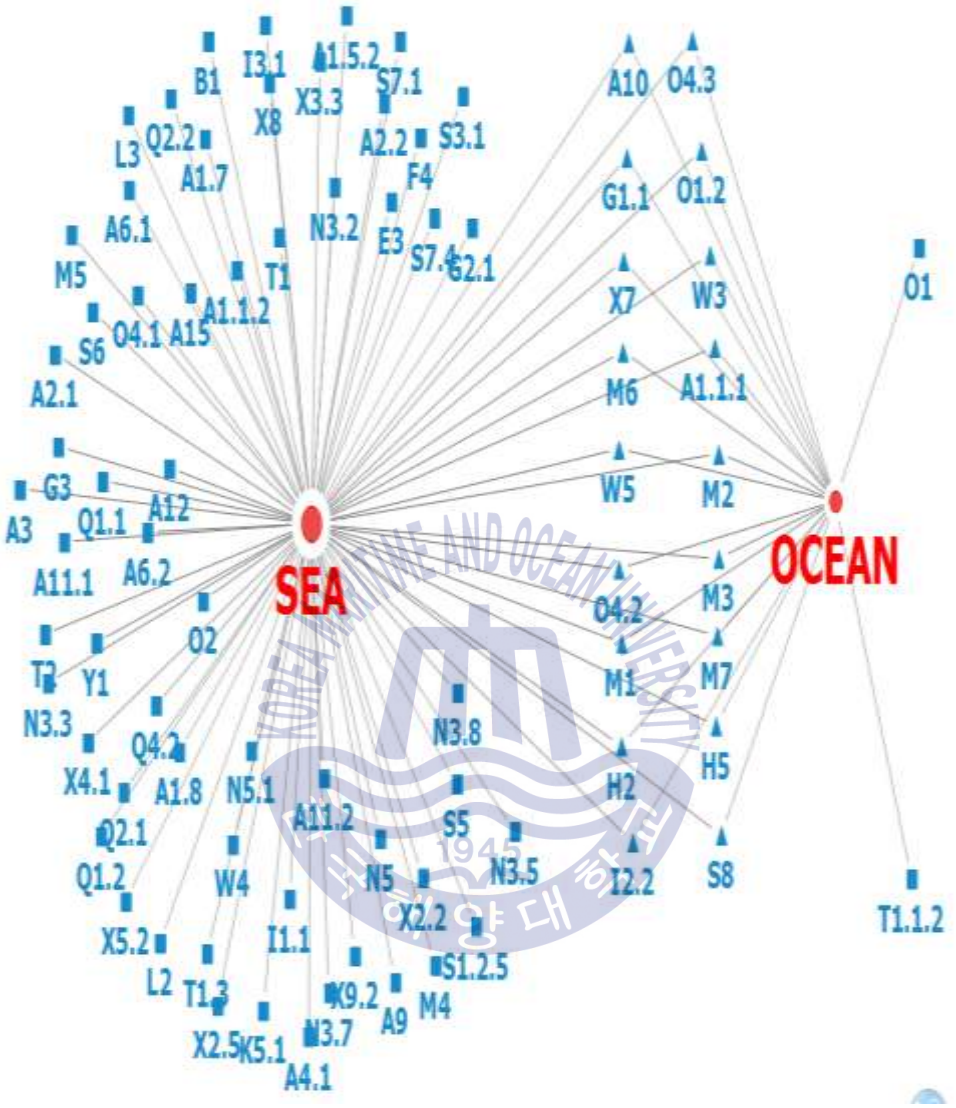


Figure 5.3a. Sub-semantic domain network of *sea* and *ocean*.

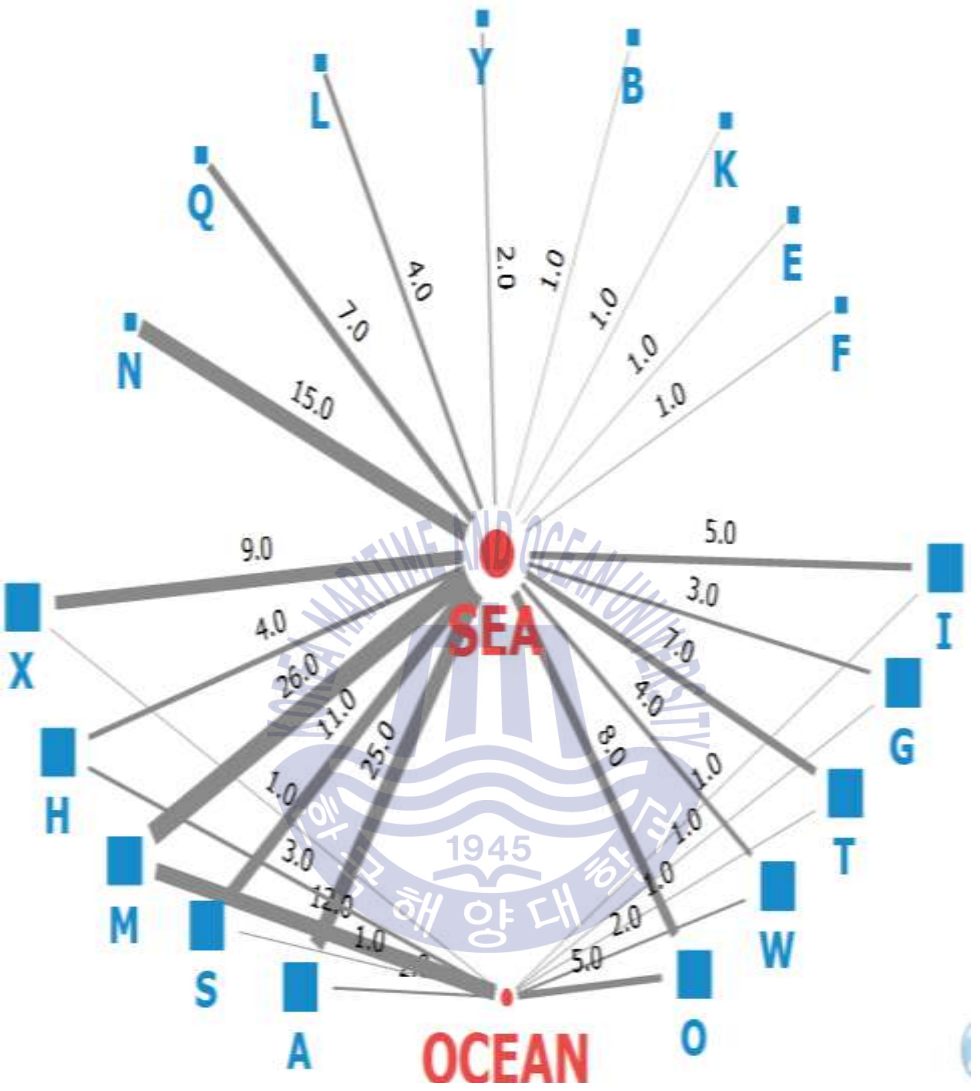


Figure 5.3b. Main Semantic domain network of *sea* and *ocean*.

5.2.4 *Safety* vs. *Security*

Figure 5.4 shows the semantic domain networks of *safety* and *security*. One is a network of more detailed subcategories; the other is a network of more general main categories. Compared and combined the results from two semantic domain networks, it can be seen that *security* has the following 7 exclusive semantic domains: H5 (Furniture and household fittings); N5.2 (Exceeding; waste); M5 (Movement / transportation: air); T4 (Time: Early / late); E4.2 (Happy / sad: Contentment); T1.3 (Time: Period); and I1 (Money general). However, since *safety* has so many exclusive subcategories, it will be much better if we move to the main semantic domains: *safety* has B (The body & the individual), C (Arts and crafts), and P (Education), three types of exclusive categories. Through semantic domain network analysis, we are able to see clearly which semantic domain is exclusive for *safety* or *security*.

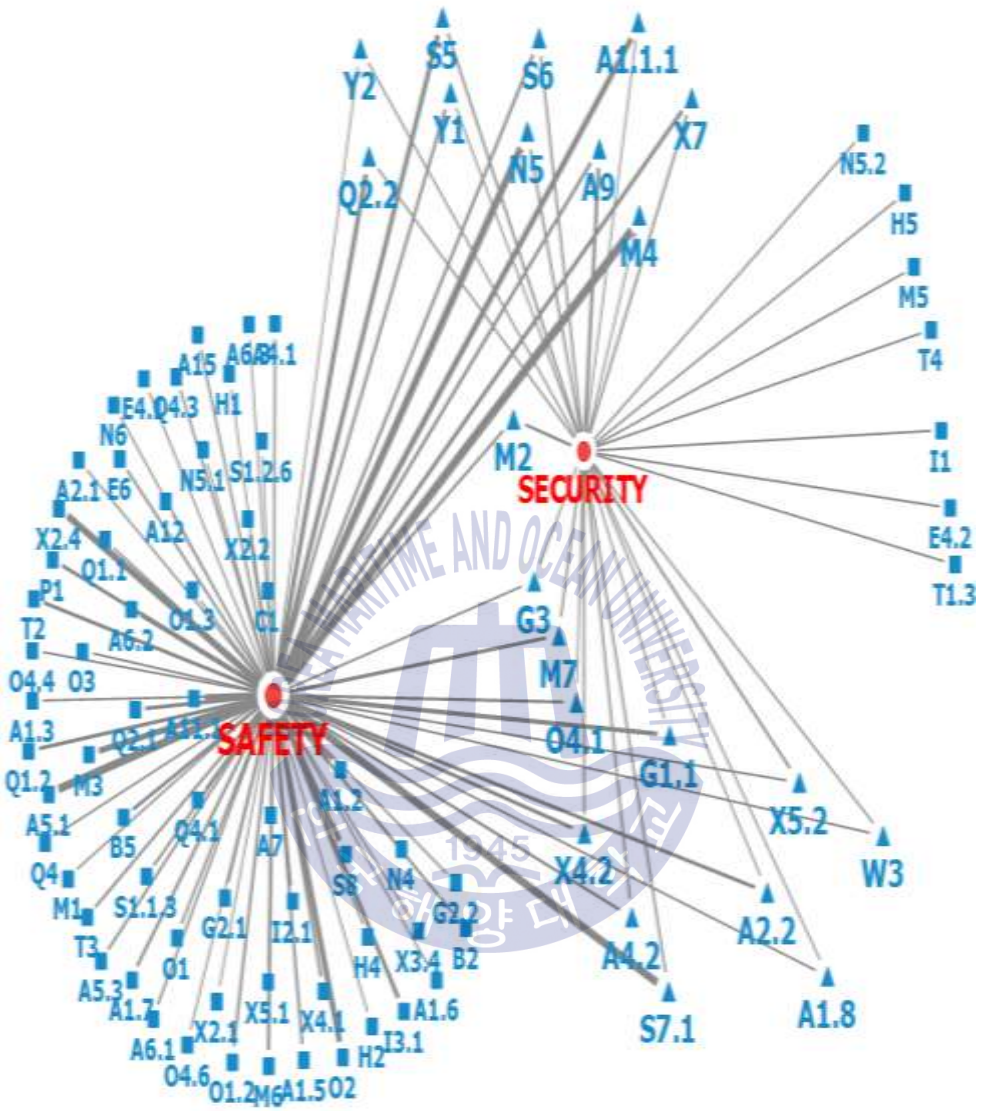


Figure 5.4a. Sub-semantic domain network of *safety* and *security*.

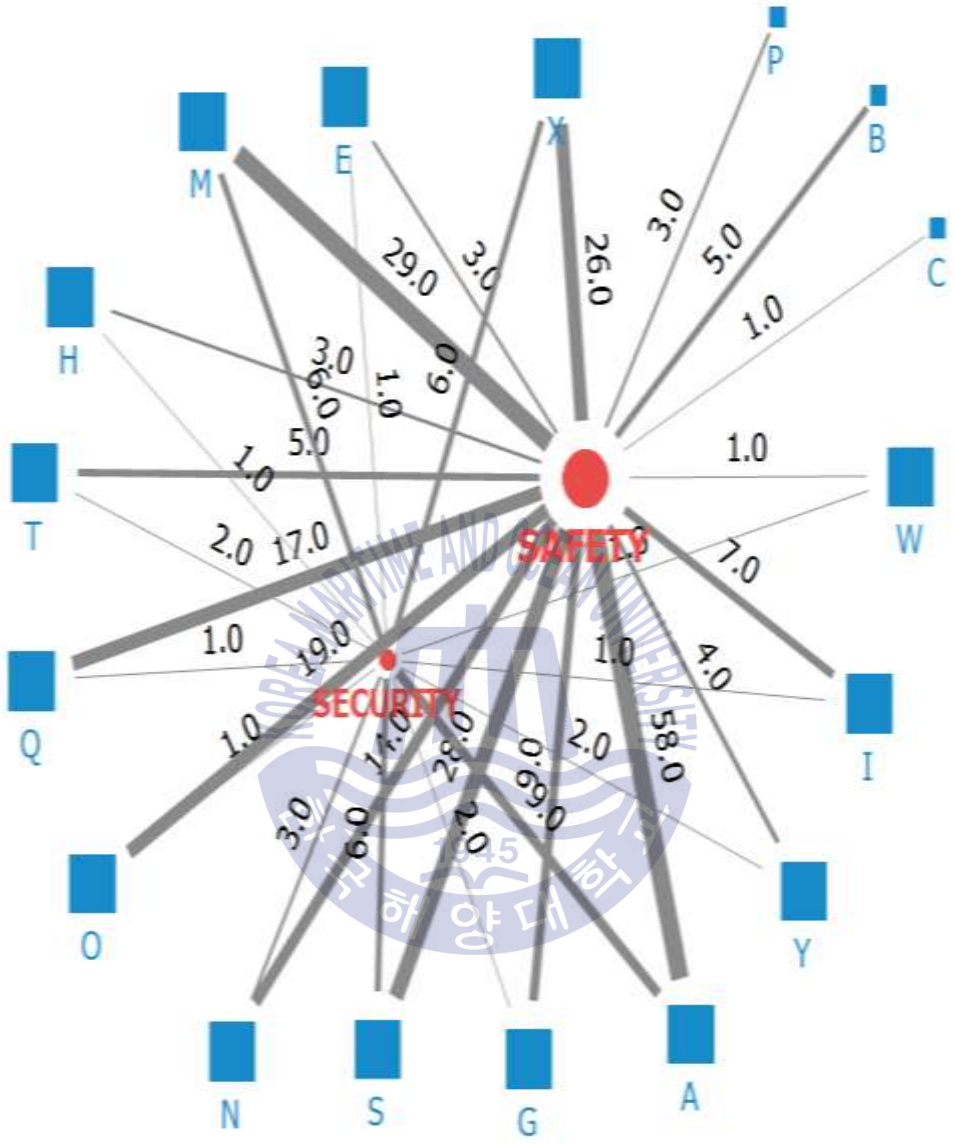


Figure 5.4b. Main Semantic domain network of *safety* and *security*.

5.2.5 *Port* vs. *Harbor*

Figure 5.5 shows the semantic domain networks of *dock* and *harbor*. One is a network of more detailed subcategories; the other is a network of more general main categories. Comparing and combining the results from these two semantic domain networks, it can be seen *harbor* has one G2.2 (General ethics) subcategory; *port* has K (Entertainment, S*port*s & Games), I (Money & Commerce), X (Psychological actions, states and processes), E (Emotional actions, states and processes), H (architecture, housing and the home), B (the body and the individual), L (Life and living things), T (time), and Y (Science & Technology), i.e., nine main exclusive semantic domains.

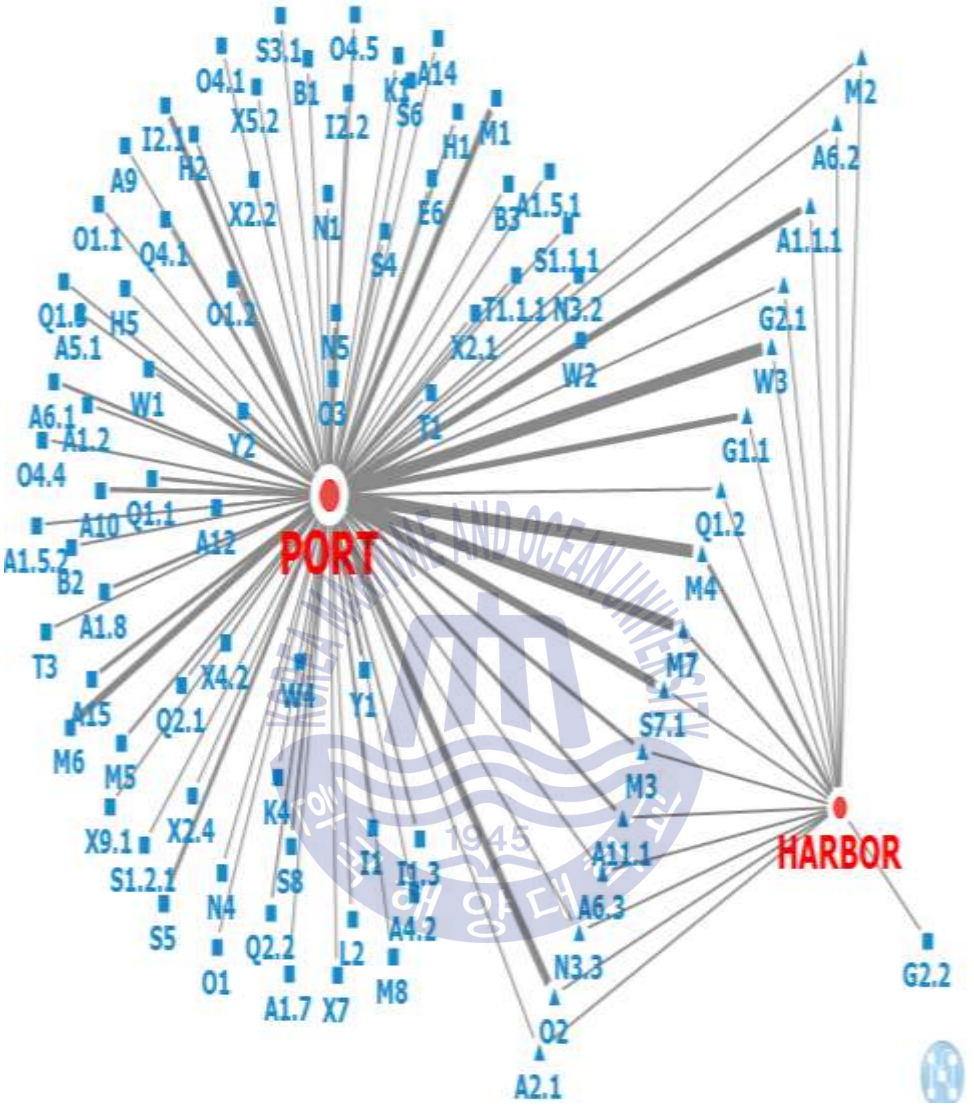


Figure 5.5a. Sub-semantic domain network of *port* and *harbor*.

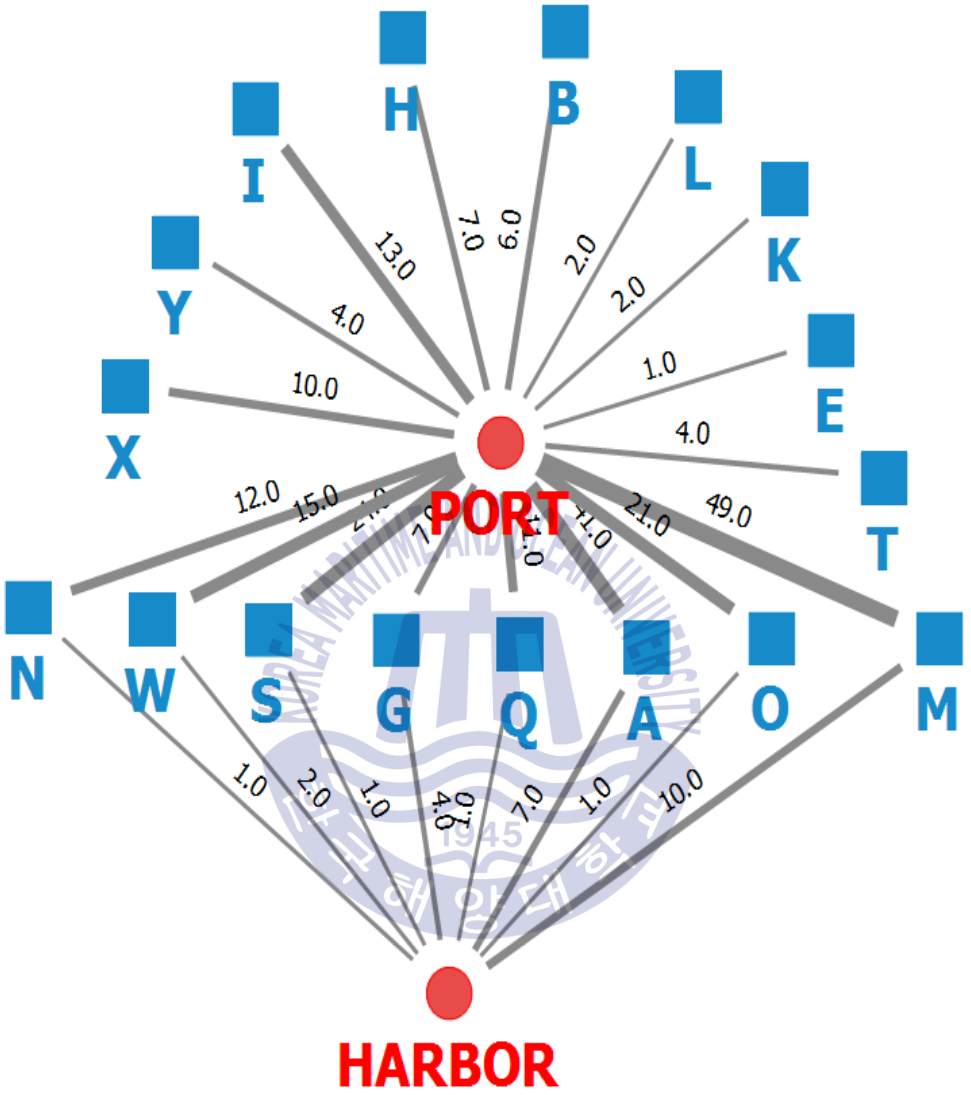


Figure 5.5b. Main Semantic domain network of *port* and *harbor*.

5.3 Analysis of Shared Semantic Domains

Section 5.2 has shown the advantages of using semantic domain network to distinguish collocates exclusive to near-synonyms. With this method, learners are able to distinguish near-synonyms through cognitive concepts. Nevertheless, what if a word belongs to a semantic domain shared by both near-synonyms? It is recommended that both the number of collocates and the frequency of collocates in that main semantic domain should be considered.

There will be three situations. If a word A belongs to a semantic domain shared by both near-synonyms B and C, then in that semantic domain,

(i) if B's number and frequency of collocates is higher than C's, it is recommended that A collocates with B.

(ii) if the number and the frequency of collocates of B and C are the same, sub-semantic domains should be considered to make the decision.

(iii) if B's number of collocates is higher than C's, while B's frequency of collocate is lower than C's, word-level should be considered to make the decision.

Two examples below can work as an illustration to this question.

Table 5.4
Main semantic domains shared by *safety* and *security*

Situation	Main semantic domains	*safety*		*security*	
		Number of collocates	Frequency of collocates	Number of collocates	Frequency of collocates
i	A	58	384	9	39
	E	3	17	1	2
	G	9	173	2	8
	H	3	21	1	3
	M	29	652	6	46
	N	14	57	3	8
	O	19	165	1	2
	Q	17	102	1	5
	S	28	607	6	30
	T	5	30	2	11
	X	26	388	6	31
	Y	4	95	2	16
ii	W	1	255	1	11
iii	I	7	41	1	66

Table 5.4 shows the main semantic domains shared by *safety* and *security*. For the first situation, *safety* tends to collocate more frequently in most shared semantic domains. This illustrates that in Maritime English, users prefer to use *safety* in their language usage. For the second situation, in the main semantic domain W, both near-synonyms have one word *maritime* (W3) collocates. However, the sub-semantic domain W3 tends to collocate with *safety* with *maritime* at a frequency of 255 compared to 11 for *security*. In the third situation, in the main semantic domain I, *security* only collocates with *financial* which belongs to I1 (Money Generally), but *safety* tends to collocate with I2.1 (Business Generally) words such

as *office* and *company* and I3.1 (Work and Employment Generally) words such as *personnel*, *job*, *worker*, *occupational*, and *fisherman*. We can easily conclude that words related to money tend to collocate *security*, while other words related to business, work and employment usually collocate with *safety* in the maritime field.

Table 5.5
Main semantic domains shared by MARITIME and MARINE

Situation	Main semantic domains	*maritime*		*marine*	
		Number of collocates	Frequency of collocates	Number of collocates	Frequency of collocates
i	A	25	401	35	407
	B	2	11	6	72
	E	3	34	1	3
	G	13	197	2	5
	N	2	13	8	29
	O	2	8	27	176
	P	5	34	1	2
	X	7	25	10	38
ii	F	1	3	1	5
	K	1	4	1	8
	W	7	45	7	414
iii	I	10	114	15	81
	M	16	404	26	232
	Q	8	103	12	87
	S	21	94	20	118

Table 5.5 shows the main semantic domains shared by *maritime* and *marine*. For the first situation, *marine* tends to collocate with words in semantic domains A, B, N, O, and X; while *maritime* tends to collocate with words in semantic domains E, G, and P. For the second situation, consider the main semantic domain W as an

example. It can be seen from Table 5.6 that both *maritime* and *marine* collocate with *pollution*, *environment* and *environmental*, however, these words tend to prefer *marine* with higher frequency (74>2, 59>2, 267>4). Therefore it can be concluded that W5 (Green Issues) tends to collocate with *marine*. In addition, W4 (Weather) tends to collocate with *marine* too. While W1 (The Universe) and W3 (Geographical Terms) tend to collocate with *maritime*. For the third situation, take main semantic domain Q as an example. It can be seen that since both near-synonyms have collocates in sub-semantic domains such as Q1.2, Q2.1 and Q2.2, it has to be decided on the specific word. In addition, it seems that Q4.3 tends to collocate with *marine*.

Table 5.6
Sub-semantic domains of collocates of *maritime* and *marine*

maritime			*marine*		
Collocates	Sub-domains	Freq	Collocates	Sub-domains	Freq
POLLUTION	W5	2	POLLUTION	W5	74
ENVIRONMENTAL	W5	2	ENVIRONMENTAL	W5	59
ENVIRONMENT	W5	4	ENVIRONMENT	W5	267
SATELLITE	W1	2	CONSERVATION	W5	2
WORLD	W1	7	ECOSYSTEM	W5	3
COASTAL	W3	3	FLOODING	W4	3
GLOBAL	W3	25	WEATHER	W4	6
DOCUMENT	Q1.2	2	ARCHIVE	Q1.2	3
COMMUNICATIONS	Q2.1	2	CHARTS	Q1.2	4
CONSULTATIVE	Q2.1	16	APPLICATION(S)	Q1.2	3
CLAIM	Q2.2	25	CHARTERS	Q1.2	2
DECLARATION	Q2.2	14	ADDRESSING	Q1.2	2
CLAIMS	Q2.2	16	RESPONSE	Q2.1	9
MEDIA	Q4	3	COMMUNICATIONS	Q2.1	15

RADIO(S)	Q4.3	25	REPORTABLE	Q2.2	2
			PROPOSED	Q2.2	2
			RADIO(S)	Q4.3	41
			RADIOTELEGRAPH	Q4.3	2
			RADIOTELEPHONE	Q4.3	2

5.4 Advantages and Limitations of Semantic Domain Network Analysis

Compared with collocation network analysis, semantic domain network analysis clearly has its advantages, especially when involving a large number of collocates of one or several synonyms. Semantic domain network analysis can clearly show the exclusive semantic domains and shared semantic domains for each node. With this method, Maritime English learners are able to view the collocational relationship between words more clearly. If collocation network is an in-depth understanding of words, then semantic domain network is a more general and comprehensive understanding of word relationships.

Of course, semantic domain network has its limitation as well. It requires the analyzer to understand how to differentiate different semantic domains with the help of Wmatrix. However, considering the fact that Wmatrix is a handy tool with more than 90% confidence rate, this problem seems easy to solve.

Chapter 6. Conclusion

6.1 Summary

Language can be viewed as a complex network if it is presented as a system of interacting linguistic units. Network analysis provides mechanisms that can reveal new patterns in a complex structure and can thus be applied to the study of the patterns in language structures. This, in turn, may contribute to a better understanding of the organization and evolution of a language.

The objective of the work in this study was to answer the following three questions. First, what are the differences and similarities between different near-synonyms in English? Second, can collocation network analysis provide a new level to explain the distinction of near-synonyms from a micro-scopic perspective? Third, is semantic domain network analysis useful to distinguish one near-synonym from the other at the macro-scopic level? In pursuit of the research questions, I first illustrated how the idea of incorporating collocate in corpus linguistics, Maritime English, near-synonyms, semantic domains and language network were studied. Then important concepts such as Maritime English, English for Specific Purposes (ESP), corpus linguistics, synonymy, collocation, language network analysis and semantic domains were introduced. Third, I compiled a 2.5 million word Maritime English Corpus (MEC) and proposed a new method of tagging English multi-word compounds, discussing the comparison of with and without multi-word compounds with regard to tokens, types, STTR and mean word length. Fourth, I examined

collocates of five groups of near-synonyms, i.e., *ship* vs. *vessel*, *maritime* vs. *marine*, *ocean* vs. *sea*, *safety* vs. *security*, and *harbor* vs. *port*. drawing data through WordSmith 6.0, tagging semantic domains in Wmatrix 3.0 and conducting network analyses using NetMiner 4.0, expecting to find a new and easy way to distinguish near-synonyms.

With the results and discussions in Chapter 4 and Chapter 5, I was able to answer my three research questions.

For the first question, the traditional way of distinguishing near-synonyms through dictionary definitions cannot be applied to the maritime industry. Generally speaking, maritime near-synonyms showed clear preference to specific collocates. Therefore a new perspective is needed to view the behaviors of maritime words.

For the second question, as a special visualization method, collocation network analysis can provide learners with a direct vision of the relationships between words. Compared with traditional collocation tables, learners are able to more quickly identify the collocate and find the relationship between several node words. In addition, it is much easier for learners to find the collocates exclusive to a specific word, hereby helping them to understand the meaning specific to that word. However, it has its limitations. For those which have many collocates in the corpus (i.e., *ship* vs. *vessel*), the network will be too complex to analyze. In addition, for those frequently-used words such as *sea*, *safety*, *maritime*, etc., there are so many

collocates existing in the corpus so that we can barely see which collocates are exclusive to them.

For the third question, semantic domain network analysis has proven to be an effective way to show the differences among near-synonyms. Semantic domain information discussed in this thesis can help beginning learners of L1 and L2 understand the specific meanings of near-synonyms. This method can contribute to understanding usages by showing that a large amount of collocates can be easily analysed through the visualization of semantic domain networks, which provides a better explanation than corpus description analysis.

If we accept that a collocation network shows us relationships of words, then semantic domain networks are able to give us guidance cognitively: when a person has a specific word, how he should process it in his mind and therefore find the more appropriate synonym to collocate with. When the learners want to distinguish near-synonyms through words that only collocate with them (i.e., through the exclusiveness), main semantic domain network analysis shows us the exclusive domains to a certain near-synonym. Therefore defines the concepts exclusive to that near-synonym. When the learners want to distinguish near-synonyms through a semantic domain shared by both near-synonyms, it is recommended that both the number of collocates and the frequency of collocates in that main semantic domain should be considered. There are three situations. If a word A belongs to a semantic domain shared by both near-synonyms B and C, then in that semantic domain, (i) if

B's number and frequency of collocates is higher than C's, it is recommended that A collocates with B; (ii) if the number and the frequency of collocates of B and C are the same, sub-semantic domain should be considered to make the decision; (iii) if B's number of collocate is higher than C's, while B's frequency of collocate is lower than C's, the word-level should be considered to make the decision.

Generally speaking, we see only the vein of a tree leaf through the traditional way of sentence-level analysis. We see the full leaf through collocation network analysis. We see the tree, even the whole forest, through semantic domain network analysis.

6.2 Limitations and Implications

Although I believe that the application of network analysis to the study of language has enormous potential, it is also worthwhile to point out its limits. First of all, there are other types of universal trends in language which cannot be explained by this approach, either because they are not related to statistical features of language networks or because they are due to the other causal factors. Second, as Simon (1955, p. 440) has already argued, the ubiquity of certain statistical distributions, such as MI3 and the mechanisms that can generate them may lead to the fact that they do not completely capture the fine-grained uniqueness of language. Third, the corpus built in this thesis is limited to copyright, time and resources. With a larger corpus, the results and conclusions generated in this thesis

may be modified. After all, there is no absolute answer to a language phenomenon. I have proposed a framework of study that helps to show the differences of near-synonyms. However, it would be interesting to go deeper into other areas. For example, how do language networks grow through language acquisition? Or are there statistical differences among networks for different languages? Or how do we explain the network through the angle of cognitive linguistics?

References

Albrechtsen, E. (2002). A generic comparison of industrial safety and information security (Term paper in the PhD course “Risk and Vulnerability”). Trondheim: Norwegian University of Science and Technology.

Albert, R., Jeong, H., & Barabasi, A. (1999). Diameter of the World Wide Web. *Nature*, *401*, 130-131.

Alonso, A., Millon, C., & Williams, G. (2011). Collocation networks and their application to an e-advanced learner’s dictionary of verbs in science. In *Proceedings of New Applications for New Users* (pp.12-22). Electronic Lexicography in the 21st Century, Bled, Slovenia: Institute for Applied Slovene Studies.

Amancio, D., Fabbri, R., Oliveira, O., Nunes, M., & da Fontoura Costa, L. (2011). Opinion discrimination using complex network features. In L. da Fontoura Costa, A. Evsukoff, G. Mangioni, & R. Menezes (Eds.), *Complex Networks. Communications in Computer and Information Science*, *116*, 154-162). Berlin, Heidelberg: Springer.

Anderson, J. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, *7*, 1-16.

Baker, C., Fillmore, C., & Lowe, J. (1998). The Berkeley FrameNet Project. In *Proceedings of the 17th International Conference on Computational linguistics* (pp. 86-90). New Jersey, USA: COLNG 1998.

Barabasi, A. (2002). *Linked: The new science of networks*. Cambridge: Perseus.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, *8*(4), 243-257.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biemann, C., Roos, S., & Weihe, K. (2012). Quantifying semantics using complex network analysis. In *Proceedings of the 24th International Conference on Computational Linguistics* (pp. 263-278). Mumbai, India: Indian Institute of Technology Bombay.

Bocanegra-Valle, A. (2012). Maritime English. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3570-3583). Hoboken: Wiley-Blackwell.

Boholm, M., Möller, N., & Hansson, S. (2015). The concept of risk, safety and security: Applications in everyday language. *Risk Analysis*, *36*(2), 320-338.

Brezina, V., McEnery, T., & Wattam, S., (2015). Collocations in context. *International Journal of Corpus Linguistics*, *20*(2), 139–173.

British National Corpus [Data set]. (n.d.). Retrieved from http://www.natcorp.ox.ac.uk/

Burns, A., McDermid, J., & Dobson, J. (1992). On the meaning of safety and security. *Computer Journal*, *35*(1), 3-15.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. New York: Praeger.

Choudhury, M. & Mukherjee, A. (2009). The structure and dynamics of linguistic networks. In N. Ganguly, A. Deutsch, & A. Mukherjee (Eds.), *Dynamics on and of complex networks: Applications Biology, Computer Science, Economics, and the Social Sciences* (pp. 145–166). Berlin: Springer.

Christakis, N., & Fowler, J. (2010). *Connected: The surprising power of our social networks and how they shape our lives*. New York: Little Brown and Company.

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22-29.

Clark, E. (1992). Conventionality and contrast: Pragmatic principles with lexical consequences. In E. Kittay & A. Lehrer (Eds.), *Frames, fields, and contrasts:*

New essays in semantic and lexical organization (pp. 171-188). Hillsdale: Lawrence Erlbaum Associates.

Conzett, J. (1997). Integrating collocation into a reading and writing course. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 70-87). Cambridge: Cambridge University Press.

Corpus of Contemporary American English [Data set]. (n.d.). Retrieved from http://corpus.byu.edu/coca/

Corpus of Historical American English [Data set]. (n.d.). Retrieved from http://corpus.byu.edu/coha/

Corpus of Global Web-Based English [Data set]. (n.d.). Retrieved from http://corpus.byu.edu/glowbe/

Cruse, D. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. (2000). *Meaning in language: An introduction to semantics and pragmatics*. New York: Oxford University Press.

Cruse, D. (2006). *A glossary of semantics and pragmatics*. Edinburgh: Edinburgh University Press.

Crystal, D. (2007). *Words words words*. Oxford: Oxford University Press.

Cyram. (2016). Netminer 4.0. Seoul: Cyram Co. Ltd.

Daille, B. (1995). Combined approach for terminology extraction: Lexical statistics and linguistic filtering. *UCREL Technical Papers 5*. Lancaster: University of Lancaster.

Dutertre B., & Stavridou V. (1998). From security to safety and back. In *Proceedings of the Computer Security, Dependability, and Assurance: From Needs to Solutions* (pp. 182-195).. York, UK: IEEE Computer Society.

Eguiluz, V., Cecchi, G., Chialvo, D., Baliki, M., & Apkarian, A. (2005). Scale-free structure of brain functional networks. *Physical Review Letters*, *94*(1), 018102.

Evert, S. (2004). *The statistics of word co-occurrences: Word pairs and collocations* (Ph.D. dissertation, University of Stuttgart, Germany).

Faber, P., León Araúz, P., & Reimerink, A. (2014). Representing environmental knowledge in EcoLexicon. In E. Bárcena, T. Read, & J. Arus (Eds.), *Languages for specific purposes in the digital era* (pp.267-301). Gewerbestrasse: Springer International Publishing.

Fellbaum, C. (1987). *Wordnet: An electronic lexical database*. Cambridge: MIT Press.

Finegan, E. (2007). *English: Its structure and use*. Boston: Heinle & Heinle.

Firth, J. (1957). A Synopsis of linguistic theory 1930-1955. In F. Palmer (Ed.), *Selected papers of J. R. Firth 1952-1959* (pp. 168-205). London: Longman.

Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic discourse* (pp. 95-114). London: Pearson.

Girvan, M., & Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 7821-7826.

Goodman, N. (1952). On likeness of meaning. In L. Linsky (Ed.), *Semantics and the philosophy of language* (pp. 67-74). Urbana: University of Illinois Press.

Gove, P. (1984). *Merriam-Webster's new dictionary of synonyms*. Massachusetts: G. & C. Merriam Co.

Greenbaum, S.（1974). Some verb-intensifier collocations in American and British English. *American Speech, 49*, 79-89.

Gries, S., & Dagmar S. (2009). Behavioral profiles: A corpus-based approach towards cognitive semantic analysis. In V. Evans, & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 57-75). Amsterdam: John Benjamins.

Gries, S., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonym. *ICAME Journal*, *34*, 121-150.

Gries, S. (2013). 50-something years of work on collocations: What is or should be next .... *International Journal of Corpus Linguistics*, *18*(1), 137-165.

Halliday, M. (1961). Categories of a theory of grammar. *Word*, *17*(3), 241-292.

Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Harley, H. (2006). *English words: A linguistic introduction*. Oxford: Blackwell.

Hayakawa, S. (1994). *Choose the right word: A contemporary guide to selecting the precise word for every situation*. New York: Harper Collins Publishers.

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research, 37* (4), 401-411.

Heyer, G., Lauter, M., Quasthoff, U., Wittig, T., & Wolff, C. (2001). Learning relations using collocations. In *Proceedings of the IJCAI Workshop on Ontology Learning*. Seattle: The 17th International Joint Conference on Artificial Intelligence.

Hirst, G. (1995). Near-synonymy and the structure of lexical knowledge. In *Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity* (pp. 51-56). Papers from the 1995 AAI symposium, California, USA: Stanford University.

Hoey, M. (2005). *Lexical priming. A new theory of words and language*. London: Routledge.

Honston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hornby, A. (2002). *Oxford advanced learner's English-Chinese dictionary (Extended 4th ed.)*. Oxford: Oxford University Press

Huxley, A. (1932). *Brave new world*. New York: Harper & Bros.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. Cambridge: Cambridge University Press.

Inkpen, D., & Hirst, G. (2006). Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics*, *32*(2), 223-262.

Jackson, H., & Amvela, E. (2000). *Words, meaning, and vocabulary*. London: Cassell.

Jackson, H. (1988). *Words and their meaning*. London: Longman.

Jhang, S., & Lee, S. (2013). Visualization of collocational networks: Maritime English keywords. *Language Research*, *49*(3), 781-802.

Jhang, S., & Lee, S. (2014). Collocational patterns of maritime English vocabulary from a statistical point of view. *Journal of Language Sciences*, *21*(1), 173-193.

Johns, A. M. (2013). The history of English for specific purposes research. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 5-30). Chichester: John Wiley & Sons, Ltd.

Johnson, B. (1995). Some features of maritime telex service communication. *English for Specific Purposes*, *14*(2), 127-136.

Jones, K. S. (1971). Automatic keyword classification for information retrieval. *The Library Quarterly: Information, Community, Policy*, *41*(4), 33-98.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, *31*, 7-15.

Ke, J., & Yao, Y. (2008). Analysing language development from a network approach. *Journal of Quantitative Linguistics*, *15*(1), 70-99.

Ke, J. (2007). Complex networks and human language. *Computer Science*, arXiv:cs.CL/0701135 [arXiv].

Koester, A. (2010). Building small specialized corpora. In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp.66-79). London: Routledge Handbooks.

Lee, S. (2016). *Network analysis of Maritime English Corpus with multi-word compounds: Keyword networks and collocation networks* (Unpublished doctoral dissertation). Korea Maritime and Ocean University, Busan, Korea.

Lewis, M. (1997). *Implementing the Lexical Approach: Putting theory into practice*. Hove: Language Teaching Publications.

Liang, W., Shi, Y., Tse, C., Liu, J., Wang, Y., & Cui, X. (2009). Comparison of co-occurrence networks of the Chinese and English languages. *Physica A Statistical Mechanics & Its Application*, *388*(23), 4901-4909.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

Lu, W., Lee, S., & Jhang, S. (2017). Keyness in maritime institutional law text. *Linguistic Research*, *34*(1), 51-76.

Lyons, J. (1981). *Language, meaning and context*. London: Fontana.

Mackin, R. (1978). On collocations: Words shall be known by the company they keep. In Strevens, P. (Ed.), *In honor of A. S. Hornby* (pp. 149-165). Oxford: Oxford University Press.

MacLaury, R. (2002). Introducing vantage theory. *Language Sciences*, *24*(5–6), 493-536.

Macmillan Publishers Ltd. (2002). *Macmillan English dictionary for advanced learners*. London: Macmillan Publishers Ltd.

Magnusson C., & Vanharanta, H. (2003). Visualizing sequences of texts using collocation networks. In P. Perner & A. Rosenfeld (Eds.), *Machine learning and data mining in pattern recognition* (pp. 276-283). Berlin: Springer.

Masucci, A., & Rodgers, G. (2006). Network properties of written human language. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, *74*, 1-8.

McArthur, T. (1981). *Longman lexicon of contemporary English*. London: Longman.

McEnery, T. (2006). *Swearing in English: Bad language, purity and power from 1586 to the present*. London: Routledge.

McEnery, T. & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice.* Cambridge: Cambridge University Press Textbooks.

Melville, H. (1851). Moby-Dick. New York: Harper & Bros.

Michigan Corpus of Academic Spoken English (MICASE) [data set]. (n.d.). Retrieved from http://quod.lib.umich.edu/m/micase/

Microsoft. (2011). Microsoft Visual FoxPro 6.0. Microsoft Corporation.

Miller, G. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39-41.

Montoya, J., & Sol, R. (2002). Small world patterns in food webs. Journal of *Theoretical Biology*, *214*(3), 405-412.

Moretti, F. (2013). *Distant reading*. New York: Verso.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Neely, J. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General*, *106*, 226-54.

Neely, J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). New York: Lawrence Erlbaum.

Newman, M. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review, 64,* 01613, 1-8.

Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press.

Northcott, J. (2013). Legal English. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 213-226). Chichester: John Wiley & Sons, Ltd.

Now Corpus (News on the Web) [Data set]. (n.d.). Retrieved from http://corpus.byu.edu/now/

Oakes, M. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Olohan M (2004) *Introducing corpora in translation studies*. London: Routledge.

Ottenheimer, H. J. (2006). *The anthropology of language: An introduction to linguistic anthropology*. Belmont: Wadsworth.

Palmer, F. (1981). *Semantics*. Cambridge: Cambridge University Press.

Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: The Institute for Research in English Teaching.

Pearson Longman. (2008). *Longman dictionary of contemporary English (5th ed.)*. London: Pearson Longman.

Phillips, M. (1983). *Lexical macrostructure in science text* (Unpublished doctoral dissertation). University of Birmingham, UK.

Piètre-Cambacédès L., & Chaudet C. (2010). The SEMA referential framework: Avoiding ambiguities in the terms "security" and "safety." *International Journal of Critical Infrastructure Protection*, *3*(2), 55-66.

Plag, I., Braun, M., Lappe, S., & Schramm, M. (2009). *Introduction to English linguistics*. Berlin: Mouton de Gruyter.

Princeton University. (2010). WordNet [Computer software]. Princeton University, NJ.

Quine, W. (1961). *From a logical point of view*. Boston: Harvard University Press.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519-549.

Rayson, P., Archer, D., Piao, S., & McEnery, T. (2004). The UCREL semantic analysis system. In *Proceedings of the workshop on beyond named entity recognition semantic labeling for NLP tasks in association with 4th*

International Conference on Language Resources and Evaluation (pp.7-12), Lisbon, Portugal: Senseval.

Rayson, P., & Mariani, J. (2009). Visualizing corpus linguistics. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference* (pp. 20-23). Liverpool: Corpus Linguistics Conference.

Robinson, P. (1991). *ESP Today: A practitioner's guide*. New York: Prentice Hall International (UK) Ltd.

Roget, P. M. (1911). *Roget's thesaurus*. New Jersey: MICRA, Inc.

Room, A. (1981). *Dictionary of distinguishables*. Boston: Routhledge & Kegan Paul.

Saha Roy, R., Ganguly, N., Choudhury, M., & Kumar Singh, N. (2011). Complex network analysis reveals kernel-periphery structure in web search queries. In *Proceedings of the 2nd International ACM SIGIR (Association for Computing Machinery Special Interest Group on Information Retrieval) Workshop on Query Representation and Understanding* (pp. 3-5). Beijing: ACM SIGIR.

Schmidt, H. (n.d.) TreeTagger [Computer software]. Retrieved from http://corpus.byu.edu/now/http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

Schmitt. N. (2000). *Vocabulary in language teaching*. New York: Cambridge University Press.

Scott, J. (2000). *Social network analysis: A handbook*. London: Sage Publications.

Scott, M. (2015). WordSmith Tools Version 6 [Computer software]. Liverpool: Lexical Analysis Software.

Scott, M. & Tribble, C. (2006). *Textual patterns: Keyword and corpus analysis in language education*. Amsterdam: John Benjamins.

Shen, J. (2010). On English teaching in maritime colleges. *English Language Teaching, 4*, 176-179.

Simon, H. (1955). On a class of skew distribution Functions. *Biometrika, 42*, 425-440.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). The search for units of meaning. *Textus, 9*(1), 75-106.

Sinclair, J. (2004). *Trust the text: language, corpus and discourse*. London: Routledge.

Sinclair, J. (2005). Corpus and Text-Basic Principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1-16). Oxford: Oxbow Books.

Sinclair, J., Jones, S., & Daley, R. (2004). *English collocation studies: The OSTIR report*. London: Continuum.

Sole, R. V., Corominas-Murtra, B., Valverde, S., & Steels, L. (2005). Language networks: Their structure, function and evolution (Technical Report). *Santa Fe Institute Working Paper, 12*(42), 1-9.

Sole, R. V., & Pastor-Satorras, R. (2002). Complex networks in genomics and proteomics. In S. Bornholdt & H. G. Schuster (Eds.), *Handbook of graphs and networks* (pp. 147-169). Berlin: John Wiley-VCH.

Sporns, O., Chialvo, D., Kaiser, M., & Hilgetag, C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Science, 8*(9), 418-425.

Stevanak, J., Larue, D., & Carr, L. (2010). Distinguishing fact from fiction: Pattern recognition in texts using complex networks. *Computer Science*. arXiv:1007. 3254v2.

Strevens, P. (1988). ESP after twenty years: A re-appraisal. In M. Tickoo (Ed.), *ESP: State of the art* (pp. 1-13). Singapore: SEAMEO Regional Centre.

Stuart, K. & Botella, A. (2009). Corpus linguistics, network analysis and co-occurrence Matrices. *International Journal of English Studies, Special Issue*, 1-28.

Taylor, R. (2003). Near synonyms as co-extensive categories: ‘high’ and ‘tall’ revisited. *Language Sciences, 25*(3), 263-284.

Tognini-Bonelli, E. (1996). *Corpus theory and practice*. Birmingham: TWC Publishers.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Trenker, P. (2000). Maritime English: An attempt at an imperfect definition. In *Proceedings of* S*econd Asian IMLA Workshop on Maritime English*. Dalian, China (pp.1-8): Dalian Maritime University.

Trenker, P. (2009). Maritime English requirements and the revised STCW. In *Proceedings of International Maritime English Conference*, *21*, 5-10.

Valverde, S., Ferrer-Cancho, R., and Sol'e, R. V. (2002). Scale free networks from optimal design. *Europhysics Letters*, *60*, 512-517.

Watts, D., & Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, *393*, 440-442.

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, *32*(3), 259-276.

Williams, G. (1998). Collocation networks: Interlocking patterns of lexis in a corpus of plant biology re*sea*rch articles. *International Journal of Corpus Linguistics*, *3*(1), 151-171.

Williams, G. (2002). In search of representativity in specialized corpora: Categorization through collocation. *International Journal of Corpus Linguistics*, *7*(1), 43-64.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, *27*(1), 103-129.

Zhou, S., Hu, G., Zhang, Z., & Guan, J. (2008). An empirical study of Chinese language networks. *Physica A Statistical Mechanics & Its Applications*, *387*(12), 3039-3047.

Ziarati, R., Ziarati, M., Calbas, B., & Moussly, L. (2008). Improving safety standards at sea and ports by developing standards for Maritime English. *Martel Pro*, 1-8.

Appendix: Collocates of Near-synonyms

Collocates of *VESSEL* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ABANDON(ABANDONED)	M1	M	7.72	6
2	ABOARD	M6	M	9.57	6
3	ACCOMMODATE	H4	H	12.12	8
4	ACCOMPANYING	S3.1	S	7.75	2
5	ACT	A1.1.1	A	3.69	3
6	ADOPTED	A9	A	4.42	3
7	ADVISING	Q2.2	Q	6.18	2
8	AFFECTED	A2.1	A	3.16	2
9	AFFECTING	A2.1	A	5.08	2
10	AFLOAT	O4.1	O	9.26	5
11	AGROUND	M6	M	9.38	5
12	AIRCRAFT	M5	M	14.67	27
13	ALLOWS	S7.4	S	4.31	2
14	ALONE	A14	A	6.45	3
15	ANCHOR	M4	M	6.58	2
16	ANCORED(ANCHORING)	M1	M	14.05	22
17	APPEARS	A8	A	3.92	2
18	APPROACHED(APPROACHING)	M1	M	11.19	19
19	APPROXIMATE	A5.3	A	3.52	3
20	ARRIVAL(ARRIVED, ARRIVES, ARRIVING)	M1	M	6.54	3
21	ASSIST	S8	S	4.28	2
22	ASSISTED	S8	S	8.93	5
23	ASSISTING	S8	S	5.70	2
24	ASTERN	M6	M	8.45	9
25	AUXILIARY	Y1	Y	4.52	3
26	AVERAGE	A6.2	A	5.39	3
27	BEARS	L2	L	7.70	2
28	BEGAN(BEGINS)	T2	T	10.78	14
29	BIND	S6	S	9.90	2
30	BOARD	S7.1	S	14.04	44

31	BOTTOM	M6	M	6.52	3
32	BOUND	A1.7	A	7.72	4
33	BUILT	I4	I	12.62	30
34	BULK	N5	N	6.23	4
35	BURNING(BURNED)	O4.6	O	11.80	14
36	CALL(CALLING, CALLED, CALLS)	Q2.2	Q	4.02	2
37	CAME	M1	M	8.84	9
38	CAPABLE	X9.1	X	3.97	2
39	CARGO	O1	O	8.91	26
40	CARRY(CARRIED, CARRIES, CARRYING)	M2	M	13.06	25
41	CATEGORY	A4.1	A	7.57	3
42	CAUSING	A2.2	A	6.44	4
43	CEASES	T2	T	6.60	2
44	CENTRE	M6	M	6.52	2
45	CENTURY	T1.3	T	5.42	2
46	CERTAIN	A4.2	A	5.22	3
47	CERTIFICATE(S)	Q1.2	Q	9.58	2
48	CHANGED	A2.1	A	4.01	2
49	CHARTER	Q1.2	Q	5.08	3
50	CHARTERED	I3.2	I	11.75	12
51	CHECKS	X2.4	X	4.72	2
52	CLEAN	O4.2	O	5.35	3
53	CLEAR(CLEARED)	M2	M	6.61	3
54	CLIENT	I2.2	I	7.42	2
55	CLOSE	N3.3	N	4.96	4
56	COASTAL	W3	W	9.93	9
57	COMING	M1	M	6.13	5
58	COMMENCED	T2	T	9.27	6
59	COMMENCING	T2	T	6.37	2
60	COMMERCIAL	I2.1	I	17.20	56
61	COMMITTING	A1.1.1	A	7.92	2
62	COMPANY	I2.1	I	8.39	12
63	COMPLEMENT	N5.1	N	6.55	2
64	COMPLY(COMPLIES)	A6.1	A	9.42	7
65	CONCERNED	E6	E	9.07	9
66	CONSIDERED	X2.1	X	3.02	2

67	CONSTRAINED	A1.7	A	14.70	13
68	CONSTRUCT(CONSTRUCTED)	A1.1.1	A	8.41	7
69	CONTAINING(CONTAINED)	A1.8	A	8.13	6
70	CONTINUE(CONTINUED)	T2	T	10.94	18
71	CRANES	O2	O	4.90	2
72	CREW(S)	M4	M	6.31	25
73	CROSSES	M1	M	8.33	2
74	CROSSING	M1	M	10.25	9
75	CRUISE	M4	M	9.20	10
76	CURRENTLY	T1.1.2	T	7.87	4
77	DAMAGED	A1.1.2	A	6.75	8
78	DEAD	L1	L	4.56	2
79	DEEP-SEA	W3	W	7.80	2
80	DEFENCE	S8	S	5.16	2
81	DEPARTS(DEPARTIRED, DEPARTURING)	M1	M	11.67	13
82	DEPLOYED	M2	M	6.30	2
83	DEPLOYMENT	M2	M	6.08	2
84	DESCRIBED	Q2.2	Q	3.99	2
85	DESIGN(DESIGNS, DESIGNED)	C1	C	4.02	3
86	DETAILS	A4.2	A	3.12	2
87	DEVIATED	A6.1	A	10.09	3
88	DISABLED	B2	B	7.50	3
89	DISTANCE	N3.3	N	6.48	2
90	DISTRESSED	E6	E	10.60	7
91	DOCUMENTS	Q1.2	Q	7.53	4
92	DOMESTIC	H4	H	5.03	2
93	DRAFT	Q1.2	Q	10.65	9
94	DRAGGED	M2	M	7.11	2
95	DRAUGHT	W4	W	5.07	2
96	DRAWING	C1	C	9.50	5
97	DRIFTED	M1	M	6.04	2
98	DRINKING	F2	F	8.13	3
99	ENABLE	S8	S	7.76	4
100	ENGAGED	S4	S	16.29	56
101	ENSURES	A7	A	5.44	2
102	ENTER(ENTERS, ENTERED, ENTERING)	M1	M	3.81	2

103	ENTIRE	N5.1	N	7.81	5
104	EQUIPMENT	Y1	Y	4.58	9
105	EQUIPPED	A9	A	8.08	5
106	EVENTUALLY	N4	N	8.00	5
107	EXCEEDING	N5.2	N	8.57	4
108	EXCURSION	M1	M	9.38	2
109	EXISTING	A3	A	10.83	14
110	EXPECTED	X2.6	X	6.69	4
111	EXPERIENCED	X2.2	X	4.97	3
112	EXPERIENCING	A2.1	A	8.18	2
113	FACILITIES	H1	H	3.54	2
114	FAIR	G2.2	G	4.45	2
115	FASTER	N3.8	N	5.72	2
116	FINISHED	T2	T	5.08	2
117	FIRE-FIGHTING	O4.6	O	3.22	2
118	FISHING	K5.1	K	7.13	8
119	FITTED	N3.2	N	9.29	18
120	FLAG	Q1.1	Q	13.62	30
121	FLEET	M4	M	3.79	2
122	FLEXIBLE	O4.5	O	4.04	2
123	FLOATING	M1	M	6.85	4
124	FLOODING	W4	W	8.34	7
125	FLYING	M5	M	15.76	33
126	FOREIGN	M7	M	17.29	51
127	FORWARD	M6	M	5.38	5
128	FOUNDERED	T2	T	7.90	3
129	FREIGHT	M2	M	10.61	9
130	FURTHER	N5	N	7.33	10
131	GENERAL-CARGO	O1	O	13.48	21
132	GOING	M1	M	9.35	9
133	GOT	A9	A	7.42	4
134	GOVERNMENT	G1.1	G	4.75	3
135	GRANT	I1	I	9.50	4
136	GREATER	A5.1	A	9.01	7
137	GROUNDED	M4	M	5.92	2
138	GROUNDINGS	P1	P	9.18	2

139	GROUNDS	M6	M	7.91	7
140	GUARANTEES	A7	A	7.11	2
141	GUARD	A15	A	7.35	7
142	HANDLING	A1.1.1	A	7.97	5
143	HEAD	B1	B	4.98	2
144	HEADED(HEADING)	M1	M	5.57	3
145	HEELED	B5	B	9.09	3
146	HELD	M2	M	5.13	4
147	HIGH-SPEED	N3.8	N	8.05	5
148	HOLDING	M2	M	7.65	8
149	HUGE	N3.2	N	6.22	2
150	IMMEDIATELY	N3.8	N	6.79	6
151	INBOUND	M6	M	13.05	11
152	INCOMING	M1	M	9.56	4
153	INDIVIDUAL	N5	N	6.90	5
154	INDUSTRY	I4	I	5.33	7
155	INNOVATIVE	T3	T	8.71	2
156	INSPECTING(INSPECTED)	X2.4	X	4.92	2
157	INSPECTION	X2.4	X	6.84	6
158	INTACT	N5.1	N	4.27	3
159	INTEGRITY	G2.2	G	4.55	2
160	INTENDED	X7	X	6.92	6
161	INTENDING	X7	X	9.33	6
162	INTENDS	X7	X	6.46	2
163	INVOLVED	A1.8	A	10.42	13
164	ISSUED	A9	A	3.60	3
165	JOINING	A2.2	A	8.45	9
166	KEEP	A9	A	7.63	5
167	LACKED	A9	A	7.04	2
168	LARGE(R)	N3.2	N	13.71	21
169	LARGEST	N3.2	N	7.80	5
170	LASH	B1	B	12.42	5
171	LATITUDE	W3	W	5.79	2
172	LATTER	N4	N	7.37	5
173	LAY	M2	M	7.15	3
174	LEAVES(LEAVING, LEFT)	M1	M	11.12	7

175	LIGHT	W2	W	10.10	20
176	LINE	O4.4	O	4.84	4
177	LISTS(LISTED)	Q1.2	Q	7.59	5
178	LOCATION	M7	M	4.32	3
179	LOSSES	I1.2	I	9.39	7
180	LOST	X9.2	X	6.11	4
181	MAIN	A11.1	A	5.11	2
182	MAINTAINED	M2	M	3.96	4
183	MAINTAINS	A9	A	6.04	2
184	MAINTENANCE	A1.1.1	A	3.76	3
185	MAJOR	A11.1	A	5.32	6
186	MAKES(MAKING)	A1.1.1	A	3.83	2
187	MANOEUVRING	M2	M	5.44	5
188	MASTER(S)	S7.1	S	3.58	2
189	MEANS	Q1.1	Q	9.32	14
190	MERCHANT	I2.2	I	14.90	39
191	MEW	X3.2	X	8.71	2
192	MODERN	T3	T	7.67	4
193	MODIFICATIONS	A2.1	A	8.85	5
194	MONITOR	O2	O	5.54	3
195	MOORED(MOORING)	M2	M	12.97	12
196	MOTION	M1	M	9.76	8
197	MOTOR	O3	O	4.62	6
198	MOVED(MOVES, MOVED, MOVEMENT)	M2	M	10.27	7
199	NAME	Q2.2	Q	6.80	5
200	NAMED	Q2.2	Q	8.51	5
201	NATIONAL	M7	M	7.40	8
202	NATIONAL-FLAG	M7	M	7.90	2
203	NAVAL	M4	M	11.44	15
204	NAVIGATING(NAVIGATION)	M4	M	13.81	18
205	NEIGHBOURING	M6	M	8.01	2
206	NEW	T3	T	11.32	30
207	NORMALLY	A6.2	A	5.61	6
208	OBSERVES	X3.4	X	8.70	2
209	OBSERVING	X3.4	X	7.35	3
210	OCCUPANTS	H4	H	6.92	2

211	OFFSHORE	M6	M	7.56	4
212	OLD(OLDER)	T3	T	12.40	10
213	OPERATE(OPERATING, OPERATED, OPERATOR, OPERATES, OPERATING, OPERATION)	A1.1.1	A	12.06	30
214	ORDERS(ORDERED)	S7.1	S	6.26	3
215	OUTBOUND	M6	M	9.29	3
216	OVERTAKEN	N3.8	N	7.70	2
217	OVERTAKING	M3	M	12.92	8
218	OWNED	A9	A	13.04	16
219	OWNER(S)	A9	A	11.39	17
220	PARALLEL	A6.1	A	3.58	2
221	PARTICIPATING	S1.1.3	S	10.89	5
222	PARTICULAR	A4.2	A	6.50	8
223	PASSENGER	M3	M	6.59	10
224	PASSES(PASSED, PASSING)	M1	M	10.05	19
225	PATROL	X2.4	X	11.31	8
226	PERFORMANCE	K4	K	7.47	6
227	PERFORMING	A1.1.1	A	10.89	6
228	PERSONNEL	I3.1	I	7.47	9
229	PICK	X7	X	7.09	2
230	PLACING	M2	M	10.50	5
231	PLAN	X7	X	6.55	2
232	PLANNING	X7	X	3.29	2
233	POLICE	G2.1	G	9.08	8
234	POSITION	M6	M	4.16	49
235	POWERED	A1.1.1	A	7.52	3
236	PRESSURE	N3.5	N	7.29	7
237	PREVIOUS	N4	N	4.64	5
238	PROCEEDS(PROCEEDED)	A1.1.1	A	8.66	5
239	PRODUCES	A2.2	A	6.81	3
240	PROJECT	A1.1.1	A	3.83	2
241	PURCHASED	I2.2	I	6.66	2
242	QUALITY	A5.1	A	6.21	3
243	REACHES(REACHED)	M1	M	6.21	4
244	REASONABLY	A13.5	A	4.26	2
245	RECEIVING	A9	A	4.08	2

246	RECORD	Q1.2	Q	4.22	3
247	RECOVERY	B2	B	4.25	4
248	RECREATIONAL	K1	K	13.09	12
249	REFERRED	Q2.2	Q	4.93	6
250	REGIONAL	M7	M	3.14	2
251	REGISTRATION	Q1.2	Q	6.81	4
252	REGULATION	G2.1	G	4.41	4
253	RELIABLE	A5.1	A	5.80	2
254	REMAINED(REMAINS)	T2	T	8.19	6
255	REPAIR	A5.1	A	3.21	2
256	REPLACEMENT	A2.1	A	8.95	6
257	REPORTED	Q2.2	Q	3.51	2
258	REQUESTING	Q2.2	Q	6.90	3
259	REQUIRED	X7	X	6.35	11
260	RESCUE	S8	S	9.74	14
261	RESEARCH	X2.4	X	11.20	15
262	RESPONDED	S1.1.2	S	8.53	3
263	RESPONSE	Q2.1	Q	8.60	13
264	RESTRICTED	A1.7	A	13.49	21
265	RESULTING	A2.2	A	7.75	6
266	RETURNED	M1	M	9.25	8
267	ROLLED	M1	M	11.89	8
268	ROLLING	N3.8	N	4.29	2
269	RO-RO	M4	M	11.62	9
270	ROUTINELY	A6.2	A	8.88	3
271	RUNNING	M1	M	5.29	6
272	RUNS	M1	M	6.45	3
273	SAFE	A15	A	4.90	4
274	SAFELY	A15	A	7.04	4
275	SAILING	M4	M	14.81	48
276	SALVAGE	O2	O	5.82	7
277	SALVED	O2	O	13.34	10
278	SALVING	O2	O	14.56	10
279	SANK	M1	M	13.75	16
280	SCREW	O2	O	5.64	3
281	SEA	M4	M	5.76	5

282	SEAGOING	M4	M	15.30	21
283	SEAWORTHY	M5	M	7.50	2
284	SEND	M2	M	4.30	2
285	SERVICE	S8	S	3.73	5
286	SETTLED	M8	M	4.93	2
287	SEVERAL	N5	N	9.78	12
288	SIDE(S)	A4.1	A	3.02	3
289	SIMILAR	A6.1	A	10.17	12
290	SINK	B4	B	6.18	2
291	SINKING	M1	M	11.74	10
292	SISTER	S4	S	10.56	11
293	SITS	M8	M	8.04	2
294	SIZE(SIZES)	N3.2	N	5.49	7
295	SMALL	N3.2	N	13.80	38
296	SOUTHBOUND	M6	M	14.46	9
297	SPECIALIZED	A4.2	A	6.28	2
298	SPECIFIC	A4.2	A	8.49	8
299	SPEED	N3.8	N	5.95	13
300	STABILITY	A2.1	A	9.60	15
301	STANDBY	A1.1.1	A	9.67	6
302	STARTED	T2	T	8.52	9
303	STATES	Q2.1	Q	5.57	7
304	STAY(STAYED)	M8	M	6.32	5
305	STEEL	O1.1	O	3.17	2
306	STEMMING	T2	T	7.42	2
307	STERN	M4	M	3.98	2
308	STERN-FIRST	N4	N	8.04	2
309	STOPPED	T2	T	5.89	6
310	STORAGE	A9	A	4.81	5
311	STRAYS	A1.7	A	10.50	2
312	STRICKEN	A2.1	A	9.18	2
313	STRUCK	I3.1	I	10.48	7
314	STRUCTURE	O4.1	O	6.56	3
315	SUBJECT	X4.1	X	6.08	4
316	SUBMERGED	M6	M	4.48	2
317	SUBSEQUENTLY	N4	N	6.80	4

318	SUFFERED	E4.1	E	7.91	5
319	SUFFICIENT	N5	N	3.92	3
320	SUPPORT	S8	S	9.55	16
321	SURVEY(SURVEYS, SURVEYED)	X2.4	X	8.21	12
322	SURVIVAL	A3	A	9.78	8
323	SUSTAINED	T2	T	8.99	8
324	TAKES	A9	A	5.00	3
325	TENDER	E2	E	5.77	2
326	TERM	Q3	Q	12.32	22
327	TOOK	A9	A	4.67	3
328	TOTALLING	N5	N	8.31	2
329	TOWING	M2	M	13.83	31
330	TRADING	I2.2	I	7.85	3
331	TRAFFIC	M3	M	15.65	48
332	TRAINING	P1	P	14.11	41
333	TRANSITING	M1	M	14.76	11
334	TRAVELING	M1	M	8.54	2
335	TRIMMED	A1.1.1	A	8.80	3
336	TURNED	M2	M	3.83	2
337	TYPICAL	A4.2	A	4.27	2
338	UNDERTAKING	A1.1.1	A	5.72	2
339	UNDERWAY	A1.1.1	A	13.38	19
340	UNDERWENT	A1.1.1	A	7.11	2
341	UNEXPECTEDLY	X2.6	X	7.60	2
342	UNKNOWN	X2.2	X	10.11	6
343	UNMANNED	I3.1	I	7.45	3
344	UTILISATION	A1.5.1	A	6.92	2
345	VALUES	A11.1	A	4.10	2
346	VARIOUS	A6.3	A	8.05	6
347	VEHICLE	M3	M	6.33	3
348	VERTICAL	O4.4	O	3.11	2
349	VISITED	S1.1.1	S	7.03	2
350	WATERWAY	O1.2	O	6.08	2
351	WEIGHED	N3.5	N	8.21	3
352	WHEELHOUSE	O2	O	3.56	2
353	WOODEN	O1.1	O	9.11	5

354	WRECKED	A1.1.2	A	10.25	4

Collocates of *SHIP* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ABANDON(ABANDONED)	M1	M	18.84	52
2	ABANDONING	M1	M	13.26	8
3	ABOARD	M6	M	13.86	22
4	ACCOMMODATE	H4	H	8.10	3
5	ACCOUNT	I1	I	5.12	3
6	ACT	A1.1.1	A	6.36	4
7	ADVISING	Q2.2	Q	7.61	2
8	AGENCY	G1.1	G	4.49	2
9	AGES	T1.3	T	9.95	3
10	AGGREGATES	N5	N	4.58	3
11	AIRCRAFT	M5	M	9.38	11
12	ALTERED	A2.1	A	5.46	2
13	ANCORED(ANCHORING)	M1	M	8.34	3
14	ANNUALLY	N6	N	8.60	3
15	APPROACHED(APPROACHING)	M1	M	9.62	7
16	APPROPRIATED	A1.2	A	9.76	2
17	AREA	M7	M	3.29	3
18	ARRESTED	G2.1	G	12.03	10
19	ARRIVAL(ARRIVED, ARRIVES, ARRIVING)	M1	M	7.66	4
20	ARTICLE	Q4.2	Q	7.94	14
21	ASSIGNED	A9	A	5.29	2
22	AUTHORISED	S7.4	S	8.40	3
23	AUTOMATIC	A1.1.1	A	3.09	2
24	BAREBOAT	M4	M	8.29	3
25	BEARING	M2	M	3.42	2
26	BIG	N3.2	N	6.76	2
27	BLACKED	O4.3	O	13.54	6
28	BOARD	S7.1	S	14.38	51
29	BOARDED	M6	M	4.77	2
30	BOTTOM	M6	M	10.52	3
31	BOUND	A1.7	A	10.95	8
32	BOX	O2	O	4.72	2

33	BREAKING	A1.1.2	A	8.24	4
34	BUILT	I4	I	14.60	24
35	BUNKERED	G3	G	10.12	2
36	BUNKERS	G3	G	8.88	3
37	BUSY	A1.1.1	A	7.17	2
38	CALL(CALLING, CALLED, CALLS)	Q2.2	Q	4.77	2
39	CAPABLE	X9.1	X	3.04	2
40	CAPACITY	N3.2	N	7.60	5
41	CARGO	O1	O	12.51	51
42	CARRY(CARRIED, CARRIES, CARRYING)	M2	M	3.86	2
43	CELEBRITY	A11.1	A	10.61	2
44	CERTAIN	A4.2	A	5.44	3
45	CERTIFICATE(S)	Q1.2	Q	6.53	3
46	CHANNEL	W3	W	7.15	6
47	CHARTERED	I3.2	I	9.55	6
48	CLASSED	A4.1	A	8.85	3
49	CLASSIFICATION	A4	A	7.32	3
50	CLEAR(CLEARED)	M2	M	5.47	6
51	CLOSED	A10	A	3.99	3
52	COLLISIONS	A1.1.2	A	7.43	3
53	COMMERCIAL	I2.1	I	6.85	4
54	COMPASSES	O2	O	8.40	2
55	COMPLEMENT	N5.1	N	3.46	3
56	COMPLY(COMPLIES)	A6.1	A	8.45	6
57	CONCERNED	E6	E	11.45	20
58	CONCERNING	E6	E	3.37	2
59	CONDEMNATION	Q2.2	Q	3.32	2
60	CONDITION	O4.1	O	3.96	3
61	CONDUCTING	A1.1.1	A	5.12	2
62	CONSTRUCT(CONSTRUCTED)	A1.1.1	A	8.41	7
63	CONSTRUCTION	H1	H	8.46	8
64	CONTAINING(CONTAINED)	A1.8	A	4.31	2
65	CONTAINER	M3	M	15.91	58
66	CONTINUE(CONTINUED)	T2	T	9.62	9
67	CONTROL	S7.1	S	5.75	7
68	CONVENTION	A6.2	A	5.83	11

69	COURSE	M6	M	3.87	7
70	CREW(CREWS)	M4	M	8.33	4
71	DEEP-SEA	W3	W	9.23	2
72	DEPARTS(DEPARTIRED, DEPARTURING)	M1	M	9.20	7
73	DEPARTURE	M1	M	6.98	2
74	DESCRIBED	Q2.2	Q	5.96	3
75	DESIGN(DESIGNS, DESIGNED)	C1	C	9.51	19
76	DEVELOPING	A2.1	A	5.52	3
77	DIFFERENT	A6.1	A	5.53	4
78	DISABLED	B2	B	12.99	13
79	DISTINCTIVE	A11.2	A	11.16	5
80	DOCTOR	B3	B	12.99	13
81	DOCUMENTS	Q1.2	Q	5.68	5
82	DRAUGHT	W4	W	12.32	2
83	DRIFTED	M1	M	8.08	3
84	DRILL(S)	O2	O	10.98	10
85	DULY	A1.2	A	5.75	2
86	DYNAMICS	X4.2	X	8.61	2
87	EARLY	T4	T	5.86	3
88	EARTH	W3	W	7.52	5
89	ELECTRIC	O3	O	7.32	5
90	ELECTRONIC	O3	O	5.77	3
91	EMERGENCY	A11.1	A	6.58	7
92	EMPIRE	G1.1	G	9.02	2
93	EMPLOYED	I3.1	I	4.76	2
94	ENGAGED	S4	S	12.30	20
95	ENTER(ENTERS, ENTERED, ENTERING)	M1	M	5.63	3
96	ENTIRE	N5.1	N	8.88	9
97	ENTITLED	S7.4	S	12.38	18
98	EQUIPMENT	Y1	Y	6.08	17
99	EVACUATION	M2	M	5.41	4
100	EXERCISING	K5.1	K	10.34	4
101	EXISTING	A3	A	5.94	4
102	EXPERIENCED	X2.2	X	3.49	2
103	EXTREME	N5.1	N	5.31	2
104	FAILED	X9.2	X	3.16	2

105	FAMILY	S4	S	6.50	2
106	FEEDER	O1	O	8.22	5
107	FEELS	X2.1	X	8.93	2
108	FIGHTING	S8	S	3.23	2
109	FIRE	O4.6	O	4.99	10
110	FIRES	O4.6	O	9.67	8
111	FISHING	K5.1	K	7.77	11
112	FITTED	N3.2	N	7.97	11
113	FLAG	Q1.1	Q	11.04	11
114	FLIES	L2	L	12.63	6
115	FLYING	M5	M	18.40	58
116	FORE	S7.1	S	8.69	5
117	FOREIGN	M7	M	16.50	63
118	FUEL	O1	O	8.34	3
119	GANGWAY	M4	M	9.56	2
120	GENERAL	G3	G	3.14	3
121	GENERAL-CARGO	O1	O	15.33	29
122	GENERALLY	A6.2	A	3.64	2
123	GIVING	A9	A	3.70	2
124	GOVERNMENT	G1.1	G	9.21	14
125	GUARD	A15	A	5.20	6
126	HAND	B1	B	7.09	4
127	HANDLING	A1.1.1	A	11.18	13
128	HEADING(HEADED)	M1	M	7.09	16
129	HOSPITAL	B3	B	7.23	2
130	HOURS	T1.3	T	6.52	2
131	IDENTIFICATION	X2.2	X	11.62	11
132	IMMEDIATELY	N3.8	N	3.46	2
133	IMMERSED	M2	M	6.78	2
134	INDIVIDUAL	N5	N	7.97	6
135	INEVITABLY	A7	A	6.57	2
136	INSPECTING(INSPECTED)	X2.4	X	8.12	2
137	INSPECTION	X2.4	X	5.36	4
138	INSPECTORS	G2.1	G	6.46	2
139	INTENDING	X7	X	11.76	7
140	INTERNATIONAL	M7	M	3.49	7

141	INVOLVED	A1.8	A	7.14	7
142	ISSUED	A9	A	5.12	4
143	LARGE(R)	N3.2	N	14.60	21
144	LASH	B1	B	15.65	12
145	LEAVES(LEAVING, LEFT)	M1	M	7.73	3
146	LENGTH	N3.7	N	3.98	4
147	LIABILITY	A1.5.2	A	4.51	3
148	LIGHT	W2	W	3.41	3
149	LINE	O4.4	O	8.06	8
150	LISTS(LISTED)	Q1.2	Q	7.13	5
151	LOADER	O2	O	7.39	2
152	LOADING	M2	M	3.67	2
153	LOSSES	I1.2	I	4.24	2
154	LOST	X9.2	X	6.29	3
155	LYING	M1	M	6.74	3
156	MACHINERY	O2	O	3.59	11
157	MADE	A1.1.1	A	3.31	6
158	MAINTAINED	M2	M	4.79	3
159	MAKES(MAKING)	A1.1.1	A	5.26	2
160	MANAGEMENT	S7.1	S	18.56	103
161	MANAGER(S)	S7.1	S	10.68	7
162	MANDATORY	S6	S	7.19	4
163	MANNED	I3.1	I	6.00	2
164	MANOEUVRES	M1	M	6.50	2
165	MANOEUVRING	M2	M	6.97	4
166	MASTER(S)	S7.1	S	6.86	6
167	MASTS	H2	H	7.28	2
168	MATERIALS	O1	O	4.23	2
169	MEANS	Q1.1	Q	9.89	19
170	MEET	S3.1	S	6.04	3
171	MERCHANT	I2.2	I	16.94	60
172	MODERN	T3	T	11.86	16
173	MONTHLY	N6	N	7.37	2
174	MOORED(MOORING)	M2	M	7.98	3
175	MORTGAGE	I1.2	I	7.61	2
176	MOTHER	S4	S	8.68	3

177	MOTION	M1	M	7.04	6
178	MOTOR	O3	O	7.05	5
179	MOVED(MOVES, MOVED, MOVEMENT)	M2	M	7.56	5
180	MULTIPURPOSE	A1.5.2	A	12.23	4
181	MUSEUM	T1.1.1	T	7.87	2
182	NAME	Q2.2	Q	7.08	5
183	NAVIGATING(NAVIGATION)	M4	M	6.52	2
184	NAVY	G3	G	10.49	6
185	NEW	T3	T	13.93	33
186	NORMAL	A6.2	A	3.85	3
187	NUCLEAR	Y1	Y	15.03	29
188	NUMBER	N5	N	5.72	6
189	OCCURS	A2.1	A	5.15	2
190	OFFICERS	G1.1	G	6.86	12
191	OLD(OLDER)	T3	T	13.59	12
192	OPERATE(OPERATING, OPERATED, OPERATOR, OPERATES, OPERATING, OPERATION)	A1.1.1	A	4.22	2
193	ORDERS(ORDERED)	S7.1	S	6.48	7
194	OWNED	A9	A	15.34	37
195	PARENT	S4	S	8.14	3
196	PARTICULAR	A4.2	A	7.45	7
197	PASSENGER	M3	M	21.86	166
198	PASSES(PASSED, PASSING)	M1	M	12.74	33
199	PERFORMANCE	K4	K	4.74	3
200	PERSONNEL	I3.1	I	4.08	7
201	PORT	M4	M	8.29	19
202	POSEIDON	A4.1	A	11.21	3
203	POWERED	A1.1.1	A	12.95	12
204	PREVIOUS	N4	N	3.17	5
205	PRIMARILY	A13.2	A	4.87	2
206	PRIVATE	A10	A	7.17	4
207	PROCEEDS(PROCEEDED)	A1.1.1	A	5.78	4
208	PROPELLING	M2	M	5.23	2
209	PROPOSED	Q2.2	Q	4.75	2
210	PROPULSION	M2	M	10.59	20
211	PUBLIC	S5	S	5.04	5

212	PURPOSE	X7	X	3.08	2
213	PURSUED	A1.1.1	A	12.08	6
214	PURSUING	A1.1.1	A	8.32	2
215	RADAR	O3	O	7.40	9
216	RADIO	Q4.3	Q	4.61	7
217	REACHES(REACHED)	M1	M	8.42	4
218	REFERRED	Q2.2	Q	8.11	9
219	REFRIGERATED	O4.6	O	7.23	2
220	REGISTRATION	Q1.2	Q	4.09	2
221	REGISTRY	G1.1	G	5.00	23
222	REGULATION	G2.1	G	7.59	8
223	REMAINED(REMAINS)	T2	T	5.46	3
224	REPAIR	A5.1	A	10.87	11
225	REPAIRER	A5.1	A	12.78	4
226	REPLACING	A2.1	A	7.40	2
227	REPORTING	Q2.2	Q	16.89	40
228	REQUIRED	X7	X	6.53	9
229	RESEARCH	X2.4	X	8.07	6
230	RESTED	E3	E	6.97	2
231	RESUMED	T2	T	6.29	2
232	RIGHT	S7.4	S	6.54	2
233	ROLLED	M1	M	11.59	7
234	RO-RO	M4	M	14.33	20
235	ROUTEING	M2	M	13.02	14
236	ROUTING	A1.1.1	A	7.80	4
237	RUNNING	M1	M	5.47	5
238	SAFE	A15	A	14.81	37
239	SAFELY	A15	A	6.08	3
240	SAFETY	A15	A	13.69	51
241	SAILING	M4	M	7.51	9
242	SALVAGE	O2	O	6.10	4
243	SCREW	O2	O	7.17	4
244	SEAGOING	M4	M	14.50	44
245	SEARCH	X2.4	X	8.64	7
246	SEAWORTHY	M5	M	7.78	2
247	SECONDHAND	T3	T	9.47	2

248	SEEMED	A8	A	8.57	4
249	SENT	M2	M	4.81	2
250	SERVES	S8	S	5.41	2
251	SERVICE	S8	S	6.47	14
252	SIDE(S)	A4.1	A	8.51	14
253	SIMILAR	A6.1	A	3.43	2
254	SIMULATOR	O2	O	8.45	3
255	SINKING	M1	M	5.05	2
256	SISTER	S4	S	18.18	36
257	SIZE(SIZES)	N3.2	N	5.27	2
258	SMALL	N3.2	N	6.57	8
259	SOPHISTICATED	O4.2	O	7.68	2
260	SPARE	N5.2	N	4.22	4
261	SPECIALIZED	A4.2	A	6.50	2
262	SPECIFIC	A4.2	A	7.52	6
263	STABILITY	A2.1	A	6.37	4
264	STAFF	I3.1	I	9.04	21
265	STANDARD	A6.2	A	3.21	2
266	STANDBY	A1.1.1	A	9.34	6
267	STARTED	T2	T	8.80	9
268	STARTS	T2	T	8.51	4
269	STATE-OWNED	G1.1	G	13.29	6
270	STATION	M3	M	8.41	10
271	STAY(STAYED)	M8	M	7.53	4
272	STEAM	O1.3	O	7.96	5
273	STEERED	M2	M	6.74	2
274	STOPPED	T2	T	6.17	4
275	STORES	I2.2	I	7.63	16
276	STRUCTURE	O4.1	O	3.56	4
277	SUBJECT	X4.1	X	7.27	7
278	SUBSEQUENTLY	N4	N	5.83	3
279	SUBSTANDARD	A5.1	A	12.23	6
280	SUDDENLY	N3.8	N	8.63	4
281	SUFFERED	E4.1	E	8.19	5
282	SUPPLY	A9	A	7.29	15
283	SURFACE	M6	M	6.54	4

284	SURVEY(SURVEYS, SURVEYED)	X2.4	X	7.91	14
285	SUSPECTED	X2.1	X	5.69	2
286	SUSTAINED	T2	T	4.51	2
287	SWINGS	M1	M	11.08	4
288	TANKER	M3	M	4.60	4
289	TIME	T1	T	7.00	12
290	TONNAGE	N3.5	N	7.89	23
291	TOOK	A9	A	4.95	3
292	TOTALLING	N5	N	11.54	4
293	TOUCHED	X3.3	X	9.37	3
294	TRADING	I2.2	I	10.28	7
295	TRAFFIC	M3	M	6.42	5
296	TRAINING	P1	P	7.52	10
297	TRANSITING	M1	M	9.36	3
298	TRIALS	G2.1	G	6.47	3
299	TURNED	M2	M	5.86	3
300	TYCOON	X9.2	X	9.61	4
301	TYPE(S)	A4.1	A	4.49	4
302	UNDERTAKING	A1.1.1	A	9.97	5
303	UNREGISTERED	Q1.2	Q	10.73	5
304	USE(USED, USER, USING)	A1.5.1	A	3.51	3
305	VARIOUS	A6.3	A	5.27	3
306	VISITING	S1.1.1	S	13.93	8
307	VISITS	S1.1.1	S	9.74	4
308	VOLTAGE	O3	O	11.52	2
309	WAITING	T1.3	T	8.64	3
310	WAR	G3	G	8.15	4
311	WASH	B4	B	5.79	5
312	WOODEN	O1.1	O	7.12	3
313	WORLD	W1	W	3.67	2
314	WORLD-WIDE	W3	W	9.51	3
315	WRECKS	O4.2	O	7.31	3
316	YARD	H3	H	8.97	6
317	YAWS	X3.4	X	9.78	2

Collocates of *MARITIME* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ACCIDENT(S)	A15	A	19.61	56
2	ACTIVITY	A1.1.1	A	7.82	3
3	ADMINISTRATOR	S7.1	S	11.31	2
4	ADVENTURE	X8	X	11.48	3
5	AFFAIRS	A1.1.1	A	13.30	9
6	AGENCY(AGENCIES)	G1.1	G	6.95	3
7	AGREEMENTS	A6.1	A	8.92	4
8	ARBITRATION	G2.1	G	10.93	6
9	ASSETS	I1.1	I	7.50	2
10	AUTHORITY(IES)	G1.1	G	13.08	8
11	BILATERAL	S5	S	8.17	2
12	BOUNDARIES	M7	M	12.16	6
13	BUSINESSES	I2.1	I	10.22	3
14	CABOTAGE	A5.3	A	10.06	3
15	CARRIAGE	M2	M	9.96	6
16	CASUALTY(IES)	B2	B	13.08	8
17	CENTRE	M6	M	7.71	4
18	CLAIM	Q2.2	Q	16.44	25
19	CLAIMS	Q2.2	Q	14.01	16
20	COASTAL	W3	W	6.10	3
21	COLLEGE	P1	P	14.35	7
22	COMMAND	S7.1	S	8.61	4
23	COMMERCE	I2.1	I	12.43	5
24	COMMERCIAL	I2.1	I	13.82	17
25	COMMON	A6.2	A	6.49	3
26	COMMUNICATIONS	Q2.1	Q	4.95	2
27	COMMUNITY	S5	S	9.01	4
28	COMPANY	I2.1	I	5.07	3
29	COMPETENT	X9.1	X	6.44	3
30	CONCERNING	E6	E	11.73	9
31	CONFERENCE	S1.1.3	S	8.30	5
32	CONSULTANTS	B3	B	9.77	3
33	CONSULTATIVE	Q2.1	Q	17.68	16

34	COUNTRIES	M7	M	6.54	3
35	COVERING	A1.8	A	6.24	2
36	CURRENT	T1.1.2	T	6.35	4
37	CUSTOMS	A6.2	A	5.93	2
38	DECLARATION	Q2.2	Q	14.03	14
39	DIRECTORATE	S7.1	S	12.31	4
40	DISTRESS	E6	E	15.42	23
41	DOCUMENT	Q1.2	Q	4.89	2
42	ECONOMY	I2.1	I	11.41	6
43	EDUCATION	P1	P	14.72	11
44	ENCOUNTER	S3.1	S	8.59	2
45	ENHANCE	A5.1	A	11.26	5
46	ENSURE	A7	A	3.27	2
47	ENVIRONMENT	W5	W	7.35	4
48	ENVIRONMENTAL	W5	W	5.01	2
49	EXISTING	A3	A	8.76	5
50	EXPANDED	N3.2	N	12.84	7
51	EXPEDITE	N3.8	N	13.34	6
52	EXPERIENCE	X2.2	X	4.68	2
53	FACILITATE	S8	S	6.11	2
54	FIELD	F4	F	5.95	3
55	FISHING	K5.1	K	7.50	4
56	FORUM	S1.1.3	S	12.24	5
57	GENERAL	G3	G	9.74	9
58	GLOBAL	W3	W	17.09	25
59	GOVERNING	S7.1	S	8.77	3
60	INDUSTRY(IES)	I4	I	14.83	22
61	INTERNATIONAL	M7	M	21.11	153
62	INVESTIGATED	X2.4	X	7.20	2
63	LAW	G2.1	G	18.63	62
64	LEADING	S7.1	S	7.06	3
65	LEGAL	G2.1	G	6.80	3
66	LEGISLATION	G2.1	G	8.94	4
67	LIEN(S)	I2.2	I	21.35	49
68	LIMITS	A1.7	A	6.50	3
69	LINES	O4.4	O	6.95	4

70	MANAGEMENT	S7.1	S	5.19	3
71	MATTERS	X4.1	X	10.91	7
72	MEDIA	Q4	Q	9.69	3
73	MERCHANT	I2.2	I	7.01	3
74	MOBILE	M1	M	14.55	12
75	MUSEUM	T1.1.1	T	14.48	6
76	NATIONAL	M7	M	11.33	12
77	NATIONS	G1.1	G	14.03	17
78	NAVIGATION	M4	M	9.21	8
79	NORMAL	A6.2	A	3.95	2
80	OPERATIONS	A1.1.1	A	9.92	9
81	ORDER	S7.1	S	5.19	3
82	ORGANISATION	S5	S	14.80	17
83	PARLIAMENTARY	G1.1	G	11.31	2
84	PERSONAL	S5	S	5.61	2
85	PILOTS	M5	M	7.76	3
86	POLLUTION	W5	W	3.58	2
87	PORT	M4	M	7.14	7
88	POWERS	S7.1	S	6.67	2
89	PROMOTE	S8	S	9.50	4
90	PURPOSES	X7	X	6.59	3
91	QUALIFICATION	P1	P	9.89	3
92	RADIO(S)	Q4.3	Q	14.39	25
93	REGIONAL	M7	M	9.23	5
94	RELATED	A2.2	A	8.81	5
95	RELATIONS	A2.2	A	7.08	2
96	RELEVANT	A1.2	A	5.78	3
97	RESCUE	S8	S	10.63	9
98	RESPECTIVE	A6.1	A	5.91	2
99	RULE	G2.1	G	14.05	17
100	RULES	G2.1	G	18.31	58
101	SAFETY	A15	A	22.86	255
102	SALVAGE	O2	O	7.95	4
103	SAME	A6.1	A	5.97	4
104	SATELLITE	W1	W	5.94	2
105	SEARCH	X2.4	X	9.04	5

106	SECURITY	A15	A	12.22	11
107	SERVICES	S8	S	10.04	9
108	SHIPPING	M2	M	7.28	5
109	STANDARDS	A5.1	A	7.10	4
110	STRONG	S1.2.5	S	5.52	2
111	SUFFER	E4.1	E	7.96	2
112	TOOK	A9	A	5.05	2
113	TORT	G2.1	G	8.11	2
114	TRADE	I2.2	I	7.43	4
115	TRADITIONAL	S1.1.1	S	12.12	7
116	TRAFFIC	M3	M	16.17	31
117	TRAINING	P1	P	9.97	9
118	TRANSPORT	M3	M	22.01	137
119	TRANSPORTATION	M2	M	11.67	9
120	UNION	G1.2	G	11.30	6
121	UNIVERSITY	P1	P	11.24	4
122	USE(USING,USES)	A1.5.1	A	6.42	2
123	USER(USERS)	A1.5.1	A	7.77	3
124	WORLD	W1	W	9.79	7
125	ZONES	M7	M	8.49	3

Collocates of *MARINE* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ACCEPTED	A6.2	A	4.62	2
2	ACCIDENT(S)	A15	A	16.65	45
3	ACT	A1.1.1	A	8.42	6
4	ACTIVITY	A1.1.1	A	7.42	3
5	ADDRESSING	Q1.2	Q	8.33	2
6	ADMINISTRATIONS	S7.1	S	12.85	7
7	ADVISORY	S8	S	8.45	3
8	AGENCY(AGENCIES)	G1.1	G	5.46	2
9	ANTENNA	B1	B	5.03	2
10	APPLICATION(S)	Q1.2	Q	7.71	3
11	APPOINTED	S7.1	S	6.17	2
12	ARCHIVE	Q1.2	Q	12.99	3
13	BAND	K2	K	12.31	8

14	BANDS	S5	S	12.53	6
15	BASIC	A6.2	A	7.97	4
16	BOILERS	O2	O	8.97	3
17	BUNKER	G3	G	8.94	3
18	CASCADE	O1.2	O	9.43	2
19	CASUALTY(IES)	B2	B	19.69	57
20	CHARTERS	Q1.2	Q	7.81	2
21	CHARTS	Q1.2	Q	6.54	4
22	CHEMIST	B3	B	12.91	4
23	CLIPPER	M4	M	9.53	2
24	COMMERCIAL	I2.1	I	7.16	4
25	COMMUNICATIONS	Q2.1	Q	13.27	15
26	COMMUNITY	S5	S	10.37	6
27	COMPANY	I2.1	I	4.67	3
28	CONDUCT	S1.1.1	S	10.41	7
29	CONSERVATION	W5	W	6.97	2
30	CONSTRUCTION	H1	H	5.67	3
31	CONTAINERS	O2	O	8.02	4
32	CONTRACTOR	I2.1	I	5.23	2
33	CONVENIENCE	A1.2	A	9.95	4
34	CONVENTIONAL	A6.2	A	6.04	2
35	COORDINATION	S7.1	S	8.53	3
36	CORP	I2.1	I	11.80	6
37	CORPORATION	I2.1	I	8.81	4
38	CRAFT	M4	M	4.52	2
39	CREW	M4	M	6.28	6
40	CREWS	M4	M	5.63	2
41	DEBRIS	O1.1	O	6.31	2
42	DELPHINUS	A4.1	A	13.40	3
43	DEPARTMENT	H2	H	15.23	23
44	DESIGN	C1	C	5.75	3
45	DESIGNERS	C1	C	8.98	2
46	DIESEL(S)	O1.2	O	15.88	31
47	DIESELOIL	O1.2	O	10.66	5
48	DIRECTORATE	S7.1	S	12.88	5
49	DISASTER	A5.1	A	9.74	3

50	DISTRESS	E6	E	6.20	3
51	DIVISION	A6.1	A	10.10	6
52	ECOSYSTEM	W5	W	13.40	3
53	ELECTRICAL	O3	O	6.34	4
54	ELECTRONIC	O3	O	5.47	2
55	EMERGENCY	A11.1	A	12.79	21
56	EMPLOYER	I3.1	I	8.84	2
57	ENGINE	O2	O	7.49	8
58	ENGINEER(S)	Y1	Y	14.25	18
59	ENGINEERING	Y1	Y	14.95	20
60	ENGINES	O2	O	12.58	17
61	ENVIRONMENT	W5	W	25.13	267
62	ENVIRONMENTAL	W5	W	19.26	59
63	EQUIPMENT	Y1	Y	3.86	3
64	EVACUATION	M2	M	5.72	2
65	EXPERIENCE	X2.2	X	7.28	4
66	EXPLORER	M1	M	10.91	2
67	EXPOSED	A10	A	5.26	2
68	FACILITATE	S8	S	7.46	3
69	FACILITY(IES)	H1	H	10.58	9
70	FEDERATION	S5	S	7.81	2
71	FIREFIGHTING	O4.6	O	11.44	8
72	FISHERIES	F4	F	10.44	5
73	FLOODING	W4	W	6.40	3
74	FORMAL	A6.2	A	8.97	4
75	FORUM	S1.1.3	S	7.88	2
76	FREQUENCIES	N6	N	6.11	2
77	GENERATORS	O3	O	6.90	3
78	GRADE	A5.1	A	8.86	3
79	GROUP	S5	S	5.58	3
80	GROWTH	N3.2	N	9.88	6
81	GUARD	A15	A	17.03	44
82	GUIDANCE	S8	S	13.20	15
83	HALTER	B5	B	13.40	3
84	HAZARDS	A15	A	5.06	2
85	HEAVY	N3.5	N	4.08	2

86	HIGHFREQUENCY	N6	N	8.84	2
87	HIGHWAY	M3	M	12.15	4
88	INCIDENT(S)	A3	A	12.63	17
89	INCLUDE	A1.8	A	3.75	2
90	INDUSTRY(IES)	I4	I	8.74	3
91	INFORMATION	X2.2	X	5.12	4
92	INSPECTION(S)	X2.4	X	7.41	3
93	INSTALLATIONS	A1.1.1	A	8.61	4
94	INTERATIONAL	M7	M	12.23	2
95	INTERNAL	M6	M	6.12	3
96	INTERNATIONAL	M7	M	9.32	11
97	ISSUED	A9	A	6.57	4
98	LANTERN	O2	O	13.29	6
99	LEG	B1	B	6.88	2
100	LIFE	L1	L	11.48	10
101	LIMITED	N5	N	7.15	4
102	LOCAL	M7	M	9.14	7
103	MACHINERY	O2	O	5.80	3
104	MAMMALS	L2	L	15.65	7
105	MANGANESE	O1.1	O	9.53	3
106	MEDIUM	N3.2	N	7.67	3
107	MERCANTILE	I2.1	I	14.75	7
108	MERCHANT	I2.2	I	15.24	22
109	NATIONAL	M7	M	13.93	24
110	NAVIGATION	M4	M	17.95	66
111	NAVIGATIONAL	M4	M	10.51	8
112	OBSERVER	X3.4	X	11.86	6
113	OCCUPATIONAL	I3.1	I	10.91	4
114	OCCURRENCE(S)	A2.1	A	9.38	8
115	OFFICE	I2.1	I	10.03	7
116	OPERATIONS	A1.1.1	A	14.11	26
117	ORDER	S7.1	S	3.04	2
118	ORGANISATION	S5	S	13.84	9
119	ORGANISMS	L2	L	13.84	9
120	PERSONNEL	I3.1	I	9.20	7
121	PILOTS	M5	M	13.36	12

122	PIPELINE(S)	O2	O	9.82	6
123	PLASTIC	O1.1	O	8.22	4
124	PLYWOOD	O2	O	9.08	3
125	POLICY	X7	X	6.88	3
126	POLLUTANTS	O1	O	14.53	8
127	POLLUTION	W5	W	18.81	74
128	PORT	M4	M	7.83	9
129	PRACTICE	A1.1.1	A	5.71	3
130	PREPAREDNESS	O4.1	O	7.66	3
131	PRODUCTS	O2	O	8.78	5
132	PROFESSIONAL	I3.2	I	9.26	4
133	PROPOSED	Q2.2	Q	5.05	2
134	PROPULSION	M2	M	8.69	6
135	PROTECTION	S8	S	10.73	10
136	QUALIFICATION	P1	P	7.74	2
137	RADAR(S)	O3	O	12.63	6
138	RADIO(S)	Q4.3	Q	15.69	41
139	RADIOTELEGRAPH	Q4.3	Q	5.97	2
140	RADIOTELEPHONE	Q4.3	Q	6.34	2
141	RAILWAY	M3	M	11.42	5
142	RECEIVED	A9	A	3.88	2
143	REFRIGERATION	O4.6	O	6.93	2
144	REGIONAL	M7	M	7.87	4
145	RELEVANT	A1.2	A	3.63	2
146	REMOTE	N3.3	N	9.37	5
147	REPORTABLE	Q2.2	Q	10.65	2
148	REQUIREMENTS	S6	S	5.73	4
149	RESCUE	S8	S	9.72	8
150	RESOURCES	X4.2	X	7.43	4
151	RESPONSE	Q2.1	Q	9.47	9
152	SAFETY	A15	A	20.20	151
153	SALVAGE	O2	O	11.52	10
154	SANCTUARIES	M7	M	16.29	11
155	SANCTUARY	M7	M	14.40	6
156	SANITATION	B4	B	12.91	4
157	SCIENCE	Y1	Y	12.07	5

158	SCIENTIFIC	Y1	Y	22.28	97
159	SENT	M2	M	6.87	3
160	SERIOUS	A11.1	A	9.57	6
161	SERVICES	S8	S	12.13	16
162	SITES	M7	M	10.84	5
163	SPEED	N3.8	N	6.48	5
164	SPILL	A1.1.1	A	5.07	2
165	STANDARDS	A5.1	A	6.70	4
166	STEAM	O1.3	O	11.26	10
167	STRUCTURES	O4.1	O	10.34	6
168	SURVEYING	X2.4	X	9.94	3
169	SURVEYS	X2.4	X	9.34	5
170	SYSTEMS	X4.2	X	4.66	3
171	TECHNOLOGY	Y1	Y	14.75	18
172	TRAFFIC	M3	M	6.91	4
173	TRANSPORT	M3	M	11.07	12
174	TRANSPORTATION	M2	M	13.18	14
175	TURBINE(S)	O2	O	10.38	12
176	TYPE	A4.1	A	4.69	3
177	TYPICAL	A4.2	A	7.76	3
178	UNDERTAKE	A1.1.1	A	6.23	2
179	UNDERWRITERS	I2.1	I	9.23	2
180	UNIT	S5	S	7.46	5
181	USE(USING,USES)	A1.5.1	A	5.06	7
182	USER(USERS)	A1.5.1	A	7.37	3
183	VARIOUS	A6.3	A	6.82	4
184	VEGETATION	L3	L	14.33	8
185	VENTURE	X8	X	9.50	3
186	WEATHER	W4	W	8.48	6
187	WORKERS	I3.1	I	9.79	4

Collocates of *OCEAN* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ACT	A1.1.1	A	10.59	3
2	AREA(S)	M7	M	10.12	6
3	BED	H5	H	12.72	5
4	BLACK	O4.3	O	10.23	7
5	BOTTOM	M6	M	12.40	8
6	BOUNDARY	M7	M	4.25	2
7	CARGO	O1	O	4.56	2
8	CARRIER(S)	M3	M	9.46	4
9	CLEAN	O4.2	O	6.23	3
10	CURRENT	T1.1.2	T	9.42	5
11	ENVIRONMENT	W5	W	8.21	3
12	FALLS	M1	M	8.80	2
13	FLOOR	H2	H	15.24	9
14	GOING	M1	M	10.63	4
15	INLETS	W3	W	10.49	2
16	NATIONAL	M7	M	9.64	5
17	OIL	O1.2	O	3.40	2
18	OPEN	A10	A	10.16	6
19	PASSAGES	H2	H	11.17	3
20	POLICY	X7	X	10.29	2
21	SERVICE	S8	S	6.76	3
22	SHIPPING	M2	M	16.17	24
23	STATE(S)	G1.1	G	12.67	10
24	TOWING	M2	M	9.89	3
25	TRADE	I2.2	I	4.25	3
26	TRAFFIC	M3	M	4.66	3
27	TRANSPORTATION	M2	M	12.02	6
28	VOYAGE	M1	M	10.48	5
29	WATER	O1.2	O	11.44	18

Collocates of *SEA* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ACT	A1.1.1	A	11.07	6
2	ADJACENT	M6	M	8.58	5
3	ADVERSE	S8	S	6.42	2
4	AFFORDED	I1.1	I	8.70	2
5	ALLOWING	S7.4	S	5.76	2
6	APPLICABLE	A1.5.2	A	4.38	2
7	AREA(S)	M7	M	13.10	45
8	ARTICLE	Q4.2	Q	8.68	10
9	AUXILIARY	Y1	Y	6.41	3
10	BASIC	A6.2	A	9.09	5
11	BED	H5	H	14.17	12
12	BLACK	O4.3	O	17.97	40
13	BOTTOM	M6	M	12.48	14
14	BOUNDARY	M7	M	5.80	2
15	BREAKING	A1.1.2	A	11.10	3
16	BUOY(S)	M4	M	15.35	26
17	CALM	E3	E	13.78	5
18	CAUSED	A2.2	A	3.73	2
19	CHART(S)	Q1.2	Q	6.73	5
20	CLEAN	O4.2	O	9.73	5
21	CLOSE	N3.3	N	3.85	2
22	CLOUD	W4	W	7.25	2
23	CLUTTER	O2	O	17.74	21
24	COMBINED	A1.8	A	8.13	4
25	COMMUNITY	S5	S	8.26	2
26	CONDITION(S)	O4.1	O	18.47	80
27	CONFUSED	X2.5	X	8.40	2
28	CONNECTIONS	A2.2	A	9.23	4
29	CONVENED	S7.1	S	9.02	3
30	CONVENTION	A6.2	A	5.51	5
31	CORAL	L3	L	7.87	2
32	CROSSING	M1	M	6.72	2
33	CRUISES	M4	M	9.28	2
34	DESIGNATE	Q2.2	Q	9.27	3

35	DIAMETER	N3.7	N	5.20	2
36	DISPOSAL	X7	X	14.44	14
37	ENGINEROOM	Y1	Y	3.85	3
38	ESTABLISHED	T2	T	3.98	2
39	EXCLUDING	A1.8	A	10.06	3
40	EXPERIENCE	X2.2	X	4.44	2
41	EXPRESS	Q1.1	Q	5.96	2
42	FEDERAL	S5	S	8.06	4
43	FERRY(FERRIES)	M4	M	8.67	6
44	FISH	L2	L	13.83	16
45	FISHERIES	F4	F	15.13	8
46	FISHING	K5.1	K	13.72	12
47	FLEET	M4	M	6.96	2
48	FLOOR	H2	H	6.40	2
49	FREEDOMS	A1.7	A	10.67	4
50	FULL	N5.1	N	10.32	10
51	GENERAL	G3	G	3.00	2
52	HEAD	B1	B	8.54	3
53	HEAVY	N3.5	N	16.24	24
54	HEIGHT	N3.7	N	6.35	3
55	HIGH	N3.7	N	24.62	198
56	INLAND	M6	M	8.50	4
57	INLETS	W3	W	15.11	10
58	INTERESTS	X5.2	X	5.14	2
59	INTERNATIONAL	M7	M	7.52	7
60	ISSUES	X4.1	X	7.25	2
61	LANES	M3	M	20.87	47
62	LARGE	N3.2	N	8.18	6
63	LEVEL	N3.7	N	12.61	18
64	LOCAL	M7	M	6.88	4
65	MAIN	A11.1	A	10.40	13
66	MAKING	A1.1.1	A	4.65	2
67	MANDATORY	S6	S	7.56	3
68	MARGINAL	A11.2	A	13.17	7
69	MARINE	M4	M	6.29	5
70	MEANS	Q1.1	Q	5.79	4

71	METRE	N3.3	N	11.77	4
72	MILE	N3.3	N	8.24	3
73	MODERATE	N5	N	10.13	4
74	NORMAL	A6.2	A	6.71	4
75	OFFDUTY	I3.1	I	9.81	2
76	OIL	O1.2	O	3.20	3
77	OPEN	A10	A	14.49	30
78	OUTLINING	Q2.1	Q	11.07	2
79	PART	N5.1	N	4.77	2
80	PARTNERS	S3.1	S	18.66	25
81	PASSAGES	H2	H	7.07	2
82	PILOT(S)	M5	M	8.77	4
83	POLLUTION	W5	W	3.34	2
84	PORT(S)	M5	M	8.86	11
85	PREVENTING	S8	S	8.35	2
86	PRINCE	S7.1	S	8.11	2
87	PROTEST	Q2.2	Q	15.23	10
88	PROVIDES	A9	A	8.33	5
89	QUELLING	T2	T	14.88	2
90	RECOGNISED	X2.2	X	6.42	2
91	RED	O4.3	O	14.34	16
92	REGION(S)	M7	M	8.31	4
93	REMAINS	T2	T	5.52	2
94	RESPECTIVE	A6.1	A	5.67	2
95	ROUGH	X3.3	X	20.03	36
96	ROUTES	M6	M	7.36	3
97	ROUTINE	A6.2	A	7.49	3
98	SAFE	A15	A	8.04	3
99	SAFETY	A15	A	7.65	8
100	SERVICE	S8	S	6.63	5
101	SHIPPING	M2	M	7.32	3
102	SHORT	T1.3	T	9.56	7
103	SIDE	A4.1	A	4.91	3
104	SIMPLE	A12	A	7.31	3
105	SLIGHT	N5	N	11.61	10
106	SNAKE	L2	L	8.75	2

107	SPARROW	L2	L	12.40	2
108	SPEED	N3.8	N	9.19	9
109	STAFF	I3.1	I	9.50	6
110	STATE(S)	G1.1	G	9.67	20
111	STRONG	S1.2.5	S	7.04	3
112	SUBSTITUTING	A2.1	A	9.94	2
113	SUCCESSFUL	X9.2	X	6.15	2
114	SUCTION(S)	M2	M	10.98	3
115	SUFFICIENT	N5	N	5.81	3
116	SURFACE	M6	M	12.43	14
117	SURVIVAL	A3	A	13.78	13
118	TIME	T1	T	3.22	3
119	TIMOR	A4.1	A	11.81	2
120	TRADE	I2.2	I	5.95	3
121	TRAFFIC	M3	M	8.04	5
122	TRANSPORT	M3	M	10.45	10
123	TRANSPORTATION	M2	M	4.92	2
124	TRIALS	G2.1	G	16.51	21
125	VALVES	O2	O	11.16	11
126	VENTURE	X8	X	9.66	3
127	VESSELS	M4	M	5.76	5
128	VOLUNTEER	I3.1	I	8.19	2
129	VOYAGES	M1	M	9.21	4
130	WALL	H2	H	16.37	20
131	WATCH	T1	T	3.93	2
132	WATER	O1.2	O	21.52	205
133	WIND	W4	W	12.39	13
134	YEARS	T1.3	T	3.04	2

Collocates of *SAFETY* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ACT	A1.1.1	A	8.99	8
2	ACTION(S)	A1.1.1	A	13.26	18
3	ADDITIONAL	N5	N	6.31	4
4	ADDRESS	H4	H	4.58	2
5	ADEQUATE	N5	N	8.64	6
6	ADMINISTRATION(S)	S7.1	S	8.56	7
7	ADVISORY	S8	S	17.88	31
8	AFFECT	A2.2	A	6.13	3
9	AGENCY	G1.1	G	11.87	12
10	AIDS	B2	B	4.95	2
11	AIR	O1.3	O	4.58	4
12	ALERT	X5.1	X	5.53	2
13	AMPLE	N5	N	6.91	2
14	ANALYSIS	X2.4	X	6.14	3
15	ANNOUNCEMENT	Q2.2	Q	8.82	3
16	ANNUAL	N6	N	6.11	3
17	APPLIANCES	O2	O	5.12	2
18	APPROPRIATE	A1.2	A	8.13	8
19	ASPECTS	A4.1	A	7.32	3
20	ASSESSMENT	X2.4	X	13.36	18
21	ASSOCIATED	S5	S	5.30	3
22	AUTHORITY	G1.1	G	19.06	101
23	AUTOMATIC	A1.1.1	A	3.86	2
24	AWARENESS	X2.2	X	11.58	7
25	BASIC	A6.2	A	10.80	9
26	BOARD	S7.1	S	20.32	168
27	BOAT	M4	M	9.86	11
28	BOATING	M4	M	13.85	9
29	BRANCH	S5	S	8.67	4
30	BRIEFING(S)	Q2.1	Q	10.61	3
31	BULLETIN(S)	Q4	Q	12.20	5
32	CAPS	B5	B	8.31	2
33	CARGO	O1	O	6.14	6
34	CARRIER	M3	M	9.79	9

35	CASE	A4.1	A	10.33	14
36	CENTRE	M6	M	6.64	4
37	CERTIFICATE(S)	Q1.2	Q	9.26	6
38	CHARACTERISTICS	O4.1	O	4.15	2
39	CHECK	X2.4	X	5.67	3
40	CIRCULAR	Q1.2	Q	5.72	2
41	CODE	Y2	Y	9.80	9
42	COMMITTEE	S7.1	S	18.97	71
43	COMMON	A6.2	A	12.08	14
44	COMMUNICATIONS	Q2.1	Q	12.60	15
45	COMPANY	I2.1	I	8.24	8
46	COMPLETE	N5.1	N	6.69	4
47	CONCERN(S)	E6	E	12.38	9
48	CONCERNING	E6	E	8.11	5
49	CONDITIONS	O4.1	O	3.76	3
50	CONSIDERATIONS	X4.1	X	10.31	5
51	CONSTRUCTION	H1	H	10.99	12
52	CONTRIBUTING	A9	A	15.08	20
53	CONTROL	S7.1	S	4.10	4
54	CORRECT	A5.3	A	3.89	2
55	COURSE	M6	M	4.37	3
56	CRITICAL	A11.1	A	13.67	17
57	CULTURE	C1	C	17.05	19
58	DATA	X2.2	X	12.69	20
59	DEFICIENCIES	N5	N	11.87	8
60	DEFICIENT	N5	N	9.75	2
61	DESIGNATED	Q2.2	Q	4.24	2
62	DEVICE(S)	O2	O	13.99	17
63	DIRECTORATE	S7.1	S	11.23	4
64	DIVISION	A6.1	A	8.63	5
65	DOOR(S)	H2	H	9.05	7
66	DRILLS	O2	O	10.21	7
67	EFFECTIVE	A1.5	A	10.23	10
68	EFFECTIVENESS	A1.5	A	9.31	5
69	ELECTRICAL	O3	O	5.66	4
70	ENGINEERS	Y1	Y	4.07	2

71	ENHANCE	A5.1	A	10.97	6
72	ENHANCEMENTS	A5.1	A	9.38	2
73	ENHANCING	A5.1	A	9.72	3
74	ENSURE	A7	A	8.71	9
75	EQUIPMENT	Y1	Y	17.50	82
76	ESTABLISH	T2	T	6.02	3
77	ESTABLISHED	T2	T	8.56	7
78	EXECUTIVE	S7.1	S	8.87	5
79	FAIRWAY(S)	M3	M	9.38	2
80	FERRY	M4	M	11.13	11
81	FIRE	O4.6	O	13.24	45
82	FISHERMEN	I3.1	I	18.56	3
83	FISHINGVESSEL(S)	M4	M	16.11	29
84	FLEET	M4	M	7.60	5
85	FORMAL	A6.2	A	14.80	18
86	FORUM	S1.1.3	S	7.20	2
87	FURTHER	N5	N	8.93	9
88	GENERAL	G3	G	6.12	5
89	GOGGLES	B5	B	9.09	2
90	GOOD	A5.1	A	5.94	4
91	GREATER	A5.1	A	5.19	3
92	GUIDANCE	S8	S	3.80	2
93	GUIDE(S)	S8	S	10.88	8
94	GUIDELINES	S8	S	9.37	6
95	HAZARDS	A15	A	7.38	4
96	HEALTH	B2	B	10.04	8
97	IDENTIFIED	X2.2	X	5.45	3
98	IMPLICATIONS	A2.2	A	7.06	2
99	IMPROVE	A5.1	A	13.26	14
100	IMPROVED	A1.5	A	5.32	2
101	IMPROVEMENTS	A5.1	A	6.49	2
102	IMPROVING	A5.1	A	6.49	2
103	INCREASED	N5	N	4.94	3
104	INCREASES	N5	N	5.11	2
105	INDEPENDENT	S5	S	8.71	5
106	INDUCTION	X2.1	X	6.06	2

107	INFORMATION	X2.2	X	13.30	31
108	INITIATE	T2	T	12.45	8
109	INSPECTION(S)	X2.4	X	8.55	7
110	INSPECTORATE	X2.4	X	16.57	13
111	INSTRUCTIONS	Q2.2	Q	5.93	3
112	INTERLOCKS	A1.1.1	A	10.23	2
113	INTERNAL	M6	M	5.44	3
114	INTERNATIONAL	M7	M	15.60	55
115	INVESTIGATION	X2.4	X	14.21	28
116	INVOLVES	A1.8	A	6.36	2
117	ISSUE(S)	X4.1	X	21.71	122
118	ISSUED	A9	A	6.86	5
119	JOB	I3.1	I	6.03	2
120	JUSTIFICATION	G2.2	G	14.39	7
121	KNIFE	O2	O	10.61	3
122	LAMP	O2	O	8.09	4
123	LAWS	G2.1	G	3.73	2
124	LIMITATIONS	A1.7	A	5.59	2
125	LIMITS	A1.7	A	3.67	2
126	LINE(S)	O4.4	O	4.63	3
127	MAJOR	A11.1	A	5.16	3
128	MANAGE	S7.1	S	6.21	2
129	MANAGEMENT	S7.1	S	22.44	207
130	MANAGER	S7.1	S	10.23	8
131	MANDATORY	S6	S	4.96	2
132	MANUAL(S)	Q4.1	Q	6.80	2
133	MARGIN	N5	N	10.41	5
134	MARINE	M4	M	20.20	151
135	MARITIME	W3	W	22.86	255
136	MATERIAL(S)	O1	O	14.80	32
137	MATTERS	X4.1	X	11.79	11
138	MEASURES	X4.2	X	13.61	24
139	MEETING(S)	S1.1.3	S	5.64	3
140	MESSAGE	Q1.1	Q	10.85	8
141	MINIMUM	N5	N	6.50	4
142	MINOR	A11.1	A	10.45	7

143	MODERN	T3	T	4.52	2
144	NAVIGATION	M4	M	7.56	7
145	NAVIGATIONAL	M4	M	10.80	10
146	NECESSARY	S6	S	4.78	4
147	NEW	T3	T	8.39	10
148	NOTICES	X3.4	X	5.09	2
149	OCCUPATIONAL	I3.1	I	16.49	17
150	OFFICE	I2.1	I	8.69	6
151	OFFICER	G1.1	G	12.22	20
152	ONBOARD	M6	M	5.51	2
153	OPERATION	A1.1.1	A	3.48	3
154	OPERATIONAL	A1.1.1	A	9.20	7
155	ORGANIZATION(S)	S5	S	4.45	3
156	OVERSIGHT	A5.3	A	10.06	4
157	PARTICULAR	A4.2	A	4.56	3
158	PASSENGER	M3	M	11.88	15
159	PATROL	X2.4	X	10.20	4
160	PERSONAL	S5	S	10.54	8
161	PERSONNEL	I3.1	I	3.10	2
162	PETROLEUM	O1.2	O	12.24	9
163	PIN(S)	O2	O	7.82	3
164	PIPELINE	O2	O	7.90	3
165	PLAN	X7	X	6.68	5
166	POLICY	X7	X	10.96	9
167	PORT	M4	M	9.90	17
168	POSITIVE	A5.1	A	4.47	2
169	PRACTICAL	A1.6	A	4.64	2
170	PRACTICES	A1.1.1	A	10.03	7
171	PRECAUTION(S)	A1.3	A	17.58	35
172	PROACTIVE	X5.2	X	12.92	6
173	PROBLEMS	A12	A	6.67	4
174	PROCEDURES	X4.2	X	10.95	15
175	PROMOTE	S8	S	8.42	4
176	PROTECTION	S8	S	7.84	6
177	PROVISION(S)	O2	O	8.12	8
178	PUBLIC	S5	S	10.29	11

179	PURPOSES	X7	X	6.75	4
180	RADIO	Q4.3	Q	7.80	7
181	RADIOCOMMUNICATIONS	Q4.3	Q	11.76	5
182	REASONABLE	S1.2.6	S	4.25	2
183	REASONS	A2.2	A	9.58	6
184	RECOMMENDATION(S)	Q2.2	Q	16.01	31
185	RECORD	Q1.2	Q	4.16	2
186	REGIME	G1.1	G	5.60	2
187	REGULATION(S)	G2.1	G	11.44	18
188	REITERATES	Q2.1	Q	10.55	2
189	RELEVANT	A1.2	A	6.92	5
190	RELIEF	E4.1	E	6.57	3
191	REPLIED	Q2.2	Q	9.85	3
192	REPORT	Q2.2	Q	3.98	3
193	REQUIREMENTS	S6	S	11.31	17
194	RISK	A15	A	8.96	8
195	RORO	M4	M	4.95	2
196	ROUTINE	A6.2	A	7.89	4
197	RULES	G2.1	G	7.42	6
198	SAME	A6.1	A	3.65	3
199	SCHEME	X7	X	11.66	9
200	SCHOOL	P1	P	9.50	3
201	SEA	M4	M	7.65	8
202	SECURING	A1.7	A	4.87	2
203	SERIES	N4	N	3.89	2
204	SERIOUS	A11.1	A	4.14	2
205	SERVICE	S8	S	8.79	10
206	SEVERAL	N5	N	4.87	3
207	SHIP	M4	M	13.69	47
208	SHIPPING	M2	M	6.99	6
209	SIGNIFICANCE	A11.1	A	6.13	2
210	SIGNIFICANT	A11.1	A	8.68	6
211	SITE	M7	M	4.64	2
212	SPECIFIC	A4.2	A	6.54	4
213	STANDARD(S)	A6.2	A	15.28	34
214	STEPS	M1	M	5.09	2

215	STUDIES	P1	P	7.97	3
216	SUPERSEDES	A2.1	A	10.55	2
217	SURVEY	X2.4	X	9.26	8
218	SYSTEM	X4.2	X	11.80	27
219	TAKEN	A9	A	3.73	3
220	TANKER	M3	M	9.65	7
221	TECHNICAL	Y1	Y	3.34	2
222	TRAFFIC	M3	M	7.19	5
223	TRAINING	P1	P	12.96	23
224	TRANSPORT	M3	M	17.90	68
225	TRANSPORTATION	M2	M	22.05	127
226	VARIOUS	A6.3	A	6.14	4
227	VESSEL	M4	M	6.26	9
228	VEST	B5	B	12.31	3
229	WATER	O1.2	O	6.64	8
230	WATERWAYS	O1.2	O	5.77	2
231	WORKER	I3.1	I	9.99	3
232	ZONE(S)	M7	M	16.76	26

Collocates of *SECURITY* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ACT	A1.1.1	A	5.95	2
2	COMMITTEE	S7.1	S	6.48	2
3	COUNCIL	S7.1	S	12.33	7
4	COVERING	A1.8	A	8.12	2
5	DURATION	T1.3	T	8.61	2
6	EXCESSIVE	N5.2	N	9.79	3
7	FEATURES	A4.2	A	7.04	2
8	FINANCIAL	I1	I	22.31	66
9	FURNISHED	H5	H	11.44	3
10	GIVEN	A9	A	10.99	8
11	GIVING	A9	A	7.44	2
12	INTERESTS	X5.2	X	11.23	5
13	MAINTAINED	M2	M	8.52	3
14	MARITIME	W3	W	12.22	11
15	MEASURES	X4.2	X	13.57	12

16	NAME	Q2.2	Q	10.81	5
17	NATIONAL	M7	M	16.56	26
18	NECESSARY	S6	S	10.16	7
19	NIGHT	T4	T	13.79	9
20	OFFICER	G1.1	G	9.97	6
21	PERSPECTIVE	O4.1	O	10.12	2
22	POLICY	X7	X	10.41	4
23	PORT	M4	M	5.35	3
24	PROGRAM	Y2	Y	7.65	2
25	PROVIDE	S5	S	5.22	2
26	PROVIDED	A9	A	12.70	14
27	PUBLIC	S5	S	11.29	7
28	PURPOSES	X7	X	6.71	2
29	REASONS	A2.2	A	7.79	2
30	REDUCED	N5	N	6.28	2
31	REQUIREMENTS	S6	S	8.97	5
32	SATISFACTORY	X5.2	X	12.57	6
33	SATISFYING	E4.2	E	11.27	2
34	SHIPPING	M2	M	8.19	4
35	SPECIFIED	A4.2	A	8.23	3
36	STEWARD(S)	M5	M	16.31	5
37	STRATEGY	G3	G	9.16	2
38	SUFFICIENT	N5	N	7.93	3
39	SYSTEM	X4.2	X	3.50	2
40	TAKES	A9	A	10.25	4
41	TECHNOLOGY	Y1	Y	15.95	14
42	ZONE(S)	M7	M	11.02	5

Collocates of *PORT* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ACCESS	M1	M	10.77	12
2	ADMINISTRATION(S)	S7.1	S	6.58	2
3	AFFAIRS	A1.1.1	A	8.99	7
4	AFFECTS	A2.1	A	8.42	3
5	AFT	M6	M	11.71	16
6	ALLEYWAY	M7	M	9.71	5
7	ALTERNATIVE	A6.1	A	8.30	3
8	ANCHOR	M4	M	13.33	20
9	APPLIED	A1.5.2	A	6.93	5
10	APPROACH	X4.2	X	3.64	2
11	APPROACHING	M1	M	5.11	2
12	ARRIVAL	M1	M	7.90	5
13	ARRIVED	M1	M	3.58	2
14	ASSESSING	X2.4	X	6.39	2
15	AUTHORITY(IES)	G1.1	G	18.44	81
16	AUTONOMY	G1.1	G	10.97	3
17	AUXILIARY	Y1	Y	6.65	4
18	BANK	I1	I	8.00	4
19	BASE	M7	M	5.79	3
20	BEAM	W2	W	9.61	6
21	BILGE	M4	M	3.81	2
22	BOILER	O2	O	8.24	7
23	BOTANY	Y1	Y	12.80	4
24	BOW	O4.1	O	15.10	30
25	BRIDGE	H1	H	11.00	17
26	BULKHEAD	O2	O	7.68	5
27	BULWARK	S8	S	7.58	2
28	CALLS	Q2.2	Q	8.38	4
29	CANAL	W3	W	7.39	4
30	CENTER	M6	M	4.29	2
31	CHANNEL	W3	W	4.75	3
32	CHARGE(S)	I1.3	I	13.87	16
33	CHEMICAL	O1	O	4.19	2
34	CITY	M7	M	6.69	3

35	CLEMENTS	W4	W	13.39	4
36	COAL	O1.1	O	4.42	2
37	COAST	W3	W	5.58	3
38	COMMERCIAL	I2.1	I	9.82	14
39	COMMUNICATIONS	Q2.1	Q	5.46	3
40	COMMUNITY(IES)	S5	S	15.72	22
41	COMPETENT	X9.1	X	3.44	2
42	COMPLEX	A12	A	5.48	2
43	CONFINES	M7	M	10.14	3
44	CONGESTION	B2	B	13.10	8
45	CONSISTS	A1.8	A	6.21	3
46	CONSTRUCTION	H1	H	4.82	3
47	CONTAINER	M3	M	10.34	14
48	CONTROL	S7.1	S	10.45	18
49	CORNER	O4.4	O	7.36	3
50	CORP	I2.1	I	17.40	18
51	CORPORATION	I2.1	I	14.93	26
52	CORPORATIONS	I2.1	I	9.24	3
53	COSTS	I1.3	I	5.13	3
54	COUNTRY	G1.1	G	3.80	2
55	DEEPWATER	W3	W	16.65	18
56	DEPARTED	M1	M	5.59	2
57	DESIGNATES(DESIGNATED)	Q2.2	Q	10.76	7
58	DESTINATION	M6	M	5.67	2
59	DEVELOPMENT(S)	A2.1	A	9.05	10
60	DIESEL	O1.2	O	3.17	2
61	DIFFERENT	A6.1	A	9.58	8
62	DIRECTIONAL	M6	M	8.01	3
63	DIRECTOR	S7.1	S	8.98	7
64	DISAPPEAR	A10	A	11.27	2
65	DISCHARGING	A1.7	A	7.87	4
66	DISTANT	N3.3	N	8.76	2
67	DOOR	H2	H	4.36	3
68	DUES	I1.3	I	17.88	34
69	DUTY	S6	S	5.46	3
70	ECONOMY	I2.1	I	8.41	4

71	EFFICIENCY	X9.1	X	6.08	3
72	EMERGENCY	A11.1	A	4.77	4
73	ENCLOSED	A1.8	A	6.14	2
74	ENGINE(S)	O2	O	13.93	47
75	ENGINE-ROOM	Y1	Y	11.06	20
76	ENHANCE	A5.1	A	6.05	2
77	ENTER(ENTERED, ENTERING)	M1	M	8.76	15
78	ENTITIES	O2	O	6.66	2
79	ENTRANCE	M7	M	10.46	8
80	ENTRY	M7	M	7.89	6
81	ESPECIALLY	A14	A	5.85	3
82	EXHAUST	M3	M	15.14	31
83	EXPANSION	N3.2	N	4.87	2
84	FACILITY	H1	H	10.40	9
85	FERRY	M4	M	9.01	7
86	FIN	L2	L	13.80	10
87	FINAL	N4	N	4.04	2
88	FINANCING	I1	I	8.59	3
89	FOREIGN	M7	M	11.90	19
90	FORMALITIES	A6.2	A	12.79	10
91	FORWARD	M6	M	11.81	18
92	FRIENDLY	S1.2.1	S	9.95	3
93	FUNNEL	O2	O	5.31	2
94	GENERATOR	O3	O	14.05	30
95	GIVEN	A9	A	4.87	4
96	GROUP	S5	S	8.64	5
97	HAND	B1	B	8.98	7
98	HANDBOOK	Q4.1	Q	7.44	2
99	HANDLED	A1.1.1	A	7.05	3
100	HARD	O4.5	O	4.92	2
101	HARDY	O4.1	O	15.07	8
102	HATCH	H5	H	5.36	3
103	HEALTH	B2	B	13.84	20
104	HELM	M4	M	13.46	12
105	HOLIDAYS	K1	K	14.91	11
106	IMPERIAL	S7.1	S	9.80	2

107	INCLUDING	A1.8	A	7.85	13
108	INFORMATION	X2.2	X	6.03	6
109	INFRASTRUCTURES	I2.1	I	11.82	3
110	INLET	W3	W	10.99	11
111	INSTALLATIONS	A1.1.1	A	6.52	3
112	INSTITUTE	S5	S	7.32	3
113	INSTRUMENT	O2	O	8.32	4
114	INTERESTS	X5.2	X	10.65	9
115	INTERFACE	Q1.1	Q	14.01	11
116	INTERMEDIATE	N5	N	7.46	3
117	INTERNATIONAL	M7	M	9.20	18
118	ISLAND	W3	W	10.82	10
119	JETS	M5	M	7.93	2
120	LAKE	W3	W	7.89	2
121	LEAVE(LEAVES, LEAVING)	M1	M	14.10	24
122	LEG	B1	B	7.79	3
123	LEVYING	G1.1	G	12.27	5
124	LIFEBOAT	M4	M	12.58	12
125	LIMITS	A1.7	A	11.26	12
126	LIST	Q1.2	Q	5.98	4
127	LOCAL	M7	M	9.38	14
128	MAIN	A11.1	A	14.67	48
129	MAIN-DECK	A11.1	A	4.29	2
130	MAJOR	A11.1	A	12.69	20
131	MANAGEMENT	S7.1	S	11.94	19
132	MANAGER(S)	S7.1	S	8.87	4
133	MARINE	M4	M	7.83	9
134	MARITIME	W3	W	13.58	28
135	MASTER	S7.1	S	8.61	12
136	MEANS	Q1.1	Q	7.78	8
137	MEDICAL	B3	B	4.29	2
138	MERCHANT	I2.2	I	7.01	4
139	MET	S3.1	S	6.17	3
140	MOTOR	O3	O	9.06	9
141	NEIGHBOURING	M6	M	9.38	2
142	NEW	T3	T	9.68	18

143	NUMBER	N5	N	3.33	3
144	OFFER	A9	A	6.72	2
145	OFFICE	I2.1	I	8.52	6
146	OFFICIAL(S)	G1.1	G	11.73	9
147	OIL	O1.2	O	3.43	6
148	OPEN	A10	A	6.75	7
149	OPENING	A1.1.1	A	5.73	3
150	OPERATIONAL	A1.1.1	A	10.12	9
151	OPERATIONS	A1.1.1	A	13.27	26
152	OPERATOR	Q1.3	Q	13.31	24
153	ORGANIZATION(S)	S5	S	4.28	6
154	ORIGINAL	T3	T	5.54	4
155	OUTBOARD	M4	M	5.95	2
156	OUTLET	Q2.1	Q	10.12	6
157	PILOT(S)	M5	M	4.96	6
158	PREVIOUS	N4	N	3.77	2
159	PRE-WAR	T1.1.1	T	12.51	3
160	PRIVATE	A10	A	6.53	5
161	PROCEDURES	X4.2	X	6.03	5
162	PROGRAM(S)	Y2	Y	9.58	4
163	PROPELLER	M4	M	7.48	5
164	PROPULSION	M2	M	3.09	2
165	PROTECT	S8	S	6.68	3
166	PUBLIC	S5	S	12.86	17
167	QUARTER	N1	N	15.77	25
168	QUESTION	Q2.2	Q	5.88	2
169	RECEPTION	H2	H	7.21	3
170	REDUCING	N5	N	5.04	2
171	REGIONAL	M7	M	7.47	7
172	REGULATIONS	G2.1	G	10.48	15
173	REHABILITATE	B3	B	10.80	2
174	RELATIONSHIP	S3.1	S	8.94	3
175	RELEVANT	A1.2	A	8.21	10
176	REMOTE	N3.3	N	11.00	6
177	REPAIR	A5.1	A	7.54	3
178	RESPECTIVE	A6.1	A	7.67	6

179	RIGHT-HAND	M6	M	8.22	2
180	RIVER(S)	W3	W	9.25	5
181	RUDDER	S7.1	S	13.32	15
182	SAFE	A15	A	6.75	5
183	SAFEGUARD	A15	A	10.93	3
184	SAFETY	A15	A	9.90	17
185	SAT	M8	M	9.39	3
186	SATISFACTORY	X5.2	X	6.44	3
187	SEA	M4	M	12.17	27
188	SEABOARD	W3	W	12.25	3
189	SECURITY	A15	A	5.35	3
190	SELECTED	X7	X	6.50	2
191	SERVICE	S8	S	6.35	4
192	SERVING	S8	S	5.54	2
193	SEVERAL	N5	N	4.64	2
194	SHAFT	O2	O	5.33	3
195	SHELTERED	A10	A	7.58	2
196	SHIP	M4	M	8.29	14
197	SHIPMENT	N5	N	7.90	3
198	SHIPPING	M2	M	7.49	11
199	SHIPYARD	M4	M	7.69	4
200	SIGNALS	Q1.1	Q	4.13	2
201	SISTER	S4	S	12.73	6
202	SITUATED	M6	M	10.77	8
203	SMALL(SMALLER)	N3.2	N	9.75	17
204	SPECIFIED	A4.2	A	5.04	2
205	SPILL	A1.1.1	A	5.98	3
206	STABILISER	M3	M	11.62	6
207	STAGING	K4	K	10.62	2
208	STANDPOINT	X2.1	X	10.44	3
209	STATE(S)	G1.1	G	6.71	11
210	TANKS	O2	O	5.19	3
211	TERMINALS	M5	M	9.30	6
212	TIDEWATER	W3	W	15.74	13
213	TIME	T1	T	5.21	6
214	TOWN	M7	M	9.87	4

215	TRAFFIC	M3	M	9.06	8
216	TRUST	E6	E	6.18	2
217	TUNNEL	O2	O	8.10	4
218	USER	A1.5.1	A	6.94	3
219	VARIOUS	A6.3	A	8.63	7
220	VILA	M4	M	15.73	9
221	VISIT	S1.1.1	S	11.88	8
222	VOYAGE	M1	M	3.17	2
223	WARDEN	S7.1	S	13.03	5
224	WATER	O1.2	O	3.47	4
225	WHEEL	O2	O	5.07	2
226	WINDOW	H2	H	7.97	3
227	WING	L2	L	14.87	17
228	WORLDS	W1	W	11.56	3
229	WORLDWIDE	W3	W	7.69	2

Collocates of *HARBOR* (L1-R1, MI3≧3,Frequency≧2)

No	Collocate	Sub-Semantic Domain	Main Semantic Domain	MI3	Frequency
1	ALTERNATE	A2.1	A	11.38	2
2	ANCHORAGE	M4	M	12.86	5
3	AREA(S)	M7	M	8.28	3
4	BASINS	O2	O	15.40	3
5	BOARD(S)	S7.1	S	11.35	28
6	BRIDGEPORT	M4	M	15.32	2
7	BUREAU	G1.1	G	13.07	7
8	CHART(S)	Q1.2	Q	13.72	9
9	COMMISSIONERS	G1.1	G	16.66	3
10	COMMUTER	M3	M	12.44	2
11	CRAFT	M4	M	8.05	3
12	CRUISES	M4	M	10.30	2
13	ENTRANCE	M7	M	12.94	10
14	FAIR	G2.2	G	16.36	9
15	FREEPORT	M4	M	11.68	2
16	IMPORTANT	A11.1	A	8.65	2
17	INSTALLATIONS	A1.1.1	A	11.04	3
18	MAINTENANCE	A1.1.1	A	10.41	4

19	MAJOR	A11.1	A	10.65	3
20	NATURAL	A6.2	A	10.74	7
21	REACH	N3.3	N	9.06	2
22	REGULATIONS	G2.1	G	6.13	5
23	RIVER(S)	W3	W	14.97	4
24	TERRITORIES	M7	M	11.72	2
25	TIDAL	W3	W	8.57	2
26	TRANSPORTATION	M2	M	10.19	3
27	VARIOUS	A6.3	A	11.62	4