



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

문학석사 학위논문

해사영어 : 코퍼스 기반 어휘 분석

Maritime English : A Corpus-based Lexical Analysis



2011년 2월

한국해양대학교 대학원

영어영문학과

변 현 정



문학석사 학위논문

해사영어 : 코퍼스 기반 어휘 분석

Maritime English: A Corpus-based Lexical Analysis



2011년 2월

한국해양대학교 대학원

영어영문학과

변 현 정

本 論文을 邊賢貞의 文學碩士 學位論文으로 認准함.

위원장 류 미 립



위원 장 세 은



위원 홍 신 철



2010년 12월 24일

한국해양대학교 대학원

목 차

Abstract	vi
I. 서 론	1
1.1 연구의 필요성 및 목적	1
1.2 연구 주제	2
1.3 논문의 구성	3
1.4 제한점	3
II. 이론적 배경	5
2.1 코퍼스 언어학	5
2.1.1 코퍼스 언어학과 응용	5
2.1.2 코퍼스의 종류	6
2.1.3 코퍼스를 활용한 어휘분석 선행연구	7
2.2 코퍼스 구축	10
2.3 특수목적 영어	13
2.3.1 특수목적 영어	13
2.3.2 해사영어	15
III. 연구 방법	17
3.1 연구대상 및 도구	17
3.2 연구절차 및 방법	17

3.2.1	해사영어코퍼스(MECO)의 구축	18
3.2.1.1	MECO의 대표성	18
3.2.1.2	MECO의 크기	20
3.2.1.3	MECO의 기계해독가능형태	21
3.2.1.4	표준화된 참고자료	22
IV.	연구 결과 및 논의	23
4.1	해사영어코퍼스의 주제와 영역별 분포	23
4.2	어휘적 분석	25
4.2.1	품사별 분포도	25
4.2.2	내용어의 분포와 어휘적 특징	28
4.2.2.1	명사	28
4.2.2.2	동사	31
4.2.2.3	형용사	36
4.2.2.4	부사	38
4.2.3	기능어의 어휘적 특징	41
4.2.3.1	전치사	41
4.2.3.2	대명사	43
4.2.3.3	조동사	45
V.	결론 및 제언	49

참고문헌

부록

표 목 차

<표 1> MECO 구축에 사용된 텍스트 종류	19
<표 2> 텍스트 종류별 크기	20
<표 3> MECO와 BNC에서 나타나는 <i>ship</i> 의 빈도수	22
<표 4> MECO의 주제별 분포	23
<표 5> MECO의 영역	24
<표 6> TTR 도출방법	25
<표 7> 분류된 해사영어 품사와 품사부착세트(tagset)	25
<표 8> 내용어 종류와 빈도수	26
<표 9> 기능어의 종류와 빈도수	27
<표 10> 명사의 빈도수와 누적빈도	28
<표 11> MECO 고빈도 명사의 빈도수와 LOB에서의 빈도수	30
<표 12> 동사의 빈도수와 누적빈도	32
<표 13> MECO와 LOB에 쓰인 <i>be</i> 동사의 빈도수와 비율	34
<표 14> 형용사사의 빈도수와 누적빈도	36
<표 15> 형용사와 연어 관계를 가지는 어휘 목록	37
<표 16> 부사의 빈도수와 누적빈도	39
<표 17> <i>however</i> 위치 비교 (빈도수/백분율)	41
<표 18> MECO 전치사 목록 (1-gram)	41
<표 19> MECO 전치사 목록 (2-gram)	42
<표 20> MECO 전치사 목록 (3-gram)	42
<표 21> MECO와 LOB의 인칭대명사 비율	44

<표 22> 조동사의 빈도수	45
<표 23> MECO에 나타나는 <i>might</i> 와 <i>shall</i> 의 빈도수	46
<표 24> MECO에 나타나는 <i>might</i> 와 <i>may</i> 의 빈도수	47



그림 목 차

<그림 1> BNC에 나타나는 표준화된 참고자료	12
<그림 2> MECO와 LOB의 통계 정보	24
<그림 3> 전체 어휘에서 차지하는 내용어와 기능어의 분포도	26
<그림 4> 내용어의 분포도	27
<그림 5> 기능어의 분포도	28
<그림 6> 명사의 누적 빈도 백분율	29
<그림 7> MECO의 키워드목록	31
<그림 8> 동사의 누적 빈도 백분율	33
<그림 9> MECO와 LOB의 어휘목록	34
<그림 10> MECO와 LOB에 쓰인 <i>be</i> 동사 백분율 비교	35
<그림 11> 형용사의 누적빈도 백분율	37
<그림 12> 부사의 누적빈도 백분율	39
<그림 13> MECO / LOB 조동사 분포	46

Maritime English : A Corpus-based Lexical Analysis

Byun, Hyun-Jeong

Department of English Language and Literature

Graduate School of

Korea Maritime University

Abstract

The purpose of this thesis is to compile a corpus of Maritime English and to describe some lexical features of Maritime English by comparing them with those of General English corpora such as LOB and BNC.

Corpus Linguistics is useful to illustrate real features of language use. In particular, specialized corpora in English can make a contribution to showing how a language is used in specific areas which are different from General English. In maritime communities, there have been communicative problems even though The International Maritime Organization (IMO) offered Standard Marine Communication Phrases (SMCP). Maritime English needs to be studied more systematically to help learners of Maritime English and seafarers who need Maritime English as English for Specific Purposes (ESP). However, a maritime English corpus has not been compiled yet, so it is necessary to compile a Maritime English corpus that offers more realistic language being used in the maritime community and to analyze the linguistic characteristics of Maritime English.

For the compilation of a Maritime English corpus, six kinds of maritime written texts such as journals, maritime law, media text, set language, textbooks and other documents were selected and about one million words

were collected. Interviews and discussions with six professors who have been teaching Maritime English courses were implemented. Word frequencies in the Maritime English corpus were identified and then keywords were extracted by using a concordance, WordSmith 5.0.

The results of this study show that some lexical features of Maritime English are different from those of General English, so they can raise the awareness of lexical features of Maritime English. The word-list and keyword-list extracted from the Maritime English corpus (MECO) will be fundamental materials to help learners understand the frequent lexis in maritime texts and will be useful resources for maritime English textbooks. The high frequent word-list and the lexical characteristics of this study are expected to develop seafarers' and learners' communication skills.



I. 서론

1.1 연구 필요성 및 목적

세계적으로 해양산업이 발전함에 따라 관련 업무의 원활한 처리를 위해 해상 의사소통은 중요한 요소이다. 즉 세계무역의 80%이상이 해상운송에 의존하고 있고, 선박의 운항, 선박간의 커뮤니케이션, 안전을 위한 규약의 이행 등의 업무가 국제적으로 밀접하게 연결되어 있기 때문에 해상에서 이루어지는 의사소통은 중요하다(Taner, 2009).

MAIB(Marine Accident Investigation Branch, 2010)의 연간 보고서는 대형 해상 사고들이 빈번하게 발생하고 있음을 보여주고 있다. IMO(International Maritime Organization)의 해상 사고분석에 따르면 70% 이상의 사고는 선원들 간의 의사소통의 문제로 인해 발생한다(Chirea-Ungureanu & Georgescu, 2009). 더욱이 HEWG(Human Element Working Group)는 선원들은 영어 말하기의 문제뿐만 아니라 해양 관련 용어를 사용하는 데도 큰 어려움을 가지고 있다고 보고한다(Taner, 2009). 국적이 다른 선원들의 문화적인 다양성과 전문적인 용어의 이해 부족 등이 해상에서의 의사소통 문제를 야기하고 있다.

이러한 의사소통의 문제를 최소화하기 위한 방안으로 IMO는 SMCP(The Standard Marine Communication Phrases)와 STCW 95(Standards of Training, Certification and Watchkeeping) Convention 등 효과적인 의사소통을 위한 규정화된 해사영어(Maritime English)를 제공한다. 그러나 Trenkner(2009)에 따르면 규정화된 모듈의 제공에도 불구하고 해사영어를 사용하는 사람들은 여전히 의사소통의 문제점을 가지고 있다. 그러므로 해사영어가 실제 의사소통에 도움이 되기 위해서는 언어학적 관점에서 보다 체계적이고 실증적인 연구가 필요하다. 해사 커뮤니티에서 실제 사용되는 언어를 분석하여 빈도 높은 어휘를 제시하고, 해사영어 언어 사용의 패턴을 제시함으로써 학습자의 어휘학습과 의사소통에 도움을 줄 수 있다. 이러한 연구는 해사영어의 교육적 효과를 증대시킬 수 있는 토대를 마련할 수 있다.

학습자가 전문적, 학문적 분야에서의 업무를 능숙하게 수행할 수 있도록 준비시키는 것을 목표로 하는 특수목적영어(English for Specific Purposes, ESP)의 관점으로 해사

영어를 이해하고, 체계적으로 분석하고 연구할 필요가 있다. 과학, 기술, 엔지니어링 등 다양한 특수 분야의 연구를 위해 코퍼스가 구축되고(Rizzo, 2010), 코퍼스를 기반으로 어휘, 문법적 특징 등 언어학적 분석이 실증적으로 이루어지고 있다. 왜냐하면 다양한 특수 분야의 언어 특징을 설명하는 데는 일반 코퍼스보다 특수 코퍼스가 더 효과적이기 때문이다(Gavioli, 2005). 그러나 현재 구축된 해사영어 코퍼스가 없기 때문에 일반영어와 차이를 보이는 해사영어의 특징을 실제 데이터를 바탕으로 분석하고 증명할 수 없다. 그러므로 현재 언어학적 연구가 제한적인 해사영어의 코퍼스 구축이 필요하다.

본 연구에서는 실제 해사 업무 및 연구 분야에 실제로 사용되는 자료를 수집하여, 해사영어를 대표할 수 있는 해사영어코퍼스(Maritime English Corpus, MECO)를 구축하고 여기에서 나타나는 품사별 분포, 내용어 및 기능어의 어휘적 특징을 일반영어를 대표할 수 있는 LOB(Lancaster-Oslo/Bergen)¹⁾ 코퍼스 또는 BNC(British National Corpus)²⁾ 코퍼스와 비교 분석하여 기술해보고자 한다.

1.2 연구 주제

본 연구에서는 특수목적영어인 해사영어와 일반영어가 어휘적인 면에서 차이가 난다는 가설을 바탕으로 McEnergy and Wilson(2001)의 코퍼스 구축의 네 가지의 관점(대표성, 한정된 크기, 기계해독가능형태, 표준화된 참고자료)에 따라 해사영어코퍼스를 구축한 과정을 기술하고 해사영어의 어휘적 특징을 일반영어 텍스트로 구성된 LOB 코퍼스 또는 BNC 코퍼스와 비교 분석해보겠다. 그러므로 본 연구의 연구주제는 다음의 세 가지로 요약된다.

1. 100만 단어 규모의 해사영어코퍼스를 구축하는 과정을 위의 네 가지 관점에서 기술 하겠다.
2. 구축된 해사영어코퍼스의 어휘목록(word-list)과 키워드³⁾목록(keyword-list)을 통

1) LOB(Lancaster-Oslo/Bergen)는 전자 처리된 대표적인 참조 코퍼스로서 원어민의 영어사용을 대표할 수 있는 100만 단어 규모의 단어로 구성 되어 있는 데이터베이스이다(Kennedy, 1998).

2) BNC(British National Corpus)는 실제 사용되는 언어를 대표하기 위해 다양한 영역에서 1억 단어의 데이터를 수집하여 구축되었고, 전체 10%의 구어체를 포함하고 있어 참조코퍼스로 가치가 있다(Hunston, 2002).

3) 키워드는 연구대상의 코퍼스 어휘목록을 다른 코퍼스의 어휘목록과 비교하였을 때 그 빈도수가 현저하

해 해사영어의 특징이 LOB 코퍼스 또는 BNC 코퍼스와 어떻게 다른지를 빈도수로 비교 분석하고자 한다.

3. 해사영어에서 나타나는 품사별 분포도와 해사영어의 내용어 및 기능어의 어휘적인 특징을 빈도수와 N-gram⁴⁾ 분석을 통해 살펴보고 LOB 코퍼스 또는 BNC 코퍼스와 비교 분석하고자 한다.

1.3 논문의 구성

본 연구는 다음과 같은 구성으로 이루어진다. 제 II장에서는 본 연구의 이론적 바탕이 되는 코퍼스 언어학을 소개하고, 코퍼스 구축의 주요한 네 가지 관점을 서술한다. 그리고 해사영어를 특수목적영어의 관점에서 분석하고자 한다. ESP의 다양한 분야에서 어떻게 코퍼스가 이용되는지 몇 가지 사례연구를 통해 살펴볼 것이다. 제 III장에서는 실제로 해사영어 코퍼스를 생성한 과정을 서술하고, 제 IV장에서는 생성된 해사영어 코퍼스를 기반으로 일반영어 LOB 코퍼스 또는 BNC 코퍼스와 비교 분석을 통하여 어휘목록과 키워드목록 및 품사별 분포도를 살펴보고, 내용어와 기능어의 어휘적인 특징을 살펴본다. 마지막으로 제 V장은 논의된 연구의 결과를 정리한 결론부분으로 앞서 기술한 연구결과와 논의에서 도출된 내용들이 해사영어 교육에 이용될 수 있는 몇 가지 방안을 제시하고자 한다.

1.4 제한점

본 연구에는 다음과 같은 제한점이 있다.

첫째, MECO는 다양한 범주에서 뽑아낸 해사영어자료를 포함하고 있어 해사영어의 문어를 대표할 수는 있지만, 구어 데이터를 얻기 위해서는 항만청과 부두 등 정부기관의 허가가 필요하고 연구에 필요한 정보를 공개하지 않아 구어 데이터를 수집하는 것은 현실적으로 어려웠다.

계 높아 그 사용비중의 차이가 통계학적으로 유의미하다고 평가되는 단어를 말한다(고광윤, 2005).

4) n-gram은 대표적인 확률적 언어 모델의 하나로서 문장에서 주어진 순서에 따라 n개의 단어가 연속으로 나타나는 표현을 일컫는다(Winnie, Chris, John & Martin, 2009).

둘째, MECO의 데이터를 수집하기 위해 여섯 명의 적은 수의 전문가와의 인터뷰가 이루어진 아쉬움이 있다. 더욱 균형 있는 코퍼스구축을 위해서는 더 많은 전문가의 의견을 구할 필요가 있다.

셋째, 해사영어의 언어학적 특징을 분석함에 있어서 본 연구는 어휘적 특징을 품사 중심의 빈도수, 연어 관계, 어휘목록 중심으로 살펴보았지만 해사영어에 나타나는 파생어와 합성어의 분포와 특성 및 능동문과 수동문 등의 문장 패턴과 같은 다양한 각도에서의 분석도 필요하다.



II. 이론적 배경

2.1 코퍼스 언어학

2.1.1 코퍼스 언어학과 응용

Kennedy(1998)는 코퍼스를 언어학적 분석과 기술을 위한 기초를 제공할 수 있는 문어와 전사된 구어의 집합체라고 정의한다. 현대 언어학에서는 코퍼스를 단순한 언어의 집합체가 아닌 전자처리 과정을 거친 데이터가 컴퓨터가 판독할 수 있는 형태로 저장되어 있는 대용량의 언어자료로 정의할 수 있다. 1960년대 들어 컴퓨터의 발달로 방대한 양의 언어 데이터를 처리할 수 있게 되면서 최초로 전자적 형태를 가진 진정한 의미의 코퍼스가 등장하게 되고 이후 컴퓨터 기술이 급속하게 발전함에 따라 코퍼스 구축 및 코퍼스 기반 언어 연구는 더욱 활발하게 이루어지고 있다.

코퍼스 언어학(Corpus Linguistics)은 전산화된 언어자료의 집합체의 분석을 바탕으로 하는 언어 연구나 연구 방법론을 말한다(고광운, 2005). 오늘날 코퍼스는 언어를 분석할 수 있는 가장 새롭고 중요한 수단이 되고 있다(Gavioli, 2005). 기존의 언어학은 화자 또는 학자들의 직관에 따른 연구에 초점을 두었다면, 코퍼스 언어학은 실제적인 언어 사용을 바탕으로 언어를 연구하고 분석한다(McEnery & Wilson, 2001). 다시 말해서 원어민들이 실제로 사용하는 언어를 모은 대용량의 데이터를 사용하여 분석, 연구한다는 측면에서 기존 언어학과는 구별된다.

Biber(1998)는 코퍼스를 기반으로 다양한 언어학적 분석이 가능하다고 설명한다. 예를 들어 코퍼스는 콘코던스⁵⁾ 프로그램을 통해 어휘목록을 만들어 단어의 빈도수가 어떠한 특징을 나타내고 있는지 기술적으로 분석할 수 있다. 코퍼스 기반 연구에서의 단어의 의미는 다양한 문맥 속에서 파악할 수 있으며, 한 단어가 가지는 여러 가지 의미 역시 문맥을 통해 보여줄 수 있다.

또한 코퍼스는 단어가 다른 단어들과 어떠한 패턴으로 관계를 맺고 있는지 그리고 어떠한 단어와 연어 관계(collocation)를 이루고 있는지 분석할 수 있게 한다. 한 단어

5) 콘코던스(concordance)는 코퍼스의 데이터를 처리해 단어의 목록과 단어의 쓰임을 문맥과 함께 보여주는 소프트웨어를 일컫는다(McEnery & Wilson, 2001: 18).

가 특정한 단어와 더 자주 함께 쓰이는 규칙을 실제 텍스트를 통해 보여 주면서 사람들이 사용하는 단어 결합 패턴 방법을 설명할 수 있다. Sinclair(1991)는 연어를 둘이나 그 이상의 단어들인 한 텍스트 속에서 가깝게 나타나는 것으로 정의한다. 이런 연어의 연구는 WordSmith 5.0 프로그램을 통해 가능하며, L1, L2,..., Ln,... (왼쪽의 첫 번째 자리에 연결하는 단어, 두 번째 자리, ... n번째 자리에 연결하는 단어)와 R1, R2,..., Rn,... (오른쪽 첫 번째 자리에 연결하는 단어, 두 번째, ... n번째 자리에 연결하는 단어)의 종류를 다 나타내 줄 수 있어 빈도 높게 연결관계를 맺고 있는 표현들을 발견할 수 있다. 이러한 어휘의 특성 연구뿐만 아니라 코퍼스를 기반으로 문장속의 단어 배열, 품사별 빈도 등 문법적인 특징도 기술할 수 있다.

최근 단어의 배열을 연구하는 한 분야로서 단어열을 확률적으로 취급하는 언어 모델이 연구 개발되어 실용화되고 있는데 이것을 N-gram이라 한다(Winnie, Chris, John, & Martin, 2009). N-gram은 대표적인 확률적 언어 모델의 하나로서 문장에서 주어진 순서에 따라 n개의 단어가 연속으로 나타나는 표현들을 말한다. 단어 하나로 이루어진 1-gram은 unigram, 두 개로 이루어진 2-gram은 bigram, 세 개로 이루어진 3-gram은 trigram이라고 부른다. 예를 들어 “*The dog smelled like a skunk.*”라는 문장에서 *the dog smelled, dog smelled like, smelled like a, like a skunk*를 trigram이라고 부른다. 사실상 N-gram 모델은 언어모델을 만들어 내는데 상당한 영향을 줄 수 있어 많은 연구에 이용되고 있고 현대 통계적 언어 응용의 핵심 요소이다.

2.1.2 코퍼스의 종류

역사적으로 많은 코퍼스가 특정한 목적에 따라 구축되고 다양한 언어학적인 분석을 위한 기초를 제공하였다. 20세기 후반 컴퓨터 기술의 발달로 언어 데이터의 전자적 처리가 가능해짐으로 전자적 형태를 가진 진정한 의미의 1세대 코퍼스가 등장한다. 대표적인 예로 1960년대 Brown 대학교의 Francis와 Kučera 교수에 의해 만들어진 미국 영어 텍스트로 구성된 Brown 코퍼스와 1970년대 Lancaster 대학, Oslo 대학과 Norwegian Computing Center의 협력으로 구축된 영국영어를 대표하는 LOB 코퍼스가 있다. 이 코퍼스들은 당시 대용량으로 인식되었던 100만 단어 규모로 구축되었다. Brown 코퍼스는 15개 장르에서 약 500개의 다양한 문어 텍스트를 모아 구축되었고

LOB 코퍼스 역시 각 2000개의 단어로 이루어진 500개의 텍스트 샘플을 포함하고 있다. 이 두 코퍼스는 다양한 종류의 텍스트가 균형 있게 선택되었기 때문에 대표성이 있는 자료를 제공한다는 특징을 가지고 있다. 또한 이 두 코퍼스는 컴퓨터 기술의 발달로 고빈도 어휘를 실증할 수 있었고 문법의 대조 연구를 가능하게 했다. 그러나 Brown 코퍼스와 LOB 코퍼스는 초기 전자 코퍼스로 의미가 있고 미국영어(문어)와 영국영어(문어)의 비교에 중요한 자료가 되고 있지만, 구어 데이터를 가지고 있지 않아 원어민이 실제 사용하는 언어 전체를 대표할 수 없다는 한계점도 있다.

그 후 1990년대에는 컴퓨터 기술의 발달로 대규모의 언어자료를 바탕으로 BNC 코퍼스가 구축된다. BNC 코퍼스는 실제 사용되는 언어를 대표하기 위해 다양한 영역에서 언어 데이터를 모았으며 1억 단어의 규모로 구축되어졌다. 또한 BNC 코퍼스는 전체의 10%이지만 구어체를 포함하고 있는 가치 있는 코퍼스이다. 그러므로 BNC 코퍼스는 코퍼스 언어학자들에게 표본이 되고 있으며 일반 영국영어를 가장 잘 대표하므로 참조 코퍼스(reference corpus)로서의 가치가 있다(Hunston, 2002).

컴퓨터 기술의 발달과 함께 사전편찬을 위한 코퍼스 구축은 더욱 더 활발해졌다. 가장 대표적인 코퍼스 기반 사전 편찬 프로젝트는 Cobuild 프로젝트이다. 1980년대 새로운 영어 사전을 만들기 위해 코퍼스 구축이 시작되었고, 1987년에 Collins Cobuild English Dictionary가 출판될 시기에 Cobuild 코퍼스는 1천 3백만 단어의 규모에 이르게 되었다. 1990년대 Cobuild 프로젝트의 책임자인 John Sinclair는 Cobuild 데이터베이스를 BOE(Bank of English)라는 거대한 코퍼스로 확장시켜 1997년에 3억 개 단어가 구축되었다. 지금도 꾸준히 데이터가 추가되고 있는 이 코퍼스는 새로 생성되는 단어와 그 단어들의 쓰임을 관찰하고 분석할 수 있게 한다. Collins Cobuild 프로젝트는 코퍼스 기반으로 사전편찬, 문법 설명, 언어 교수 방법 제시를 가능하게 했을 뿐만 아니라 언어학적 연구와 상업적인 분야를 연결시키는 새 지평을 열게 되었다.

2.1.3 코퍼스를 활용한 어휘 분석 선행연구

LSWEC(Longman Spoken and Written English Corpus, Biber et al., 1999)는 약 4천만 개의 단어로 구성되어 있으며, 여섯 개의 사용영역(register)으로 이루어져 있다. 그리고 그들은 여섯 개의 사용영역 중 네 개 영역 즉, 대화(conversation), 소설

(fiction), 뉴스(news), 학술적인 글(academic prose)에서 선정된 영어자료를 기반으로 LGSWE(Longman Grammar of Spoken and Written English)라는 문법서를 만들었다. 이 문법서는 LSWE를 바탕으로 영어 사용자들이 어떻게 영어를 사용하고 있는지 다량의 실제 자료를 보여주는 문법서이다. LGSWE는 각 사용영역에 속하는 텍스트 정보를 선정해서 상호 비교 분석함으로써 문법 정보, 패턴, 구조 등이 각 사용영역마다 어떻게 다르게 나타나는지 그 결과를 정밀하게 분석한 내용으로 이루어져 있다. 예를 들어 내용어(content words)와 기능어(function words)가 대화 영역과 뉴스 영역에서 각각 어떤 비율로 나타나는지 보여준다(Biber et al., 1999: 61). 또한 네 개의 영역에 나타난 내용어(명사, 동사, 형용사, 부사)의 빈도수를 비교해서 비율분포도를 나타내고 있다(Biber et al., 1999: 65). 기능어 역시 전치사, 조동사, 대명사, 접속사, 한정사 등의 영역별 분포도를 나타내어 각 기능어들이 장르별로 다른 비율로 나타남을 보여주고 있다. 예를 들어 대화문에 전치사의 비율이 내용어의 비율보다 현저하게 높고, 조동사는 대화문과 소설에서 높은 비율 나타내는 것을 보여준다(Biber et al., 1999: 91-92). 이처럼 LGSWE는 코퍼스 분석을 통해 각 장의 문법에 대한 결과를 제시하고 359개 이상의 표와 그래프를 이용하여 어휘, 문법 구문 등의 영역별 비교 분석 자료를 제시한다. 이와 같은 일반영어의 코퍼스 기반 어휘, 문법적인 분석 결과는 해사영어가 일반 영어와 어떠한 차이가 있는지를 다각적으로 비교분석가능하게 한다.

김소라(2009)는 어린이용 국내 영자신문 네 종류의 기사를 수집하여 코퍼스를 구축하고 어린이용 BBC 인터넷 판 신문의 기사를 수집하여 코퍼스(CBBC)를 구축하여 Antconc 3.2.0w 콘코던스 프로그램으로 어휘목록을 추출하고 연어를 분석하였다. 여기서 그녀는 어휘목록을 통해 동사, 명사, 형용사, 부사, 전치사의 상위 순위의 어휘를 추출하였고, 그 목록을 CBBC와 BNC 코퍼스의 children 장르와 비교 분석하였다. 국내영자신문 어휘목록을 여덟 개의 어휘 분석 대상과 비교하여 284개의 어휘를 추출한 후 용례를 찾아 연어 표현 목록을 작성하였다. 코퍼스 프로그램을 활용하여 구축한 이와 같은 연어 표현 목록은 영자신문을 활용한 영어수업을 진행하고자 하는 교사와 학습자들에게 실제적인 참고자료가 될 수 있다.

정미애(2007)는 비즈니스 관련 실제 자료를 수집하여 코퍼스를 구축하고 분석하여 특수목적영어의 한 분야인 비즈니스 영어의 어휘적 특성을 살펴보았다. 어휘의 빈도수를 통해 상위 어휘목록을 추출하였고 텍스트 상 현저하게 사용되는 키워드를 찾아내어

참조 코퍼스(BNC_World)와 비교분석하였다. 키워드 분석을 통해 도출된 키워드는 관련 분야인 전문영역에서 자주 사용되는 어휘를 학습자에게 제시하여줌으로서 효율적인 학습에 도움을 줄 수 있다고 주장한다.

Ishikawa(2009)에 따르면 대화에서 결정적인 역할을 하는 연접어구(phraseology)는 준원어민성(nativeness)을 알 수 있는 좋은 기준이 된다. 그러나 연접어구는 언어학적 접근이 아닌 관습적으로 선택된 단어들의 조합이므로 영어가 모국어인 학습자들에게는 학습하기에 어려운 영역이므로 영어를 배우는 학습자들의 코퍼스를 N-gram 분석을 통해 양적, 객관적인 분석을 시도했다. 일본 Asian 대학교 학생들 중 일본학생들과 영어를 모국어로 하는 학생들의 에세이에 나타난 연접어구의 특징을 비교했다. 두 그룹의 어구를 2gram, 3gram과 4gram으로 비교분석하였다. 일본학생들의 에세이 코퍼스(Corpus of English Essay Written by Japanese University Students, CEEJUS)는 17만개의 출현어휘(token)로, 영어를 모국어로 하는 학생들의 에세이 코퍼스(CEENAS: Corpus of English Essay Written by Native Speakers)는 3만 7천개의 출현어휘로 이루어져있다. 두 코퍼스에서 나타난 2-4gram을 비교한 결과 두 집단 사이에 연접어구의 사용이 다른 것으로 나타났다. 일본학생들은 *I think*라는 표현을 문장을 시작할 때 지나치게 많이 쓰고, *but also, a lot*과 같은 숙어를 역시 많이 썼다. N-gram 분석으로 살펴본 원어민의 고빈도 연접어구는 *right to, I disagree, can be, on the* 등 이었다. 그리고 일본 학생들이 *not*을 많이 쓰는 반면 원어민들은 문장의 부정성을 위해 *not*을 직접적으로 쓰지 않고 다른 표현을 사용했다. 예를 들어 *I don't agree* 대신 *I disagree*를 더 많이 사용했다. 일본 학생들은 전치사구의 사용 빈도가 낮은 반면에 영어 원어민들은 명사 또는 부사구를 수식하기 위해 전치사 구를 자주 사용하였다. 코퍼스를 기반으로 한 이와 같은 연구를 통해 연접어구 사용에 대해 보다 많이 밝혀 낼 수 있었으며, 원어민과 비원어민 사이의 연접어구 사용의 차이를 더욱 분명하게 나타냈다. 학습자 코퍼스로부터 얻은 결과들이 일본의 영어 교육에 대한 새 지평을 여는 데 영향을 줄 수 있을 것이라고 Ishikawa는 말한다.

허미영(2007)은 고등학교 수준의 독해교재로 코퍼스를 구축해 사용된 영어어휘를 분석 평가하여 영문독해교재의 집필이나 학습자의 교재 선정에 도움이 되는 참고 자료를 제공하고자 하였다. WordSmith 4.0 프로그램을 이용하여 단어수와 출현어휘수의 비율(type-token ratio, TTR)를 살펴보고 각 독해교재에 반복 사용되는 어휘수를 파악하

여 독해집의 난이도를 평가하였다. 또한 영어권 코퍼스(Brown 코퍼스, LOB 코퍼스, BNC 코퍼스의 빈도분석표)와 독해교재 코퍼스의 최고빈도 어휘를 상호 비교하였다. 그 결과 영어권 코퍼스와 독해교재의 최고빈도 어휘가 일치하였다는 것을 발견하였고, 독해교재의 어휘 구성과 사용비중이 대체로 영어다운 것으로 판단하였다. 독해교재에 쓰인 인칭대명사를 영어권 코퍼스와 비교분석한 결과 독해교재 코퍼스에는 *it*을 제외한 3인칭 대명사보다 1인칭, 2인칭 대명사를 훨씬 높은 빈도로 사용하고 있었다. 독해집 코퍼스의 *be*동사와 *have*동사를 중심으로 시제를 연구한 결과 영어권 코퍼스는 과거시제와 현재시제가 거의 비슷하게 사용하였지만, 독해교재 코퍼스는 월등하게 현재시제를 과거시제보다 자주 사용하였다는 것을 발견하였다.

김은정(2007)은 고등 5종 교과서를 15권으로 소규모 코퍼스를 구축하고 사용된 양상조동사의 의미 분포와 사용 경향을 조사하였다. 참조 코퍼스로는 LSWEC를 활용하였다. 의사소통을 잘 하기 위해서는 화자의 자신의 의사를 상황에 맞게 표현하고 감정을 명확하게 나타내어야 한다는 점에서 양상 조동사는 화자의 감정을 드러낼 수 있는 중요한 문법적인 수단 중 하나이므로 양상조동사를 코퍼스 언어학적인 방법으로 분석하였다. LSWEC에서 양상조동사로 분류된 아홉 개의 조동사 중에서 의지와 예측의 의미로 사용되는 *shall*의 빈도를 살펴보면 5종 교과서와 LSWEC에서 빈도가 낮았으며 예측의 의미보다는 의지의 의미로 조금 더 많이 사용되고 있었다. 각각의 5종 교과서를 대화문과 학술산문의 영역으로 나누어 양상 조동사를 살펴본 결과 LSWEC와 유사한 분포를 보였다. 그러나 *shall*의 경우에는 두 개의 교과서의 대화문에서 의지와 예측의 *shall*의 의미가 누락되어 있었고, 학술산문 영역에서는 *shall*이 한 번도 나타나지 않았다. 이러한 교과서들은 양상조동사의 의미를 고르게 제시하지 못하는 것으로 나타났다. 그녀는 원어민의 양상조동사의 사용 경향을 교과서에 적극 반영하여 다양한 의미를 제시하고 *shall*과 같이 사용빈도가 낮은 조동사를 누락시키지 말고 고르게 제시해야 한다고 주장했다.

2.2 코퍼스의 구축

해사영어코퍼스 구축을 위해 본 연구는 McEnergy and Wilson(2001)의 코퍼스 구축의 네 가지 필수요건을 고려하였다. 코퍼스는 라틴어로 '*body*'라는 의미이고 텍스트의

집합체로 정의된다(McEnery & Wilson, 2001). 그러나 문어와 구어 텍스트를 어떤 기준도 없이 수집하여 코퍼스로 구축한다면, 코퍼스로서의 의미가 없기 때문에 McEnery and Wilson(2001)은 현대의 코퍼스를 구축할 때 아래의 (1)과 같은 네 가지의 요건을 고려해야 한다고 주장한다.

(1) 코퍼스 구축의 필수요건

- a. 대표성(representativeness)
- b. 한정된 크기(finite size)
- c. 기계해독가능 형태(machine-readable form)
- d. 표준화된 참고자료(standard reference)

첫째, 대표성은 수집된 텍스트가 코퍼스 구축의 목적과 관련된 언어를 대표하는 것을 의미한다. Leech(1991)는 수집된 언어 텍스트가 특정 언어로 일반화될 수 있을 때 코퍼스는 대표성을 가진다고 말한다. 예를 들어 초기 코퍼스 모델인 Brown 코퍼스와 LOB 코퍼스는 신중하게 디자인되었기 때문에 각각 미국영어와 영국영어의 문어를 대표할 수 있고, 일반적으로 사용되는 영국영어의 90% 문어 텍스트와 10% 구어 텍스트로 구축된 BNC 코퍼스 역시 영국영어를 대표할 수 있다(Kennedy, 1998). 해사영어코퍼스는 해사영어 연구를 위해 다양한 종류의 언어 데이터를 수집하고, 각 데이터의 크기의 균형을 고려하였다. 단순한 말뭉치가 아닌 특정 언어로 일반화할 수 있는 코퍼스의 대표성은 코퍼스 생성에서 가장 중요한 요소이다.

두 번째 요소는 코퍼스의 대표성과 관련되어 있는 코퍼스의 크기이다. 코퍼스 크기는 총 단어 수뿐만 아니라 카테고리의 수, 또 각 카테고리 내의 샘플의 수, 그리고 그 샘플에 포함되는 단어 수까지 고려해야 한다(Kennedy, 1998). 1960년대 코퍼스 1세대에는 100만 단어 규모를 대용량으로 인식하였으나 컴퓨터 기술의 발달로 현대에는 수백만 단어를 포함한 코퍼스 구축이 어렵지 않다. 이런 점에서 McEnery and Wilson(2001)은 코퍼스의 규모를 단순히 확장시키기보다 코퍼스의 대표성을 가질 수 있는 텍스트 수집을 강조하였고 본 연구에서는 소규모 코퍼스로서 합리적인 크기로 여겨지는 100만 단어 규모의 코퍼스를 구축하였다.

세 번째 요소는 코퍼스 데이터가 컴퓨터가 읽고 처리할 수 있는 형태를 가져야 한다는 것이다. 대부분의 경우 코퍼스는 표준의 마크업(mark-up) 언어, 예를 들어 sgml,

html, xml 등을 사용할 수 있다(Leech, 2004). 본 연구의 해사영어코퍼스 데이터는 WordSmith 5.0이 처리할 수 있는 형태의 text파일로 변환하였다.

마지막으로 표준화된 참고자료는 코퍼스에 나타나는 언어 데이터 이외의 텍스트 장르, 작가 정보, 출판된 시기 등의 자료의 출처를 나타내는 것을 말한다. 이것은 코퍼스를 구축하는 데 필수적인 부분은 아니지만, 언어의 다양성을 나타내는 참고자료가 코퍼스에 포함되어 있다는 것을 알려주어 데이터의 신뢰도를 높일 수 있다. 이런 참고자료는 다른 학자들에 의해 다각적으로 사용될 수 있다(McEnery & Wilson, 2001). 특히 코퍼스 텍스트의 자료의 출처를 제시하는 것은 사회 언어학자에게 매우 중요한 연구 자료가 되며 다양한 후속연구에 도움을 줄 수 있다. BNC 코퍼스는 다음의 <그림 1>과 같이 언어데이터의 종류, 출처, 만들어진 시기, 작가의 성별, 나이 등 다양한 참고자료를 자세하게 보여준다.

<그림 1> BNC에 나타나는 표준화된 참고자료

No	Filename	
1	A08 1036	
2	A0D 32	A08: The big glass. Sample containing about 40097 words from a book (domain: imaginative)
3	A0D 35	
4	A0G 13	<i>Spoken or Written: Written</i>
5	A0H 67	<i>Number of Words (tagged items): 40,186</i>
6	A0N 13	<i>Average sentence length (w-tags per s-unit): 12.5542</i>
7	A0P 42	<i>Derived text type: Fiction and verse</i>
8	A12 76	<i>Genre: W:fict:prose</i>
9	A12 96	<i>Text type: Written books and periodicals</i>
10	A12 96	<i>Publication date: 1985-1993</i>
11	A12 96	<i>Age of Author: 45-59</i>
12	A12 96	<i>Domicile of Author: UK and Ireland</i>
13	A12 16	<i>Sex of Author: Male</i>
14	A14 39	<i>Type of Author: Sole</i>
15	A14 39	<i>Age of Audience: Adult</i>

2.3 특수목적 영어

2.3.1 특수목적 영어

전통적으로 특수목적 영어는 특별한 상황에서 또는 전문적 지식이 필요한 영역에서 사용되는 영어의 연구로 정의된다(Gavioli, 2005). 일반적인 언어 능력의 향상에 목표를 둔 일반 영어와는 달리 특수목적 영어는 업무 수행 능력의 향상 또는 전문적인 지식의 습득을 목표로 두고 있다. 그러므로 특수목적 영어는 실제 사용되는 언어를 기반으로 한 실용적인 학습을 강조하고 학습 자료로서의 실제 언어 연구에 초점을 둔다. 그 결과 실제 언어의 용례를 제시할 수 있는 코퍼스가 특수 목적 영어 분야에서 효율적인 하나의 연구방법으로 활용되고 코퍼스를 기반으로 한 어휘적 특성, 문법적 특성 등의 언어학적 연구가 더욱 활발해졌다.

Gavioli(2005: 8)에 따르면 특수목적으로 구축된 코퍼스는 특별한 언어의 특징을 묘사하는 데 일반 코퍼스보다 훨씬 유용하다고 말한다. 특수목적 영어를 위해 구축된 코퍼스는 일반 영어와는 차이를 보이는 핵심 어휘와 문법적인 특징을 자세한 용례를 통해 제공할 수 있기 때문에 실용적이다. 또한 특수한 단어의 의미를 다양한 문맥 속에서 나타내어 의미 분석을 가능하게 한다. 많은 학자들은 특수 목적 영어의 연구를 위해 코퍼스를 구축하고 특수목적 영어에서 나타나는 어휘적, 문법적인 분석을 위한 코퍼스 기반 연구의 중요성을 꾸준히 강조하고 있다.

특수목적 영어로서 비즈니스 영어, 엔지니어링 영어, 메디컬 영어 분야에서 코퍼스 기반으로 한 언어학적 분석이 많이 이루어지고 있다. 최근 Nelson(2006)은 비즈니스 영어의 특징 분석을 위해 BEC(Business English Corpus)를 구축했다. 그는 코퍼스를 기반으로 비즈니스 영어의 의미 운율(semantic prosody) 관계를 분석하였다. 그가 사용한 BEC 코퍼스는 56%의 문어 텍스트와 44%의 구어 텍스트로 구성이 되어 있고 100만 단어 규모이다. BEC 코퍼스 구축을 위해 미국과 영국에서 수집된 팩스, 이메일, 보고서, 회의 내용, 신문, 뉴스, 라디오, TV, 계약서 등 다양한 종류의 텍스트가 사용되었다. 비즈니스 영어의 의미 운율을 살펴보기 위해 일반영어를 대표할 수 있는 참조 코퍼스로 BNC 코퍼스를 사용하여 비즈니스 영어와 일반 영어를 대조 분석하였다. 연구 결과 일반영어와는 다른 독특한 의미 운율을 가진 단어의 집단을 찾을 수 있었다.

예를 들어 단어 *big*은 *companies, money, people*과 같은 단어들과 31.86%의 비율로 연결관계를 맺고 있지만, BNC에서 *big*은 세 단어와의 연결관계의 비율이 12.87%로 낮게 나타났다. 비즈니스 영어의 어휘가 일반영어와 다른 의미적 결합관계를 가지는 것을 실증적으로 보여주고 있는 예이다.

Olga(2006)는 특수 목적 영어인 엔지니어링 영어를 가르치고 배우는 데 있어 코퍼스가 중요한 역할을 한다고 주장하면서 SEEC(Student Engineering English Corpus)를 구축했다. SEEC는 엔지니어링을 전공하는 학생들의 전공서에서 언어 데이터를 수집하여 13종류의 텍스트파일을 구축하였고, 총 1,986,595개의 단어를 포함하고 있다. 데이터 분석을 위해 WordSmith Tools를 사용하였고, Olga는 이 연구를 통해 어휘목록에서 100회 이상 나타나는 가장 빈도수가 높은 어족을 뽑아 학생들을 위한 엔지니어링 영어어휘 요강의 근간을 마련했다. 또한 이 연구에서 참조 코퍼스로서 COBUILD Bank of English Corpus, BNC 코퍼스와 SEEC 코퍼스의 비교를 통해 어휘의 빈도수를 분석하였다. SEEC 코퍼스에서 가장 빈도가 높은 내용어는 Bank of English Corpus와 BNC 문어 코퍼스에서는 높게 나타나지 않았다. 예를 들어 *force, pressure, system, energy, stress* 등의 준 전문(sub-technical) 어휘의 빈도가 SEEC에서는 높지만, 나머지 참조 코퍼스에서는 빈도가 전혀 높지 않았다. 흥미로운 연구결과는 빈도가 높은 준 전문 어휘는 전문적인 의미(technical sense)보다 비전문적인 의미(non-technical sense)로 더 자주 쓰인다는 것을 보여주었다는 것이다.

Marco(2000)는 메디컬 연구 논문이라는 특수한 텍스트를 모아 코퍼스를 구축하고 메디컬 논문에 나타나는 언어학적 특징을 분석하였다. 특히 연접어구(phraseology) 구조 분석에 초점을 두었다. 이 연구를 위한 코퍼스는 가장 저명한 메디컬 저널 사이트에서 선정한 100개의 논문으로부터 수집한 텍스트로 구성되었고, 총 298,457개의 단어를 포함하고 있다. 이 코퍼스에서 가장 빈도가 높게 나타난 어구는 *the ... of* 이고, 다음으로 *a ... of, be ... to* 등이 있다. 특히 *the start of*는 메디컬 텍스트에서만 나타나는 독특한 표현이다. 또한 *a ... of* 구조에서 확인할 수 있는 어휘의 목록을 Renouf and Sinclair(1991)의 Birmingham Collection of English 텍스트를 기반 연구에서 얻은 어휘와 비교 분석하였다. 메디컬 코퍼스에서 *a ... of* 구조에서 나타난 어휘는 *history, study, proportion* 등이었으나, Birmingham 텍스트에서는 *lot, kind, number* 등의 어휘가 사용되었다. 코퍼스 기반 연접어구 연구에서 다른 종류의 담화에서는 빈도

가 낮은 표현이 메디컬 연구 논문에서는 확연히 자주 쓰이는 것을 확인할 수 있었다. Marco는 학생들이 본 연구에서 분석한 구조를 알고 쓰임새를 인지한다면 메디컬 논문을 이해하고 작성하는 능력을 향상시킬 수 있는 토대가 될 수 있다고 주장한다.

2.3.2 해사영어

선박의 운항, 선박간의 커뮤니케이션, 엔지니어링, 안전을 위한 규약의 이행 등의 중요한 업무를 효율적으로 처리하기 위해 사용되는 해사영어는 다른 나라 출신 선원들간의 원활한 의사소통을 위한 중요한 요소이다. 점점 그 중요성이 강조되고 있는 해사영어를 Trenkner(2000: 7)는 다음의 (2)와 같이 정의내리고 있다.

(2) 해사영어는 국제 해양 커뮤니티간의 의사소통을 위한 수단으로서 사용되는 영어를 말하며 선박 운항의 안전과 해상 무역 편의를 제공하는 데 기여할 수 있다.

Trenkner(2000)뿐만 아니라 많은 학자들은 해사영어를 정의하고 연구하고 있다. Pritchard(2003)는 해사영어는 해사 커뮤니티에서 사용되는 언어로서 정보를 교환하고 지식을 전달하는 데 사용된다고 정의한다. 또한 Cole(1999)은 해사영어를 해사 커뮤니티의 표준화된 의사소통 수단이자 해사영어 학습과 교육을 위해 필요한 자료로 인정될 수 있는 영어라고 정의한다. 다양한 연구에서 해사영어를 정의한 것을 바탕으로 본 논문에서는 코퍼스 구축을 위해 해사영어를 해사업무에서 필요한 의사소통의 수단이자 학습자 또는 선원들이 배워야 할 전문적인 지식과 의사소통을 위한 규정화된 표현과 해상안전을 위한 국제협약에서 사용되는 영어라고 정의한다.

해사영어는 IMO에 의해 해상에서 쓰이는 공용어로 채택되었고 STCW 95 Convention은 선원들이 해사영어를 이해하고 능숙하게 사용할 수 있어야 한다고 규정한다(Ioana, 2009). 또한 IMO는 의사소통의 문제를 최소화하기 위해 SMCP와 같은 규정화된 해사영어 모듈을 제공한다.

IMO는 해사영어의 패턴에 더 많은 관심을 가지고 해사영어의 특별한 모듈, 즉 Model course 3.17 Maritime English(IMO, 2009)를 제공하지만, 여전히 의사소통의 문제로 사고가 빈번하게 발생하고 있으므로 이러한 모듈 제공이 효율적인 의사소통을

위해서는 충분하지 않다(Trenkner, 2009).

Horck(2005)에 따르면 의사소통은 서로 이해하고 협력하면서 효과적인 업무를 위한 가장 중요한 도구이므로 해사교육훈련원은 영어 교육에 더 주목할 필요가 있다고 강조한다. 그리고 Lozinska(2009)는 실제적인 해사영어 학습을 위한 자료를 인터넷을 통해 제시하지만 인터넷 상의 단어 항목의 나열만으로 해사영어에 가지는 언어학적 특징을 설명하기는 어렵고 의사소통의 문제점의 해결책을 제시할 수 없다고 주장한다. 또한 Ioana(2009)는 문맥을 통한 단어의 쓰임과 의미파악이 해사영어 학습에 도움이 된다고 말한다. 즉 학생들이 알고 있는 일반적인 의미의 단어와는 다르게 쓰이는 다양한 어휘적 특징(단어 의미, 전문 용어, 연어 관계)이 문맥을 통해 쉽게 전달되며 학생들이 더 정확하게 내용을 해석하고 이해하는 데 도움을 줄 수 있다고 한다.

위에서 살펴 본 바와 같이 해사교육에서 영어 교육의 중요성이 무엇보다 강조되고 많은 학자들이 효과적인 해사영어 교육을 위한 연구를 하고 있다. 그러나 단순히 학습 자료를 제시하거나 해사영어 교육의 필요성을 주장하고 있을 뿐 체계적인 해사영어 연구 방법을 제시하지 않고 있다. 해사영어 교육이 효율적으로 이루어지고 해사영어의 의사소통에 도움이 되기 위해서는 실제 해사 커뮤니티에서 사용되는 언어사용 패턴과 어휘를 체계적으로 분석해야 한다. 해사영어와 같이 특수 분야에서 사용되는 언어를 연구하기 위해서는 코퍼스 구축과 코퍼스를 기반으로 한 특수목적 영어 연구가 유용하다(Gavioli, 2005).

앞서 살펴 본 바와 같이 비즈니스, 엔지니어링, 메디컬 등 다양한 분야의 영어를 연구하기 위해 코퍼스를 구축하고 활용하고 있지만 구축된 해사영어 코퍼스를 바탕으로 한 분석이 없기 때문에 현재까지 해사영어의 실증적인 언어학적 연구가 제한적이다.

Ⅲ. 연구방법

3.1 연구대상 및 도구

본 연구에서 해사영어코퍼스(MECO) 구축을 위해 한국해양대학교에 재직 중인 여섯 명의 해사영어 강의 담당 교수님들의 조언을 바탕으로 데이터를 선정하였다. 해사영어 코퍼스는 해사업무 상 중요한 비중을 차지하는 국제적 규약과 전문적 지식, 해사 관련 미디어 기사, 그리고 해사영어교육에서 사용되는 문어 텍스트로 구성되어 있으며 데이터의 분류와 구성내용은 아래의 <표 1>에 제시되어 있다.

본 연구에서 구축된 MECO의 분석을 위해 콘코던스 프로그램인 WordSmith 5.0 Tools(Scott, 2009)를 사용하였다. WordSmith 5.0은 기본적으로 어휘의 쓰임을 어휘 목록(word-list), 단어 빈도수(word frequency), 키워드(keyword)와 콘코던스(concordance), 연어(collocation)를 나타내주는 소프트웨어이다. WordSmith 5.0은 콘코던스를 통해 어휘들의 문장내의 쓰임을 나타내주어 실증적인 연구의 바탕이 된다. 또한 WordSmith 5.0은 각 어휘의 빈도와 비율에 대한 정보와 함께 전체 단어수와 출현어휘수의 비율인 TTR(<표 6>참조), 문장의 수, 단어의 평균길이 등 해당 텍스트에 대한 다양한 통계 정보를 제공한다(고광윤, 2009).

전치사구의 연구를 위해 N-gram을 분석한 프로그램은 Multilingual Corpus Toolkit (Piao, 2010)으로서 문장인식, 콘코던스, 연어 추출 등 다양하고 유용한 기능을 제공한다.

3.2 연구절차 및 방법

Hong and Jhang(2010)은 2010년 7월부터 10월까지 위의 2.2장의 네 가지의 코퍼스 구축 요건에 기초하여 해사영어코퍼스(MECO)를 구축하였고, 본 연구자는 연구 프로젝트의 보조 연구원으로서 인터뷰, 데이터 수집 및 분류의 업무를 맡았다.

구축된 코퍼스 MECO는 WordSmith 5.0 프로그램으로 총 어절 수, 어휘의 종류, 문장 수 등의 기본적 통계정보를 분석하였다. MECO의 어휘목록을 WordSmith 5.0을 통해 도출했다. 해사영어와 일반영어의 비교분석을 위해 MECO와 규모가 비슷한 100만

단어 규모의 문어중심 코퍼스인 LOB 코퍼스를 MECO의 주된 참조 코퍼스로 선택하였고, BNC 코퍼스 역시 비교분석을 위해 참조 코퍼스로 사용하였다. 코퍼스 간의 비교 분석은 빈도수 비교에 주안점을 두었고, 빈도수에 따라 나타난 어휘적인 특징을 기술적으로 분석하는 연구 방법을 사용하였다. MECO의 어휘목록과 LOB의 어휘목록을 추출하여 두 코퍼스에 나타나는 어휘를 비교 분석하였으며 LOB 코퍼스를 참조한 MECO의 키워드목록을 추출하였다. 또한 CLAW5(C5) 프로그램으로 MECO에 품사를 부착(POS tagging)하는 작업을 하여 각 품사별 빈도수와 그 종류를 분석할 수 있었다. 본 연구는 품사를 내용어와 기능어로 분류하여 자세히 살펴보았고, 각 품사별 고빈도 어휘는 품사별로 편의상 재분류하였다(<부록 2-2> 참조). 재분류의 모든 절차는 Perl 프로그램을 이용하였다. 먼저 코퍼스 데이터를 출현어휘단위로 분리하였으며 분리된 어휘를 품사별로 정렬(sorting)하였고 정렬된 단어들은 빈도순으로 재정렬하였다. 이렇게 얻은 데이터를 엑셀 프로그램을 이용하여 다시 통계를 내어 그래프를 만들었다.⁶⁾

3.2.1 해사영어코퍼스(MECO)의 구축

3.2.1.1 MECO의 대표성

아래 <표 1>에서 보는 바와 같이 MECO가 대표성을 가지기 위해서 여섯 개의 해사 영어 문어 텍스트가 사용되었다. MECO 구축을 위한 적절한 텍스트 수집을 위하여 <부록 1>에 실려 있는 질문지와 같이 한국해양대학교에서 해사영어 강의를 담당하시는 여섯 분의 교수님들께 여덟 가지의 질문을 하고 얻은 조언과 의견을 바탕으로 텍스트를 수집하였다. 코퍼스 구축에 있어서 전문가의 조언을 참고로 하는 ‘민족지학적(ethnographic)’ 논점은 꾸준히 연구되어 왔다. 많은 학자들은 그 분야의 전문가의 조언이 대표성을 극대화할 수 있다고 주장한다(Flowerdew, 2005; Hyland, 1998). 특히 Flowerdew(2005)에 따르면 민족지학적인 면은 사회·문화적인 양상을 분석하는 데 중요한 역할을 한다고 한다. 따라서 MECO 구축에 있어 전문가의 조언을 바탕으로 대표성을 높임으로서 해양 문화적인 면에서 MECO의 질을 향상시킬 수 있다.

각 텍스트의 내용은 질문지의 답을 이용하여 결정되었다(<부록 1> 참조). 먼저 질문

6) 한국해양대학교 IT공학부 김재훈 교수님의 도움으로 Perl 프로그램을 이용하여 데이터를 추출하여 재정렬하였다. 이 자리를 통해 교수님의 자상한 가르침과 헌신적인 도움에 깊은 감사를 드립니다.

지의 답을 통해 텍스트를 수집한 후 전문가들이 텍스트가 적절한지 결정하고 난 뒤, 논의를 통해 전문가들에게 공통된 텍스트 주제에 동의를 구하는 두 가지 절차를 따랐다.

<표 1> MECO 구축에 사용된 텍스트 종류 (Hong and Jhang, 2010)

텍스트 종류	내용
Journals	<ul style="list-style-type: none"> • RIN-Journal of Navigation • MAIB Investigation • Marine Policy and Management
Maritime Law	<ul style="list-style-type: none"> • SOLAS • STCW • MARPOL • COLREG
Media Texts	<ul style="list-style-type: none"> • Maritime Reporter • Seaways • Safety at Sea • Lloyd's List • IMO Latest News • WMU Event • WMU News • Maritime Today News • Sea Discovery News • World Marine News • ITF Global News
Set Language (SMCP)	<ul style="list-style-type: none"> • SMCP • SMCP Sample Data
Textbooks	<ul style="list-style-type: none"> • Carriage of Goods by Sea • Elementary of Shipping • Maritime Economics • The International Maritime Language Programme • Maritime Insurance • Principles of Maritime Law • Time Charters • Shipping Law
Others	<ul style="list-style-type: none"> • IMEC Proceedings • IMO Documents: Ship Security Officer • IMO Documents: Environment • IMO Documents: Introduction to IMO • IMO Documents: Officer in Charge of a Navigation Watch

3.2.1.2 MECO의 크기

아래 <표 2>에서 나타나듯이 MECO의 크기는 대략 100만 단어이다. 코퍼스 구축에 있어 코퍼스 크기는 가장 중요한 요소 중 하나이나 대표성을 가지기 위해 코퍼스 크기에 대한 정확한 기준은 정해져 있지 않다(Kennedy, 1998). Pearson(1998)에 따르면 100만 단어 규모를 특수 코퍼스의 합리적인 크기로 권한다. 이에 따라 MECO를 100만 단어 규모로 구축하였다.

코퍼스 사이즈의 주된 논점은 각 텍스트 장르의 크기에 있다. 각 텍스트 장르의 크기가 대표성을 띄기 위해 절적하게 균형 잡혀 있지 않다면 코퍼스의 신뢰도에 영향을 미칠 수 있지만, Kennedy(1998)는 각 장르들의 균등한 크기가 대표성을 보장하지 않는다고 주장한다. 실제 해양 공동체에서의 모든 장르의 크기는 균등하게 구분되어 있지 않다. 오히려 세부 장르의 크기는 해양 공동체의 현실을 반영해야한다. 예를 들어 <표 2>에서 미디어 텍스트의 크기가 가장 큰 이유는 해사 커뮤니티의 많은 사람들이 다른 장르보다 미디어 텍스트를 더 많이 접하기 때문이다.

데이터를 수집한 후 각 데이터의 내용과 알맞은 크기, 전공서 별 데이터에 반드시 포함되어야 할 내용과 크기, 각 데이터들이 전체 코퍼스에서 차지하는 비율 등에 대해서도 여섯 분의 해사영어 담당 교수님과 여러 차례 논의가 있었다. 해사영어코퍼스 구축에 필요한 데이터의 기준을 명확하게 하기 위해 인터뷰를 통해 얻은 자료의 내용과 크기에 대한 전문가와의 토론이 필수적이므로 충분한 상의를 통해 아래의 <표 2>와 같이 텍스트의 종류와 크기를 결정하였다.

<표 2> 텍스트 종류별 크기 (Hong and Jhang, 2010)

	텍스트 종류	사이즈
1	Journals	213,100
2	Maritime Law	177,162
3	Media Texts	276,258
4	Set language (SMCP)	36,852
5	Textbooks	175,368
6	Others	152,099
	계	1,030,839

3.2.1.3 MECO의 기계해독 가능형태

WordSmith 5.0(Scott, 2009) 또는 AntConc 3.2.1.W(Laurence, 2007)와 같은 대부분의 콘코던스 프로그램은 text파일을 읽을 수 있기 때문에 MECO 데이터를 text파일로 변환하여 입력하였다.

MECO의 데이터는 품사가 부착되었다. 초기 코퍼스 연구를 위해 컴퓨터는 단어들의 동형이의어를 구별해 낼 수 없었다. 즉 언어 데이터의 문법적인 범주가 모호했기 때문에 정확한 연구에 문제점이 제기되었다. 그리하여 각 단어 마다 문맥 속에서 어떻게 쓰이는지를 보여주기 위해 '태그(tag)'가 자동적으로 달릴 수 있는 프로그램이 디자인되기 시작했다(Kennedy, 1998). 데이터에 언어학적인 정보를 붙여주는 것으로 정의되는 품사부착은 문법적 부착(grammatical tagging) 또는 단어군 부착(word-class tagging)이라고도 불린다(Garside, Leech, & McEnery, 1997). 1971년에는 87개의 태그를 가진 TAGGIT⁷⁾ (Greene & Rubin, 1971) 프로그램이 개발되어 Brown 코퍼스에 품사부착이 되었고, 1978년에서 1983년 사이에는 새로운 자동 품사부착 시스템인 CLAW1 프로그램이 개발되어 언어 데이터의 통사적 특징들을 나타내주어 많은 언어학적 연구를 가능하게 하였다(Kennedy, 1998). 오늘날 더 정확한 코퍼스 언어 분석을 위해 새로운 품사부착 프로그램이 꾸준히 개발되고 있다.

MECO 품사부착을 위해 Lancaster University에 의해 개발된 C5 품사부착세트(tagset)(<부록 2> 참조)를 사용했다(Garside, 1996). C5로 품사부착된 문장은 다음의 (3)과 같은 언어학적 정보가 나타난다.

(3) the_AT0 language_NN1 performance_NN1 with_PRP testing_VVG
Maritime_AJ0 English_NN1 as_CJS related_VVN to_PRP
(MECO_IMEC_Proceeding)

이러한 품사부착은 구체적인 언어학적 정보를 제공하므로 사용자들의 언어 연구에 도움을 준다. 예를 들어 아래 <표 3>에서와 같이 'ship'은 MECO와 BNC 코퍼스에서 품사부착 비율이 다르게 나타난다. MECO에서는 주로 단수/복수 명사로 쓰이나 과거형

7) TAGGIT은 1971년 Greene & Rubin이 언어데이터에 품사를 부착하기 위한 목적으로 개발한 프로그램이다(Kennedy, 1998).

동사, *to* 부정사, 과거분사로 쓰이는 비율이 낮다. 그러나 BNC 코퍼스에서는 단수/복수 명사로 많이 쓰일 뿐만 아니라 과거형 동사, *to* 부정사, 과거분사로 쓰이는 비율이 MECO보다 높게 나타났다. 그러므로 품사부착은 학습자와 선원들이 어휘 품사 사용의 차이점을 인지하고 실제 사용할 수 있는 유용한 정보를 줄 수 있다.

<표 3> MECO와 BNC에서 나타나는 'ship'의 빈도수

POS tagging sets	MECO		BNC	
	Raw Frequency	Relative* Frequency	Raw Frequency	Relative Frequency
NN1_singular noun	3146	53.77%	3919	53%
NN2_plural noun	2358	40.31%	2375	32.1%
NPO_proper noun	2	0.03%	7	0.1%
VVB_lexical verb	32	0.55%	25	0.34%
VVD_past tense form	1	0.02%	89	1.2%
VVI_infinitive form	36	0.62%	336	4.53%
VVG_-ing form	225	3.846%	312	4.2%
VVN_past participle	50	0.854%	336	4.53%
계	5850	100%	7399	100%

*상대적 빈도(Relative Frequency) = $\frac{\text{빈도수 (raw frequency)}}{\text{총 빈도수 (total)}} \times 100$

3.2.1.4 표준화된 참고자료

MECO는 앞에서 보여준 <표 1>을 통해 수집한 데이터의 출처를 보여주고 있다.

IV. 연구결과 및 논의

4.1 해사영어코퍼스의 주제와 영역별 분포

MECO의 모든 텍스트의 내용을 분석하고 분류한 결과 아래 <표 4>와 같이 MECO는 여덟 개의 주제로 구성되어 있다. 그 중 해사뉴스(Maritime News)는 분포 상 가장 높은 비중을 차지하고 있다. SMCP를 포함한 해사영어는 가장 낮은 비율을 보인다. 흥미로운 점은 해사법(Maritime Law)이 상대적으로 비율이 높다는 것이다. 그 이유는 해사법이 해사 공동체에서 가장 중요한 분야 중 하나로 국제적인 관계에 영향을 주는 다양한 규약들을 포함하고 있기 때문이다.

<표 4> MECO의 주제별 분포 (Hong and Jhang, 2010)

	주제	크기
1	Maritime Accidents	93,092 (9.30%)
2	Science & Technology	146,709 (14.23%)
3	Shipping	110,074 (10.68%)
4	Business & Economics	109,948 (10.67%)
5	IMO Conventions	80,079 (7.77%)
6	Maritime Law	193,646 (18.79%)
7	Standard English	65,985 (6.40%)
8	Maritime News	231,306 (22.44%)
	계	1,030,839 (100%)

여덟 개의 주제로 MECO를 분류하여 분포를 살펴본 후 다시 텍스트를 분석하여 영역별로 분류한 결과 MECO는 아래 <표 5>와 같이 학문, 공식문서, 미디어의 세 개 영역으로 나뉜다. 저널과 교과서 장르가 학문으로 분류되기 때문에 학문 영역의 비율이 가장 높다. MECO는 해사 공동체의 특별한 분야에서 사용되는 문어 텍스트만을 모았지만, MECO는 실제 언어 사용을 나타내기 위해 선원간의 대화에서 얻을 수 있는 구어데이터를 보강해야 할 필요가 있다.

<표 5> MECO의 영역 (Hong and Jhang, 2010)

영역	크기
학문(Academic)	427,475 (41.47%)
공식 문서(Official Documents)	276,258 (26.70%)
미디어(Media)	327,106 (31.73%)
계	1,030,839 (100%)

MECO와 참조 코퍼스인 LOB 코퍼스의 기본적인 통계 정보는 아래 <그림 2>에 나타나 있다.

<그림 2> MECO와 LOB의 통계 정보

MECO.lst		Lob.lst		
	N	Overall		
tokens (running words) in text	1,030,839		tokens (running words) in text	1,025,234
types (distinct words)	26,848		types (distinct words)	41,181
type/token ratio (TTR)	2.72		type/token ratio (TTR)	4.07
standardised TTR	37.40		standardised TTR	43.60
standardised TTR std.dev.	63.07		standardised TTR std.dev.	55.18
standardised TTR basis	1,000		standardised TTR basis	1,000
mean word length (in characters)	4.94		mean word length (in characters)	4.56
word length std.dev.	2.90		word length std.dev.	2.51
sentences	44,031		sentences	54,992
mean (in words)	22.38		mean (in words)	18.41
std.dev.	20.49		std.dev.	14.35

MECO의 텍스트가 1,030,839개의 단어를 가지고 있다는 것은 바로 MECO가 1,030,839개의 출현어휘수(token)를 가지고 있음을 의미한다. 그러나 많은 단어들이 반복되어 사용되고 단지 26,848개의 다른 단어가 텍스트에서 사용된다. 반복하여 사용되지 않는 단어를 단어수(type)라고 한다. 단어수와 출현어휘수의 비율이 type-token ratio(TTR)이다. TTR을 구하는 방법은 다음 <표 6>과 같다:

<표 6> TTR 도출 방법

$$\begin{aligned} \text{Type-Token Ratio} &= (\text{number of types}/\text{number of tokens}) \times 100 \\ &= (26,848/1,030,839) \times 100 = 2.72\% \end{aligned}$$

TTR을 통해 코퍼스에 사용된 어휘의 다양성을 측정할 수 있다(Rizzo, 2010). 다시 말해 TTR이 높으면 높을수록 어휘가 훨씬 더 다양하다는 것을 의미한다. 위의 <그림 2>에서 MECO가 LOB의 4.07 TTR과는 달리 다소 낮은 2.72의 TTR을 보이는 것은 MECO가 어휘 다양성이 일반영어보다 더 낮다는 것을 의미한다. 이러한 이유는 MECO는 특수한 텍스트로 구성된 코퍼스이므로 기술적, 전문적인 용어들이 구체적인 개념을 전달하기 위해 반복적으로 사용되기 때문이다.

4.2 어휘적 분석

4.2.1 품사별 분포도

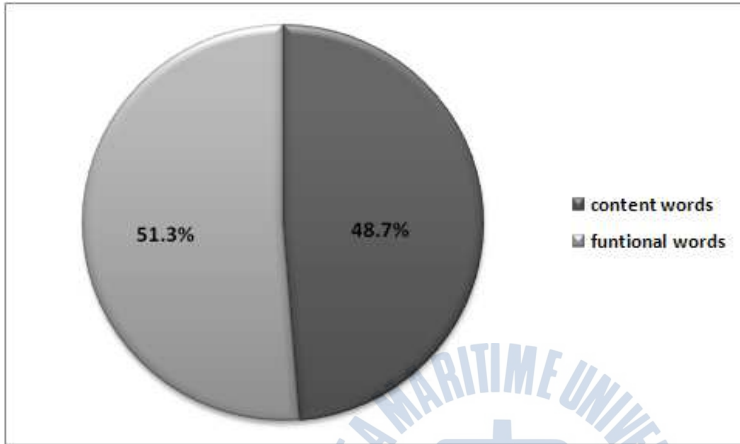
MECO의 품사별 분포도를 살펴보기 위해 C5 품사부착세트를 기준으로 각 품사를 분류하였다. 각 품사에 속하는 품사부착세트는 아래<표 7>과 같다. 여기에 언급되지 않은 품사부착세트는 <부록 2-1, 2-2>에 자세하게 나타나 있다.

<표 7> 분류된 해사영어 품사와 품사부착세트(tagset)

품사	본 논문 품사부착분류	C5 품사부착세트
명사	NN	NN0, NN1, NN2, NP0
동사	VB	VBB, VBD, VBG, VBI, VBN, VBZ, VDB, VDD, VDG, VDI, VDN, VDZ, VHB, VHD, VHG, VHI, VHN, VHZ, VVD, VVG, VVI, VVN, VVZ
부사	AV	AV0, AVP, AVQ
형용사	AJ	AJ0, AJC, AJS
전치사	PR	PRP
조동사	VM	VM0
대명사	DT (DP)	PNP, PNI, PNQ, PNX, DPS, DT0, DTQ, EX0

해사영어의 품사별 분포도의 분석 결과, 전체 단어 중 내용어는 523,766개이고 기능어는 551,849개로 나타났으며, 전체에서 내용어와 기능어가 차지하는 비율은 <그림 3>과 같다.

<그림 3> 전체 어휘에서 차지하는 내용어와 기능어의 분포도

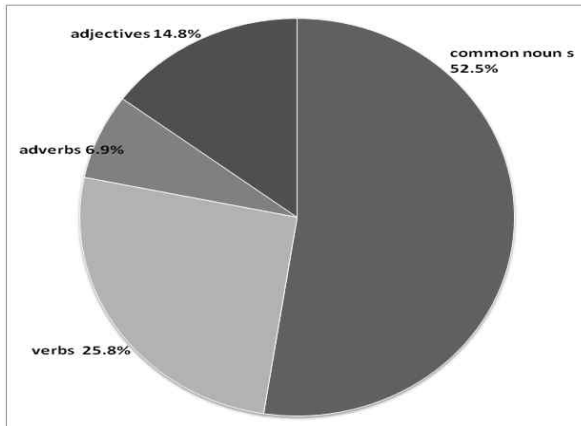


내용어의 품사는 아래 <표 8>와 같이 명사, 동사, 형용사, 부사의 순으로 나타났고 <표 8>의 백분율을 나타낸 내용어의 분포도는 아래 <그림 4>와 같다.

<표 8> 내용어 종류와 빈도수

본 논문의 품사부착분류	품사	빈도수	백분율
NN	보통명사	274,824	52.5%
VB	동사	134,875	25.8%
AJ	형용사	77,727	14.8%
AV	부사	36,340	6.9%
계		523,776	100%

〈그림 4〉 내용어의 분포도

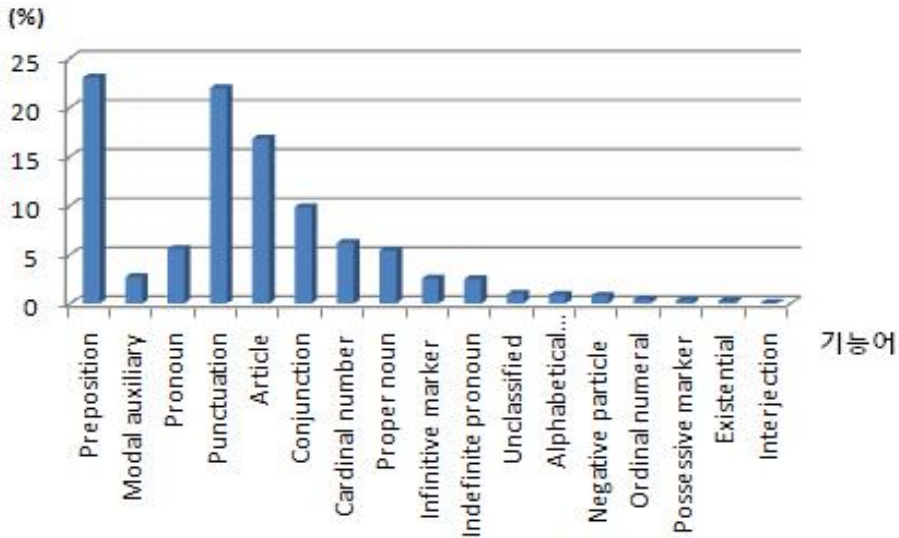


아래 <표 9>는 MECO에 나타나는 기능어의 빈도수와 백분율을 보여준다. 본 논문에서 논의할 기능어는 조동사, 대명사, 전치사이다. 조동사, 대명사, 전치사의 분포도는 아래 <표 9>와 같이 전치사, 대명사, 조동사 순으로 나타난다.

〈표 9〉 기능어의 종류와 빈도수

본 논문 품사부착분류	품사	빈도수	백분율(%)
PR	전치사	127,049	23.02
VM	조동사	14,772	2.67
DT	대명사	30,684	5.56
PU	문장부호	121,024	21.93
AT	관사	92,605	16.78
CJ	접속사	54,035	9.8
CR	기수	34,154	6.19
NP	고유명사	29,734	5.39
TO	부정사 표시	14,026	2.54
PN	부정대명사	13,602	2.46
UN	미분류	5,288	0.96
ZZ	알파벳기호	4,840	0.88
XX	부정접사	4,385	0.79
OR	서수	2,334	0.42
PO	소유격 표시	1,707	0.31
EX	존재 기능어	1,245	0.23
IT	감탄사	365	0.07
계		551,849	100

<그림 5> 기능어의 분포도



4.2.2 내용어의 분포와 어휘적 특징

4.2.2.1 명사

본 논문에서는 C5에서 NN0, NN1, NN2로 품사 부착된 일반명사를 NN으로 재분류하여 다룬다. <표 10>과 같이 해사영어 코퍼스에서 명사는 총 14,892개로 나타났고 총 빈도수는 274,824개이다. 본 연구는 아래 <표 10>에서와 같이 명사의 빈도수와 함께 누적 빈도⁸⁾를 산출하여 나타내었다.

<표 10> 명사의 빈도수와 누적빈도

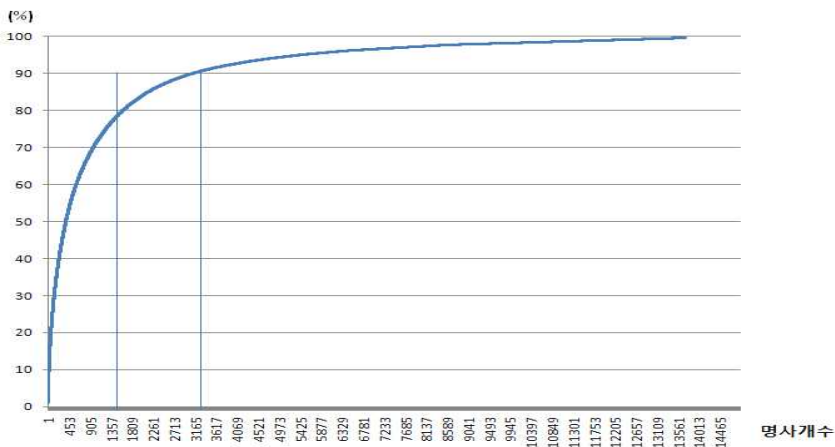
순위	명사	빈도수	누적빈도(%)
1	ship	3,131	1.13
2	ships	2,356	1.99
3	vessel	2,332	2.84

8) 누적빈도계산방법 : $\frac{\text{앞 순위단어 빈도수} + \text{구하고자 하는 단어 빈도수}}{\text{전체 명사 수}} \times 100$

4	cargo	1,645	3.44
5	vessels	1,466	3.97
6	system	1,403	4.48
7	safety	1,378	4.98
8	time	1,340	5.47
9	regulation	1,266	5.93
10	oil	1,169	6.36
11	port	1,150	6.78
12	convention	1,138	7.19
13	fire	1,063	7.58
14	sea	1,049	7.96
15	requirements	1,011	8.33
16	water	929	8.66
17	part	924	9.00
18	position	922	9.34
19	information	914	9.67
20	spaces	889	9.99
...			...
1610	implications	32	80.00
...			...
3112	peers	10	90.00
...			...
14892	ichthyology	1	100

명사의 누적빈도 80%는 1,610개, 90%는 3,112이고 80%와 90%의 10%의 차이는 어휘수가 약 2배로 증가된다. 이러한 누적빈도에 따른 어휘수의 차이는 아래 <그림 6>에서 잘 보여준다. 80%까지의 어휘목록은 <부록 5>에 둔다.

<그림 6> 명사의 누적빈도 백분율



MECO의 상위 20위까지의 일반명사를 LOB 코퍼스와 비교해 본 결과 아래 <표 11>에서와 같이 MECO 고빈도 일반명사의 빈도수가 LOB 코퍼스보다 현격하게 높았다.

<표 11> MECO 고빈도 명사의 빈도수와 LOB에서의 빈도수

순위	명사	MECO 빈도수	LOB 빈도수
1	ship*	3,131	65
2	ships*	2,356	38
3	vessel*	2,332	22
4	cargo*	1,645	7
5	vessels*	1,466	13
6	system*	1,403	301
7	safety*	1,378	49
8	time	1,340	1,769
9	regulation*	1,266	9
10	oil*	1,169	51
11	port*	1,150	23
12	convention*	1,138	26
13	fire*	1,063	140
14	sea*	1,049	190
15	requirements*	1,011	43
16	water	929	431
17	part	924	606
18	position	922	266
19	information	914	146
20	spaces*	889	10

* 표시는 MECO의 500개의 키워드 중 상위 20위 안에 든 고빈도 일반명사

위 <표 11>에서 * 표시가 없는 단어 중에서 *time*은 아래 <그림 7>의 MECO의 키워드목록에는 나타나지 않고 *water*는 키워드 목록에서 151위, *part*는 461위, *position*은 66위, *information*은 35위를 차지한다. <부록 6>에서 보는 바와 같이 MECO와 LOB 코퍼스의 고빈도 명사를 32위까지 뽑아 비교한 결과 *time*, *water*, *part*의 세 개의 명

사가 공통으로 나타났다. 공통으로 나타난 세 개의 명사는 키워드목록에 나타나지 않거나 나타나더라도 비교적 낮은 순위를 차지한다. 결국 MECO와 LOB 코퍼스의 빈도수가 현격하게 차이난다면 날수록 키워드목록에서 상위로 나타나므로 * 표가 있는 단어들은 해사영어교육에서 우선적으로 다루어야 할 일반명사이다. 또한 키워드목록에서 명사의 비중이 높은 이유는 해사영어가 특수한 해사 분야 즉 해사, 선박, 해상안전 규정과 관련되어있는 전문용어로 이루어져 있기 때문이다.

<그림 7> MECO의 키워드목록

MECO.kws			
N	Key word	Freq.	%
1	SHIP	3,459	0.34
2	VESSEL	2,549	0.25
3	SHALL	3,665	0.36
4	SHIPS	2,342	0.23
5	MARITIME	2,072	0.20
6	CARGO	1,845	0.18
7	S	2,724	0.26
8	VESSELS	1,531	0.15
9	PORT	1,540	0.15
10	SHIPPING	1,272	0.12
11	REGULATION	1,270	0.12
12	SAFETY	1,440	0.14
13	IMO	1,085	0.11
14	OIL	1,265	0.12
15	CONVENTION	1,139	0.11
16	MARINE	1,044	0.10
17	INTERNATIONAL	1,332	0.13
18	SPACES	899	0.09
19	REQUIREMENTS	1,037	0.10
20	FIRE	1,324	0.13
21	NAVIGATION	831	0.08
22	DECK	843	0.08
23	SEA	1,312	0.13
24	CREW	823	0.08
25	SYSTEM	1,493	0.14
26	SYSTEMS	863	0.08
27	CODE	653	0.06
28	II	563	0.05
29	PARAGRAPH	692	0.07
30	TRAFFIC	744	0.07

4.2.2.2 동사

MECO에서 동사는 총 6,593개로 나타났고, 총 빈도수는 134,875개이다.

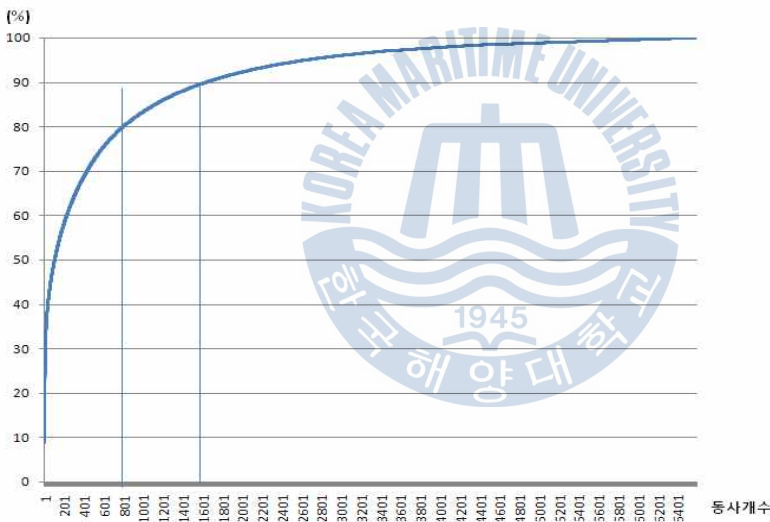
〈표 12〉 동사의 빈도수와 누적빈도

순위	동사	빈도수	누적빈도(%)
1	is	11,806	8.75
2	be	9,873	16.07
3	are	5,608	20.23
4	was	3,482	22.81
5	have	2,838	24.92
6	has	2,510	26.78
7	been	1,756	28.08
8	were	1,497	29.19
9	had	1,120	30.02
10	used	1,094	30.83
11	being	877	31.48
12	required	830	32.10
13	made	756	32.66
14	provided	729	33.20
15	do	695	33.71
16	said	665	34.21
17	using	561	34.62
18	provide	512	35.00
19	see	490	35.37
20	take	486	35.73
...
797	reports	26	80
...
1636	keel	10	90
...
6591	screen	1	99.99
...
6593	posting	1	100

아래 <그림 8>에서 동사의 누적빈도 80%는 약 800개의 단어로 나타났으며 90%는

그것의 약 2배 정도인 1,600개로 나타났다. 일반 영어의 경우 제 2언어 학습의 초기 목표의 기본 어휘 수는 약 2,000개로 전체의 약 80%를 차지한다고 주장하고 있다 (Nation, 1990). 그리고 Nation(1990)과 Coxhead(2000)는 특별한 목적을 갖는 학술 영어에서는 학습자들은 빈도가 높은 단어를 1,000개 더 합한 3,000개의 단어를 학습해야 할 뿐만 아니라 낮은 빈도의 단어들도 처리할 수 있는 전략을 습득해야 한다고 주장한다. 그러므로 본 연구에서는 품사별 빈도수를 통해 해사영어 코퍼스의 80%와 90%의 누적빈도를 산출하였고 각 품사별 80%와 90%를 차지하는 어휘를 제시하여 학습자들의 해사영어 학습에 도움이 되는 기본 어휘목록을 추출하였다. 이에 대한 목록은 <부록 7>에 둔다.

<그림 8> 동사의 누적빈도 백분율



위의 <표 12>에서 동사의 1순위에서 4순위까지 *is, be, are, was*가 나타나고 7순위와 8순위는 *were*와 *been*이, 11위는 *being*이 나타나므로 해사영어코퍼스에 사용된 *be* 동사의 시제를 일반영어 LOB 코퍼스와 비교하는 것은 의미가 있다. 해사영어 코퍼스와 일반영어 코퍼스인 LOB 코퍼스의 어휘목록을 비교분석하여 *be*동사의 빈도를 살펴본 것은 아래 <그림 9>와 같다.

<그림 9> MECO와 LOB의 어휘목록

MECO.lst						LOB.lst					
N	Word	Freq.	%	Texts	%	N	Word	Freq.	%	Texts	%
1	THE	73,877	7.17	75	100.00	1	THE	68,379	6.67	15	100.00
2	OF	38,611	3.75	75	100.00	2	OF	35,769	3.49	15	100.00
3	AND	29,447	2.86	75	100.00	3	AND	27,932	2.72	15	100.00
4	TO	27,786	2.70	75	100.00	4	TO	26,907	2.62	15	100.00
5	IN	21,728	2.11	75	100.00	5	A	23,127	2.26	15	100.00
6	A	19,935	1.93	75	100.00	6	IN	21,328	2.08	15	100.00
7	IS	12,482	1.21	75	100.00	7	THAT	11,197	1.09	15	100.00
8	FOR	10,749	1.04	75	100.00	8	IS	10,995	1.07	15	100.00
9	BE	10,262	1.00	74	98.67	9	WAS	10,502	1.02	15	100.00
10	ON	8,015	0.78	75	100.00	10	IT	10,031	0.98	15	100.00
11	THAT	7,901	0.77	75	100.00	11	FOR	9,318	0.91	15	100.00
12	AS	7,085	0.69	75	100.00	12	HE	8,797	0.86	15	100.00
13	BY	7,042	0.68	75	100.00	13	AS	7,339	0.72	15	100.00
14	WITH	6,861	0.67	75	100.00	14	BE	7,201	0.70	15	100.00
15	OR	6,506	0.63	74	98.67	15	WITH	7,201	0.70	15	100.00
16	ARE	5,959	0.58	75	100.00	16	ON	7,054	0.69	15	100.00
17	THIS	4,841	0.47	74	98.67	17	I	6,896	0.67	15	100.00
18	AT	4,784	0.46	75	100.00	18	HIS	6,272	0.61	15	100.00
19	FROM	4,673	0.45	75	100.00	19	AT	6,047	0.59	15	100.00
20	IT	4,470	0.43	75	100.00	20	BY	5,838	0.57	15	100.00
21	WHICH	4,265	0.41	74	98.67	21	HAD	5,391	0.53	15	100.00
22	NOT	4,216	0.41	74	98.67	22	THIS	5,292	0.52	15	100.00
23	AN	3,978	0.39	75	100.00	23	NOT	5,158	0.50	15	100.00
24	WAS	3,704	0.36	71	94.67	24	BUT	4,965	0.48	15	100.00
25	SHALL	3,665	0.36	39	52.00	25	FROM	4,695	0.46	15	100.00
26	SHIP	3,459	0.34	57	76.00	26	HAVE	4,598	0.45	15	100.00
27	WILL	3,166	0.31	68	90.67	27	ARE	4,546	0.44	15	100.00
28	HAVE	3,000	0.29	75	100.00	28	WHICH	4,466	0.44	15	100.00
29	S	2,724	0.26	73	97.33	29	HER	4,034	0.39	15	100.00
30	HAS	2,722	0.26	75	100.00	30	MDASH	3,927	0.38	15	100.00

위의 <그림 9>에서 보는 바와 같이 일반 영어를 대표하는 LOB 코퍼스에서도 *is*, *be*, *are* 등의 *be*동사가 상위에 나타났다.

MECO와 LOB 코퍼스에 쓰인 *be*동사의 빈도수와 백분율을 비교하여 살펴보면 아래의 <표 13>과 같다.

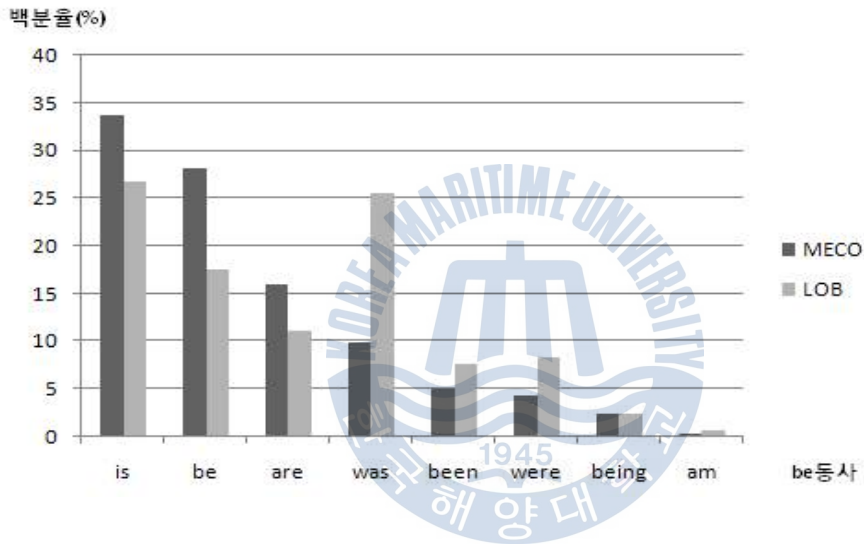
<표 13> MECO와 LOB에 쓰인 *be*동사의 빈도수와 비율

	MECO		LOB	
is	11,806	33.74%	10,995	26.8%
be	9,873	28.2%	7,201	17.5%
are	5,608	16.02%	4,546	11.1%

was	3,482	9.95%	10,502	25.6%
been	1,756	5.02%	3,116	7.6%
were	1,497	4.28%	3,401	8.3%
being	877	2.51%	968	2.4%
am	97	0.28%	304	0.7%
계	34,996	100%	41,033	100%

위의 <표 13>에 나타난 수치를 막대그래프로 나타내면 아래의 <그림 10>과 같다.

<그림 10> MECO와 LOB에 쓰인 be동사 백분율 비교



위의 <그림 10>의 MECO와 LOB 코퍼스에 쓰인 be동사 백분율을 비교하면 앞의 세 막대그래프에 나타난 *is*, *be*, *are*은 LOB 코퍼스보다 MECO에서 백분율이 높으나 다음 세 개의 막대그래프에서는 반대로 MECO보다 LOB 코퍼스에서 더 높은 백분율을 보여 준다. 여기서 해사영어에서는 *is*, *are*와 같은 현재형이 *was*, *were*와 같은 과거형 보다 더 많이 쓰인다는 것을 알 수 있다. MECO의 텍스트는 전문적, 학문적인 내용으로 구성되어 있으므로 일반영어와는 달리 지식과 정보를 전달하기 위해 현재시제가 많이 쓰인 것으로 예측가능하다. 그러므로 해사영어 텍스트에서는 일반영어에서보다 현재형이 훨씬 두드러지게 사용됨을 알 수 있었다.

4.2.2.3 형용사

해사영어코퍼스에 사용된 형용사는 총 6,140개로 나타났고, 총 빈도수는 77,727개이다. 빈도순으로 나타나는 형용사의 종류와 누적빈도는 아래 <표 14>에 나타난다.

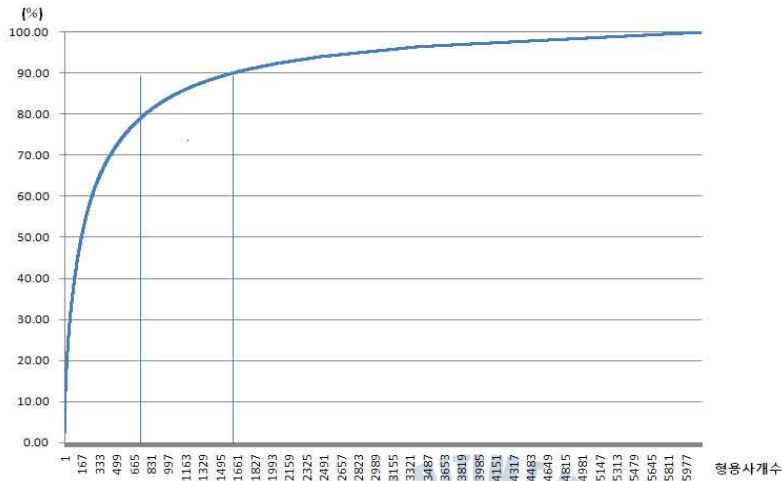
<표 14> 형용사의 빈도수와 누적빈도

순위	부사	빈도수	누적%
1	maritime	2,004	2.58
2	other	1,737	4.81
3	international	1,285	6.47
4	new	1,102	7.88
5	marine	948	9.10
6	main	629	9.91
7	English	614	10.70
8	general	601	11.48
9	different	555	12.19
10	available	503	12.84
11	following	478	13.45
12	special	453	14.04
13	possible	450	14.61
14	necessary	440	15.18
15	present	429	15.73
16	safe	426	16.28
17	high	420	16.82
18	appropriate	404	17.34
19	important	401	17.86
20	operational	387	18.35
...
775	inclined	16	80
...
1616	en-route	5	90
...
6140	nonintrusive	1	100

누적빈도수에 따른 백분율은 아래 <그림 11>과 같이 나타나고 형용사의 누적빈도 80%는 775개, 90%는 1,616개이고 80%와 90%의 10%의 차이에서 어휘수가 약 2배

로 증가된다. 80%까지의 어휘목록은 <부록 8>에 든다.

<그림 11> 형용사의 누적빈도 백분율



MECO에서 높은 빈도수를 나타내는 20개의 형용사를 조사하여 일반영어에서는 고빈도로 나타나지 않는 형용사 *maritime, marine, available, safe* 네 개와 MECO와 BNC 코퍼스⁹⁾에 공통적으로 상위에 나타나는 형용사 *high*를 선정하여 이러한 형용사가 어떤 단어와 연어 관계(collocation)를 가지는지 살펴보았다. 아래 <표 15>의 연어 관계를 이루는 어휘는 WordSmith 5.0 콘코던스 프로그램의 패턴 기능을 사용하여 각 형용사의 바로 뒤 자리(R1)에 나타난 상위 빈도 어휘 열 개를 뽑은 것이다. 밑줄 표시는 두 코퍼스에서 공통적으로 발견된 어휘이다.

<표 15> 형용사와 연어 관계를 가지는 어휘 목록

형용사	연어 관계	
	MECO	BNC
maritime	English, <u>safety</u> , <u>organization</u> , university, industry, security, <u>transport</u> , education, community, <u>law</u>	museum, <u>organization</u> , <u>safety</u> , Antarctica, <u>transport</u> , history, trade, <u>law</u> , board, nation

9) 여기서 BNC 코퍼스를 이용하여 연어를 살펴본 이유는 MECO에 나타나는 형용사가 LOB에서는 빈도수가 상대적으로 너무 낮았기 때문에 데이터양이 많은 BNC 코퍼스와 비교하였다.

marine	<u>environment</u> , traffic, insurance, pollution, engineering, communication, claims, evacuation, accidents, services	mammals, life, fish, pollution, invertebrates, <u>environments</u> , organisms, erosion, corps, terraces
available	to, for, on, in from, at, and <u>means</u> , the, depth	evidence, resources, information, data, space, <u>means</u> , time, media, sources, land
safe	speed, navigation, working, operation, distance, passing, access, <u>passage</u> , manning, escape	environment, haven, distance, keeping, side, place, sex, bet, <u>passage</u> , return
high	<u>level</u> , <u>speed</u> , seas, pressure, quality, <u>school</u> , water, <u>risk</u> , context, traffic	court, street, quality, <u>levels</u> , proportion, standard, <u>speed</u> , <u>school</u> , degree, <u>risk</u>

MECO와 BNC 코퍼스에서 다섯 개의 형용사와 연어 관계를 가지는 어휘를 살펴본 결과 *marine*, *available*, *safe* 형용사와 연어 관계를 가지는 공통된 어휘는 한 개 또는 두 개이지만 *maritime*과 *high* 형용사와 연어 관계를 가지는 공통된 어휘는 네 개로 나타났다. 그러나 *available*은 MECO에서는 *means*와 *depth*를 제외하고 기능어와 연어 관계를 이루고 있지만 BNC 코퍼스에서는 모두 명사와 연어 관계를 가진다. MECO에서 다섯 개의 형용사와 연결하는 단어는 대체로 MECO의 주제인 해사법, 해운, 해상사고, IMO 규약과 관련이 있는 것을 발견하였다. 해사영어텍스트에서 사용되는 형용사와 연어 관계를 이루는 어휘는 일반영어에서는 제한적으로 쓰이고 있다. 이렇게 MECO와 BNC 코퍼스에서 형용사와 연어 관계를 가지는 어휘가 다르기 때문에 연어를 기반으로 한 연구와 조사결과들이 해사영어 특징의 좋은 예가 될 수 있고 교육적으로 활용이 될 수 있다.

4.2.2.4 부사

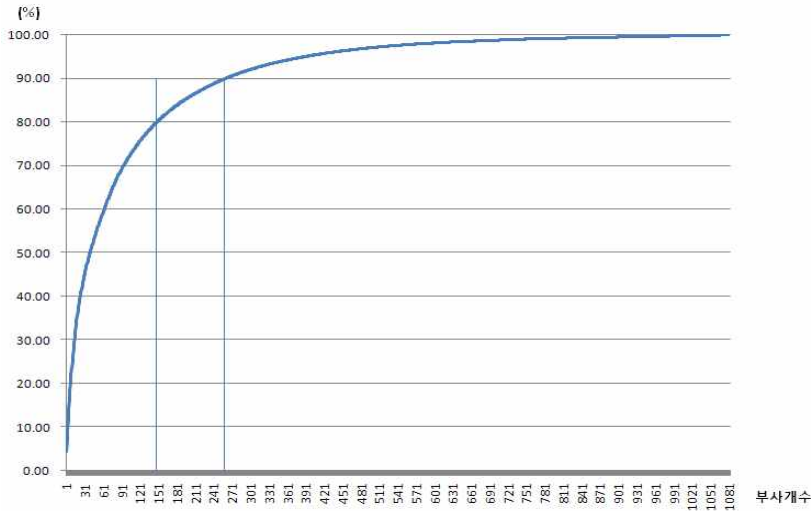
해사영어코퍼스에서 부사는 총 1,083개로 나타났고, 총 빈도수는 36,340개이다. 아래 <표 16>은 부사의 빈도수와 누적빈도를 함께 나타내고 있다. 여기서 *in*, *on*, *out*, *up* 등의 어휘는 부착된 품사를 참고하여 전치사와 중복이 되지 않는 빈도수를 나타낸 것이다.

<표 16> 부사의 빈도수와 누적빈도

순위	부사	빈도수	누적빈도(%)
1	also	1,664	4.58
2	in	929	7.14
3	more	869	9.53
4	on	862	11.90
5	only	852	14.24
6	out	764	16.35
7	so	737	18.37
8	however	720	20.35
9	up	647	22.14
10	board	583	23.74
11	then	525	25.18
12	at	522	26.62
13	where	513	28.03
14	how	500	29.41
15	as	483	30.74
16	least	482	32.06
17	very	475	33.37
18	when	447	34.60
19	now	398	35.70
20	most	394	36.78
...
149	perhaps	50	80.00
...
260	exactly	21	90.00
...
1083	unjustly	1	100

아래 <그림 12>에서 부사의 누적빈도 80%는 약 150개의 단어로 나타났으며 90%는 260개로 나타났다. 80%까지의 어휘목록은 <부록 9>에 둔다.

<그림 12> 부사의 누적빈도 백분율



Eckersley & Eckersley(1960)은 문장에서 부사의 위치를 문두, 문중, 문미의 세 가지로 나누어 제시하고 있다. 부사는 위에서 분류한 세 가지의 위치에 자유롭게 나타나는 특징을 갖는다. 홍임표(2002)에 따르면 고빈도 부사 목록에 나타난 *however*는 일반영어에서 대체로 문중에 위치한다. 효과적인 의사소통을 위해 일관성과 명료함은 필수적인 요소이고 이런 요소를 가질 수 있는 방법 중 하나는 *but*, *because*등과 같은 연결어의 사용으로 담화의 논리적, 의미적 관계를 나타내는 것이다(Altenberg & Tapper, 1998). 그러므로 본 연구에서 살펴보고자 하는 *however*와 같은 연결어는 담화상의 길잡이 역할을 할 수 있다(Leech & Svartvik, 1994). 이러한 점에서 해사영어 학습자 또는 선원들의 효과적인 의사소통을 위해 해사영어에서의 *however*의 위치와 쓰임이 일반영어처럼 문중에 위치하는지 살펴보고 그 차이점을 학습자 스스로 인지하게 하는 것은 의의가 있다.

본 연구에서는 MECO와 LOB 코퍼스에 나타나는 *however*의 위치를 아래 <표 17>과 같이 비교하였다.

<표 17> *however* 위치 비교 (빈도/백분율)¹⁰⁾

코퍼스 위치	MECO	LOB
문두	528(69.5%)	114(20.1%)
문중	226(29.7%)	446(78.8%)
문미	6(0.8%)	6(1.1%)
계	760(100%)	566(100%)

위 <표 17>에서 보는 바와 같이 LOB 코퍼스에서는 *however*가 문중에 현저하게 많이 쓰이는 것과는 달리 해사영어에서의 *however*는 문두에서 가장 많이 쓰였다. 해사영어에서는 *however*의 위치가 일반영어와는 다른 것을 알 수 있었다.

4.2.3 기능어의 어휘적 특징

4.2.3.1 전치사

전치사의 종류와 빈도를 알아보기 위해서 데이터를 하나씩 세어 볼 수 없으므로 n-gram을 사용해 2-gram과 3-gram을 분류한 후 전치사구만 따로 뽑아냈다. 아래 <표 18>, <표 19>, <표 20>은 MECO에 나타나는 각 상위 10위의 전치사를 나타내어 LOB 코퍼스에의 전치사와 그 빈도수를 비교한 것이다. 전치사의 목록은 <부록 10>, <부록 11>, <부록 12>에 둔다.

<표 18> MECO 전치사 목록 (1-gram)

순위	전치사	MECO 빈도수	LOB 빈도수
1	of	35,258	35,769
2	in	18,153	21,328
3	to	10,077	26,907
4	for	9,726	9,318
5	by	6,424	5,838

10) 각 코퍼스에 나타난 *however*의 개수가 달라 편의상 백분율로 환산하여 비교하였다.

6	on	6,288	7,064
7	with	5,790	7,201
8	from	4,266	4,695
9	at	3,988	6,047
10	as	1,619	7,339

<표 19> MECO 전치사 목록 (2-gram)

순위	전치사구	MECO빈도수	LOB 빈도수
1	such as	676	193
2	due to	414	151
3	as to	331	162
4	up to	242	257
5	according to	200	131
6	out of	178	586
7	because of	116	143
8	prior to	108	16
9	Apart from	37	100
10	ahead of	32	23

<표 20> MECO 전치사 목록 (3-gram)

순위	전치사구	MECO 빈도수	LOB 빈도수
1	in accordance with	324	26
2	in addition to	151	52
3	in respect of	96	23
4	in terms of	91	65
5	in charge of	87	16
6	in relation to	81	41
7	by means of	80	47
8	in case of	63	5
9	with respect to	57	19
10	as part of	55	22

비교한 빈도수를 살펴보면 위의 <표 18>에 나타나는 전치사의 빈도수는 *for*와 *by*를 제외하고는 LOB 코퍼스에서 더 높게 나타난다. 그러나 위의 <표 19>와 <표 20>에 나타나는 전치사의 빈도수는 MECO에서 더 높게 나타나는 것을 알 수 있었다. 일반영

어를 대표하는 LOB 코퍼스는 시, 소설 등 다양한 텍스트로 구성되어 있어 사용되는 어휘와 표현이 다양하지만 MECO를 구성하고 있는 텍스트가 다양하지 않고 특수하기 때문에 어휘나 어구(phrase)를 반복해서 많이 쓰고 있다.

한편 모든 영어 전치사구¹¹⁾을 확인한 결과 LOB 코퍼스에서는 나타나지만 해사영어 코퍼스에 나타나지 않는 전치사군은 아래 (4)와 (5)에 나타난 것과 같다.

(4) 2-gram: *devoid of, upwards of, void of, preliminary to, preparatory to, previous to*

(5) 3-gram: *in back of, on grounds of*

의사소통에서 연결어구(숙어와 연어)의 중요성은 꾸준히 강조되고 있다. Cowie(1998)은 문어와 구어에서 중요한 역할을 하는 연결어구는 준원어민성의 좋은 척도로 보았다. 최근에 Bozdchová(2008)가 다양한 연결어구는 모든 현대의 언어에서 가장 유용하고 가장 많이 쓰이는 표현이라고 말하고, Moon(2008)은 단어를 하나하나 연구하는 것은 더 이상 흥미롭지 않다고 주장하며 연결어구의 중요성을 강조한다. 이런 전치사구와 같은 연결어구 표현은 비원어민에게는 어려운 부분이므로 학습자는 원어민이 덩어리로 된 표현들을 자연스럽게 만들어 내는 것처럼 복잡한 연결어구를 익혀 자유자재로 쓸 수 있어야 한다. 이런 점에서 본 연구에서 나타내는 고빈도 전치사구 목록을 통해 학습자가 우선적으로 사용해야 할 표현의 순위를 제공할 수 있고 가장 많이 사용되는 어구를 익힘으로써 영어 원어민과 원활한 의사소통할 수 있는 능력을 향상시키는 교육적 활용도가 높을 것으로 생각된다.

4.2.3.2 대명사

앞선 4.2.2.2에서 살펴본 <그림 9>의 MECO와 LOB 코퍼스의 어휘목록을 비교하면 고빈도 단어 중에 LOB 코퍼스에는 나타나는 *he, his, I, her* 등의 인칭대명사를 MECO에서는 찾을 수 없었다. 그래서 MECO에서 일반영어와 두드러진 차이점을 나타내는

11) 모든 전치사구는 *Longman Grammar of Spoken and Written English*(Biber, Johansson, Leech, Conard, & Finegan, 1999) 74-75쪽의 전치사구 목록을 참고하였다.

대명사의 사용을 살펴보았다.

MECO와 LOB 코퍼스의 인칭대명사의 빈도는 아래의 <표 21>에 나타난다.

<표 21> MECO와 LOB의 인칭대명사 비율

인칭 대명사	MECO	LOB
I	0.12%	0.67%
we	0.13%	0.28%
you	0.09%	0.35%
he	0.12%	0.85%
she	0.03%	0.38%
they	0.15%	0.34%
it	0.43%	0.97%

위의 <표 21>에서 보는 바와 같이 MECO에서는 *I, he, she* 등의 대명사 빈도가 낮았다. 이는 MECO가 학술지, 전공 서적 등의 텍스트로 이루어져 있으므로 인칭대명사 빈도가 일반영어 보다 낮고, *it, they*와 같은 대명사의 빈도가 높다. 이는 MECO의 전문적인 텍스트의 특성으로 미루어 보아 전문 학술지가 가지는 대명사의 쓰임의 특징 때문으로 유추된다.

특히 MECO에서 나타난 *she*의 쓰임을 콘코던스를 통해 살펴보면 약 70% 이상이 *vessel, ship* 등을 지시하는 대명사로 쓰이는 것이 발견되었다. 아래 (6)과 (7)은 MECO의 각 장르에 면색으로 밑줄 친 *she*가 사물을 지칭하는 대명사로 나타난 예문이다.

- (6) Every vessel shall reduce her speed to a minimum at which **she** can be kept on course when a fog signal is heard apparently foreword of the beam or when a dose-quarters situation cannot be avoided with another vessel foreword of her beam. (Textbooks_International Regulations for Preventing Collisions at Sea)
- (7) If the ship does not pass the survey **she** may have her class own-graded or suspended until the she has carried out the necessary corrective measures; failure in the latter respect will cause her name to be expunged from the register (Textbooks_Marine Insurance).

4.2.3.3 조동사

앞선 4.2.2.1에서 살펴본 <그림 7>에 나타난 해사영어코퍼스의 키워드목록에서 상당히 주목할 만한 어휘는 바로 *shall*이다. Thompson(2002)에 따르면 조동사의 사용은 학문적인 텍스트에서 사용되는 수사적인 기능과 관련되어 있다. 아래 <표 22>는 MECO와 LOB 코퍼스를 비교했을 때 조동사의 쓰임의 다른 패턴을 보여준다.

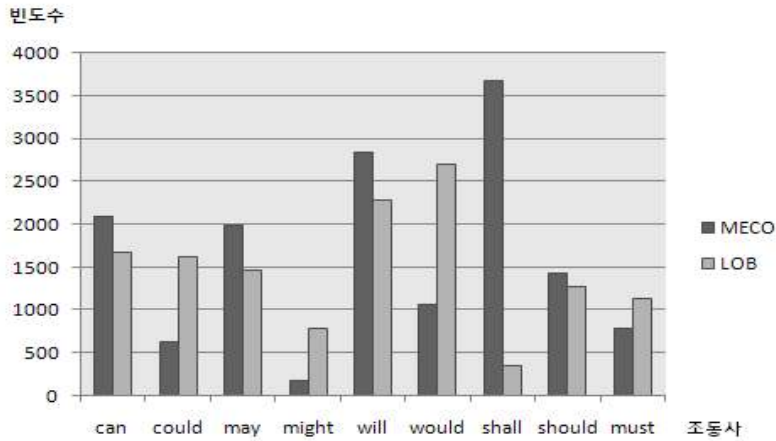
<표 22> 조동사의 빈도수

Modal verbs	MECO	LOB
can	2,090	1,678
could	631	1,615
may	1,985	1,454
might	169	781
will	2,830	2,275
would	1,058	2,688
shall	3,665	350
should	1,434	1,276
must	789	1,128

위의 <표 22>에서 보는 바와 같이 MECO에서 *shall*은 가장 많이 사용되었지만 *might*는 현저하게 적게 사용되었다. 특히 면색으로 표시되어 있는 조동사 *shall*은 LOB 코퍼스에서는 가장 낮은 빈도수를 보여주지만 MECO에서는 가장 높은 빈도수를 보여주고 있다.

이러한 두 코퍼스간의 조동사의 쓰임의 차이점은 아래 <그림 13>에서와 같이 막대 그래프 분포로 일목요연하게 나타낼 수 있다.

<그림 13> MECO / LOB 조동사 분포



MECO에서 고빈도로 사용되는 조동사 *shall*의 분포는 MECO의 장르 중 해사법에 집중되어 있고 아래 (8)과 같은 예문에서 *shall*은 의지를 표현하고 있다.

- (8) The training manual, which may comprise several volumes, **shall** contain the instructions and information required in paragraph 2.3.4 in easily understood terms and illustrated where possible (IMO, 2004: 260).

그러나 현저하게 적게 사용된 *might*의 패턴은 *shall*과는 다르게 장르별로 고르게 분포 되어 있지만 *shall*의 빈도 패턴은 해사법 장르에 집중되어 있는 것을 아래 <표 23>에서 나타내고 있다.

<표 23> MECO에 나타나는 *might*와 *shall*의 빈도수 (Hong & Jhang, 2010)

types	<i>might</i>	<i>shall</i>
Journals	36	36
Maritime Law	17	3,259
Media	39	28
Others	39	40
Set Language	5	5
Textbooks	33	297
계	169	3,665

위와 같이 *shall*이 특히 해사법 장르에서 고빈도로 쓰이고 있으므로 일반법 텍스트에서의 *shall*의 쓰임과 비교할 가치가 있다. 일반법에서 나타나는 *shall*의 빈도수를 비교하기 위해 BNC 코퍼스의 문어 텍스트 중 *politic_law_edu* 장르를 이용하였다. BNC 코퍼스의 *politic_law_edu* 장르는 4,703,303 단어로 이루어져 있으며 그 중 *shall* 빈도수는 1,849개로 나타났다. 100만 단어 당 *shall*의 빈도수를 계산한 수치는 393.13개이다. 분석 결과 일반법 텍스트에서는 조동사의 *shall* 빈도가 해사법처럼 높지 않은 것으로 나타났다. 이처럼 *shall*은 해사 텍스트에서 특히 해사 규약 또는 법에서 사용이 선호되므로 학습자와 선원들은 해사영어와 일반영어사이의 분명한 차이를 나타내는 조동사의 특징을 학습할 필요가 있다.

MECO에서 현저하게 적게 사용된 *might* 경우에는 빈도 패턴이 아래 (9)와 같은 예문처럼 모호한 의사소통을 피하기 위해 SMCP에서 사용하지 않도록 IMO에서 권장하는 것과 관련이 있다(Pritchard, 2003).

- (9) a. Do not say: *I might enter the fairway.*
 b. Say: INTENTION. *I will enter the fairway.* (Hong & Jhang, 2010)

그러나 조동사 중에서 *may* 또한 *might*처럼 모호한 의사소통을 야기할 수 있으나 MECO에서 LOB 코퍼스보다 빈도수가 높게 나타났다. 그 이유는 아래 <표 24>처럼 MECO의 다양한 장르에서 *may*의 빈도수가 높았기 때문이다. 그러나 면색으로 표시된 Set Language 장르에서는 빈도수가 매우 낮게 나타나 규정화된 해사영어 표현에서는 *may*와 *might*와 같은 조동사의 쓰임이 제한되는 것을 알 수 있었다.

<표 24> MECO에 나타난 *might* 와 *may*의 빈도수

types	<i>might</i>	<i>may</i>
Journals	36	220
Maritime Law	17	619
Media	39	318
Others	39	291
Set Language	2	21
Textbooks	33	516
계	169	1,985

요약하면 해사영어 학습자와 선원들은 해사 텍스트에서는 조동사의 정확한 사용법을 알아야 할 필요가 있다. 특히 SMCP는 선박과 선박 그리고 선박과 항구의 실제 의사소통에서 공식적인 표현이 사용될 때 모호한 의사소통을 야기할 수 있는 *may*, *might*, *could* 등의 조동사를 사용하지 않도록 권하고 있기 때문이다.



V. 결론 및 제언

본 논문에서는 문어중심의 해사영어 코퍼스인 MECO를 구축하였다. MECO가 대표성을 가지기 위해 여섯 명의 해사영어 담당 교수의 전문가 조언을 바탕으로 텍스트를 균형 있게 수집하였으며 특수목적영어의 기초 연구에 필요한 코퍼스 크기인 100만 단어 규모로 구축하였다. 또한 컴퓨터가 해독할 수 있는 형태인 text파일로 변환하였으며 각 데이터의 출처를 나타내어 참고자료로 제시하였다.

구축된 MECO의 어휘적인 특징을 살펴보기 위해 코퍼스 분석 프로그램을 활용하여 일반영어 코퍼스인 LOB 코퍼스와 BNC 코퍼스와의 비교 분석을 하였다. 먼저 코퍼스 분석을 통해 해사영어 코퍼스가 얼마나 다양하게 어휘를 포함하고 있는지를 단어수와 출현수의 도출결과를 TTR로 살펴보았다. 다음으로 어휘목록, 키워드목록, 콘코던스를 통해 각 품사별 분포도와 내용어 및 기능어의 어휘적인 특징, 특징적인 예문, 연어 관계를 비교 분석하여 해사영어의 어휘적인 특징을 살펴보았다. 이와 같은 연구에서 얻은 결과는 다음과 같다.

MECO와 LOB 코퍼스의 통계정보를 비교한 결과 MECO의 TTR이 LOB 코퍼스보다 훨씬 낮게 나타나 MECO가 특수한 전문 분야의 텍스트에 기초하고 있어 어휘의 다양성이 일반영어보다 낮다는 것을 알 수 있었다.

품사별 분포도를 살펴본 결과 전체의 어휘 중 내용어가 48.7% 기능어가 51.3%로 비슷한 분포를 보여주었다. 높은 어휘 빈도순으로는 내용어에서는 명사, 동사, 형용사, 부사의 순이고 본 연구에서 다룬 기능어에서는 전치사, 대명사, 조동사의 순으로 나타났다.

내용어의 어휘적인 특징으로 명사에서는 고빈도를 보여주는 MECO 일반명사의 빈도수가 LOB 코퍼스보다 더 현격하게 높아 그 차이가 많이 날수록 키워드목록의 상위에 나타난 점이 특징적이었다. 이는 MECO와 LOB 코퍼스의 고빈도 명사를 32위까지 뽑아 비교한 결과 *time, water, part*의 세 개의 명사는 키워드목록에 나타나지 않거나 나타나더라도 비교적 낮은 순위를 차지하는 것에서 알 수 있었다.

동사에서는 MECO의 동사의 시제가 일반영어와 흥미로운 차이를 보여주었다. MECO의 동사 시제에서의 특이한 점을 동사목록에서 고빈도를 보여주는 *be*동사를 통

해 살펴보았다. MECO의 고빈도 동사목록에서 *be*동사는 *is, be, are, was, been, were, being* 순으로 나타났다. LOB 코퍼스와 비교한 결과 MECO의 *is, be, are, being*은 LOB 코퍼스보다 높은 백분율을 보여주는 반면에 *was, were, been*의 백분율은 LOB 코퍼스보다 낮았다. 그러므로 해사영어 텍스트에서는 일반영어에서보다 현재형이 훨씬 두드러지게 사용되는 것을 알 수 있었다.

형용사에서는 MECO에서만 높은 빈도수를 보여주는 형용사 *maritime, marine, available, safe*와 BNC 코퍼스에서도 고빈도로 나타나는 형용사 *high*를 선정하여 어떤 어휘와 연어 관계를 가지는지를 살펴본 결과 MECO의 고빈도 형용사는 일반영어와는 다른 어휘들과 연어 관계를 가졌다. 또한 형용사와 연어 관계를 이루는 명사들은 대체로 MECO의 주제인 해사법, 해운, 해상사고, IMO 규약과 관련이 있는 어휘와 연어 관계를 가지는 것을 발견하였다.

또한 부사의 특징에서는 부사의 위치를 살펴보았다. 한 예로 부사 연결어인 *however*의 문장에서의 위치를 살펴본 결과 MECO에서는 일반영어와는 다른 쓰임을 발견할 수 있었다. MECO와 LOB 코퍼스에서 부사 *however*이 문두, 문중, 문미에 오는 빈도수를 분석한 결과 LOB 코퍼스에서는 *however*이 문중에서 현저하게 많이 쓰이지만 MECO에서는 *however*이 문두에서 가장 많이 쓰인 것을 알 수 있었다. 이렇게 해사영어 학습자들은 MECO에 나타난 연결 부사 *however*의 위치와 쓰임이 일반영어와 차이가 있다는 것을 데이터 분석을 통해 인식(awareness)할 수 있다. 특히 고광윤과 이성운(2008)이 지적하고 있듯이 *however*를 문두의 위치에서 과잉 사용하는 한국 학생들에게 인식시킬 필요가 있다.

기능어의 어휘적인 특징은 전치사, 대명사, 조동사를 분석하여 살펴보았다. MECO에 나타난 전치사를 LOB 코퍼스와 비교한 결과 한 단어로 되어있는 전치사의 빈도수는 큰 차이를 보이지 않았으나, 2-gram, 3-gram의 전치사군에 나타난 표현들은 MECO에서 훨씬 높은 빈도수를 나타나기 때문에 해사영어 학습자는 전치사를 학습해야 할 필요가 있다.

그리고 MECO에 나타난 대명사를 LOB 코퍼스와 비교한 결과 인칭대명사의 빈도수가 일반영어보다 낮았고, *I, he, she*의 대명사의 빈도는 현저하게 낮았다.

마지막으로 MECO와 LOB 코퍼스의 조동사 빈도수를 분석한 결과 MECO에서 *shall*이 현저하게 많이 사용되었음을 알 수 있었다. 이는 해사법 또는 해사 규약에서 *shall*이 선호

되는 조동사이기 때문에 빈도가 높게 나타난 것으로 보인다. *shall*이 고빈도로 사용되는 것이 해사법 텍스트의 특징인지 살펴보기 위해 BNC 코퍼스의 politic_law_edu 장르와 비교한 결과 해사법에서만 *shall*의 사용이 현저하게 높게 나타났다는 것을 알 수 있었다.

지금까지 코퍼스기반으로 연구를 통해 살펴 본 해사영어의 어휘적 특징은 일반영어와 차이가 있다는 것을 실증적으로 보여주고 있다. 본 연구의 어휘목록과 키워드목록을 통해 도출된 어휘는 해사영어의 특수한 용어를 포함하고 있기 때문에 일반영어와는 차이를 보이는 해사영어를 이해하기 위한 기본적인 자료가 될 수 있다. 또한 해사영어 어휘에 대한 전반적인 개요를 보여줄 수 있다는 점에서 의미가 있고 추후 해사영어교육을 위한 교재의 토대가 될 수 있을 뿐만 아니라 교사와 학습자가 유용하게 활용할 수 있을 것이다. 이와 같은 연구결과가 실제 해사 커뮤니티에서 해사영어 학습자와 선원들의 실제적인 의사소통 향상에 도움을 줄 것으로 기대하는 마음에서 다음의 두 가지 구체적인 제안을 하고자 한다.

첫째, 내용어에서 보여준 어휘별 누적빈도는 학습자의 수준을 고려한 어휘 목록을 제공할 수 있으므로 어떤 어휘를 먼저 가르쳐야 하는가에 대한 현실적인 도움을 줄 수 있다. 초급수준의 해사영어수업에서는 누적빈도 백분율 80%까지의 어휘를, 중급수준의 해사영어수업에서는 누적빈도 백분율 90%까지의 어휘를 우선적으로 학습할 수 있도록 하는 것이 바람직하겠다. 둘째, 코퍼스를 기반으로 하는 자료중심의 언어 교육은 학습자가 직접 프로그램을 통해 언어자료를 분석하여 결과를 찾을 수 있는 유용함이 있다. 다시 말해 자료중심의 학습법은 학습자가 학습의 중심이 되어 스스로 묻고 답할 수 있는 자기 주도적 학습을 가능하게 한다는 장점이 있다(Benson, 2001; Hong, 2010a; 2010b; Schmidt, 1990). 그러므로 Hong(2009)에서도 언급하고 있듯이 일반영어 교육에서 코퍼스 분석 또는 학습자 중심의 방법론에 기반을 둔 다양한 프로그램들이 현재 개발되어 있듯이, 특수목적영어인 해사영어 교육에서도 체계적인 교육을 위해서는 코퍼스의 자료중심 학습법을 위한 웹 기반 언어 학습 프로그램이 차후 개발되기를 기대한다.

참 고 문 헌

<간접인용>

- Bozdchová, I. (2008). Review of František Čermák, *Frazeologie a idiomatika Česká a obecná Czech and General Phraseology*. *International Journal of Lexicography*, 21, 198-202.
- Cowie, A. P. (1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Eckersley, C. E., & Eckersley, J. M. (1960). *A Comprehensive English Grammar for Foreign Students*. London: Longman.
- Garside, R. (1996). The robust tagging of unrestricted text: the BNC experience. In Thomas, J. and Short, M. (eds.), *Using Corpora for Language Research: Studies in Honor of Geoffrey Leech*, 167-180. London: Longman.
- Greene, B. B. & Rubin, G. M. (1971). *Automated Grammatical Tagging of English*. Providence, RI: Department of Computer Science, Brown University.
- Hyland, K. (1998). *Hedging Scientific Research Articles*. Amsterdam: John Benjamins Publishing Company.
- Leech, G & Svartvik, J. (1994). *A Communicative Grammar of English*. 2nd edition. London: Longman.
- Moon, R. (2008). Sinclair, phraseology, and lexicography. *International Journal of Lexicography*, 21, 243-254.
- Nation, I. S. P. (1990) *Teaching and Learning Vocabulary*. New York: Newbury House.
- Pearson, J. (1998). *Terms in Context*. Amsterdam: John Benjamins Publishing Company.

<직접인용>

- 고광윤. (2005). 한국적 영어학과 영어코퍼스의 활용. *영어학 연구*, 19, 21-35.
- 고광윤. (2009). 코퍼스 분석을 위한 프로그램 선택: WordSmith Tools 4.0 MonoConc Pro 2.2를 중심으로. *영어학 연구*, 27, 1-22.

- 고광윤, 이성운. (2008). 한국 영자신문 영어의 코퍼스 바탕 연구. 한국 현대언어학회(언어연구), 23, 601-619.
- 김소라. (2009). 코퍼스 기반 아동용 영자신문 어휘 분석과 연어 표현집 개발. 석사학위논문, 청주대학교 교육대학원.
- 김은정. (2007). 고등학교 교과서 양상조동사의 코퍼스 언어학적 분석. 석사학위논문, 연세대학교 교육대학원.
- 정미애. (2007). 코퍼스 분석을 통해 살펴본 비즈니스 영어의 어휘적 특성. 한국언어학회, 32(4), 751-777.
- 허미영. (2007). 영어 독해교재의 코퍼스 바탕 연구 - 어휘 분석을 중심으로 -. 석사학위논문, 연세대학교 교육대학원.
- 홍임표. (2002). 영어 부사의 연구. 석사학위논문, 경기대학교 교육대학원.
- Altenberg, B. & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In Granger, S. (ed.) *Learner English on Computer*, 80-93. London and New York: Longman.
- Benson, P. (2001). *Teaching and Researching Autonomy in Language Learning*. London: Pearson Education.
- Biber, D. (1998). *Corpus Linguistics Investigation Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*, England: Pearson Education.
- Chirea-Ungureanu, C., & Goergescu, M. (2009). Managing cultural diversity. *Proceedings of International Maritime English Conference*, 21, 49-56.
- Coxhead, A. (2000). A new academic wordlist, *TESOL Quarterly*, 34(2), 213-238.
- Durrant, P. (2009). Investigating the validity of a collocation list for student English for academic purposes. *English for Specific Purposes*, 28, 157-169.
- Flowerdew, L. (1993). An educational, or process, approach to the teaching of professional genres. *ELT Journal*, 47(4), 305-316.
- Flowerdew, L. (2003). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against

- corpus-based methodologies. *English for Specific Purposes*, 24, 321-332.
- Garside, R., Leech, G., & McEnery, A. M. (1997). *Corpus Annotation: Linguistics Information from Computer Text Corpora*, London: Longman.
- Gavioli, L. (2005). *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins Publishing Company.
- Hong, S. C. (2009). *The Pedagogical Use of Corpus Data Based on Two Case Studies: the Dong-A for Korean Learners and the Chemnitz for German Learners*, Doctorial dissertation, Lancaster: Lancaster University.
- Hong, S. C. (2010a). EFL learners' consciousness-raising through a corpus-based approach. *English Teaching*, 65(1), 57-86.
- Hong, S. C. (2010b). How can corpus data be applied for EFL learners and teachers. *English Language Teaching*, 22(2), 55-78.
- Hong, S. C., & Jhang, S. E. (2010). The compilation of a maritime English corpus for ESP learners. *Korean Journal of English Language and Linguistics*, 10(4), 963-985.
- Horck, J. (2005, October). Communication skills are vital to safe operation. *Alert*, Issue 9.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hutchson, T., & Waters, A. (1987). *English for Specific Purposes: A Learning Centered Approach*. Cambridge: Cambridge University Press.
- IMO. (1995). *International Convention of Standards of Training, Watchkeeping and Certification 1978 as Amended 1995*. London: IMO.
- IMO. (2004). *SOLAS*. London: IMO.
- IMO. (2009). *Module Course 3.17: Maritime English. Reading*. London: IMO.
- Ioana, R. V. (2009). Translation: a useful tool in maritime English teaching. *Proceedings of International Maritime English Conference*, 21, 111-117.
- Ishikawa, S. (2009). Phraseology overused and underused by Japanese learners of English: A contrastive interlanguage analysis. *Corpus Linguistics and Lexicography*,

87–100.

- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.
- Laurence, A. (2007). *AntconC 3.2.1 (Window)*. Tokyo: Waseda University.
- Leech, G. (2004). Adding linguistic annotation. In Wynne, M. (ed.), *Developing Linguistic Corpora: A Guide to Good practice*, 17–29. Oxford: Oxford University Press.
- Marco, M. J. L. (2000). Collocational frameworks in medical research papers: genre-based study. *English for Specific Purposes*, 19(1), 63–86.
- McEnery, T., & Wilson, A. (2001). *Corpus Linguistics*. 2nd edition. Edinburgh: Edinburgh University Press.
- Nelson, M. (2006). Semantic associations in business English: a corpus-based analysis. *English for Specific Purposes*, 25, 217–234.
- Olga, M. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25, 235–256.
- Pritchard, B. (2003). *Maritime Communications and IMP SMCP 2001*. Rijeka: Rijeka University Press.
- Rizzo, C. R. (2010). Getting on with corpus compilation: from theory to practice. *ESP World*, 1(9), 1–23.
- Schmidt, R. (1990). Consciousness in second language learning. *Applied Linguistics*, 11, 129–158.
- Schriever, U. (2009). Acceptance of opposition to and competency levels of in Maritime English as seen by seafarers. *Proceedings of International Maritime English Conference*, 21, 151–164.
- Scott, M. (2001). Comparing corpora and identifying keywords, collocations, frequency distributions through the WordSmith Tools suite of computer programs. In Henry, M. A. and Rosebury, R. (eds.) *Small Corpus in ELT*. 47–67. Amsterdam: John Benjamins Publishing Company.
- Scott, M. (2009). *Oxford WordSmith Tools, Version 5.0*. Oxford: Oxford University Press.

- Sinclair, J. (1991). *Corpus, Concordance and Collocation*, Oxford: Oxford University Press.
- Sinclair, J. (2004). Corpus and text: basic principle. In Wynne, M. (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. 1-16. Oxford: Oxford University Press.
- Taner, A. R. Ziarati. (2009). Evaluation, assessment, and testing in Maritime English: Measuring students' competence and performance. *Proceedings of International Maritime English Conference*, 21, 39-48.
- Thompson, P. (2002). Modal verbs in academic writing. In Kettelman, B., & Marko, G. (eds.), *Teaching and Learning by Doing Corpus Analysis*. 305-325, Amsterdam: Rodopi.
- Trenkner, P. (2000). Maritime English: an attempt at an imperfect definition. presented at *The Second IMLA Workshop on Maritime English*. Dalian: Dalian Maritime University.
- Trenkner, P. (2009). Maritime English requirements and the revised STCW. *Proceedings of International Maritime English Conference*, 21, 5-10.
- Winnie, C., Chris, G., John, S., & Martin, W. (2009). Uncovering the extent of the phraseological tendency: Towards a systemic analysis of concgrams. *Applied Linguistics*, 30: 236-252.

<온라인 인용>

<http://corpora.lancs.ac.uk/BNCweb/home.html>

<http://en.wikipedia.org/wiki/N-gram>

<http://sites.google.com/site/scottpiaosite/software/mlct>

http://www.lexically.net/downloads/version5/HTML/index.html?type_token_ratio_proc.htm

부 록

<부록1> 해사영어 코퍼스 구축을 위한 인터뷰 질문내용	58
<부록2-1> CLAW5 품사부착세트(tagset)	59
<부록2-2> MECO에 사용된 대표 품사부착세트(tagset)와 C5 품사부착 세트(tagset) 종류	60
<부록3> MECO 상위 100위까지의 어휘목록 (Word-list)	61
<부록4> MECO 상위 100위까지의 키워드목록 (Keyword-list)	62
<부록5> 누적빈도 백분율 80%까지의 명사 목록	63
<부록6> MECO와 LOB의 고빈도 상위 32위까지 명사 목록	76
<부록7> 누적빈도 백분율 80%까지의 동사 목록	77
<부록8> 누적빈도 백분율 80%까지의 형용사 목록	83
<부록9> 누적빈도 백분율 80%까지의 부사 목록	90
<부록10> 전치사 목록(1-gram)	91
<부록11> 전치사 목록(2-gram)	92
<부록12> 전치사 목록(3-gram)	93

<부록 1>

해사영어코퍼스 구축을 위한 인터뷰 질문내용

1. Name: _____ / Affiliation: _____

2. Specialized areas: _____

3. How would you define Maritime English(ME)?

4. Do you think it is necessary to teach or learn ME based on real communications and authentic materials?

5. What is your main focus in ME in your teaching ME at Korea Maritime University? (For example, structure of ME/General English, terminologies, SMCP, and so on)

6. With regard to the above question 4, if your answer is "yes", can you recommend any written materials (books, journals, newspaper, or websites) which represent your specialised areas?

7. Do you have any intention to use the MECO (Maritime English Corpus) compiled from authentic materials widely used in the maritime community for teaching ME to your students?

8. Do you have any other comments on the project of the MECO?

This interview will be used for the MECO project which is conducted in the Department of English Language and Literature at Korea Maritime University in 2010. For this reason, it will not be published or used without the interviewee's permission.

<Interview>

Date: _____ Time: _____ Place: _____

<부록 2-1>

CLAW5 품사부착세트(tagset)

Tag	Description
AJO	Adjective (general or positive) (e.g. good, old, beautiful)
AJC	Comparative adjective (e.g. better, older)
AJS	Superlative adjective (e.g. best, oldest)
ATO	Article (e.g. the, a, an, no)
...	
NN0	Common noun, neutral for number (e.g. aircraft, data, committee)
NN1	Singular common noun (e.g. pencil, goose, time)
NN2	Plural common noun (e.g. pencils, geese, times)
NPO	Proper noun (e.g. London, Michael, Mars, IBM)
...	
PNP	Personal pronoun (e.g. I, you, them, ours)
PNQ	Wh-pronoun (e.g. who, whoever, whom)
PNX	Reflexive pronoun (e.g. myself, yourself, itself, ourselves)
...	
VBB	The present tense forms of the verb BE, except for is, 's: e.g. am, are, 'm, 're and be (subjunctive or imperative)
VBD	The past tense forms of the verb BE: was and were
VBG	The -ing form of the verb BE: being
VBI	The infinitive form of the verb BE: be
...	
VDB	The finite base form of the verb DO: do
VDD	The past tense form of the verb DO: did
VHB	The finite base form of the verb HAVE: have, 've
VHD	The past tense form of the verb HAVE: had
...	
VM0	Modal auxiliary verb (e.g. will, would, can, could, 'll, 'd)
...	
CJC	Coordinating conjunction (e.g. and, or but)
CJS	Subordinating conjunction (e.g. although, when)
...	
DPS	Possessive determiner-pronoun (e.g. your, their, his)
...	

<부록 2-2>

MECO에 사용된 대표 품사부착세트(tagset)와 C5 품사부착세트(tagset) 종류

본 논문용 대표 품사부착세트	C5 품사부착세트 종류
PU	! / (/) / - / . / ... / : / ?
AJ	AJ0/ AJ021/ AJ022 / AJ031 / AJ032/ AJ033/ AJC / AJS
AV	AV0 / AV021/ AV022 / AV031 / AV032 / AV033 / AV041/ AV042/ AV043/ AV044 / AVP / AVQ / AVQ31 / AVQ32 / AVQ33
AT	AT0
CJ	CJC / CJS/ CJS21 / CJS22 / CJS31 / CJS32 / CJS33 / CJS41 / CJS42 /CJS43 / CJS44 / CJT / CRD / CRD21 / CRD22
DT	DT0 / DTQ / DTQ31 / DTQ32 / DTQ33
EX	EX0
IT	ITJ
NN	NN0 / NN021 / NN022 / NN031 / NN032 /NN033 NN1 / NN121 / NN122 / NN2 / NN221 / NN222
NP	NPO
OR	ORD
PN	PNI / PNI21 / PNI22 / PNP / PNQ / PNQ31 PNQ32 / PNQ33 / PNX / PNX21 / PNX22
PO	POS
PR	PRF / PRP / PRP21 / PRP 22/ PRP 31 / PRP 32 / PRP 33 / PRP 41 / PRP42 / PRP43 / PRP44
TO	TO0
UN	UNC
VB	VBB / VBD / VBG / VBI / VBN / VBZ /VDB / VDD / VDG VDI VDN / VDZ / VHB / VHD / VHG / VHI / VHN / VHZ VM0 / VVB / VVD / VVG / VVI/ VVZ
XX	XX0
ZZ	ZZ0 / ZZ021 /Z022

<부록 3>

MECO 상위 100위까지의 어휘목록 (Word-list)

순위	어휘	순위	어휘	순위	어휘
1	THE	35	ALL	69	SEA
2	OF	36	SUCH	70	USED
3	AND	37	THEIR	71	SHIPPING
4	TO	38	MARITIME	72	REGULATION
5	IN	39	OTHER	73	OIL
6	A	40	CAN	74	HE
7	IS	41	ITS	75	WHERE
8	FOR	42	ONE	76	USE
9	BE	43	BEEN	77	ENGLISH
10	ON	44	WHEN	78	HAD
11	THAT	45	CARGO	79	UNDER
12	AS	46	ALSO	80	CONVENTION
13	BY	47	THAN	81	WOULD
14	WITH	48	MORE	82	OUT
15	OR	49	IF	83	SO
16	ARE	50	THEY	84	IMO
17	THIS	51	ANY	85	BETWEEN
18	AT	52	WERE	86	ONLY
19	FROM	53	TIME	87	WATER
20	IT	54	PORT	88	MARINE
21	WHICH	55	VESSELS	89	REQUIREMENTS
22	NOT	56	SYSTEM	90	PART
23	AN	57	INTO	91	POSITION
24	WAS	58	BUT	92	EACH
25	SHALL	59	NO	93	REQUIRED
26	SHIP	60	THESE	94	UP
27	WILL	61	TWO	95	INFORMATION
28	HAVE	62	SAFETY	96	YOU
29	S	63	SHOULD	97	FIRST
30	HAS	64	WE	98	GENERAL
31	VESSEL	65	THERE	99	BEING
32	SHIPS	66	NEW	100	PROVIDED
33	I	67	INTERNATIONAL		
34	MAY	68	FIRE		

<부록 4>

MECO 상위 100위까지 키워드 목록 (Keyword-list)

순위	키워드	순위	키워드	순위	키워드
1	SHIP	35	RO	71	BALLAST
2	VESSEL	36	PORTS	72	ROUTE
3	SHALL	37	REQUIRED	73	LIABILITY
4	SHIPS	38	PROVIDED	74	DAMAGE
5	MARITIME	39	STCW	75	CONSTRUCTION
6	CARGO	40	COMMUNICATION	76	FUEL
7	VESSELS	42	DATA	77	BASED
8	PORT	43	COLLISION	78	TONNAGE
9	SHIPPING	44	EMERGENCY	79	CERTIFICATE
10	REGULATION	45	ORGANIZATION	80	PILOT
11	SAFETY	46	ANNEX	81	CONTROL
12	IMO	47	BE	82	BULK
13	OIL	48	CONTRACT	83	NAVIGATIONAL
14	CONVENTION	49	BOARD	84	OFFSHORE
15	MARINE	50	ADMINISTRATION	85	RESCUE
16	INTERNATIONAL	51	SECURITY	86	FREIGHT
17	SPACES	52	PIRACY	87	ENVIRONMENT
18	REQUIREMENTS	53	TANK	88	MASTER
19	FIRE	54	LANGUAGE	89	GPS
20	NAVIGATION	55	MACHINERY	90	DISCHARGE
21	DECK	56	OPERATIONAL	91	TRANSPORT
22	SEA	57	RADAR	92	ARTICLE
23	CREW	58	REGULATIONS	93	RISK
24	SYSTEM	59	LIMITATION	94	SPEED
25	SYSTEMS	60	CONTAINER	95	PROJECT
26	CODE	61	OPERATIONS	96	ENSURE
27	PARAGRAPH	63	GLOBAL	97	MSC
28	TRAFFIC	64	POLLUTION	98	MERCHANT
29	SEAFARERS	65	POSITION	99	SPACE
30	OR	66	TANKERS	100	RESOLUTION
31	ENGLISH	67	MANAGEMENT		
32	INFORMATION	68	TRAINING		
33	PASSENGER	69	CHARTER		
34	EQUIPMENT	70	OPERATION		

<부록 5>

누적빈도 백분율 80%까지의 명사 목록

1	ship	26	number	51	regulations	76	certificate	101	purpose
2	ships	27	training	52	operations	77	transport	102	goods
3	vessel	28	navigation	53	way	78	force	103	rule
4	cargo	29	use	54	seafarers	79	areas	104	length
5	vessels	30	traffic	55	communication	80	period	105	point
6	system	31	crew	56	security	81	level	106	arrangements
7	safety	32	equipment	57	management	82	contract	107	route
8	time	33	company	58	article	83	limitation	108	coast
9	regulation	34	control	59	machinery	84	means	109	environment
10	oil	35	data	60	speed	85	radar	110	tank
11	port	36	area	61	years	86	knowledge	111	process
12	convention	37	paragraph	62	conditions	87	risk	112	date
13	fire	38	case	63	officer	88	section	113	provisions
14	sea	39	industry	64	world	89	report	114	rules
15	requirements	40	administration	65	emergency	90	collision	115	pilot
16	water	41	power	66	side	91	range	116	pollution
17	part	42	students	67	ports	92	chapter	117	action
18	position	43	course	68	line	93	group	118	capacity
19	information	44	code	69	damage	94	engine	119	research
20	spaces	45	space	70	figure	95	market	120	container
21	shipping	46	english	71	construction	96	development	121	protection
22	language	47	operation	72	work	97	fuel	122	imo
23	systems	48	organization	73	services	98	officers	123	bridge
24	service	49	passenger	74	standards	99	state	124	party
25	deck	50	year	75	master	100	liability	125	piracy

126	class	151	effect	176	rate	201	protocol	226	competence
127	resolution	152	project	177	committee	202	countries	227	others
128	bulk	153	government	178	paper	203	tanker	228	technology
129	annex	154	tanks	179	companies	204	facilities	229	assessment
130	merchant	155	analysis	180	situation	205	method	230	source
131	loss	156	radio	181	passengers	206	fact	231	fleet
132	gas	157	application	182	end	207	aircraft	232	culture
133	person	158	results	183	personnel	208	problems	233	boat
134	design	159	procedures	184	stability	209	structure	234	result
135	parties	160	accident	185	engineering	210	issues	235	day
136	members	161	carrier	186	category	211	approach	236	target
137	ballast	162	wind	187	signal	212	hours	237	weather
138	months	163	place	188	life	213	station	238	basis
139	air	164	type	189	act	214	access	239	detection
140	trade	165	watch	190	need	215	energy	240	practice
141	tankers	166	survey	191	test	216	gear	241	waters
142	measures	167	words	192	discharge	217	plan	242	doors
143	value	168	university	193	cases	218	skills	243	member
144	persons	169	problem	194	stations	219	pressure	244	direction
145	distance	170	ability	195	table	220	guidelines	245	change
146	business	171	conference	196	freight	221	prevention	246	signals
147	performance	172	tonnage	197	separation	222	miles	247	team
148	fishing	173	model	198	charter	223	carriage	248	list
149	law	174	learning	199	activities	224	cargoes	249	methods
150	study	175	amendments	200	people	225	lights	250	material

251	search	276	department	301	support	326	condition	351	interest
252	room	277	responsibility	302	investigation	327	costs	352	draft
253	event	278	carriers	303	programme	328	avoidance	353	emissions
254	types	279	governments	304	entry	329	scheme	354	region
255	experience	280	transportation	305	channel	330	levels	355	sector
256	voyage	281	circumstances	306	substances	331	size	356	name
257	assistance	282	agreement	307	no.	332	owner	357	account
258	gps	283	factor	308	parts	333	changes	358	units
259	accommodation	284	efficiency	309	error	334	degrees	359	days
260	role	285	appliances	310	starboard	335	unit	360	transfer
261	metres	286	owners	311	education	336	distress	361	provision
262	terms	287	shipowner	312	quarter	337	impact	362	inspection
263	points	288	factors	313	door	338	flag	363	word
264	quality	289	assembly	314	location	339	message	364	shipowners
265	question	290	communications	315	escape	340	institute	365	decision
266	rescue	291	cost	316	chain	341	importance	366	integrity
267	issue	292	implementation	317	duties	342	routes	367	certification
268	certificates	293	form	318	light	343	guidance	368	east
269	craft	294	accidents	319	attention	344	introduction	369	operators
270	maintenance	295	bill	320	alarm	345	visibility	370	nature
271	term	296	order	321	supply	346	centre	371	right
272	solution	297	lines	322	reference	347	delivery	372	growth
273	bulkhead	298	claims	323	secretary-general	348	survival	373	student
274	bulkheads	299	times	324	teacher	349	materials	374	response
275	engines	300	propulsion	325	insurance	350	anchor	375	questions

376	guard	401	function	426	workers	451	paragraphs	476	ais
377	pirates	402	union	427	navy	452	loading	477	requirement
378	lading	403	teachers	428	increase	453	details	478	differences
379	hull	404	instructions	429	purposes	454	groups	479	projects
380	rates	405	steel	430	operating	455	field	480	situations
381	minutes	406	courses	431	crews	456	trainees	481	load
382	authority	407	satellite	432	policy	457	north	482	surface
383	consideration	408	session	433	terminal	458	lifeboat	483	meeting
384	seafarer	409	orders	434	hand	459	participants	484	values
385	boats	410	states	435	mode	460	strategy	485	organizations
386	nations	411	vhf	436	record	461	demand	486	incidents
387	book	412	pumps	437	engineer	462	office	487	display
388	approval	413	board	438	efforts	463	studies	488	teaching
389	country	414	failure	439	testing	464	frequency	489	learners
390	standard	415	task	440	amount	465	openings	490	adoption
391	phase	416	gulf	441	claim	466	integration	491	handling
392	ladder	417	vocabulary	442	court	467	products	492	understanding
393	production	418	positions	443	facility	468	south	493	director
394	scale	419	infrastructure	444	installations	469	compliance	494	view
395	subdivision	420	ventilation	445	text	470	installation	495	command
396	danger	421	flow	446	incident	471	tug	496	height
397	head	422	price	447	phrases	472	depth	497	containers
398	steering	423	extent	448	zone	473	president	498	distribution
399	volume	424	errors	449	example	474	subject	499	canal
400	knots	425	awareness	450	council	475	characteristics	500	cadets

501	authorities	526	pump	551	draught	576	accuracy	601	principles
502	strategies	527	community	552	planning	577	aid	602	berth
503	release	528	recycling	553	framework	578	skipper	603	chart
504	choice	529	movement	554	conduct	579	future	604	reasons
505	storage	530	harbour	555	procedure	580	shell	605	west
506	evidence	531	agency	556	transit	581	recommendations	606	compartments
507	decks	532	investment	557	effects	582	notice	607	focus
508	measurements	533	writing	558	body	583	items	608	land
509	actions	534	set	559	angle	584	monitoring	609	document
510	mind	535	observation	560	bottom	585	tonnes	610	stern
511	acts	536	conflict	561	road	586	stages	611	techniques
512	ro.ro	537	vehicle	562	robbery	587	contribution	612	agents
513	passage	538	fault	563	bow	588	solutions	613	helicopter
514	receiver	539	climate	564	opportunity	589	attacks	614	measurement
515	executive	540	algorithm	565	tons	590	context	615	scope
516	share	541	warning	566	vector	591	degree	616	review
517	attitude	542	reduction	567	ocean	592	reason	617	difference
518	association	543	winch	568	series	593	casualty	618	call
519	masters	544	tests	569	option	594	flooding	619	intervals
520	cent	545	platform	570	instruments	595	limits	620	objectives
521	manager	546	functions	571	status	596	onboard	621	reliability
522	resources	547	shore	572	instrument	597	availability	622	manual
523	per	548	tasks	573	devices	598	diesel	623	concern
524	activity	549	week	574	reports	599	matter	624	rudder
525	operator	550	lack	575	contracts	600	coal	625	targets

626	salvage	651	generation	676	insurer	701	quantity	726	noise
627	aspects	652	division	677	fairway	702	p.	727	program
628	crisis	653	rights	678	zones	703	legislation	728	outcome
629	dwt	654	content	679	barge	704	buoy	729	customer
630	concept	655	objective	680	recovery	705	satellites	730	prices
631	instruction	656	tool	681	sound	706	speech	731	combination
632	background	657	element	682	inspections	707	strength	732	examples
633	amendment	658	temperature	683	sample	708	award	733	events
634	stage	659	min	684	identification	709	voyages	734	elements
635	decisions	660	classroom	685	title	710	lane	735	divisions
636	surveys	661	society	686	bell	711	intention	736	sentence
637	staff	662	valves	687	competition	712	branch	737	ratings
638	plans	663	month	688	velocity	713	institutions	738	components
639	co-operation	664	aids	689	partners	714	compartment	739	regard
640	propeller	665	bilge	690	models	715	product	740	periods
641	agent	666	advantage	691	bay	716	benefits	741	things
642	man	667	network	692	briefing	717	practices	742	features
643	secretary	668	embarkation	693	economy	718	buoys	743	buyer
644	duty	669	mariners	694	sides	719	forces	744	property
645	seas	670	acceptance	695	recognition	720	meaning	745	exercises
646	markets	671	kind	696	possibility	721	tables	746	sales
647	satisfaction	672	tracking	697	compensation	722	http	747	risks
648	sailing	673	applications	698	limit	723	manner	748	reception
649	arrangement	674	hour	699	positioning	724	care	749	uwb
650	statement	675	rail	700	sources	725	charterer	750	filter

751	o.	776	bank	801	bearings	826	extinction	851	wire
752	academy	777	ft	802	ratification	827	bearing	852	exercise
753	user	778	valve	803	progress	828	matters	853	motor
754	classification	779	terminals	804	income	829	science	854	speakers
755	principle	780	customers	805	track	830	tugs	855	object
756	industries	781	competency	806	cause	831	tools	856	turn
757	mm	782	criteria	807	opportunities	832	guide	857	rest
758	developments	783	shipper	808	coverage	833	hold	858	selection
759	smoke	784	health	809	administrations	834	rig	859	relationship
760	sectors	785	opinion	810	liferafts	835	cursor	860	schemes
761	drills	786	opening	811	languages	836	options	861	forms
762	shipbuilding	787	sale	812	ones	837	examination	862	interval
763	reporting	788	documents	813	rooms	838	segment	863	idea
764	boundaries	789	presentation	814	patterns	839	guide	864	density
765	ways	790	workshop	815	discussion	840	jobs	865	responsibilities
766	charts	791	weight	816	vicinity	841	controls	866	respect
767	advantages	792	corporation	817	sentences	842	presence	867	substance
768	destination	793	majority	818	repairs	843	shipyard	868	calculations
769	liner	794	definitions	819	managers	844	solas	869	classes
770	expansion	795	apparatus	820	definition	845	effectiveness	870	pilots
771	enforcement	796	cultures	821	lighting	846	ins	871	house
772	exchange	797	advice	822	hazards	847	description	872	ministry
773	pattern	798	commission	823	needs	848	--	873	m.
774	steps	799	republic	824	directions	849	home	874	margin
775	ice	800	pipe	825	ownership	850	itf	875	step

876	translation	901	compass	926	detail	951	grain	976	damages
877	cable	902	frame	927	liquids	952	shipboard	977	letter
878	numbers	903	sense	928	ground	953	benefit	978	input
879	determination	904	regions	929	recommendation	954	motion	979	improvement
880	objects	905	starting	930	articles	955	users	980	foundation
881	feedback	906	history	931	building	956	vol.	981	potential
882	top	907	gases	932	behaviour	957	connection	982	kw
883	superintendent	908	lifeboats	933	conventions	958	software	983	device
884	chains	909	observations	934	shaft	959	outcomes	984	waterline
885	race	910	evaluation	935	correlation	960	challenges	985	demands
886	school	911	trading	936	mechanism	961	success	986	schedule
887	seamen	912	average	937	explosion	962	pairs	987	contents
888	contact	913	pipes	938	capital	963	aim	988	holds
889	age	914	drivers	939	charters	964	priority	989	output
890	vehicles	915	capability	940	look-out	965	database	990	river
891	indicators	916	profit	941	estimation	966	distances	991	programmes
892	alternative	917	engineers	942	base	967	processing	992	back
893	switchboard	918	endorsement	943	cycle	968	sight	993	parameters
894	tax	919	technologies	944	stowage	969	effort	994	prediction
895	ferry	920	messages	945	seminar	970	loran	995	shapes
896	reef	921	chairman	946	variety	971	collisions	996	wheelhouse
897	ducts	922	dynamics	947	relationships	972	shippers	997	petroleum
898	plating	923	launching	948	galley	973	capabilities	998	request
899	lessons	924	dimensions	949	domain	974	extension	999	tide
900	completion	925	acquisition	950	rating	975	regime	1000	hazard

1001	ideas	1026	agencies	1051	night	1076	revision	1101	bunker
1002	injuries	1027	notes	1052	percent	1077	cabin	1102	structures
1003	accession	1028	exhaust	1053	beam	1078	generator	1103	mate
1004	proposal	1029	heading	1054	simulation	1079	minister	1104	scene
1005	interaction	1030	consequences	1055	captain	1080	money	1105	plant
1006	college	1031	aviation	1056	obe	1081	proceedings	1106	residues
1007	pirate	1032	papers	1057	fund	1082	ambiguity	1107	ceremony
1008	jurisdiction	1033	discourse	1058	bodies	1083	advance	1108	minimum
1009	vision	1034	address	1059	headquarters	1084	disaster	1109	calculation
1010	measure	1035	duration	1060	appendix	1085	organizations	1110	drilling
1011	topics	1036	career	1061	subjects	1086	turbines	1111	cooperation
1012	weeks	1037	employees	1062	carbon	1087	attack	1112	cash
1013	mareng	1038	itineraries	1063	logistics	1088	readiness	1113	obstacle
1014	arrival	1039	delay	1064	limitations	1089	job	1114	quaternion
1015	interface	1040	preparation	1065	respondents	1090	cycles	1115	me
1016	interpretation	1041	start	1066	relations	1091	family	1116	membership
1017	piping	1042	manning	1067	version	1092	categories	1117	latitude
1018	concerns	1043	conclusions	1068	steam	1093	conclusion	1118	percentage
1019	texts	1044	raim	1069	gis	1094	modes	1119	centres
1020	diameter	1045	exception	1070	horizon	1095	teams	1120	cables
1021	motivation	1046	organization	1071	men	1096	voice	1121	interests
1022	sulphur	1047	individual	1072	antenna	1097	alteration	1122	trim
1023	launch	1048	washing	1073	exploration	1098	ignition	1123	employment
1024	literature	1049	confidence	1074	index	1099	negotiations	1124	boiler
1025	treatment	1050	manufacturers	1075	total	1100	establishment	1125	grammar

1126	threat	1151	unions	1176	representatives	1201	transmission	1226	disadvantages
1127	indication	1152	departure	1177	publication	1202	angles	1227	precautions
1128	pumping	1153	lakes	1178	shape	1203	entrance	1228	permission
1129	topic	1154	commitment	1179	lifejackets	1204	trust	1229	challenge
1130	shock	1155	obligations	1180	answers	1205	computer	1230	lot
1131	difficulties	1156	hands	1181	difficulty	1206	pronunciation	1231	vapour
1132	corridors	1157	panel	1182	edition	1207	species	1232	ore
1133	movements	1158	spill	1183	dock	1208	copy	1233	city
1134	component	1159	laws	1184	surveyor	1209	privacy	1234	colleagues
1135	scenario	1160	turning	1185	alternatives	1210	scenarios	1235	muster
1136	firm	1161	seller	1186	fog	1211	expenses	1236	indicator
1137	answer	1162	phrase	1187	lanes	1212	whole	1237	latency
1138	trial	1163	msc	1188	site	1213	dangers	1238	mask
1139	revenue	1164	emphasis	1189	hydrogen	1214	help	1239	charge
1140	assets	1165	skill	1190	trainee	1215	balance	1240	ratio
1141	war	1166	clause	1191	significance	1216	boundary	1241	blocks
1142	instructors	1167	register	1192	mile	1217	coastguard	1242	algorithms
1143	authors	1168	findings	1193	beginning	1218	general	1243	ton
1144	tier	1169	audit	1194	credit	1219	representation	1244	registration
1145	charges	1170	forum	1195	proficiency	1220	doubt	1245	th.
1146	half	1171	congress	1196	feature	1221	responses	1246	t.
1147	peak	1172	routine	1197	payment	1222	figures	1247	consumption
1148	preference	1173	repair	1198	experts	1223	comparison	1248	leadership
1149	officials	1174	volumes	1199	meetings	1224	oils	1249	probability
1150	liferaft	1175	manoeuvre	1200	earth	1225	statistics	1250	rise

1251	representative	1276	bias	1301	injury	1326	barrier	1351	attributes
1252	suppliers	1277	modifications	1302	%	1327	individuals	1352	revenues
1253	whistle	1278	width	1303	news	1328	chamber	1353	kalman
1254	removal	1279	path	1304	alterations	1329	relief	1354	federation
1255	grounding	1280	casualties	1305	surveyors	1330	statements	1355	gravity
1256	technique	1281	wheel	1306	aspect	1331	barges	1356	funds
1257	press	1282	note	1307	anchorage	1332	conflicts	1357	goal
1258	will	1283	winds	1308	wave	1333	theory	1358	channels
1259	sets	1284	game	1309	cylinder	1334	resolutions	1359	well
1260	circuit	1285	insurers	1310	workload	1335	closure	1360	isolation
1261	evacuation	1286	records	1311	freeboard	1336	superstructure	1361	fleets
1262	mechanisms	1287	diagram	1312	co	1337	disasters	1362	businesses
1263	facts	1288	iron	1313	clearance	1338	greenhouse	1363	sensors
1264	stairways	1289	origin	1314	theme	1339	fitness	1364	foam
1265	specialist	1290	goals	1315	bills	1340	companys	1365	spot
1266	heat	1291	renewal	1316	heater	1341	seawater	1366	km
1267	worlds	1292	considerations	1317	genre	1342	emission	1367	criterion
1268	apr	1293	proposals	1318	sign	1343	campaign	1368	stores
1269	insulation	1294	appliance	1319	delegates	1344	comments	1369	liquid
1270	suppression	1295	drill	1320	strait	1345	lives	1370	labour
1271	agreements	1296	mixtures	1321	journal	1346	hostage	1371	center
1272	echo	1297	plane	1322	masthead	1347	sub-paragraph	1372	assumptions
1273	truck	1298	publications	1323	deal	1348	causes	1373	return
1274	funding	1299	extinguishers	1324	ferries	1349	quantities	1374	reform
1275	atmosphere	1300	announcement	1325	tow	1350	feet	1375	shipyards

1376	boxes	1401	questionnaire	1426	transmitter	1451	caution	1476	contributions
1377	trades	1402	trend	1427	locations	1452	morning	1477	deadweight
1378	consequence	1403	face	1428	contractor	1453	employer	1478	star
1379	mining	1404	cover	1429	coordination	1454	crossing	1479	stress
1380	stock	1405	flight	1430	column	1455	trajectory	1480	geometry
1381	maximum	1406	50%	1431	longitude	1456	dredging	1481	met
1382	fuels	1407	salvor	1432	fish	1457	charterparty	1482	bunkers
1383	keywords	1408	duct	1433	processes	1458	restrictions	1483	innovation
1384	pitch	1409	pricing	1434	segments	1459	a.	1484	shortage
1385	influence	1410	manoeuvres	1435	propellers	1460	sections	1485	discussions
1386	formula	1411	art	1436	buyers	1461	lead	1486	restoration
1387	obligation	1412	pulse	1437	minute	1462	teu	1487	waves
1388	stairway	1413	resistance	1438	choices	1463	export	1488	hatch
1389	elevation	1414	fighting	1439	alarms	1464	possession	1489	hoist
1390	hoses	1415	modification	1440	improvements	1465	marpol	1490	assertiveness
1391	circle	1416	budget	1441	banks	1466	memory	1491	cleaning
1392	population	1417	identity	1442	partnership	1467	claimant	1492	alerts
1393	update	1418	institution	1443	chemicals	1468	seaman	1493	proof
1394	collection	1419	simulator	1444	processor	1469	marks	1494	wall
1395	dispute	1420	charterers	1445	affairs	1470	manoeuvrability	1495	survivors
1396	trip	1421	arm	1446	roles	1471	allocation	1496	sensor
1397	absence	1422	drainage	1447	qualifications	1472	families	1497	screen
1398	manoeuvring	1423	mepc	1448	instructor	1473	present	1498	negotiation
1399	concentration	1424	losses	1449	investigations	1474	waterways	1499	deposit
1400	transshipment	1425	constraints	1450	edge	1475	frames	1500	stakeholders

1501	awards	1526	medium	1551	expressions	1576	interference	1601	washings
1502	yard	1527	architecture	1552	m	1577	police	1602	exposure
1503	corrections	1528	core	1553	turbine	1578	pipeline	1603	forepeak
1504	eta	1529	equation	1554	deployment	1579	firms	1604	sub-committee
1505	receipt	1530	maib	1555	congestion	1580	shares	1605	net
1506	variables	1531	mixture	1556	delays	1581	denunciation	1606	electricity
1507	approaches	1532	keel	1557	holder	1582	perspective	1607	internet
1508	partner	1533	degr.	1558	storm	1583	club	1608	tracks
1509	estimates	1534	plates	1559	mariner	1584	marketing	1609	formation
1510	breadth	1535	summary	1560	observer	1585	estimate	1610	implications
1511	suction	1536	assumption	1561	spillage	1586	necessity		
1512	places	1537	moment	1562	specifications	1587	metre		
1513	customs	1538	shareholders	1563	page	1588	expertise		
1514	marker	1539	provider	1564	shipments	1589	experiment		
1515	buoyancy	1540	shipment	1565	mark	1590	obstacles		
1516	setting	1541	autonomy	1566	references	1591	speaker		
1517	weapons	1542	portion	1567	listening	1592	registry		
1518	signature	1543	dsc	1568	boilers	1593	periodical		
1519	carrying	1544	piston	1569	discharges	1594	arbitration		
1520	candidate	1545	corporations	1570	matrix	1595	video		
1521	vapours	1546	windows	1571	remarks	1596	stem		
1522	misrepresentation	1547	move	1572	terrain	1597	fire-extinguishing		
1523	quarters	1548	expression	1573	precision	1598	initiative		
1524	surfaces	1549	deviation	1574	death	1599	conversation		
1525	expiry	1550	hub	1575	proximity	1600	checks		

<부록 6>

MECO와 LOB의 고빈도 상위 32위까지 명사 목록

순위	MECO 명사	LOB 명사
1	ship	time
2	ships	man
3	vessel	years
4	cargo	back
5	vessels	people
6	system	work
7	safety	year
8	time	life
9	regulation	pound
10	oil	men
11	port	world
12	convention	house
13	fire	part
14	sea	course
15	requirements	home
16	water	fact
17	part	case
18	position	hand
19	information	place
20	spaces	government
21	shipping	number
22	language	sir
23	system	room
24	service	four
25	deck	days
26	number	half
27	training	war
28	navigation	water
29	use	head
30	traffic	school
31	crew	side
32	equipment	children

<부록 7>

누적빈도 백분율 80%까지의 동사 목록

1	is	26	taken	51	applied	76	following	101	entitled
2	be	27	does	52	located	77	included	102	improve
3	are	28	fitted	53	taking	78	specified	103	enter
4	was	29	based	54	keep	79	includes	104	presented
5	have	30	carried	55	carrying	80	go	105	support
6	has	31	having	56	prevent	81	put	106	understand
7	been	32	include	57	issued	82	operating	107	prescribed
8	were	33	considered	58	shown	83	working	108	containing
9	had	34	apply	59	developed	84	refer	109	applies
10	used	35	set	60	reduce	85	requires	110	obtained
11	being	36	become	61	agreed	86	caused	111	complying
12	required	37	designed	62	engaged	87	known	112	brought
13	made	38	held	63	work	88	compared	113	help
14	provided	39	expected	64	referred	89	reported	114	read
15	do	40	need	65	making	90	completed	115	reduced
16	said	41	means	66	intended	91	consider	116	received
17	using	42	shipping	67	called	92	find	117	indicate
18	provide	43	carry	68	constructed	93	continue	118	operated
19	see	44	give	69	providing	94	operate	119	indicated
20	take	45	found	70	defined	95	say	120	closed
21	make	46	avoid	71	seen	96	develop	121	remain
22	ensure	47	require	72	established	97	allow	122	conducted
23	given	48	meet	73	approved	98	identified	123	come
24	use	49	comply	74	determined	99	follows	124	done
25	adopted	50	did	75	provides	100	entered	125	maintain

126	related	151	associated	176	address	201	establish	226	bring
127	met	152	noted	177	control	202	's	227	run
128	added	153	show	178	signed	203	proceed	228	became
129	followed	154	lead	179	cause	204	receive	229	follow
130	determine	155	built	180	teaching	205	equipped	230	told
131	installed	156	recommended	181	learn	206	discharged	231	transferred
132	know	157	giving	182	increased	207	deemed	232	takes
133	maintained	158	placed	183	contained	208	permit	233	writing
134	get	159	described	184	am	209	connected	234	serve
135	rescue	160	sailing	185	involved	210	announced	235	allowed
136	hold	161	learning	186	serving	211	perform	236	stand
137	shows	162	regarded	187	exceed	212	identify	237	reducing
138	took	163	makes	188	create	213	involving	238	contain
139	speaking	164	enable	189	consists	214	leading	239	passed
140	relating	165	employed	190	covered	215	understood	240	proposed
141	accepted	166	increase	191	occurred	216	introduced	241	restricted
142	discussed	167	owned	192	paid	217	resulting	242	leaving
143	needs	168	led	193	going	218	arranged	243	developing
144	kept	169	note	194	requested	219	obtain	244	stowed
145	stated	170	assigned	195	delivered	220	published	245	result
146	achieved	171	limit	196	act	221	recorded	246	observed
147	pass	172	check	197	indicates	222	measured	247	reach
148	amended	173	assist	198	permitted	223	communicate	248	states
149	needed	174	started	199	assumed	224	ensuring	249	comes
150	change	175	achieve	200	move	225	call	250	form

251	ranging	276	protect	301	ships	326	doing	351	laid
252	left	277	drawn	302	depends	327	opened	352	supply
253	tested	278	becomes	303	answer	328	protected	353	determining
254	secured	279	let	304	created	329	produced	354	write
255	asked	280	extend	305	drilling	330	exercise	355	awarded
256	entering	281	confirmed	306	committed	331	replaced	356	indicating
257	launched	282	meeting	307	present	332	supported	357	extended
258	testing	283	bearing	308	applying	333	produce	358	suggested
259	cover	284	mean	309	reached	334	attended	359	fly
260	recognized	285	running	310	offer	335	facilitate	360	transfer
261	speak	286	allows	311	spill	336	standing	361	concluded
262	expressed	287	look	312	complete	337	affect	362	came
263	proceeding	288	preventing	313	combined	338	reflect	363	test
264	mentioned	289	exceeding	314	seems	339	implemented	364	informed
265	keeping	290	think	315	raised	340	addressed	365	supplied
266	situated	291	stop	316	leave	341	constitute	366	contains
267	gives	292	build	317	ship	342	examined	367	contact
268	limited	293	continued	318	offered	343	enters	368	released
269	performed	294	pay	319	aimed	344	moved	369	causing
270	remains	295	inform	320	accept	345	controlled	370	discharge
271	represents	296	ended	321	written	346	discuss	371	heard
272	start	297	occur	322	enhance	347	generated	372	estimated
273	annex	298	forming	323	appear	348	showing	373	ask
274	lost	299	calculated	324	uses	349	want	374	conduct
275	served	300	assess	325	towing	350	loading	375	navigating

376	covers	401	went	426	coming	451	passing	476	exist
377	registered	402	chosen	427	stopped	452	manoeuvre	477	monitoring
378	began	403	encourage	428	affected	453	vary	478	treated
379	moving	404	offers	429	avoided	454	add	479	bulk
380	removed	405	planned	430	planning	455	depend	480	focused
381	sent	406	happened	431	try	456	launching	481	experienced
382	changed	407	submitted	432	starting	457	approaching	482	formed
383	continues	408	involves	433	getting	458	saw	483	replace
384	loaded	409	marked	434	mining	459	turn	484	creating
385	exists	410	cross	435	suffered	460	watchkeeping	485	looking
386	resulted	411	dealing	436	join	461	explained	486	undertake
387	divided	412	selected	437	costs	462	advised	487	relate
388	listed	413	holding	438	decide	463	attached	488	estimate
389	report	414	managed	439	decided	464	assessed	489	racing
390	expect	415	deal	440	fall	465	securing	490	generate
391	refers	416	manoeuvring	441	trained	466	becoming	491	fill
392	appears	417	understanding	442	prepare	467	load	492	seeking
393	represent	418	detected	443	allowing	468	worked	493	discharging
394	representing	419	open	444	separated	469	navigate	494	deliver
395	play	420	begin	445	shipped	470	demonstrated	495	seek
396	raise	421	exhibit	446	acting	471	scheduled	496	demonstrate
397	minutes	422	promote	447	meant	472	adopt	497	contribute
398	believe	423	respond	448	fixed	473	covering	498	granted
399	represented	424	named	449	writes	474	happen	499	minimize
400	focus	425	maintaining	450	facing	475	stay	500	damaged

501	operates	526	seem	551	choose	576	revised	601	fell
502	handle	527	involve	552	attend	577	designated	602	place
503	steering	528	switch	553	express	578	share	603	enabling
504	experience	529	agree	554	remained	579	extending	604	runs
505	establishing	530	explain	555	joining	580	stopping	605	market
506	taught	531	holds	556	like	581	positioned	606	displayed
507	turned	532	reading	557	meets	582	commenced	607	improving
508	return	533	tell	558	plotting	583	arrived	608	points
509	assume	534	checked	559	consist	584	secure	609	transmit
510	arise	535	prohibited	560	consisting	585	enables	610	failed
511	introduce	536	undertaken	561	presents	586	evaluate	611	review
512	implementing	537	prepared	562	including	587	gave	612	setting
513	inspected	538	learned	563	increasing	588	implement	613	sail
514	teach	539	close	564	improved	589	damage	614	closing
515	suggest	540	collected	565	leads	590	handling	615	pointed
516	accommodate	541	invited	566	sets	591	appointed	616	recognised
517	remove	542	joined	567	says	592	define	617	trying
518	transported	543	feel	568	investigate	593	checking	618	remember
519	claims	544	occurs	569	termed	594	prove	619	distributed
520	receiving	545	lies	570	ordered	595	face	620	directed
521	driven	546	brings	571	argued	596	state	621	adding
522	goes	547	examine	572	revealed	597	finding	622	manage
523	stored	548	tend	573	exposed	598	considers	623	deployed
524	considering	549	nominated	574	showed	599	alter	624	recover
525	aims	550	preparing	575	escape	600	differ	625	organized

626	requiring	651	service	676	recognize	701	measure	726	suspended
627	send	652	detect	677	authorized	702	imposed	727	reflected
628	addressing	653	studying	678	investigated	703	increases		
629	turning	654	falling	679	approach	704	conclude		
630	boarded	655	acknowledged	680	plan	705	bringing		
631	extends	656	avoiding	681	gained	706	focuses		
632	request	657	abaft	682	draft	707	commands		
633	carries	658	notify	683	encountered	708	contributing		
634	benefit	659	updated	684	comparing	709	excluded		
635	code	660	adapted	685	bound	710	declared		
636	satisfied	661	controlling	686	complied	711	derived		
637	influence	662	reviewed	687	expand	712	issue		
638	saying	663	crossing	688	got	713	governed		
639	comment	664	moves	689	overcome	714	thought		
640	reporting	665	compare	690	measuring	715	embark		
641	transmitted	666	intends	691	confirm	716	rendered		
642	leaves	667	won	692	proved	717	sound		
643	producing	668	supporting	693	reflects	718	deposited		
644	process	669	begins	694	handled	719	wrote		
645	reaching	670	sold	695	calling	720	cut		
646	evaluated	671	promoting	696	gain	721	describe		
647	predicted	672	satisfy	697	offering	722	interfere		
648	submit	673	participate	698	predict	723	dealt		
649	monitor	674	identifying	699	integrated	724	display		
650	reduces	675	wait	700	recovered	725	listening		

<부록 8>

누적빈도 백분율 80%까지의 형용사 목록

1	maritime	26	small	51	capable	76	great	101	proper
2	other	27	additional	52	lower	77	personal	102	wide
3	international	28	able	53	human	78	key	103	required
4	new	29	current	54	standard	79	minimum	104	initial
5	marine	30	effective	55	sufficient	80	horizontal	105	ready
6	main	31	total	56	offshore	81	nautical	106	original
7	english	32	clear	57	common	82	united	107	successful
8	general	33	major	58	basic	83	heavy	108	previous
9	different	34	navigational	59	higher	84	european	109	involved
10	available	35	technical	60	environmental	85	concerned	110	independent
11	following	36	specific	61	vertical	86	naval	111	senior
12	special	37	full	62	certain	87	individual	112	approved
13	possible	38	significant	63	low	88	difficult	113	complete
14	necessary	39	fixed	64	single	89	practicable	114	final
15	present	40	particular	65	watertight	90	financial	115	natural
16	safe	41	electrical	66	essential	91	responsible	116	working
17	high	42	relevant	67	economic	92	recent	117	commercial
18	appropriate	43	further	68	greater	93	applicable	118	coastal
19	important	44	chief	69	direct	94	maximum	119	limited
20	operational	45	open	70	cultural	95	potential	120	future
21	large	46	dangerous	71	long	96	equivalent	121	actual
22	global	47	various	72	legal	97	automatic	122	strong
23	good	48	similar	73	public	98	armed	123	ltd
24	short	49	existing	74	best	99	only	124	free
25	national	50	local	75	suitable	100	likely	125	due

126	efficient	151	average	176	revised	201	complex	226	aft
127	adequate	152	central	177	electric	202	federal	227	cardinal
128	chinese	153	seagoing	178	useful	203	developing	228	entire
129	proposed	154	life-saving	179	restricted	204	subsequent	229	reasonable
130	contracting	155	medical	180	integrated	205	fire-extinguishing	230	old
131	close	156	modern	181	aware	206	permanent	231	comprehensive
132	mandatory	157	communicative	182	gross	207	british	232	towing
133	dry	158	larger	183	flammable	208	combined	233	adjacent
134	early	159	crude	184	traditional	209	equal	234	closed
135	positive	160	regular	185	visual	210	white	235	corresponding
136	professional	161	true	186	right	211	given	236	valid
137	internal	162	whole	187	hydraulic	212	narrow	237	assured
138	better	163	latest	188	immediate	213	related	238	intermediate
139	normal	164	separate	189	liable	214	terminal	239	written
140	continuous	165	largest	190	alternative	215	physical	240	harmful
141	overall	166	associated	191	spatial	216	dynamic	241	severe
142	foreign	167	liquid	192	joint	217	smaller	242	linguistic
143	net	168	enclosed	193	correct	218	specialized	243	regulatory
144	external	169	active	194	double	219	chemical	244	strategic
145	above	170	serious	195	easy	220	increased	245	acceptable
146	regional	171	social	196	hazardous	221	floating	246	remaining
147	private	172	annual	197	auxiliary	222	upper	247	simple
148	electronic	173	deep	198	forward	223	american	248	substantial
149	practical	174	portable	199	poor	224	formal	249	clean
150	critical	175	relative	200	real	225	stern	250	advanced

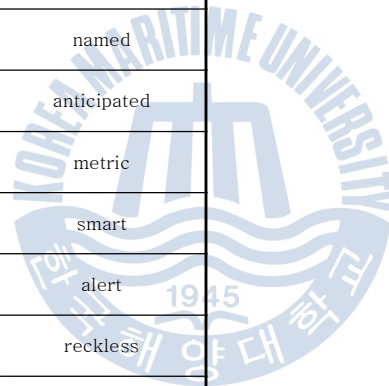
251	reliable	276	satisfactory	301	educational	326	subsea	351	registered
252	detailed	277	used	302	primary	327	familiar	352	french
253	noxious	278	level	303	green	328	brief	353	fundamental
254	domestic	279	genetic	304	designated	329	satisfied	354	minor
255	increasing	280	left	305	mean	330	huge	355	nearest
256	somali	281	military	306	accurate	331	intercultural	356	fresh
257	longer	282	typical	307	inc.	332	inert	357	tidal
258	accessible	283	red	308	extensive	333	swedish	358	corporate
259	civil	284	respective	309	shallow	334	crucial	359	digital
260	conventional	285	competitive	310	considerable	335	leading	360	exact
261	unique	286	interesting	311	thermal	336	valuable	361	autonomous
262	transverse	287	inland	312	fire-fighting	337	planned	362	australian
263	fast	288	sound	313	competent	338	optimum	363	servqual
264	structural	289	negative	314	predicted	339	russian	364	authentic
265	political	290	scientific	315	improved	340	voluntary	365	middle
266	daily	291	vital	316	expected	341	bottom	366	royal
267	static	292	contractual	317	slow	342	young	367	frequent
268	reduced	293	highest	318	estimated	343	apparent	368	binding
269	light	294	remote	319	eastern	344	obvious	369	technological
270	precise	295	past	320	visible	345	excellent	370	online
271	passive	296	bad	321	rapid	346	managing	371	opposite
272	multiple	297	power-driven	322	functional	347	prevailing	372	innovative
273	ro-ro	298	unknown	323	sure	348	official	373	adverse
274	constant	299	unable	324	growing	349	experienced	374	exceptional
275	native	300	long-term	325	industrial	350	protective	375	late

376	mechanical	401	extreme	426	optional	451	strict	476	two-way
377	administrative	402	norwegian	427	inertial	452	asian	477	absolute
378	northern	403	--	428	criminal	453	shore-based	478	emerging
379	sustainable	404	southern	429	manned	454	ltd.	479	random
380	injured	405	impracticable	430	meteorological	455	prior	480	recommended
381	tracking	406	statutory	431	vice	456	thorough	481	driving
382	manual	407	black	432	objective	457	usual	482	improving
383	nuclear	408	deepest	433	fair	458	residual	483	hard
384	expensive	409	outer	434	big	459	inner	484	powerful
385	top	410	variable	435	western	460	worldwide	485	underwater
386	consistent	411	on-board	436	continued	461	constructive	486	inc
387	territorial	412	intended	437	real-time	462	toxic	487	novel
388	solid	413	oily	438	temporary	463	partial	488	flat
389	resulting	414	ordinary	439	numerous	464	occupational	489	multicultural
390	arab	415	labour	440	non-combustible	465	seafaring	490	modal
391	specified	416	principal	441	cubic	466	insufficient	491	FALSE
392	deepwater	417	complicated	442	routeing	467	hydrographic	492	transitional
393	mobile	418	extended	443	defective	468	unnecessary	493	elementary
394	greatest	419	wider	444	inshore	469	well	494	yellow
395	sole	420	known	445	generating	470	watch-keeping	495	academic
396	diplomatic	421	cross-cultural	446	combustible	471	changing	496	excessive
397	impossible	422	protected	447	prime	472	dedicated	497	optimal
398	ongoing	423	abstract	448	unscented	473	mutual	498	nonlinear
399	longitudinal	424	junior	449	nigerian	474	shorter	499	desired
400	indian	425	qualified	450	hot	475	interested	500	steady

501	stable	526	biggest	551	prolonged	576	systematic	601	popular
502	broad	527	superior	552	kinematic	577	robust	602	convenient
503	geographical	528	all-round	553	preliminary	578	sliding	603	italian
504	magnetic	529	aggregate	554	little	579	spare	604	virtual
505	observed	530	earliest	555	goal-based	580	compatible	605	integral
506	geographic	531	experimental	556	quick	581	meaningful	606	situational
507	near	532	secondary	557	dalian	582	friendly	607	enhanced
508	polar	533	intact	558	updated	583	shortest	608	specialised
509	charted	534	associate	559	blue	584	seaworthy	609	market-based
510	parallel	535	out-of-class	560	helpful	585	agreed	610	lateral
511	fit	536	closing	561	cognitive	586	baltic	611	weighted
512	assertive	537	content	562	filipino	587	prospective	612	stringent
513	after	538	audible	563	high-level	588	loaded	613	so-called
514	moulded	539	moderate	564	brazilian	589	tropical	614	unlimited
515	said	540	predictive	565	african	590	extra	615	congested
516	differential	541	renewable	566	distinct	591	solar	616	worst
517	japanese	542	compulsory	567	ample	592	illegal	617	favourable
518	diving	543	dead	568	far	593	universal	618	selected
519	statistical	544	lowest	569	scheduled	594	thick	619	beneficial
520	flexible	545	unlikely	570	sensitive	595	desirable	620	faster
521	guilty	546	grammatical	571	feasible	596	navigable	621	historical
522	dependent	547	pleased	572	power-operated	597	refrigerated	622	careful
523	evident	548	theoretical	573	permissible	598	consultative	623	stand-alone
524	attractive	549	arctic	574	dutch	599	preferred	624	atmospheric
525	earlier	550	non-governmental	575	continuing	600	mental	625	wrong

626	intermodal	651	uniform	676	canadian	701	easier
627	deckhand	652	instrumental	677	short-term	702	sequential
628	gaussian	653	square	678	angular	703	amended
629	empty	654	underlying	679	rigid	704	express
630	genre-based	655	supplementary	680	warm	705	iranian
631	liquefied	656	recognized	681	segregated	706	unified
632	inflatable	657	assistant	682	realistic	707	unreasonable
633	reported	658	costly	683	vulnerable	708	mere
634	isolated	659	damaged	684	nearby	709	tightly-coupled
635	legitimate	660	secure	685	packaged	710	later
636	productive	661	steering	686	linear	711	oral
637	multi-lingual	662	multinational	687	enormous	712	sea-going
638	numerical	663	added	688	interim	713	medium
639	danish	664	aquatic	689	rough	714	nominal
640	outstanding	665	operable	690	adrift	715	discursive
641	insured	666	windward	691	evolutionary	716	conspicuous
642	unattended	667	ideal	692	korean	717	gps-only
643	outside	668	identical	693	line-of-sight	718	prescribed
644	established	669	cold	694	lexical	719	like
645	straight	670	stated	695	sheltered	720	imo-class
646	assumed	671	bigger	696	encouraging	721	inadequate
647	weak	672	lesser	697	german	722	thin
648	very	673	airborne	698	smooth	723	diluted
649	extraordinary	674	vocational	699	urgent	724	subjective
650	connected	675	periodic	700	consolidated	725	round

726	mediterranean	751	widespread
727	diverse	752	accidental
728	seaborne	753	marked
729	inferior	754	turkish
730	pure	755	surrounding
731	listening	756	approaching
732	anti-piracy	757	continental
733	participating	758	unusual
734	exclusive	759	imminent
735	interior	760	timely
736	raw	761	named
737	organizational	762	anticipated
738	massive	763	metric
739	governmental	764	smart
740	legislative	765	alert
741	state-owned	766	reckless
742	approximate	767	standardized
743	governing	768	weekly
744	self-propelled	769	verbal
745	psychological	770	circular
746	passing	771	recreational
747	ranging	772	floodable
748	artificial	773	invasive
749	pertinent	774	prominent
750	exposed	775	inclined



<부록 9>

누적빈도 백분율 80%까지의 부사 목록

1	also	26	both	51	generally	76	course	101	readily	126	primarily
2	in	27	about	52	i.e.	77	otherwise	102	furthermore	127	aft
3	more	28	either	53	immediately	78	clearly	103	previously	128	upwards
4	on	29	even	54	am	79	particular	104	significantly	129	additionally
5	only	30	often	55	addition	80	around	105	ashore	130	closely
6	out	31	e.g.	56	especially	81	off	106	please	131	hence
7	so	32	down	57	effectively	82	easily	107	specifically	132	successfully
8	however	33	over	58	today	83	much	108	there	133	instance
9	up	34	still	59	back	84	et	109	safely	134	nearly
10	board	35	always	60	later	85	automatically	110	better	135	away
11	then	36	example	61	directly	86	finally	111	astern	136	completely
12	at	37	together	62	almost	87	sometimes	112	earlier	137	no
13	where	38	far	63	currently	88	recently	113	enough	138	considerably
14	how	39	above	64	ahead	89	mainly	114	quite	139	right
15	as	40	just	65	respectively	90	probably	115	ever	140	consequently
16	least	41	fully	66	again	91	quickly	116	overboard	141	potentially
17	very	42	further	67	normally	92	all	117	best	142	subsequently
18	when	43	below	68	once	93	never	118	environmentally	143	frequently
19	now	44	usually	69	to	94	by	119	rapidly	144	thereby
20	most	45	already	70	here	95	under	120	simply	145	accordingly
21	therefore	46	yet	71	forward	96	underway	121	long	146	actually
22	order	47	particularly	72	approximately	97	general	122	al	147	continuously
23	thus	48	etc	73	of	98	relatively	123	offshore	148	correctly
24	for	49	too	74	why	99	properly	124	widely	149	perhaps
25	well	50	less	75	etc.	100	highly	125	longer		

<부록 10>

전치사 목록(1-gram)

순위	전치사	빈도수	순위	전치사	빈도수
1	of	3,5258	26	except	257
2	in	18,153	27	among	255
3	to	10,077	28	off	239
4	for	9,726	29	outside	205
5	by	6,424	30	towards	146
6	on	6,288	31	regarding	126
7	with	5,790	32	like	122
8	from	4,266	33	concerning	120
9	at	3,988	34	despite	101
10	as	1,619	35	since	96
11	into	1,478	36	involving	93
12	under	1,177	37	considering	93
13	between	1,081	38	behind	84
14	after	924	39	beyond	75
15	over	751	40	onto	74
16	during	726	41	inside	73
17	about	724	42	beginning	62
18	following	715	43	barring	27
19	including	599	44	assuming	23
20	before	598	45	down	15
21	above	560	46	round	12
22	against	440	47	beneath	11
23	without	383	48	pending	6
24	below	361	49	succeeding	2
25	upon	277			

<부록 11>

전치사 목록(2-gram)

순위	전치사	빈도수
1	such as	676
2	due to	414
3	as to	331
4	up to	242
5	according to	200
6	out of	178
7	because of	116
8	prior to	108
9	apart from	37
10	ahead of	32
11	adjacent to	31
12	outside of	29
13	regardless of	26
14	as for	26
15	referring to	25
16	irrespective of	22
17	thanks to	4
18	subsequence to	2
19	save for	1
20	aside form	1
21	as per	1
22	along with	1

<부록 12>

전치사 목록(3-gram)

순위	전치사구	빈도수	순위	전치사구	빈도수
1	in accordance with	324	24	by way of	19
2	in addition to	151	25	in spite of	16
3	in respect of	96	26	in front of	14
4	in terms of	91	27	in contact with	11
5	in charge of	87	28	with reference to	9
6	in relation to	81	29	in light of	6
7	by means of	80	30	as well as	6
8	in case of	63	31	for lack of	5
9	with respect to	57	32	in lieu of	5
10	as part of	55	33	in place of	5
11	as far as	53	34	in comparison with	5
12	into force of	42	35	in return for	5
13	with regard to	42	36	in need of	4
14	in line with	40	37	on account of	4
15	in compliance with	38	38	by reference to	4
16	on behalf of	34	39	in exchange for	4
17	in conjunction with	28	40	in consequence of	3
18	by virtue of	27	41	in conformity with	2
19	on top of	27	42	for want of	1
20	in view of	25	43	in search of	1
21	in connection with	21	44	at variance with	1
22	in favour of	21	45	as distinct from	1
23	in excess of	20			