2017년 한국항해항만학회 추계학술대회
2017년 11월 2일(목)~3일(금) 부산관광공사 아르피나

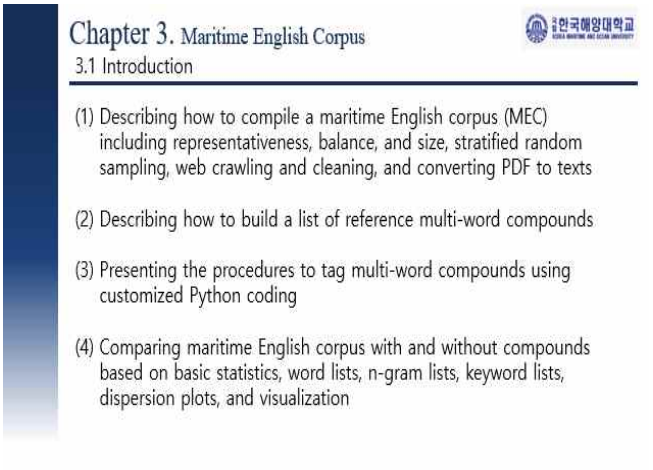
해사영어 전문용어에 관한 연구

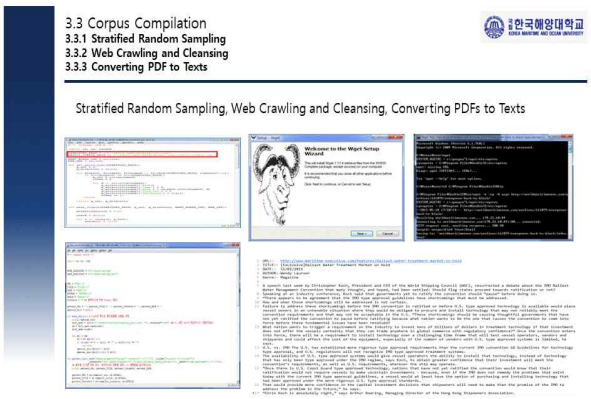
이성민*

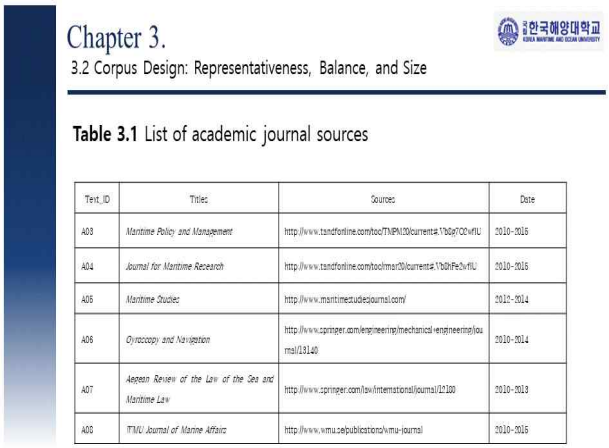
*한국해양대학교 강사

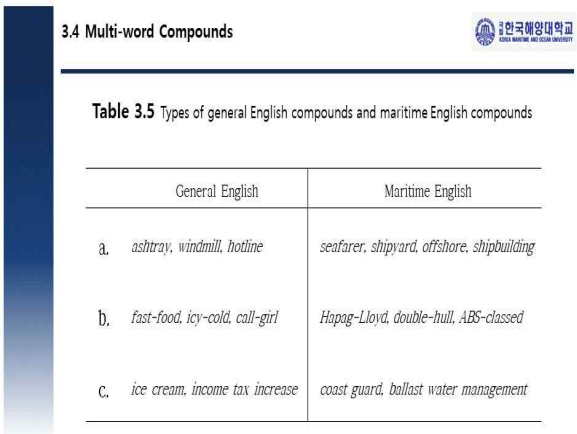
요 약 : 본 연구에서는 해사영어어휘의 특징인 b*allast water, fore peak bulkhead, container, freight station charges*와 같은 n-gram의 복수 단어로 구성된 합성어 (multi-word compounds) 태깅(tagging)처리가 포함된 해사영어코퍼스를 구축하였다. 해사영어코퍼스는 백만 단어씩 수집한 학술, 법, 신문, 교과서 4개 하위 코퍼스로 구성된 총 400만 단어의 해사영어코퍼스로 구성되어 있다.

핵심용어 : 해사영어, 코퍼스언어학, 합성어, 자연언어처리, 태깅









† 교신저자 : roy7942@hanmail.net

No	All the vocabulary items	No	List of single words	No	List of multi-word compounds
1	a1	1	a1	1	aboft the beam
2	aa	2	aa	2	aboncon chip
3	aak	3	aak	3	abonconed goocs
4	aalboot	4	aalboot	4	abonconment
5	ab	5	ab	5	abonconment clouse
6	abaok	6	abaok	6	abc analysis
7	abaft	7	abaft	7	abel brown
8	abaft the beam	8	abancon	8	abel tester
9	abancon	9	abanconment	9	able seaman
10	abancon chip	10	abatement	10	able-bodied seaman
11	abanconed goocs	11	abeam	11	about ship
12	abanconment	12	abiation	12	above-deck girder
13	abanconment clouse	13	able seaman	13	above-water hull
14	abatement	14	aboard	14	abscnce flag
15	abc analysis	15	abo-veoeck	15	absent flag
16	abeam	16	abov	16	absentee pennant
17	abel brown	17	abrasion	17	absolute accuracy
18	abel tester	18	abreast	18	absolute bearing
19	ablation	19	absorption	19	absolute contrabanc
20	able seaman	19	aburton	19	absolute pressure
.	.	.	.	.	.
16,736	zulu	6,553		6,147	zone price

Table 3.9
All the vocabulary items from three different sources

Table 3.10 Statistical results before tagging multi-word compounds

Genres	Token	Type	Type/Token Ratio (TTR)	Standardised TTR
Academy	989,943	79,639	8.47	56.17
News	997,835	39,790	4.12	55.03
Laws	985,621	15,500	1.63	30.05
Textbooks	989,610	53,617	5.64	54.11
Total	3,963,009	138,389	3.64	48.87

Fig. 3.6 Creation of multi-word compound tagged MEC

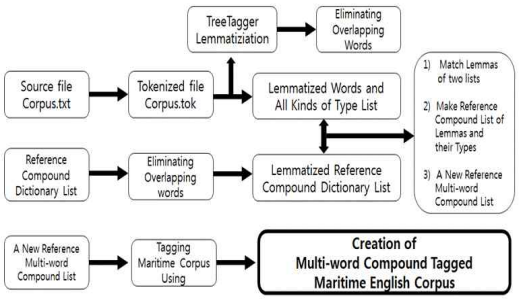


Table 3.11 Statistics results after tagging multi-word compounds

Genres	Token	Type	Type/Token Ratio (TTR)	Standardised TTR
Academy	985,284	80,317	8.59	56.31
News	983,350	42,792	4.50	55.74
Laws	971,471	16,464	1.76	30.53
Textbooks	984,180	54,416	5.76	54.34
Total	3,924,285	142,041	3.77	49.29

Fig. 3.7 Result of compound tagged special vocabulary terms

en_route means that the ship is underway at_sea on a course or courses, including deviation from the shortest direct_route, which, as far as practicable for navigation purposes, will cause any discharge to be spread over as great an area of the sea as is reasonable and practicable. except as provided in paragraph 2 of this regulation, in ships delivered after 31 December 1979, as defined in regulation 1.28.2, of 4,000 gross_tonnage and above other than oil_tankers, and in oil_tankers delivered after 31 December 1979, as defined in regulation 1.28.2, of 150 gross_tonnage and above, no ballast_water shall be carried in any oil fuel tank.
where the need to carry large quantities of oil fuel render it necessary to carry ballast_water which is not a clean_ballast in any oil fuel tank, such ballast_water shall be discharged to reception_facilities or into the sea in compliance with regulation 15 of this annex using the equipment specified in regulation 14.2 of this annex, and an entry shall be made in the oil_record_book to this effect.

Table 3.13 Comparison of untagged and tagged MEC word lists

Table 3.13 Comparison of the untagged and tagged MEC word lists

N	Untagged Law Corpus	Freq	N	Tagged Law Corpus	Freq
1	THE	74169	1	THE	74060
2	OF	46478	2	OF	45993
3	#	35184	3	#	35180
4	AND	28336	4	AND	28232
5	TO	26214	5	TO	26132
6	IN	22031	6	IN	21969
7	BE	17869	7	BE	17869
8	A	17350	8	A	17344
9	OR	16544	9	OR	16497
10	SHALL	16380	10	SHALL	16380
174	WITHOUT	692	174	BALLAST_WATER	686
769	PACKED	165	769	BULK_CARGOES	177
1,777	CARTRIDGE	54	1,777	CHIEF_ENGINEER	54
3,240	SIGNATORY	21	3,240	MOULDED_DEPTH	27
3,863	LEAKAGES	15	3,863	ABANDON_SHIP	15
.	.	.	.	.	.
.	.	.	.	.	.
15,381	YEMEN	1	16,433	YERSINIA	1

Table 3.15 Comparison of vocabulary on the keyword list

N	Untagged Law corpus	Keyness	N	Tagged Law corpus	Keyness
1	SHALL	35496.65	1	SHALL	35355.64
2	OR	15452.64	2	OR	15576.60
3	BE	10440.35	3	BE	10484.42
4	PARAGRAPH	4446.24	4	PARAGRAPH	4479.26
5	CONVENTION	4186.93	5	CONVENTION	4363.91
6	PROVISIONS	4127.62	6	PROVISIONS	4071.53
7	ANY	3826.46	7	ANY	3854.71
8	ACCORDANCE	3816.66	8	ACCORDANCE	3846.40
9	PROVIDED	3260.59	9	CERTIFICATE	3250.75
10	CERTIFICATE	3219.56	10	PROVIDED	3239.50
11	SPACES	3019.47	11	SPACES	3174.75
34	COMPETENT	1849.52	34	DANGEROUS_GOODS	1844.39
69	OTHER	1041.42	69	BALLAST_WATER	1048.02
424	VOYAGES	219.70	424	LIQUEFIED_GASES	222.15
.	.	.	.	.	.
1,968	ETHYLENE	24.04	2,183	RATING	24.05
1,969	ABOARD	-23.92	2,184	ELEMENT	-23.53
.	.	.	.	.	.
2,759	WAS	-3444.55	3,010	WAS	-3422.31

Table 3.14 Comparison of the untagged and tagged MEC 4-grams

N	Untagged Law Corpus	Freq	N	Tagged Law Corpus	Freq
1	IN ACCORDANCE WITH THE	803	1	IN ACCORDANCE WITH THE	803
2	WITH THE REQUIREMENTS OF	401	2	FOR THE PURPOSE OF	388
3	FOR THE PURPOSE OF	388	3	THE PROVISIONS OF THIS	374
4	THE PROVISIONS OF THIS	374	4	THE DATE ON WHICH	326
5	WITH THE PROVISIONS OF	362	5	CONSTRUCTED ON OR AFTER	316
6	THE DATE ON WHICH	325	6	IN THE CASE OF	307
7	CONSTRUCTED ON OR AFTER	315	7	WITH THE PROVISIONS OF	299
8	IN THE CASE OF	306	8	BY THE COMPETENT AUTHORITY	290
9	BY THE COMPETENT AUTHORITY	290	9	WITH THE REQUIREMENTS OF	269
10	TO THE PROVISIONS OF	286	10	ADOPTED BY THE ORGANIZATION	262
11	ADOPTED BY THE ORGANIZATION	262	.	.	.
.	.	.	156	TO THE PROVISIONS OF	78
271	TAKE INTO ACCOUNT THE	58	271	SHIPS BALLAST_WATER AND SEDIMENTS	33
1,859	A MEMBER OF THE	19	1,859	FIT TO PROCEED_TO_SEA WITHOUT	19
4,577	CAPABLE OF WITHSTANDING A	11	4,577	GROSS_TONNAGE CALCULATED IN ACCORDANCE	11
13,454	MODIFICATION OF A MAJOR	6	13,454	OF CHIEF_ENGINEER OFFICERS AND	6
16,981	THE HOLDER HAS ATTENDED	5	16,981	THE PORT_STATE AUTHORITY TONNAGE	5
.	.	.	.	.	.
20,715	ZONE IN AN AREA	5	20,665	YOU MAY RESUME THIS	5