

특수목적영어 교육을 위한 해사영어코퍼스 구축

이성민[○], 김재훈, 장세은
한국해양대학교

roy7942@hanmai.net, jhoon@kmou.ac.kr, jhang@kmou.ac.kr

A Compilation of Maritime English Corpus for English for Specific Purposes Education

Sung-Min Lee[○], Jae-Hoon Kim, Se-Eun Jhang
Korea Maritime and Ocean University

요 약

본 연구는 특수목적영어인 해사영어코퍼스의 구축을 목적으로 한다. 구축과정에서 코퍼스 구축에 필요한 대표성과 균형성을 고려하여 네 가지 장르인 학술, 뉴스, 법, 책으로 나누고 각 하위코퍼스를 백만 단어씩 구축하였다. 코퍼스 구축과정에서 웹사이트와 PDF형태의 자료에서 텍스트만을 수집하고 정제하기 위하여 파이썬(Python) 프로그래밍 코딩을 하였고 무료 공개 프로그램도 병행하였다. 앞으로 해사영어코퍼스는 해사영어어휘교육에 필요한 단어목록제공이나 예문 검색 등을 통한 자료중심학습법에 활용될 수 있을 것이다. 또한 본 연구의 코퍼스구축 과정은 다른 분야의 ESP코퍼스 구축에도 응용 될 수 있을 것이다.

주제어: ESP, Corpus Linguistics, Corpus Compiling, Maritime English Corpus, Python

1. 서론

글로벌전문인력양성의 필요성이 대두되면서 특수목적영어(English for Specific Purposes, 이하 ESP) 관련 연구와 교육은 갈수록 중요성을 더해가고 있다. ESP교육과 관련하여 Paltridge and Starfield(2013)는 말하기, 듣기, 읽기, 쓰기 및 어휘를 통합한 ESP 교육시스템이 필요하다고 주장하였다. 이러한 ESP 교육시스템의 구현을 위해서는 코퍼스언어학과 자연언어처리 방법론으로 구축된 특수목적영어코퍼스 DB를 토대로 수업모델을 개발이 필요한 시점이라고 본다. 따라서 본 연구에서는 ESP의 한 분야인 해사영어(maritime English)의 코퍼스 구축을 목적으로 한다.

국제 해양 커뮤니티간의 의사소통을 위한 수단으로 사용되는 해사영어는 해사분야의 국제 언어로서 항해의 안전과 해상무역을 기여하기 위하여 사용되는 언어를 말하며 언어의 특수성 때문에 ESP의 한 분야로 분류되어 있고 그 중요성은 심화되고 있는 추세이다. 국제해사기구(International Maritime Organization, 이하 IMO)에서는 해상운송의 안전과 효율성을 위하여 영어로 작성된 표준해사통신용어(Standard Marine Communications Phrases, 이하 SMCP)를 제시하며 모든 선박운항관련 종사자들이 영어로 의사소통을 하도록 강제하고 있다. 이처럼 해사영어는 산업적으로 매우 중요한 분야이므로 자연언어처리를 활용한 체계적인 방법으로 코퍼스를 구축하여 연구할 필요가 있다고 본다.

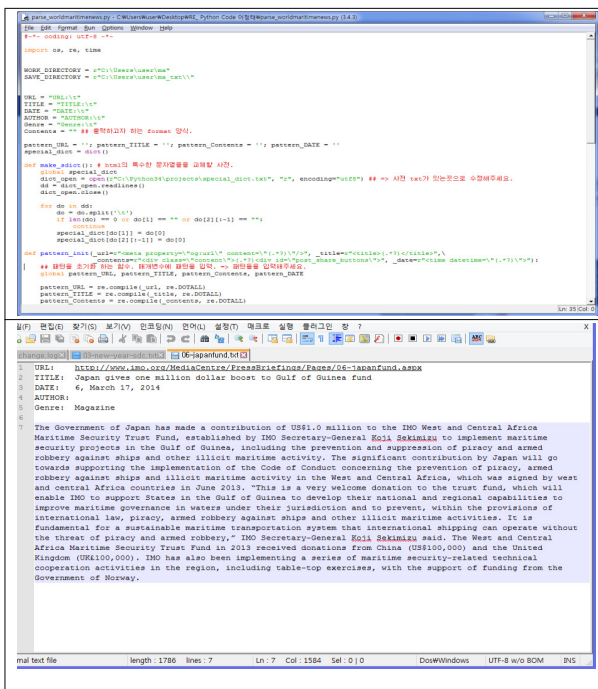
2. 관련 연구

특수목적영어인 해사영어로 코퍼스를 구축한 관련연구

로는 외국의 사례는 아직 나오지 않았으며 국내에서는 한국해양대학교를 중심으로 활발하게 진행하여 왔다. Hong and Jhang(2010)의 소규모 해사영어코퍼스 구축연구, 장세은·변현정(2011)의 코퍼스를 활용한 해사영어어휘 분석, Hong(2012)의 해사영어의 N-gram분석연구, 장세은·이성민(2013)의 코퍼스기반 고등학교 해사영어교과서의 어휘비교분석, Jhang and Lee(2013)의 해사영어키워드의 언어네트워크 시각화, Jhang and Lee(2013)의 해사영어의 클러스터와 키클러스터의 의미분석 연구, Ryoo(2013)의 해사영어 어휘의 언어의 어휘문법연구, 장세은·이성민(2014)의 고등학교 해사영어 교과서와 해기사 해사영어시험 간의 상관관계 분석, 장세은·이성민(2014)의 해사영어 어휘의 통계적 언어 유형 등이 있다.

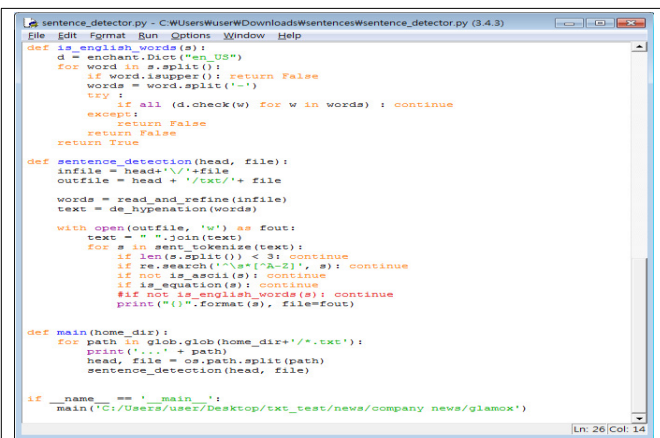
3. 연구 방법

코퍼스 구축을 위해서 온라인에서 데이터의 수집과 정제작업을 수행하였다. 웹 크롤러인 Wget(<http://www.gnu.org/software/wget/>)으로 원하는 데이터를 수집하고 추출한 파일에서 영어문장을 추출하는 파이썬(Python) 코드로 텍스트만을 추출하되 주소(URL), 제목(TITLE), 날짜 DATE), 저자(AUTHOR), 장르(GENRE)에 관한 정보를 별도로 저장하였다.



<그림1> 웹 텍스트 추출 및 정제작업

학술논문을 수집하기 위해서는 Springer의 Springer e-journal을 활용하여 해사산업분야의 관련 저널을 찾아 수록된 모든 논문을 수집하였다. 학술논문은 모두 이동식문서형식(Portable Document Format, 이하 PDF)로 저장하고 그 후 nPDF (<http://iblogbox.com/npdf/>)를 사용하여 텍스트만을 저장하고 오류를 제거하기 위하여 파이썬 코딩으로 정제작업용 프로그램을 작성하여 텍스트 정제작업을 실시하였다.



<그림2>. PDF 텍스트의 정제작업 파이썬 코딩

4. 결론

본 논문에서는 국제적으로 중요하게 다루어지는 ESP의 한 분야인 해사영어를 대상으로 해사영어코퍼스를 구축하였다. 코퍼스의 대표성을 위하여 전문가의 자문을 받

아 텍스트소스를 선정하였고 장르의 균형을 맞추기 위하여 학술, 뉴스, 법, 책으로 나누었다. 웹사이트와 학술 논문에서 자료를 추출하기 위하여 Wget, nPDF와 같은 무료 공개 프로그램을 사용하였고 파이썬 코드를 사용하여 텍스트추출과 편집 그리고 정제작업을 실시하였다. 구축이 완료된 해사영어코퍼스의 크기는 장르별로 백만 단어씩 총 4백만 단어이다. 본 연구의 결과물은 추후 콘코던스 프로그램이 개발된다면 해사영어코퍼스를 활용한 자료중심학습법이 가능하여 학습의 실효성을 증가시킬 수 있으며 학습자가 자기주도적인 학습법을 통해서 학습의 효과를 향상시키는데 사용될 수 있다.

참고문헌

- [1] B. Paltridge and S. Starfield, The Handbook of English for Specific Purposes, West Sussex: Wiley-Blackwell, 2013.
- [2] S. C. Hong, and S. E. Jhang, The Compilation of a Maritime English Corpus for ESP Learners, Korean Journal of English Language and Linguistics, 10, 4, pp.963-985, 2010
- [3] 장세은 · 변현정, “코퍼스를 활용한 해사영어 어휘 분석”, 새한영어영문학회, 제53권, 제4호, pp. 247-266, 2011
- [4] S. C. Hong, An N-gram Analysis of Maritime English, The Journal of Linguistics Science, 61, 283-302, 2012.
- [5] 장세은 · 이성민, “고등학교 해사영어교과서의 코퍼스기반 어휘 분석”, 언어과학, 제20권, 201호, pp. 165-183, 2013.
- [6] S. E. Jhang and S. M. Lee, Visualization of Collocational Networks: Maritime English Keywords, Language Research, 49, 3, pp.781-802. 2913.
- [7] S. E. Jhang and S. M. Lee, Clusters and Key Clusters in the Maritime English Corpus, Journal of Language Sciences, 20, 4, pp.199-219, 2013.
- [8] M. L. Ryoo, Lexico-semantic Associations in Maritime English as ESP: A Corpus-based Study, English Language Teaching, 25.4, pp.107-128, 2013.
- [9] 장세은 · 이성민, “고등학교 해사영어 교과서와 해기사 해사영어시험 간의 상관관계 분석”, 새한영어영문학회, 제56권, 제2호, pp.151-167, 2014.
- [10] 장세은 · 이성민, “해사영어 어휘의 통계적 언어 유형”, 언어과학, 제21권, 제1호, pp.173-193, 2014.