

『Asia Pacific Journal of Corpus Research』 1 권 1 호의 연구 동향과 연구 방법에 관한 고찰

정채관

(인천대학교/코퍼스연구소)

Jung, C. K. (2020). A review of emerging research trends within Asia Pacific Journal of Corpus Research (Vol. 1, No. 1) and some insights on research issues. *Asia Pacific Journal of Corpus Research*, 1(1), 127-132.

The purpose of this review is to provide local readers, more specifically, Korean student readers who are not all that familiar with the English language a general overview of research articles that have been published in Asia Pacific Journal of Corpus Research vol. 1, no. 1. A brief summary of each research article focusing on research methods and then followed by an overall review and some insights on research issues will be presented.

Keywords: Corpus Linguistics, Covid-19, Singapore English, CEFR-J Vocabulary, Social Media Discourse

아시아 아시아 태평양 지역을 중심으로 하는 코퍼스 분야 전문 국제학술지 『아시아태평양코퍼스연구』 제 1 권 1 호에 논문이 엄선되어 출간되었다. 저자의 소속 기관이 위치한 나라는 말레이시아, 싱가포르, 일본, 영국이다. 우선 Joharry와 Turiman(2020)는 코로나 19 로 인해 말레이시아에서 이동제한이 이뤄진 3 개월 동안 말레이시아 전국 단위 신문 편집자에게 보내진 온라인 편지 303 개를 코퍼스로 구축하고, #LancsBox v.5.x(Brezina, Weill-Tessier & McEnery, 2020)을 사용하여 코로나 19 와 연관된 언어 사용 패턴을 분석하고, 언어 네트워크를 시각화하였다. 세계적으로 언론이나 소셜미디어 네트워크에 코로나 19 상황에 관한 다양한 반응이 쏟아져 나오고 있는 상황에서 최근 말레이시아의 분위기를 가늠할 수 있는 의미 있는 연구라고 할 수 있다. 이 연구에서 사용된 영국 랭커스터대학에서 개발한 #LancsBox v.5.x는 아직 국내에 많이 알려지지 않은 언어 데이터 분석 도구이다. 이 도구는 언어 데이터를 시각화하고, 다양한 언어를 분석할 수 있으며, 자동으로 품사 태깅이 되고, 윈도우, 매킨토시, 리눅스 등 다양한 운영 체제에서 사용할 수 있는 것으로 알려져 있다. 기존 코퍼스 연구에서 자주 사용하던 AntConc 3.5.8(Anthony, 2019)나 WordSmith Tools 8(Scott, 2020)와 성능 비교도 흥미로울 것으로 보인다.

Lee(2020)은 1993 년부터 2016 년까지 싱가포르에서 발행된 싱가포르 영자 신문을 시간 추이에 따라 사용 언어가 어떻게 변화했는지 변천 과정을 살펴보았다. Lee(2020)은 기존에 보고된 Schneider(2007)의 탈식민지 영어 동적 모델과 싱가포르식 영어가 점차 안정화되어 가는 시기에서의 연구와 비교하여 영국식 영어에서 싱가포르식 영어로의 진화적 상황을 고찰하였다. 이 연구에서도 역시 국내 코퍼스 연구자들에게는 잘 알려지지 않은 Sketch Engine(Kilgarriff, Baisa, Busta, Jakubicek, Kovar, Michelfeit, Rychly & Suchomel, 2014)을 사용하였는데, Sketch Engine은 2015 년 작고한 Adam Kaligarriff가 2003 년에 개발하여 보급하기 시작한 웹 기반 코퍼스 관리 및 텍스트 분석 도구이다. Sketch Engine은 유료이긴 하나, 세계 각국의 연구자들이 사용하는 520 개에 달하는 참고코퍼스가 기본으로 탑재되어 있어서 참고코퍼스와 개인 연구자가 소규모로 구축한 연구코퍼스와 비교 연구도 가능한 유용한 도구이다.¹

Ohashi와 Katagiri(2020)는 Tono(2012)가 유럽언어공통기준(Common European Framework of Reference for Language, CEFR)를 일본인 영어학습자 상황에 맞춰 개발한 CEFR-J와 West(1953)이 제안한

¹ Sketch Engine 에서 보유하고 있는 참고코퍼스는 다음 링크를 참고한다.
<https://www.sketchengine.eu/corpora-and-languages/corpus-list>

일반영어어휘목록(General Service List of English(GSL), Coxhead (2000)가 제안한 학술영어어휘목록(Academic Word List, AWL)을 사용하여 일본 초등학교 영어 교실에서 사용되는 어휘 사용 양상 실태를 분석하였다. 이들은 초등학교 교사들이 학술이 더 종합적이고 체계적인 영어 학습을 할 수 있도록 CEFR-J 레벨 1에 속한 더 다양한 어휘를 더 많이 사용해야 한다고 주장한다. 이 연구에서 주목할 점은 사실 이 연구가 가능하도록 바탕이 된 코퍼스 기반 일본인 영어학습자 코퍼스 분석 연구이다(Tono, 2012). 우리나라는 일본과 달리 현재까지 국가 주도로 우리나라의 다양한 연령대와 학년 군을 분류하여 이들의 특성이 반영되어 구축된 전국 단위 한국인 영어학습자 코퍼스가 없다. 따라서 우리나라는 미국의 민간 업체인 교육시험서비스(Educational Testing Service, ETS)가 개발한 토익(Test of English for International Communication, TOEIC)으로 전 국민의 영어 능력이 가늠되고 있다. 외화 낭비 문제를 떠나 어느 정도가 우리나라 사람의 영어 능력 정도를 나타내는지 알 수 있는 기준 마련이 필요하며, 기초 자료 수집을 위해서는 무엇보다 국가 차원의 전국 단위 우리나라 영어학습자 구어와 문어 코퍼스 구축이 시급해 보인다.

Solomon(2020)은 기존의 코퍼스 연구에서 한 걸음 더 나아가 언어 사용 패턴을 기반으로 인공지능과 챗봇과 같은 곳에서 코퍼스가 어떻게 활용될 수 있는지에 관한 연구를 제안한다. 4차 산업혁명(정채관, 안계명, 홍선호, 이완기, 심창용, 이재희, 김해동, 김명희, 김선웅, 2018)의 도래와 이 시기에 강조되는 빅데이터, 그리고 빅데이터를 사용 목적사용에 맞춰 고도로 정제된 코퍼스는 현재는 물론 미래의 코퍼스 연구가 나아가야 할 방향이 분명하다. 현재 우리나라는 국립국어원을 중심으로 4차 산업혁명 대비 국어 빅데이터 구축사업이 한창이다. 일상대화 말뭉치를 모아 고품질 AI 데이터 코퍼스로 정제하여, 이를 대화형 인공지능 개발을 위한 자료로 활용하겠다는 것이다. 영어가 이미 국제공용어로서 세계인과 소통하는 대표적인 언어라는 점을 고려하면, 국어 빅데이터 구축사업뿐 아니라 우리나라 사람들이 사용하는 영어 빅데이터를 구축하여, 우리나라 사람들의 영어 역량을 개선할 수 있는 기초 자료를 확보하고, 이를 토대로 대화형 인공지능을 개발하는 것도 국가 경쟁력 강화를 위해 도움이 될 것으로 보인다.

이상에서 『아시아태평양코퍼스연구』 1 권 1 호에 출간된 논문을 살펴 보았다. 국내에서 출간된 논문과 두드러지게 다른 점은 연구자들이 적극적으로 유·무료 웹 기반 코퍼스 분석 도구를 사용하여 연구한다는 것이다. 국내에서는 데스크톱 전용인 무료 AntConc 계열이나 유료 WordSmith Tools 계열을 코퍼스 분석 도구로 자주 사용되는 경향이 있다. AntConc 계열이 무료라는 점과 WordSmith Tools 계열보다 상대적으로 다루기 쉽다는 점 때문

에 국내 코퍼스 연구자에게 자주 사용하는 도구이다(고광윤, 2019; 권혁승, 정채관, 2012). 인터넷 연결이 필수이고 개발자에 의해 이미 정해진 기능에서 사용해야 하는 웹 기반 코퍼스 분석 도구보다는 인터넷 연결이 필요 없고, 개인 연구자가 보고자 하는 더 다양한 부분을 분석할 수 있는 점은 분명히 데스크톱 코퍼스 분석 도구가 가진 유용성이라고 할 수 있을 것이다(권혁승, 정채관, 김재훈, 2018). 웹 기반 Sketch Engine은 유료이긴 하지만 연구자가 개별적으로 구해야 하는 브라운코퍼스(Brown Corpus, BC), 영국국립코퍼스(British National Corpus, BNC), 2014 영국국립구어코퍼스(British National Corpus 2014 Spoken, Spoken BNC2014)와 같은 브랜드 참고 코퍼스를 비롯하여, 영국학술구어영어(British Academic Spoken English, BASE), 영국학술문어영어(British Academic Written English, BAWE), 케임브리지학술영어(Cambridge Academic English, CAE), 학술논문코퍼스(Corpus of Academic Journal Articles, CAJA), 의료웹코퍼스(Medical Web Corpus, MWC) 등과 같은 특수코퍼스가 기본 탑재되어 있다는 점을 고려하면, 국내에서도 이러한 자원을 바탕으로 연구할 부분이 상당히 많을 것으로 예상된다. 또한, AntConc 계열이나 WordSmith Tools 계열과 같이 기존의 데스크톱 코퍼스 분석 도구로는 할 수 없고, 다른 도구를 사용해야 하는 품사 태깅 같은 추가적인 작업을 한가지 도구에서 수행할 수 있는 무료 웹 기반 #LancsBox v.5.x와 같은 새로운 분석 도구의 활용도 앞으로는 국내 코퍼스 연구자들이 조금 더 적극적으로 시도해 볼 필요가 있어 보인다.

참고문헌

- 고광윤(Goh, G. Y.). (2019). *안트콩크 북(The AntConc Book)*. 의왕: 래러비러닝(Uiwang: Larrabee Learning).
- 권혁승, 정채관(Kwon, H. S., & Jung, C. K.). (2012). *코퍼스 언어학 입문(Corpus Linguistics Introduction)*. 서울: 한국문화사(Seoul: Hankook Publishing House)
- 권혁승, 정채관, 김재훈(Kwon, H. S., Jung, C. K., & Kim, J. H.). (2018). *코퍼스 언어학 기초(Corpus Linguistics Basic)*. 서울: 한국문화사(Seoul: Hankook Publishing House).
- 정채관, 안계명, 홍선호, 이완기, 심창용, 이재희, 김해동, 김명희, 김선웅(Jung, C. K., Ahn, K. M., Hong, S. H., Lee, W. K., Sim, C. Y., Lee, J. H., Kim, H. D., Kim, M. H., & Kim, S. W.) (2018). 4 차

산업혁명과 미래 영어교육(*The 4th Industrial Revolution and English Education in the Future*). 서울: 한국문화사(Seoul: Hankook Publishing House).

- Anthony, L. (2019). AntConc 3.5.8. [software]. Available at: <https://www.laurenceanthony.net/software/antconc>.
- Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 5.x. [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Joharry, S. A., & Turiman, S. (2020). Collocation networks and COVID-19 in letters to the editor: A Malaysian case study. *Asia Pacific Journal of Corpus Research*, 1(1), 1-30.
- Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfeit, J., Rychly, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1, 7-36.
- Lee, K. M. (2020). Relative clauses in a modern diachronic Corpus of Singapore English. *Asia Pacific Journal of Corpus Research*, 1(1), 31-60.
- Ohashi, Y., & Katagiri, N. (2020). The ratios of CEFR-J vocabulary usage compared with GSL and AWL in elementary EFL classrooms and suggestions of vocabulary items to be taught. *Asia Pacific Journal of Corpus Research*, 1(1), 61-94.
- Schneider, E. W. (2007). *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Scott, M. (2020). WordSmith Tools 8. [software]. Available at: <https://lexically.net/wordsmith>.
- Solomon, B. S. (2020). A novel theory of support in social media discourse. *Asia Pacific Journal of Corpus Research*, 1(1), 95-125.
- Tono, Y. (2012). Eigo totatsudo shihyo CEFR-J guidebook (*The CEFR-J Handbook. A Resource Book for Using CAN-DO Descriptors for English Language Teaching*). Tokyo: Taishukan.
- West, M. (1953). *A General Service List of English Words*. London: Longman.

THE AUTHOR

Chae Kwan Jung is an Assistant Professor in the Department of English Language & Literature and Head of Institute for Corpus Research (www.icr.or.kr) at College of Humanities, Incheon National University. He studied and conducted research in the field of Engineering, Effective Online Tutoring, Applied Linguistics and English Language Teaching at the universities of Birmingham, Oxford, and Warwick in the UK respectively. He is the co-author of *Corpus Linguistics Introduction* (2012), *Corpus Linguistics Basic* (2018), and the 4th Industrial Revolution and English Education in the Future.

THE AUTHOR'S ADDRESS**First and Corresponding Author****Chae Kwan Jung**

Professor

Department of English Language & Literature/Institute for Corpus Research

Incheon National University

119 Academy-ro, Yeonsu-gu, Incheon, 22012, SOUTH KOREA

E-mail: ckjung@inu.ac.kr

Received: 10 August 2020

Received in Revised Form: 15 August 2020

Accepted: 20 August 2020