1. 서론

이 연구는 약 500만 어절 규모의 분야별 한국어 논문 자료를 수집, 전산화하여 분석하는 것을 목표로 한다. 분석도구로는 ‘AntConc’의 concordance, N-grams, word list 등이 사용되며 시각화를 위해 ‘Pajek’을 사용하였다. 연구의 목적은 격식적, 학문적 글쓰기의 정점에 있는 논문의 어휘, 문장구조를 분석하고 각 분야별 논문의 글쓰기 방식의 차이를 확인하는 것이다.

논문의 언어 사용 양상 분석을 통해 얻을 수 있는 효과는 크게 외국어 화자에 대한 것과 모국어 화자에 대한 것으로 나눌 수 있다. 먼저 외국어로서 한국어를 배우는 학습자에게 있어서 학문적 텍스트를 읽고 쓰는 것은 매우 어려운 일이다. 이러한 자료들이 어렵게 느껴지는 이유는 사용되는 어휘와 표현이 일상적인 것과 크게 다르기 때문일 것이다. 본 연구를 통해 전산화된 자료는 학문적 텍스트에 대한 실제적 교육 자료가 될 수 있을 것이다. 모국어 화자에게 있어서도 논문은 어려운 텍스트에 속하는데, 이는 다른 담화 유형에 비해 우리가 논문을 비교적 늦게 접하기 때문일 것이다. 모국어 화자는 연구 결과를 토대로 다른 격식적인 텍스트나 학문적인 텍스트와 비교했을 때 논문이 가지는 특징이 무엇인지 세밀하게 분석해보고 이를 논문의 생산과 수용에 활용할 수 있을 것이다.

이를 위해 연구에서는 구체적으로 분야별 논문의 특징적인 단어나 문장 구조를 추출하고 비교해 볼 것이다. 분야별 글쓰기 방식에 차이가 있다는(또는 그래야만 한다는) 주장은 오래전부터 논의 되었으며 특히 이공계열, 인문계열 간의 차이가 중요한 것으로 여겨지고 있다. 이에 따라 실제로 많은 글쓰기 수업에서 전공 분야에 맞는 글쓰기 수업을 진행하고 있기도 하다. 본 연구에서는 이러한 차이가 실제로 존재하는지, 그렇다면 기존의 통념과 일치하는지 알아보고자 한다. 이를 통해 분야별 특성에 맞는 글쓰기 수업 방식이 강화되거나 보완되는 효과를 거둘 수 있을 것이다. 이와 더불어 본 연구에서는 분야에 관계 없이 논문 텍스트 전체를 관통하는 특징은 무엇인지 분석해 볼 것이다. 논문 글쓰기의 특징으로 피동형의 잦은 사용 등이 언급되곤 하는데 이러한 일반적인 인식이 실제 자료와 일치하는지 확인해보고 새롭게 도출해낼 수 있는 전반적인 특징이 있는지 살펴볼 것이다.