

심리학적 언어분석 프로그램 개발을 위한 융합연구: 기존 프로그램의 비교와 관련 문헌의 동향 분석

김영준¹, 최원일², 김태훈^{3*}

¹아주대학교 심리학과 강사, ²광주과학기술원 기초교육학부 교수, ³경남대학교 심리학과 교수

A Convergence Study for Development of Psychological Language Analysis Program: Comparison of Existing Programs and Trend Analysis of Related Literature

Youngjun Kim¹, Wonil Choi², Tae Hoon Kim^{3*}

¹Instructor, Division of Psychology, Ajou University

²Professor, Division of Liberal Arts and Sciences, GIST College

³Professor, Division of Psychology, Kyungnam University

요약 내용어 기반 빈도 분석은 의도적 기만이나 반어적 표현에 분명한 한계가 있지만, 많이 사용되는 한국어 분석 프로그램인 KLIWC는 기능어 분석을, KrKwic는 동시출현빈도를 시각화하는 방법으로 발전했다. 하지만 개발된 지 십수 년이 지나 여러 문제점으로 개선이 필요한 상황이다. 그래서 KLIWC와 KrKwic를 분석하여 새 심리학적 언어분석 프로그램을 개발하고자 하였다. 첫째로 두 프로그램의 특징을 분석하였다. 특히, 기능어 분석기능 제고를 위해서 KLIWC와 한국어 형태소 분석기의 형태소 분류를 비교하였고, 심리적 분석의 강화를 위해 심리사전의 구조와 체계를 분석하였다. 분석 결과 한나눔 품사 분석기가 가장 세분화되었지만, 인칭대명사에서는 KLIWC가, 어미와 어말어미에서는 KKMA의 품사 분류가 더 세분화되어 있어, 기능어 분석 강화를 위해 여러 품사 분석기의 통합적 사용을 제안하였다. 둘째로 이 프로그램들로 텍스트를 분석한 연구들의 연구동향을 분석하였다. 분석 결과 두 프로그램이 복합학 분야 등 다양한 학술분야에서 사용되고 있었다. 특히 논문과 보고서의 분석에는 KrKwic가 많이 사용되었고, 글쓴이의 생각, 정서, 성격 비교 연구에는 KLIWC가 많이 사용되었다. 이 결과를 바탕으로 새로운 심리학적 언어분석 프로그램의 필요성과 개발 방향에 대해 제안하였다.

주제어 : 한국어 분석 프로그램, 심리사전, 기능어, KLIWC, KrKwic

Abstract While content word-based frequency analysis has obvious limitations to intentional deception or irony, KLIWC has evolved into functional word analysis and KrKwic has evolved as a way to visualize co-occurrence frequencies. However, after more than 10 years of development, several issues still need improvement. Therefore, we tried to develop a new psychological language analysis program by analyzing KLIWC and KrKwic. First, the two programs were analyzed. In particular, the morpheme classification of KLIWC and the Korean morpheme analyzer was compared to enhance the functional word analysis function, and the psychological dictionary were analyzed to strengthen the psychological analysis. As a result of the analysis, the Hannanum part-of-speech analyzer was the most subdivided, but KLIWC for personal pronouns and KKMA for endings and endings were more subdivided, suggesting the integrated use of multiple part-of-speech analyzers to strengthen functional word analysis. Second, the research trends of studies that analyzed texts with these programs were analyzed. As a result of the analysis, the two programs were used in various academic fields, including the field of Interdisciplinary Studies. In particular, KrKwic was used a lot for the analysis of papers and reports, and KLIWC was used a lot for the comparative study of the writer's thoughts, emotions, and personality. Based on these results, the necessity and direction of development of a new psychological language analysis program were suggested.

Key Words : Korean language analysis program, Psychological Dictionary, Functional words, KLIWC, KrKwic

*This work was supported by the Ministry of Education of the Republic of Korea (NRF-2020S1A3A2A02103899).

*Corresponding Author : Tae Hoon Kim(taehoonk@kyungnam.ac.kr)

Received August 6, 2021

Revised November 4, 2021

Accepted November 20, 2021

Published November 28, 2021

1. 서론

언어는 사람들이 자신의 심리와 그 상태를 다른 사람에게 전달할 수 있는 가장 보편적인 방법이다. 기억, 개념적 사고, 문제해결, 추리 등의 인지 과정 밑바탕에 언어적 표상 체계가 자리 잡고 있기 때문에, 언어 연구는 인지 구조와 과정의 본질을 탐구하는 인지심리학의 핵심적인 연구 주제이다[1].

대화에서 화자가 자신을 표현하고 청자가 다른 사람을 이해할 수 있는 이유는 문법이 있어 가능한데, 놉 촘스키 [2,3]의 생성문법 이론을 중심으로 하는 학자들의 보편 문법과 언어 사용 경험을 중시하는 인지문법[4-6] 모두 주고 받는 말과 글에 어떠한 규칙이 있어야만 한다는 공통점을 가지고 있다. 특히 의미의 기본 단위인 형태소와 단어를 이해하는 과정에 대한 심성어휘집(lexicon) 접속 과정에 대한 연구[7,8]는 글분석 프로그램의 개발에 중요하다. General Inquirer, TAS/C 등 초기 글분석 프로그램은 언어 의미의 기본 단위인 단어 수준에서 주로 이루어져 왔는데, 정서 단어나 구체적 혹은 추상적 처리 과정을 나타내는 단어의 수를 세는 방식이었다[9,10]. 이러한 연구는 내용분석, 텍스트 마이닝 등 내용어 탐지에 기초한다. 하지만, 내용어 기반 탐지 방식의 분석은 의도적으로 숨기거나 반대 의미의 단어를 사용하는 것과 같은 기만적 상황에서 분명한 한계를 보인다는 점에서 제한점을 가지고 있다.

내용어 중심의 단순 계산 방식을 넘어서려는 시도 중 하나인 글분석 프로그램인 LIWC(Linguistic Inquiry and Word Count)는 말과 글이 개인의 심리적 특성을 나타낸다는 연구 결과를 바탕으로 개발된 언어분석 프로그램이다[11-13]. 일상생활이나 실험적 상황에서 쓰여진 글에 포함된 단어가 언어학적 변인과 심리학적 변인의 각각에 포함된 빈도와 그 비율을 계산하는 단순한 방식이다. 하지만 초기의 글분석 프로그램이 내용어(content word) 중심의 분석이었던 것과는 다르게, LIWC는 단어 길이, 전치사, 문장 당 단어 수, 인칭대명사 등 언어 구조(Linguistic structure)를 만드는데 사용되는 기능어(functional word), 즉 통사적인 변인의 분석도 포함하고 있다[11]. LIWC처럼 기능어 분석이 가능한 심리학적 한국어 글분석 프로그램인 KLIWC는 LIWC에 한국어의 특성과 문화를 반영하여 개발되었다[14].

기능어 분석은 기존의 내용어 분석의 한계를 흥미롭게 보여준다. Slatcher 등의 연구[15]에서는 2004년 미국 대통령 후보자, John Kerry, John Edwards, George W. Bush, Dick Cheney의 언어를 분석하였다. 이중 Cheney는 가장 지적인 언어를 구사했고, Edwards는 가장 여성적 언어를 사용했으며, Bush는 가장 나이 든 사람처럼 말했고, Kerry의 언어는 우울증에 걸린 사람의 언어에 가장 가까웠다. 이런 흥미로운 결과는 고빈도 단어 중심의 내용분석이 아니라, LIWC를 사용하여 언어 차원의 기능어와 심리적 상태를 나타내는 내용어의 비중을 계산하는 방식으로 도출되었다. 예를 들어, 지적으로 말하는 사람의 언어에는 인지적 단어(예, think, because)의 사용 비율이 더 높게 나타났다[12,16,17]. 구체적으로 배제 단어(but, except 등), 잠정적 표현(maybe, perhaps 등), 부정어(no, never 등), 불일치(should, would 등) 단어를 더 많이 사용하고 포함 단어(with, and 등)는 더 적게 사용하였다. LIWC와 KLIWC의 기능어 분석은 대통령 연설문의 심리적 특성 [14,18], 성차[19,20], 거짓말[16,21], 인터넷 댓글 [22,23] 등 다양한 분야에 적용되어왔다.

2001년 개발 배포된 LIWC는 2007년과 2015년에 사전을 업그레이드했다. 하지만, KLIWC는 개발된 지 15년 이상이 지난 현재까지 사전의 업그레이드가 이루어지지 않아서, 시대의 변화에 따른 언어 사용의 변화를 반영하지 못하고 있다. 그리고 그 기간 한국어 형태소 분석 기술의 발달로 분석 속도와 정확성에서 많은 발전이 있었기 때문에, 이를 한국어 분석 프로그램에 반영할 필요도 분명해 보인다.

사회관계망 분석 방법은 기존의 빈도 기반 내용어 분석의 한계에 다른 방식으로 접근할 수 있게 했다. 한국어 분석 프로그램인 KrKwic는 사회관계망 분석 접근을 통해 핵심어 간의 관계를 의미 있게 설명하고 시각화하고 있다. KrKwic는 사회 관계망 분석 알고리즘에 기반해 개발된 영어권 언어분석 프로그램인 FullText를 한국어 텍스트 분석을 위해 변형된 언어분석 프로그램이다[24]. 이 프로그램은 어떤 특정한 단어들어 어떤 다른 특정한 단어들과 함께 사용될 때, 그 메시지의 의미가 달라지는 것에 주목하여, 동시 출현 빈도를 분석하여 시각화하는 절차를 포함하고 있다. 하지만 도스 환경에서 구동되고, 시각화에 별도 프로그램을 사용해야 하는

문제점이 있다.

따라서 본 연구는 새로운 한국어 분석 프로그램을 개발하기 위해, 특성이 전혀 다른 두 한국어 분석 프로그램을 분석할 필요가 있다. 본 연구는 KLIWC와 KrKwic의 장단점을 비교하였다. 특히, 한국어 형태소 분석과 구문 분석 등 자연어 처리 기술을 도입하여 프로그램의 완성도를 높이는 방안을 고안하였다. 또한 두 한국어 분석프로그램을 사용한 국내 연구를 연도별 사용빈도, 피인용 횟수, 학술 연구분야, 분석 텍스트의 종류 별로 비교하였다. 이 결과를 바탕으로 새로운 한국어 분석 프로그램 개발 방향을 제안하고자 하였다.

2. 이론적 배경

2.1 한국어 분석 프로그램

한국어 분석 프로그램 개발을 위해 기존의 한글 텍스트를 분석할 수 있는 프로그램을 살펴봐야 할 필요성이 있다. 한글 텍스트 분석에 많이 사용되는 대표적인 프로그램인, KLIWC와 KrKwic를 살펴보았다.

2.1.1 KLIWC

광범위하게 사용되는 한국어 분석 프로그램인 KLIWC의 개발 과정을 고찰한 연구[14]를 보면, 내용이 분석뿐 아니라 기능어 분석이 포함되어 있다. 특히 한국어의 기능적 특성을 언어학적 변인으로 개발할 필요성을 제기하였고, 이를 위해 한국어의 언어학적 특성을 반영하였다. 문법적 측면에서 한국어는 여러 개의 형태소가 교착하여 하나의 어절을 구성하는 교착어(agglutinative language)로, 하나의 형태소가 독립된 어휘로 나타나는 비율이 높은 굴절어(inflexional language) 계통의 영어와 다르게 매우 정교한 형태소 분석이 필요하다. 또한 사회언어학과 화용론적 측면에서 한국어는 높임말의 사용에 있어 2, 3인칭 대명사의 사용이 극히 제한되는 되는 특성이 있고, 대명사보다 명사구를 반복하는 경향이 강하며, 논항의 삭제도 흔히 발생한다. 논항의 위치도 엄격히 고정되지 않는 반자유순어(semi-free order language)이기 때문에, 논항의 위치가 해당 논항의 통사적 기능을 나타내지 않을 수 있다. 이러한 특성은 품사 태깅(POS tagging)과 구문 분석(parsing)에서 중의성을 증가시키는 요인들이다.

이처럼 언어학적 변인의 개발에서 나타날 수 있는 품사 태깅의 문제를 해결하기 위해서, KLIWC는 9품사(명사, 대명사, 수사, 동사, 형용사, 관형사, 부사, 감탄사, 조사)를 계층적으로 세분화해 47개의 범주로 분석해주는 부산대학교 한국어정보처리 연구실의 형태소 분석기[25,26]를 사용하였다. 이에 더해 ‘문장 당 형태소’, ‘문장 당 어절’ 등의 LIWC의 언어학적 변인도 포함하였다.

언어 사전을 보면, 기본형과 활용형으로 구성되어 있다 [14]. 이는 언어적 변인 외에도 심리적 변인의 태깅을 위한 것으로, 기본형 사전과 한국어 활용형 규칙과 불규칙 사전을 이용하여 새로운 단어의 분석을 가능하도록 할 수 있게 활용형 분석 체계가 설계되었다.

이렇게 개발된 KLIWC는 실제로 언어적 차원, 심리적 차원, 개인 관심사 차원의 80여 개의 변인의 빈도와 비중을 제공한다. 이 변인을 사용하여, 심리학 분야뿐만 아니라 교육학, 신문방송학 분야의 연구에서도 다양하게 사용되고 있다.

2.1.2 KrKwic

KrKwic(Korean Key Words In Context)는 영어권에서 개발된 내용분석 프로그램인 FullText를 한국어로 작성된 메시지의 내용 분석을 위해 변형한 프로그램이다 [24]. KrKwic는 사회연결망 내용 분석을 통해, 여러 뉴스 사이에서 공유된 의미 연합(semantic association) 구조 분석에 특화되어 있다. 예를 들어, ‘매스’와 ‘커뮤니케이션’이 함께 사용되어 ‘매스 커뮤니케이션’라는 개념이 생성된다. 이처럼 특정 단어가 서로 연결될 때 가지는 특별한 의미를 분석하는데 사용된다. 특히 방대한 양의 자료를 분석할 때, 유용하게 핵심어를 추출할 수 있다. 그리고 이러한 단어들의 출현 빈도와 동시 출현 빈도를 제공한다.

하지만, KrKwic 프로그램은 어휘사전을 가지고 있지 않아서, 동의어 혹은 특정 단어의 파생어를 자동으로 하나의 범주로 묶어주는 처리를 하지 못한다. 또한 품사 단위로 분해하지 않고 어절 단위 빈도 정보를 제공하기 때문에, 연구자의 반복된 데이터 클리닝이 요구된다.

KrKwic는 단어 간의 동시 출현 빈도를 기반으로 단어 간의 의미 연합 연결망을 시각화할 수 있는 절차를 제공한다. 시각화는 연결 구조를 직관적으로 이해하는데 중요한 역할을 한다. 하지만 시각화를 위해서는 UciNet 등의 사회 연결망 분석 전문 소프트웨어를 사용해야 하는 점에서 아쉬움이 있다.

2.2 한국어 자연어 처리 기술

인공지능을 위한 기술 중 하나인 자연어 처리 기술은 컴퓨터가 사람처럼 사람의 언어를 이해하고 표현하게 해주는 기술에 관한 연구 분야다. 컴퓨터가 사람의 언어를 이해할 수 있게 사람의 언어 문법에 기초하여, 품사 태깅, 구문 분석, 교정 등 자연어 처리 기술이 연구되었다.

2.2.1 말뭉치(Corpus)

한국어 디지털 언어자원 구축을 목표로 하는 “21세기 세종 계획”을 통해 구축된 세종 말뭉치는 다양한 한국어 언어자원들을 활용한 자연어 처리 기술 연구에 활용되고 있다[27]. 이외에도 부산대학교의 KorLex, 울산대학교의 U-WIN, 한국과학기술원(KAIST)의 CoreNet 등의 어휘 의미망도 구축되어 있다[28].

2.2.2 자연어 처리 기술

자연어 처리 기술 중 가장 기본이 되는 기술에는 품사 태깅(POS tagging, part of speech tagging)과 문법 규칙에 따라 구성 성분으로 분석하는 구문 분석(parsing, 형태 분석)이 있다. 그리고 띄어쓰기와 철자 교정(spelling), 단어 의미의 증의성 해소(Word Sense Disambiguation), 상호 참조(Co-reference Resolution), 개체명 인식(Named Entity Recognition, NER), 요약(Summerization) 등의 응용 자연어 처리 기술은 영어와 비슷한 수준이거나 더 좋은 성능을 나타내고 있다[28]. 특히, 품사 태깅의 정확도가 90%를 넘는 등 높은 수준의 성능을 보여주고 있다[29-32].

2.2.3 한국어 형태소 분석기

자연어 처리를 위한 형태소 분석기는 Stanford NLP를 시작으로 다양한 오픈소스 라이브러리가 다수 배포되었다. 한국어 형태소 분석기도 1995년 KTS를 시작으로 한나눔, 꼬꼬마 KoNLP, MeCab-Ko(은전한닢), KOMORAN 등 다수의 형태소 분석기가 C/C++, JAVA, R로 개발되었다.

NLPK는 교육용으로 개발된 파이썬 패키지인데, 다양한 어간 추출 알고리즘과 50개가 넘는 다양한 언어 말뭉치를 포함하고 있어, 한국어의 분석도 가능하다. 하지만, 언어 분석에 관심 있는 언어학자, 심리학자, 언론학자 등을 포함해 많은 연구 분야의 학자들에게 프로그래밍 능력이 요구되는 점은 사용에 걸림돌이 될 수 있다. 이에 컴퓨터와 프로그래밍 비전문가도 쉽게 배우고 연구에 사용할 수 있는

한국어 분석 패키지가 개발되었다. 대표적으로 파이썬 프로그래밍 언어로 개발된 KoNLPy를 들 수 있다.

KoNLPy는 쉽고 간단한 사용법, 확장 가능성, 상세한 문서, 개방성을 설계 철학으로 하여 개발되었다[33]. KoNLPy 0.3.3에서는 꼬꼬마, 한나눔, MeCab-ko 등 세 가지 형태소 분석기를 사용할 수 있고, 최신 버전인 KoNLPy 0.5.2는 코모란. Okt가 추가되었다(<https://konlpy.org/en/latest/>). 각각 konlpy. tag에서 Kkma, Hannanum, Mecab, Komoran, Okt 클래스를 호출하여 사용할 수 있으며, 모두 동일한 형태의 입력을 받아 동일한 형태로 출력되도록 통일된 인터페이스를 사용하고 있다. 최근 한국어 처리의 통합 인터페이스를 지향하는 프로젝트 KoalaNLP는 11개의 한국어 형태소 분석기의 서로 다른 개발언어, 자료 입력 방식, 분석 결과 등을 동일한 인터페이스로 사용할 수 있게 하는 것을 목적으로 하고 있다(<https://koalanlp.github.io/koalanlp/>).

지금까지 살펴본 바와 같이 한국어 분석 프로그램은 내용어 중심의 분석에서 기능어 분석으로 확장되고 있고, 자연어 처리 기술의 발달로 다양한 한국어 형태소 분석기를 통합적으로 사용하는 것이 가능하다. 또한 빅데이터 시대를 맞아 인터넷 댓글과 SNS 등 다양한 분야에서 텍스트 분석의 요구가 높아지고 있다. 따라서 본 연구는 여러 형태소 분석기를 통합적으로 사용할 수 있는 KoNLPy를 사용하여, 내용어뿐만 아니라 기능어 분석의 정확성을 높일 수 있는 방향을 제안하고자 하였다.

3. 연구 방법

3.1 연구 전략과 대상

본 연구는 크게 두 분석으로 나누었다. 하나는 한국어 분석 프로그램의 장단점 분석이고, 다른 하나는 이 두 분석 프로그램을 사용한 연구의 동향 분석이다.

첫째로 KLIWC, KrKwic 등 한국어 분석 프로그램의 특징을 개발과정을 담고 있는 논문을 통해 분석하였다. 둘째로 국내 학술 DB에서 이 분석 프로그램들을 사용한 연구의 동향을 정량적으로 살펴보았다. 이를 위해 검색 기간은 제한하지 않았고, 2021년 4월 초에 검색하였다. “KLIWC”, “K-LIWC”, “KrKwic.”를 기본 검색어로 국내 대표적인 논문 검색 플랫폼인 RISS (Research Information Sharing Service)에서 국내 논문을 대상으로 검색하였다. 또한, “한국어”와 “글분석 프로그램”을 검색어로 한국어 분석 프로그램 연구를 검색하였다. 261개의 논문이

검색되었는데, 우선 KCI 등재후보지 이상의 논문을 대상으로 하였고, 한국어 교육과 훈련을 위한 프로그램처럼 분석 프로그램을 사용하지 않은 논문은 분석 대상에서 제외하였다. 또한 중복 논문도 제외하였다. 마지막으로 KLIWC와 KrKwic의 개발 과정을 담고 있는 논문도 정성적으로 분석하기 때문에 정량적 분석에서는 제외하였다. 총 173개의 논문이 분석되었다.

3.2 연구방법 및 분석 절차

일차적으로 KLIWC와 KrKwic의 개발 과정을 담고 있는 연구를 분석하여 프로그램의 특징을 비교했고, 이차적으로는 이 프로그램을 사용한 한국어 분석 연구의 동향을 출판연도, 학술연구 분야, 연구 대상으로 구분하여 분석하였다. 이를 토대로 심리 과정 분석을 위한 한국어 분석 프로그램의 개발 방향을 제안하였다.

4. 결과

4.1 한국어 분석 프로그램 분석 비교

4.1.1 언어학적 변인

한국어 분석의 일차적 단계는 품사로 구분하는 것이다. KLIWC는 어간에 교착하는 조사와 어미의 수가 2만 개에 달하고 불규칙적 활용이 다양하기 때문에, 기본형 사전과 불규칙 활용형 사전을 나누어 구축하여 자료의 검색과 갱신의 효율성을 추구하였다[14]. 하지만 이러한 사전 시스템의 본질적인 한계는 주기적인 갱신을 통해 사전을 최신상태로 유지해야 한다는 것이다. 실제로 KLIWC의 언어 사전은 현재까지 갱신이 이루어지지 않아, 언어 사용의 변화와 새로운 어휘의 처리에서 빈번한 오류를 보여 개선이 시급한 것으로 보인다. KrKwic는 품사 분석을 하지 않고 어절 단위의 단어의 출현 빈도와 동시 출현 빈도를 추출하고 있다[26]. 따라서 언어학적 변인의 분석은 KLIWC의 품사 태깅과 꼬꼬마(KKMA), 한나눔(Hannanum), 코모란(Komorán 3.0), Etri의 한국어 품사 분석기와 세종품사(Sejong)의 품사 분류를 비교한 결과를 Table 1에 나타내었다. 품사의 분류는 체언(P1: Che-eon/Substantive), 용언(P2: Yong-eon/Predicate), 수식언(P3: Modifier), 독립언(P4: Independent), 관계언(P5: Relative), 의존형태(P6: Dependence), 부호(P7: Sign), 분석불능(P8: Inability to analyze), 한글이외(P9: Others)의 9개로 구분하였다.

가장 눈에 띄는 특징은 한나눔의 품사 분류가 다른 품사 분석기의 품사 분류보다 더 세분화되어 있다는 것이다. 용언에서 나타나는 특징은 한나눔(Hannanum)은 일반명사(common noun)를 동작성 명사(Motion noun), 상태성 명사(Stative noun), 비서술성 명사(Non-declarative noun), 비서술성 직위 명사(Non-declarative title noun)로 구분하고 있고, 의존명사도 비단위성 의존명사(Non-unitary dependent nouns), 비단위성 의존명사(-하다)(Non-unitary dependent nouns(-hada)), 단위성 의존명사(Unitary dependent noun)로 세분화하고 있다. 동작성 명사에는 ‘활용’, ‘예방’, ‘처리’와 같은 일반명사가 포함되고, 상태성 명사에는 ‘필요’, “와 같은 일반명사가 포함되고, 비서술성 명사에는 ‘백신’, ‘첨단’, ‘국내’와 같은 일반명사가 포함되고, 비서술성 직위명사에는 ‘연구원’, ‘기자’, ‘박사’와 같은 일반명사가 포함된다. 이외에도 용어(Yong-eon), 수식언(Modifier), 관계언(Relative)에서 그 특징이 확인되었다. 하지만, 인칭대명사(Personal pronouns)의 경우에는 KLIWC가 더 세분화되어 있고, 의존형태(Dependence)의 어미(Ending)는 코코마의 품사 분류가 더 세분화되어 있다. KLIWC의 인칭대명사 품사 분류는 1인칭(1st person singular), 2인칭(2nd person singular), 3인칭(3rd person singular), 1인칭 복수(1st person plural), 2인칭 복수(2nd person plural), 3인칭 복수(3rd person plural)로 매우 세분화되어 있고, KKMA의 어미는 평서형 종결어미(Declarative sentence-closing ending), 의문형 종결어미(Interrogative sentence-closing ending), 명령형 종결어미(Imperative sentence-closing ending), 청유형 종결어미(Conjugational sentence-closing ending), 감탄형 종결어미(Exclamatory sentence-closing ending), 존칭형 종결어미(Honorific sentence-closing ending)로 구분되어 있고, 선어미말 어미(Seonemal ending/Pre-final ending)도 존칭 선어미말 어미(Honorific Address Seonemal ending), 시제 선어미말 어미(Tense seonemal ending), 공손 선어미말 어미(Politeness seonemal ending)로 구분되어 있다.

Table 1. Part-of-speech(POS) comparison table between KLIWC and Korean morpheme analyzers

POS	KLIWC	Sejong	KKMA	Hannaum	Komorán 3.0	Etri
P1	Common noun	Common noun	Common noun	Motion noun	Common noun	Common noun
				Stative noun		
				Non-declarative noun		
				Non-declarative title noun		
	Proper noun	Proper noun	Proper noun	Last name	Proper noun	Proper noun
				First name		
				Last name +first name		
				Other (general)		
	Dependent noun	Dependent noun	Common dependent nouns	Non-unitary dependent nouns	Dependent noun	Dependent noun
				Non-unitary dependent nouns(-hada)		
			Unit dependent noun	Unitary dependent noun		
	Cardinal numeral	Numeral	Numeral	Cardinal numeral	Numeral	Numeral
	Ordinal numeral			Ordinal numeral		
	Demonstrative pronouns	Pronouns	Pronouns	Demonstrative pronouns	Pronouns	Pronouns
	Personal pronouns			Personal pronouns		
1st person singular	/	/	/	/	/	
2nd person singular						
3rd person singular						
1st person plural						
2nd person plural						
3rd person plural						
P2	Verb	Verb	Verb	Demonstrative Verb	Verb	Verb
				Common Verb		
	Adjective	Adjective	Adjective	Demonstrative Adjectives	Adjective	Adjective
				State-property Adjective		
	/	Auxiliary predicate element	Auxiliary verb auxiliary Adjective	Auxiliary predicate element	Auxiliary predicate element	Auxiliary predicate element
'Ida' postposition	Positive copula	Positive copula, 'Ida' postposition	predicative particle	Positive copula	Positive copula	
/	Negative copula	Negative copula, 'Anida' Adjective	/	Negative copula	Negative copula	

P3	Determiner	Determiner	Common determiner	Demonstrative determiner	Determiner	Demonstrative determiner	
				State-property determiner		State determiner	
			Numeral determiner			Numeral determiner	
	Adverb	Common adverb	Common adverb	Demonstrative adverb	Common adverb	Common adverb	
			Common adverb				
		Conjunctive adverb	Conjunctive adverb	Conjunctive adverb	Conjunctive adverb	Conjunctive adverb	
P4	Interjection	Interjection	Interjection	Interjection	Interjection	Interjection	
P5	Postpositional particle; Postpositional marker	Prepositional particle for nominative	Prepositional particle for nominative	Prepositional particle for nominative	Prepositional particle for nominative	Prepositional particle for nominative	
		Complement case marker	Complement case marker	Complement case marker	Complement case marker	Complement case marker	
		Adnominal case marker	Adnominal case marker	Adnominal case marker	Adnominal case marker	Adnominal case marker	
		Object case marker	Object case marker	Object case marker	Object case marker	Object case marker	
		Adverbial case marker	Adverbial case marker	Adverbial case marker	Adverbial case marker	Adverbial case marker	
		Vocative case marker	Vocative case marker	Vocative case marker	Vocative case marker	Vocative case marker	
		Quotative case marker	Quotative case marker	Quotative case marker	Quotative case marker	Quotative case marker	
		Conjunctive postpositional particle	Conjunctive postpositional particle	Comitative postpositional particle	Connective Postposition	Conjunctive postpositional particle	Conjunctive postpositional particle
Auxiliary postpositional particle	Auxiliary postpositional particle	General used auxiliary postpositional particle	Sentence-closing auxiliary postpositional particle	Auxiliary postpositional particle	Auxiliary postpositional particle		
P6	Seonemal ending/ Pre-final ending	Seonemal ending	Honorific Address Seonemal ending	Seonemal ending	Seonemal ending	Seonemal ending	
			Tense seonemal ending				
			Politeness seonemal ending				
	Ending	Sentence-closing ending	Declarative sentence-closing ending		Sentence-closing ending	Sentence-closing ending	Sentence-closing ending
				Interrogative sentence-closing ending			
				Imperative			

			sentence-closing ending			
			Conjugational sentence-closing ending			
			Exclamatory sentence-closing ending			
			Honorific sentence-closing ending			
	Connective ending	Coordinative ending	Coordinative ending	Coordinative ending	Coordinative ending	Coordinative ending
		Subordinate connective ending	Subordinate connective ending			
		Auxiliary connective ending	Auxiliary connective ending			
	Nominalization ending	Nominalization ending	Nominalization ending	Nominalization ending	Nominalization ending	Nominalization ending
	Adnominal suffix	Adnominal suffix	Adnominal suffix	Adnominal suffix	Adnominal suffix	Adnominal suffix
Prefix	Che-eon prefix	Che-eon prefix	Prefix	Che-eon prefix	Che-eon prefix	Che-eon prefix
		Verb prefix				
Suffix	Noun-derived suffix	Noun-derived suffix	Noun-derived suffixes after units	Noun-derived suffix		Noun-derived suffix
			Noun-derived suffix after action noun			
			Noun-derived suffixes after common nouns			
			Noun-derived suffix after state noun			
			Noun-derived suffix after personal pronouns 1, 3			
			Noun-derived suffix after all nouns			
	Verb derivation suffix	Verb derivation suffix	Verb derivation suffix	Verb derivation suffix		Verb derivation suffix
			Verb derivation suffix after action noun			
			Verb derivation suffix after common noun			
	Adjective-derived suffix	Adjective-derived suffix	Adjective-derived suffix after state nouns	Adjective-derived suffix	Adjective-derived suffix	Adjective-derived suffix

				Adjective-derived suffixes after common nouns		
			Adverb-derived suffix	Adverb-derived suffix after adjective		
				Adverb-derived suffix after state noun		
			Other Suffix			
		Stem	Stem		Stem	Stem
P7		Period, question mark, exclamation point	Period, question mark, exclamation point	Period	Period, question mark, exclamation point	Period, question mark, exclamation point
		Comma, center dot, colon, slash	Comma, center dot, colon, slash	Comma	Comma, center dot, colon, slash	Comma, center dot, colon, slash
		Quotation marks, parentheses, lines	Quotation marks, parentheses, lines	Opening quotes and quotation marks	Quotation marks, parentheses, lines	Quotation marks, parentheses, lines
				Closing quotation marks and quotation marks		
		Ellipsis	Ellipsis	Ellipsis	Ellipsis	
		Hyphens (tilde, hidden, missing)	Hyphens (tilde, hidden, missing)	Hyphens	Hyphens (tilde, hidden, missing)	Hyphens (tilde)
		Other Symbols (Logical Math Symbols, Money Symbols)	Other Symbols (Logical Math Symbols, Money Symbols)	Unit symbol	Other Symbols (Logical Math Symbols, Money Symbols)	Other symbols
				Other symbols		
P8	Noun presumed category	Noun presumed category	Noun presumed category		Inability to analyze	Inability to analyze
		Verb Guessing Category	Verb Guessing Category			
		Inability to analyze	Inability to analyze			
P9	English	Foreign language	Foreign language	Foreign language	Foreign language	Foreign language
	Chinese character	Chinese character	Chinese character		Chinese character	Chinese character
		Number	Number		Number	Number

4.1.2 심리적 변인과 사용자 변인

심리 변인은 ‘글쓴이가 얼마나 지적이고 논리적인가?’, ‘슬프거나 우울한가?’, ‘즐겁거나 기뻐하고 있는가?’, ‘중요하게 여기는 사회적 관계는 무엇인가?’ 등 글쓴이의 생각과 정서를 엿볼 수 있는 단초가 된다. 예를 들어, 입학에 대한 인상과 느낌을 주제로 한 글쓰기에서 우울한 사람은 그렇지 않은 사람보다 부정적인 정서 단어를 더 많이 사용하였다[34]. 정서적인 격변(emotional upheavals)에 대한 글쓰기 효과 연구를 보면[35], 외상 경험에 대한 3일 이상의 글쓰기에서 긍정적인 정서 단어의 사용이 증가하고, 부정적인 정서 단어의 사용은 줄어들고, 원인과 사고와 같은 인지 과정의 단어 사용이 증가하였다. 그 외에도 여성은 남성보다 정서 단어와 사회적 과정의 단어를 더 많이 사용하였고[19,20], 거짓말을 할 때 부정적인 정서 단어를 더 많이 사용하고 [16,21], 악성 댓글을 쓸 때 그렇지 않은 댓글을 쓸 때보다 부정적인 정서 단어를 더 많이 사용하고, 인지 과정의 단어를 더 적게 사용하는 특성을 나타낸다[22,23]. 이처럼 글쓴이의 심리적 과정에 대한 탐색에서 심리사전의 활용은 매우 중요하다. 따라서 새로운 한국어 분석 프로그램의 심리사전을 한층 더 강화할 필요가 있다. 기존 한국어 분석 프로그램 중에 심리사전을 포함하고 있는 프로그램은 KLIWC가 유일하여, KLIWC의 심리사전을 분석하였다. 또한 KLIWC는 LIWC에 뿌리를 두고 있어, LIWC의 심리사전도 참조하였다.

Table 2에 아주대학교에서 연구용으로 운영 중인 KLIWC의 심리사전과 LIWC의 심리사전을 나타내었다. KLIWC의 심리사전을 보면, 3 계층구조(L1: Upper level, L2: Middle level, L3: Lower level.), 47개 범주로 구성되어 있다. 총 16348개 단어가 포함되어 있고, 8970개의 단어가 분류되어 있다. 상위 계층 범주(L1)는 하위 계층 범주(L2와 L3)를 포함하는 구조로 어휘가 중복 집계되는 방식으로 빈도가 계산된다. KLIWC 심리사전은 LIWC의 2001버전의 심리사전과 전체적인 구조가 유사하였다. 실제로 이창환 등[14]은 LIWC의 심리학적 과정 차원(정서, 인지, 감각, 사회과정)을 포함하는 한국어 어휘집에 대한 연구가 부족함으로 지적하였고, 한국에서 자주 사용되는 단어로 심리 단어집을 구성하였다고 밝히고 있다. 따라서 본 연구에서는 LIWC의 갱신된 사전을 모두 참조하여, KLIWC의 심리사전의 어휘를 재분류하여 미분류 단어들 줄이고자 한다. 또한 소설/비소설, 영화와 드라마, 신문, 논문, 유튜브, 방송사 뉴스 등에서 새로 수집된 단어와

심리 과정과 관련성이 높은 어휘에 대한 선행연구들을 개관하여 새로운 어휘들을 추가하는 할 필요가 있다.

범주 단위로 살펴보면, KLIWC의 긍정적인 정서(Positive emotion) 범주는 ‘긍정적인 느낌(Positive feeling)’과 ‘낙천성 또는 활동성(Optimism or activity)’으로 나누어져 있다. 긍정적인 느낌(Positive feeling)에는 “즐거움”, “사랑”, “행복함” 등이 포함되어 있다[14]. 기쁨과 즐거움, 만족과 안도 등 긍정적인 정서를 나타내는 단어가 구분 없이 모두 포함되어 있다. 낙천성 또는 활동성(Optimism or activity) 변인은 긍정적인 정서와 연결된 행동과 연결된 단어들로 구성되어 있다.

접근 동기-회피 동기와 만족과의 상호작용에 대한 연구를 보면, 접근 동기인 사람이 회피 동기인 사람보다 만족감에 대해 강하게 묶여 있다[36]. 따라서 긍정적인 느낌을 기쁨과 즐거움, 만족과 안도 등으로 더 세분화한다면, 글쓴이의 동기에 뿌리를 두고 있는 정서 탐색에 효과적일 것으로 생각한다[37]. 또한 한국어 정서 체크리스트를 이용한 연구[38]의 단어들을 참조하여, 정서 범주의 변인의 단어들을 재분류할 필요성도 있다. 정서 범주에 대해서는 KLIWC의 3 계층 구조를 유지하되, 모든 정서 단어에 대해서 하위 계층까지 분류하여, 미분류율을 낮추는 것이 필요하다.

인지 과정(Cognitive process)을 보면, LIWC 2007과 2015 버전에서 제한(inhibition) 범주가 제외되었다. 그러나 ‘제한(inhibition)’은 억제 작용과 관련된 범주로 인지 과정에 중요하다. 예를 들어, 무관련 정보의 억제는 인지 과정에서 매우 중요하기 때문에[39][40], 유지해야 할 것으로 보인다.

감각 과정은 오감각을 나타내는 어휘들을 분류하여, 중분류에 시각, 청각, 미각, 후각, 촉각 변인을 신설할 필요가 있다. 감각 명사[41], 지각 동사에 대한 기존 연구(시각: 김인화[42]; 송현주[43]; 청각: 정수진[41]; 미각: 이은희[44], 정수진[45,46]; 후각: 노운채[47]; 촉각: 김준기[48]; 천시권[49])를 참조하여, 단일 대분류의 ‘감각지각적인 과정(Sensory-Perception process)’을 세분화할 필요가 있다. 또한 공감각 어휘에 대한 선행 연구[50][51]를 참조할 때, 공감각 변인의 가능성도 열어두어야 할 것으로 생각된다.

자기활동(Self activity)과 여가활동(Leisure activities)의 중분류 변인들이 LIWC의 2007, 2015 버전에서 삭제되었다. 하지만 이들 변인은 글쓴이의 관심사가 무엇인지 탐색할 수 있는 변인들로, 한국어 분석 연구에서의 변별력을 검토

한 후 삭제 여부를 고려해야 할 것이다. 반면, 2015 버전에는 Drives(동기), affiliation(소속감), power(권력), reward(보상), risk(위험) 변인이 새롭게 추가되었다. 새로운 심리사전을 구축한다면, 기존의 성취(Achievement)를 포함하여, 동기 변인을 대분류로 하고, 중분류에 소속감, 성취, 권력, 보상, 위험 분류를 포함하는 위계적 구조를 추가할 필요가 있다고 판단된다.

KLIWC의 어눌한 말(Slurred speech) 변인은 범주 분류만 있고 실제 단어가 포함되어 있지 않았다. LIWC는 nonflu와 filler 등 담화 표지어(discourse marker)를 포함하고 있는데, 이 단어들과 한국어 담화 표지어 연구 [52-57]의 참조하여 한국어 구어 표현에 적합하게 수정하여 추가할 필요가 있다.

Table 2. Psychological variables of KLIWC and LIWC.

KLIWC			LIWC		
L1	L2	L3	2001	2007	2015
Affective & Emotional process			Affect	Affect	Affect
	Positive emotion		Posemo	Posemo	Posemo
		Positive feeling	Posfeel		
		Optimism or activity	Optim		
Negative emotion			Negemo	Negemo	Negemo
		Anxiety	Anx	Anx	Anx
		Anger	Anger	Anger	Anger
		Sad/depress	Sad	Sad	Sad
Cognitive process			CogMech	CogMech	CogProc
		Cause	Cause	Cause	Cause
		Insigt	Insight	Insight	Insight
		Expect	Discrep	Discrep	Discrep
		Inhibition	Inhib		
		Guess	Tentat	Tentat	Tentat
		Conviction	Certain	Certain	Certain
Sensory-Perception process			Senses	Percept	Percept
Social process			Social	Social	Social
		Face			
		Communication	Comm		
		Refer to others	Othref		
		Reer friend	Friends	Friends	Friend
		Family	Family	Family	Family
		Humans	Humans	Humans	
Self activity					
		School	School		

	Job-Work	Job		
	Achievement	Achieve	Achiev	Achieve
Leisure activities		Leisure	Leisure	Leisure
	Home	Home	Home	Home
	Sports	Sports		
	TV-Movie	TV		
	Music	Music		
Money-Financial issues		Money	Money	Money
Physical state & function		Physcal	Bio	Bio
	Body condition and symptoms	Body	Body	Body
	Gender- Sexual characteristics	Sexual	Sexual	Sexual
	Eating- Ddrinking- Diet	Eating	Ingest	Ingest
	Sleep-Dream	Sleep		
Metaphysical issues		Metaph		
	Death	Death	Death	Death
	Religion	Relig	Relig	Reling
Slang				Informal
	Swear word	Swear	Swear	Swear
Slurred speech				

4.1.3 사용 환경과 전처리 등 사용성

연구용으로 제작된 KLIWC와 KrKwic는 사용성이 낮다. KLIWC는 맞춤법, 오타자, 분류 불가능 기호 등 오류를 유발할 수 있는 요인을 분석자료를 탑재하기 전에 제거해야 분석 결과를 얻을 수 있다. KrKwic는 어절 단위 분석으로 고빈도 단어의 추출 시 기호뿐 아니라 조사, 어미 등 언어학적 변인을 제거하는 절차를 반복해야 한다. 이러한 점은 연구자의 사용 가능성에 부정적일 영향을 줄 수밖에 없을 것이다.

사용 환경 측면 역시 프로그램의 사용을 어렵게 하고 있다. KLIWC는 웹응용 프로그램으로 접근성이 우수한 반면, 웹서버와 DB 서버가 구축 운용되어야 한다. 윈도우 환경에서 개발된 KLIWC는 15년이 지나면서 운영체제의 지원 종료 등으로 인한 바이러스 취약성이 나타났고, 랜섬웨어에 감염되는 등 보안 문제가 심각하다. KrKwic의 경우 도스 환경에서 개발되어 도스 명령어를 통해 실행해야 한다. 또한 현재의 64 비트 윈도우 환경에서 직접 사용이 불가능하고, 에뮬레이터(emulator)를 사용해야 하는 점들이 사용자의 불편을 초래하고 있어, 사용자의 접근성과 사용성을 높이기 위하여 변화하는 IT환경에 적합한 형태로 개발이 시급하다.

4.2 한국어 분석 프로그램을 사용한 연구 분석

4.2.1 연도별 연구 동향

한국어 분석 프로그램을 사용한 연구들은 연도별로 Table 3과 Fig. 1에 나타내었다. 2004년 처음 발표되었고

21년 3월까지 증가하고 있다. 특히 2013년을 기점으로 한국어 분석 프로그램을 사용한 연구논문이 매년 두 자릿수 이상 출판되는 등 급격한 증가세를 나타내고 있다. 이는 빅데이터 시대에 내용분석 도구의 필요성이 증가하고 있기 때문일 것이다. 특히 KrKwic는 중심 내용을 담고 있는 핵심어를 추출하고, 그 단어 간의 관계(의미연결망)를 효과적으로 시각화할 수 있어 많이 사용되고 있다. Fig. 2는 본 연구에서 분석하고 있는 173편의 연구의 평균 KCI 피인용수(Mean Times Cited from KCI)를 KLIWC와 KrKwic로 구분하여 나타내고 있다. 평균 KCI 피인용수는 개별 논문이 KCI의 다른 논문에 인용된 수를 평균한 것이다. 36편의 KLIWC 논문이 285회 인용되어 한 편당 평균 7.92회 인용되었고, 136편의 KrKwic 논문은 1532회 인용되어 한 편당 평균 11.26회 인용되었다. 이 결과들은 한국어 분석 프로그램이 연구자들에게 높은 관심을 받고 있다는 것을 나타낸다.

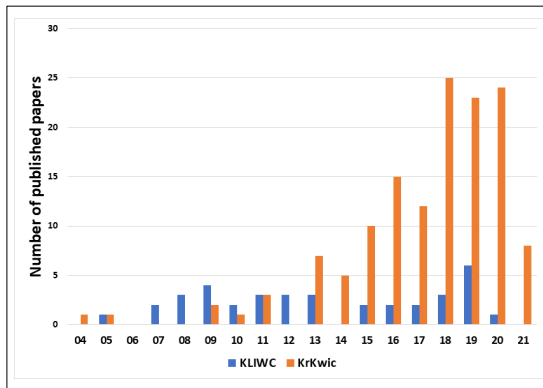


Fig. 1. Number of papers by year using KLIWC and KrKwic. In 2009, studies using both KLIWC and KrKwic were duplicated

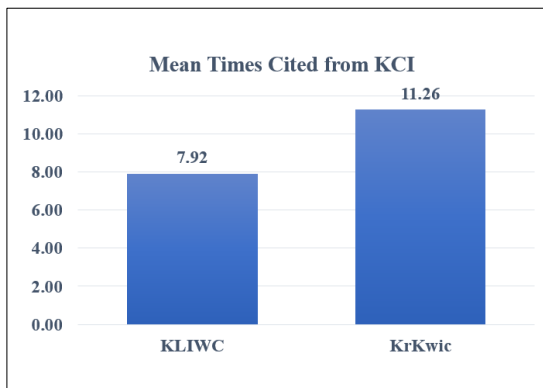


Fig. 2. Mean Times Cited from KCI.

4.2.2 연구 영역별 연구 동향

한국어 분석 프로그램을 사용한 연구가 얼마나 다양한 연구 분야에서 활용되고 있는지를 알아보기 위해, 출판된 논문이 속한 연구영역에 대해 분석하였다. 그 결과를 한국 연구재단의 대분류에 따라 공학(R1: Engineering), 농수해양학(R2: Agriculture and Maritime Field), 복합학(R3: Interdisciplinary Studies), 사회과학(R4: Social Science), 예술체육학(R5: Arts and Kinesiology), 의학학(R6: Medicine and Pharmacy), 인문학(R7, Humanities), 자연과학(R8: Natural Science) 구분하여 Fig. 3과 4에 나타내었다.

Fig. 3을 보면, KLIWC를 사용한 연구를 분야를 나타내고 있다. 통계학, 심리과학, 교육학에서 많이 사용되었다. 복합학 분야인 인지과학(Cognitive Science), 학제간 연구(Interdisciplinary Research)에서 KLIWC를 이용한 연구가 발표되었다. 이 영역에 발표된 논문은 거짓말[21], 대통령의 언어스타일[18]에 나타난 언어적, 심리적 특성을 비교하였고, 포털사이트에서 행복과 불행이라는 단어와 동시 발생한 단어들[58]을 비교하였다.

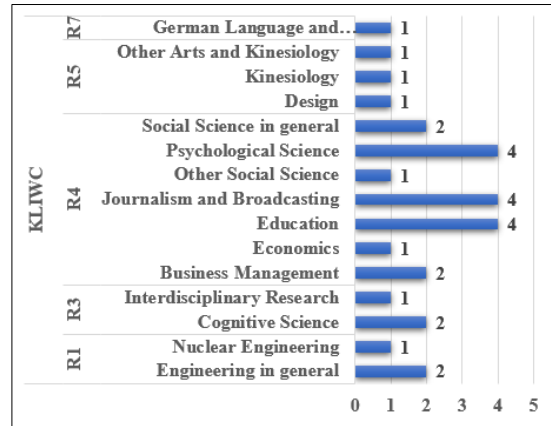


Fig. 3. Classification of papers by research area using KLIWC

Fig. 4는 KrKwic를 사용한 연구를 학술연구 분야별로 나타내었다. 교육학과 관광학에서 집중적으로 사용되었다. 복합학 분야 과학기술학(Science and Technology Studies), 감성과학(Affective Science), 문헌정보학(Library and Information Science), 학제간연구(Interdisciplinary Research)에서 KrKwic를 이용한 연구가 발표되었다. 이 영역에 발표된 논문은 춘천[59],

카프카의 변신에 대한 연구[60], 기록관리학[61], BTS에 대한 청소년의 정서탐색[62], 제주이주[63] 등의 대한 키워드를 시각화하여 현상을 탐색하거나 연구동향을 분석하였다.

종합하면, KLIWC를 사용한 연구는 집단별 언어적, 심리적 특성을 비교 연구가 많은 비중을 차지하고 있고, KrKwic를 사용한 연구들은 핵심어 추출해 시각화하거나 동시출현단어의 관계를 통해 연구동향을 분석하고 있는 차이점이 있다. 또한 한국연구재단의 학술연구 분야 분류표상 8개의 대분류 모두에서 한국어 분석 프로그램을 연구에 사용하고 있다는 것을 보여준다.

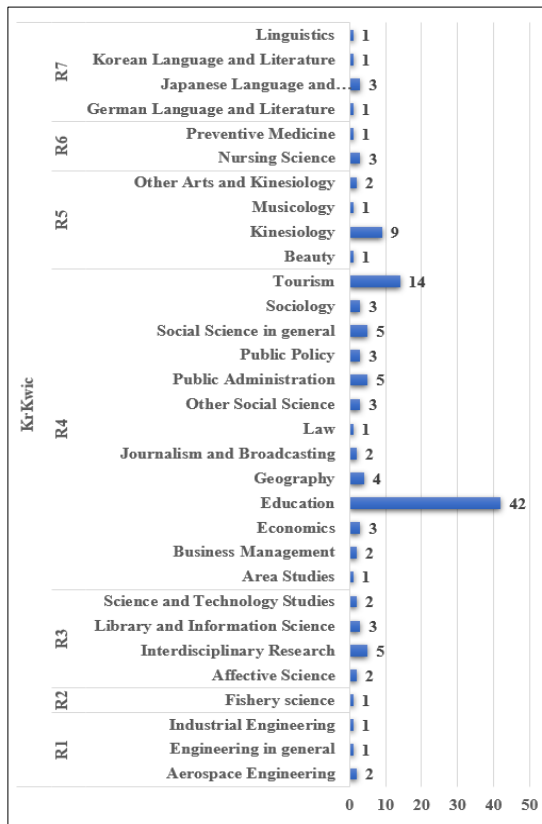


Fig. 4. Classification of papers by research area using KrKwic

4.2.3 분석 대상별 연구 동향

빅데이터 시대에 한국어 분석 프로그램과 같은 분석 도구의 도움 없이, 방대한 글과 말을 분석하는 것은 사실 불가능하기 때문에, 분석 대상에 대한 분석은 한국어 분석 프로그램 개발 계획에 중요한 통찰을 제공할 수 있다. 분석

결과 논문과 보고서가 61건(35.3%)으로 가장 많이 분석된 글 자료였고, 신문기사와 행정문서도 많이 연구되었다. SNS, 사용후기, 댓글 등을 연구한 논문도 다수 확인되었다. 또한 대화, 토론, 연설, 방송언어와 같은 말 언어는 녹취되어 분석되었다. 경험과 전문성이 요구되는 상담 축어록, 설문 응답, 인터뷰 녹취도 분석 되었다.

특기할 만한 점은 KrKwic를 사용한 연구는 논문과 보고서를 분석 대상으로 하여 관련 분야의 의미연결망 분석을 주로 수행하였고, KLIWC는 글쓴이의 생각, 정서, 성격 등 심리 과정을 분석하기 위한 글쓰기를 요구하고 대조군과 그 차이를 비교하는 연구에서 많이 사용되었다는 것이다. 이러한 차이는 KLIWC와 KrKwic의 사용목적에 뚜렷한 차이점이 있음을 시사한다.

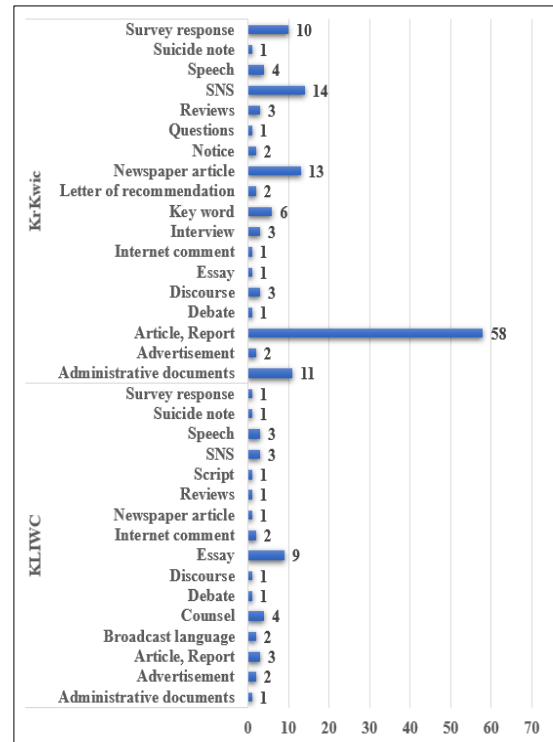


Fig. 5. Classification of papers by text type using KLIWC and KrKwic.

5. 결론 및 시사점

5.1 언어적 변인: 통합 품사 태깅 체계의 설계

글쓴이의 의도나 정서적 상태처럼, 심리학적 분석을 위해서는 기존의 내용어 중심의 분석뿐 아니라 KLIWC처럼

언어학적 변인의 분석이 매우 중요하다. 새로운 분석 프로그램에서는 다양한 언어학적 변인의 분석이 가능하도록 개발되어야 한다. 예를 들어, 1인칭의 낮은 사용 비중과 선어말어미의 높은 사용 비중은 거짓말의 지표로 사용되었다 [16,21]. 특히 선어말어미는 한국어의 특징으로 높임말과 시제 표현을 나타낸다는 측면에서 분류를 세분화할 필요가 있다. 따라서 가장 세분화된 품사 분류가 가능한 한나눔 품사분석기를 사용하는 것이 최선으로 보이나, 선어말어미어미의 분류는 꼬꼬마 분석기가 더 세분화되어 있다.

이러한 점을 고려할 때, 여러 품사 분석기를 통합해 사용할 필요성이 있다고 생각한다. 실제로 KoNLPy, KoalaNLP는 다양한 품사 분석기를 통합적으로 사용하려고 시도하였고, 사용자의 목적에 따라 여러 형태소 분석기를 사용할 수 있다. 따라서 언어학적 기능어 분석을 효과적으로 수행할 수 있도록, 가장 세분화된 품사 분류가 가능한 한나눔을 기본으로 하고, 선어말어미의 경우 꼬꼬마 품사 분석의 분류를 적용해야 할 것으로 보인다. 이처럼 새로운 한국어 분석 프로그램을 위해서 통합적인 품사 태깅 체계를 설계해야 할 것이다.

5.2 심리적 변인의 보강

KLIWC의 심리 변인의 분석 결과 개발 후 갱신이 없었다는 점에서 한국어 분석 프로그램을 사용한 연구에서 확인되는 트렌드를 반영하는 새로운 범주의 변인을 추가하거나 용도가 낮은 변인을 제거하는 변화는 불가피할 것으로 보인다. 정서 변인의 경우 부정적 정서가 불안, 화, 우울의 축으로 세분된 것과 다르게 긍정적인 정서는 즐거움과 만족처럼 결이 다른 긍정 정서를 구분하지 않고 있다[14]. 정서의 동기적 측면을 고려할 때[36,37], 긍정적인 정서를 기쁨/즐거움과 만족/안도의 축으로 구분할 필요가 있다. 또한 LIWC의 2015 버전에서 affiliation(소속감), power(권력), reward(보상), risk(위험) 등의 동기 변인이 추가되었다. 이러한 변화를 반영하여, 동기 대분류 변인을 신설하고, 중분류로 기존의 성취에 더불어 새로 4개의 동기 변인을 추가하는 것이 필요하다. 동기 변인을 추가하면 정서 변인과 함께 글쓴이의 심리 상태를 보다 체계적으로 설명할 수 있고, 행동에 대한 예측도 조심스럽지만 가능해질 것으로 기대한다.

감각지각 과정의 세분화도 중요한 요구라고 생각한다. 시각, 청각, 촉각, 후각, 미각의 오감각을 나타내는 명사와 동사로 변인을 구성하는 것은 심리학적 글분석에서 중요한

의미가 있다고 생각한다. 감각과 지각은 주위 환경의 변화를 눈, 귀, 피부, 코, 혀의 다섯 감각 기관을 통해 외부의 자극을 알아차리는 과정으로, 정보처리 양식에서의 개인차가 나타날 수 있다. 예를 들어, 학습 상황에서 감각을 선호하는 학습자는 구체적 지각을 선호하고 직관을 선호하는 학습자는 추상적인 아이디어를 선호한다[64]. 또한 신체화 증상 연구 [65]를 보면, 부정적 정서와 감정표현불능증 등의 취약성은 신체 귀인과 신체감각 증폭지각 등의 인지과정의 매개를 통해 신체화가 유발된다는 설명을 하고 있다. 이외에도 감각 관련 표현의 변화는 외부 환경과의 상호작용 과정을 추론할 수 있는 단서를 제공하기 때문에 중요한 도구로 사용될 수 있다.

마지막으로 KLIWC의 심리사전의 계층 구조를 유지하면서, LIWC 2007, 2015에서 갱신된 사전의 변화를 반영하고, 포함된 단어에 대한 재분류도 함께 진행하면서 미분류율을 낮추는 방식으로 심리사전을 정교화해야 할 것이다.

5.3 활용성

본 연구에서는 한국어 분석 프로그램 KLIWC와 KrKwic를 비교 분석하고, 이 분석 프로그램을 사용한 연구를 고찰하여 새로운 한국어 분석 프로그램을 위한 개발 방안을 제안하고자 하였다.

문헌 고찰의 주요 결과를 정리하면, 첫째 한국어 분석 프로그램을 사용한 연구들은 2004년 처음 발표된 후로 꾸준히 증가하고 있고, 2013년을 기점으로 두 자릿수 이상으로 급격히 늘어나고 있다. 이는 빅데이터 시대에 내용분석 도구의 필요성이 증가한 것의 효과로 보인다. 둘째, 한국 연구재단의 학술연구분야 분류표상 8개의 대분류 모두에서 한국어 분석 프로그램을 사용한 연구가 나타났다. 특히, 사회과학 분야의 교육학에서 많은 연구(45건, 25.7%)가 수행되었고, 그 외에도 원자력공학, 수산학, 디자인, 미용, 지구과학 다양한 분야에서 한국어 분석 프로그램을 연구에 사용하고 있다. 셋째, 학술지 기준으로 볼 때, 흥미롭게도 전체 연구의 1/4이 집중된 교육학 분야의 연구가 31개의 다양한 학술지에 고루 분포되어 있었다. 마지막으로 분석 대상에 대한 분석 결과를 보면, 논문과 보고서가 61건 (34.9%)으로 가장 많이 분석된 글 자료였고, 신문기사와 행정문서의 분석도 많이 수행되었다.

KLIWC과 KrKwic 두 프로그램의 사용 특징 비교하면, 논문과 보고서의 분석에 KrKwic가 많이 사용되었고, 글쓴이의 생각, 정서, 성격 등 심리 과정을 분석하기 위한

글쓰기를 요구하고 대조군과 그 차이를 비교하는 연구에서는 KLIWC가 많이 사용되었다.

다양한 분야에서 한국어 분석 프로그램의 사용 요구가 증가하고 있는 시점에서, 기존 프로그램의 문제점을 개선 및 보완하는 새로운 한국어 분석 프로그램의 개발이 필수적일 것이다. 특히 KLIWC의 장점인 품사 기반의 언어학적 분석과 심리사전 기반의 심리변인의 심층적 분석기능을 제공 하면서, KrKwic의 장점인 연결망 시각화와 같은 직관적인 정보를 제공하는 기능을 포함해야 할 것이다.

5.4 기대 효과

본 연구를 통해 기대되는 효과는 크게 세 가지이다. 첫째, 통합적 품사 태깅 체계는 세분화된 품사 변인의 사용 비중을 제공할 수 있어서 기존의 KLIWC의 분석에서 탐지하지 못했던 선어말어미, 어미의 사용 특성을 탐지할 수 있을 것으로 기대한다. 둘째, 심리사전의 개발 과정에서 동기-정서-행동으로 이어지는 심리적 과정에 대한 어휘집이 구축 될 것으로 기대된다. 이를 통해 단편적인 정서 분석에서 체계적인 분석이 가능한 새로운 언어행동 분석의 도구 개발이 기대된다. 셋째, 빅데이터 시대의 요구에 부응하여 다양한 분야의 연구에서 사용할 수 있는 한국어 분석 프로그램 개발을 촉진할 수 있을 것으로 생각한다.

REFERENCES

- [1] J. M. Lee. (1998). *Problems II of Cognitive Psychology: Language and Cognition*, Seoul: Hakjisa.
- [2] N. Chomsky. (1965). *Aspects of the theory of syntax*, Cambridge, MA: MIT Press.
- [3] R. Jackendoff. (2002). *Foundations of language: Brain, meaning, grammar, evolution*, Oxford: Oxford University Press, 2002
- [4] C. Beckner & J. Bybee. (2009). A Usage-Based Account of Constituency and Reanalysis. *Language Learning*, 9, 27-46.
- [5] M. Barlow & S. Kemmer. (2000). *Usage-based models of language*, Stanford, CA: Center for the Study of Language and Information.
- [6] M. Carreiras, B. C. Armstrong, M. Perea, & R. Frost. (2014). The what, when, where, and how of visual word recognition. *Trends in cognitive sciences*, 18(2), 90-98.
- [7] W. D. Marslen-Wilson & A. Welsh. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, 10(1), 29-63.
- [8] J. Bybee. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-733.
- [9] E. Mergenthaler. (1996). Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of consulting and clinical psychology*, 64(6), 1306-1315.
- [10] P. J. Stone, D. C. Dunphy, & M. S. Smith. (1966). *The General Inquirer: A Computer Approach to Content Analysis*, Cambridge: MIT Press.
- [11] J. W. Pennebaker, M. E. Francis, & R. J. Booth. (2001). *Linguistic inquiry and word count: LIWC 2001*, Mahway: Lawrence Erlbaum Associates.
- [12] J. W. Pennebaker & L. A. King. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296-1312.
- [13] J. W. Pennebaker, M. R. Mehl, & K. G. Niederhoffer. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1), 547-577.
- [14] C. H. Lee, J. -M. Sim, A. Yoon. (2005). The Review about the Development of Korean Linguistic Inquiry and Word Count. *The Korean Society for Cognitive Science*, 16(2), 93-121.
- [15] R. B. Slatcher, C. K. Chung, J. W. Pennebaker, & L. D. Stone. (2007). Winning words: Individual differences in linguistic style among US presidential and vice presidential candidates. *Journal of Research in Personality*, 41(1), 63-75.
- [16] M. L. Newman, J. W. Pennebaker, D. S. Berry, & J. M. Richards. (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5), 665-675.
- [17] Y. R. Tausczik & J. W. Pennebaker. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24-54.
- [18] Y. Kim & K. Kim. (2019). Language style of the President of the Republic of Korea: linguistic and psychological properties in Presidents' Address. *The Korean Society for Cognitive Science*, 30(3), 105-132.
- [19] K. Kim, J. Bae, & C. H. Lee. (2009). Gender Differences in Language Use: An Analysis by

- KLIWC. *Journal of the Korean Data Analysis Society*, 11(3), 1307-1318.
- [20] M. L. Newman, C. J. Groom, L. D. Handelman, & J. W. Pennebaker. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3), 211-236.
- [21] Y.-I. Kim, Y. Kim, & K. Kim. (2016). Detecting a deceptive attitude in non-pressure situations using K-LIWC. *The Korean Society for Cognitive Science*, 27(2), 247-273.
- [22] Y.-I. Kim, Y. Kim, Y. Kim, & K. Kim. (2019). The Characteristics of Malicious Comments: Comparisons of the Internet News Comments in Korean and English. *Journal of The Korea Contents Association*, 19(1), 548-558.
- [23] Y. Kim, Y.-I. Kim, & K. Kim. (2013). Differences of Linguistic and Psychological Dimensions between Internet Malicious and Normal Comments. *Journal of the Korean Data Analysis Society*, 15(6), 3191-3201.
- [24] H. W. Park & L. Leydesdorff. (2004). Understanding the KrKwic: A computer program for the analysis of Korean text. *Journal of the Korean Data Analysis Society*, 6(5), 1377-1387.
- [25] M.-J. Kim & H.-C. Kwon. (1992). An Automatic Indexing Method Using the Characteristics of Korean. *Proceedings of Korean Institute of Information Scientists and Engineers*, (pp. 1005-1008). Seoul: Korean Institute of Information Scientists and Engineers.
- [26] Y. J. Chung, J. H. Park, A. S. Yoon, & H. C. Kwon. (2000). An Improved Korean Morphological Analyzer. *Artificial Intelligence and Soft Computing*, July 24-26, 2000, (pp. 466-471). Banff, Alberta, Canada: IASTED/ACTA Press 2000.
- [27] Y. J. Hwang & J. Choi. (2016). A closer look at the 21st century Sejong Corpus. *Saegugeosaenghwal*, 26(2), 73-86.
- [28] M. Han, S. Park, H. Lee, J. Yeon, & S. Lee. (2015). Natural language processing on Korean language : A survey. *Proceedings of Korean Institute of Information Scientists and Engineers*, (pp. 681-683). Seoul: Korean Institute of Information Scientists and Engineers.
- [29] S.-H. Na, C.-H. Kim, & Y.-K. Kim. (2014). Lattice-based Discriminative Approach for Korean Morphological Analysis. *Journal of KISS : Software and Applications*, 41(7), 523-532.
- [30] J.-C. Shin & C.-Y. Ock. (2012). A Stage Transition Model for Korean Part-of-Speech and Homograph Tagging. *Journal of KISS : Software and Applications*, 39(11), 889-901.
- [31] K. Shim. (2013). Morpheme Restoration for Syllable-based Korean POS Tagging. *Journal of KISS : Software and Applications*, 40(3), 182-189.
- [32] K. Shim. (2014). Syllable-based Probabilistic Models for Korean Morphological Analysis. *Journal of KIISE*, 41(9), 642-651.
- [33] E. Park & S. Cho. (2014). KoNLPy: Korean natural language processing in Python. *Annual Conference on Human and Language Technology*, (pp. 133-136). Seoul: Human and Language Technology.
- [34] S. Rude, E. M. Gortner, & J. W. Pennebaker. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18(8), 1121-1133.
- [35] J. W. Pennebaker & M. E. Francis. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion*, 10(6), 601-626.
- [36] J. A. Updegraff, S. L. Gable, & S. E. Taylor. (2004). What Makes Experiences Satisfying? The Interaction of Approach-Avoidance Motivations and Emotions in Well-Being. *Journal of Personality and Social Psychology*, 86(3), 496-504.
- [37] E. T. Higgins. (1997). Beyond pleasure and pain. *American psychologist*, 52(12), 1280-1300.
- [38] Y. Hong, Y. Nam, & Y. Lee. (2016). Developing Korean Affect Word List and Its Application. *Korean Journal of Cognitive Science*, 27(3), 377-406.
- [39] K. K. Harnishfeger, The development of cognitive inhibition: Theories, definitions, and research evidence, *Interference and inhibition in cognition*, (pp.175-204). Academic Press. 1995.
- [40] C. M. MacLeod. (2007). The concept of inhibition in cognition. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition*, (p.3-23). American Psychological Association.
- [41] S. J. Jeong. (2012). A Cognitive Analysis of the Meaning Extension of Sensual Nouns in Korean. *Hanminjok Emunhak*, 60, 271-290.
- [42] I. Kim. (1987). *Research for Color Terms of Current Korean Language*. Master dissertation. The Graduate School of Ewha Womans University. Seoul.
- [43] H. J. Song. (2003). A Study on the Meaning

- Extension of Color Adjectives in Korean. *The Journal of Linguistics Science*, 24, 131-148.
- [44] E. Lee. (2019). A Study on the Presentation Status and Improvement Plan of Taste Expression Vocabulary in Korean Textbooks. *Studies of Korean & Chinese Humanities*, 64, 111-133.
- [45] S. J. Jeong. (2003). A Cognitive Interpretation of the Meaning of the Korean Sweet Taste Expressions. *The Journal of Linguistics Science*, 24, 303-320.
- [46] S. J. Jeong. (2005). The Aspects of the Meaning Extension of the Taste-terms. *Korean Semantics* 18, 149-174.
- [47] Y. Nho. (2010). Verbes olfactifs du français : du phénomène physiologique au phénomène linguistique. *Etudes de la Culture Francaise et de Arts en France*, 34, 27-59.
- [48] C. Kim. (2002). A Study on the Meaning of Words Related to Temperature. *The Korean Language and Literature*, 78, 21-45.
- [49] S. Cheon. (1980). Correlation system of temperature vocabulary. *The Journal of Korean Language and Literature Education*, 12, 1-14.
- [50] J. -H. Kim. (2001). A Cognitive Linguistic Approach on Synaesthetic Expressions in Korean. *Discourse and Cognition* 8(2), 23-46.
- [51] I. -S. Chung. (1997). Synaesthetic Transfer of Korean Adjectives. *Studies in Modern Grammar*, 11, 163-180.
- [52] M. H. Kim. (2005). The Development of Korean Interrogatives as Discourse Markers. *Discourse and Cognition* 12(2), 41-63.
- [53] K. H. Suh. (2002). An Interactional Account of the Korean Demonstrative ku in Conversation. *Language and Linguistics*, 29, 137-158.
- [54] H. Lee. (1999). The Pragmatics of the Discourse Particle mwe in Korean. *Discourse and Cognition* 6(1), 137-157.
- [55] H. Lee. (1997). The pragmatics of the discourse particle woay in Korean. *Discourse and Cognition* 4(1), 1-20.
- [56] H. Lee. (1996). The pragmatics of the discourse particle kuray in Korean. *Discourse and Cognition* 3, 1-26.
- [57] C. Yoon & T. Kim. (2015). A Study on the Discourse Markers Used in the Listening Texts of TOPIK. *The Journal of Linguistics Science* 75, 231-250.
- [58] S. Park, K. Do, H. Kim, G. Park, J. Yun, & K. Kim. (2018). An Exploratory Study of Happiness and Unhappiness Among Koreans based on Text Mining Techniques. *The Journal of the Korea Contents Association*, 18(7), 10-27
- [59] H. Kim, Y. Jang, H. Park, & S. Ryu. (2011). A Study on Tourists' Urban City Image : Focused on ChunCheon Case. *Review of Culture and Economy*, 14(1), 95-122.
- [60] C. T. Lim & K. Lee. (2020). Semantic Network Analysis on the Research Tendency of "'Die Verwandlung'" by Franz Kafka in KoreaII. *Cultural Exchange and Multicultural Education*, 9(4), 375-392.
- [61] Y. Choi. (2015). A Study on the Research Trends of Archival Studies in Korea: : Focused on Research Papers between 2004 and 2013. *The Korean Journal of Archival Studies*, 43, 147-177.
- [62] M. J. Cheong, J. Kim, & S. I. Kim. (2018). A Study on the Emotional Search and Psychological Schema of Teenagers Enjoying the Songs of BTS and the U.N. Speech Text - Recovery of Self-Esteem -. *Culture and Convergence*, 40(8), 555-578.
- [63] M. Chang, Y. Heo, & H. -S. Lim. (2019). 'Moving to Jeju': An Exploratory *Keyword Analysis Using Social Big Data*. *Journal of Tourism & Industry Research*, 39(1), 15-26.
- [64] M. Kim, S. Ahn, J. Lee, D. Ahn, & Y. Lee. (2006). Characteristics of Medical Students' Learning Styles and Their Relationships to GPA. *Korean Journal of Medical Education*, 18(2), 133-140.
- [65] H. K. Shin. (2000). The Effect of Negative Affectivity, Alexithymia, Somato-sensory Amplification, and Somatic Attribution on Somatization: Test of a Linear Mediation Model. *Korean Journal of Clinical Psychology*, 19(1), 17-32.

김 영 준(Youngjun Kim)

[정회원]



- 2006년 8월 : 아주대학교 심리학과 (문학석사)
- 2015년 2월 : 아주대학교 심리학과 (철학박사)
- 2019년 9월 ~ 현재 : 아주대학교 심리학과 강사

- 관심분야 : 인지/언어심리, 인지공학
- E-Mail : arete0@ajou.ac.kr

최 원 일(Wonil Choi)

[정회원]



- 2004년 2월 : 고려대학교 심리학과 (문학석사)
- 2013년 5월 : 노스캐롤라이나 주립대학 심리학과 (철학박사)
- 2016년 9월 ~ 현재 : 광주과학기술원 기초교육학부 교수

- 관심분야 : 인지/언어심리, 인간-로봇 상호작용
- E-Mail : wichoi@gist.ac..kr

김 태 훈(Tae Hoon Kim)

[정회원]



- 1999년 2월 : 고려대학교 심리학과 (문학석사)
- 2009년 12월 : 오하이오주립대학교 심리학과 (PhD)
- 2012년 3월 ~ 현재 : 경남대학교 심리학과 교수

- 관심분야 : 인지심리, 혁신적 사고, 판단과 의사결정
- E-Mail : taehoonk@kyungnam.ac.kr