

질적 자료의 정량화 방안: 감성분석 기법을 활용한 챗봇 기반 과업활동의 만족도 분석 중심으로*

신동광** · 이장호*** · 김혜영 · 양혜진

교수, 광주교육대학교 · 교수, 중앙대학교 ·

교수, 중앙대학교 · 연구원, 중앙대학교

Shin, Dongkwang, Lee, Jang Ho, Kim, Heyoung & Yang, Hyejin (2021). A proposal for analyzing qualitative data quantitatively based on the satisfaction analysis on chatbot-based tasks using sentiment analysis. *Journal of the Korea English Education Society*, 20(3), 167-185.

The present study examines the perception of undergraduate students about chatbot-based English writing tasks, by employing a sentiment analysis approach. To achieve this, the participants were asked to communicate with a chatbot, called Mitsuku, based on seven sub-tasks. The data collected for this study include: 1) English utterances of the participants, generated through the chatbot-based tasks; 2) open-ended responses of the participants in Korean to the chatbot-based tasks; and 3) results from two 5-point Likert-scale questionnaire items, assessing the perceived effectiveness by the participants of chatbot-based tasks for their English learning, as well as their interests in those tasks. The results revealed that the participants were generally satisfied with the chatbot-based tasks. The results from the different tools were moderately or highly correlated overall, and the data of the participants' English utterances, obtained from the sentiment analysis, was the most significant predictor of their expectations for the effectiveness of chatbot-based tasks for their English learning. The present study provides methodological implications for the employment of the sentiment analysis as an alternative or supplement for the traditional Likert-scale items.

[chatbot/sentiment analysis/Likert-scale questionnaire/Tone Analyzer/KNU
Korean Sentiment Lexicon/챗봇/감성분석/리커트 척도 설문/감성
분석기/KNU 한국어 감성사전]

* 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018S1A5A2A03037255).

** 제1저자

*** 교신저자

I. 서론

차수미, 김정렬과 남승우(2021) 연구에서는 최근 국내 학술지에 게재된 영어 교육 논문 중 챗봇과 관련한 연구를 분석하였고 관련 연구의 수가 40여 편에 이른다고 보고하고 있다. 이는 그 만큼 영어교육 분야에서 영어 교육·학습뿐 아니라 평가 등에 챗봇을 적용하려는 시도가 다양하게 이루어지고 있다는 증거이다. 지금까지 챗봇 연구의 대부분은 챗봇의 특징을 비교·분석(김영우, 2020; Coniam, 2014; Lee, Yang, Shin, & Kim, 2020)하거나 시범 적용을 통한 챗봇의 교수학습 적용가능성(신동광, 2019; Fryer & Carpenter, 2006)을 탐색하는 연구가 주를 이루고 있다. 김재상(2017)의 연구에 따르면 언어교육에 있어 챗봇의 활용은 다양한 장점을 가지고 있는데 이러한 연구들이 공통적으로 제시하는 요점은 외국어에 대한 불안을 낮춰줄 수 있다는 점이다. 즉 교사와의 상호작용에서는 발화의 두려움이 상대적으로 높게 나타나는 경향을 보인 반면 챗봇과의 대화에서는 발화 시도 시의 긴장감이 낮아지고 발화 오류 발생에도 큰 어려움이나 긴장 없이 지속적인 재시도를 보인다는 것이다. Shen, Wang과 Shen(2009)은 학습 상황에서 긍정적인 학습 태도와 같은 감정적 요인은 학습자의 학습 동기와 그들의 학습 결과에 매우 큰 영향을 미친다고 주장하였다.

그러한 이유로 앞서 언급한 챗봇 연구에서도 실험연구인 경우, 사후 설문조사에서 리커트 척도(Likert scale)를 기반으로 만족도 또는 흥미도를 조사하거나 개방형 설문문항을 통해 챗봇 기반 과제의 참여 소감 등을 조사하고 있다. 물론 연구마다 그 결과가 유사하거나 상이하기도 하다. 예를 들어, 양혜진, 김혜영, 신동광과 이장호(2019)의 연구에서는 연구에 참여한 초등학생들의 과반수가 챗봇 기반 과제의 참여에 흥미를 느끼는 등 챗봇 사용에 매우 긍정적으로 반응을 보인 반면, 일본에서 연구를 진행한 Kanda, Hirano, Eaton와 Ishiguro(2004)의 연구에서는 초등학생을 대상으로 한 실험에서 챗봇에 대한 학습자들의 초기 관심은 시간이 지날수록 현저히 낮아지는 양상을 보였다. 이는 만족도, 흥미도, 동기유발과 같은 정의적 요인의 측정에는 많은 환경적인 요인이 개입될 수밖에 없고 그만큼 측정이 쉽지 않다는 것을 시사한다. 특히 연구자들이 가장 손쉽게 채택할 수 있고 실제 적용하는 방식은 리커트 척도에 기반한 설문문항이다. 하지만 리커트 척도의 경우에도 중간점(5점과 같은 홀수 척도의 경우 3점) 사용 유무 등과 같이 결과에 영향을 주는 변수가 존재한다(장덕현, 조성겸, 2017). 이를 보완하기 위해, 개방형 질문(open-ended question)을 추가로 사용하기도 하지만 개방형 질문의 해석 또한 주관성을 배제하기 힘들다는 한계가 있다.

본 연구는 챗봇 기반 과업이라는 특정한 환경 속에서 학습자들이 챗봇과 직접 영어로 이야기하면서 작성한 영문 대화문과 이 과업에 참여한 소감에 대한 학습자들의 한국어 응답을 감성분석을 통해 살펴보고자 한다. 더불어 5점 리커트 척도에 기반한 설문문항으로 챗봇 기반 과업의 영어 학습효과에 대한 기대

치와 흥미도 및 참여도에 대해 조사한 후, 해당 결과를 감성분석 기법을 통해 도출된 자료와 비교·분석함으로써 측정 도구 간 유사도 혹은 편차를 살펴보고자 한다. 이를 통해 대화문이나 소감과 같은 질적 자료를 보다 객관적으로 정량화 하는 방법을 제안해 보고자 한다.

II. 이론적 배경

1. 챗봇 활용 언어교육 관련 선행연구

챗봇(Chatbot)은 보통 인간과 상호작용할 수 있는 인공지능 로봇 프로그램을 의미하며 이 용어는 Mauldin(1994)이 처음 제안한 ‘ChatterBot’에서 시작되었다. 최근에는 챗봇에 대한 언어교육 분야의 관심이 집중되면서 다양한 연구들이 집중적으로 진행되고 있다(차수미 외, 2021).

챗봇을 활용한 외국어 교육과 관련한 국외의 초기 연구로 Coniam(2014)은 ‘Cybelle,’ ‘Dave,’ ‘George,’ ‘Jenny’ 및 ‘Lucy’라는 5개의 챗봇을 평가하였다. 주로 언어의 정확성에 초점을 두고 평가가 이루어졌으며 특히 단어, 문장 및 문단 수준으로 구분하여 언어 구조를 분석하였다. 분석 결과를 살펴보면, 5종의 챗봇은 모두 어휘와 문법 면에서 다른 특성을 보였으며 문법적인 면에서는 전반적으로 우수한 능력을 보여 큰 차이를 보이지 않았지만 어휘 사용의 측면에서는 상당한 편차를 보였다. Coniam은 당시 개발된 대표적인 챗봇들이 언어학습의 관점에서 사용자와 자연스런 대화를 수행하는 데 한계가 있다는 점을 지적하였다.

Abbasi와 Kazi(2014)는 챗봇과 Google 검색 엔진을 통해 대학생의 학습 결과와 기억력을 비교하였다. 이 연구에서는 ‘OOPLChatbot’이라는 QnA 챗봇의 프로토타입을 개발하여 실험에 적용하였다. 연구에서는 챗봇을 사용하는 실험 집단($n = 36$)과 검색 엔진을 사용하는 통제 집단($n = 36$)으로 나누어 그 효과성을 검증하였다. 이를 위해 실험 중 과제 수행 후 5분, 2시간, 1일의 세 가지의 시간 프레임을 설정하여 두 집단의 학습 결과를 측정하였다. 두 집단 모두 사전 테스트와 사후 평가 간에 상당한 학습 향상을 보였으나 실험 집단의 점수가 통제 집단의 점수보다 유의한 수준에서 더 높게 나타났다. 또한, 세 가지 시간 프레임에 따른 분석 결과, 챗봇 사용 집단은 Google 검색 엔진을 사용자 집단에 비해 더 학습 내용을 더 오래 기억한다는 것을 보여 보여주었다. 즉, 챗봇과의 대화를 통한 맥락적 이해는 학습자의 장기 기억에 기여한다는 것이 확인되었다.

국내 초기 챗봇 관련 연구는 AMAZON의 Echo와 같은 범용 인공지능형 스피커를 언어교육에 활용하기 위한 과업 개발이 주를 이루었다. 김재상(2017)의 연구는 AMAZON의 Echo에 기반하여 초등 학습자 수준에 적합한 3단계의 챗봇

기반 과업을 제안하였다. 1단계는 통제과업(controlled task)으로 챗봇에게 요청할 질문 표현을 제시하고 그대로 읽고 정보를 수집하게 하는 것이었다(예, How is the weather in London?). 2단계에서는 안내과업(guided task)으로 일정한 패턴 표현만 제시하고 학습자 스스로 질문을 완성하여 과업을 수행하게 하였다(예, What is ~? 패턴 활용하기 + ‘workaholic’의 의미를 알아보세요.). 3단계는 수행할 과업에 대한 주제만 제시하고 학습자 스스로 문제해결을 시도하는 독립과업(independent task)으로 구성되었다(예, 미국의 수도는 어디인지 알아보세요.).

신동광(2019)의 연구에서는 Echo보다는 심층적인 대화(문자 채팅)가 가능한 2개의 챗봇인 Mitsuku와 Cleverbot를 영어 쓰기 지도에 시범 적용하고 그 결과를 분석하였다. 신동광은 학습자의 참여도 편차를 줄이기 위해 챗봇 기반 과업을 7개의 세부 지시(예, 인사하기, 자기 소개하기, 직업 알아보기 등)로 구성하여 제시하였다. 연구 결과, 범용 챗봇인 Mitsuku와 Cleverbot이 구사하는 영어 표현은 한국인 고등학생이나 대학생 학습자 수준에 적절한 수준이었고 학습자들도 챗봇 기반 과업에 대해 대부분 긍정적인 반응을 보였다.

2. 감성분석 적용 교육 관련 선행연구

감성분석은 ‘Affective Computing’(Picard, 1995)이라는 개념에서 시작되었다. Affective Computing은 인간의 감정 상태를 기계를 통해 분석하고 해석하여 그 결과에 적절한 반응을 제시하는 연구라고 정의할 수 있다. 감성분석 방식은 보통 감성사전(sentiment dictionary)에 따른 사전기반 기법(lexicon based approach)과 빅데이터에 의한 기계학습 기법(machine-learning approach)으로 구분된다. 최근에는 기계학습 기법의 적용이 더 증가하는 추세이지만 가장 널리 사용되는 기법은 사전기반 기법과 기계학습 기법을 혼용하여 적용하는 방식이다(Zhou & Ye, 2020).

언어교육에서 감성분석을 적용한 사례는 극히 제한적이지만 최근 교육 분야에서의 감성분석 연구들을 살펴보면 주로 학습과정 중 학습자의 감성 상태를 분석하여 피드백을 제공하거나 학습의 성패를 예측하는 연구가 대부분이며 일부 연구는 학습과정에 대한 학습자의 평가 도구로서 사용되는 것을 확인할 수 있다. 본 장에서는 이 두 가지 주제를 중심으로 선행연구를 살펴보려고 한다.

1) 학습과정 중 학습자의 감성 상태 분석 연구

감성분석의 적용 방식 중 하나는 온라인 학습·평가 시스템에 결합하여 학습자들의 반응을 감성분석을 통해 실시간으로 분석하고 그 결과를 바탕으로 자동 피드백을 제공하거나 학업성취 정도를 예측하는 것이다. 또한, 이러한 감성분석은 학습자의 만족도를 즉각적으로 분석하여 교사나 학습시스템 관리자가 이를 모니터링하여 교수학습의 개선에 활용할 수도 있다(Ortigosa, Martín, & Carro,

2014).

Lin, Chen, Sun과 Tsai(2014)는 그들의 연구에서 학습과정 중 감성분석을 통한 시각적 피드백을 제공하였다. 현재 일반적으로 제공되는 시각적 피드백은 크게 가상 에이전트, 워드 클라우드 및 대시 보드의 유형이 대부분이며 이들의 연구에서는 가상 에이전트를 채택하였다. 이러한 시각적 피드백은 학습 결과에 긍정적인 영향을 미치며, 학습자들로 하여금 학습 시스템과 상호작용한다고 느끼게 만드는 것으로 나타났다.

Arguedas 외(2018)의 연구 또한 사용자의 감정 상태를 분석하고 이를 바탕으로 피드백 및 비계(scaffolding)를 제공하는 모델을 제안하였다. 이를 위해 Arguedas 외는 감성사전 기반 감성분석을 실시하였고 이러한 정보는 협력적 학습 분위기 조성 및 학습과정 시의 장애요소를 식별하는데 활용되었다. 학습자들은 이러한 감성적 피드백이 학습 맥락에서 주의력을 높이고 기억 및 추론 능력을 향상시키는데 도움을 주었다고 평가하였다.

Liu 외(2019)는 토론 위주 수업에서 학습자들의 감정 변화를 지속적으로 추적하는 감성분석 모델을 제안하였다. Liu 외는 종단적 감정 변화 측정 모델(Temporal Emotion-Aspect Model, TEAM)을 통한 분석을 바탕으로 몇 가지 흥미로운 결과를 제시하였다. 첫 번째는 부정적인 감정 요인은 학습환경보다는 학습내용에 관련된 부분일수록 더 명확하게 나타난다는 것이고, 두 번째는 학습자들의 감정 표현은 학습과정 중에 나타나기보다는 학기 초나 학기 말에 보다 두드러진다는 것이다. 끝으로 상대적으로 학업성취도가 낮은 학생일수록 동료 학습자들과의 정서적 관계성을 맺는데 소극적이었고 학기 말에 정서적으로 더 혼란스런 모습을 보였다. 즉 참여도가 낮을수록 한 학기 동안 무엇을 했는지, 무엇을 얻었는지를 알 수 없는 회의적인 상태가 될 가능성이 높다는 것이다.

2) 학습자의 학습내용 및 강의평가 연구

실시간 감성분석이 아니더라도 하나의 강좌가 끝나고 학습자들의 정의적 반응을 분석하여 학습내용 및 강의 전반에 대한 만족도를 분석하고 그 결과를 강의 개선에 반영하는 것은 감성분석을 적용할 수 있는 또 다른 방법이라고 볼 수 있다.

학습자의 만족도 분석은 온라인 학습 시스템의 설계를 위해 매우 중요한 과정 중 하나이다. Elia, Solazzo, Lorenzo와 Passiante(2019)는 온라인 협업 학습 시스템의 만족도를 평가하기 위해 학습자들이 참여하는 콘텐츠, 서비스, 상호작용을 위한 커뮤니티 등에서 생성된 대규모의 학습자 로그 데이터를 활용하여 시스템에 대한 만족 수준을 감성분석 기법을 적용하여 분석하였다. 이들이 제안한 모델은 시스템 만족도뿐만 아니라 협업 학습 시스템이라는 점을 고려하여 개인 학습자별 기여도도 분석할 수 있도록 고안되었다. 이 연구는 이러한 감성분석 기법이 설문조사와 같은 전통적인 분석 기법에 비해 소요되는 시간과 노

력이 적고 실시간 분석 처리 또한 가능하다는 장점과 더불어 학습과정 중 자연스럽게 학습자들이 생성하는 데이터를 활용함으로써 분석 결과에 대한 신뢰도 또한 높일 수 있다는 점을 강조하였다.

Alm와 Nkomo(2020)는 영어교육 프로그램의 만족도 조사를 위해 감성분석을 활용하였다. 이를 위해 챗봇 대화 기능이 포함된 온라인 학습 도구 사용 후 학습자들의 챗봇 기능에 대한 평가를 살펴보기 위해 관련 커뮤니티 게시판의 댓글을 수집하여 감성분석 기법으로 분석하였다. 이 연구는 Microsoft Azure Cognitive Services에서 제공하는 감성분석 도구를 활용하여 Duolingo, Eggfun, Memrise 및 Mondly, 네 가지 학습 어플리케이션에 대한 학생들의 만족도를 분석하였다. 분석 결과, 학습자들은 챗봇과의 대화에 대해 어느 정도 호기심을 가지고 대화에 참여하고자 하는 의지를 보인 것으로 확인되었다. 하지만 챗봇과의 대화내용이 학습목표와 일치하지 않거나 사용 중 나타난 기술적인 오류 등에 대해서는 부정적인 반응을 보였다.

위에서 살펴본 감성분석의 적용 방식과는 다른 새로운 시도를 보인 연구도 있었다. 신동광(2020)의 연구에서는 감성분석 기법을 수능 영어 문항 유형 중 하나인 ‘심경·분위기 파악’ 문항의 타당도 분석에 적용하였다. 이 분석에는 기계학습 방식을 채택한 IBM의 ‘Tone Analyzer’(Bhuiyan, 2017)와 Higuchi(2016)가 개발한 ‘KH-Coder’에 감성사전을 탑재하여 분석하는 방식을 적용하여 그 결과를 비교하였다. 분석 결과에서는 Tone Analyzer가 67%의 정답 예측률을 보였고 KH-Coder는 57%의 예측률을 보여 기계학습 방식이 보다 정확한 예측을 가능하게 한다는 것을 입증하였다. 중요한 시사점 중 하나는 두 방식 모두 감정을 나타내는 용어가 지문에 포함되어 있을 경우 높은 예측률을 보였지만 감정이 직접적으로 드러나지 않고 지문 전반에 나타난 상황적 맥락을 유추해야 정답을 찾을 수 있는 문항에서는 매우 낮은 예측률을 보였다는 것이다. 그런데 ‘심경·분위기 파악’ 유형의 기출 문항 대부분에서는 감정을 나타내는 어휘를 중심으로 문항이 출제되어 특별히 추론 능력을 요구하지는 않았다. 따라서 이 문항 유형의 취지가 추론적 능력을 측정하는 문항임을 감안하면 타당도에 상당한 문제가 있었다는 것을 알 수 있다. 이에 신동광의 연구에서는 출제 시 이 문항유형의 타당도 검증을 위해 감성분석 기법의 활용을 제안하였다. 즉 감성분석기를 통해 분석이 어려운 문항이라면 오히려 추론적 문항으로 잘 출제되었다고 가정할 수 있다는 것이다.

위의 선행연구에서 보듯, 타 분야와는 달리 영어교육 분야에서는 감성분석 기법의 적용 사례가 매우 제한적이라는 것을 알 수 있다. 본 연구에서는 챗봇 기반 과업 참여에 대한 만족도 분석에 대하여 기존의 리커트 척도와 감성분석을 함께 적용하여 유사도 또는 편차를 확인해 보고자 한다. 이에 다음과 같은 세부 연구 질문을 설정하였다.

첫째, 챗봇과의 대화문 감성분석, 챗봇 과업 참여 소감문 감성분석, 그리고

챗봇 과업 참여 만족도 설문조사의 결과는 어떠한가?

둘째, 각 만족도 분석 도구에 따라 도출된 수치들은 어느 정도의 상관관계를 보이는가?

셋째, 영어 학습효과에 대한 기대치를 예측하는데 가장 효과적인 만족도 분석 도구는 무엇인가?

III. 연구 방법

1. 데이터 수집

연구에서 분석한 데이터는 2019년 2학기에 A대학에서 영작문 과목을 수강한 초등영어교육을 전공한 여학생 18명과 남학생 10명, 총 28명의 3학년 학생들이 온라인 쓰기 활동의 하나로 1회에 한해 챗봇 기반 과업을 수행하면서 작성한 대화문과 과업 수행 후 실시한 설문조사 자료이다. 실험에 참여한 학생들은 대학수학능력시험 영어영역에서 1등급 수준으로 챗봇과 영어로 대화를 나누는데 큰 어려움이 없는 영어능력을 가지고 있었다.

2. 실험 도구 및 절차

1) 챗봇 및 챗봇 기반 과업

본 연구에서는 과업 수행을 위한 챗봇으로 Mitsuku, 현재는 Kuki(<https://chat.kuki.ai/chat>)라 불리는 개방형 챗봇을 사용하였다. 이는 챗봇 개발의 월드컵이라고 할 수 있는 뢰브너 상(Loebner Prize)을 2013, 2016, 2017, 2018, 2019년에 걸쳐 5회나 수상한 가장 우수한 챗봇 중 하나이다(Guinness World Records, 2020). Mitsuku는 20대의 여성 캐릭터로 설정되어 있으며 인간과 유사한 대화가 가능하다. 특히 자체 데이터베이스에 없는 특정한 정보를 질의하면 웹검색을 통해 요청한 정보를 제공하는 특징을 가지고 있다. 챗봇 기반 과업은 자유대화보다는 학습자 간 대화 주제나 발화량의 편차를 줄이기 위해 아래의 표 1과 같이 7개의 세부 과업으로 나누어 대화를 진행하도록 고안하였다.

표 1
챗봇 기반 세부 과업

단계	수행 과업
1	인사하기
2	자기소개(이름 등)하기 및 미추쿠의 신상정보 알아보기(예, 사는 곳, 취미, 결혼 여부, 가족 등)

- 3 하고 있는 일(직업) 알아보기, 사람들을 돕기 위해서 미추쿠가 할 수 있는 일 알아보기
- 4 미추쿠가 좋아하는 것들(이유 포함) 알아보기, 그 밖의 궁금한 점 물어보기
- 5 태국의 대표적인 관광지, 축제, 음식, 기후 등의 유무형의 문화적 정보 알아보기
- 6 스페인의 대표적인 관광지, 축제, 음식, 기후 등의 유무형의 문화적 정보 알아보기
- 7 대화 마무리하고 인사하기

위의 챗봇 기반 과업은 표 1의 7단계의 채팅이 끝난 후 ‘태국’과 ‘스페인’의 문화를 비교하는 에세이 쓰기가 포함되어 1시간 동안 수행되었지만 본 연구에서는 챗봇과의 대화문 분석에만 초점을 두었다. 챗봇 발화를 제외한 순수한 학생들의 발화량은 평균 266.9 단어(표준편차: 182.8)였고 가장 길게 대화한 학생의 경우 총 932 단어, 가장 짧은 대화의 경우 62개 단어의 발화량을 보였다.

2) 대화문 감성분석 도구

위의 과업 수행 시 Mitsuku와 나눈 대화 중 학습자의 발화 데이터는 영문 감성분석 자료로 활용되었다. 이렇게 수집된 대화문에 나타난 학습자의 감성분석은 IBM의 Tone Anlayzer라는 감성분석기를 통하여 분석하였다. Bhuiyan(2017)의 Tone Anlayzer 소개에 따르면 Tone Anlayzer는 7개의 감성—분노(Anger), 두려움(Fear), 기쁨(Joy), 슬픔(Sadness), 분석적(Analytical), 확신적(Confident), 잠정적(Tentative)—의 분석이 가능하며 분석 결과는 그림 1과 같이 감성의 강도에 따라 0-0.5(매우 약함), 0.5-0.75(해당 감성 포함) 0.75-1(매우 강함)의 3단계로 제시된다.

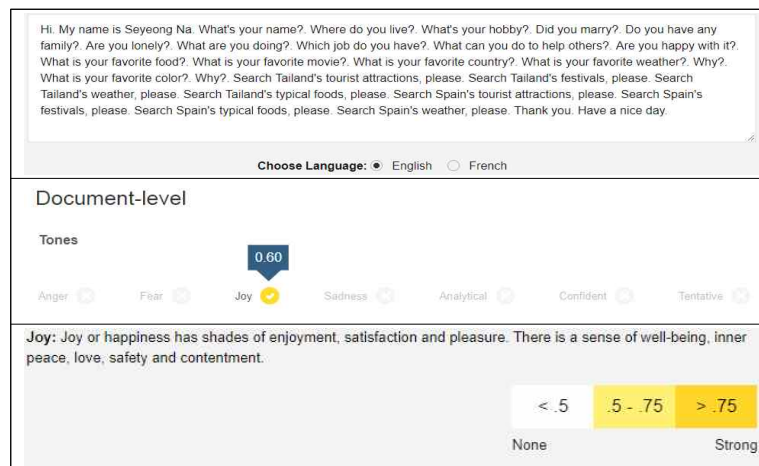


그림 1
Tone Analyzer 분석 예시

Tone Analyzer는 다양한 종류의 감성을 분석할 수 있으나 일반적으로 한국어 감성분석이나 리커트 척도 설문문항은 긍정, 중립, 부정의 속성만 분석 가능한 점을 고려하여 기쁨(Joy)은 긍정으로, 잠정적(Tentative)는 중립으로, 분노(Anger), 두려움(Fear), 슬픔(Sadness)은 부정으로 분류하였으며 이를 5점 리커트 척도의 설문문항과 비교 가능하도록 표 2와 같이 변환 기준을 설정하였다. 예를 들어, 긍정에 해당하는 수치가 0.6인 경우, 아래 표의 변환 기준에 따라 4점 척도로 변환하였다.

표 2
대화문 감성분석 수치의 척도화 점수 변환 기준

척도	1	2	3	4	5
의미	부정2	부정1	중립	긍정1	긍정2
기준	0.75초과~1	0.5~0.75	0.5~1	0.5~0.75	0.75초과~1

3) 챗봇 기반 과업 참여 소감문 감성분석 도구

챗봇 기반 과업 참여 소감문에 대한 감성분석을 위해 본 연구에서는 R 프로그램(R Development Core Team, 2011)에 KNU 한국어 감성사전(KNU Korean Sentiment Lexicon)을 탑재하여 감성분석을 시도하였다. Data Intelligence Lab(2017)이 개발한 KNU 한국어 감성사전은 표준국어대사전에 수록된 형용사, 부사, 동사, 명사를 긍정, 중립, 부정으로 분류한 후 전문 패널의 검토를 거쳐 ‘매우 부정,’ ‘부정,’ ‘중립,’ ‘긍정,’ ‘매우 긍정’ 5개 단계로 세분화하여 개발되었다. KNU 한국어 감성사전에 포함된 감성의 종류 및 강도, 그리고 각 강도에 포함된 한국어 어휘 수와 관련 예시는 아래 표 3에서 정리된 바와 같다.

표 3
KNU 한국어 감성사전의 구성

감성의 종류 및 강도	한국어 감성사전 수록 어휘 수	예시
매우 긍정	2,597	가능성이 있다고, 기특하고
긍정	2,266	보탬이, 깜찍하고
중립	154	가까스로, 수줍다
부정	5,029	사납게, 차거나 싫은
매우 부정	4,797	못마땅할, 불안한
계	14,843	

총 14,843개의 감성어휘로 구성된 KNU 한국어 감성사전은 5개 단계로 구성되어 앞서 소개한 챗봇 대화문의 5개 단계와의 비교가 용이하였다. 다음의 그림 2는 실제 R 프로그램에 KNU 한국어 감성사전을 탑재하여 학습자들의 챗봇

기반 과업 참여 소감문을 분석한 예시이다.

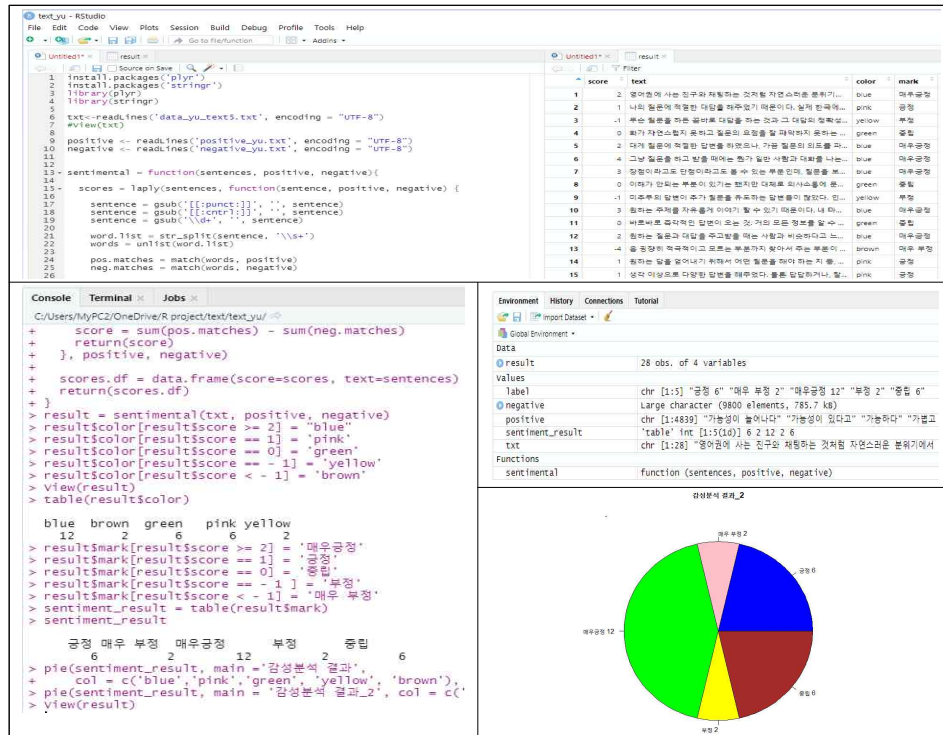


그림 2
R 프로그램을 이용한 한국어 감성분석 예시

4) 챗봇 기반 과업 참여 만족도 설문 도구

챗봇 기반 과업 참여 만족도 측정에 활용된 세 번째 도구는 5점 리커트 척도에 기반한 설문조사였다. 설문지는 학습자의 배경정보 관련 문항들과 함께 다음의 두 문항을 포함하였다:

1. 미추쿠와의 대화가 영어 능력 향상에 어느 정도 도움이 될 수 있다고 생각하십니까?

- ① 전혀 도움이 되지 않는다. ② 거의 도움이 되지 않는다. ③ 보통이다.
④ 어느 정도 도움이 된다. ⑤ 매우 도움이 된다.

2. 인공지능 챗봇을 활용한 학습이 기존의 학습방법과 비교하여 흥미도나 참여도 유발에 어느 정도 도움이 된다고 생각하십니까?

- ① 더 나은 점이 없다. ② 거의 나은 점이 없다. ③ 보통이다.

④ 약간은 더 나은 것 같다. ⑤ 훨씬 나은 것 같다.

1번 문항은 챗봇 기반 과업 활동이 영어 학습효과에 어느 정도 기여할 수 있을지를 묻는 기대치 추정관련 질문이고, 2번 문항은 기존의 영어학습 방법에 비해 챗봇 기반 과업이 학습자의 흥미도나 참여도 유발에 어느 정도 기여했는지 대한 질문이다.

‘챗봇 대화 감성분석,’ ‘챗봇 과업 참여 소감 감성분석,’ ‘영어 학습효과 기대 정도 설문분석,’ ‘챗봇 대화의 흥미 및 참여도 유발 정도 설문분석’에 대한 관계 및 비교는 빈도분석과 상관분석을 통해 살펴보았고 추가로 ‘영어학습효과 기대치 설문분석’을 종속변인으로 하고 나머지 3개의 결과를 독립변인으로 하여 3개의 독립변인이 종속변인에 미치는 영향은 다중회귀분석을 통해 예측하였다.

IV. 연구 결과 및 논의

1. 챗봇 기반 과업 참여 만족도 분석에 대한 기술 통계

본 연구에서는 챗봇 기반 과업 참여 만족도를 비교·분석하기 위해 챗봇 사용자의 대화문에 대한 영문 감성분석, 소감문에 대한 국문 감성분석, 챗봇 기반 과업의 영어 학습효과 기대에 대한 설문문항 분석, 그리고 챗봇 기반 과업의 흥미도 및 동기유발 정도에 대한 설문문항을 실시하였고 그 결과의 기술 통계는 다음의 표 4과 같다.

표 4
챗봇 기반 과업 만족도에 대한 분석 도구별 기술 통계 결과

N=28	척도					M/SD
	1 (매우 부정)	2 (부정)	3 (중립)	4 (긍정)	5 (매우 긍정)	
대화문 감성	-	-	23(82%)	5(18%)	-	3.2/0.4
소감문 감성	2(7%)	2(7%)	6(21.5%)	6(21.5%)	12(43%)	3.9/1.2
영어 학습효과 기대치	-	-	4(14%)	17(61%)	7(25%)	4.1/0.6
흥미도 및 동기유발 정도	1(4%)	2(7%)	1(4%)	11(39%)	13(46%)	4.2/1.0

챗봇 대화문 분석의 경우 23명의 대화문이 중립적으로 나타났고 5명의 대화문은 긍정적으로 나타나 대화문에 나타난 감성은 매우 단조로웠다. 챗봇 기반 과업 참여에 대한 소감문에서는 12명의 소감문이 매우 긍정적으로 나타났고 6명은 긍정, 다른 6명은 중립으로 분석되었다. 부정과 매우 부정은 각각 2명으로 나타나 절반 이상의 학습자들은 긍정적인 태도를 가지고 있는 것으로 파악되었다. 표 5는 ‘매우 긍정’부터 ‘매우 부정’까지의 5단계별 소감문 예시이다.

표 5
챗봇 기반 과업 참여 소감문 분석 결과 예시

매우 긍정	s1: 영어권에 사는 친구와 채팅하는 것처럼 자연스러운 분위기에서 영어를 사용할 수 있다는 점에서 미추쿠와의 대화가 영어 능력 향상에 도움이 된다고 느꼈다. 군말이나 옆으로 세는 것이 없이 원하는 것을 바로 얻을 수 있다는 점은 좋은 것 같다. 몇몇 무례한 질문에 대해서는 대답하지 않도록 하는 것이 좋을 것 같다. 이곳은 질문에도 다 대답해주다보니 몇몇 아이들은 수업의 목표에 달성하지 못할 수도 있을 것 같다. 물어보는 모든 질문에 대부분 친절하게 대답해준 것이 어색했다. 실제 사람이었다면, 몇몇 질문에 대해서는 답변을 하지 않았을 것이다.
긍정	s2: 생각 이상으로 다양한 답변을 해주었다. 물론 답답하거나, 잘 대답하지 못하는 부분도 있었지만 그래도 이 정도면 기술의 발전이 대단하다 느꼈다. 영어로 누군가와 타자를 치면서 채팅을 자유롭게 하는 기회가 흔하지 않다. 이렇게 자연스러운 대화를 반복하다 보면 영어 실력이 분명 향상될 것이라고 생각한다. 상대방에 감정이나 반응을 보지 못하기 때문에 오히려 더 편할 수도 있다. 하지만 아직 인간의 감정이나, 정서 혹은 복잡한 부분을 표현하기에는 부족함이 있다고 생각한다.
중립	s3: 이해가 안 되는 부분이 있기는 했지만 대체로 의사소통에 문제가 없었다. 사람과 유사한 반응을 보여 흥미로웠다.
부정	s4: 미추쿠의 답변이 추가 질문을 유도하는 답변들이 많았다. 인간에 비해 넓고 다양한 지식을 가지고 있어 질문에 대해 원하는 답변을 얻을 수 있다. 질문과 연관성이 적은 답변을 하는 것, 가끔 Too much information을 제공해 사람과 하는 대화와 다르다고 느낄 때가 있었다.
매우 부정	s5: 말의 맥락을 정확히 파악하지 못하고, 평풍이 잘 안 되며, 관용구 이해 능력이 떨어진다. 미추쿠가 쓰는 표현은 비교적 기본적이고 정석적인 표현들을 사용한다고 생각하기 때문이다. 상대가 로봇이라는 점을 이미 알고 있으므로, 내 말을 받는 상대가 기분 상해 할 것을 염려하며 대화할 일이 거의 없는 것 같다. (이것이 교육현장에 잘못 적용된다면 아이들의 배려심 함양에 장애로 작용할 수도 있다는 생각 또한 든다) 관용구를 더 학습시켜야 할 것 같고, 모든 문장에 일일이 대답할 필요는 없다는 사실을 숙지시켜 주었으면 좋겠다.

영어 학습효과에 미치는 기대 정도를 묻는 설문문항에 대해서는 응답자의 25%는 매우 긍정적인 영향을 줄 것이라 예측했고 61%는 긍정, 나머지 14%는 중립으로 응답하였다. 흥미도 및 참여도 유발 효과에 대한 설문문항에서는 46%는 매우 긍정으로 응답하였고 39%는 긍정으로 나타나 학습자들은 전반적으로 챗봇 기반 과업이 흥미롭고 참여의 집중도를 유발하는데 기존의 영어학습 방법

에 비해 효과적이라고 느끼는 것으로 나타났다. 즉, 챗봇 과업이 영어 학습효과에 어느 정도 영향을 줄 것인가에 대해서 긍정적으로 예측한 학습자가 86%인 동시에 흥미도와 참여도 유발에 긍정적인 효과가 있다고 응답한 학습자는 85%라는 점을 고려해볼 때, 학습자들은 챗봇 과업에 대한 높은 학습효과를 기대하며 적극적으로 과업에 참여했다는 점을 알 수 있었다.

2. 챗봇 기반 과업 만족도에 대한 분석 도구 간 상관관계 분석

기술 통계 외에도 본 연구에서는 각 만족도 분석 도구에 따라 도출된 수치들이 어느 정도의 상관관계를 보이는지를 분석하여 표 6에 나타내었다.

표 6
챗봇 기반 과업 만족도 분석 도구 간 상관관계 분석 결과(Spearman's rho)

		대화	소감	학습효과 기대	흥미·참여도 유발
대화	<i>r</i>	1			
	<i>p</i> -value	-			
소감	<i>r</i>	0.486**	1		
	<i>p</i> -value	0.009	-		
학습효과 기대	<i>r</i>	0.696**	0.778**	1	
	<i>p</i> -value	0.000	0.000	-	
흥미·참여도 유발	<i>r</i>	0.472*	0.945**	0.778**	1
	<i>p</i> -value	0.011	0.000	0.000	-

Note. * $p < 0.05$, ** $p < 0.01$.

표 6의 결과를 보면 챗봇 기반 과업 참여 소감문에 대한 한국어 감성분석 결과와 흥미도 및 참여도 유발 효과에 대한 설문문항 관련 수치들 간의 상관도가 가장 높았다($r = .945$, $p < 0.01$). 이를 통해, 대체로 감정 표현이 보다 직설적으로 드러난 소감문의 분석 결과와 설문문항 간의 상관관계가 높은 것을 확인할 수 있었다. 반면 영문 대화문의 감성분석 결과는 국문 소감문 감성분석 결과와 상관계수가 0.486($p = 0.009$)으로 나타나 중간정도의 상관도를 보였고 설문문항과의 상관계수도 상대적으로 낮았다. 실제 상관관계분석이라는 것이 경향성의 유사도를 측정하는 것이고 대화문의 경우 앞서 살펴본 표 4의 기술 통계에서도 알 수 있듯이 중립적인 성향이 82%를 차지하다 보니 다양한 감정 단계의 결과가 포함된 타 분석 도구와의 상관도는 상대적으로 낮을 수밖에 없다고 사료된다. 하지만 모든 측정 도구에 따라 도출된 수치들 간에는 통계적으로 유의한 상관관계가 있는 것으로 나타났다.

3. 챗봇 기반 과업의 영어 학습효과 기대에 대한 예측 변인

마지막으로 챗봇 기반 과업이 영어 학습효과에 어느 정도 기여할 수 있을 지에 대한 학습자들의 기대치를 종속변수로 설정하고 나머지 3개의 만족도 분석 도구에 따라 도출된 수치들을 독립변수로 하여 다중회귀분석을 수행하였다. 즉 ① 챗봇 대화 감성분석, ② 챗봇 과업 참여 소감 감성분석, 그리고 ③ 챗봇 대화의 흥미 및 참여도 유발 정도 설문문항 관련 변인들 중에서 어느 것이 영어 학습효과 기대치를 예측하는데 가장 효과적인 지를 분석하였다. 다음의 표 7은 영어 학습효과 예측 위한 3개 변인에 대한 다중회귀분석 결과를 요약한 것이다.

표 7
영어 학습효과 기대치의 예측을 위한 3개 변인에 대한 다중회귀분석 결과

Model	Predictor Variables	Unstandardized Coefficients		Standardized Coefficients		
		<i>B</i>	<i>Std. Error</i>	β	<i>t</i>	<i>Sig.</i>
영어 학습 효과 기대	(Constant)	0.41	0.50		0.82	0.42
	챗봇 대화 감성분석	0.68	0.16	0.42	4.33	0.00
	챗봇 과업 참여 소감 감성분석	0.08	0.12	0.17	0.68	0.50
	챗봇 대화의 흥미 및 참여도 유발 정도 설문분석	0.29	0.14	0.49	2.04	0.05

$R^2 = 0.82$, *Std. Error* = 0.28

Note. * $p < 0.05$, *B*: 회귀계수, β : 영향력

표 7에서 알 수 있듯이 3개 만족도 분석 도구 중 영어 학습효과 기대치에 유의한 영향($\beta = 0.42$, $p < 0.001$)을 준 것은 챗봇 대화의 감성분석 뿐이었다. 분석 결과는 챗봇 대화의 감성분석 척도가 한 단계 오를 때 5점 척도의 영어 학습효과에 대한 기대치가 0.68단계 오른다고 해석할 수 있으며 결정계수인 R^2 값은 0.82로 나타나 3개의 독립변수가 종속변수 변화의 82%를 설명한다고 할 수 있다. 결과적으로 이는 타 분석 도구와의 상관관계수가 낮았던 챗봇 대화 감성분석의 경우, (다른 변인의 영향을 통계적으로 통제했을 경우) 영어 학습효과 기대치의 예측 도구로 활용될 수 있다는 것을 의미한다.

V. 결론 및 제언

본 연구는 챗봇 기반 과업이라는 특정 맥락 속에서 학습자들이 해당 과업에 대해 어떠한 인식을 가지고 있는지에 대해 알아보고자 하였다. 이를 위해 기존의 대다수 연구들에서 사용되는 리커트 척도 설문분석과 더불어 본 연구에서는 최근 교육학 분야에서 폭넓게 활용되는 감성분석 방법(Arguedas 외, 2018; Lin 외, 2014; Liu 외, 2019)을 활용하여 학습자가 챗봇과의 과업 중 작성한 영문 대화문과 과업 수행 후 작성한 국문 소감문을 추가로 분석하여 이들 간에 어떠한 관계가 있는지 살펴보았다. 이 중 소감문과 리커트 척도 설문문항의 결과에 따르면 학습자들은 챗봇 기반 과업에 대해 전반적으로 만족한다는 점을 확인할 수 있었으며, 이는 전반적으로 선행연구(양혜진 외, 2019)의 결과와 같았다.

본 연구는 물론 본 연구에서 수집·분석한 문헌의 범위로 제한되기는 하지만 Alm와 Nkomo(2020)의 연구와 더불어 영어교육 분야에서 감성분석을 기반으로 학습자의 인식을 살펴본 최초의 연구 중 하나라는 점에서 의의가 있으며, 그 결과를 바탕으로 향후 영어교육연구에 다음과 같은 방향을 제시할 수 있겠다. 우선, 소감문 분석(한국어 감성분석)의 경우 향후 기존의 리커트 척도의 대체방법으로 활용될 수 있을 것으로 기대된다. 리커트 척도의 경우 학습자가 연구자가 작성한 문항에 대해 어떻게 생각하는지에 대해 하나의 척도를 선택하도록 강요하는 방식이기 때문에 정확하게 학습자가 어떻게 특정 과업 또는 학습활동 등에 대해 인식하는지를 살펴보기에는 제한적일 수밖에 없다. 그러나 소감문의 경우, 학습자가 본인의 생각을 표현하는 데 있어 보다 융통성을 제공해 주며 연구자가 생각하지 못했던 감성을 가지고 있을 수도 있다는 점에서 기존 리커트 척도에 비해 좀 더 정확하게 학습자 인식을 파악하게 해 줄 수 있다. 또한, 여기에 감성분석을 적용할 경우 연구자의 주관에 따른 해석의 가능성을 줄이는 동시에 질적 자료를 정량화할 수 있다는 점에서 그 활용가치가 높다고 볼 수 있다. 그리고 과업 수행 후 작성한 소감문 뿐만 아니라 ‘think-aloud protocol(사고 발화프로토콜)’(Jääskeläinen, 2010)과 같이 학습자의 인식을 실시간으로 살펴보는 연구 도구 등을 사용하여 감성분석을 실시하는 방법도 향후 활용될 있을 것으로 기대된다.

영문으로 구성된 대화문은 학습자의 영어효과 기대치와 관련하여 가장 예측력이 높은 변인으로 확인되었는데, 이는 학습자가 해당 과업을 수행할 때 보여주는 감성의 수치가 향후 해당 과업에 대한 기대효과와 상관관계를 보여준다는 점을 의미한다. 따라서 학습자가 영어 출력(output)을 만드는 과업의 경우 감성분석을 활용하여 해당 학습자가 가지고 있는 과업에 대한 만족도, 혹은 영어실력 향상에 대한 기대치를 살펴볼 수 있을 것이다. 다만 본 연구의 경우 다른 수치들과의 직접적인 비교를 위해 분노, 두려움, 슬픔과 같은 감성을 부정적 감성으로 묶어서 살펴보았으며 해당 수치들을 5점 척도로 변환하였다는 점에서 한계를 가지고 있다고 볼 수 있다. 향후 연구에서는 이러한 단점을 보완하여 Tone

Analyzer(Bhuiyan, 2017)가 제공하는 7개의 감성분석 결과와 해당 수치들을 산출된 그대로 활용하여 좀 더 구체적으로 학습자의 인식과 만족도를 조사하는 연구가 필요하다고 생각된다.

참고문헌

- 김영우. (2020). 영어 학습을 위한 챗봇 및 챗봇 빌더 분석. *멀티미디어언어교육*, 23(4), 161-182.
- 김재상. (2017). *인간과 인공지능기기의 상호작용이 EFL 환경에서 초등학생들의 협업적 언어습득에 미치는 영향*. 석사학위논문, 광주교육대학교, 광주.
- 신동광. (2019). 영어 쓰기 능력 향상을 위한 AI 챗봇 활용 방안 탐색. *교원교육*, 35(1), 41-55.
- 신동광. (2020). 감성분석에 기반한 수능 영어영역 ‘심경·분위기 파악’ 문항의 타당도 분석 연구. *영어교과교육*, 19(4), 159-179.
- 양혜진, 김혜영, 신동광, 이장호. (2019). 인공지능 음성챗봇기반 초등학교 영어 말하기 수업 연구. *멀티미디어언어교육*, 22(4), 184-205.
- 장덕현, 조성겸. (2017). 리커트형 척도의 중간점은 필요한가?: 중간점 있는 척도와 없는 척도 간 측정 결과 비교. *조사연구*, 18(4), 1-24.
- 차수미, 김정렬, 남승우. (2021). 영어교육 관련 AI 챗봇 연구 논문의 동향 분석. *영어교과교육*, 20(1), 203-225.
- Abbasi, S., & Kazi, H. (2014). Measuring effectiveness of learning chatbot systems on student's learning outcome and memory retention. *Asian Journal of Applied Science and Engineering*, 3(7), 57-66.
- Alm, A., & Nkomo, L. M. (2020). Chatbot experiences of informal language learners: A sentiment analysis. *International Journal of Computer-Assisted Language Learning and Teaching*, 10(4), 51-65.
- Arguedas, M., Xhafa, F., Casillas, L., Daradoumis, T., Peña, A., & Caballé, S. (2018). A model for providing emotion awareness and feedback using fuzzy logic in online learning. *Soft Computing*, 22(3), 963-977.
- Bhuiyan, M. (2017). *Watson Tone Analyzer: 7 new tones to help understand how your customers are feeling*. Retrieved May 12, 2021, from the World Wide Web: <https://www.ibm.com/blogs/watson/2017/04/watson-tone-analyzer-7-new-tones-help-understand-customers-feeling/>.
- Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text & Talk*, 34(5), 545-567.
- Data Intelligence Lab. (2017). *KNU Korean Sentiment Lexicon*. Retrieved May 12,

- 2021, from the World Wide Web: https://github.com/park1200656/KnuSentiLex/blob/master/KnuSentiLex/SentiWord_Dict.txt.
- Elia, G., Solazzo, G., Lorenzo, G., & Passiante, G. (2019). Assessing learners' satisfaction in collaborative online courses through a big data approach. *Computers in Human Behavior*, 92, 589-599.
- Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8-14.
- Guinness World Records. (2020). *Most Loebner Prize wins*. Retrieved May 12, 2021, from the World Wide Web: <https://www.guinnessworldrecords.com/world-records/603076-most-loebner-prize-wins/>.
- Higuchi, K. (2016). *KH-Coder version 3* [Computer Software]. Retrieved May 12, 2021, from the World Wide Web: <https://github.com/ko-ichi-h/khcoder/releases/tag/3.Beta.03>.
- Jääskeläinen, R. (2010). Think-aloud protocol. In Y. Gambier & L. Van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1) (pp. 371-374). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1-2), 61-84.
- Lee, J. H., Yang, H., Shin, D., & Kim, H. (2020). Technology for the language teacher: Chatbots. *ELT Journal*, 74(3), 338-344.
- Lin, H.-C. K., Chen, N.-S., Sun, R.-T., & Tsai, I.-H. (2014). Usability of affective interfaces for a digital arts tutoring system. *Behaviour & Information Technology*, 33(2), 105-116.
- Liu, Z., Yang, C., Rüdian, S., Liu, S., Zhao, L., & Wang, T. (2019). Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums. *Interactive Learning Environments*, 27(5-6), 598-627.
- Mauldin, M. L. (1994). Chatterbots, TinyMUDS, and the Turing Test: Entering the Loebner Prize Competition. *Proceedings of the Eleventh National Conference on Artificial Intelligence* (pp. 16-21). Seattle, Washington: AAAI Press.
- Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31, 527-541.
- Picard, R. W. (1995). *Affective computing* (M.I.T Technical Report No. 321). Retrieved May 12, 2021, from the World Wide Web: <http://affect.media.mit.edu/pdfs/95.picard.pdf>.

- R Development Core Team (2011). *R: A language and environment for statistical computing* [Computer Software]. Retrieved May 12, 2021, from the World Wide Web: <http://www.R-project.org/>.
- Shen, L., Wang, M., & Shen, R. (2009). Affective e-learning: Using “emotional” data to improve learning in pervasive learning environment. *Educational Technology & Society*, 12(2), 176-189.
- Zhou, J., & Ye, J.-M. (2020). Sentiment analysis in education research: A review of journal publications. *Interactive Learning Environments*. Advance online publication. Retrieved May 12, 2021, from the World Wide Web: <https://doi.org/10.1080/10494820.2020.1826985>.

적용가능 수준 (Applicable Levels): Elementary/Secondary/Tertiary

신동광

광주교육대학교 영어교육과

(61204) 광주광역시 북구 필문대로 55

Tel: (062) 520-4211

Email: sdhera@gmail.com

이장호

중앙대학교 영어교육과

(06974) 서울특별시 동작구 흑석로 84

Tel: (02) 820-5122

Email: jangholee@cau.ac.kr

김혜영

중앙대학교 영어교육과

(06974) 서울특별시 동작구 흑석로 84

Tel: (02) 820-5399

Email: englishnet@cau.ac.kr

양혜진

중앙대학교 영어교육과

(06974) 서울특별시 동작구 흑석로 84

Tel: (02) 820-5399

Email: hjyang1112@gmail.com

질적 자료의 정량화 방안: 감성분석 기법을 활용한 챗봇 기반 과업활동의 만족도 분석 중심으로 185

Received June 30, 2021

Revised August 5, 2021

Accepted August 22, 2021