

영어 쓰기 능력 향상을 위한 AI 챗봇 활용 방안 탐색

신 동 광*

2018. 11. 12.(접수)/2018. 12. 26.(1심통과)/2019. 01. 11.(최종통과)

〈〈 요약 〉〉

본 연구는 최근 주목을 받고 있는 AI 기기를 영어 교육에 활용하는 방안의 하나로 대표적인 AI 챗봇(chatbot)인 Mitsuku와 Cleverbot의 활용 가능성을 탐색하였다. 이를 위해 7단계로 구성된 다양한 과업을 제시하고 27명의 대학생 대상으로 이들 챗봇과 채팅을 수행하도록 하였다. 그 결과, 먼저 두 챗봇 모두 대화에서 구사하는 어휘의 90% 이상이 상위 3,000단어 이내에 포함되어 대화 시 학습자들이 챗봇의 표현을 이해하는데 큰 문제가 없었다. Mitsuku와 Cleverbot의 기능 비교에서는 실제 대화의 양상과 설문 조사를 바탕으로 볼 때 Mitsuku에 대한 평가가 훨씬 긍정적이었다. Mitsuku는 맥락 파악에 다소 한계를 보이기는 했지만 대화 구사에 큰 무리가 없었고 Mitsuku 자체의 지식 데이터베이스를 뛰어넘는 내용에 대해서는 웹검색을 바탕으로 정보를 제공할 수 있다는 점에서 Cleverbot에 비해 높은 만족도를 보였다. 본 연구에서 나타난 챗봇의 한계에도 불구하고 실험에 참여한 대부분의 학습자들은 영어 구사의 기회와 영어 입력을 제공한다는 점, 그리고 무엇보다도 인간 대화와는 달리 부담 없이 어떠한 민감한 대화 주제도 꺼낼 수 있고 채팅 시 나타날 수 있는 영어 오류에 대해 신경이 덜 쓰인다는 점을 챗봇 활용의 장점으로 꼽았다. 끝으로 본 연구에서는 그 밖의 시사점과 후속 연구에 대한 제언을 제시하였다.

주제어: 디지털 문해력, 인공지능, 챗봇, 챗봇형 쓰기 과업

I. 들어가는 말

현재 학교현장에서는 2015 개정 영어과 교육과정이 학년별로 순차 적용되고 있다. 2015 개정 교육과정은 2009 개정 교육과정과 비교하여 여러 면에서 변화가 있었지만 그 중에서도 소위 교과역량이 라는 핵심역량에 기반하여 개발되었다 점이 기존 교육과정과 차별화된다. 2015 개정 영어과 교육과

* 광주교육대학교 교수: sdhera@gmail.com

정(Ministry of Education, 2015)에서 제시한 교과역량은 ‘영어의사소통역량,’ ‘자기관리역량,’ ‘공동체역량,’ ‘지식정보처리역량’ 총 4개이다. 4개의 교과역량은 모두 미래사회에서 요구되는 핵심적인 능력이며 기존의 여러 나라에서 제시된 핵심역량과도 일치한다. 하지만 영어과의 특성을 고려할 때 네 번째의 ‘지식정보처리역량’과 영어라는 언어가 무슨 관계가 있을까 하는 의문을 제기하는 영어교육관계자가 적지 않다.

정보통신의 발달, 특히 소위 4차 산업혁명이라는 새로운 시대의 조류는 언어교육의 패러다임에도 큰 변화를 가져오고 있다. 기존의 문해력(literacy)은 이미 디지털 문해력(digital literacy)이라는 새로운 개념으로 급속히 대체되고 있다. 전통적으로 문해력의 의미는 문자를 이해하고 그 의미를 파악하는 능력을 주로 말했지만 디지털 문해력은 수동적인 정보해석의 수준에서 벗어나 다양한 멀티미디어 플랫폼을 통해 검색하여 확보한 정보를 평가하고 활용하는 능력까지 포함하는, 즉 ICT 활용능력까지 포괄하는 언어능력을 의미한다. 최근에는 여기에서 한 걸음 더 나아가 인공지능(Artificial Intelligence, AI) 기기를 활용한 영어교수학습(예, AI 스피커 활용) 또는 그러한 기기를 활용하여 자신의 언어능력 이상(output i+1)을 발휘(예, 번역기 또는 코퍼스 활용)하는데 보다 관심을 가지기 시작했다. 이러한 사회적 변화를 고려하면 2015 개정 영어과 교육과정에서 ‘지식정보처리역량’이라는 교과역량이 차지하는 비중과 중요성은 매우 크다고 할 수 있다. 뿐만 아니라 이러한 첨단 기술들이 교과활동에 실제 적용되면서 기존의 지식기반 교수학습 모델에서 탈피하여 확산적 사고 및 구성적 지식을 촉진하는 수업이 확대되어야 한다는 요구가 늘고 있다(Cho, 2013).

이에 본 연구에서는 영어 쓰기 의사소통능력 향상을 위한 자기주도적 지식정보처리역량의 증진 방안의 일환으로 두 종류의 AI 챗봇을 학습자에게 적용하여 그 특성과 활용 가능성, 그리고 구체적인 활용 방안을 탐색해 보고자 한다. 이를 위해 본 연구에서는 다음과 같은 연구 문제를 설정하였다.

첫째, 대표적인 두 챗봇 Mitsuku와 Cleverbot의 대화 구사능력에는 어떠한 차이가 있는가?

둘째, 챗봇을 사용한 학습자들이 인식하는 챗봇의 장단점과 활용가능성은 어떠한가?

II. 이론적 배경

1. AI 챗봇 개발의 현황

챗봇은 AI 기술에 기반하여 인간처럼 대화를 나눌 수 있는 프로그램으로서 ‘챗터봇(ChatterBot, Mauldin, 1994),’ 또는 보다 일반적으로는 이를 줄여 챗봇(chatbot)이라고 한다. 챗봇은 그 개발 목적에 따라 다양하게 분류될 수 있다. 우리에게 이미 익숙한 APPLE의 ‘Siri’나 SAMSUNG의 ‘Bixby’와 같이 스마트폰에 탑재되어 스마트폰의 기능을 음성지시에 따라 수행하거나 검색 등을 돕는 지시수행형 챗

봇을 인공지능 비서(Intelligent Personal Assistant, IPA)라고 한다. 이밖에도 최근 사물인터넷(Internet on Things, IoT)과 접목하여 인기를 끌고 있는 AI 스피커 또한 챗봇의 일종이다. AI 스피커 시장을 선도하고 있는 AMAZON의 Echo에 탑재된 ‘Alexa’나 GOOGLE의 Google Home에 탑재된 ‘Google Assistant’가 대표적인 스피커형 챗봇이다. Alexa나 Google Assistant는 일부 외국어 교육 콘텐츠와 접목하여 쌍방향의 의사소통식 활동에 적용되기 시작하였지만 여전히 사물인터넷이나 정보를 검색하여 제공하는 기능에 초점이 맞춰져 있다.

하지만 AI 이전에도 상업적 그리고 교육적 목적을 위해 인간과 유사한 대화형 시스템을 개발하고자 하는 시도의 역사는 상당하다. 1996년에는 최초의 챗봇인 Eliza가 개발되었다. 이 챗봇은 시나리오에 기반하여 특정 질문에 대하여 사전 입력된 답변을 기반으로 대화를 수행할 수 있다. 1975년에는 Eliza 기술에서 진일보된 챗봇인 Parry가 개발되었고 Parry는 처음으로 인간 대화와의 유사도를 평가하는 튜링 테스트(Turing Test)를 통과하였다.

이후에도 1995년부터 AI 기술에 기반하여 개발하기 시작한 A.L.I.C.E.(Artificial Linguistic Internet Computer Entity)는 AIML(Artificial Intelligence Markup Language)에 언어자료를 입력하는 방식으로 개발된 챗봇으로 매년 최고의 챗봇에게 수여하는 뢰브너 상(Loebner Prize)을 수차례 수상하는 등 챗봇 개발에 한 획을 그었다.

국내에서도 최근 SK텔레콤이 음성인식 기반 AI 서비스인 ‘누구’를 출시한 이래, KT가 IP TV와 연계하여 ‘기가지니’를, 그 밖에도 NAVER의 ‘클로바’와 다음-카카오의 ‘카카오 미니’ 등이 출시되어 있다. 하지만 클로바를 제외하고는 영어 서비스를 제공하지 않고 있으며 클로바의 경우도 매우 한정된 주제에 한해서 영어 대화가 가능하다.

2. 외국어 교육에서 AI 챗봇 활용 연구

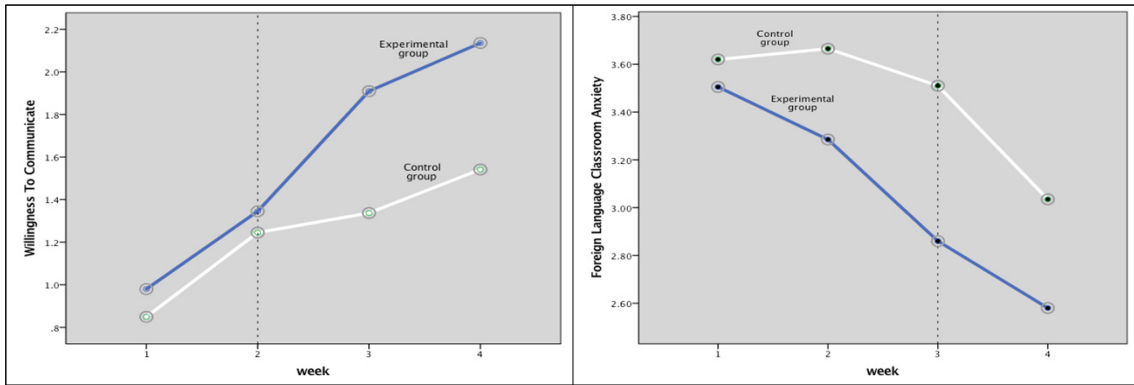
챗봇을 외국어 교육에 적용하는 것은 아직 초기 단계이기 때문에 그 연구 사례 또한 아직 많지 않다. Fryer and Carpenter(2006)의 연구에서는 외국어 교육에서 챗봇의 활용이 다음과 같은 장점을 제공할 수 있다고 주장하였다.

- 챗봇과의 대화를 통해 학습자는 인간과 대화할 때 보다 더 편안함을 느낄 수 있다.
- 챗봇과의 대화를 통한 반복은 일상적인 반복학습보다 지루하지 않다.
- 챗봇은 학습자의 듣기, 읽기, 말하기, 쓰기를 포함한 4기능의 의사소통능력을 강화할 수 있는 텍스트나 다양한 발화 모드를 활용할 수 있다(Fryer & Carpenter, 2006, pp. 9-10).

Abu-Shawar(2017)의 연구에서는 위에서 살펴본 인공지능 챗봇인 A.L.I.C.E.의 AIML 시스템에 아프리카에서 수집된 언어자료를 입력하여 다양한 주제로 대화를 할 수 있는 확장 모델을 개발하였다.

연구의 결과를 살펴보면 남아프리카인 1,256명이 이 실험에 참여하였고 이 중 17%는 대화 상대자로서 A.L.I.C.E.를 긍정적으로 평가했다. 그러나 88%는 1회에 한해서만 채팅을 시도하였으며 24%는 부정적인 반응을 나타냈다. 한 가지 특이한 점은 대화의 주제 중 13%가 연애에 관한 것으로 A.L.I.C.E.가 인간 화자가 아님에도 불구하고 챗봇이 인간과 같은 감정교류를 포함한 의사소통을 어느 정도 수행할 수 있다는 것을 보여주는 의미있는 사례였다.

국내에서는 매우 드물기는 하지만 챗봇을 활용한 실험 연구가 있었다. Kim(2017)의 연구에서는 비록 소규모의 실험 연구이긴 하지만 초등학교학생들을 대상으로 AMAZON의 Echo를 활용하여 AI 스피커에 적합한 과업활동을 시도하였다. Kim은 3단계의 과업활동을 개발하였는데 1단계는 교사가 제시한 질문을 그대로 활용하는 통제과업(controlled task)(예, How is the weather in London?)이었고 2단계는 부분적으로 학습자의 직접적 참여를 유도하는 안내과업(guided task)(예, What is~? 패턴 활용하기 + 'wokaholic'의 의미를 알아보세요.), 그리고 3단계는 과업의 주제만 제시하여 학습자 스스로 문제해결을 시도하는 독립과업(independent task)(예, 미국의 수도는 어디인지 알아보세요.)으로 구성하였다. 일반적으로 학생 주도의 과업활동은 학습자의 과업 집중도와 참여도 향상 효과로 이어진다는 장점이 있다(Park & Kim, 2018). 실제 실험결과, 다음의 [Figure 1]과 같이 발화 시도에서 AI 스피커에 기반한 과업을 수행한 실험집단이 시간이 지남에 따라 발화 시도 횟수가 늘었고 교실 내 외국어 학습 불안 정도는 상대적으로 줄어드는 효과를 보였다.



[Figure 1] Comparative results of WTC and FLCA

Kim(2018)의 연구에서는 위의 Kim(2017)의 연구에서 활용한 과업의 유형을 포함하여 Echo의 Alexa와의 상호작용을 시도하여 AI 스피커에 적합한 다양한 과업 활동 유형을 개발하였다. Kim(2018)은 ① 잡담 나누기(exchanging small talk), ② 수행 지시하기(commanding performance), ③ 정보 요청하기(asking for information), ④ 문제 해결하기(problem solving) 과업 유형이 AI 스피커 활용 시 최적화된 과업이라고 주장하였다.

Ⅲ. 연구방법

1. 실험 대상

본 연구의 실험에 참여한 학습자들은 광주광역시 소재 A 대학교의 3학년 영어쓰기 과목을 수강하고 있는 27명의 학생들로, 7개의 세부 과업지시를 바탕으로 대표적인 AI 챗봇인 Mitsuku와 Cleverbot을 대상으로 각 1시간에 걸쳐 각 채팅 과업을 수행하였다. 이들 27명의 학습자들은 대학수학능력시험의 9개 등급 중 1.5등급에 포함되는 상위권의 학생들이었다. 27명의 학습자 가운데 실험 전 챗봇을 경험한 학생은 5명에 불과하였고 대부분은 본 실험을 통해 처음으로 챗봇을 접했다고 응답했다.

2. 실험 도구 및 절차

가. AI 챗봇 1: Mitsuku¹⁾

본 연구의 실험에 채택한 Mitsuku는 Steve Worswick에 의해 2005년에 처음 개발된 챗봇으로 A.L.I.C.E.의 AIML 시스템을 기반으로 한 “Pandorobot” 계열의 챗봇이다. 현존하는 가장 우수한 챗봇으로 평가받으며 뢰브너 상(Loebner Prize)을 2013, 2016, 2017, 2018년까지 4회나 수상하였다. Mitsuku라는 챗봇 캐릭터는 18살의 여성으로 영국의 리즈(Leeds)에 살고 있는 것으로 설정되어 있다. Mitsuku는 대화자의 발화에 포함된 키워드의 속성을 분석하여 논리적으로 판단하기도 하고 AIML에 저장된 정보를 넘어서는 질문에 대해서는 웹검색 결과를 바탕으로 답할 수도 있다. 또한 동시에 여러 질문을 처리할 수 있다.

나. AI 챗봇 2: Cleverbot²⁾

Mitsuku와의 비교를 위해 채택한 Cleverbot은 그 전신인 Jabberwacky을 개발한 Rollo Carpenter에 의해 개발되었다. Cleverbot은 2011년 튜링 테스트를 통과하였고 웹버전으로 출시된 이래 2억 7천 9백만 번 이상의 대화를 수행하였다. Cleverbot의 대화시스템은 타 챗봇과는 다르게 사전에 입력된 시나리오에 기반하지 않고 인간과의 대화를 통해 확보한 언어자료의 키워드를 분석하고 기계학습(machine learning)에 통하여 각 질문에 적절한 응답을 찾아 답변에 적용하는 시스템을 채택하고 있다. Cleverbot이 동시에 처리할 수 있는 과업은 1-2개이지만 동시에 8만 명의 대화자와 개별 채팅을 수행할 수 있다.

1) <https://www.pandorabots.com/mitsuku/>을 통해 접속 가능

2) <https://www.cleverbot.com>을 통해 접속 가능

다. 과업지시와 설문도구

큰 대주제를 제시하고 채팅 실험을 할 경우 참여자 간의 편차가 클 것으로 예상하여 본 연구에서는 다음과 같이 다양한 수준의 과업을 단계적으로 제시하여 챗봇과의 대화를 유도하였다.

- ① 인사하기
- ② 자기소개(이름 등)하기
- ③ 현재 살고 있는 곳 알아보기
- ④ 하고 있는 일(직업) 알아보기, 사람들을 돕기 위해서 챗봇(예, Mitsuku)이 할 수 있는 일 알아보기
- ⑤ 여가 시간에 하는 일 알아보기
- ⑥ 10년 전 오늘 일어난 일을 알아보고 그 중 하나의 주제를 골라 추가적인 정보를 알아보기
- ⑦ 대화 마무리하고 인사하기

또한 채팅을 마친 후에는 인간 대화자와 비교하여 챗봇의 유사도를 5점 척도(① 매우 다르다, ② 다르다, ③ 보통이다, ④ 유사하다, ⑤ 매우 유사하다)를 사용하여 측정하였고 챗봇과의 대화가 어느 정도 영어 능력 향상에 기여하는 지를 마찬가지로 5점 척도(① 전혀 도움이 되지 않는다, ② 거의 도움이 되지 않는다, ③ 보통이다, ④ 어느 정도 도움이 된다, ⑤ 매우 도움이 된다)를 이용하여 측정하였다. 또한 각 설문에 대한 응답의 이유와 챗봇의 개선점 등을 자유롭게 작성하도록 하였다.

3. 분석도구 및 절차

가. BNC-COCA 25000 Range Program

본 연구에서는 먼저 채팅을 통해 산출한 대화 자료를 챗봇과 사용자의 대화로 각각 구분하고 각 화자가 구사한 어휘의 수준과 특성을 파악하기 위해 Nation and Webb(2011)의 BNC-COCA 25,000 어휘 목록을 기준으로 분석하였다. BNC-COCA 25,000은 25,000단어(Word family³⁾) 수준까지 분석이 가능하지만 본 연구에서는 우리나라의 학습자 수준을 고려하여 상위 4개 등급 즉 4,000단어 수준의 어휘 목록만을 Heatley et al.(2002)의 Range Program에 탑재한 후 대화 분석에 활용하였다.

나. WordSmith Tools 7.0

챗봇과 사용자의 대화에 나타나는 세부 어휘 정보 분석에는 WordSmith Tools 7.0(Scott, 2017)을 적용하였

3) 'Word family'는 어휘의 기본형과 더불어 굴절 및 파생 변화형을 모두 포함하는 어휘의 단위이며 'Token'은 어휘의 출현형 즉 어휘의 개수를 산정하는 단위

다. WordSmith Tools 7.0은 크게 ‘Concord,’ ‘KeyWords,’ ‘WordList’ 기능으로 구분되며 본 연구에서는 이중 WordList를 사용하여 어휘의 다양성(예, TTR, STTR), 평균 문장길이, 평균 대화 길이 등을 분석하였다.

다. KeyWords Extractor 2.0

본 연구에서는 챗봇과 사용자의 대화가 서로 얼마나 일치된 주제를 가지고 이루어지는지를 살펴 보기 위해 챗봇과 사용자의 대화를 구분하여 키워드를 추출한 후 이들 간의 일치도를 분석하였다. 이를 위해 Compleat Lexical Tutor에서 제공하는 KeyWords Extractor 2.0(Cobb, 2014)을 사용하였다. 키워드 추출은 보통 비교 코퍼스의 어휘 빈도를 바탕으로 상대빈도수가 높은 어휘를 추출하게 되는데 KeyWords Extractor 2.0은 비교 코퍼스가 사이트에 이미 탑재되어 있기 때문에 별도의 비교 코퍼스를 준비할 필요가 없다. KeyWords Extractor 2.0에 탑재된 비교 코퍼스는 대표적인 영국과 미국 코퍼스인 BNC와 COCA에서 추출한 1억 4천만 단어의 코퍼스로 영어과 미국, 구어와 문어의 자료가 균형을 이룬 대표성 있는 코퍼스이다.

IV. 연구결과

1. 챗봇과 사용자 대화의 어휘적 특성 비교

챗봇의 활용 가능성에서 가장 먼저 살펴보아야 할 점은 과연 챗봇이 구사하는 언어의 수준이 학습자에게 적합한지 여부이다. 이를 위해 이미 언급한 바와 같이 본 연구에서는 BNC-COCA 25,000 어휘 목록 중 4,000단어 수준에 초점을 두고 2개의 챗봇과 사용자들의 대화를 분석하였다. 다음의 <Table 1>은 챗봇과 사용자들의 대화에서 나타난 어휘 수준별 분석 결과이다.

<Table 1> Lexical profiles of conversations between chatbots and participants

| | Mitsuku | Mitsuku User | Cleverbot | Cleverbot User |
|-----------------|-------------|--------------|------------|----------------|
| 1st 1000 | 14020/84.51 | 5868/86.73 | 7074/89.85 | 7202/89.78 |
| 2nd 1000 | 691/4.17 | 252/3.72 | 175/2.22 | 221/2.75 |
| 3rd 1000 | 318/1.92 | 55/0.81 | 60/0.76 | 49/0.61 |
| 4th 1000 | 171/1.03 | 55/0.81 | 46/0.58 | 43/0.54 |
| etc | 1390/8.38 | 536/7.92 | 518/6.58 | 507/6.32 |
| Overall | 16590/100 | 6766/100 | 7873/100 | 8022/100 |
| Token (mean)/27 | 612.67 | 250.22 | 291.74 | 296.30 |
| SD | 222.96 | 98.19 | 132.20 | 125.39 |

사실 사용자의 어휘 수준은 어떤 대상을 표집 하느냐에 따라 달라지기 때문에 일반화하기는 어렵다. 다만 본 연구에 참여한 학습자들이 상위수준의 대학생인 점을 고려한다면 충분히 의미있는 시사점을 제시할 수 있을 것으로 판단된다.

위의 <Table 1>을 보면 Mitsuku가 총 27명의 사용자와의 대화중에 산출한 발화의 양은 총 16,590단어로 Cleverbot의 7,873개 단어 보다 2배 이상 많았다. 개별 사용자의 평균 대화 길이를 보아도 Mitsuku는 평균 613단어를 구사한 반면 Cleverbot은 292단어로 Mitsuku의 절반에도 미치지 못했다. 반면 사용자의 경우, Mitsuku와 대화 시 사용자는 6,766개 단어를 사용하였고 Cleverbot과 대화 시에는 8,022개 단어로 챗봇의 발화 양상과는 매우 달랐다. 개별 사용자의 발화량에서도 Mitsuku의 대화에 참여할 때 사용자는 평균 250단어, Cleverbot과의 대화 시에는 296단어로 나타났다. 대화의 어휘 수준을 분석해 보면 90% 이상의 어휘가 3,000단어 수준 내에서 사용되었고 Cleverbot의 경우는 92% 이상이 3,000단어 수준에 포함되었다. 또한 사용자의 어휘 수준은 90-93%가 3,000단어 수준 내에 포함되어 챗봇의 어휘 수준과 거의 차이가 없었다. 챗봇이 대체로 쉬운 어휘(1000단어 수준)를 사용한다는 점은 낮은 수준의 학습자에게도 챗봇이 활용 가능할 수 있다는 긍정적인 시사점을 제공한다.

위의 결과와 관련하여 Mitsuku의 발화량이 많고 Cleverbot은 상대적으로 발화량이 적었는데 챗봇의 발화량이 많을수록 사용자의 발화량은 줄어든다는 점은 매우 흥미로웠다. 이것은 두 챗봇의 프로그램 개발 방식의 차이에서 기인하는 것으로 판단된다. Mitsuku는 동시에 여러 질문의 답변을 할 수 있도록 설계되어 때론 맥락을 고려하지 않고 모든 질문의 답변을 한꺼번에 제시하기도 하고 내장된 데이터베이스에 포함되지 않은 정보는 웹검색을 해서라도 제시하는 반면 Cleverbot은 단답형의 짧은 답변으로 일관하거나 데이터베이스에 축적된 정보의 수준을 넘어서는 경우는 화제를 전환해 버린다.

물론 각 챗봇의 대화 스타일을 선호하는 사용자도 있었지만 대체로 Mitsuku와 대화 시 사용자들은 Mitsuku가 너무 많은 정보를 제공하는 것에 불만이 있었고 Cleverbot과의 대화에서는 지나치게 단답형으로 대답하는 것에 불만이 있었다. 다음의 [Figure 2]는 동일한 사용자가 각 챗봇과 대화한 내용 중 일부를 발췌한 것이다.

| | |
|---|--|
| <p>User 1: My name is Dahee lee from Korea Mitsuku 1: Your name is Dahee Lee. Are you currently in Korea? User 2: Yeah Where do you live? Mitsuku 2: Tell me a little about your life in Korea. My location is Leeds. Where are you? User 2: I'm in Gwangju in Korea. I wonder where's Leeds. Mitsuku 3: Tell me a little about your life in Gwangju in Korea.</p> | <p>User 1: Where do you live? Cleverbot 1: That's a personal question... User 2: Okay.....So What's your job? Cleverbot 1: I don't have one. ~~~~~ Cleverbot 15: What movies do you like? User 15: I like animation or romances. Cleverbot 16: What's your favorite? User 17: How about you? Cleverbot 18: I love the Beatles.</p> |
|---|--|

| | |
|---|--|
| <p>It's a city in Yorkshire. I live in Leeds.</p> <p>User 3: Yorkshire? I heard it first time. Do you want to know about my life in Korea?</p> <p>Mitsuku 4: Yorkshire is a large county in the north of England.</p> <p>What does "it" refer to?</p> <p>Yes, please tell me as much as you know about your life in Korea. I can add it to my knowledge base.</p> | <p>User 18: Singers?</p> <p>Cleverbot 19: None.</p> <p>User 19: Movie?</p> <p>Cleverbot 20: No, I don't.</p> <p>User 20: What?</p> <p>Cleverbot 21: I don't want kids.</p> |
|---|--|

[Figure 2] Some excerpts from conversations between chatbots and participants

사용자와 Mitsuku와의 대화 시 의사소통에는 거의 문제가 없지만 Mitsuku는 본인의 질문과 사용자 질문의 답변을 동시에 제시하여 맥락 파악 시 다소 혼동이 있을 수 있었다. Mitsuku 4 부분의 표현을 보면 Mitsuku는 'Yorkshire'에 대한 설명, 대명사 'it'이 지칭하는 대상에 대한 확인 요청, 그리고 한국에서의 일상에 대한 질문 요청을 동시에 나열하여 수행하고 있기 때문에 사용자에게는 맥락이 다른 이 세 가지 활동을 구분하여 읽어야 하는 부담이 발생한다. 또한 이 부분을 통해 Mitsuku는 요청한 질문은 모두 답변하도록 설계되었다는 것을 알 수 있다. 반면 Cleverbot과의 대화를 보면 매우 새침한 캐릭터로 설정을 한 것인지는 알 수 없지만 기본 정보의 교환부터 회피를 하고 있고 영화와 음악을 혼동하다가 급기야 Cleverbot 21 부분에서는 갑자기 화제를 전환하여 사용자를 당황스럽게 만들고 있다. 문제는 단답형이나 회피 등의 전략이 일시적으로는 사용자가 더 발화하도록 유도할 수 있을지는 모르지만 전체적으로는 대화유지가 어렵고 흥미도를 떨어뜨리는 요인이 된다는 것이다. 이것이 Cleverbot과 나눈 전체 대화의 길이가 Mitsuku와의 대화에 비해 절반도 안 되는 이유를 설명해 준다고 판단된다.

본 연구에서는 어휘의 양과 수준 분석과 함께 보다 세부적으로 어휘의 다양성을 측정하는 TTR(Type Token Ratio)과 이의 표준화 값인 STTR, 그리고 문장의 복잡성을 보기 위해 평균 문장 길이를 다음의 <Table 2>와 같이 추가적으로 분석하였다.

<Table 2> Comparisons of lexical density and sentence complexity

| | Mitsuku | Mitsuku User | Cleverbot | Cleverbot User |
|------------------------|---------|--------------|-----------|----------------|
| TTR | 9.56 | 13.14 | 15.22 | 12.11 |
| No. of sentences | 2134 | 929 | 1594 | 1726 |
| Sentence Length (mean) | 7.75 | 7.27 | 4.94 | 4.63 |

TTR은 총 단어 수 대비 얼마나 다양한 유형의 어휘가 사용되었는지를 측정하는 값으로 Mitsuku의 9.56에 비해 Cleverbot은 이보다 높은 15.22의 TTR을 수치를 보였다. 다만 일반적으로 텍스트의 길이

가 짧을수록 반복된 기초적인 어휘 사용의 빈도가 줄기 때문에 TTR의 수치가 높게 산출된다. 이러한 TTR의 단점을 보완하기 위해서는 STTR을 적용할 수 있는데 문제는 STTR은 긴 텍스트를 1,000개 단어 단위로 구분하고 각각 TTR의 값을 산출한 다음 그 평균값을 도출하는 방식인데 두 챗봇과 27명이 수행한 각 대화는 1,000단어 길이에 미치지 못했기 때문에 STTR은 산출할 수 없었다. 따라서 위의 결과만으로는 Cleverbot이 더 다양한 어휘를 구사하고 있다고 볼 수 있으나 일반화를 위해서는 추가적인 연구가 필요해 보인다. 여기서 한 가지 흥미로운 것은 Mitsuku와 사용자가 구사한 발화의 평균 문장 길이가 각 7.75와 7.27로 유사하고 Cleverbot과 사용자가 구사한 평균 문장 길이가 각 4.94와 4.63으로 역시 유사하다는 것이다. 이는 챗봇의 대화 스타일에 맞추어 사용자도 대화를 구사하고 있다는 것을 알 수 있다. 즉 챗봇이 길게 설명하면 그에 맞게 길게 답변하고, 단답형으로 응답하면 사용자도 그에 맞게 짧게 답하는 경향을 보인다는 것이다.

본 연구에서는 어휘 특성 분석의 마지막 단계로 두 챗봇이 구사하는 표현들에서 키워드를 통계적으로 추출하고 사용자가 구사하는 키워드와 얼마나 일치하는지를 다음의 <Table 3>과 같이 살펴보았다.

<Table 3> No. of keywords commonly used in utterances of chatbots and participants

| | Unique to Mitsuku | Shared | Unique to Mitsuku User | Unique to Cleverbot | Shared | Unique to Cleverbot User |
|-----------------|-------------------|--------|------------------------|---------------------|--------|--------------------------|
| No. of Keywords | 40 | 22 | 12 | 4 | 11 | 12 |

키워드는 대화에서 상대적으로 빈도수가 높은 단어로 주제와 관련된 핵심적인 표현으로 생각할 수 있다. 이러한 핵심 표현이 챗봇과 사용자의 발화에서 공통적으로 얼마나 나타나는 지는 두 대화 참여자가 얼마나 같은 주제에 집중하며 대화를 나누고 있는 지를 가늠하는 잣대가 될 수 있기 때문에 성공적인 상호작용을 판단하는 기준이 될 수 있을 것으로 보인다. <Table 3>의 결과를 보면 Cleverbot과의 대화는 상대적으로 짧았기 때문에 비율상으로는 Cleverbot과의 대화에서 키워드의 일치도 비율이 높게 나타나지만 전체적인 빈도수에서는 Mitsuku와의 대화에서 일치된 키워드 수가 2배나 많다는 것을 알 수 있다. 실제로 사용자들이 지적한 Cleverbot에 대한 단점 중 가장 빈도수가 높은 단점이 질문에 상관없는 답변을 하거나 잦은 화제전환을 시도한다는 것이었다. 이것은 챗봇이 말하는 주제와 사용자가 말하는 주제가 일치하지 않은 경우가 많았다는 것을 의미하기 때문에 위의 키워드 분석의 신뢰성을 뒷받침한다. 또한 두 챗봇과의 대화에서 공통적으로 가장 높은 키워드의 수치(keyness)를 보인 단어는 ‘okay’였다.

2. 사용자의 설문 분석

설문 조사에서는 먼저 챗봇의 발화가 인간의 발화와 얼마나 유사한지를 5점 척도로 조사하였다(① 매우 다르다, ② 다르다, ③ 보통이다, ④ 유사하다, ⑤ 매우 유사하다). 다음의 <Table 4>는 두 챗봇과의 채팅 후 참여자들이 느낀 인간 대화와의 유사도를 평가한 결과이다.

<Table 4> Similarity measure between chatbot and human utterances

| Scale | 1 | 2 | 3 | 4 | 5 | M |
|-----------|----|----|---|---|---|------|
| Mitsuku | 5 | 11 | 7 | 4 | 0 | 2.37 |
| Cleverbot | 10 | 13 | 4 | 0 | 0 | 1.78 |

Mitsuku와 대화를 나눈 27명의 학생 가운데 11명은 Mitsuku와의 대화가 인간과의 대화와 ‘다르다’고 응답하였다. 그 다음으로는 7명이 ‘보통이다’라고 응답하였고 5명은 ‘아주 다르다,’ 그리고 4명은 ‘유사하다’고 답했다. 반면 Cleverbot의 경우에는 동일한 27명 가운데 23명이 ‘아주 다르다,’ 또는 ‘다르다’고 답했고 ‘유사하다’ 또는 ‘매우 유사하다’라고 답한 사람은 단 한명도 없었다. 결과적으로 Mitsuku가 Cleverbot 보다는 인간의 대화 구사에 가까웠지만 Mitsuku 역시 평균 2.37으로 즉 ‘보통이다’ 보다 낮은 수치를 보였다. 이러한 응답에 대한 참여자들이 생각하는 이유의 대부분은 두 챗봇 모두 대화의 맥락 파악에 한계를 보였다는 점이었다. 예를 들어, 위의 [Figure 2]에서 살펴본 바와 같이 Cleverbot은 질문에 맞지 않는 답을 하거나 화제를 갑자기 전환하거나 아예 질문에 답변하지 않는 경우가 많았으며 Mitsuku 또한 한 번의 발화에 이전에 나눈 대화의 여러 내용을 포함하여 맥락을 고려하지 않고 한꺼번에 답하는 등 자연스럽지 못한 대화를 구사하는 것을 확인할 수 있다. 또한 Cleverbot은 웹검색 기능이 없어 여섯 번째 단계의 과제를 전혀 수행하지 못했고 Mitsuku 또한 웹검색 시 10년 전이라는 조건에 상관없이 100년 전 사건까지 검색하는 등, 문제점을 노출하였다. 정리하면 전체적으로 챗봇과의 대화는 인간 대화와의 유사도에서 낮은 평가를 받았다.

첫 번째 질문과는 달리, 두 번째 질문인 챗봇과의 대화가 영어 학습에 미치는 효과성에 대해서는 참여자들이 <Table 5>와 같이 상대적으로 긍정적으로 답했다.

<Table 5> Effect of conversations with a chatbot on the improvement of students' English proficiency

| Scale | 1 | 2 | 3 | 4 | 5 | M |
|-----------|---|---|---|----|---|------|
| Mitsuku | 3 | 4 | 6 | 14 | 0 | 3.15 |
| Cleverbot | 2 | 7 | 5 | 13 | 0 | 3.07 |

Mitsuku와의 대화가 영어 학습에 ‘어느 정도 도움이 된다’고 답한 참여자는 14명이었고 ‘전혀 도움

이 되지 않는다'와 '거의 도움이 되지 않는다'고 답한 경우는 7명에 불과하였다. 대체로 부정적인 견해를 보였던 Cleverbot의 경우도 13명이 '어느 정도 도움이 된다'고 답했다. 그 이유로는 대부분 챗봇의 답변 여부나 적절성에 상관없이 챗봇과의 채팅에 참여하는 것 자체가 영어 구사의 기회를 제공하고 챗봇이 구사하는 표현 가운데도 참여자들이 몰랐던 유용한 표현들이 포함되어 있어 도움이 되었다고 답했다. 다음의 [Figure 3]은 Mitsuku가 제공한 간접적 그리고 직접적 문법 피드백의 예시이다.

| | |
|---|---|
| User: Yes, concentrate <u>about</u> our talk, Mitsuku! | Mitsuku: Describe your surroundings to me. |
| Mitsuku: That's good information. Ah. I am concentrating <u>on</u> all these conversations I am currently having. | User: I am taking a <u>English</u> class. |
| | Mitsuku: It's "AN English" not "A English". |
| | User: Oh! Thanks for telling that. |

[Figure 3] Some grammatical feedback by Mitsuku

[Figure 3]의 좌측 예문에서 사용자는 “concentrate on”이라는 표현을 “concentrate about”으로 잘못 사용하고 있고 이에 대해 의도한 것으로 보이지는 않지만 이어지는 대화에서 Mitsuku가 ‘about’을 ‘on’으로 수정하여 구사함으로써 간접적인 피드백 효과를 제공하는 것을 확인할 수 있다. 또한 우측 예문에서는 모음으로 시작하는 명사 앞의 부정관사를 ‘a’가 아닌 ‘an’으로 사용해야 한다는 것을 직접적으로 수정해주고 있다. 이와 같이 챗봇과의 대화 속에서 직간접적으로 얻는 언어적 지식이 두 번째 설문에서는 긍정적인 영향을 준 것으로 나타났다.

V. 결 론

본 연구에서는 영어 쓰기 능력 향상을 위해 AI 챗봇 활용을 탐색하고자 두 가지 연구 질문을 설정하여 이를 체계적으로 살펴보았다. 첫 번째는 대표적인 두 챗봇 Mitsuku와 Cleverbot의 대화 구사 능력에는 어떠한 차이가 있는지 진단하는 것이고 두 번째는 챗봇을 사용한 학습자들이 인식하는 챗봇의 장단점과 그 활용가능성에 관한 것이었다.

전체적으로 Mitsuku의 기능에 대한 만족도가 Cleverbot 보다 높았다. Mitsuku는 사용자의 발화에 대해 맥락적 이해는 다소 부족했지만 문자적 의미에 초점을 둔 이해 능력은 뛰어났고 모든 질문에 충실히 답하고자 하는 모습을 보였다. Cleverbot은 대화에 활용가능한 지식 데이터베이스가 Mitsuku에 비해 현저히 부족해 보였고 웹검색에 기반한 정보 제공 기능도 없어 대화가 단절되는 상황이 많았다. 또한 대화의 단절 시 갑자기 화제를 전환하거나 불친절한 표현을 구사하는 점은 사용자의 불쾌감을 유발하는 요인이었다. 하지만 두 챗봇이 대화에서 구사하는 어휘 수준은 90% 이상이 상위 3,000단

어 이내도 대학생 수준에서는 거의 어려움이 없었고 우리나라의 영어과 교육과정에서 설정하고 있는 영어 기본 어휘의 양이 3,000단어라는 점을 고려하면 고등학교 수준에서도 챗봇에서 제공하는 영어 입력에 대한 이해에는 큰 무리가 없을 것으로 예상된다.

인간 대화와의 유사도와 영어 학습에서의 효과성에 대한 설문 조사에서는 과반수의 참여자들이 AI 챗봇의 현재 대화 구사 수준에 대해 아직까지는 인간의 수준에는 미치지 못한다고 응답한 동시에 영어 쓰기 능력 향상에는 어느 정도 도움이 된다는 긍정적인 반응을 보였다. 챗봇의 개선점에 대한 가장 많은 지적은 맥락 파악 능력의 개선과 지식 데이터베이스의 확충이었다. 그 밖에도 챗봇이 대화 중 사용자의 표현에 대해 보다 직접적으로 피드백을 제공해 주면 좋겠다고 의견도 있었다. 이러한 챗봇의 한계에도 불구하고 영어 구사의 기회와 영어 입력을 제공한다는 점, 그리고 무엇보다도 인간 대화와는 달리 부담 없이 어떠한 민감한 대화 주제도 꺼낼 수 있고 채팅 시 나타날 수 있는 영어 오류에 대해 신경이 덜 쓰인다는 점에 대해서 대부분 학습자들이 챗봇의 장점으로 공감하였다.

본 연구는 실제 장기간의 연구 기간을 설정하여 챗봇기반 학습 활동이 영어 능력 향상에 영향을 미치는지를 측정하지는 못했기 때문에 연구 범위 설정의 한계가 존재하나 다양한 영어교육의 시사점을 이끌어 냈다는 점은 의미가 있다. 본 연구의 결과 분석에 의하면 현재 챗봇의 개발 수준을 고려할 때, 영어 쓰기 활동에 보다 효율적으로 활용하기 위해서는 자유 대화보다는 본 연구에서 채택한 방식과 같이 구체적인 세부 과업을 제시하는 하는 것이 효과적이고 추후 연구에서는 학습자의 수준을 다양화하여 그에 따른 챗봇의 대화 구사 방식과 영어 입력의 수준, 그리고 영어 능숙도별 학습자들이 구사하는 대화 전략의 차이 등을 분석할 필요가 있다.

References

- Abu-Shawar, B. (2017). Integrating CALL systems with chatbots as conversational partners. *Computación y Sistemas, 21*(4), 615-626.
- Cho, K. (2013). Critical study for teacher appraisal for professional development index national university of education. *Korean Journal of Teacher Education, 29*(1), 53-70.
- Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning and Technology, 10*(3), 8-14.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. Retrieved November 3, 2018, from the World Wide Web: <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>.
- Kim, H. (2018, June). Designing L2 interactive tasks with an artificial intelligence robot. Paper presented at *the Asia TEFL International Conference*. University of Macau, China: Asia TEFL.
- Kim, J. (2017). *The effects of human-AI assistant intelligences on children's collaborative language acquisition*. Unpublished master thesis. Gwangju National University of Education.
- Mauldin, M. (1994). Chatterbots, tinymuds, and the turing test: Entering the Loebner Prize competition. *In Proceedings of the 11th National Conference on Artificial Intelligence*. Seattle, Washington: AAAI Press.
- Ministry of Education. (2015). *2015 revised national curriculum of English*. Seoul: Ministry of Education.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle Cengage Learning.
- Park, C., & Kim, J. R. (2018). An analysis of task conditions for teaching excellence of elementary English classes using on-task analytic criteria. *Korean Journal of Teacher Education, 34*(1), 121-142.
- Scott, M. (2017). *WordSmith Tools version 7*[Computer Software]. Stroud: Lexical Analysis Software.

Abstract

Exploring the Feasibility of AI Chatbots as a Tool for Improving Learners' Writing Competence of English

Shin, Dong-Kwang (Associate Professor, Gwangju National University of Education)

The present study attempted to explore the feasibility of using the most representative AI chatbots Mitsuku and Cleverbot, as one of the ways to utilize AI devices for teaching and learning English. For this, 27 university students chatted with the two chatbots. During the chatting seven various sub-tasks were given to the students. As a result, both chatbots used 90%, or more of their words, within the top 3,000 words during their conversations, and it was thus not difficult for the students to understand the two chatbots' expressions. In the comparison between Mitsuku and Cleverbot, the evaluation of Mitsuku was much more positive. Although Mitsuku demonstrated some limitations regarding its ability for context recognition, she could properly respond to the students using her own knowledge database and even conducted web searches. In spite of the limitations of the chatbot the participants thought that the chatbot could provide a lot of English-use opportunities and English input, and reduce foreign language anxiety of the learner. Finally, this study addresses additional implications and suggestions for further research.

Key words : Digital Literacy, Artificial Intelligence (AI), Chatbot, Chatbot-based Writing Task

