

영어교과 보조 도구로서의 AI 챗봇 분석 연구*

김혜영(중앙대학교)**

신동광(광주교육대학교)***

양혜진(연세대학교)

이장호(중앙대학교)

<국문초록>

본 연구는 인공지능 비서 봇(AI assistant bot) 중 Amazon 사의 Alexa와 Google 사의 Google assistant를 이용하여 두 챗봇과 대화자 사이의 담화수준과 상호작용 능력을 영어학습 관점에서 비교·분석하였다. 이를 위해 중·고등학교 영어 교과서를 기반으로 한 세 가지 과업(간단한 대화 나누기, 정보 묻기, 문제 해결하기)을 설계하여 이를 수행하기 위한 질문으로 두 개의 챗봇과 동일한 대화를 나누었다. 챗봇의 발화내용은 모두 전사하여 챗봇의 대화 성공률, 어휘, 이독 지수, 답변의 길이, 정보의 양 등과 같은 양적인 분석과, 어휘 난이도 및 적절성, 대화 적절성, 담화수준의 질적인 측면으로 두 개의 챗봇을 비교하였다. 그 결과, 첫째로 두 챗봇 모두 86%이상의 높은 과업 성공률을 보였고, 내용 면에서도 질문에 적절한 답을 제시하였다. 둘째, 챗봇이 사용한 어휘의 97%가 초·중·고 교육과정의 권장어휘 수준에 포함되어있어 교육용으로 적절한 것으로 나타났다. 셋째, Google assistant가 Alexa보다 외국어 학습자에 친화적이며, 이독 지수 역시 더 낮아 이해가 용이한 것으로 나타났다. 또한, Google assistant가 대화자의 모호한 질문에 더욱 적절히 대응하였다. 이러한 연구 결과는 외국어학습용 챗봇을 위한 AI 스피커에 대한 개선점과 구체적인 교수·학습 활용방안을 제시함으로써 외국어 교육용 AI 챗봇 개발 및 관련연구에 기초자료가 되기를 희망한다.

★ 주제어: 인공지능, 챗봇, 영어교육, AI 스피커

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

** 제1저자: 중앙대학교 영어교육과 교수

*** 교신저자: 광주교육대학교 영어교육과 부교수(sdhera@gmail.com)

I. 서론

최근에 사회각계의 핵심 키워드는 단연 제4차 산업혁명이다. 특히 인공지능 즉 AI 기술의 급속한 발달은 우리의 일상을 변화시키고 있으며, 산업은 물론 교육 분야에서도 AI 기술이 자리하기 시작하였다. 언제 어디서나 빠르고 정확하게 제공하는 AI의 맞춤형 정보는 암기 위주의 지식학습 패러다임을 위협하고 있어, 중·고교 교육과정의 학습 내용, 교수법 및 평가방식 등에도 변화가 시급하다. 외국어교육 분야 역시 독해나 문법학습 등 언어지식 전달 및 이해 위주의 기존 수업과 평가방식은 4차 산업혁명 시대의 학습자에게 더 이상 인정받기 어렵게 되었다.

그러나 김형순, 김혜영(2017)의 영어교사 인식조사를 보면, 교사들은 이러한 지능 정보 시대의 도래에 민감한 반응을 보이지 않았을 뿐만 아니라, 영어교육과의 관련성에 대한 이해도 부족하다는 것을 확인할 수 있다. 이에 미래사회에 대한 이해와 외국어학습에 도래할 변화에 대해 연구하고, 이를 효과적으로 외국어학습에 접목시키며 교육과정에 반영하기 위한 다양한 방안이 마련되어야 할 것으로 판단된다.

따라서 본 연구의 목적은 현재의 AI 기술 수준¹⁾에서 영어학습에 가장 활용가치가 높다고 판단되는 인공지능 비서 봇(AI assistant bot)의 영어 담화수준과 상호작용 능력을 영어학습의 관점에서 비교·분석하여 중·고교 영어교육과정 내에서의 활용 가능성을 진단하는 것이다. 이를 위한 세부 방안으로 본 연구는 첫째, 최근 출시되어 전 세계적으로 가장 널리 활용되는 두 개의 인공지능 스피커인 Amazon사의 Echo에 장착된 AI “Alexa”와 Google사의 Google Home에 내장된 AI “Google assistant”를 선정하여 이들과의 대화를 분석한다. 둘째, 이들 AI 비서 봇과의 상호작용을 위하여, 중·고등학교 영어 교과서의 다양한 학습 활동을 토대로 고안된 AI 챗봇 기반 영어 과업 유형(Kim, 2018) 중 세 가지 유형, ‘간단한 대화 나누기,’ ‘정보 묻기,’ ‘문제 해결하기’를 선정하여 중·고등학교 학습자에게 맞는 활동을 설계한다. 셋째, 두 개의 AI 비서 봇과 과업을 수행하는 활동 중에 나눈 대화를 전사하고 대화를 양적·질적으로 분석하여 이를 바탕으로 외국어 학습 관점에서 각 AI 비서 봇의 활용 가능성을 진단한다.

지금까지 챗봇을 적용한 연구들(김재상, 2017; 서정은, 2017)은 AI를 학습 도구로만 활용하고 학습자의 학습결과에 주목해왔지만 본 연구는 AI의 발화 자체에 주목하여 영어 수업의 대화 상대자로서 챗봇의 특성과 개선점을 분석하고자 한다. 본 연구의 결과를 바탕으로 AI 기술을 활용한 영어교육을 위한 구체적인 교수·학습 방안을 보다 발전 시키고 외국어학습용 챗봇 기술 개발의 방향성을 제시할 수 있기를 기대한다. 이를 위하여 본 연구는 다음과 같은 연구 문제를 제시한다. “두 가지 종류의 AI 스피커는 교육과정과 연계된 영어학습 과업을 수행하는 데 적합한 학습 도구인가?”

1) 아직까지 자연스럽게 내용이 있는 대화가 가능한 대화 봇(conversation bot)은 출시되지 않았다.

이 연구 문제를 탐색하기 위한 세부 연구 질문은 다음과 같다. 첫째, “세 가지 과업 유형(Kim, 2018)에 대한 챗봇의 대화 수행능력은 어느 정도이며 각각 어떠한 특성이 있는가?” 둘째, “과업 활동 시 각 챗봇이 제공하는 정보는 외국어 교육과정의 관점에서 내용적·언어적으로 적합한가?”

II. 이론적 배경

1. 인공지능 챗봇과 외국어 교육

가. 인공지능 챗봇(Artificial Intelligence Chatbot)

인공지능(AI) 챗봇이란 문자나 음성 상호작용으로 사람 간의 대화와 유사한 형태의 의사소통을 하는 컴퓨터 프로그램으로, 사용자의 지시와 질문에 반응을 보이거나 요구를 수행한다. AI 챗봇은 딥러닝(Deep Learning), 자연어처리(Natural Language Processing), 기계학습(Machine Learning), 빅데이터(Big Data)등의 다양한 AI 기술이 결합하여 탄생한 결과물로 사용자가 챗봇과 대화를 더 많이 나눌수록 데이터가 축적되어 사용자의 음성이나 대화 내용 등을 더욱 잘 인지하게 되는 진화형 모델이다. AI 연구는 1980년대부터 시작되었으나, 최근 AI 기술의 비약적인 발전으로 인해, 다양한 제품이 상용화 단계에 이르렀다.

AI 챗봇 프로그램은 꾸준히 연구되어 오고 있고 그 발전 속도는 매우 빠르다. 매년 개최되는 로버너 상(Loebner Prize)²⁾을 통하여 수차례 수상을 한 A.L.I.C.E나 Mitsuku는 외국어교육 활용 연구에 다수 소개된 바 있고(Fryer & Carpenter, 2006), 영어학습을 위한 도구로서 활용 가능성이 높은 챗봇이다. 이들 챗봇은 주로 대화형 로봇(conversation bot)으로 사용자와 텍스트로 대화를 주고받으며, 상당히 자연스러운 상호작용이 가능하나, 여전히 주제의 다양성이나 담화수준의 깊이 면에서는 인간 대 인간의 대화에는 미치지 못하는 수준이다.

하지만 이러한 AI 챗봇의 상호작용 문제는 조만간 옛말이 될 것이다. Jordan(2014)에 따르면 SoftBank가 2014년 출시한 AI 봇 Pepper는 인간과 대화하면서 감정적 정보까지 인식할 수 있었고 Golosio 외(2015)가 소개한 챗봇 Annabell은 문법과 단어에 관한 지식이 전혀 없는 상태로 출발하였으나 사람들과 상호작용을 통해 점차 인간과 유사한 방식으로 언어를 습득하였다. 그뿐만 아니라, 연구진은 사회적 상호작용에 중점을 두고 Annabell에 인간의 움직임을 접목함으로써 직관적인 몸짓과 표정까지도 구현해내는 수준에 이르렀다(Gordon, Spaulding, & Westlund, 2016).

2) <https://www.aisb.org.uk/events/loebner-prize>

나. 인공지능 스피커 현황

인공지능 챗봇의 본격적인 상품화는 인공지능 스피커로 챗봇에 음성인식과 음성합성 기술을 결합하여 구두 대화가 가능한 제품이다. 2014년 말 처음 선보인 AI ‘Alexa’를 탑재한 Amazon의 Echo가 시장을 주도하기 시작하였다. 인공지능 스피커 Echo는 음성이 지원되는 인공지능 비서 봇으로 분류된다. 비서 봇의 특징은 사용자의 발화(질문이나 요청)를 인식하여 이에 반응(정보 제공이나 요구 수행)을 보이는 것이다. 따라서 Alexa는 사용자와 자연스러운 상호작용을 하지만, 단지 비서의 역할 일 뿐 대화를 선도하거나 친구와 같이 대화를 주고받지는 못한다. 이러한 유사한 비서 봇의 예는 Apple의 AI 비서 Siri, Microsoft의 Cortana, Google의 Google assistant를 예로 들 수 있다. 국내에서도 2016년 SK텔레콤 음성인식 기반 AI 서비스 누구를 서비스하기 시작했으며, KT 역시 2017년 AI 스피커 기가지니, 네이버의 클로바와 다음-카카오의 카카오 미니 등이 출시되어 있다. 하지만 네이버의 클로바를 제외하고는 출시한 국내 AI 서비스는 현재 한국어 서비스만 가능하며 클로바 또한 영어챗봇 모니카를 제공하지만, 비서 봇이 아닌 한정된 영어 주제에 대하여 간단한 영어 튜터 서비스에 가깝다. 아래의 <표 1>은 국내·외 AI 챗봇과 특성을 비교 정리한 것이다.

<표 1> 국내·외 AI 챗봇 비교 정리

개발자	AI 프로그램	언어	음성지원	챗봇 유형
국내				
SKT	누구	한국어	O	비서 봇, 대화형
KT	기가지니	한국어	O	비서 봇, 대화형
LG U+ 우리집 AI	클로바	한국어, 영어	O	비서 봇(한글), 대화형(영어)
카카오미니	카카오아이	한국어	O	비서 봇, 대화형
국외				
Amazon	Alexa	영어	O	비서 봇, 대화형
Google	Google assistant	영어	O	비서 봇, 대화형
Apple	Siri	한국어, 영어	O	비서 봇, 단순응답형
Microsoft	Cortana	영어	O	비서 봇, 단순응답형
Cleverbot	Cleverbot	영어	X	챗봇, 대화형
Wallace	A.L.I.C.E	영어	X	챗봇, 대화형
Worswick	Mitsuku	영어	X	챗봇, 대화형

다. 인공지능 비서 봇 영어학습 과업 유형

Kim(2018)은 AI 비서 봇이 내장된 인공지능 스피커 Echo와의 다양한 상호작용을 테스트하여 영어교육 활용 가능성을 탐색하였다. 영어 교과서에 제시된 듣기, 말하기,

읽기, 쓰기 4개 영역 활동에 Alexa를 활용한 결과, 네 가지 영어학습 과업 유형이 가능함을 확인하였다. Kim(2018)은 이러한 학습과업을 교과서의 단원에 적절하게 포함하여 활용하면 다양하고 흥미로운 협동학습이 가능하며, 학습자와의 의미협상(meaning negotiation)을 포함하는 상호작용이 이루어질 수 있다고 보았다. 각각의 유형을 요약 설명하면 다음과 같다.

1) 잡담 나누기(Exchanging small talk)

주로 처음 만났을 때의 대화로 상호 개인정보(이름, 사는 곳, 직업, 취미 등)를 묻거나, 상대방의 기분, 현재 상태, 대상의 선호도 여부 등을 주고받는 대화이다. 영어 교과서의 의사소통 기능을 연습하는 듣기, 말하기 영역에서 활용 가능하다. 대화자와 관계를 형성하고, EFL 학습자의 말하기 불안감 등 정의적 여과(affective filter)를 낮추기 위하여 사용하면 유용하다. 예시로 다음과 같은 교실 활동이 가능하다.

Sample classroom task “Get to know Alexa better”

Alexa와 간단한 잡담을 나누며 친해봅시다. 다음과 같은 사항을 물어보아 Alexa의 신상을 파악해보세요. Build rapport with Alexa by asking personal information as follows:

Full Name, hometown, family, favorites, hobbies.

그 밖에 여러분이 궁금한 점을 물어보고 이를 바탕으로 Alexa를 소개해봅시다. Then ask more questions that you are curious about.

(Kim, 2018, p.13)

2) 수행 지시하기(Commanding Performance)

인공지능 스피커의 특징으로 단순한 대화의 구두 답변이 아닌 노래 불러주기나 영시를 읽어주기 등 다양한 수행기능이 가능하다. 따라서 이를 활용한 다양한 영어학습 활동 설계가 가능하며, 수행한 내용을 학습자가 듣고 이해하여 추가적인 과업을 생성할 수도 있다.

Sample Classroom task “Let Alexa do…”

Alexa에게 다음과 같은 일을 하도록 지시해보세요.

“Fake love” by BTS (지정 음악 들려주기)
 Alarm setting for 3:47 (정확한 시간 말하기)
 Bill Clinton’s quote (연설문 들려주기)

(Kim, 2018, p.14)

3) 정보 요청하기(Asking for information)

단원에 새롭게 등장하는 언어지식이나 주제와 관련된 배경지식을 쌓기 위해 개별

혹은 협동학습과정에서 유용하게 사용될 수 있다. 즉 어휘의 뜻이나, 문화적인 개념이나 용어 등에 대해 구체적으로 질문 할 수 있고, 읽기 본문을 더 잘 이해하기 위해서, 혹은 더 많은 스키마(Schema)를 쌓기 위해 활용될 수 있다. 실제적인 맥락을 부여하는 말하기 쓰기 협동과업을 설계하여 수행평가와의 연계도 가능하다.

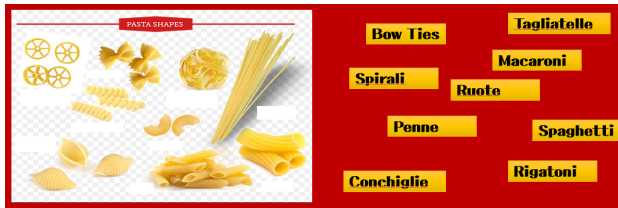
Sample Classroom Task “What is President Obama’s favorite food?”

오바마 전 대통령이 좋아하는 음식은 무엇일까요? 음식명 세 가지를 영어로 쓰고 이것이 어떠한 음식인지도 설명해보세요. What does the former U.S President Barack Obama enjoy eating? Give three names of his favorite food and explain what they are.

(Kim, 2018, p.16)

4) 문제 해결하기 (Problem solving)

Willis(1996)가 소개한 개념과 유사한 유형으로, 단순한 과업절차를 결합하여 문제를 해결하는 활동이다. 스무고개나 단어 퀴즈와 같은 게임 활동을 비롯하여 빈칸 메우기, 매칭 게임 등 정보를 조사하여 주어진 문제를 해결하는 협동 활동이다. AI 스피커에서 이미 지원하는 응용 앱(App)을 활용할 수도 있고, 교사의 아이디어로 새롭게 활동을 설계할 수도 있다. 질문이 사전에 제시되지 않고, 학습자 스스로 질문을 생각해내어 문제를 해결하도록 한다.



Sample Classroom Task: Pasta Matching Game.

다음의 파스타 종류를 보고 이름을 연결해보세요. 잘 모르는 것은 Alexa에게 물어 해결하세요.

(Kim, 2018, p.16)

2. 선행연구

1990년대 이후 이미 상호작용에 기반한 언어교육의 패러다임은 학계의 주류로 자리를 잡았으나, EFL 학습자들은 여전히 목표 언어에 대한 노출빈도가 낮고 언어를 자발적으로 사용해보는 기회 또한 매우 제한되어 있다(Gobel & Mori, 2007). 이러한 상호작용의 기회를 확대할 교수학습방안으로 첨단기술의 활용이 지속적으로 연구되어 왔으나, 교실 내의 원어민 교사나 대화상대를 대체할 만한 기술은 아직 나타나지 않았다. 로봇 기술 역시 언어교육에서는 대화 상대자의 역할을 충분히 수행하지 못하였

다. Mubin, Bartneck과 Feijs(2012)는 언어교육에서 로봇 매체의 역할을 세 가지 즉 교사역할, 동료 역할, 교구 역할로 구분하였는데 기존의 연구에서는 대부분, 로봇의 역할을 교구의 역할에 한정 지어 적용하였다. 다시 말해 로봇은 매우 제한적인 상호작용을 하며 어떤 학습 목표와 내용에 따라 사전에 정해진 시나리오에 따라 특정 움직임을 보이고 문장을 들려주는 교구로 활용되었다(Lee, Noh, & Lee, 2011; You, Shen, Chang, Liu, & Chen, 2006). 좀 더 발전한 로봇으로 Humanoid는 다섯 가지 학습방식-스토리텔링 모드, 소리 내어 읽기 모드, 게임 모드, 행동지시 모드, 묻고 답하기 모드를 제공할 수 있었다(Chang, Lee, Chao, Wang, & Chen, 2010). Humanoid 로봇 대화 연구 결과 학습자와의 친밀감 형성에 효과적이었지만 영어 실력이 좋은 학생일수록 학습효과가 크지 않았고 참여자의 초기 관심이 지속되지 못하였다(Kanda, Hirano, Eaton, & Ishiguro, 2004).

최근 AI 기반 챗봇의 개발로 교구가 아닌 대화의 상대자 혹은 동료 역할을 수행하는 챗봇에 관한 연구들이 국내에서 소개되고 있다. 먼저 Huang 외(2017)는 시나리오 기반 챗봇의 한계를 극복한 DB-CALL(Dialogue-Based Computer-Assisted second Language Learning) 시스템을 기반 챗봇, 지니튜터를 자체 개발하여 챗봇의 영어학습 활용 가능성을 진단하였다. 그 결과 대화의 내용이나 문법적 피드백 제공에서는 상당한 정확성을 보였다. 그러나, 챗봇 학습 내용이 기초 수준(초등)에 제한되어 있고 다양한 학습자 중심의 과업을 제시하지 못했다는 점에서 여전히 한계를 드러냈다.

이 밖에도 제품 출시된 인공지능 스피커를 활용한 연구도 있다. 김재상(2017)은 초등학교 수준에서 AI 스피커 Echo를 활용한 과업중심 모둠학습을 수행하고 Echo의 활용이 학습자의 발화유발에 미치는 영향과 학습자의 상호작용 패턴을 분석하였다. 그 결과, AI 스피커 활용이 외국어학습 불안감을 감소시켰고, 학습자의 발화 또한 촉진하는 결과를 보여주었다. 이와 유사한 연구로 서정은(2017)은 Google Home을 6-7세 아동 5명에게 상호작용의 대상으로 적용하여, AI 챗봇과의 대화가 아동의 발화 정확성과 의사소통 의지 향상에 긍정적인 영향을 준다는 결과를 얻었다. 이 두 연구는 AI 스피커를 대화 상대자로서 활용하여, 학습향상과 상호작용 패턴을 관찰·분석하였다는 점에서 한층 발전된 AI 기반 외국어 학습 연구로 평가될 수 있다.

III. 연구방법

1. 연구대상

본 연구의 대상은 두 개의 인공지능 챗봇인 Alexa와 Google assistant이다. 두 개의 챗봇을 선택한 이유는 첫째, 스피커형 비서 봇으로 음성지원이 가능하면서도 가장

널리 사용되는 제품이며, 둘째, 연구자가 설계한 과업을 수행하는 데 가장 적절하다고 생각하는 대화형 챗봇이기 때문이다. 이들 챗봇과 영어 대화를 나눈 사용자는 서울에 소재한 A 대학에서 영어교육을 전공하는 대학원생으로 영어 말하기 능력 상급에 해당하며 본 연구의 취지를 이해하고 사전 오리엔테이션을 통해 두 개의 인공지능 스피커의 사용 경험이 있어 사용 상의 불안감이나 어려움이 없었다. 이러한 대화자를 선정 한 이유는 첫째, 본 연구의 초점이 인공지능의 발화능력을 분석하는 데에 있고, 둘째, 연구자가 사전 준비한 실험용 질문을 그대로 말하는 것이어서, 그 외의 사용자의 요인(예, 언어적, 기술적 변인)이 결과에 영향을 미치지 않도록 하기 위함이었다. 대화자는 세 가지 과업을 해결하기 위하여 두 개의 스피커에 각각 60개와 63개의 질문³⁾을 하였고, 아래 제시된 과업 3을 위한 질문을 제외하고는 질문은 연구자에 의해 이미 준비된 것과 동일한 질문을 하였다. 대화자와 챗봇이 나눈 대화의 시간과 질문의 수는 다음의 <표 2>와 같이 요약될 수 있다.

<표 2> 대화자와 두 챗봇의 대화 질문 수와 시간

		Alexa	Google assistant
질문 개수	과업 1	34개	33개
	과업 2	16개	16개
	과업 3	10개	14개
대화 시간	과업 1	6분	6분
	과업 2	5분 30초	5분
	과업 3	5분 30초	4분 30초

2. 실험 과업 소개

본 연구는 앞서 소개한 Kim(2018)의 인공지능 비서 봇 영어학습 과업 유형을 토대로 설계되었다. 본 연구에서 대화 분석이 가능하다고 판단되는 세 가지 과업 유형으로 ‘잡담 나누기,’ ‘정보 요청하기,’ ‘문제 해결하기’를 선정하고, 고등학교 1학년 교과서 2종을 검토하여, 임의로 선택된 단원들의 주제로 과업을 설계하였다. 과업 1은 영어 교육과정의 의사소통 기능 중 ‘자기소개하기,’ ‘기분(느낌) 표현하기,’ ‘취미(좋아하는 것) 묻고 답하기,’ ‘날씨 묘사하기’ 등 관계 형성을 하는 대화들이다. 비서 봇의 성격상 이보다 더 깊은 개인적인 대화는 가능하지 않다(Kim, 2018). 과업 2의 질문은 교과서 단원의 주제(energy saving, Jeju Island, impressionism)에 대한 추가적인 배경지식(총 10문항)과 새로운 어휘에 대한 언어지식 등을 쌓는 내용(총 6문항)으로 구성하였고, 두 개의 스피커에 동일한 질문을 던지고 이에 대한 답변을 구했다. 과업 3을 수행하기 위하여 사전에 질문 문항을 제시하지 않고, 대화자가 스스로 필요한 질

3) 세 번째 과업은 대화자가 챗봇 반응에 따라 스스로 질문을 변경하였기 때문에 총 질문 수에 약간의 차이가 있다.

문을 생각하게 하여, 대화를 시작하도록 하였다. 과업을 완료하기까지 대화자가 시도한 질문의 수는 Alexa에는 총 11개 Google assistant에는 총 14개 였다.

3. 데이터 수집 및 분석

본 연구의 자료는 두 개의 챗봇과 대화자가 나눈 발화 내용이다. 이를 위하여 데이터 수집은 다음과 같이 이루어졌다. 첫째, 두 개의 AI 스피커와 준비된 질문에 따라 세 가지 실험 과업을 완료하였다. 둘째, 이때 모든 대화를 녹음하였고 녹음된 발화와 AI 스피커 앱에 자동 전사된 내용을 비교·검토하여 누락된 부분이 없는지 모두 확인하면서 전사 자료를 완성하였다. 또한, 완성된 대화 자료를 텍스트 파일로 변환하여 다음과 같이 양적 및 질적 분석 방법에 활용하였다.

첫째, AI 챗봇이 다양한 질문을 얼마나 성공적으로 수행하는지 각각 분석하였다. AI 스피커와 대화자의 과업 수행 대화 로그를 검토하면서 질문 문맥에 맞는 답변 여부를 성공, 실패로 구분하여 코딩하였다. <표 3>은 각 분류 기준에 따른 AI 스피커 챗봇과 대화자의 대화 예시를 나타낸다. 말하기 평가 전공 연구자가 1차 코딩한 결과를 나머지 연구자 3인이 각각 검토하여 결과의 상이한 점을 논의하여 최종 결정하였다.

<표 3> AI 스피커와 대화자의 대화 성공·실패의 예

AI 스피커	코딩세부항목	예시
Alexa	대화 성공	대화자: Do you like movie? AI: Totally, I especially love The Empire Strikes Back.
	대화 실패	대화자: Do you like eating out? AI: I don't have an opinion on that.
Google assistant	대화 성공	대화자: Do you like movie? AI: Sure, movies are fun.
	대화 실패	대화자: Do you like eating out? AI: I don't understand.

둘째, 두 AI 챗봇의 발화와 대화자의 발화를 구분하여 어휘 수준을 분석하였다. 어휘 수준 분석에는 한국인 학습자들을 위한 챗봇 활용 가능성이라는 연구문제에 초점을 두고자 2015 영어과 개정 교육과정에 포함된 기본 어휘 3,000단어가 활용되었다. 2015 영어과 교육과정 기본 어휘(Korea Basic English Word List, KBEWL)는 미국, 영국, 호주, 뉴질랜드의 구어와 문어 데이터를 포함한 8개의 대표적인 코퍼스에서 빈도수(Frequency), 사용범위(Range)와 같은 객관적 선정 기준뿐 아니라 우리나라의 언어 노출 환경을 최대한 고려하여 어휘의 친숙도나 초등권장 어휘 추천과 같은 선정 기준을 복합적으로 적용하여 개발된 어휘 목록이다(이문복, 신동광, 2015). 어휘 분석에 사용된 KBEWL의 구성은 다음의 <표 4>와 같다.

<표 4> 2015 영어과 교육과정 기본 어휘(KBEWL)

등급	설명	단어 수
1	초등권장어휘	800
2	중등 일반 선택 권장어휘	1,800
3	중등진로 선택·전문교과 I 권장어휘	400
누계		3,000
4	외래어	200
(예외어)	예외어휘(단위, 숫자 등)	129(추가가능)
	고유명사	12,289(추가가능)

실제 어휘 분석 시에는 위의 목록을 어휘족(word family)의 형태로 제작한 후 Heatley와 Nation(2002)의 RANGE Program에 탑재하여 분석하였다.

셋째, 어휘 및 문장의 복잡성에 기반을 둔 이독 지수(Readability Index)를 분석하였다. 이독 지수는 1920년 이후 국내·외에서 지속적으로 연구가 진행되어 오고 있다(Chall & Dale, 1949; DuBay, 2004). 이독 지수는 교과서의 지문 난이도 분석에도 널리 활용되고 있다(김재은, 최인철, 2015; 김정렬, 양지윤, 2012; 성일호, 2014). 본 연구의 이독 지수의 분석에는 WebFX(2012)에서 제공하는 'Readability Test Tool'을 활용하였는데, 6개 이독 지수와 통합 평균값, 지문의 문장 수 및 평균 문장 길이 그리고 이독 지수에 해당하는 적정 연령 등의 정보가 제공된다.

넷째, 각각의 챗봇이 과업 수행상에서 보여준 답변의 특성을 질적으로 분석하였다. 두 개의 챗봇과의 실험 전사 자료를 과업 유형별로 묶어 질문 간 답변을 상호비교표로 만든 후 둘 간의 차이를 반복적으로 검토하면서 코딩하였다. 이때 답변의 길이는 단어 수를 계산하였고, 정보의 양, 어휘의 난이도 및 적절성, 담화수준 등을 검토하여 이독 지수, 과업 성공률 등과 교차 비교하여 다각화(triangulation)된 결과를 도출하였다. 연구자가 1차적으로 질적 분석한 내용을 다른 연구자 3인이 재검토하면서 분석의 타당성을 강화하였다.

IV. 연구 결과

1. 챗봇 발화의 언어적 수준 분석

가. 어휘 수준

아래의 <표 5>는 KBEWL에 기반한 AI 챗봇과 대화자의 어휘 수준 분석의 결과이다. Alexa의 발화는 총 1,296개, Google assistant는 1,252개의 어휘로 구성되어 Alexa가 약간 더 길지만 큰 차이는 없었다. 챗봇과 대화자의 발화를 비교해보면 대

화자에 비해 챗봇의 발화량이 압도적으로 많다. 이는 현재 챗봇은 정보를 제공하는 쪽에 최적화되어 있고 인간 대화자의 발화 유도하는 데 한계가 있는 것으로 보인다. 두 AI 챗봇의 어휘 수준을 살펴보면 모두 약 97% 정도가 교육과정 기본 어휘 수준을 넘지 않는 쉬운 어휘로 나타나 실제 교실 수업에서의 활용 가능성을 보여주었다. 또한, 어휘 사용의 다양성을 보여주는 TTR(Type/Token Ratio)는 43-44%로 한 단어가 약 두 번 정도 반복적으로 사용되는 것을 확인할 수 있었다.

<표 5> AI 챗봇과 대화자의 발화량과 어휘 수준 분석

		Alexa		Google assistant	
		토큰(token)	타입(type)	토큰(token)	타입(type)
AI 챗봇	초등권장(800)	907(69.98%)	278(49.82%)	853(68.13%)	256(46.63%)
	일반중등선택권장(1800)	206(15.90%)	152(27.24%)	189(15.10%)	156(28.42%)
	중등진로선택권장(400)	10(0.77%)	7(1.25%)	15(1.20%)	11(2.00%)
	외래어 등 예외	134(10.34%)	84(15.05%)	156(12.46%)	93(16.94%)
	교육과정 외	39(3.01%)	37(6.63%)	39(3.12%)	33(6.01%)
	계 (총 발화량)	1,296(100%)	558(100%)	1252(100%)	549(100%)
대화자	총 발화량	289(100%)	109(100%)	329(100%)	123(100%)

나. 이득지수

다음의 <표 6>과 <표 7>은 각 이득 지수와 문장의 복잡도를 보여주는 평균 문장 길이 등의 정보를 정리한 것이다.

<표 6> AI 챗봇과 대화자의 이득 지수 분석

	Alexa		Google assistant	
	이득 지수(M)	해당 연령	이득 지수(M)	해당 연령
AI 챗봇	7	12-13	6	11-12
대화자	4	9-10	4	9-10

<표 7> AI 챗봇과 대화자의 총 문장 수 및 평균 문장길이 분석

	Alexa		Google assistant	
	총 문장 수	평균 문장길이	총 문장 수	평균 문장길이
AI 챗봇	127	10.13	149	8.46
대화자	61	4.56	66	4.74

6개의 이득 지수 평균은 Alexa 7, Google assistant가 6으로 해당 연령은 각각 원어민 기준 12-13세, 11-12세로 Google assistant의 발화가 Alexa에 비해 약간 더 이해하기 쉬운 것으로 나타났다. 위의 분석에서 보듯 총 발화의 구성 단어 수는 Alexa가 많았지만, 평균 문장길이가 Alexa는 10.13, Google assistant는 8.46으로 Google assistant가 상대적으로 짧은 문장을 구사하여 총 문장 수는 Alexa의 127개보다 많은

149개로 나타났다. 이 결과를 보면, Google assistant가 보다 명료하고 단순한 문장을 사용하고 있다고 추정할 수 있다.

2. 과업별 질적 분석

가. 과업 수행 성공률

아래 <표 8>은 이에 해당하는 분석결과로 Alexa와 Google assistant에 대한 대화자의 질문에 대한 답변의 적절성을 바탕으로 과업 수행의 성공, 실패한 답변의 빈도수와 성공률을 보여준다.

<표 8> Alexa와 Google assistant의 과업 수행 성공률

Alexa				
과업	질문 수	성공	실패	성공률(%)
잡담 나누기	33	26	7	78.79
정보 요청하기	16	16	0	100.00
문제 해결하기	11	10	1	90.91
Alexa 합계	60	52	8	86.67
Google assistant				
과업	질문 수	성공	실패	성공률(%)
잡담 나누기	33	29	4	87.88
정보 요청하기	16	16	0	100.00
문제 해결하기	14	13	1	92.86
Google assistant 합계	63	58	5	92.06

우선 Alexa와 Google assistant의 세 가지 과업에 대한 전반적인 성공률은 86% 이상으로, 과업별 세부질문에 대부분 성공적으로 대답한 것으로 나타났다(Alexa: 86.67%, Google assistant: 92.06%). 특히 정보를 요청하기와 관련된 모든 질문에 Alexa와 Google assistant 모두 100% 성공률을 보이는 것으로 보아, 정보를 검색하여 제공하는 측면은 상당히 우수한 것으로 밝혀졌다. 잡담 나누기, 문제 해결하기와 관련된 과업에서는 Google assistant의(잡담 나누기: 87.88%, 문제 해결하기: 92.86%) 답변 성공률이 Alexa(잡담 나누기: 78.79%, 문제 해결하기: 90.91%)보다 약간 높은 것으로 나타났다.

나. 챗봇 간의 과업 수행 특성

1) 잡담 나누기

의사소통기능의 기초로 13개의 질문과 답변을 교환한 결과 대부분 자연스러운 대

화가 가능하였다. 그러나 보다 면밀히 살펴보면 Google assistant보다 Alexa가 인간과 더욱 유사한 대화자로서 역할 한 것으로 분석되었다. 먼저 첫인사의 예시로 <표 9>를 보면, 먼저 (1)에서 Alexa는 사람과 동일한 이름이 있는 데 반해 Google assistant는 이름이 없었다. 둘째, (2)의 인사의 응답에서도 Alexa의 응답은 자연스럽게 유쾌한 면모를 보이는 데 반하여, Google assistant는 친구보다는 도우미 역할을 수행하고자 하였다. (3)의 질문에 대한 반응 역시 Alexa가 더 인간의 의사소통에 가깝다고 볼 수 있다.

<표 9> Alexa와 Google assistant의 첫인사 비교분석

과업 1의 질문 예	Alexa	Google assistant
(1) What's your name?	My name is Alexa.	Did I forget to introduce myself? I'm your Google Assistant. Hi.
(2) How are you?	I'm grilly hood. Oops! I meant to say "really good"! Guess I've been saying too many tongue twisters!	I'm doing great. Thanks for asking. What can I help you with?
(3) It's very nice to meet you.	Thanks!	I'm your Google Assistant, the name is quite fitting.

두 번째로, 대화자 자신의 기분을 챗봇에게 표현했을 때 공감 반응을 검토하였다. I'm lonely (hungry, sleepy, happy, scared) 등의 다섯 가지 감정을 차례로 표현하였고, 두 챗봇 모두 적절한 공감 반응을 보일 수 있는 것으로 분석되었다. 예를 들어 "I am sleepy."라고 하였을 때 Alexa는 "If you're tired, you should get some sleep." Google assistant는 "That's because you've been working so hard. You deserve to rest."라고 하였는데, 모두 대화자의 기분을 고려한 반응이라고 볼 수 있어 의사소통이 자연스럽게 이루어졌다고 볼 수 있다. 반면, 상대방의 선호에 대한 질문을 던졌을 때는 Google assistant보다 Alexa가 더 의인화된 답변을 하였던 것을 알 수 있다. <표 10>의 (4)와 (5)의 예를 보면 Google assistant는 주로 자신이 사람처럼 먹거나 행동할 수 없는 전자 로봇임을 강조하는 답변을 하지만 Alexa는 좋아하는 음식과 취미가 있는 가상의 인간임을 표현하였다.

<표 10> Alexa와 Google assistant의 선호 묻기 답변 비교분석

과업 1의 질문 예	Alexa	Google assistant
(4) Do you like pasta?	Absolutely, my favorite of all is De Cecco.	Pasta sounds good, but I get my energy from answering questions. And electricity.
(5) Do you like reading?	I love to read! Every book feels like a new adventure.	Reading can be an adventure. I can help you find a good book.

2) 정보 묻기

단원의 주제와 그 주제에 관련된 문화적인 지식, 배경지식을 학습은 매 단원의 목표의 하나이자 읽기 지문을 보다 잘 이해하는 데 중요한 역할을 한다. 두 챗봇은 이러한 역할을 수행하는 데 있어 매우 적절하였다. 본 연구에서 선정한 교과서 주제와 내용 관련된 질문들에 대해 Alexa와 Google assistant는 적절한 정보를 제공하였다. 답변의 길이는 내용에 따라 5개-105개의 단어였으며, 유사한 정보 출처로부터 검색한 것으로 파악되었다. 챗봇 간의 차이는 크지 않았으나, Alexa의 답변이 대개 Google assistant보다 길었다. 특히 how-to에 대한 질문에 있어 Google assistant는 보다 간결한 답변을 제공하는 경향이 있었다. 그 예시를 보면 다음의 <표 11>과 같다.

<표 11> Alexa와 Google assistant의 내용 지식 정보 답변 비교분석

과업 2의 질문 예	Alexa	Google assistant
(6) How can I save electricity?	According to wikiHow, to save electricity, open your curtains whenever you can to let in sunlight instead of using artificial lights. Switch to energy-efficient CFL or LED light bulb and appliances, and try to unplug any devices or appliances that aren't in use. In addition, you can apply weather-stripping and caulk around your home to seal any cracks. This will help keep the temperature stable, saving you money on your heating and cooling costs. For more details on this and other how-to's, ask to learn more from wikiHow. (88단어)	Here's a summary from the website wikihow.com: Steps 1. Embrace natural light. Open up your curtains and let the sun shine in. 2. Change your bulbs, and 3. Turn off the lights. To read more, look for the link in your Google Home app. (42단어)

(6)에서 보는 것처럼 Alexa의 답변은 88단어로 42단어(실제 답변은 22단어)인 Google assistant에 비해 한 번에 훨씬 더 많은 정보를 제공한다. 유사한 질문인 “How can I”에서 모두 Alexa는 97-103 단어, Google assistant는 56-57단어로 답을 말하여 동일한 패턴을 보였다.

언어적인 지식과 관련된 정보의 경우에도 유사한 경향성이 있었다. 즉 Alexa는 Google assistant보다 한 번에 더 길고, 많은 분량의 정보를 제공하는 경향이 있었다. 아래 <표 12>의 (7)번의 예시에 보는 것처럼 Alexa의 정의는 총 31개 단어로 더 길고 자세한 내용을 담고 있었고 어휘 수준도 Google assistant보다 높았다. 마찬가지로 (8)번에 반의어를 묻는 질문 역시 Alexa는 대개 5개의 예시를 제공하는 데 반하여, Google assistant는 3개의 예시를 일정하게 제시하였다. 이는 앞서 <표 6>의 이

독 지수 결과와도 일치하는 것으로 Google assistant의 사용 언어가 더 짧고 이해하기 쉽다는 것을 알 수 있다.

<표 12> Alexa와 Google assistant의 언어지식 정보 답변 비교분석

과업 2의 질문 예	Alexa	Google assistant
(7) What's the definition of 'upcycling?'	Upcycling, also known as creative reuse, is the process of transforming by-products, waste materials, useless, or unwanted products into new materials or products of better quality or for better environmental value. (31단어)	Here's the definition of upcycle: Reuse (discarded objects or material) in such a way as to create a product of a higher quality or value than the original. (23단어)
(8) What's the antonym of 'aggressive?'	"Aggressive"'s antonyms are "defensive", "amicable", "passive", "unassertive" and "coy". (5개)	Here are some antonyms for aggressive: Meek, Friendly, Peaceable. (3개)

3) 문제 해결하기

교과서 단원의 내용 중 말하기, 쓰기 활동 영역에서 다루는 통합 활동과 유사하게 설계된 두 개의 문제해결 과업을 대화자에게 제시하였다. 본 연구에서 대화자는 각 과업당 7-8개의 질문을 각각의 챗봇에 던졌는데 Google assistant가 보다 높은 성공률과 적합한 답변을 제공하였다.

첫 번째, 다소 모호한 질문에서 Google assistant가 대응을 더 잘하였다. <표 13>의 (9)와 (10)과 같은 다소 모호한 질문을 대화자가 던졌을 때, Alexa의 경우 아예 답을 하지 못하였지만, Google assistant는 매우 정확하고 자연스러운 답변을 주었다.

<표 13> Alexa와 Google assistant의 모호한 질문 답변 비교분석

과업 3의 질문 예	Alexa	Google assistant
(9) What do I need when I travel abroad?	Hum, I don't know that one.	I found 14 on the website neamb.com. Here are the first 4. Make sure your passport and visa are up to date., Check for travel warnings and advisories., Get vaccinations and fill prescriptions., and Register your trip..
(10) What's special about Vincent Van Gogh's paintings?	Sorry, I don't know that one.	I'm not sure, but I can tell you the answer to a similar question: What is unique about Van Gogh?

둘째, Google assistant는 모호한 질문에 대해 답변을 포기하기보다 이해를 확인하는 적극적인 상호작용을 이어갔다. 이러한 차례를 주고받으면서 대화를 이어가는 상

호작용은 Alexa에서는 확인된 바 없었다. 다음은 질문 (9)에 이어진 대화이다.

Google assistant: I found 14 on the website neamb.com. Here are the first 4. Make sure (omission) and Register your trip. Want me to tell you more?

대화자: Sure.

Google assistant: Okay. Here's 4 more: Stock your wallet., Buy travel insurance (omission), To see the whole list, follow the link I just sent to your phone.

아래는 질문 (10)에 이어진 대화이다.

Google assistant: I'm not sure, but I can tell you the answer to a similar question:

What is unique about Van Gogh? Would that help?

대화자: yes please.

Google assistant: On the website artble.com, they say:

Vincent van Gogh was a unique artist who worked with a sense of urgency...

<표 13>의 질문 (9)와 (10)은 모두 Alexa는 답변을 포기하였는데 반해서, Google assistant는 확인요청 후 보다 정확한 답변을 하는 것을 볼 수 있다.

V. 논의 및 제언

서론에서도 명시했지만 본 연구에서는 두 가지 세부 연구 문제를 설정하였다. 첫째는 Kim(2018)이 제안한 AI 스피커에 최적화된 과업 유형 중 세 가지 유형에 대한 챗봇의 대화 수행능력을 진단하는 것이고, 둘째는 학습자의 과업 활동 시 각 챗봇이 제공하는 정보는 외국어 교육과정의 관점에서 내용적·언어적으로 적합한지에 대해 살펴보는 것이었다. 먼저 이 두 가지 연구 문제에 대한 실험결과를 간단히 정리하면 다음과 같다.

첫째, 'Alexa'와 'Google assistant,' 이 두 가지 AI는 현재 상태에서도 우리나라 영어 수업에서 활용 가능한 도구로 판단된다. 그 근거로는 교육과정과 연계된 과업을 수행하는 데 있어 86% 이상의 높은 성공률을 보였다. 두 번째 연구문제인 주제나 내용 면에서도 우리나라 학습자에게 적합한 수준의 정보를 제시한 것으로 분석되었다. 두 종류의 AI가 구사한 발화의 어휘 수준만 보더라도 전체 발화의 97%가 초·중·고 교육과정의 권장어휘 수준에 포함되어 있다는 점에서도 적합성이 인정된다고 볼 수 있다. Hu와 Nation(2000)에 따르면 가장 이상적인 텍스트 이해도는 98%이며 최소 95% 정도의 이해도가 확보되어야 의미 있는 학습(meaningful learning)과 나머지 2-5%의 신출 어휘에 대한 우연적 학습(incidental learning)이 가능하다고 주장하였기

때문에 97%라는 수치는 매우 긍정적인 결과라고 볼 수 있다. 또한, 실험 중 제시된 과업 유형별로 수행 성공률을 살펴보면 먼저 ‘정보 묻기’ 유형에서 두 AI 비서 봇 모두 100%의 성공적인 답변을 제공하였다. 이는 AI 비서 봇이 해당 단원의 주제에 대한 배경지식을 쌓고, 신출 어휘의 의미 등을 파악하여 언어적인 지식을 확장하는 도구로서도 매우 적합하다고 판정할 수 있는 근거가 된다는 점에서 의미가 있다. 반면 ‘잡담 나누기’에서는 대화의 실패가 가끔 있었는데(Alexa 성공률: 84.1%, Google assistant 성공률: 75.8%), 이는 높은 실패율은 아니나 이 정도 실패율만으로도 실제 대화에서는 대화의 단절을 유발할 여지가 충분하기 때문에 아직까지는 주제를 한정하지 않은 자유 대화를 나누기에는 어려움이 예상된다. 지금까지는 두 개의 AI 비서 봇의 공통적인 실험결과라면 다음에 제시할 내용은 두 AI 비서 봇의 챗봇의 차별적인 특성을 정리한 것이다.

먼저 첫 번째 두드러진 차이점은 Google assistant가 Alexa보다 조금 더 외국어 학습자 친화적이라는 점이다. Google assistant는 발화 시 상대적으로 짧은 문장을 구사하여 듣기이해에 도움이 되었고 하나의 질문에 대한 답변에서 제공하는 정보량도 가능한 정보를 모두 제공하는 Alexa보다 적절하여 대화자의 인지 부담을 낮추는 효과가 있었다. 두 번째로는, 발화의 이독 지수 역시 Google assistant가 영어 원어인 11-12세 수준으로 분석이 되어, Alexa의 13-14세 수준에 비해 더 쉬운 영어를 구사한다는 점이다. 세 번째는 Google assistant가 모호한 질문에 대한 대응력이 더 우수하다는 것인데 즉, 학습자가 구사한 질문이 정확하지 않더라도 가능한 수준의 이해를 통한 답변을 시도한다는 점은 챗봇의 기능으로서 매우 중요한 부분이라고 판단된다.

하지만 학습 목적에 따라서는 Alexa도 상대적인 강점을 가진 것으로 분석되었다. 예컨대 ‘잡담 나누기’에서 좀 더 인간에 가까운 표현을 구사하기도 하고 ‘정보 요청하기’ 유형에서는 더 많은 정보를 제공함으로써 언어 입력량 측면에서는 보다 효과적이라고 말할 수 있다. 그러므로 어떤 AI가 더 우수하다고 단정 짓기보다는 대화자의 영어 능숙도, 사용 학년, 활동설계 등에 따라 적합성이 달리 판단해야 할 것으로 보인다.

본 연구는 비록 새로운 시도이며 다각적인 분석을 제시하려 노력하였으나 연구 결과의 일반화 측면에서는 제한점을 인정할 수밖에 없다. 첫째, 연구의 성격이 실험적 성격이 강했기 때문에 데이터의 규모가 크지 않았다. 따라서, 보다 일반화된 결과를 도출하기 위해서는 다양한 영어 수준과 성격의 학습자들과 과업을 통해 대규모 데이터의 구축이 필요할 것으로 보인다. 또한, 본 연구는 상당 수준의 영어 능숙도 가진 대화자가 AI 비서 봇과 대화를 시도한 실험으로 비영어권의 다양한 발음이나 부적절한 표현 구사, 또는 다수의 대화자 참여 등과 같은 고려하지 못한 구두(oral) 변인이 추가되었을 때 AI 비서 봇의 대화 인식률이 어떻게 달라질지에 대해 확인하지 못했다. 하지만 이러한 연구의 한계에도 불구하고 본 연구는 다음과 같은 시사점을 도출

할 수 있었다.

첫째, 외국어학습, 외국어 교육과정에 근거한 과업을 수행할 수 있는 목적형 챗봇 프로그램 개발이 필요하다. 현재의 영어 챗봇은 주로 비목적형이기 때문에 목표 대상 학습자들의 요구에 맞는 과업수행능력이 부족하다. 특히 ‘잡담 나누기’와 같은 기초 의사소통기능을 연습하는 일상대화 질문에 대한 적절한 수준의 답변을 제공할 수 있는 설계가 필요하다. 둘째, 외국어 학습자 코퍼스 구축 및 분석을 통해 외국어 학습자의 발화 특성을 도출하고 이를 챗봇 개발에 반영할 필요가 있다. 이와 관련하여 세부적으로 개선될 챗봇의 기능을 예로 들면, 대화자가 알고자 하는 내용에 대한 질문이 모호하거나, 단어 및 형식사용이 다소 오류가 있을지라도 챗봇은 확인요청 (clarification request) 등의 의사소통전략을 사용하여 문제를 해결할 수 있는 능력이 있어야 한다. 셋째, 학습자에게 더 많은 발화의 기회를 제공하고 발화를 유도할 수 있도록 대화를 주도하는 기능이 추가될 필요가 있다. 현재의 기술 수준으로는 아직 완벽하진 않지만 <표 13>의 (9)와 (10)처럼 관련 주제에 대하여 연계된 추가 질문하기 등이 이와 같은 기능의 대안이 될 수 있을 것이다.

끝으로, 본 연구는 부족한 부분이 있기는 하지만 지금까지 막연하다고 생각했던 AI 비서 봇의 외국어학습 활용에 대한 가능성과 개선점이 본격적으로 논의되었다는 점에서 의의가 있으며, 이를 통해 보다 구체적인 교수·학습 활용방안을 마련할 수 있게 되었다고 생각된다. 앞으로 실제 수업 활용 연구 등이 활발히 진행되어 외국어학습용 챗봇 개발과 활용을 통한 학교 영어교육의 발전을 기대한다.

참고문헌

- 김정렬, 양지윤 (2012). Coh-Metrix를 통한 초·중등 영어교과서 연계성 분석. **영어교육**, 67(2), 319-341.
- 김재상 (2017). 인간과 인공지능 기기의 상호작용이 EFL 환경에서 초등학생들의 협업적 언어습득에 미치는 영향. 광주교육대학교 대학원 석사학위논문.
- 김재은, 최인철 (2015). 고등학교 영어 교과서, EBS 수능 연계 교재, 대학수학능력시험의 코퍼스기반 난이도 비교 분석. **멀티미디어언어교육**, 18(1), 59-92.
- 김형순, 김혜영 (2017). 4차 산업혁명시대의 영어교사 미래준비도 연구. **멀티미디어언어교육**, 20(3), 179-205.
- 서정은 (2017). 인공지능 음성인식 시스템 기반 유아 영어학습 사례연구. 중앙대학교 대학원 석사학위논문.
- 성일호 (2014). 이독성공식과 Coh-Metrix를 활용한 우리나라 고등학교 영어교과서 이독성 분석. **영어영문학연구**, 40(4), 299-320.
- 이문복, 신동광 (2015). 2015 영어과 교육과정 기본 어휘 목록 개발. **영어교과교육**, 14(4), 115-134.
- Chall, J. S., & Dale, E. (1949). The concept of readability. *Elementary English*, 26, 23-33.
- Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society*, 13(2), 13-24.
- DuBay, W. (2004). *The principles of readability*. CA: Impact Information.
- Fryer, L., & Carpenter, R. (2006). Bots as language learning tool. *Language Learning and Technology*, 10(3), 8-14.
- Gobel, P., & Mori, S. (2007). Success and failure in the EFL classroom: Exploring students' attributional beliefs in language learning. In L. Roberts, A. Gürel, S. Tatar, & L. Martı (Eds.), *EUROSLA Yearbook 7* (pp. 149-169). Amsterdam: John Benjamins Publishing Company.
- Golosio, B., Cangelosi, A., Gamotina, O., & Masala, G. L. (2015). A cognitive neural architecture able to learn and communicate through natural language. *Plos One*, 10(11), 1-37.
- Gordon, G., Spaulding, S., & Westlund, J. K. (2016). Affective personalization of a social robot tutor for children's second language skills. *In Proceedings of*

- the 30th AAAI Conference on Artificial Intelligence* (pp. 3951-3957). Phoenix, USA: AAAI Press.
- Heatley, A., & Nation, I. S. P. (2002). *RANGE and FREQUENCY programs*. Retrieved June 15, 2018, from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Range_GSL_AWL.zip.
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403 - 430.
- Huang, J. X., Lee, K. S., Kwon, O. W., & Kim, Y. K. (2017). A chatbot for a dialogue-based second language learning system. In K. Borthwick, L. Bradley, & S. Thouësny (Eds), *In Proceedings of the EUROCALL 2017* (pp. 151-156). University of Southampton, UK: Research-publishing.net.
- Jordan, A. (2014, June). Meet Pepper, Softbank's Emotion-Reading Robot. *Tokyo Weekender*. Retrieved September 19, 2018, from <https://www.tokyoweekender.com/2014/06/meet-pepper-softbanks-emotion-reading-robot/>.
- 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*, 61-84.
- Kim, H. (2018, June). Designing L2 interactive tasks with an artificial intelligence robot. *Paper presented at Asia TEFL International Conference*. University of Macau, China. Retrieved September 21, 2018 from <https://www.slideshare.net/heyoungkim/designing-l2-interactive-tasks-with-a-n-artificial-intelligence-robot>.
- Lee, S., Noh, H., & Lee, J. (2011). On the effectiveness of robot-assisted language learning. *ReCALL, 23*(1), 25-58.
- Mubin, O., Bartneck, C., & Feijs, L. (2012). Improving speech recognition with the robot interaction language. *Disruptive Science and Technology, 1*(2), 79-88.
- WebFX. (2012). *Readability Test Tool*. Retrieved June 15, 2018, from <https://www.webfx.com/tools/read-able/>.
- Willis, D. (1996). *A framework for task-based learning*. London: Longman.
- You, Z. J., Shen, C. Y., Chang, C. W., Liu, B. J., & Chen, G. D. (2006). A robot as a teaching assistant in an English class. *In Proceedings of the 6th International Conference on Advanced Learning Technologies* (pp. 87 - 91). Washington, DC: IEEE Computer Society.

논문 접수: 2018년 10월 8일

논문 심사: 2018년 11월 9일

게재 승인: 2018년 11월 21일

<ABSTRACT>

A study of AI chatbot as an assistant tool for school English curriculum

Kim Heyoung(Chung-Ang University)

Shin Dong Kwang(Gwangju National University of Education)

Yang Hyejin(Yonsei University)

Lee Jang Ho(Chung-Ang University)

This study aimed to investigate the possibility of using artificial intelligence (AI) chatbot in English class curriculum of secondary schools in Korea. A participant asked questions to the two representative AI chatbots. The questions were about exchanging small talk, asking for information, and solving problem, adopted from middle and high school English textbooks. The audio recorded conversation were transcribed and analyzed quantitatively in order to examine success rate, vocabulary, readability index, and the length and the amount of information from the chatbot. Quantitative analyses were also carried out to examine the difficulty of the vocabulary, the adequacy and quality of the conversation with the chatbot. The results revealed that both chatbots showed a high success rate of 86% or more in carrying out the given tasks, and provided adequate responses to the participant's questions. Next, 97% of the vocabulary used by the chatbots was included in the recommended vocabulary level of English class curriculum in Korea. Furthermore, the Google assistant was more user-friendly than Alexa as its utterance was easier for them to comprehend. Google assistant also presented excellence in responding to the participant's ambiguous questions. Finally, this study provided practical suggestions for integrating chatbots into English class curriculum, and for future development in AI speakers as tools for foreign language education.

★ **Key words:** Artificial intelligence (AI), Chatbot, English education, AI speaker