SIRI and Language Learning: A Case Study

Andrew Kirby

New York University

Advanced Individual Project

LANED-GE.2039.002

May 9th, 2016

Introduction

In this project, the effectiveness of using a software-based 'virtual assistant' in a second-language learning context will be evaluated. This particular assistant is SIRI, an interactive program designed by Apple Ltd. to assist users with various computing and mobile phone tasks. It is connected to a search engine. The purpose for conducting research on this particular program is related to the widespread use of smartphones in everyday life. Language learning is connected to the habits and behaviors of the learners; tasks in educational contexts should be adapted to them as such. Younger students may feel more comfortable using such devices in a school environment.

As technology becomes more ubiquitous in everyday life, tasks in the classroom will change in order to reflect this ubiquity. It is my prediction that human and computer program interaction will increase in the near future. Adaptations will be made on both ends; while many companies create so-called 'user-experiences' that allow the user to behave as they would when interacting with a human, their designs are often imperfect. People too must change in order to interact with the technology around them. Language learners, whose human to human interactions are already impeded by their lack of linguistic knowledge, can in turn face difficulties when interacting with technology that are both similar and different to that of a human.

As interactive programs spread from mobile devices to home appliances, automobiles, and beyond. For language learners, and for educators, generalizing tasks to new forms of media is necessary in order to maintain fidelity to methods of learning, behavior, and interaction which are familiar to their students. SIRI, a virtual assistant created by Apple, is one of the first steps in this process. This assistant is designed to gather information, execute certain device functions,

and provide a form of interaction and feedback to the user. It is possible that students and language learners may find themselves using this technology in the future.

Mobile-Assisted Language Learning (MALL) is a relatively new field in which mobile devices are used to assist with language learning. These devices provide a variety of methods in which a learner may acquire knowledge: through apps, accessing websites, or doing work that was once done on paper or a computer with a newer interface. An advantage of MALL is its novelty value: many have never used their phone for learning; for many, it is the vector from which entertainment and communication springs. Another advantage is its convenience: mobile devices are available anywhere anytime, provided there is a signal and enough battery power. Potentially, learning could happen around the clock.

To better understand the applications of virtual assistants, and artificial intelligence-based tools generally, I have designed a series of tasks which should demonstrate the utility (or lack thereof) of such tools for the purposes of language learning. What is needed is a way of determining how much a language learner can improve specific skills through the interaction an artificial intelligence that has been designed for another purpose; in this case, for gathering information and facilitating operation of a device. As of this time, little consideration has been given to the potential utility of artificial intelligences in education, despite the growing wave of automation which will eventually affect all sectors of the economy. With these ideas in mind, I have concocted a series of research questions for investigation.

Research Questions

The questions that are raised in this project are the following:

1. To what extent did a short-term, task-oriented virtual assistant intervention help high school ESL students to improve their ability to
    a. demonstrate receptive knowledge of wh-questions?

   b. compose simple present tense sentences in written form about a given topic?
2. What obstacles did the learners encounter in terms of technology and in terms of language use?

The plan is to investigate these questions in a public school setting with high school students. The students will complete a series of structured tasks using SIRI. The products of these tasks, along with a multiple choice test, will be used for analysis and discussion. I will consider implications for both educators and for researchers who plan to conduct their own studies on human interaction with artificial intelligences in a language learning context. Can virtual assistants be used in the classroom to promote language learning?

To what extent are these assistants effective? To what extent did these obstacles prevent successful implementation of the intervention? What modifications, if any, need to be made to better adapt this technology to a learning environment? These considerations will be addressed in the discussion section.

Literature Review

Before advancing to the present study, I will briefly review the current literature in three domains: MALL (Mobile-Assisted Language Learning), conversation analysis in terms of human-device interaction, and the Output Hypothesis. The relationship between these three elements form the theoretical basis for which the interaction between a learner and a virtual assistant can be described in a language learning context. While research concerning CALL (Computer-Assisted Language Learning) has been around for decades, the subset of research concerning MALL (Mobile-Assisted Language Learning) is much more recent. The utility of mobile devices in general have not been fully realized in educational contexts (Yang 2013). Studies have been conducted on the use of mobile devices and programs on these devices; the

responses to the use of such devices as a learning medium by learners have been generally positive (Wang & Vazquez 2012). The gap in the literature here is the use of a virtual assistant artificial intelligence program in a language learning context; this study intends to break ground in that specific area.

In terms of conversation analysis, the first step here is to review how, if at all, pre-existing conversational skills are adapted and oriented towards interaction either with devices or simply using the device as a medium. For example, when two people are interacting with one another through text messages, they will use turn-taking strategies and symbols in an attempt to emulate 'real-life' conventions in face-to-face interaction (Gonzalez-Lloret 2015). However, when working with a virtual assistant, the person is interacting with a form of 'Ambient Intelligence' which is defined as "the presence of a digital environment that is sensitive, adaptive, and responsive to the presence of people" (Cook, Augusto, & Jakkula 2009). The degree to which the virtual assistant SIRI fulfills this definition is in question: its sensitivity and adaptability are part of its functions, but it does not yet (as of 2016) have the specifications to tolerate error or work with more abstract demands for information gathering.

Relatively older forms of theory are still applicable in such new contexts. While the media with which language learners interact may change, the problems which they encounter do not. As language learners encounter obstacles in these interactions, they will seek out ways to negotiate these obstacles; through their output, they will accumulate language knowledge. This is known as the Output Hypothesis (Swain 2000). In the context of interacting with SIRI, an adaptive and responsive computer program, the obstacle provided SIRI's ability (or lack thereof) to reconcile its pre-programmed notions of 'authentic' English with that of the learner. This obstacle should then facilitate self-correction and adaptation on the part of the learner; after

repeatedly surmounting this obstacle, the learner should hopefully be able to retain new language knowledge.

In summary, the language learner can improve his or her language abilities by interacting with SIRI in a human-like way; the problems that he or she may encounter in this process will lead to learning. In the context of conversation analysis, the learners are interacting with SIRI, a kind of ambient intelligence under a the loose pretext that this intelligence will respond as a person would. The language learning component enters when the learners encounter obstacles as they attempt to communicate with SIRI and succeed in finding effective ways to successfully interact. My study contributes to this literature by exploring new applications of MALL in this context.

Description of Context

This research project took place at El Puente Academy for Peace and Justice, a public high school in Brooklyn, New York. The two students I invited to participate, Joffre and Ericksson, are both English Language Learners in the 9th grade, and are native speakers of Spanish. Joffre has recently come to the US from Ecuador and speaks and writes little English. Ericksson, from the Dominican Republic, knows more English and is able to consistently and correctly form questions in English, but has trouble with expressing himself in English through writing.

While the students at the time of this project were attending freshman-year classes and receiving instruction in the English language, they were receiving little to no formal English language instruction. These students themselves do not own smartphones or tablets, but they are generally familiar with how to use such devices.

Methodology

To conduct this research, the virtual assistant SIRI, an Apple product, was used alongside two pen-and-paper methods. The students were allowed to use Google Translate to fill any gaps in knowledge that might occur while working on the tasks. During the tests, the students were not allowed to use Google Translate or any tools. This was done in order to measure information retention and their ability to independently produce English orally and in writing.

The tasks consisted of a series of topics such as "Tell me about Times Square", followed by a series of bullet points in which a group of parameters for what exactly the students should say, such as 'location' and 'travel time' or 'population'. These instructions were translated if the student could not understand them. The student would then try, to the best of his or her ability, to fulfill all of the parameters of the written task. The purpose of this task design was to give structure to the experience of interacting with SIRI. In addition, the parameters were chosen specifically because it was information that could be obtained through the use of specific questions, that were then measured in the multiple choice test. The written product also provided data (a series of approximately five simple, present tense sentences) which could be qualitatively analyzed, as the quantitative multiple choice component was limited solely to evaluating the students' grammatical and syntactical knowledge concerning question generation. There was no time limit requirement; each task took about 20 minutes to complete.

Because interacting with SIRI was often problematic, the students were asked to try speaking each question to SIRI no more than three times. If, after three attempts, the student was unable to receive any information, the student was allowed to type in his or her question directly, and then produce information accordingly. The purpose of instructing students to prioritize orally attempting to obtain information from SIRI is to utilize the feedback function of the program. If

the student's pronunciation or grammar and syntax were too inaccurate, SIRI would give incorrect results or no results at all.

Data collection

Two types of data were collected. One is a multiple choice test, which measures quantitatively the learner's ability to correctly form questions such as "How can I get to ____" or "What restaurants are in ____". The other form of data that was collected was a written task which provided way to measure language ability qualitatively. This component was a series of statements about a given topic about which the student was required to gather information through the use of SIRI. This part is more open: the learners were encouraged to provide information on a series of points, but the way in which they produced this information could vary.

The multiple choice test was selected to provide a quantitative measurement of the student's ability to posit wh-questions and to provide a picture of the difficulty with which the student may have trouble interacting with SIRI. The same multiple choice test was used for the pre- and post-test, and I compared the results.

The written task was selected for a number of reasons. First, the written task is open to a wide range of answers; even if the student had answered all of the multiple choice questions correctly, progress or variance can still be measured through the written component. In addition, a qualitative measurement may help to inform gaps found in the quantitative multiple choice test. And, as stated above, it may also be useful in comparison to the multiple choice test and can assist with the inference of correlations in the data from the quantitative component.

In addition, an informal interview was conducted at the end of the study. The purpose of this informal interview was to take into account certain emotional factors that could influence the

learners ability to make progress. Finally, I also took notes throughout the process, observing the responses SIRI gave, any obstacles the students encountered, and how the students responded emotionally to the device.

Data analysis

To answer research question 1a – to what extent did a short-term, task-oriented virtual assistant intervention help high school ESL students to improve their ability to demonstrate receptive knowledge of wh-questions – the scores on the pre- and post-tests on the multiple choice component were compared, and then a conclusion was derived. To answer research question 1b – to what extent did a short-term, task-oriented virtual assistant intervention help high school ESL students to improve their ability to compose simple present tense sentences in written form about a given topic – the writing task of the pre- and post-test were compared. Sentences with fewer errors in grammar, an increase in diversity or quantity of vocabulary, and an increased ability to complete all of the bullet points of the task were considered to be signs of improvement. An observed decrease or no change in the aforementioned areas were considered to be signs of regression or inconclusive.

To answer research question 2 – what obstacles did the learners encounter in terms of technology and in terms of language use? – I conducted informal interviews after the administration of the tasks which consisted of the following questions: "Did you enjoy learning English in this way? Why?", "Was this easy or difficult for you? Why?" and "Would you do this again in the future? Why or why not?". This information was documented and outlined in the results. I also documented examples of SIRI's responses and feedback, and whether or not it was helpful to the learners when attempting to obtain information.

Results

Research Question 1a) Wh-questions

For the multiple choice tests, the results are as follows:

Ericksson pre-test: 11/14	Joffre pre-test: 5/14
Ericksson post-test: 5/14	Joffre post-test: 5/14

For Ericksson, there was a significant drop in his test score. While this drop could indicate a serious regression over a one-month period, it is unlikely that SIRI or this experience actively damaged the learner. Thus the results of this test are inconclusive.

For Joffre, although he received the same total score, he answered different questions correctly. Two inferences can be made. First, it is possible that Joffre did not demonstrate an increase in receptive knowledge of wh-questions, and was guessing on the multiple-choice test. It is also possible that he was guessing on the pre-test, while attempting sincerely on the post-test. These results can be correlated with the written results, in which some small improvements can be observed.

Research Question 1b) Writing Task

For the writing pre- and post-test, Ericksson has demonstrated some slight improvement in his ability to create complex sentences. For example, in the pre-test, Ericksson wrote "I like it Bushwick because...", whereas on the post-test, he wrote "I like Times Square because...". This can be interpreted as a sign of improvement because he did not incorrectly insert the word 'it' into the sentence to refer to an already-designated object. In addition, Ericksson also was able to create a sentence about one item of the task in the post-test which he did not create in the pre-test. In the pre-test, he simply copied the address of a restaurant from SIRI without creating a sentence. In the post-test, he wrote a complete correct sentence on his own: "a restaurant near central times square is junior's restaurant." This is indicative of an increase in understanding of

syntax and grammar, and, generally speaking, how to better compose simple sentences using the present tense. Another observed difference occurred in the speed and comfort with which Ericksson used SIRI; he was more quickly able to get the program to understand his questions and operate the device to obtain information.

For Joffre, it is once again difficult to interpret the results. In the pre-test, his sentences are completely unrelated to the topic, despite having had the task instructions translated to him in his native language. The sentences appear to have been copied from an IMDB (Internet Movie Database) website profile of a Spanish-language television show, with no original sentences of his own. The copying continued in the post-test, this time from a Wikipedia article on Times Square. However, he contributed phrases of his own, such as "neighborhood in Times Square is johns of Times Square" and "like Times Square is beatiful". On one hand, the results suggest a small improvement by way of the emergence of original phrases in English, where none had appeared before. This could indicate a small step toward understanding of how to construct simple sentences in the present tense. However, the improvement is so small, and the comparison is originally based on a misunderstanding of the task. Therefore, it is difficult to come to any meaningful measurement of the result.**Research Question 2) Obstacles in Implementation**

At the start, Joffre was completely unable to get information from SIRI using the voice function. Both of these problems can be attributed to his lack of acquaintance with technology and his lack of knowledge of the target language. In the post-test, Joffre was able to independently obtain results from SIRI and produce sentences, despite having had chronic problems with pronouncing words in a way that SIRI would be able to understand him. A correlation can also be drawn to the multiple choice results: while it is likely that Joffre was guessing on the multiple choice, it is possible that he retained information on how to form

queries that SIRI could understand after repeated challenges. Like Ericksson, Joffre also was better able to use SIRI and get it to understand him over time; overall, however, he frequently had problems with getting the device to understand his pronunciation throughout the project.

Initially Ericksson was able to use SIRI and other electronic resources more successfully than Joffre, who displayed frequent frustration and was often unable to orally direct SIRI to retrieve certain pieces of information. His pronunciation was closer to that of a standard native English speaker (as defined by the creators of SIRI) than Joffre's pronunciation, and he therefore had more success with oral commands.

The learners' lack of experience with the technology appeared to be a factor in their success. While both students gradually became more able to use SIRI and the iPad, problems continued to arise, even in the post-test. Both students reported that they enjoyed using the device. Ericksson enjoyed using it to read through extensive information and the images that accompanied it; Joffre stated that he enjoyed the fact that SIRI could listen to him and produce information rapidly.

SIRI is designed to give a form of error feedback that is not geared toward language learners, but toward general users. This feedback comes in the form of incorrect search results, or a message: “Sorry, I didn’t get that.” SIRI is extremely intolerant of mispronunciation, but it will sometimes parse incorrect grammar or syntax and generate a useful result. For example, SIRI will show that it heard “How can I getting to Central Park?” and generate directions to Central Park. Because the students were focusing on locations in New York City, location names that were not part of the program’s English lexicon (such as Dos Hermanos Tortilla Factory) would prove to be problematic when attempting to gather information.

The intolerance of SIRI toward inaccurate pronunciation provided both students, especially Joffre, with an obstacle that they could overcome, if very slowly. Based on the increasing speed with which the students were able to successfully complete intended oral commands, the idea of the Output Hypothesis is displayed as being predictive of success, thanks to SIRI's rejection of the command or incorrect production of information. With that in mind, SIRI by itself could not tell the students what their problem was or how to fix it; this job still must be accomplished by a person or a program that is designed to give such feedback.

Another key factor influencing the relevance of these results is the lack of formal English language education for both students; the improvement in performance may have been much more pronounced had the students been receiving instruction on grammar concerning how to form questions, how to use English vocabulary, and how to form sentences. Being confronted by and negotiating the obstacles alone was not enough to produce drastic improvements in the knowledge of the target language.

Discussion

Research Question 1a) Wh- Questions

Because the results of both learners' multiple choice tests were inconclusive, it is not possible to say whether or not the intervention improved their ability to form wh- questions. While success on the written component it dependent upon the student's ability to start a search orally, this search does not necessarily need to come in the form of a conventional question. In order to provide a better answer to this question, the task would need to be redesigned to combine formal instruction on the formulation of wh- questions with the use of SIRI.

There was some observed improvement in pronunciation by way of an increased acceptance rate by SIRI of the learners' questions; SIRI's intolerance of poor pronunciation may have spurred the students to produce more phonetically accurate English, but there is little evidence to show this. Overall, the improvement in pronunciation was not meaningful enough to assign value to this specific intervention over any other.

Research Question 1b) Writing Task

As shown in the results, some slight improvement occurred in the students' writing samples after the intervention had been administered. However, how much of this can be attributed directly to this intervention, to the use of SIRI? There were moments in which SIRI produced usable model sentences for the students. For example, when Eriksson asked for the population of Santo Domingo, SIRI produced the sentence, "The population of Santo Domingo is 1,506,233," whick Ericksson dutifully copied. The task was designed with the expectation that this specific answer would be produced; replicating this was a source of difficulty; finding other items of information in which a learner will ask a question about a specific topical item and consistently receive the same answer in the form of a complete sentence was not possible in this study, but could be made possible in other tasks.

Research Question 2) Obstacles in Implementation

SIRI is extremely easy to access and pick up — only one push of a button is required after it has been activated in the device’s settings menu. Both learners were easily able to activate it and navigate the results after having struggled to obtain them. They also enjoyed the visual depictions of logistical directions on the map and images of restaurants and tourist attractions. For other learners who may not be as ready to accept technology, using SIRI may prove to be difficult or emotionally unpleasant. This is because, as noted in the literature review, people tend

to adapt face-to-face interactions to different media; in this study, the learners also attempted to speak to SIRI as though it were an actual person, knowing all too well that SIRI was not a believable simulation. This disjunction may prove jarring to learners who cannot accept this incomplete simulation of human-to-human interaction.

Implications for Classroom Language Education

Concerning a learning environment, the adaptation of the task to SIRI's parameters was almost completely successful; the same questions tried by the teacher and used in the multiple choice test were later employed (when the student could generate them through translation or from their own knowledge) successfully, and SIRI returned the correct information. However, it is not a substitute for formal instruction, but provides a way for formal instruction to be applied. In terms of feedback, the intolerance and inflexibility of SIRI will only show what has gone wrong, without giving language learners a way to produce the correct sentence that he or she needs to get the required information. SIRI cannot see beyond the learner's error; it is unlikely that modifying the task could facilitate the production of meaningful corrective feedback from SIRI.

Could a sophisticated AI or virtual assistant program be designed for the purpose of language learning, and within it adaptive, self-contained formal instruction modules? Based on the observed performance of SIRI in this study, it is only a matter of time. At this time, is SIRI a replacement for well-designed language learning tasks and formal instruction? The lack of progress measured in this study suggests that the answer is no. Might SIRI make for an interesting activity for intermediate-level students who show a proclivity for unstructured in-class operation of mobile devices? Yes, again provided that the task is meaningfully designed

and adapted to parallel formal language instruction.

Implications for Research

As interactive programs and virtual assistants and artificial intelligences grow more sophisticated, more research needs to be done on their utility in language learning contexts. Future tasks may consist of students using SIRI or similar, more advanced programs to physically navigate complex areas of a city in English, or to shop for goods or services, or to plan out a day using the calendar function. With this in mind, the limitations of this study are ultimately found both in the task and in the properties of SIRI itself. Measurement tools may also vary; it can prove useful to measure oral language production, or to examine more complex written language production.

References

Cook, D. J., Augusto, J. C., & Jakkula, V. R. (2009). Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing*, *5*(4), 277-298.

Gonzalez-Lloret, M. (2015). Conversation analysis in computer-assisted language learning. *Calico Journal*, *32*(3), 569.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. *Sociocultural theory and second language learning*, *97*, 114.

Wang, S., & Vásquez, C. (2012). Web 2.0 and second language learning: What does the research tell us. *Calico Journal*, *29*(3), 412-430.

Yang, J. (2013). Mobile assisted language learning: review of the recent applications of emerging mobile technologies. *English Language Teaching*, *6*(7), 19.

Appendix

Blank test and task items:

Language test Name: _____

1) ____ ____ ____ live in Los Angeles?
   3.8 million people live in Los Angeles.
   a) How people many
   b) How many people
   c) How much people
   d) People is how

2) ____ ____ ____ ____ to New Jersey?
   Take the PATH train from West 4th Street.
   a) How can I get
   b) Where is I go
   c) How do I went
   d) Where did go I

3) ____ ____ ____ ____ in Washington DC?
   You can see the White House.
   a) Where does see I
   b) What can I see
   c) What I can seeing
   d) I can doing see

4) ____ ____ ____ Angelina Jolie?
   Angelina Jolie is 40 years old.
   a) How is old
   b) How age are
   c) How old is
   d) How much years

5) ____ ____ Barack Obama's ____ ?
Barack Obama's birthday is on August 4th, 1961.
   a) When is Barack Obama's birthday?
   b) What do Barack Obama's birthday?
   c) When birthday Barack Obama's is?
   d) When born Barack Obama's is?

6) ____ ____ ____ ____ in Queens?
You can go to the MoMA PS1 art museum.
   a) What can I doing
   b) What can I do
   c) What do I can
   d) What do can I

7) ____ ____ ____ see the movie Kung Fu Panda 3?
Kung Fu Panda 3 is playing at the Roxy Theater.
   a) Where is can
   b) Can where I
   c) Where can I
   d) How I can

8) ____ ____ ____ Ben Affleck in?
Ben Affleck is in Batman vs Superman.
   a) What movie is
   b) Who is movie
   c) Which is movie
   d) What is the

9) ____ ____ ____ it take to get to Central Park?
It takes 45 minutes to get to Central Park.
   a) Where does time
   b) How much minutes
   c) What time does
   d) How long does

10) ____ ____ ____ Chinese restaurants in Astoria, Queens?
There are 20 Chinese restaurants in Astoria, Queens.
a) Do there any
b) Is there a
c) What is the
d) Are there any

11) ____ ____ Michael Jordan from?
Michael Jordan is from Brooklyn, New York.
a) Who is
b) Where are
c) Is where
d) Where is

12) ____ ____ ____ ____ food in my neighborhood?
I found some restaurants.
a) Where do I going
b) Where can I find
c) How do I eat
d) I can get how

13) ____ ____ ____ ____ in the Bronx?
You can go to the Bronx Zoo.
a) What can I do
b) What can I doing
c) What I can do
d) What am I do

14) ____ ____ ____ Mexico City in?
Mexico City is in Mexico.
a) Where is country
b) Is country what
c) What is the
d) What country is

Written Task:

1. Tell me about Times Square. Use SIRI to get information about it.
   Write sentences about the following:
   - Location
   - 1 Thing to do
   - 1 Thing to see
   - 1 Restaurant
   - Directions and travel time from home
   - Like/dislike? Why?

2. Tell me about your favorite celebrity. Use SIRI to get information about it.
   Write sentences about the following:
   - Name of the celebrity
   - Hometown
   - Age & birthday
   - 1 Movie he/she is in
   - Why you like him/her

3. Tell me about your hometown. Use SIRI to get information about it.
   Write sentences about the following:
   - Name of hometown
   - Country of hometown
   - 1 Thing to do
   - Travel time to your hometown from NYC
   - Number of people there
   - Why you like it

4. You are going out with your friend. Use SIRI to get information about a place and directions.
   Write sentences about it below.
   - Name of friend
   - Place you can meet
   - 1 Restaurant you can go to
   - Directions to restaurant from home
   - Travel time
   - Food you can eat

5. Tell me about Central Park. Use SIRI to get information about it. Write sentences about it below.
   - Location
   - 1 Thing to do
   - 1 Thing to see
   - 1 Restaurant
   - Directions and travel time from home
   - Like/dislike? Why?

6. Tell me about your neighborhood. Use SIRI to get information about it. Write sentences about it below.
   - Name and location
   - 1 Thing to do
   - 1 Thing to see
   - 1 Restaurant
   - Like/dislike? Why?