석 사 학 위 논 문
Master's Thesis

Vocabgram: 비디오를 통한 표현 어휘 학습

Vocabgram: Facilitating the Development of Productive Vocabulary using Videos

2019

조 경 제 (Jo, Kyung Je)

한 국 과 학 기 술 원

Korea Advanced Institute of Science and Technology

석 사 학 위 논 문

Vocabgram: 비디오를 통한 표현 어휘 학습

2019

조 경 제

한 국 과 학 기 술 원

전산학부

Vocabgram: 비디오를 통한 표현 어휘 학습

조 경 제

위 논문은 한국과학기술원 석사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2019년 12월 17일

심사위원장 김 주 호 (인)

심 사 위 원 김 유 진 (인)

심 사 위 원 오 혜 연 (인)

Vocabgram: Facilitating the Development of Productive Vocabulary using Videos

Kyung Je Jo

Advisor: Juho Kim

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Master of Science in Computer Science

Daejeon, Korea
December 17, 2019

Approved by

Juho Kim
Professor of School of Computing

The study was conducted in accordance with Code of Research Ethics[1].

[1] Declaration of Ethical Conduct in Research: I, as a graduate student of Korea Advanced Institute of Science and Technology, hereby declare that I have not committed any act that may damage the credibility of my research. This includes, but is not limited to, falsification, thesis written by someone else, distortion of research findings, and plagiarism. I confirm that my thesis contains honest conclusions based on my own careful research under the guidance of my advisor.

MCS
20183576

조경제. Vocabgram: 비디오를 통한 표현 어휘 학습. 전산학부 . 2019년. 22+iv 쪽. 지도교수: 김주호. (영문 논문)
Kyung Je Jo. Vocabgram: Facilitating the Development of Productive Vocabulary using Videos. School of Computing . 2019. 22+iv pages. Advisor: Juho Kim. (Text in English)

초 록

어휘 지식은 크게 "이해 어휘"와 "표현 어휘"에 대한 지식으로 나눌 수 있다. 이해 어휘는 단어의 뜻을 이해하는 능력이고 표현 어휘는 단어를 자유롭게 구사할 수 있는 능력이다. 두 종류의 지식이 언어 구사 능력에 갖는 중요성은 비슷하지만, 대부분의 어휘 학습은 이해 어휘을 돕는 데에 초점이 맞추어져 있다. 선행 연구 조사를 통해 이러한 배경에 이해 어휘 학습을 위한 자료를 제작하고 채점하는 방법이 표현 어휘 학습보다 쉽기 때문이라는 것을 파악하였다. 이 연구에서는 자연어 처리 기술 및 자동 학습 자료 제작 방법을 통해 표현 어휘 학습을 위한 어휘 학습 시스템, Vocabgram 을 제시한다. 시스템의 학습 효과를 검증하기 위해 베이스라인 시스템을 개발하여 비교하는 실험을 진행했다. 실험 결과, Vocabgram 통해 학습한 학습자 집단이 새로 배운 어휘에 대해 34% 높은 자신감을 보였고, 이해 어휘 지식 시험에서는 20.05% 높은 결과를 보였다.

핵 심 낱 말 어휘 습득, 표현 어휘, 외국어 학습, 맥락기반학습

Abstract

Vocabulary knowledge can be divided into receptive—being able to understand the meaning of a word—and productive knowledge—being able to produce a word on one's own account. While both receptive and productive knowledge are equivalently important factors to language proficiency, existing vocabulary learning tasks tend to focus on receptive learning due to difficulties in designing and grading productive tasks. In this work, we present Vocabgram, a two-staged vocabulary learning system that enhances the depth of vocabulary knowledge. Our Vocabgram exposes learners to multiple occurrences of an unknown words in authentic context and provide a series of structured learning activities to ensure effective learning. To assess our system, we implemented a baseline interface to compare with. The results from the user study with 18 participants show that users that learned words with Vocabgram reported a higher level of confidence in regards to the knowledge of new words and performed 20.05% better on a receptive knowledge assessment task.

Keywords Vocabulary acquisition, productive vocabulary, foreign language learning, contextual learning

Contents

List of Tables

List of Figures

Chapter 1. Introduction

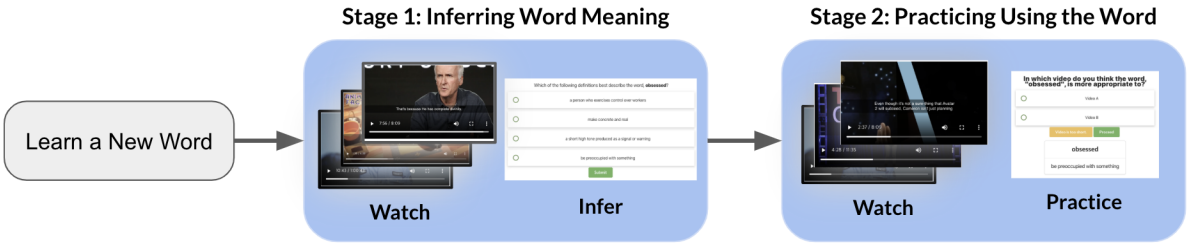

Figure 1.1: An overview of Vocabgram, a video learning interface that combines two types of learning activities to facilitate the development of productive vocabulary knowledge. The first stage encourages learners to elaborately process the context to infer the meaning of an unknown word. Then, in the second stage, learners practice the use of the word in given authentic scenarios.

The level of vocabulary knowledge is considered to be one of the most significant measures to language proficiency [1]. "Knowing a word" is often divided into receptive (or input-based)—knowing the meaning of a word— and productive (or output-based) knowledge —being able to produce a word in speaking or writing [2, 3]. While both types of knowledge are equivalently important factors of one's language proficiency, many studies have shown that people tend to have a larger size of receptive vocabulary than productive vocabulary [4, 5]. For example, one may have the knowledge to understand many words in reading tasks while not being able to fully utilize the knowledge in productive tasks such as in writing or speaking. The findings also imply that two types of knowledge often do not necessarily develop simultaneously. In fact, other researchers have shown that the types of learning affect the type of vocabulary knowledge one acquires (i.e., productive learning tasks led to a greater learning gain of productive knowledge) [6, 7].

There are two main reasons why receptive knowledge far outstrips productive knowledge for many learners. Firstly, productive knowledge is often harder to acquire than receptive knowledge as it requires knowledge beyond just the meaning and the form of a word [2]. In order to develop language proficiency, not only the breadth of vocabulary knowledge (e.g., number of words they know receptively) need to be widened but also the depth of knowledge (i.e., precisely knowing the meaning and use of a word) should be deepened. Secondly, while the type of learning affects the type of knowledge to be developed, receptive learning is more common than productive learning [8, 4]. Receptive learning is more commonly supported in existing systems where learners read, listen, or watch authentic materials such as text, audio, or videos in the target language to develop their language skills.

There has been extensive research on strategies to support the effective acquisition of productive vocabulary. Schmitt claims that vocabulary is not acquired instantaneously but rather learned gradually over time through numerous encounters in authentic situations [9]. Other researchers assert that exposures alone are not enough and vocabulary acquisition is enhanced when learners' attention is directed towards new words [10, 11]. They emphasize that it is important to provide opportunities to learners to elaborately process unfamiliar words in context (i.e., inferring the meaning of the word from the context or analyzing the context the word is used in). In addition, some studies show that structured productive

practices are invaluable to support learners to achieve productive mastery [12, 13].

Videos are a underutilized source of authentic input that offers many opportunities for productive learning. With video platforms such as YouTube and Netflix, learners have a high accessibility to a vast amount of videos that depict the language use in various real-life situations. Zhu et al. have shown that TV episodes and movies cover a wide range of vocabulary that can be pedagogically meaningful to foriegn language learners [14]. Furthermore, learners can construct a more comprehensive knowledge of a word (e.g., pronunciation or spelling of a word) watching videos as they are a medium of multi-modal input. Videos illustrate the language use associated with a rich context that range from relationship to emotion to setting. While these benefits that are offered in videos have the potential for productive learning, the use of videos in most of existing language learning interfaces is often limited to receptive learning.

In this paper, we present Vocabgram, a video learning interface that combines vocabulary learning activities to facilitate the development of productive learning. The core idea behind is to utilize videos as a productive learning resource. We make use of transcripts and audio of videos to generate a series of structured learning activities to support effective learning. The main distinction of Vocabgram to existing systems is that we turn videos into a resource for productive learning. In Vocabgram, we provide a two-staged vocabulary learning process that includes two types of learning activities in a series. Learners first infer and learn the meaning of a target word by watching several videos, and then complete practicing activities to master the productive knowledge of the target word.

To assess the validity of Vocabgram, we conducted a user study with 18 participants where we compared Vocabgram to a baseline interface. Results suggest that users who used Vocabgram gained greater knowledge of productive vocabulary. In addition, users that used Vocabgram showed a higher self-reported knowledge on new words, which indicates a greater confidence gain. On the receptive knowledge test, users that used Vocabgram performed 20.05% better than the baseline group.

The contributions of this work are as follows:

- A prototype vocabulary learning interface, Vocabgram, that turns videos into a productive learning resource. Vocabgram provides a two-staged learning that includes inferencing and practicing stages to facilitate the development of productive vocabulary knowledge.
- A computational pipeline that enables generation of activities using authentic videos. We show how to combine various techniques which include Natural Language Processing (NLP), forced alignment, and Word Sence Disambiguation (WSD) to generate activities.
- Quantitative and qualitative findings from user study suggesting that Vocabgram enhances the depth of vocabulary knowledge and learners' perceived confidence in new words.

Chapter 2. Related Work

In this section, we review research on 1) types of vocabulary knowledge, 2) theories of vocabulary acquisition in language learning, 3) vocabulary learning in context, and 3) existing vocabulary learning systems.

2.1 Types of Vocabulary Knowledge

Vocabulary knowledge can often be divided into receptive and productive vocabulary knowledge (or input- and output-based knowledge). While both types of knowledge are important factors of one's language proficiency, the goal of many vocabulary learning tasks is primarily on developing receptive knowledge [8]. Webb explains that receptive tasks are more common as they are easier to design, grade, and complete than productive tasks [4]. Unfortunately, many studies support that the two types of knowledge do not necessarily develop simultaneously. For example, many studies have shown that people have a larger size of receptive vocabulary compared to that of productive vocabulary [15, 5, 16]. Also, many researchers suggest that productive vocabulary knowledge is harder to acquire than receptive ones since learners need to know beyond a word's form and meaning if they are to use the word productively [2, 8].

Earlier studies show that types of learning affect the type of vocabulary knowledge one acquires [6, 7]. Receptive learning (e.g., reading glossed sentences or learning from word pairs) is found to be more effective for developing receptive knowledge, whereas productive learning (e.g., writing sentences or completing gap-filling questions) is effective for developing productive knowledge. However, productive learning is often found to be superior to receptive learning since receptive knowledge is often naturally acquired during productive learning [4]. In addition, given that vocabulary learning is predominantly receptive, Griffin and Harley suggest the need for attention to productive learning [6].

Inspired by the findings, this work proposes Vocabgram that provides a two-staged vocabulary learning to facilitate the development of productive vocabulary through utilizing videos.

2.2 Vocabulary Acquisition Theory

There has been extensive research on how one acquires vocabulary knowledge. An accepted view is that vocabulary is acquired through either incidental or intentional learning [17, 8, 18]. Incidental vocabulary learning is when the vocabulary is acquired through an interaction in which the primary focus is not on learning the word. On the other hand, intentional learning is when learning occurs with a deliberate attempt to commit vocabulary knowledge into memory (e.g., learning with flashcards).

2.2.1 Incidental Vocabulary Learning

Earlier researchers argue that incidental learning is responsible for the most of vocabulary acquisition, given the amount of vocabulary one naturally acquires without explicit instruction [19, 20]. Even though how incidental learning takes place remains an open question [21], repeated exposures to authentic inputs, such as extensive reading, are suggested as the key to how learners become familiar with new

vocabulary in various aspects(e.g., its form, meaning, and use) [2, 22, 23, 4].

2.2.2 Intentional Vocabulary Learning

Other groups of researchers suggest intentional learning as an effective way of vocabulary acquisition [24, 25, 18]. Some researchers assert that the incidental way of acquiring vocabulary is a slow and error-prone process because learners may not always notice unfamiliar words (e.g., skipping unknown words in reading tasks) or not capable to infer the correct meaning from context [2, 26, 16, 27]. Furthermore, many researchers argue that simply encouraging learners to spend more time on reading and listening, hoping it will lead to incidental learning, may not be sufficient [28, 29, 30, 31]. Instead, they suggest directing learners' attention toward new words using learning tasks (e.g., reading tasks along with activities related to a word's meaning, form, or use) is the key to enhancing vocabulary acquisition. Hulstijn points out that such activity should be designed to guide learners to elaborately analyze and process context and practice inferring the meaning of the word based on their processing [17].

Multiple exposures and directing learners' attention to unfamiliar words are found to be key in vocabulary acquisition. In addition, intentional learning can be an effective way of acquiring accurate knowledge of vocabulary. Building upon these findings, Vocabgram is designed to expose learners to multiple occurrences of words and direct their attention to those words through a series of activities.

2.3 Vocabulary Learning in Context

Many studies show evidence that learning vocabulary is most effective when learned in diverse context [4, 10, 32]. The number of exposures, that researchers believe, one needs to gain knowledge of a word varies from six [33] to eight [34] to ten or above [35, 11, 36]. However, many researchers view that the quality of the context is more important than the number of exposures [32, 37, 18]. For example, as Krashen suggests, learning may be most effective when the new input is surrounded by familiar elements, which is also known as "i plus one" level (e.g., when learners know the rest of words in a sentence except for the target word) [38]. Sternberg lists several context cues that are applied in the vocabulary acquisition process which includes cues such as temporal (i.e., when or how frequently is a word used?) or spatial cues (i.e., where can the word be used?) [32]. In addition, Sternberg also lists variables that affect learning such as the importance of the word in understanding the context, the degree of how the meaning of an unknown word can be inferred from the surrounding context, and the density of unknown words in the given context. Inspired from these findings, in implementing Vocabgram, we select words and videos considering the surrounding context.

2.4 Existing Vocabulary Learning Systems

Many systems are shown to be effective in helping learners improve linguistic components such as vocabulary, grammar, and pronunciation. For example, in Duolingo [39], Anki [40], and Memrise [41], learners complete activities on vocabulary and grammar to improve their language skills. An earlier research indicated that 34 hours of Duolingo is equivalent to a full semester of college-level Spanish class [**?**]. Edge et al. introduced MicroMandarin which uses the location of learners teach relevant vocabulary [42]. Cai et al. developed Wait-Learning displays a vocabulary learning quiz at waiting time during

instant messaging [43]. Zhu et al. introduced ViVo which is a video-augmented dictionary that helps learners to acquire knowledge on unknown words that appear in videos [**?**]. Ingenium enables learners to reconstruct visually the relationship between words in a sentence to understand Latin grammar [44]. TransPhoner enables learners learn the pronunciation of words effectively by transforming a word to phonetically similar and highly imageable keywords in learners' native language [45].

Other works showed that well-designed games are also another great source for language learning with their rich context within the fictional world. Crystallize found that collaborative actions in games improved user interactions as well as learning [46]. Toup et al. showed that games can help learners improve communication in crisis response teams [47].

Video learning has become increasingly popular due to its authentic conversations and naturally captured rich contextual information such as relationship between speakers, setting, intention or emotion. Platforms such as FluentU [48], VoiceTube [49], and Viki [50] provide interactions to help learners improve language skills. However, existing video learning systems tend to be focused on receptive (input-based) knowledge, which provides learners lack of opportunities for practicing. To the best of our knowledge, Little research has explored utilizing the benefits of videos to support productive knowledge. Culbertson et al. applied speech recognition technology to enable learners to develop on their paralingusitic ability such as intonation and pitch with the expressions that appear in foreign language videos [51]. In this work, we present Vocabgram, which uses authentic videos to support acquisition of productive knowledge through generating vocabulary learning questions from transcripts and providing opportunities for practicing the words learners have learned.

Chapter 3. Vocabgram

Our major goal in this work is to facilitate the development of productive vocabulary. To achieve our goal, we present Vocabgram, a video-based learning interface that supports learners to acquire the knowledge of productive vocabulary through two stages. In this chapter, we first list our design goals of Vocabgram and provide rationale of activity selection. Then, we describe the two stages of Vocabgram on how learners acquire vocabulary knowledge. Finally, we explain the technical details of Vocabgram.

We had following design goals in designing Vocabgram:

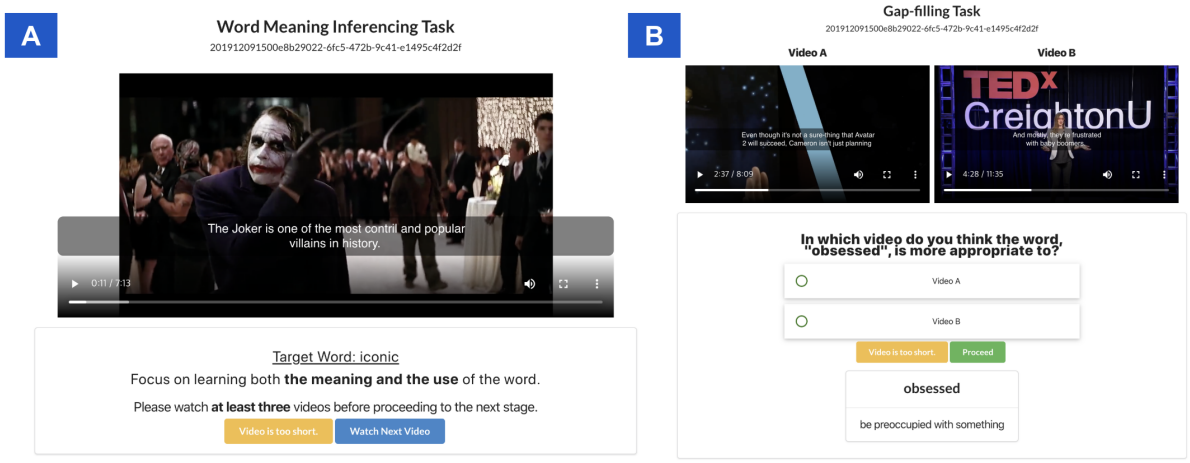

Figure 3.1: A system overview of Vocabgram. Vocabgram has two main stages of vocabulary acquisition: word meaning inferring stage (A) and practicing stage (B). The first stage (A) provides a target word without its meaning and expect learners to watch videos to infer the meaning from the context. The second stage (B) provides a target word and complete a gap-filling activity.

- **To expose learners to authentic use of words in a diverse real-life context.**
- **To guide learners' attention towards unfamiliar words.**
- **To assess learning in both receptive and productive vocabulary knowledge.**

To address these design goals, each of the two stages in Vocabgram provides a vocabulary learning task, inferencing task in the first stage and practicing task in the second stage (Figure 1.1). In the first stage, learners are encouraged to elaborately process the context and infer the meaning of an unknown word. Although the inferencing task is often more recognized as a receptive learning activity, we chose to provide the task to enable initial learning and enhance familiarity with the target word. In earlier studies, inferencing task is shown to be pedagogcically meaningful as it encourages learners to elaborately process the context [32]. In addition, inferencing task is one of the most frequent and preferred strategies which learners use in practice to deal with unknown words [9]. Among various types of productive learning task, we chose to use a gap-filling activity in the second stage because it is one of the most common types of activity and easy to assess.

As learners complete a two-staged learning process, they are eventually exposed to multiple occurrences of the word in different contexts and thus expected to become familiar with new words. Each

learning session is also intentionally designed to learn an individual word at a time which ensures learners' attention to be fully directed towards the single unknown word. At the end of each stage, learners are asked to complete a simple activity (e.g., word matching, gap-filling, etc.) which assesses their understanding. Due to natural handicap in the target language, learners may be incapable to notice any errors in their understanding unless there is appropriate feedback. While generating feedback can be challenging and costly, the results from two activities may play a role as feedback to learners.

3.1 Stage 1: Word Meaning Inferencing Task

The first stage is designed to provide learners opportunities to infer the meaning of an unknown word. In this stage, learners are exposed to multiple occurrences of a target word in different videos. Initially, a target word is given to learners without its meaning as depicted in A of Figure 3.1. So, learners are expected to infer the meaning of the unknown word through watching multiple videos and thoroughly processing their context. After learners watch at least three videos, they have an option to either watch more occurrences in different videos or to take a quiz on the meaning of the word. We ask learners to watch at least three videos to ensure that they are exposed to the use of each target word in many different contexts throughout the two stages. In this design, learners will be exposed to a minimum of six different contexts for each target word throughout their learning session.

In Vocabgram, the quiz is designed as a multiple-choice question that includes a single correct definition of the word. Distractors of each question are selected randomly from the database. When learners choose the correct definition, then they may proceed to learn the next target word or to the next stage if learners have gone through all the target words. If learners choose an incorrect answer, they are asked to watch more videos. Although we could have designed the quiz in a different form such as a freeform question, we chose to design a multiple-choice question as it is easy to grade and provide direct feedback to learners.

3.2 Stage 2: Gap-filling Task

The second stage is designed to provide learners with experience of practicing the use of target words in authentic situations. At this stage, learners are first given a target word with its meaning on the bottom of the page (B of Figure 3.1). Then, learners are asked to watch two videos in which one word from each video is made blank in the subtitle and the audio is muted at the moment the word is used in the video. Then, learners need to choose the video in which the given target word seems to be more appropriate. Learners are asked to repeat the quiz with the same target word until they are correct for three times. Then, they may proceed to practice other target words.

In both stages, videos are initially set to play only for 5 seconds before and after the word is used in the video, which we empirically found to be an appropriate amount of time to grasp the context. But, learners may extend the length of the video segment by clicking "Video is too short" button if they feel the video plays for a too short amount of time to understand the context. Learners may also watch videos as many times as they wish before proceeding to the next step.

3.3 Question Generation

To ensure effective learning, each stage includes a vocabulary learning task. This section describes the two important technical steps to generate the two types of questions in a scalable way: 1) identifying the word's meaning and 2) finding an exact timestamp of words.

3.3.1 Word Meaning Identification

To ensure that learners to effectively acquire vocabulary knowledge, Vocabgram needs to be able to provide correct definition of words. While each word occurrence in a specific situation illustrates a particular meaning of the word, a challenge is that polysemy is common. For example, the word "bass" should be counted differently when it refers to musical frequency and when it is used to refer to a type of fish. Thus, the surrounding context of each word needs to be taken into consideration to identify the correct meaning of a word. To accurately extract words' meaning, we employed an NLTK's [52] Word Sense Disambiguation (WSD), which is an algorithm that identifies the definition of words based on the surrounding context [53]. Each word is then stored into the database with its definition so that occurrences of each word can be accurately retrieved on request.

3.3.2 Word Timestamp Annotation

It is important to identify the exact timestamp of words for many reasons. For example, the timestamp is required to mute the audio at the right time in the gap-filling activity. Furthermore, many subtitles that are uploaded with videos online are out-of-sync with the content which may detract learning experience. Common subtitles provide timestamps in a longer unit of words such as sentences. Crowdsourcing can be a solution to identify the exact timestamps of each word, but employing crowdsourcing alone is both costly and unscalable. We, thus, use a human-machine hybrid approach. We first apply forced alignment algorithm which is widely used to synchronize audio and text. In this work, we used the algorithm of Python's aeneas library []. Because existing forced alignment algorithms are known to show a poorer performance when the synchronization involves a long segment of audio and text, we applied the algorithm in a smaller segment using the original timestamps of subtitles. Then, we use crowdsourcing to verify the timestamp of each word is correctly identified. In this work, we applied our approach to identifying the exact timestamp of 217-word occurrences. In the crowdsourcing stage, a total of 166 timestamps were fixed and crowds spent approximately one minute on average per verification.

The data gathered from the two steps are then used to generate questions for both stages. In question generation, word's definition is used as a distractor in multiple-choice questions, a ground truth in verifying learners' answers or feedback to learners. Word timestamps, on the other hand, are used to generate gap-filling activities (e.g., audio is muted based on word timestamp) or to provide learners videos at an exact time in both stages.

Chapter 4. Evaluation

4.1 Participants

For our study, we recruited 18 participants (male=11, female=7) for the user study through on-campus online community postings. All participants were students at an English medium institution in Korea and had English education through K-12. Except for 2 participants who have resided in an English speaking country for between 6 months to a year, the remaining 16 participants had no experience of living abroad. For the study, participants received an hourly rate of US$10.

4.2 Comparison Interface

To evaluate the effectiveness of vocabulary learning activities in video learning, we developed a baseline interface that uses video but do not include activities. As a baseline for comparison, we adapted the design of existing language learning systems. The design on the baseline interface is a replicated and modified version of ViVo, a video-based vocabulary learning interface [**?**]. Whereas the original version of ViVo uses a video-augmented dictionary (i.e. digital dictionary appears as a layer on top of the video), our baseline interface displays a dictionary on the bottom of the page. The design, such as color and buttons, also followed the design of Vocabgram. The purpose of these modifications is to minimize any possible effects from design factors on the learning experience. Our decision to use ViVo is based on the following reasons. Firstly, ViVo focuses on teaching receptive vocabulary rather than productive alike to many of the existing vocabulary learning systems. Secondly, ViVo makes use of videos as the medium for vocabulary learning, which aligns with Vocabgram. Lastly, ViVo has shown not only its effectiveness in learning gain but also found to be highly engaging. We included this baseline interface to allow us to evaluate the validity of Vocabgram on the acquisition of productive vocabulary and learners' confidence in their producibility of words in actual conversation.

4.3 Study Design

We adopted a between-subjects study design which we evenly divided participants into two groups, one for baseline and the other for Vocabgram. Groups were balanced based on the predicted vocabulary size reported on recruitment. The main reason behind our decision on using a between-subjects study design is to avoid learning and ordering effects. Since we are evaluating the effect on vocabulary acquisition, participants' prior knowledge may be a significant factor. Thus, to minimize the impact from individual differences (e.g., level of language proficiency), we used a pre-determined set of videos and target words that were carefully selected following the procedure described below.

4.3.1 Selection of Videos and Target Words

Three factors were considered in selecting videos and target words for the study: 1) the overall difficulty of the language used in videos, 2) the total number of occurrences of the target words in our dataset, and 3) the density of unknown words in the sentence where the target word is used in. To

measure the overall difficulty of videos, we analyzed the difficulty level of the words used in each video. Based on the findings of Nation and Waring that most adult foreign language learners have a vocabulary size less than 5000 [54], top 5000 frequent word list generated from COCA dataset [55], one of the largest English corpus, was used to classify each word in a binary whether it is "likely to be known" or "likely to be unknown". Then, we used the proportion of each class to sort videos in the order of difficulty level. To ensure that video content is likely to be comprehensible to participants, we filtered videos in which more than 20% of words are classified as "likely to be unknown". Then, We generated a list of candidate target words with words that were used in any video of the initial dataset but not found in the 5000-words list. For each candidate word, we computed the number of occurrences using string matching. Since the density of unknown words can be a significant factor in vocabulary acquisition [32], we only counted the occurrence if the density was low. Based on Laufer's suggestion that 95% coverage is sufficient to allow a reasonable condition for guessing the meaning of the unknown words [56], we only included occurrences if more than 95% of words in the sentence were found in the 5000-words list. Then, we extracted target words if they appear in at least 10 times in our dataset.

Table 4.1: A list of original words and replaced words that are used for the user study

Original Word	Replaced Word
subscribe	expore
obsessed	proburned
breed	pouled
iconic	contril
obey	rasty

Following the procedures described above, a total of 5 target words selected which are: subscribe, obsessed, breed, iconic, and obey. Then, each word was replaced as a pseudoword—words that do not have a meaning but are pronounceable— to avoid any effects from participants' prior knowledge (Table 4.1). We chose to use nouns and verbs for target words because they are two most common and basic form of words used. A total of 138 distinct videos were used in the study.

4.4 Procedure

Participants were randomly assigned to either baseline interface or Vocabgram. During the study, they were neither told about the other condition nor whether the target words are pseudowords. Each participant was given a brief introduction of the study and a tutorial session which walked through each stage and component of the system with visual screenshots. Participants were told to freely ask questions at any time during the study. Then, they were asked to begin a learning session using their assigned interface. Target words and videos were presented in random order. For each target word, participants were allowed to watch as many videos as they wish and they were allowed to replay videos. Each participant in baseline condition was shown the target word once and each participant that used Vocabgram was exposed to each target word twice through two stages. Each session was conducted in a group of 4 to 6 participants at a time who were assigned to the same condition. A session ended with a post-test and a semi-structured interview which was intended to better understand their interface usage patterns and elicit subjective feedback about the interface.

4.5 Vocabulary Test

The post-test included two types of vocabulary tests: Vocabulary Knowledge Scale (VKS) and word meaning matching test. Participants were asked to complete the post-test immediately after the learning session.

4.5.1 Vocabulary Knowledge Scale

The first test was based on VKS which measures both self-perceived and demonstrated knowledge of specific words using a 5-point scale. We considered VKS as a valid measure as it is designed to measure both receptive and productive vocabulary development and track the acquisition of new words in the short-term [57, 58, 59]. Each score indicates the following level of vocabulary knowledge: 1 (not familiar at all), 2 (familiar but the meaning is not known), 3 (correct synonym or translation is given), 4 (word is used with semantic appropriateness in a sentence), and 5 (word is used with grammatical and semantic appropriateness in a sentence).

4.5.2 Word Meaning Matching Test

The design of the second test follows a conventional receptive knowledge assessment [], which asks learners to find the correct definition of a specific word from given lists of definitions. Participants were asked to choose the correct definition for each target word from a given list of 10 definitions, which included 5 distractors.

Chapter 5. Results

In this section, we report and discuss the results of our user study on 1) vocabulary gain, 2) qualitative results, and 3) user interaction.

5.1 Learning Gain

Table 5.1: The descriptive statistics for VKS scores (range: 5 to 25) and results of word meaning matching test (range: 1 to 5), and self-reported knowledge for each condition (range: 5 to 25).

		VKS scores		Word Meaning Matching		Self-reported Knowledge	
Condition	N	M [5, 25]	SD	M [0,5]	SD	M [5,25]	SD
Baseline	9	11.89	5.37	3.89	1.66	12.00	5.15
Vocabgram	9	14.11	4.08	4.67	0.67	16.11	4.22

Participants' learning outcome was measured by two vocabulary tests described earlier. VKS was scored by two raters, who resided in an English speaking country for more than 10 years and had prior experience with teaching ESL students, and resolved conflict with discussion. Pearson's r for interrater reliability was .89 for the VKS scores, which indicates a high reliability. Each vocabulary item was scored on a scale of 1 to 5. Thus, the possible range of scores was between 5 to 25. The word meaning matching test was scored with the ground truth definition. A correct answer received 1 point and an incorrect answer received 0 points. The total possible range of word meaning matching test was 0 to 5.

The statistics for two post-tests are displayed in Table 5.1. The average score of participants that used Vocabgram was higher (M=14.11, SD=4.08) on average than participants that used the baseline interface (M=11.89, SD=5.37). For comparison of VKS scores of two groups, we used Mann-Whitney tests. The result showed that the difference was marginally significant (U=786, p=0.07). The group of participants that used Vocabgram also had a higher average score on the word matching test (M=4.67, SD=0.67) than the other group (M=3.89, SD=1.66). To evaluate the differences in users' performance on word matching tests in two groups, we used a t-test, which revealed that the difference is statistically significant (t(88)=-2.13, p<.05). We chose to use a Mann-Whitney test to analyze the result of VKS test because the score of VKS test is not equidistance, unlike that of word meaning matching score.

Table 5.2: Number of vocabulary items for each VKS score per subgroup

		# of vocabulary items per VKS score subgroup				
Condition	# of vocabulary items	1	2	3	4	5
Baseline	45	14	13	10	3	5
Vocabgram	45	2	21	13	1	8

Table 5.3: Number of vocabulary items for each self-reported VKS score per subgroup

Condition	# of vocabulary items	# of vocabulary items per VKS score subgroup				
		1	2	3	4	5
Baseline	45	14	6	20	3	2
Vocabgram	45	2	8	19	10	6

Because statistics of VKS scores do not present the performance on individual target words, we present a detailed overview of the results in subgroups in Table 5.3. The participant group that used Vocabgram outnumbered the other group for the subgroup of a score of 5, which indicates a higher learning gain in productive knowledge.

The results suggest that Vocabgram helps learners to acquire both receptive and productive knowledge more effectively. Several factors may have led to a statistically insignificant impact on product knowledge. Firstly, gap-filling activities, which were designed to enhance productive ability, might have been too easy that it did not require elaborate processing of the context. For example, whereas the difficulty of gap-filling activity may vary depending on whether a word is given or the number of distractors, the current work adapted a simple version that involved a single distractor and provided a word. Secondly, gap-filling questions may not be sufficient for learners to acquire knowledge to produce new words. Thus, future work may address this by expanding the variety of questions. Lastly, the difference between individual participant's proficiency level may have an impact when they were asked to produce a sentence using the word in English.

Table 5.1 presents the statistics on participants' self-perceived knowledge on target words. The participant group that used Vocabgram reported a higher average score (M=16.11, SD=4.22) than the other group that used the baseline interface (M=12.00, 5.15). We ran the Mann-Whitney test to evaluate the difference which revealed that the difference is statically significant (U=626, $p<.05$). While the self-perceived score does not indicate higher learning gain, the result may imply that participants felt more confident after using Vocabgram. Many earlier studies have reported the importance of self-confidence in foreign language learning (e.g., [60]).

5.2 Qualitative Results

During the post-study interview, users shared both positive and negative comments on video-based vocabulary learning. Positive comments mostly reported the benefits of having a rich input of both visual and audio aids. For example, one user stated, "videos help me understand in what context the word is used." On the other hand, some users showed concerns about the inefficiency in the approach. One user reported, "I had to spend about half an hour to learn 5 words whereas I can learn 10 words in a much short time using word cards." However, there are pros and cons of using other approaches or adapting different medium as the resource. While video-based learning may take a longer time to acquire knowledge on a specific word, the learning session may have more values in the longer-term as learners acquire knowledge through associating with the rich contexts provided in a video that ranges from settings to relationship to emotion. Future work may address the inefficiency in many directions. For example, we may give more freedom to learners to choose how many videos they wish to watch per word.

There were also other practical comments on design choices. Many users expressed that the system provides them a translated definition of each target word in their native language. One user specifically expressed favor to the English definition of words as it describes the meaning without any loss in the meaning from translation. Thus, we may add a feature to allow users to freely translate contents in the system.

5.3 Usage Patterns

Each participant learned 5 target words throughout the session. Baseline condition took approximately 24 minutes on average (M=24.67, SD=4.24) while participants that used Vocabgram spent about 36 minutes (M=36, SD=8.33). Participants that used the baseline interface watched an average of 3.43 videos (SD=0.72). In Vocabgram, participants watched on average 6.45 videos (inference stage: M=3.33, SD=0.72, practicing stage: M=3.12, SD=0.38). The number of videos watched in the practicing stage also indicates the number of practice tests each participant completed. The low average value indicates that most of the participants performed well at this stage.

Chapter 6. Limitation & Future Work

In this section, we discuss limitations and potential directions for future work. We discuss on 1) long-term learning gain 2) learning multi-word units 3) selection of target words and videos.

6.1 Long-term Vocabulary Gain

This work only evaluated on short-term vocabulary learning using a single immediate post-test. In future work, we plan to include multiple measures such as a delayed post-test to ascertain vocabulary learning over time. Furthermore, Vocabgram is currently designed to support a single vocabulary learning session through two stages, but many studies have shown that vocabulary knowledge needs to be reactivated through spaced and continuous exposures to be effectively acquried [17]. Thus, one potential direction is to expand our approach to support a longitudinal acquisition of vocabulary knowledge by adding features that enable spaced repetition.

6.2 Multi-word Units

This work mainly focused on teaching vocabulary learning, but being able to understand and produce multi-word units such as phrasal verbs and idioms is often an important factor of language proficiency [61]. Future work may extend our approach to address multi-word units. While some components of Vocabgram (e.g., generating gap-filling activities) seem to be directly applicable to multi-word units, extending our approach to teach a more complex structure may pose new challenges. For example, whereas definitions of words are readily available in various resources such as an online dictionary, finding definitions or explanations of multi-word units may not be as accessible.

6.3 Target Words and Videos

In the evaluation, we used target words and videos that were pre-determined in advance by researchers based on a set of criteria. In real use, learners would have more freedom to choose videos and words they wish to learn. We expect if videos and words are better matched with learners' interest and level of knowledge, the learning could be more engaging and effective.

6.4 Video as a Productive Learning Material

This work is an exploratory research that investigated the potential of videos as a productive learning material. While many of existing language learning systems use videos as their resources, videos still remain as a underutilized resource. Most significantly, videos are mainly used for receptive learning in existing systems where learners are expected to develop linguistic skills through watching videos. However, videos offer many opportunities for productive learning such as the association of the language use with rich context in diverse authentic situations. Future work may extend and utilize videos further to support productive language learning using videos.

Chapter 7. Conclusion

This work investigates the effects of utilizing videos for acquiring productive vocabulary knowledge. We present Vocabgram, which involves employs a two-staged vocabulary learning process using authentic videos. The two stages of Vocabgram exposes learners to multiple occurrences of words in diverse contexts and provides a series of structured vocabulary learning activities to ensure learning. To assess the effectiveness of Vocabgram, we compared it to a baseline interface through conducting a between-subjects study with 18 participants. The result shows that learners that used Vocabgram gained higher knowledge of vocabulary than the other group. In addition, learners who learned with Vocabgram reported a higher level of confidence in their vocabulary knowledge.

Bibliography

[1] W. Grabe, "Current developments in second language reading research," *TESOL Quarterly*, vol. 25, no. 3, pp. 375–406, 1991.

[2] I. S. P. Nation, *Learning Vocabulary in Another Language*. Cambridge Applied Linguistics, Cambridge University Press, 2001.

[3] J. Milton, *Measuring Second Language Vocabulary Acquisition*. Second Language Acquisition, Channel View Publications, 2009.

[4] S. Webb, "Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge," *Studies in Second Language Acquisition*, vol. 27, no. 1, p. 33–52, 2005.

[5] B. Laufer and P. Nation, "A vocabulary-size test of controlled productive ability," *Language Testing*, vol. 16, no. 1, pp. 33–51, 1999.

[6] G. Griffin and T. A. Harley, "List learning of second language vocabulary," *Applied Psycholinguistics*, vol. 17, no. 4, p. 443–460, 1996.

[7] R. Waring, "The negative effects of learning words in semantic sets: A replication," *System*, vol. 25, no. 2, pp. 261 – 274, 1997.

[8] N. Schmitt, "Review article: Instructed second language vocabulary learning," *Language Teaching Research*, vol. 12, no. 3, pp. 329–363, 2008.

[9] N. Schmitt, *Vocabulary in Language Teaching*. Cambridge language education, Cambridge University Press, 2000.

[10] J. Hulstijn, *Incidental Learning in Second Language Acquisition*. 11 2012.

[11] S. Webb, "The Effects of Repetition on Vocabulary Knowledge," *Applied Linguistics*, vol. 28, pp. 46–65, 03 2007.

[12] S. H. Lee and J. Muncie, "From receptive to productive: Improving esl learners' use of vocabulary in a postreading composition task," *TESOL Quarterly*, vol. 40, no. 2, pp. 295–320, 2006.

[13] S. H. Lee, "Esl learners' vocabulary use in writing and the effects of explicit vocabulary instruction," *System*, vol. 31, no. 4, pp. 537 – 561, 2003.

[14] Y. Zhu, Y. Wang, C. Yu, S. Shi, Y. Zhang, S. He, P. Zhao, X. Ma, and Y. Shi, "Vivo: Video-augmented dictionary for vocabulary learning," pp. 5568–5579, 05 2017.

[15] B. Laufer and P. Nation, "Vocabulary Size and Use: Lexical Richness in L2 Written Production," *Applied Linguistics*, vol. 16, pp. 307–322, 09 1995.

[16] B. Laufer, "Focus on form in second language vocabulary learning," *EUROSLA Yearbook*, vol. 5, no. 1, pp. 223–250, 2005.

[17] J. H. Hulstijn, *Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity*, p. 258–286. Cambridge Applied Linguistics, Cambridge University Press, 2001.

[18] S. Webb, "Receptive and productive vocabulary sizes of l2 learners," *Studies in Second Language Acquisition*, vol. 30, no. 1, p. 79–95, 2008.

[19] S. Krashen, "We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis," *The Modern Language Journal*, vol. 73, no. 4, pp. 440–464, 1989.

[20] P. A. Nagy, William E.; Herman, *Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction.*, p. 19–35. Lawrence Erlbaum Associates, Inc, 1987.

[21] W. E. Nagy, "Learning word meanings from context: How broadly generalizable? technical report no. 347.," 1985.

[22] J. Harmer, *The Practice of English Language Teaching*. Always learning, Pearson Education, 2015.

[23] R. R. Day, J. Bamford, W. A. Renandya, G. M. Jacobs, and V. W.-S. Yu, "Extensive reading in the second language classroom," *RELC Journal*, vol. 29, no. 2, pp. 187–191, 1998.

[24] B. Laufer, "The development of l2 lexis in the expression of the advanced learner," *The Modern Language Journal*, vol. 75, no. 4, pp. 440–448, 1991.

[25] B. Laufer and T. S. Paribakht, "The relationship between passive and active vocabularies: Effects of languagelearning context," *Language Learning*, vol. 48, no. 3, pp. 365–391, 1998.

[26] J. Read, "Second language vocabulary assessment: Current practices and new directions," *International Journal of English Studies*, vol. 7, pp. 105–126, Jan. 2009.

[27] R. D. S. Brown, Ronan; Waring, "Incidental Vocabulary Acquisition from Reading, Reading-While-Listening, and Listening to Stories," *Reading in a Foreign Language*, vol. 20, pp. 136–163, 10 2008.

[28] D. M. Chun and J. L. Plass, "Effects of multimedia annotations on vocabulary acquisition," *The Modern Language Journal*, vol. 80, no. 2, pp. 183–198, 1996.

[29] J. H. Hulstijn, "Retention of inferred and given word meanings: experiments in incidental learning," 1992.

[30] J. H. Hultsijn, M. Hollander, and T. Greidnus, "Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words," *The Modern Language Journal*, vol. 80, no. 3, pp. 327–339, 1996.

[31] E. Peters, J. H. Hulstijn, L. Sercu, and M. Lutjeharms, "Learning l2 german vocabulary through reading: The effect of three enhancement techniques compared," *Language Learning*, vol. 59, no. 1, pp. 113–151, 2009.

[32] R. J. Sternberg, "Most vocabulary is learned from context.," 1987.

[33] S. Rott, "The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading," *Studies in Second Language Acquisition*, vol. 21, no. 4, p. 589–619, 1999.

[34] M. Horst, T. M. Cobb, and P. Meara, "Beyond a clockwork orange: Acquiring second language vocabulary through reading.," 1998.

[35] T. Saragi, I. Nation, and G. Meister, "Vocabulary learning and reading," *System*, vol. 6, no. 2, pp. 72 – 78, 1978.

[36] R. Waring and M. Takaki, "At what rate do learners learn and retain new vocabulary from reading a graded reader," 2001.

[37] I. Beck, M. Mckeown, and E. McCaslin, "Vocabulary development: All contexts are not created equal," *Elementary School Journal - ELEM SCH J*, vol. 83, 01 1983.

[38] S. Krashen, *Second language acquisition and second language learning.* Language teaching methodology series, Pergamon Press, 1981.

[39] "Duolingo." https://www.duolingo.com/. Accessed: 2019-12-30.

[40] "Anki." https://apps.ankiweb.net/. Accessed: 2019-12-30.

[41] "Memrise." https://www.memrise.com/. Accessed: 2019-12-30.

[42] D. Edge, E. Searle, K. Chiu, J. Zhao, and J. A. Landay, "Micromandarin: Mobile language learning in context," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, (New York, NY, USA), pp. 3169–3178, ACM, 2011.

[43] C. J. Cai, P. J. Guo, J. R. Glass, and R. C. Miller, "Wait-learning: Leveraging wait time for second language education," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), pp. 3701–3710, ACM, 2015.

[44] S. Zhou, I. J. Livingston, M. Schiefsky, S. M. Shieber, and K. Z. Gajos, "Ingenium: Engaging novice students with latin grammar," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (New York, NY, USA), pp. 944–956, ACM, 2016.

[45] M. Savva, A. X. Chang, C. D. Manning, and P. Hanrahan, "Transphoner: Automated mnemonic keyword generation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, (New York, NY, USA), pp. 3725–3734, ACM, 2014.

[46] G. Culbertson, E. Andersen, W. White, D. Zhang, and M. Jung, "Crystallize: An immersive, collaborative game for second language learning," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, (New York, NY, USA), pp. 636–647, ACM, 2016.

[47] Z. O. Toups, A. Kerne, and W. A. Hamilton, "The team coordination game: Zero-fidelity simulation abstracted from fire emergency response practice," *ACM Trans. Comput.-Hum. Interact.*, vol. 18, pp. 23:1–23:37, Dec. 2011.

[48] "Fluentu." https://www.fluentu.com/. Accessed: 2019-12-30.

[49] "Voicetube." https://www.voicetube.com/. Accessed: 2019-12-30.

[50] "Viki." https://www.viki.com/. Accessed: 2019-12-30.

[51] G. Culbertson, S. Shen, M. F. Jung, and E. Andersen, "Facilitating development of pragmatic competence through a voice-driven video learning interface," in *CHI*, 2017.

[52] S. Bird, "Nltk: The natural language toolkit," in *Proceedings of the COLING/ACL on Interactive Presentation Sessions*, COLING-ACL '06, (Stroudsburg, PA, USA), pp. 69–72, Association for Computational Linguistics, 2006.

[53] M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone," in *Proceedings of the 5th Annual International Conference on Systems Documentation*, SIGDOC '86, (New York, NY, USA), pp. 24–26, ACM, 1986.

[54] N. Schmitt and M. McCarthy, "Vocabulary size, text coverage and word lists," *Applied Psycholinguistics*, p. 6–19, 1997.

[55] M. Davies, "Corpus of Contemporary American English (COCA)," 2015.

[56] B. Laufer, *How Much Lexis is Necessary for Reading Comprehension?*, pp. 126–132. London: Palgrave Macmillan UK, 1992.

[57] Y. Kim, "The role of task-induced involvement and learner proficiency in l2 vocabulary acquisition," *Language Learning*, vol. 61, no. s1, pp. 100–140, 2011.

[58] J. Read, *Assessing Vocabulary*. Cambridge Language Assessment, Cambridge University Press, 2000.

[59] T. Paribakht and M. Wesche, "Reading comprehension and second language development in a comprehension-based esl program," *TESL Canada Journal*, vol. 11, pp. 09–29, Oct. 1993.

[60] R. Clément, Z. Dörnyei, and K. A. Noels, "Motivation, self-confidence, and group cohesion in the foreign language classroom," *Language Learning*, vol. 44, no. 3, pp. 417–448, 1994.

[61] A. Wray, "Formulaic sequences in second language teaching: principle and practice," *Applied Linguistics*, vol. 21, pp. 463–489, 12 2000.

Acknowledgment

I acknowledge and thank my advisor and colleagues, without whom I could have not done this work. First, and most of all, I would like to thank Juho Kim, for his patience, expertise, guidance, and a great amount of feedback and ideas. My colleagues in KIXLAB for providing feedback at every stage and creating the good environment that enabled me to conduct this research. My family for their unconditional care, love, and support.

Curriculum Vitae

Name : Kyung Je Jo (조경제)

Date of Birth : May 21, 1995

Address : 88, Yanghyeon-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea

E-mail : kyungjejo@kaist.ac.kr

Educations

2009. 9. – 2013. 6. Overseas Family School, Singapore

2013. 9. – 2018. 2. Korea Advanced Institute of Science and Technology, Republic of Korea (B.S. in School of Computing)

2018. 3. – 2020. 2. Korea Advanced Institute of Science and Technology, Republic of Korea (M.S. in School of Computing)

Career

2017. 1. – 2018. 2. Research intern at KIXLAB, KAIST

Publications

1. **Kyung Je Jo**, Hyeonggeun Yun, Juho Kim, "Supporting Instruction of Formulaic Sequences Using Videos at Scale," *Work-in-Progress of Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale.*
2. **Kyung Je Jo**, John Joon Young Chung, Juho Kim, "Exprgram: A Language Learning Interface for Mastering Pragmatic Competence," *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.*
3. **Kyung Je Jo**, John Joon Young Chung, Juho Kim, "Exprgram: A Video-based Language Learning Interface Powered by Learnersourced Video Annotations," *HCOMP 2017 Workshop on Human Computation for Image and Video Analysis (GroupSight 2017).* **Best Paper Runner-Up**