So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

# A New Technology Design for Personalized Incidental Vocabulary Learning using Lifelog Image Analysis

**Mohammad Nehal HASNINE[a*], Kousuke MOURI[b], Gokhan AKCHAPINAR[c], Mahmoud Mohamed Hussien AHMED[d] & Hiroshi UEDA[a]**
[a]*Research Center for Computing and Multimedia Studies, Hosei University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
[c]*Department of Computer Education & Instructional Technology, Hacettepe University, Turkey*
[d]*Department of Educational Technology, South Valley University, Egypt*
*nehal.hasnine.79@hosei.ac.jp

**Abstract:** Incidental vocabulary learning is the process of learning foreign vocabulary without the intention of doing so. In language learning pedagogy, incidental learning is considered to be an effective way to enhance foreign vocabulary from context. Conventional vocabulary learning systems in ubiquitous learning scenarios are developed primarily for supporting intentional vocabulary learning. When learning foreign vocabulary using conventional tools, it is not feasible to learn incidental vocabulary. Moreover, there does not exist a general framework to formulate lifelong images along with typical ubiquitous learning logs such as location, time, and demographics. Therefore, this research precisely looked into the scope of formulating lifelog images as sensor data for enhancing incidental vocabulary knowledge of English as Foreign Language (EFL) learners. The main pursuit of this article is to introduce a model for enriching ubiquitous learning literature, in which incidental vocabulary can be learned. In this study, a technology-enhance environment for incidental vocabulary learning in the EFL context is presented. The research objectives are to- i) design a new technology that is capable of analyzing lifelong images in order to generate a bag of incidental vocabularies, ii) determining the top-5 vocabularies that could be recommended to the learner, and iii) to automatically create learning material for each of the recommended words. We employed visual content analytics on lifelong learning images using object detection method. This is known to be an applied artificial intelligence method that is applicable across a variety of fields, such as natural language processing, computer vision, and others. We aimed to apply this applied AI method to design a new technology-enhanced learning environment.

**Keywords:** AI in language learning, applied AI, incidental vocabulary learning, informal learning, lifelog image analysis, technology-enhance language learning

## 1. Introduction

For language learners, vocabulary learning is an indispensable activity to achieve competence in the target language. Vocabulary enriches learners' integrated language skills such as listening, speaking, reading; which later on facilitates fluent conversation and effective writing (Ahmad, 2012). In language learning pedagogy, several theories and approaches have tried to account for the specific way that learners' learning takes place. It seems, however, that vocabulary learning depends on the type of cognitive process in which the learner is engaged (Ramos & Dario, 2015). Two commonly discussed pedagogical methods that are directly associated with learners' cognitive processes are intentional learning and incidental learning. Intentional learning is defined as being designed, planned for, or intended by teachers or students (Yali, 2010). In contrast, incidental learning is the process of learning something without the intention of doing so (Ahmad, 2012). It is also learning one thing while intending to learn another (Richards & Schmidt, 2013). This type of learning strategies can be used to enhance vocabulary swiftly. Research indicates that, in comparison with the intentional vocabulary learning method, the incidental learning method can be effective. According to (Ahmad, 2012), intentional vocabulary learning often learned based on synonyms, antonyms, word substitution, multiple-choice, scrambled words, and crossword puzzles, regardless of context, is not so effective, because learners are more prone to rote learning. Hence, research emphasis needs to be given on incidental vocabulary learning. Lifelog images are those often captured to record learners' learning experiences. Other research fields addressed lifelog images as visual lifelogging as they are logged by wearable devices such as GoPro, MeCam, Looxcie, or Google Glass. Without arguing, each of the lifelog images has a story to tell for the capturer, hence, they are a rich source of data. Research already evident that, lifelog images are the source of valuable information because the pictures taken offer considerable potential for knowledge mining concerning how people live their lives; hence, they open up new opportunities for many potential applications in fields including healthcare, security, leisure, and the quantified self (Bolaños et al., 2017). With the convenience of smartphones and wearable technologies, high temporal resolution, and are more suitable to record specific moments from language learners, such as cultural experience, authentic learning experience, problem-based learning, task-based learning experiences, etc. Therefore, language learners' lifelog images together with other ubiquitous logs need to be analyzed to improve language learning particularly vocabulary learning.

To our best knowledge, as yet, there does not exist any framework that formulates lifelong images along with typical ubiquitous learning logs such as location, time, and user information. Therefore, we aim to design a new technology in technology-enhance language learning with the objective to support incidental vocabulary learning from lifelogging. With this study, we aim to overcome one criticism about conventional ubiquitous learning systems namely the scope of learning new vocabulary is limited.

## 2. Literature Review

### *2.1 Previous works by this research team*

Researching about the incidental vocabulary learning using ubiquitous learning log and other sensor data is our ongoing project. Some recent outcomes that related to this work are- in 2019, we introduced a location-based word recommendation system for incidental vocabulary learning in English as Foreign Language (EFL) context (Mohammad Nehal Hasnine et al., 2019). This system uses EFL learners' vocabulary learning histories and location data as the input; then it leverages association rule mining for discovering new knowledge and non-trivial patterns hidden in learning history; finally generates words that are associated to that particular location. In 2019, our other study used images' visual contents as contextual clues to generate special learning contexts where new vocabularies believed to be learned from sentences (Hasnine, Flanagan, et al., 2019). To develop this technology, we leveraged the power of automatic image captioning. Later on, in ICCE 2019, we introduced a model that detects the textual information from lifelog images (Mouri et al., 2019). This model relies on several APIs to detect objects and textual clues from images. These clues were used to support foreign language learning by bridging cyber space learning with physical learning.

### *2.2 Related works by global researchers*

In order to support intentional vocabulary learning using ubiquitous and computer-aided technologies, SCROLL, UEVL, MALL, and U-Arabic systems are commonly cited. These tools use location data, context information, word learning histories etc. to support intentional learning. However, these tools do not support incidental learning at all; the idea of learning experience sharing is noticed, though. For incidental vocabulary learning, study by (Song & Fox, 2008) used personal digital assistant as the learning technology to teach English words to university students. A one-year-long case studies on undergraduate students revealed that personal device assistants can be used in more flexible, novel and extended ways for English as a Foreign Language (EFL) vocabulary teaching and learning in higher education, taking student needs and contexts into consideration. Besides, video captioning has been tested for incidental vocabulary learning by (Montero Perez et al., 2014). This study suggests that captioning did not affect much on comprehension nor meaning recall. Electronic dictionaries as the tool are used in the early 2000. Study conducted by (Laufer, 2000) revealed that students who learned by electronic text performed significantly better than the paper text group in learning low frequency words.

Our meta-analysis on previous studies indicates that, by far, time, place, contextual memos are used as contextual clues to support vocabulary learning. However, lifelog images as contextual clues is overlooked. Recently, wearable cameras on the market like GoPro, MeCam, Looxcie, or Google Glass are relatively getting popularity for high temporal resolution and are more suitable to record specific moments, such as cultural experiences. Therefore, lifelog images can be used as contextual clues to support foreign language learning. Consequently, we propose this study where foreign language learners lifelog image are analyzed to support incidental vocabulary learning. This, we believe, will bring new dimension in the design of technology-enhance language learning using applied AI.

## 3. Technology Design

### *3.1 Overview of the Model*

We begin by clearly introducing the objectives of this study. This technology is designed for foreign language learners to support their incidental vocabulary learning. Therefore, this technology aims to broaden learners' opportunities to engage with new vocabulary, create learning materials, and reflect on their memory. Unlike existing ubiquitous language learning systems where place, time, and handwritten memos are used the contextual clues; the newness of this proposed technology is the usage of lifelogging images as the primary contextual clue. We leveraged lifelog images because the visual contents of lifelogging contain powerful social interaction, the analysis of social interactions in lifelogging data is of fundamental importance to understanding human behavior; and the presence of people and social interactions are consistently associated with our memory. Hence, we hypothesized that as lifelog images contain powerful information, they need to be analyzed for educational uses (Bolaños et al., 2017). We introduce the design of the model in Figure 1. Later in the 3.3, we discuss each of the key components of the model.

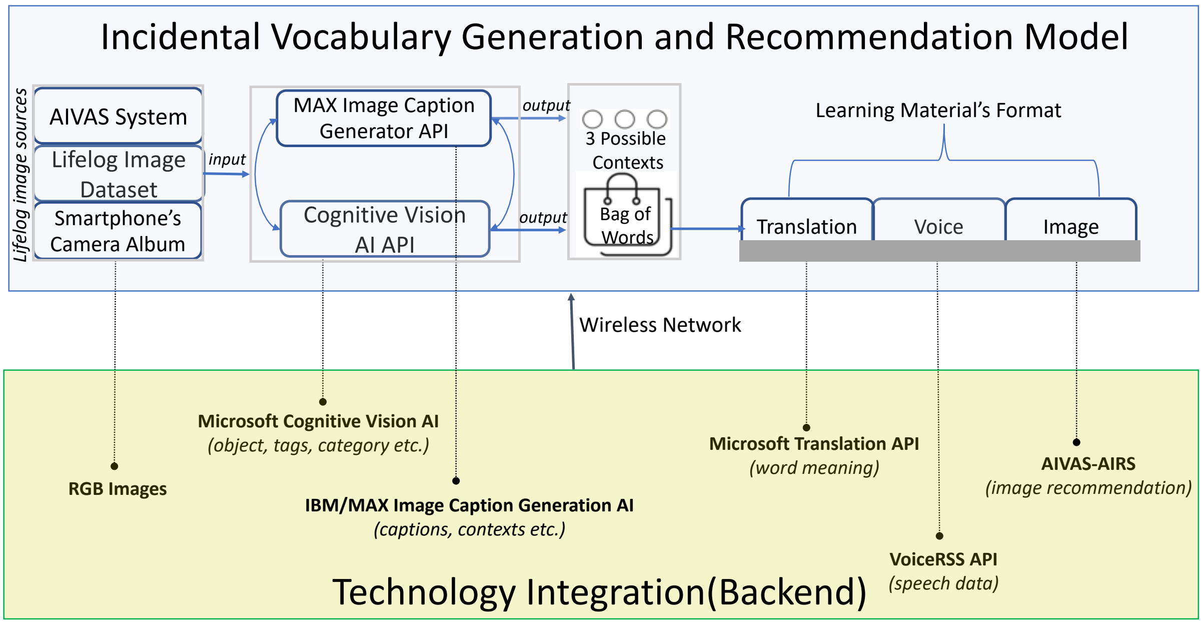


*Figure 1.* The Design of the Model

### *3.2 Methods and Materials*

The proposed model begins by reading data from three modular data sources. At present, there are three main sources for reading lifelog images. The first source is, the database of the AIVAS (Appropriate Image-based Vocabulary Learning System) system; the second is a dataset; the third is manual uploading. After the data is read, the image analysis task is begun. This task is done by two AI-based models namely an image captioning model and a cognitive vision model. The image captioning model generates three possible contexts in the form of sentences; whereas the cognitive vision model produces a bag of words. The bag of words contains all of the possible vocabularies that the learner could learn as incidental learning. After that, the model determines top-5 incidental vocabularies based on the word's confidence level. That is, five words with the highest confidence are recommended for learning. Finally, for each of the recommended words, learning material is created automatically. For the automatic creation of the learning materials, we used our previously developed system (M. N. Hasnine et al., 2016; Mohammad Nehal Hasnine, 2018; Mohammad Nehal Hasnine et al., 2017).

### *3.3 Technical Specifications of the Components*

The proposed design heavily relies on external APIs and an online language learning tool. Here, we provide a brief introduction to the technologies that are used to design this new technology.

- AIVAS system: It is a web-based language learning tool that assists learners in creating learning materials (M. N. Hasnine et al., 2016; Mohammad Nehal Hasnine et al., 2015).
- Lifelog image dataset: This is a dataset that contains about 10000 lifelog images in total. These images were captured by foreign language learners using lifelog camera and smartphone camera technology. A mobile technology called SCROLL (System for Capturing and Reminding of Learning Logs) is used to capture the images. The dataset (Ogata et al., 2018) is available for research purposes.
- Camera album and local directory: Our model can read image data from a local computer and camera album of a smartphone. This is for the manual testing.
- MAX Image Caption Generation API: This is an API[1] developed by IBM. This API uses artificial intelligence for understanding scenes and uses NLP methods for describing scenes in natural language. The model we used for our study is based on Show and Tell[2] image-to-text generation model.
- Microsoft Cognitive Vision API: This API is developed by Microsoft used for extracting rich information from RGB images. This API uses Microsoft's computer vision services for analyzing image scenes. The input of this API is a natural image. The output is a bag of vocabulary. Each of the vocabularies listed in the bag is represented with a tag, object name, description, and category of the object.
- Microsoft Translation API: This is a cloud-based service for machine translation offered by Microsoft[3]. This API takes a word or a phrase as input and returns the corresponding translation of the target language that the user wishes to get.
- VoiceRSS API: This is a text to speech API. This API[4] read textual content for converting to speech. This is a free API that can be used for research purposes.
- AIVAS-AIRS: AIRS stands for Appropriate Image Recommendation System. In AIVAS, we use AIRS system for determining appropriate images for representing a word. This system uses the image ranking method (Mohammad Nehal Hasnine, 2018) for recommending appropriate images. This system is capable of ranking images from a set of corresponding images for a word that is downloaded from image search engine.

## 4. System Development

Based on the proposed model (described in Section 3), we developed a prototype of the incidental vocabulary learning system for supporting foreign language learners. At present, the system produces English vocabularies from lifelogging images. The system captures conventional ubiquitous logs such as time, place, intentional vocabulary together with a bag of incidental vocabulary. Figure 2 shows the functions to analyze lifelog image scenes for producing top-5 incidental vocabulary. The top-5 incidental vocabulary is determined by the model's confidence for each of the vocabulary in the bag.

---

[1] https://github.com/IBM/MAX-Image-Caption-Generator
[2] https://github.com/tensorflow/models/tree/master/research/im2txt
[3] https://www.microsoft.com/en-us/translator/business/translator-api/
[4] http://www.voicerss.org/api/documentation.aspx

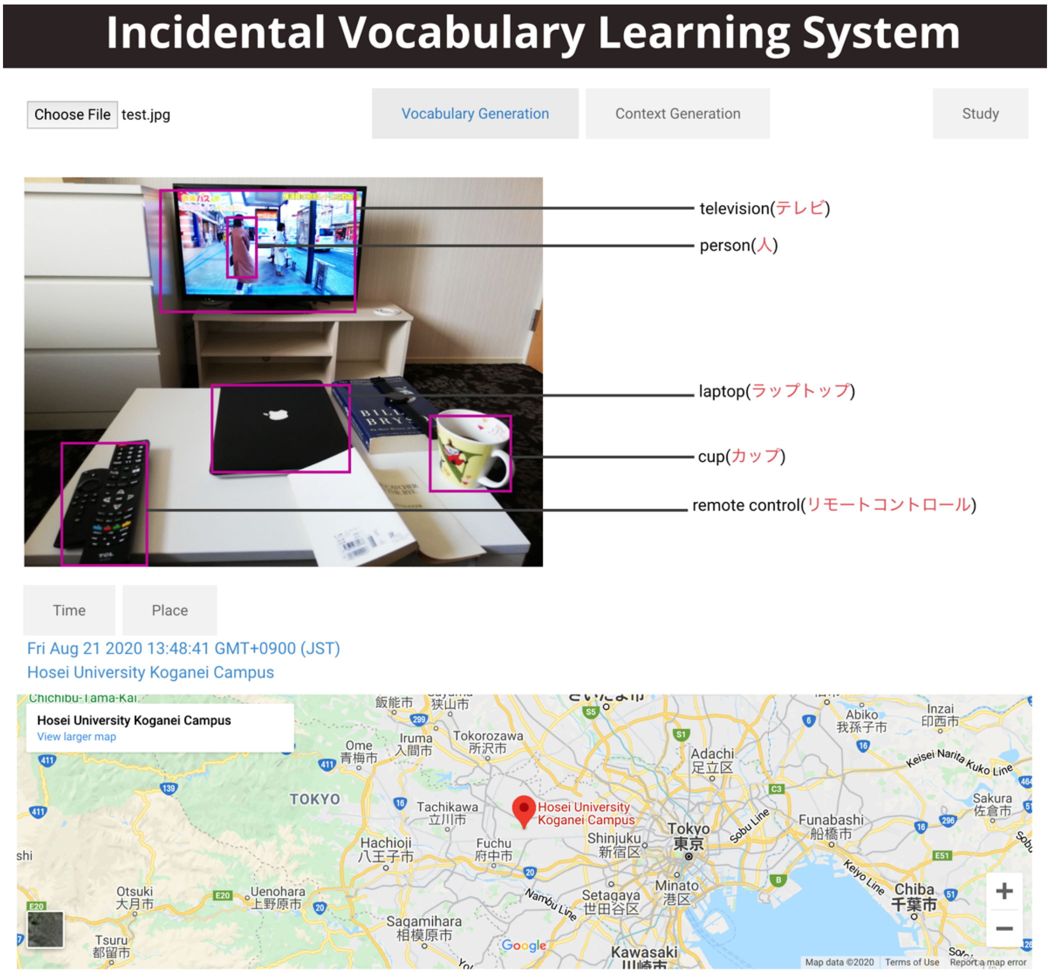


*Figure 2.* Incidental Vocabulary Generation from a Lifelog Image

Figure 3 is the interface where incidental vocabularies are recommended and those are studied by the learners. In this interface, a learning material can be created automatically for each vocabulary. Upon clicking on a word, a learning material will automatically be created and displayed on the interface. For creating a learning material, we used the learning material creation module of the AIVAS system (Mohammad Nehal Hasnine, 2018; Mohammad Nehal Hasnine et al., 2017).

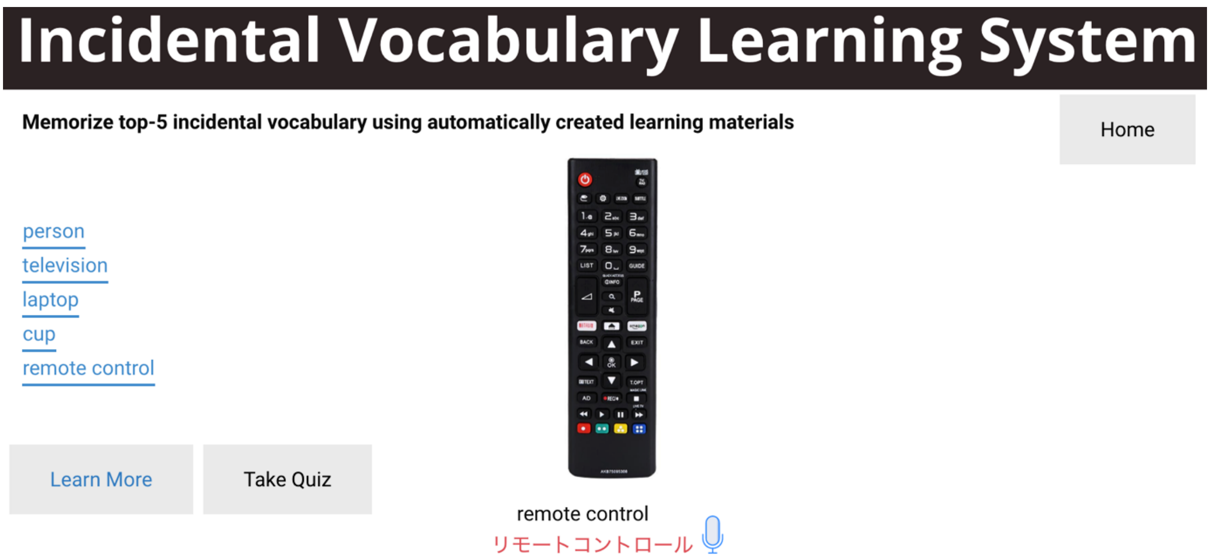


*Figure 3.* An Automatically-generated Material for Learning for a New Word

## 5. Conclusion

Incidental vocabulary learning is an important aspect of language learning pedagogy. Yet, it is not feasible to learn incidental vocabulary using conventional learning tools. On top of that, research on conventional ubiquitous learning overlooked the power of visual lifelogging as the potential learning logs. It seems traditional definitions of ubiquitous learning is much closely associated with learning by logging and contextualization. Therefore, in this paper, we introduced a technology design for incidental vocabulary learning in the EFL context. As smartphone camera technologies along with other wearable technologies such as GoPro, MeCam, Looxcie, or Google Glass are becoming mainstream research topics, our proposed model would be a good fit for it.

In this paper, we introduced a new design for technology-enhanced language learning that aims to support incidental vocabulary learning in the EFL context. We first implemented the model in our local server. Then, we developed a prototype of the system. Interface displayed in Figure 2 is to analyze the model's output with regards to top-5 objects to learn from a lifelogging image. The interface displayed in Figure 3 is for generating learning material for each word. The second interface can be used for memorizing new words and taking quizzes for reflection. With this technology, our goal is to give a learner scope to encounter with new vocabularies and learn them using multimedia annotations.

## References


Ahmad, J. (2012). *Intentional vs. Incidental vocabulary learning*. 9.

Bolaños, M., Dimiccoli, M., & Radeva, P. (2017). Toward Storytelling From Visual Lifelogging: An Overview. *IEEE Transactions on Human-Machine Systems*, *47*(1), 77–90.

Hasnine, M. N., Hirai, Y., Ishikawa, M., Miyakoda, H., Kaneko, K., & Pemberton, L. (2016). An image recommender system that suggests appropriate images in creation of self-learning items for abstract nouns. *International Journal of Management and Applied Science*, *2*(5), 38–44.

Hasnine, Mohammad Nehal. (2018). *Recommendation of appropriate images for vocabulary learning*. Ph.D. thesis, TUAT repository, http://hdl.handle.net/10636/00001480.

Hasnine, Mohammad Nehal, Akçapınar, G., Flanagan, B., Mouri, K., Ogata, H., Kaneko, K., & Uosaki, N. (2019). Design of a Location-based Word Recommendation System Based on Association Rule Mining Analysis. *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 250–253.

Hasnine, Mohammad Nehal, Hirai, Y., Ishikawa, M., Miyakoda, H., & Kaneko, K. (2015). Learning effects investigation of an on-demand vocabulary learning materials creation system based on appropriate images. *Proceedings of the 4th ICT International Student Project Conference*.

Hasnine, Mohammad Nehal, Ishikawa, M., Hirai, Y., Miyakoda, H., & Kaneko, K. (2017). An algorithm to evaluate appropriateness of still images for learning concrete nouns of a new foreign language. *IEICE Transactions on Information and Systems*, *100*(9), 2156–2164.

Laufer, B. (2000). Electronic dictionaries and incidental vocabulary acquisition: Does technology make a difference. *EURALEX*, 849–854.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning and Technology*, *18*(1), 118–141.

Ogata, H., Uosaki, N., Mouri, K., Hasnine, M. N., Abou-Khalil, V., & Flanagan, B. (2018). SCROLL Dataset in the context of ubiquitous language learning. *Workshop Proceedings of the 26th International Conference on Computer in Education*, 418–423.

Ramos, R., & Dario, F. (2015). Incidental Vocabulary Learning in Second Language Acquisition: A Literature Review. *Profile Issues in Teachers` Professional Development*, *17*(1), 157–166.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge.

Song, Y., & Fox, R. (2008). Using PDA for undergraduate student incidental vocabulary testing. *ReCALL*, *20*(3), 290–314.

Yali, G. (2010). L2 Vocabulary Acquisition Through Reading–Incidental Learning and Intentional Learning. *Chinese Journal of Applied Linguistics (Foreign Language Teaching & Research Press)*, *33*(1).