Learning what students want to learn

Author(s): Victoria Abou-Khalil
Graduate School of Informatics, Kyoto University
v.aboukhalil@gmail.com

Author(s): Brendan Flanagan
Academic Center for Computing and Media studies, Kyoto University
flanagan.brendanjohn.4n@kyoto-u.ac.jp

Author(s): Hiroaki Ogata
Academic Center for Computing and Media studies, Kyoto University
hiroaki.ogata@gmail.com

ABSTRACT: Polysemous words are words that have different meanings in different contexts. This type of word is confusing for language students. In order to be able to provide the language learners with the right translation in the right context, it is important to understand the meaning intended by the student. This paper proposes a method to predict the meaning intended by the student based on students' past learned words, current location and time. The method proposed uses records from the SCROLL system (System for Capturing and Reminding Of Learning Log) to analyze the activity of students. We assume that the uploaded vocabulary of a student can be used to predict the meaning intended by the student when looking up a polysemous word. The identification of the intended meaning in the student's current context could be then used to provide the student with the appropriate translation possibly improving the learning.

Keywords: Learning Analytics, Ubiquitous learning, Polysemous words, Computer Supported Language Learning

1 INTRODUCTION

Words change their meaning from one context to the other. To *book* a hotel for example is not the same thing as reading a *book*. This type of words, called polysemous words, represent about 40% of the entries of an English dictionary (Nagy, 1995). The traditional way of learning polysemous words is to present to the student the different definitions of the word in different contexts. From the pedagogical perspective, it is suggested that the meanings should not be presented as one bloc to the student but should be put in a context and restricted to a particular area (Hayward, 1984). However, in a self learning environment, teachers have less or no control over the learned content. With the rise of smart phones, the use of mobile devices in language learning is a growing trend (Godwin, 2016). Students have more control over their learning pace (Benson, 2005) and the vocabulary they want to learn, but don't have opportunities to communicate which meaning of the word they are looking for.

This paper proposes to take advantage of student past learned words to understand the particular meaning queried by the learner to be able to potentially provide them with the correct translation in their intended context.

2 METHOD

2.1 SCROLL system

During this study we use records from the SCROLL System (System for Capturing and Reminding Of Learning Log). Scroll is a digital record of what language students have learned in daily life. It allows the learners to log the new words or sentences they learned along with photos, audios, videos and location (Ogata, 2011). SCROLL captures what learners are learning as well as its contextual data. The users are then reminded of what they learned in the right place and the right time. Moreover, students receive personalized quizzes to fortify the learning. Figure 1 is a screenshot from the SCROLL system that shows a log inserted by a student for the word *Karaoke*. The student appended a picture and a location when creating the log. A Japanese translation of the word *Karaoke* is automatically provided to the student, and the time is automatically registered.

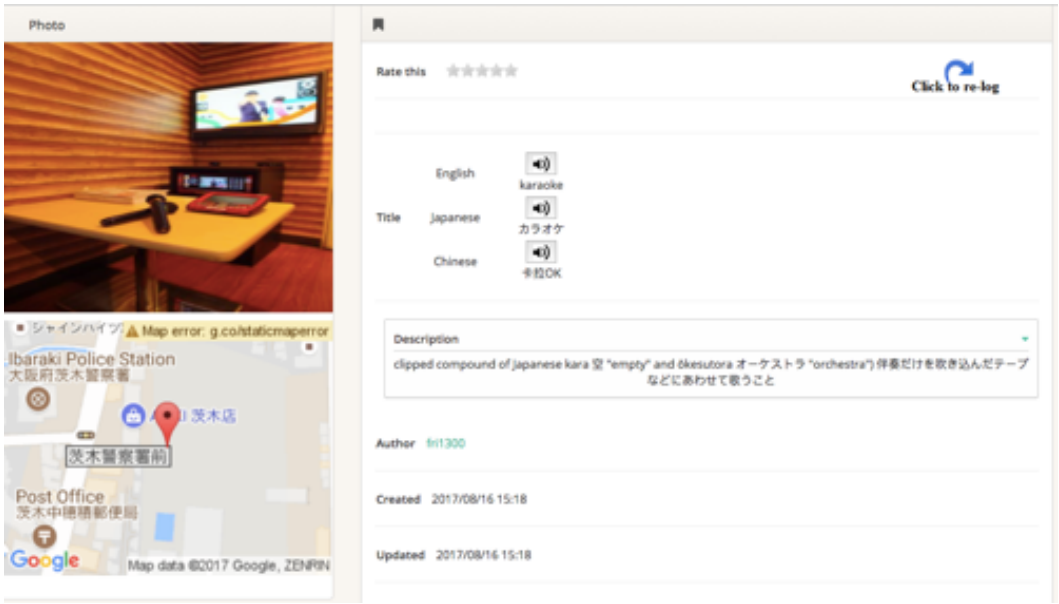

Figure 1: Screenshot from the SCROLL system showing a log inserted by a student.

In order to determine the intended meaning of a polysemous word, we will be the following meta-data included in the logs:

Vocabulary: words that students have learned in the past
User: author identification
Pictures: Uploaded picture of the learned word

Currently SCROLL has 1648 users and contains 24355 logs. The system is used mainly by students learning Japanese.

2.2 Contextual false friends learning

When using the SCROLL system, Japanese language learners insert logs containing a word in English and learn its Japanese translation. However, if the word is a polysemous word, students get a translation that does not usually reflect the context, the different meanings and the nuances of the word. Moreover, in a self learning environment students don't get the opportunity to explain which meaning of the word they are intending to learn. In order to provide learners with the right translation in the right context, we have to understand their intended meaning. We propose to analyze the past activity of students to extract the meaning they are looking for. We assume that the past activity of a student gives a context that can be used to extract the meaning intended by the student when looking up a polysemous word. The following meta-data is used to predict the meaning the student is looking for:

Past knowledge (words): to measure the similarity between the current word the student is looking up and past words the student looked up in the system.

In order to identify the intended meaning of the word, we measure the semantic similarity between the vocabulary of the student and different meaning of the word. The words similarity is measured using Word2Vec, a word embedding technique that reconstruct linguist context of words. The corpus is trained by cortical.io and includes the full texts of Wikipedia. Figure 2 shows an example. The student looked up the word *driver*. However, the word driver has different meanings and different translations in Japanese for each of those meaning:

Person who drives a vehicle: ドライバ (doraiba)

Train driver: 運転士 (untenshi)

Mass driver : マスドライバー (masu doraiba)

Computing : 仮想デバイスドライバ (kasoo debaisu doraiba)

Golf club: ドライバー (doraiba)

Screwdriver (British English): ドライバー (doraiba)

Figure 2: Screenshot from the SCROLL system showing the student past knowledge of a student looking for the meaning of the word driver

In order to understand which meaning the student is looking for, we look at the previous knowledge. In this case, the word looked up just before the word *driver* is *stepladder*. The semantic distance between the words *srewdriver* and *stepladder* is smaller than the semantic distance between the word *stepladder* and each of the words: *car driver*, *train driver*, *computer driver*, *mass driver* and *golf club*. We can conclude that the meaning the student is looking for is *screwdriver*, and that the translation that should be given is ドライバー (doraiba).

3 RESULTS

In order to evaluate the identification of intended words, we collected 60 polysemous words from the SCROLL system. The collected words have pictures associated to them, uploaded by the user. For each of those words, we measured the semantic similarity between every different meaning of the word and the vocabulary of the user. The meaning that has the biggest semantic similarity with the vocabulary of the user is supposed to be the meaning intended by the user. In order to evaluate the accuracy of the result we manually compare the obtained meaning with the uploaded picture. 85.714 of the obtained meanings corresponded to the picture uploaded by the user.

The semantic similarity was similar between the vocabulary of users and meanings that are semantically close (e.g.: Fan*: 'an apparatus with rotating blades that creates a current of air for cooling or ventilation* and *a handheld device, typically folding' and 'shaped like a segment of a circle when spread out, that is waved so as to cool the person holding it'*). In that case, the results were often erroneous.

4 DISCUSSION

We proposed a method to determine the intended by the student when they are looking up a polysemous word. The method is applied in the context of mobile learning. The main features of mobile learning are accessibility, immediacy, interactivity and situating of instructional activities (Ogata, 2004), benefit students during the learning process. However, mobile learning does not allow language students to share their intended meaning while looking up a polysemous word. In order to do so, the proposed method uses records from the SCROLL system (System for Capturing and Reminding Of Learning Log) to analyze the previous activity of students. We assumed that the students' past learned words can be used to predict the meaning intended by the student when looking up a polysemous word. The identification of the intended meaning is then used to provide the student with the appropriate translation, based on the intended meaning. The evaluation showed that the expected intended meaning corresponded to the real intended meaning of the student for 85.714 of the words. In future works more factors should be taken into consideration for the identification of the intended meaning such as:

Time: to measure the similarity between the current word the student is looking up and the set of words the student looked up at the same period of time.

Location: to measure the similarity between the word the student is looking up and the set of words the student looked up at the same location.

5 CONCLUSION

We proposed a method to determine the intended by the student when they are looking up a polysemous word. The method is applied in the context of mobile learning. Mobile learning does not allow language students to share their intended meaning while looking up a polysemous word. In order to do so, the proposed method uses records from the SCROLL system (System for Capturing and Reminding Of Learning Log) to analyze the previous activity of students. We assumed that the students' past learned words can be used to predict the meaning intended by the student when looking up a polysemous word. We identified the intended meaning by computing the sematic similarity between the vocabulary uploaded by the student and each of the different meanings of the word. The results show that the intended meaning corresponded to the picture uploaded by the student 85.714 % of the times.

Acknowledgments

This research was also partially supported by Grant-in-Aid for Scientific Research (S) from Japan Society for the Promotion of Science (24220002, 2012-2016) and from Monbukagakusho (16H06304, 2016-2021).

REFERENCES

Benson, P. (2005). Autonomy and information technology in the educational discourse of the information. Information technology and innovation in language education, 1:173.

Godwin-Jones, R. (2016). Looking back and ahead: 20 years of technologies for language learning.

Hayward, T. and Moulin, A. (1984). False friends invigorated. In Lexeter'83 Proceedings, pages 190–198.

Nagy, W. E. (1995). *On the role of context in first-and second-language vocabulary learning*. Champaign, Ill.: University of Illinois at Urbana-Champaign, Center for the Study of Reading.

Ogata, H., Li, M., Hou, B., Uosaki, N., El-Bishouty, M. M., and Yano, Y. (2011). Scroll: Supporting to share and reuse ubiquitous learning log in the context of language learning. Research & Practice in Technology Enhanced Learning, 6(2).

Ogata, H. and Yano, Y. (2004). Knowledge awareness for a computerassisted language learning using handhelds. International Journal of Continuing Engineering Education and Life Long Learning, 14(4-5):435–449.