Augmenting the Web for Second Language Vocabulary Learning

Andrew Trusty & Khai N. Truong
Department of Computer Science
University of Toronto
Toronto, ON, Canada, M5S 3E4
{atrusty, khai}@cs.toronto.edu

ABSTRACT

The busyness of everyday life means that those with casual interest in additional learning opportunities are often unable to schedule regular time and effort for studying. In this paper, we explore how to augment information technologies that people use on a daily basis to create micro-learning opportunities. In particular, we examine how a person's existing Web browsing experience—with first language Web pages—can be augmented to teach them second language vocabulary. We present a prototype, ALOE, which runs inside the Firefox Web browser and dynamically augments Web pages by replacing a selected set of English words with their foreign translations. The foreign translations are embedded in the rich context of a Web page's existing English text to promote incidental learning and guessing from context of the translated words. Through a two month user evaluation of ALOE, we found that most participants were able to learn an average of 50 new French vocabulary words.

Author Keywords

Computer assisted language learning, user interface, Web

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

General Terms

Design, Experimentation

INTRODUCTION

There are many reasons why an individual may want to learn a second language (L2). This is not surprising given how integral communication is to people's daily lives. It is the method by which individuals relate and learn from each other. Furthermore, as modern technologies continue to shrink the "distance" between people and cultures, being able to communicate with and understand each other has become increasingly more valuable.

Learning a new language requires time and effort on a recurring basis. The more practice a person has with a new language, the better she will learn. Unfortunately, the busyness of everyday life means that those with casual interest in studying a second language are often unable to schedule regular time and effort towards learning.

However, there is growing agreement that most L2 vocabulary acquisition occurs incidentally [5, 20, 26, 28, 30, 36] and evidence that repeated exposure with minimal mental processing by the learner can have memory effects [15]. From this perspective, we propose breaking traditional language study tasks into many micro learning moments which can be distributed throughout a person's day-to-day life in an unobtrusive way. As people spend increasing amounts of time on the Web, a person's existing Web browsing experience can be augmented to support the learning of L2 vocabulary in many contexts. The novelty of this approach lies in the fact that we aim to teach second language vocabulary while people are reading Web pages that are written in their native language (L1). In particular, ALOE only translates a couple of words in each sentence on a page and leaves the majority of the text in L1. Thus, the central idea behind ALOE is that users will have a rich L1 context from which they can infer the meaning of the foreign words, and incidental learning of vocabulary will occur through repeated exposure to these foreign words.

In this paper, we present the design and evaluation of the ALOE L2 vocabulary learning prototype which runs inside the Firefox Web browser and augments a user's Web browsing experience to provide second language vocabulary learning opportunities. Through interviews with six second language educators from our institute, we learned that the biggest challenges with teaching a foreign language include motivating and challenging students to learn beyond the classroom and grounding their learning in a context that facilitates learning. Based on these findings, we then designed the ALOE system. Through a two month deployment study in which 18 participants used ALOE on their personal computers, we show that the ALOE prototype and our approach can facilitate L2 vocabulary learning. Most participants enjoyed using ALOE; and on average, learned 50 new French vocabulary words during the study.

RELATED WORK

Research in Computer Assisted Language Learning (CALL) investigates how to improve the effectiveness and efficiency of language learning while also making it easier and more convenient for both teachers and students to access and use language learning materials [12, 16]. Recent efforts in CALL have focused on language learning on different platforms, including specifically through the Web.

CALL on Different Platforms

In the mobile domain, researchers have explored the use of SMS for sending L2 vocabulary definitions to learners [3, 15, 18, 33], providing adaptive L2 news article recommendations [4], and creating sensor augmented physical spaces that trigger the presentation of learning materials on the user's mobile device [1, 18, 21, 22]. Data gathered from these studies have shown positive results including user interest and enjoyment in using the system, new language learning, increased user motivation and improved performance over traditional study materials.

Fallahkhair, Pemberton & Griffiths [6] explored augmenting L2 television programs with L1 language learning materials that were accessible using a mobile phone. Through a qualitative in-lab evaluation, they showed that participants enjoyed the proposed concept and believed it could be useful for language learning, although there was no data to reflect whether or not participants actually improved their language skills. By using L2 television programs, their work limits potential users to advanced learners who would have the skills necessary to follow such programs.

CALL through the Web

In recent years, the Web has been a major source of L2 language learning material both in terms of existing L2 material and intentionally prepared L2 educational resources made available by teachers at all levels. Realizing the value of the Web as a language learning corpus, many CALL researchers have investigated how to best use this new abundance of material for language learners [16]. There are two approaches: Web mining and Web augmentation. Web mining attempts to extract language learning data from the Web. Web augmentation attempts to overlay language learning functionality on top of the Web.

Fletcher [7] provided a detailed overview of how the Web mining approach is currently used in language learning along with the future possibilities it provides. He emphasized the many benefits of L2 material on the Web: freshness, diversity, large scope, low cost and convenience. Teachers have found beneficial results from having their students compile and study personalized corpora [8, 9, 11].

The Web augmentation approach has seen the most work on augmenting L2 texts to provide L1 annotations [24, 25, 27, 35]. Their augmentation techniques most often use affordances of the Web such as hyperlinks, pop ups and frames to overlay translations, dictionary definitions, grammatical explanations and cultural information that is easily accessible [25]. These annotations have been found to be beneficial to several aspects of language learning [25] and improvements in reading comprehension [27].

Wible *et al.* [34] and Pan & Huang [23] explored a Web-based L2 reading environment with the SRP and WBL projects respectively. Students who used these systems to read outside of class had better English reading comprehension, higher English learning motivation, and greater learner satisfaction when compared to students reading un-annotated paper equivalents. Wible *et al.* [34] also developed two browser based tools, Collocator and WordSpider, which annotated L2 Web pages by highlighting word collocations and provided in-page L2 contextual word clues respectively. Similarly, the Gymn@zilla project overlays the L2 Web with text and picture annotations and supports the creation of personal word lists and closure exercises [31] but no user study of the prototype was conducted.

Augmenting L1 Text with L2 Vocabulary

Similar to what we propose in this paper are the LEARN Project [2] and graded-reader [32]. Both projects promote the reading of L1 texts augmented with L2 vocabulary. However, the LEARN project requires users to manually find and enter text for it to translate while the graded-reader only works on the Bible. Furthermore, neither has been studied to determine how effective they are for language learning.

The LEARN Project is a standalone application, that when given any English text, would replace specific English words with L2 vocabulary words [2]. Translated words were highlighted and there was also support for fill-in-the-blank word guessing and accommodating learner advancement by translating more words. Language support was dependent on the presence of "word experts" for each language which guided the substitution of English words. The focus of the research was concentrated on the design and use of these "word experts" whose task was to examine the context of a target English word and then output the best translation.

Graded-reader is also a standalone application but was designed specifically as a Greek learning aid for reading the Greek version of the New Testament of the Bible [32]. It presents the New Testament in English with certain words or phrases translated into Greek. What makes graded-reader unique is the way in which it chooses what words and phrases to translate. Tauber approached the problem algorithmically with the goal of ordering the vocabulary presented to the learner in such a way that it maximizes the learner's ability to read a corpus. Tauber showed that word frequency does not always provide the best ordering when emphasizing the readability and comprehension of sentences in the corpus as opposed to words. The drawback of graded-reader is that it currently can only work on the text of the Bible because it depends on a manual tagging of grammatical, semantic and discourse related features of a text.

Figure 1: The ALOE prototype. (a) A page from the New York Times Web site translated by ALOE; the French words appear in a yellow highlight. (b) The settings page which allows the user to choose any combination of color and style to highlight translated words.

FORMATIVE INTERVIEW STUDY

To develop an understanding of the challenges involved with teaching a foreign language and general guidelines for designing a language learning environment, we interviewed 6 language professors and instructors from our institute.

We learned from these educators that motivating and challenging students, particularly beyond the classroom, is extremely important. In particular, they discussed how time on task is a very important principle of language learning, because the more exposure and practice a person has with a new language, the better she will learn it. One professor was very clear that it matters most that students study and remain motivated to study, rather than how they study.

Additionally, the educators emphasized the importance of grounding the learning in specific contexts that are of interest to each student. One strategy that educators use is to let students choose and study the second language texts they find most interesting. In this way, as students read, they can use their knowledge and interest in the subject matter as context and motivation to help them to better process and learn the language.

THE ALOE PROTOTYPE

The Web has an abundance of material which CALL researchers believe can be used for language learning [16]. Based on findings from the aforementioned formative interview study, we designed a system, called ALOE, to help language learners develop second language vocabulary while they are reading text that is written in their native language and is of personal interest to them (see Figure 1a).

High-Level Design

The ALOE prototype runs in a user's Web browser and translates words from English to a second language on Web pages as a user views the page. ALOE only translates a couple words in each sentence on a page and leaves the majority of the text in English. This approach does not result

in a learning environment with the "thick authenticity" as described by Shaffer and Resnick [29]. However, it assumes that Web pages viewed by the user are often of personal interest to her and will have a rich English context from which she can infer the meanings of the foreign words. Over time, new foreign words could be exposed enough times such that the practice of inferring their meanings may result in learning.

In particular, ALOE aims to support the learning of nouns. Nouns are among the most common type of words in a language; for example, over half the words in the Oxford English Dictionary are nouns[1]. Additionally, verbs and other parts of speech are highly dependent on context, which makes them harder to translate correctly.

Once ALOE translates a word on a Web page that a user is viewing, the word is highlighted on the page with a user-specified style. The default highlight is a simple dashed underline of the word in the same color of the word. This default was chosen because it allows translated words to stand out in a page while only slightly affecting a page's existing design. At the same time, it differentiates translated words from hyperlinks which traditionally have a solid underline. Users can change this style at any time by accessing the ALOE settings page; choosing any combination of color and style (background, underline, border) and line type for underlines and borders (dashed, dotted, solid) to highlight translated words (see Figure 1b).

The user can hover the mouse over a translated word to view a multiple choice question asking the meaning of the word (see Figure 2). The two incorrect choices are generated by a simple heuristic that chooses two words with a similar length to and the same part of speech as the correct word. Once the user answers the multiple choice question, ALOE indicates if she was right or wrong and reveals the correct translation.

---

[1] http://www.oxforddictionaries.com/page/93

Figure 2: When the user hovers the mouse over a translated word to view a multiple choice question asking the meaning of the word.

ALOE also provides a link to display the full definition of the foreign word using WordReference.com which provides additional information about the word, including a longer definition with synonyms and example phrases.

We chose to present a multiple choice question on mouse hovers based on Nagata's study that found multiple-choice glossing with immediate feedback to be more effective at promoting learning than simple glosses with just the translation [19]. The multiple choice questions challenge users, make them spend more time processing the word, and encourage them to attempt to guess the correct translation. Guessing helps the memory retain word meanings better than meanings provided by simple glosses [13, 14].

To prevent users from making mistakes on important and sensitive Web sites, ALOE does not run on secure (https) Web sites. Additionally, we also provide users with three different disable options. It could be disabled for a single URL, a single top-level domain, or temporarily disabled everywhere for eight hours. We designed the global disable to be temporary in order to prevent users from forgetting to re-enable ALOE.

Implementation

We designed ALOE as a client-server application. We developed the server as a webpy[2] based Python Web application that performs the following functions:

- Hosts the new user setup and configuration pages where users input their name, email and choose the style used for translated words.
- Administers the Web-based vocabulary quizzes.
- Acts as a caching middle-man between the Google Translate Web service and the ALOE client. As a middle-man, the server allows the client to make batch translation requests for quicker response times because Google Translate only interprets one piece of text at a time. It also returns the multiple choice options for every translated word.
- Collects and stores all client usage data in a MongoDB database.

- Serves static HTML and Javascript files used by the client.

We implemented the client as an extension for Firefox 3. Specifically, the client is a Greasemonkey-based user script that is packaged as an extension[3]. Packaging the script enables it to run automatically on Web pages without requiring users to install Greasemonkey[4].

When the ALOE client user script runs on a page, it loads the control panel and then parses the HTML looking for words to translate. The client looks for a set of high frequency English words[5]. Once all the translatable words are found, the client fetches the translations and their multiple choice questions from the browser cache (if the word has been seen before) or the ALOE server. It then replaces the target English words on the page with their translations and adds code to perform logging and to support the multiple choice and WordReference.com definition pop ups.

USER EVALUATION

To evaluate ALOE, we conducted a two month study in which participants used ALOE on their personal computers to augment their normal Web browsing activities.

Method

At the start of the study, each participant first met with a researcher to have ALOE installed on their computer. They were then given a walkthrough of how ALOE could be used. In order to explore how the participants would naturally interact with ALOE, the researcher told them that they could use ALOE however they liked and disable it whenever they felt it was necessary.

We used 3 quizzes to evaluate the users' vocabulary recognition and recall skills. The first quiz was given before the users started using ALOE. The second was given after one month and the third was given a month after the second quiz was completed. The participants were emailed links to the online quizzes at the designated times. They were given no set deadline for when quizzes had to be completed and progress in the quiz was saved so that they could complete part of the quiz and return later to finish it. We did send reminder emails to participants who did not complete the quiz after a few days. Because the quizzes contained about 600 questions, these measures were used to better

---

[2] http://webpy.org/

[3] Anthony Lieuallen's Greasemonkey script compiler from: http://arantius.com/misc/greasemonkey/script-compiler

[4] http://www.greasespot.net/

[5] The current proof-of-concept implementation only translates a target set words. This word set is composed of ~1500 nouns that we manually selected based on two criteria: high frequency (as measured in the British National Corpus) and low ambiguity (or words that do not have many different meanings and usage context). A full implementation would use these same criteria to automatically determine additional nouns to support after the user has learned a significant amount of the words in the target set.

accommodate the participants and prevent them from dropping out of the study.

Each quiz consisted of the full set of French words (singular form only) that could appear while using ALOE. The words were presented in a random order. For each of the French words, users had three options. They were asked to write the English translation if they knew it or write the letter R if they recognized the word but were unsure of its translation or leave it blank if they neither knew it nor recognized it. The instructions specifically asked that users not guess so that we could better account for words they thought they knew and words they recognized. No feedback was given during or after the quiz so as not to affect users' vocabulary knowledge. We graded the quizzes manually to accommodate misspellings and synonyms.

After the second quiz, the participants were randomly divided into two equal groups. The first group had the ALOE system remotely disabled so that we could study how users' vocabulary retention and recall was affected after a month of not using ALOE. The second group was given the option to either continue or stop using ALOE so that we could determine if users enjoyed using ALOE and wanted to continue using it. Both groups took the third quiz a month after the second quiz.

As soon as any participant stopped using ALOE, we conducted a semi-structured interview with the participant to explore their usage of ALOE and elicit feedback. We also encouraged all of the participants to contact us with feedback at any time during the study.

Throughout the two month study, ALOE logged extensive usage data for each participant. For every page on which ALOE was active, it logged how long the page was open, which words were translated on the page, which words the participants hovered over and answered multiple choice questions for and the multiple choice answers they chose. ALOE also logged which words participants viewed full definitions for and how often ALOE itself was disabled.

Participants

There were 21 participants recruited by word of mouth and flyers posted around our university's campus. We recruited participants who had some previous French experience, owned personal computers and regularly browsed the Web. As compensation for participating, participants names were entered into two raffles where three $100 gift certificates were given out in each raffle. The raffles took place after the second quiz and after the third quiz as a means of motivating the participants to complete the quizzes.

Two participants were dropped in the first month after one did not complete the first quiz and the second stopped using the browser in which ALOE was installed. A third participant was dropped after the participants were split into two groups because he/she did not complete the second quiz. No data for these three participants is included in our analysis. There was also one participant who did not complete the third quiz but whose other data was included in our analysis.

The 18 participants who completed the study included 7 males and 11 females. Fifteen of the participants were students (undergraduate or graduate) or recent university graduates and used ALOE on their personal laptops. By their own estimates, this group spent 4.5 hours a day browsing the Web (SD=3.13). The remaining three participants were middle-aged (40-55), employed and university educated and used ALOE on their personal desktop computers. These three participants gave estimates of their average Web browsing time that ranged from 40 minutes to an hour and a half a day.

All 18 participants had some prior French education ranging from primary (6/18) and secondary (14/18) education classes to University level (4/18) and professional (2/18) classes. Five of the participants were involved in some form of independent French practice (4/18) or University level French class (2/18) during the two months that they were participating in the study. The independent French practice ranged from watching French movies and listening to French radio to attending local French language meetups and talking to a friend in French. One participant was fluent in Spanish though none of the other participants had significant language experience in a Western European language besides French and English. The participants' motivations for participating in the study included improving their French skills (14/18), curiosity about the software (4/18), maintaining their existing French skills (4/18) and testing their French skills (2/18).

RESULTS

In this section, we report the results of our study. We first present how much time the participants spent using ALOE in total and on a day to day basis. We then describe how the participants interacted with ALOE. Finally, we analyze the quiz results to see how ALOE affected participants L2 vocabulary knowledge.

We split the 18 study participants into two groups. The first group consisted of eight participants who stopped using ALOE after completing the second quiz. The second group of ten participants was given the option of continuing to use ALOE after the second quiz. Of the participants given the option to continue, seven chose to continue and three chose to stop using ALOE. For our analysis, the three who voluntarily stopped are grouped with the eight who were forcibly stopped after the second quiz. We will discuss why the three participants stopped after we examine how the participants used the disable feature of ALOE.

Usage by Time

Overall, ALOE logged 1,334 hours of data across all the participants in the time before the second quiz. This can be broken down into 389 disabled hours, when participants were browsing pages ALOE did not translate, and 945 active hours on pages that ALOE had translated. These totals are only
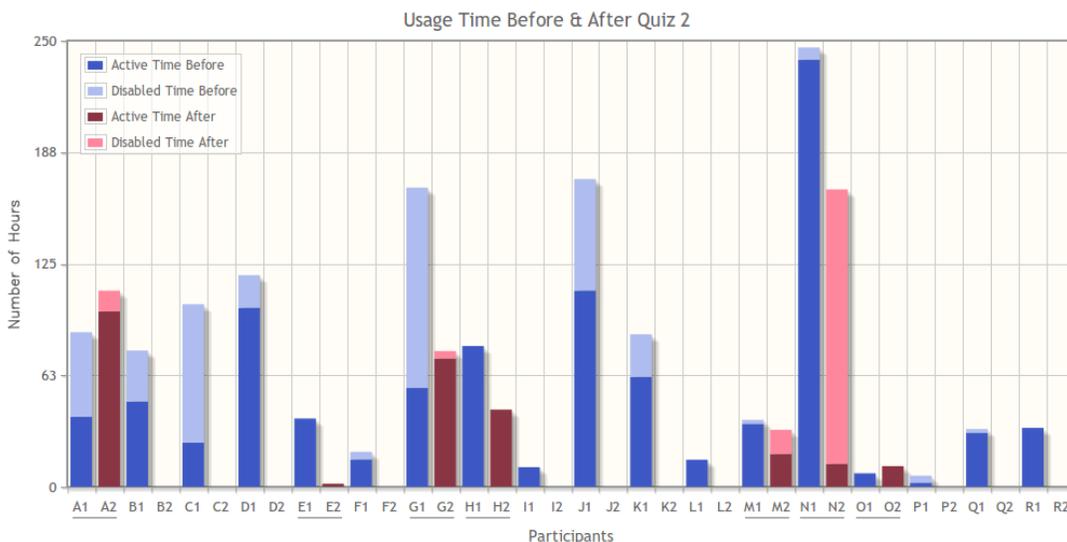
Figure 3: Total hours of usage by each participant broken into the times when ALOE was active and disabled. The seven participants who used ALOE for a second month are underlined and have a second bar showing their second month totals.

approximations of the actual time the participants spent browsing the Web. The logging monitors how long pages are open and focused in the Firefox Web browser. It does not take into account times when a participant used another Web browser, left their browser running while not at their computer or time spent on secure (https) Web pages.

Figure 3 shows how much active and disabled time each participant logged before completing the second quiz and afterwards for the seven participants who continued using ALOE. These time periods vary for each participant because some participants completed the second quiz when it was released and a few waited over a week before finishing it. A few of the participants also had extended periods of time (from a couple days to a week) where they were away from their computers while travelling. The graph shows that the participants varied widely both in their active hours (AVG=52.5; SD=54.1) and disabled hours (AVG=21.6; SD=31.7) before quiz two as well as after (active: AVG=36.9; SD=36.1) (disabled: AVG=36.6; SD=65.7).

For the most part, participants' time logs matched well with their self-reported estimates of Web browsing time. The notable exceptions are P and Q who reported spending many hours online each day but logged relatively few in ALOE. Participant Q reported that much of his Web browsing time during the study was spent working on a computer in a lab that did not have ALOE installed. Additionally, participant P would often use the Google Chrome Web browser when he did not want to use ALOE.

In terms of disabling ALOE, nine participants had ALOE disabled less than 10% of the time before quiz two and five of them never disabled ALOE. On the other side of the spectrum, four of the participants had ALOE disabled over half of the time before quiz two. Participants most often used the global disable option followed closely by the

domain disable and less frequently disabled single pages. Once a participant used one of the disable options, it was rare for them to re-enable ALOE.

We further analyzed the participants' active and disabled time by examining their day to day usage. Not counting days when participants logged no time, the daily active usage averaged 1.85 hours (SD=0.75) and the daily disabled time averaged 0.69 hours (SD=0.45). The only apparent trends in the day to day graphs appear to be that active time levels out after twenty days and that most participants held off on taking advantage of the disable feature until the fourth day. The periodic changes in frequency of use could be the effect of weekends but the participants started on different days of the week so the weekdays and weekends do not line up exactly.

In comparing the second month to the first month for the group of seven participants who chose to continue using ALOE for an extra month, we found that the daily active usage decreased by 0.45 hours and the daily disabled usage increased by 0.17 hours. For just the first month, the seven second month participants averaged 0.69 more active hours and 0.02 less disabled hours than the first month only participants.

From the interviews, we found a few reasons why the participants disabled or did not use ALOE. Ten participants cited work-related tasks, seven cited time-sensitive tasks, three cited Web site-specific issues with ALOE and one participant reported their mood as a reason. All of these reasons stem from the participants' belief that ALOE decreased the speed at which they could perform tasks on the Web. Most often the participants attributed the decrease in speed to extra time they had to devote to understanding text translated by ALOE. About a third of the participants reported that ALOE slowed down their Web browser or
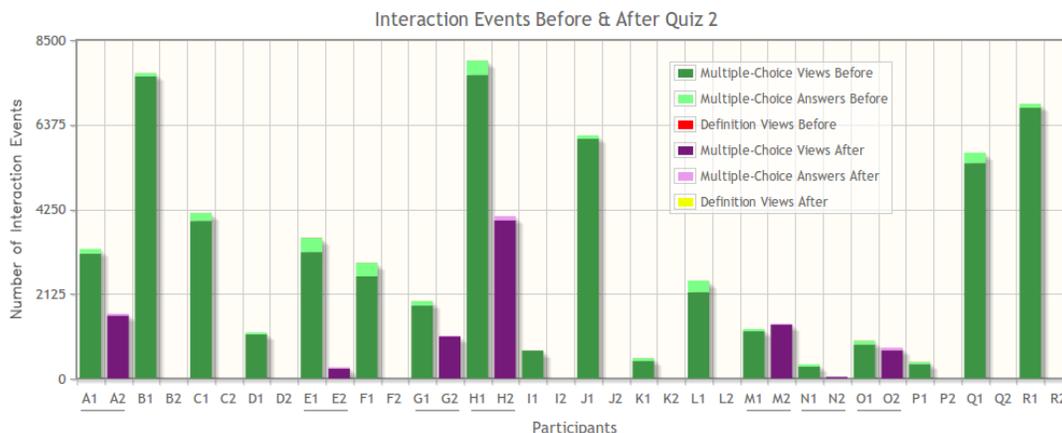
Figure 4: Total number of interaction events by each participant broken into the different types of interactions. The seven participants who used ALOE for a second month are underlined and have a second bar showing their second month totals.

caused Web pages to load slower. One participant commented that scrolling with ALOE running was slower. Even though we built the ALOE prototype with an emphasis for speed, the amount of text processing it performs can cause noticeable slow-downs in the Web browser when users have many Web pages open at once or open multiple Web pages in a quick succession. However, the slow scrolling issue is most likely caused by the logging code which monitors the scrolling to determine which translated words on a Web page are visible to the user.

Interaction

Figure 4 shows the total number of interactions recorded by ALOE for each participant during the same time periods as Figure 2. Participants hovering their mouse over translated words to view the multiple choice question made up the large majority of interactions. But, Figure 4 also shows that very few of the multiple choice interactions resulted in participants actually answering the multiple choice question. The counts for viewing WordReference.com definitions are also included in the graph, but because the participants used this feature an average of only 1.22 times, it is not visible.

There are two explanations for why participants answered very few multiple choice questions. First, participants reported that the multiple choice questions were so easy in context that they could determine which choice was the correct one without needing to answer the question. Second, the multiple choice popup is instantly triggered when the mouse cursor passes over a translated word. Given the density of translated words on a page, there are probably many false positive interactions counted when participants accidentally triggered the multiple choice display while moving their mouse within and across the Web browser window.

Dividing the total number of interactions and the total active time for each participant during this time period, we see that the participants' interactions per hour varied widely from participant N's minimum of 1.49 to participant R's maximum of 208.92 (AVG=104.12; SD=70.74). With the exception of

participants O and I, the participants who used ALOE the least amount of active time (participants F, C, L, P, Q, R) had the highest number of interactions per hour (> 160). However, this trend does not hold true for the participants with the most active time who are more evenly dispersed.

Comparing the second month interactions to the first month for the seven second month participants, we can see that most of them dropped their total interactions by half or more. The day to day plots below will explain this in more detail. But in terms of the hourly interactions metric, there was not a significant decrease (AVG=-5.57) from these participants' hourly interactions in the first month.

In the post-study interview, we asked participants to estimate how often they stopped to think about or interact with translated words. The answers were positioned on a 5-point Likert scale ranging from very seldom to very often. Most participants felt that they thought about or interacted with words every now and then or often (AVG=3.3, SD=1.1). Because this data is subjective and includes estimates of times participants thought about but did not interact with the words, the participants' answers do not match with the interactions per hour metric we previously calculated and cannot be compared to that metric. From our interviews, we found that many participants would guess the translated words from context without interacting with the word. Only when they were unsure of their guess would they view the multiple choice question. As one participant put it, *"the [multiple choice] answers are obvious most of the time"* and because of this many of the participants did not feel the need to answer the multiple choice questions.

Both the amount of multiple choice views and the amount of multiple choice answers start elevated and level off after a few days; this can be attributed to participants playing with ALOE while it is new to them. Otherwise, both types of interactions stay pretty stable throughout the first month with an average of 108.69 multiple choice views (SD=36.39) and 5.08 multiple choice answers (SD=7.06) per day.

For the group of seven participants who continued using ALOE for the second month, both types of interactions have decreased on average by a bit more than half. The amount of multiple choice views averages 52.90 (SD 20.54) and the amount of multiple choice answers averages 5.08 (SD 7.06).

The decrease in interactions is probably due to the static nature of the set of words being translated. A few of the participants noted that they always saw the same set of words translated, especially on pages where the content did not change much, and so once they learned these words they no longer interacted with them. Therefore, as the participants' French vocabulary increased, they encountered fewer unknown words to interact with.

The interviews shed some light on what types of activities participants were doing when ALOE was active and all these interactions occurred. In contrast to the work-related and time-sensitive tasks that occurred when most participants disabled ALOE, most of the participants (14/18) reported using ALOE the most while casually browsing the Web. This included activities such as reading news and blogs, searching, shopping, sending email and using social networking sites. Three participants could not identify any specific activity where they used ALOE more often and one participant reported using it the most on pages with lots of text.

Quiz Marks

Figure 5 shows the quiz marks for all participants on all three quizzes. The green bars represent the number of correct translations and align to the left axis. The red bars represent the number of incorrect translations and align to the right axis. Each participant's three quiz marks are grouped together such that the bars with the lightest colors represents the first quiz and the bars with darkest colors represents the third quiz. The empty space in between the green and red bars accounts for quiz words participants skipped or recognized but did not provide a translation. The

participant labels for the seven participants who continued using ALOE after quiz two are underlined.

Although the participants were instructed not to guess on the quizzes, it was clearly apparent from grading the first quiz that every participant (except participants H and P) was providing answers they were not completely sure of. One participant (participant K) put question marks in some of her answers. Other participants would provide multiple guesses for one word and many would guess incorrectly on "false friends", French words that look similar to English words but have different meanings. Because participant P and H had so few correct translations and even fewer wrong translations on quiz one, we contacted them to determine if they had restrained themselves from guessing. Both confirmed that they had taken care not to guess. As participant H put it:

"I wasn't 100% sure, and the instruction said "don't guess"."

As long as all the participants stuck with a consistent quiz strategy for all three quizzes, we can extract useful information from their quiz marks. Participant C's quiz scores are therefore problematic because she reported that she stopped guessing after quiz one. This shows up prominently in the graph as she is the only participant to get fewer words correct on the second quiz. Therefore, we do not have her marks factored into any of the following analyses.

On average, the participants answered 50 more words correctly (SD=26.38) and 2.59 more words incorrectly (SD=24.47) on quiz two. On quiz three, the seven participants who used ALOE for an extra month answered 2.27 fewer words correctly (SD=25.05) and 4.14 more words incorrectly (SD=20.41). The eleven participants who stopped using ALOE averaged 2 more words correct (SD=17.31) and 2.89 fewer word incorrect (SD=21.18) on quiz three.

Using statistical methods, we analyzed the groups and their scores to determine if there were any statistically significant differences. By using a t-test to compare the two groups, we
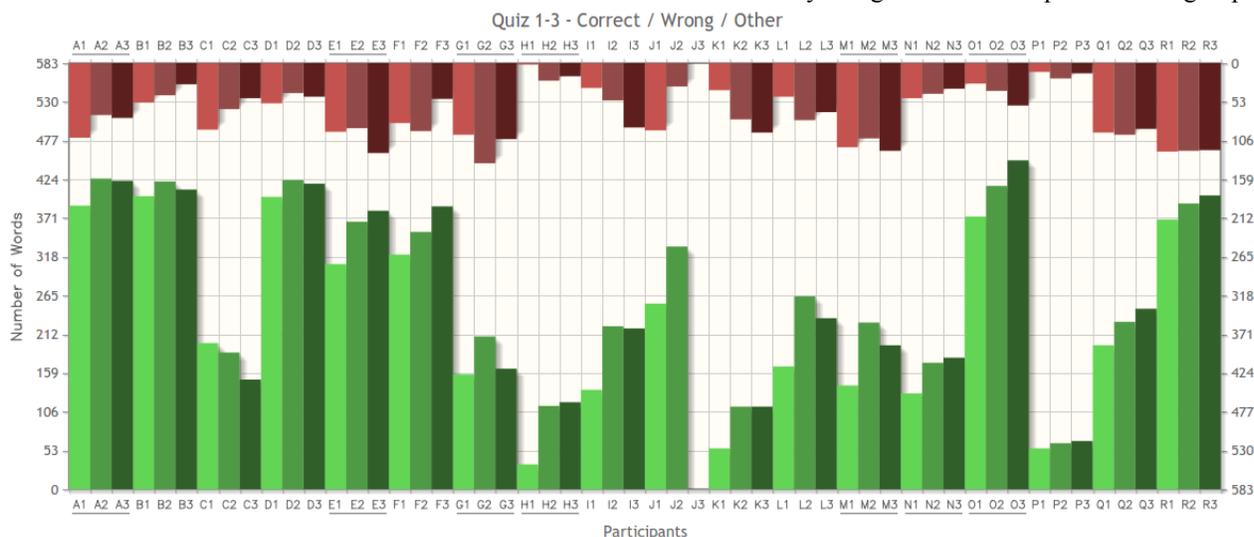

Figure 5: Quiz 1, 2 and 3 scores for all of the participants. The seven participants who used ALOE for a second month are underlined and participant J never completed quiz 3.

found no significant difference (p > 0.05) in how the groups' amounts of correct marks changed from quiz two to quiz three. We then ran a repeated measure ANOVA on both group's correct marks in all three quizzes. This found that there was a significant difference (p < 0.001) between marks for quiz one and quiz two and also between marks for quiz one and quiz three. Marks for quiz two and quiz three were not found to be significantly different. This suggests there is no difference between the group of participants who stopped using ALOE and those who continued using ALOE, although the sample sizes were small so these results should be considered preliminary. As a whole, the participants improved between quiz one and quiz two but neither improved nor did worse on quiz three.

The fact that an extra month of using ALOE did not significantly improve participants' vocabulary is likely related to the decrease in participant interactions during the second month. Because the participants encountered the same set of words repeatedly, even after they had already learned the words, they had far fewer opportunities to learn new words.

As mentioned in the Participants section, five of the participants (A, E, O, J and F) were involved in some form of independent French practice or University level French class during the two months that they were participating in the study. To understand if additional learning methods impacted the participants' results, we compared quiz score results for participants A, E, O, J and F to the rest of the participants and found no significant difference on number of words learned between quiz 1 and quiz 2, or quiz 1 and quiz 3, or quiz 2 and quiz 3 (p > 0.05 using a t-test).

Finally, we also analyzed if there was a correlation between learning with prior knowledge. We split the participants' quiz results into two groups based on their self-designated French level (beginner or intermediate). We found no significant difference in learning between each group (p > 0.05 using a t-test).

Overall, all the statistics reported above imply that the idea supported by ALOE, learning L2 vocabulary by inserting L2 words into the context of L1 Web pages, produces promising initial results. Participants were able to learn new French vocabulary and retain it for a month with negligible loss of recall.

CONCLUSIONS & FUTURE WORK

In this work, we explored the feasibility of integrating L2 learning with everyday Web browsing experiences. Most of the participants found using the ALOE prototype to be an enjoyable way to learn L2 vocabulary and wanted to continue using it after the study. Using ALOE, participants in our user study were able to learn an average of fifty new French vocabulary words in a month with little change to their normal Web browsing experience.

Although findings from our study are promising, we note that the evaluation itself has potential learning impact and could have influenced the results. Furthermore, the evaluation was conducted over a 2-month period using a proof-of-concept prototype which operated on a limited set of words. Thus, the long-term effect of language learning using this approach remains to be investigated.

We believe that adding an intelligent tutoring system to the back-end of the system is an important next-step for this work. Because the current prototype operates on a limited set of words, participants began to encounter the same set of words over and over again. Based on the participants' feedback and the lack of improvement and decrease in interactions in the two month participant group, it is also clear that the system would need to support a much larger set of words in order to find opportunities to teach users new words on the pages that they visit. However, as the set of words grows, the system must begin to consider the level difficulty of different words, gauge the user's level of knowledge, and present words at the appropriate level of language proficiency.

Additionally, ALOE needs to adapt as participants learn in order to continually challenge them. The adaptation should be able to recognize when people have learned a word so that ALOE can stop translating it and instead translate a new word. It would also be interesting to see how increasing the difficulty of the multiple choice questions would affect vocabulary learning. Most of the study participants found that they could easily identify the correct multiple choice answer without actually answering. From a usability perspective, this speeds up how quickly they are able to read and understand Web pages but it also promotes shallow processing of the translated words which is not as conducive to vocabulary acquisition.

REFERENCES

1. Beaudin, J. Intille, S., Tapia, E.M., Rockinson, R., Morris, M. Context-Sensitive Microlearning of Foreign Language Vocabulary on a Mobile Device. Ambient Intelligence, 2007, p. 55–72.
2. Berleant, D., Lovelady, S., Viswanathan, K. A foreign vocabulary learning aid for the networked world of tomorrow: the LEARN project. ACM SIGICE Bulletin, vol. 19, 1994, pp. 22-29.
3. Cavus N., Ibrahim, D. m-Learning: An experiment in using SMS to support learning new English language words. British Journal of Educational Technology, vol. 40, 2009, pp. 78-91.
4. Chen, C., Hsu, S.. Li, Y., Peng, C. Personalized Intelligent M-learning System for Supporting Effective English

Learning. 2006 IEEE International Conference on Systems, Man and Cybernetics, vol. 11, 2006, pp. 4898-4903.

5. Chun, D.M., Plass, J.L. Effects of multimedia annotations on vocabulary acquisition. The Modern Language Journal, vol. 80, 1996, pp. 183–198.

6. Fallahkhair, S., Pemberton, L., Griffiths, R. Development of a cross-platform ubiquitous language learning service via mobile phone and interactive television. Journal of Computer Assisted Learning, vol. 23, 2007, pp. 312-325.

7. Fletcher, W.H. Concordancing the Web: Promise and Problems, Tools and Techniques. Corpus linguistics and the Web, 2007, p. 25–45.

8. Friedman, G.L. Learner-created lexical databases using web-based source material. ELT Journal, vol. 63, 2008, pp. 126-136.

9. Guo, S., Zhang, G. Building a customised Google-based collocation collector to enhance language learning. British Journal of Educational Technology, vol. 38, 2007, pp. 747-751.

10. Helft, M. Using Computing Might, Google Improves Translation Tool. The New York Times, 2010.

11. Hirata, Y., Hirata, Y. Independent research project with web-derived corpora for language learning. The JALT CALL Journal, vol. 3, 2007, pp. 33–48.

12. Hubbard, P. A General Introduction to Computer Assisted Language Learning. Computer Assisted Language Learning: Critical Concepts in Linguistics. Volume I – Foundations of CALL, 2009, pp. 1-20.

13. Hulstijn, J. Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud and H. Béjoint, editors, Vocabulary and Applied Linguistics, pages 113–125. Macmillan, London, 1992.

14. Hulstijn, J.H., Hollander, M., Greidanus, T. 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(iii), 1996, pp. 327–339.

15. Kolers, P.A., Ostry, D.J.. Time course of loss of information regarding pattern analyzing operations. Journal of Verbal Learning and Verbal Behavior, vol. 13, 1974, pp. 599–612.

16. Levy, M., Stockwell, G. CALL Dimensions: Options and Issues in Computer-Assisted Language Learning. Lawrence Erlbaum Associates, N.J., 2006, 328 pages.

17. Li, C. SMS-based Vocabulary Learning for ESL Students. Chun Li. 2009.

18. Liu, T. A context-aware ubiquitous learning environment for language listening and speaking. Journal of Computer Assisted Learning, vol. 25, 2009, pp. 515-527.

19. Nagata, N. The effectiveness of computer-assisted interactive glosses. Foreign Language Annals, vol. 32(4), 1999, pp. 469–479.

20. Nation, I.S.P. Teaching and Learning Vocabulary. Heinle & Heinle, Boston, MA, 1990.

21. Ogata, H., Chengjiu Yin, Paredes, R.G., Nobuji, J., Saito, A., Yano, Y., Oishi, Y., Ueda, T. Supporting Mobile Language Learning outside Classrooms. Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06), 2006, pp. 928-932.

22. Ogata, H., Yano, Y. Context-aware support for computer-supported ubiquitous learning. Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, 2004, pp. 27-34.

23. Pan, C., Huang, P. The development and evaluation of a web-based college English outside reading system for EFL freshmen. JALTCALL, vol. 5, 2009, pp. 3-22.

24. Peters, E. Manipulating L2 learners' online dictionary use and its effect on L2 word retention. Language Learning & Technology, vol. 11, 2007, p. 36–58.

25. Ridder, I.D. Visible or invisible links: does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? Language Learning & Technology, vol. 6, 2002, pp. 123-146.

26. Rott, S. The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. Studies in Second Language Acquisition, vol. 21(4), 1999, pp. 589–619.

27. Sankó, G. The Effects of Hypertextual Input Modification on L2 Vocabulary Acquisition and Retention. University of Pécs Roundtable 2006: Empirical Studies in English Applied Linguistics, M. Nikolov and J. Horváth, Lingua Franca Csoport, 2006, pp. 157-178.

28. Segler, T., Pain, H., Sorace, A. Second Language Vocabulary Acquisition and Learning Strategies in ICALL Environments. Computer Assisted Language Learning, vol. 15(4), 2002, pp.409-422.

29. Shaffer, D.W., Resnick, M. "Thick" Authenticity: New Media and Authentic Learning. Journal of Interactive Learning Research, vol. 10(2), 1999, pp. 195-215.

30. Sternberg, R.J. Most vocabulary is learned from context. In M.G.McKeown and M.E. Curtis, editors, The Nature of Vocabulary Acquisition, pages 89–105. Lawrence Erlbaum, Hillsdale, N.J., 1987.

31. Streiter, O., Knapp, J., Voltmer, L., Zielinski, D. Browsers for autonomous and contextualized language learning: tools and theories. ITRE 2005. 3rd International Conference on Information Technology: Research and Education, 2005, pp. 343-347.

32. Tauber, J. A New Kind Of Graded Reader. Feb. 10, 2008.

33. Thornton, P., Houser, C. Using mobile phones in English education in Japan. Journal of Computer Assisted Learning, vol. 21(3), 2005, pp. 217-228.

34. Wible, D. Automating repeated exposure to target vocabulary for second language learners. Proceedings IEEE International Conference on Advanced Learning Technologies, 2001, pp. 127-128.

35. Yanguas, I. Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. Language Learning and Technology, vol. 13, 2009, p. 48–67.

36. Yongqi, G., Johnson, R.K. Vocabulary learning strategies and language learning outcomes. Language Learning, vol. 46(4), 1996, pp. 643–679.