

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF TORONTO

Augmenting the L1 Web for L2 Vocabulary Learning

Andrew Trusty
April 2010

Abstract

We examine how a person's existing Web browsing experience on first language Web pages can be augmented to teach second language vocabulary. We present a prototype, ALOE, which runs inside the Firefox web browser and dynamically augments Web pages by replacing a select set of English words with their French translations. The French translations are embedded in the rich context of a Web pages existing English text which promotes the incidental learning and guessing from context of the translated words. Our two month user evaluation of the ALOE prototype demonstrates that our approach works in practice. Most of the participants enjoyed using ALOE and they were able to learn an average of fifty new French vocabulary words.

Table of Contents

Introduction	1
Motivation.....	1
Contributions.....	2
Overview.....	3
Related Work.....	4
CALL and Mobile Devices.....	4
CALL and the Web.....	5
Web Mining	5
Web Augmentation.....	6
Augmenting L1 with L2.....	10
Interviews	12
Prototype.....	15
Design Goals.....	15
Prototype System.....	16
Implementation	22
User Evaluation.....	23
Participants	24
Results	26
Usage By Time	26
Interactions	30
Quizzes.....	35
Interviews	37
Conclusions & Future Work.....	41
Bibliography	42
Appendix A - Definitions	45
Appendix B - Participant Demographics.....	46
Appendix C - Quiz Examples.....	47
Appendix D - Semi-structured Interview Script.....	49
Appendix E - Translation Process Sequence Diagram.....	50
Appendix F - French Words Translated by ALOE.....	51

1 - Introduction

Motivation

Each person has their own reasons for wanting to learn a language. Whether they want to learn a few Italian phrases to use while they stay at a villa in Tuscany or they want to impress the cute Flugbegleiterin (stewardess) on Lufthansa during their business trips to Germany or just polish up their old skills they learned in school, there is invariably some innate motivation driving humans to learn other languages. 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. And with modern technologies increasingly shrinking the "distance" between people and cultures across the globe, being able to communicate and understand each other will make all the difference in the world.

But just because everyone desires to learn a new language does not mean that they will ever get around to it. In the hierarchy of goals, learning new languages most often gets relegated to the bottom of the list, something to be done in the "future" when people have more time. But how often does someone have the urge to pull out a big deck of vocabulary flashcards and spend a solid half hour reviewing them? Or to study the intricacies of past perfect and conditional French verb conjugations? The sad truth is that learning a new language requires devoting time, attention and focus on a recurring basis. The more exposure and practice someone has with a new language, the better they will learn.

The advent of ubiquitous computing and our daily interactions with dozens if not hundreds of powerful computer systems provides an ideal environment for promoting language learning in any context. If each interaction can provide just one language learning opportunity, then it is easy to envision someone improving their language skills on a day to day basis. But someone can not just switch their computer to a Kanji interface and expect to learn Japanese. There is a delicate balance between the presentation of learning opportunities and the inconvenience these opportunities cause as impediments to normal activities.

We argue that traditional language study tasks can be broken into small micro learning opportunities. And these opportunities can be distributed throughout a person's day-to-day life in an unintrusive way that maximizes learning. To that end, this thesis explores a Web-based approach to providing micro learning opportunities for second language learners. At its core, the Web is a medium for communication and it is the largest, fastest growing corpus of machine readable authentic language material. As people are spending increasing amounts of time on the Web and accessing it with more and more devices, the Web is the perfect testbed for enabling the idea of ubiquitous language learning.

The idea of using the Web for language learning is not new though. Many previous research projects have studied how to best move traditional language learning activities online and how such activities can be adapted to best take advantages of the affordances that the Web provides. Too often these studies are constrained by their traditional models of language learning and do not push the limits of the Web to enable new language learning models.

In this thesis, we examine how a person's existing Web browsing experience can be augmented to teach second language vocabulary. The novelty of this approach lies in the fact that we aim to teach second language vocabulary while people are reading pages that are written in their native language. There are many studies that tout the effectiveness of reading in a second language to learn vocabulary but none that address the effectiveness of our approach. This is not surprising given that such an idea would be extremely challenging to implement using paper study materials but the Web makes it practical.

The goal of our research is to study the usability and effectiveness of learning second language vocabulary on first language Web pages. We aim to answer whether people are able to learn, how much they learn and if they enjoy learning using the browser-based prototype that we designed, called ALOE. We also hope to uncover interface design issues that arise from this approach with an emphasis on exploring how to integrate learning opportunities without inconveniencing learners during their normal Web activities.

Contributions

Web Browser Extension for L2 Vocabulary Learning

We present 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.

User Evaluation of L2 Vocabulary Learning on the L1 Web

Our two month user evaluation of the ALOE prototype demonstrates that our L2 vocabulary learning approach works in practice. Most of the participants enjoyed using ALOE and they were able to learn an average of fifty new French vocabulary words. We found that most of the participants wanted to continue using the ALOE prototype as-is but would have benefited from improvements in speed, Website compatibility, learning adaptability and customizability of ALOE.

Overview

In chapter 2, we discuss related work in the field of computer assisted language learning with a focus on systems designed for ubiquitous language learning and language learning on the Web.

In chapter 3, we discuss the results of our interviews with six second language educators from the University of Toronto. Our interviews focused on exploring the challenges faced by second language educators, how they used technology in their classes and how our idea for ALOE might fit in the language learning environment.

In chapter 4, we discuss the design and implementation of ALOE, our second language vocabulary learning prototype for the Mozilla Firefox Web browser.

In chapter 5, we present our user evaluation plan for testing the usability and effectiveness of ALOE as a second language vocabulary learning tool and our results from running the evaluation.

In chapter 6, we summarize the results of our work and motivate future research in this area.

2 – Related Work

Computer Assisted Language Learning (CALL) is a fast evolving field that is changing at the same pace of the technology that enables it. From its humble beginnings running on mainframes in the 1950's and 60's, CALL applications have been developed for almost every computing device imaginable ranging from microcomputers and laser videodiscs to mp3 players, mobile phones and DVD players (Hubbard 2009). CALL is also one of the few domains that motivates cutting edge research in multiple disciplines such as speech recognition, natural language processing, machine translation and artificial intelligence (Gamper & Knapp 2002).

The field of CALL is beneficial to language learners because it improves the effectiveness and efficiency of language learners while also making it easier and more convenient for both teachers and students to access and use language learning materials (Hubbard 2009). These benefits serve to more fully engage and motivate language learners and even enable them to develop their own autonomy in language learning (Hubbard 2009).

Many researchers have investigated augmenting existing student practices and environments to aid in language learning. By developing these new language learning practices and environments, researchers hope to extend language learning beyond the traditional classroom setting and encourage students to spend more time honing their language learning skills.

CALL and Mobile Devices

A number of projects have concentrated on augmenting user's mobile devices in order to take advantage of the anytime, anywhere affordances such devices provide. For example, CALL applications on mobile phones have been created to study the use of SMS for sending L2 vocabulary definitions to learners (Lu 2008, Cavus & Ibrahim 2009, Li 2009), providing adaptive L2 news article recommendations (Chen & Hsu 2006) and creating sensor augmented physical spaces that trigger the presentation of learning materials on the user's mobile device (Liu 2009, Beaudin *et al.* 2007, Ogata *et al.* 2006, Ogata & Yano 2004). The 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.

The TAMALLE system built by Fallahkhair, Pemberton & Griffiths (2007) is particularly interesting for their use of both the mobile phone and the television in an attempt to create a seamless and unobtrusive language learning environment from people's television watching habits similar to what ALOE attempts with the web. By using authentic L2 television programs, their work limits potential users to

advanced learners who have the skills necessary to follow such programs. The TAMALLE implementation presented also had no ability to automatically attach language learning materials to arbitrary television programs and depended on manually augmented television programs. In their qualitative in-lab evaluation, participants enjoyed using TAMALLE and found it useful for language learning, though there was no data to reflect whether or not participants actually improved their language skills. The biggest advantage was found to be that TAMALLE integrated learning into an activity the user already enjoyed and ironically the biggest disadvantage was that it sometimes served as a distraction from watching television.

CALL and the Web

A lot of recent CALL research has been focused on using the internet to enable new CALL applications. The Internet has provided a flexible new platform for defining new types of interactions between teachers, students and language learning materials. Teachers can run classes on the other side of the world and still have face to face interactions with students. Students can easily find pen pals by connecting with native L2 speakers. And the web has become the largest corpus of language learning materials ever realized. In a few clicks, a student can find enough L2 games, exercises, quizzes, grammar rules and course materials for almost any topic and language level to last them a lifetime of study. Not to mention the overwhelming amount of informal and authentic L2 material available from L2 speakers and cultures such as blogs, podcasts, YouTube videos, tweets and websites of every shape and kind.

Realizing the value of the Web as a language learning corpus, a number of researchers have examined how to make the web-corpus useful for language learners. There are basically two approaches to using the web as a corpus, web mining and web augmentation. Web mining attempts to extract language learning data from the web. Web augmentation, which ALOE uses, attempts to overlay language learning functionality on top of the web.

Web Mining

Fletcher (2005) goes into detail about the value of the web as a corpus and characterizes the different ways it is being used by teachers, students and researchers taking the Web mining approach to the Web as a corpus. Fletcher motivates future research on the Web as a corpus with the following arguments:

“

- Freshness and spontaneity: the content of compiled corpora ages quickly,

while texts on contemporary issues and authentic examples of current, non-standard, or emerging language usage thrive online.

- Completeness and scope: existing corpora may lack a text genre or content domain of interest, or else may not provide sufficient examples of an expression or construction easily located online; some very productive contemporary genres (blogs, wikis, discussion forums...) exist only on the Net.
- Linguistic diversity: languages and language varieties for which no corpora have been compiled are found online.
- Cost and convenience: the Web is virtually free, and desktop computers to retrieve and process webpages are available to researchers and students alike.
- Representativeness: as the proportion of information, communication and entertainment delivered via the Net grows, language on and of the Web increasingly reflects and enriches our tongue.

”

Fletcher does caution though that the Web is not without its drawbacks. With a sort of creolized web English contaminating non-English content and the commonplace occurrence of misspellings and the cryptic language of acronyms and netspeak that has evolved from chat rooms, the Web is fraught with obstacles that could confuse and impede language learners (Fletcher 2005). But this has not stopped researchers from mining the Web or teachers from sending their students to compile personalized corpora with beneficial results (Fletcher 2005, Guo & Zhang 2007, Friedman 2008, Hirata & Hirata 2007 to name a few).

Web Augmentation

A number of studies have looked at how different approaches to text augmentation affect learner behavior and language learning (Ridder 2002, Sankó 2006, Peters 2007, Yanguas 2009 to name a few). 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 (Ridder 2002). The availability of these annotations is “considered to be one of the preeminent advantages of language learning via computers, and consulting any of these extras is no longer seen as a major interruption of the language-learning activity” (Ridder 2002). And usually, these annotations have been found to be beneficial to several aspects of language learning (Ridder 2002) and improvements in reading comprehension (Sankó 2006, Yanguas 2009). It is less clear whether such annotations significantly affect recall (Yanguas 2009) and retention (Peters 2007). Peters attributes the lack of solid

evidence for improved retention and recall to a shallow processing of unknown words by learners. Many researchers believe that such shallow processing is caused by the ease of use of digital annotations although Peters attributes it to the type of task being performed (Peters 2007). Given that the materials being used in most of these studies were L2-only, it's necessary to exercise caution when extrapolating how these results translate into the L1 augmented domain of ALOE.

Before digging into the details of studies on web augmentation techniques and their results, it will be helpful to first consider the results of Hsu's study of student's experiences using the un-augmented web for language learning (Hsu 2005). In the study, 46 university students taking first year English were required to read English web sites and communicate with native English speakers using email or chat outside of the class. The student's were initially frustrated and even intimidated by English only web pages because they didn't have the skills necessary to understand the pages quickly and efficiently. The reactions of the students to this sudden immersion in a foreign language show that surfing the web for language learning can be a daunting experience. In this light, the following web augmentation projects can be seen as a way to alleviate such problems and make the foreign language Web friendlier for language learners.

Wible *et al.* (2001) and Pan & Huang (2009) explored the use of the web-based learning environment for learners to use outside of the classroom with SRP and WBL respectively. SRP was designed to find reading material in the Brown corpus for learners based on a set of target vocabulary words and the language level of each student. WBL provided the biographies of modern celebrities for students to read with embedded hyperlinks to additional resources. There was no study of SRP but the study of WBL found that use of their web-based learning system resulted in significantly better English reading comprehension, higher English learning motivation, and greater learner satisfaction when compared to the paper equivalent. The students using WBL also had very positive views of it and the Pan & Huang attributed its success to the fact that it provides for flexible accessibility, individual pacing and assistive resources.

Wible *et al.* (2004) also developed two browser based tools, Collocator and WordSpider, which run as a browser plugin and augment L2 web pages with L2 annotations. Collocator automatically runs on every Web page the learner visits and detects and highlights word collocations. The learner can then select any of the highlighted collocations to list additional examples of that collocation. Collocations are one of the most persistent areas of difficulty for learners improving their L2 vocabulary. The purpose of Collocator is thus to condense massive amounts of contextualized exposure into a single reading experience (Wible *et al.* 2004). WordSpider on the other hand, like ALOE, attempts to help learners acquire new L2 vocabulary. Based on the pedagogical motivation of teaching learners how to guess from context, Wible *et al.* do not show an L1 gloss or multiple choice as ALOE does. Instead, when a learner selects an unknown word, WordSpider searches for semantically related words in the context of the current page in order to provide

contextual clues for the learner. The semantically related words are highlighted and when the learner hovers over them, WordSpider provides a pop-up describing the relation between the highlighted word and the unknown word.

The Gymn@zilla project is an open-source web application that allows language learners to surf the L2 web with text and picture annotations overlaid onto every page (Streiter *et al.* 2005). In addition to the annotated browsing, they provide support for creating personal word lists and automatically generated cloze exercises using learner's word lists. Gymn@zilla acts as a proxy in order to retrieve and annotate Web pages on demand in response to learner requests. For every page it retrieves, it uses natural language processing modules to apply word stemming, word-segmentation, parts of speech tagging and word meaning disambiguation in order to accurately map words onto L1 dictionary entries used in their annotations. Gymn@zilla also parses the images in the page and tries to relate images to words via the image file name and directory. It then modifies the page and adds annotations to the recognized words which are highlighted with a dashed grey underline. The annotations include the following:

- the L2 word along with any L1 dictionary entries
- a related image if one was found
- a link to add the L2 word to the user's personal word list
- links to search for the L2 word with Google search and Google Image search

Using the mouse to hover over an annotated word brings up the annotation and highlights all other occurrences of the word on the page. Highlighting the other occurrences allows learners to explore the different contexts the word can be used in. Using the personal word list, learners can review L2 words they have saved along with the context they encountered the word in. Gymn@zilla is also able to automatically generate L2-only cloze exercises with the word list words at varying levels of difficulty. The different levels of difficulty are obtained by controlling the amount of character scrambling that is done to generate the options for each cloze question. Gymn@zilla supports a number of different L2-L1 language pairs (though the L1 choice is restricted by the L2 chosen). The current version of Gymn@zilla that is available online¹ supports over twenty separate L2 languages and many country specific dialects.

Gymn@zilla has the ambitious goal of integrating many modern approaches to language learning, including learner autonomy, personalized learning, contextualized language learning and learning with authentic material, by facilitating the incidental and intentional learning of L2 languages on the web. Their approach has a couple drawbacks though. Streiter *et al.* (2005) note that their method is often not fast enough and if used in isolation, would take a considerable amount of use to master the vocabulary needed for fluent and correct conversations. The slowness is due to the fact that about half the words of any text occur only once (Streiter *et al.* 2005). This makes it hard for a learner to gain the required amount of

¹ <http://140.127.211.214/cgi-bin/gz-cgi/gzstart.pl>

exposure needed to learn those words. Gymn@zilla does try to address this problem by providing the Google search links in the annotation pop ups so that learners can actively seek further examples of words.

Streiter *et al.* (2005) also point out the same problem that Hsu discovered regarding the stumbling block that complete L2 immersion can become to novice L2 students. To this, Streiter *et al.* can only suggest that novices take a more intentional approach and actively practice and learn the unknown L2 vocabulary they encounter until they overcome their inhibitions.

Sadly, there have been no user studies evaluating the effectiveness of Gymn@zilla for language learning. Streiter *et al.* mention a proposed future implementation would work as a plugin for Firefox like ALOE.

Outside of research, there are a plethora of options for providing instant glossing and translation of web page content. The Mozilla add-on site² alone lists 400+ language support extensions that provide toolbars, pop ups and other types of annotations, translations and language related services for the Firefox web browser and the recently launched Google Chrome extension site³ features dozens more. Almost all of the translation and gloss packages are powered by freely available services such as Google Translate, BabelFish, Wikipedia, Wiktionary and many other language specific dictionaries online.

On the commercial front, there has been an equal explosion in language learning offerings on the web. Most of the commercial sites offer language learning games, flashcards, audio and video material, quizzes and other exercises. There is also a growing selection of sites that match L2 language learners with L2 native speakers. WordChamp and UltraLingua in particular are most relevant to ALOE. Both WordChamp's WebReader service and UltraLingua's WebLex service support extensive L2 reading using the proxy model of Gymn@zilla. And also like Gymn@zilla, there are no user studies exploring the effectiveness of these tools for language learning.

Using WordChamp's WebReader, a learner can read web pages in any of the over 100 languages WordChamp supports and access hover over annotations for any word on the page in any of the supported languages. The WebReader is much more responsive than Gymn@zilla, probably due to the fact that the WebReader does not use NLP to find the one correct translation but instead just shows all the possible translations and lets the learner decide which is correct. WebReader also allows the learner to play short audio pronunciations of each translation, add words to a practice list and submit and view translations from other users.

UltraLingua's WebLex does the same thing as WebReader but lacks many of WebReader's features including audio pronunciations, user translations and word practice lists. WebLex also only supports 10 different L1-L2 language pairs, has problems displaying accented characters and doesn't automatically proxy links so the learner has to manually input the address for each page they want to read with WebLex annotations.

² <https://addons.mozilla.org/>

³ <https://chrome.google.com/extensions/>

Augmenting L1 with L2

It is clear that the majority of research on CALL applications developed to improve vocabulary through reading concentrate on authentic L2 texts and how to augment them to make them accessible and useful for language learners. There are two projects though, The LEARN Project (Berleant *et al.* 1994) and graded-reader (Tauber 2008), that take the same reversed approach as ALOE by promoting reading of L1 texts augmented with L2 vocabulary. Though neither have been studied to determine how effective they are for language learning.

The LEARN Project was a standalone application for IBM PCs and Sun OS machines. Given any English text, it would replace specific English words with L2 vocabulary words. The LEARN Project supported Chinese, Bengali, German, Spanish and Greek as L2 languages and could translate into multiple languages in the same text. 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. By their measures, the “word experts” they used chose the correct translation 90% of the time.

In designing this application of “word experts”, Berleant and his colleagues realized and addressed many of the same ideas, questions and issues encountered by ALOE. The following excerpt shows that they realized the potential of using arbitrary English text as L2 language learning material:

“

1. LEARN can circumvent the difficulty that many people have in scheduling study time since no study time needs to be scheduled when using LEARN, if it is used to process text the user would want to read even if it was presented completely in English.
2. LEARN places foreign words in meaningful contexts of interest to the reader. As Miller and Gildea point out, ‘The key is to see words in intelligible contexts’, and ‘to facilitate vocabulary growth . . . read as much as possible’

”

And they were also aware of the drawbacks of the approach. The LEARN Project saw that using unrestricted text would make their technique broadly applicable but would also possibly cause weak results on specific problems and domains as compared to the use of vetted teaching materials. The question of how to handle mixing of two languages syntax also arose. They chose not to attempt translating English syntax because they felt it was not necessary for learning vocabulary, it would drastically increase the complexity of the software to support grammars for every language and there is no concept of correct or incorrect syntax when mixing two languages on a word by word basis.

Beyond just translating specific words, The LEARN Project had a number of additional features to help learners acquire vocabulary. Translated words were highlighted in a different color and learners could optionally have the system display the L1 word in brackets next to the translation or view all the L1 words translated in a given text on a separate screen. There was also support for a guessing game for each of the translated words. The game involved learners being shown the first letter of the L1 word and allowed users to guess the next letter, and keep guessing until they got the whole word (the system would show the correct letter when they guessed wrong). There was also an implementation of character transliteration for Greek instead of word translation to help teach foreign character sets. In order to motivate the user, the system would translate the most frequent L1 words first and then translate a few more every time it was run and give feedback to the user on how many words they had gotten so far. Users also had the option of selecting how many words they wanted translated. It is also interesting to note that in their future work they saw the need for more extensive interactive hint facilities and a browser based implementation like many of the previously discussed projects realized in one form or another years later.

Graded-reader is also a standalone application but with a very restricted domain. Tauber designed it as a Greek learning aid for reading the Greek version of the New Testament of the Bible. It presents the New Testament in English with certain words or phrases translated into Greek. What makes graded-reader most 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, the New Testament in Greek in his case. Tauber shows 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. Tauber then sets up the ordering of vocabulary as a Travelling Salesman Problem and uses simulated annealing to find good orderings.

There are a couple drawbacks to graded-reader. Its ability to perform substitutions on the clause level in an intelligent way, *e.g.* not making a mess of the word ordering, is dependent on the manual tagging of grammatical, semantic and discourse related features of the Greek New Testament made available by the Opentext.org project and Tauber's own manually tagging of the English text used. The current state of graded-reader is a set of python scripts that take a corpus and the necessary tagging data and spit out an optimized word ordering. There is no learner friendly viewer that automatically translates words. But from the name Tauber chose, it is possible to see that he envisions the technique being more broadly applicable to the task of automatically creating graded language readers.

3 – Interviews

After coming up with the idea for ALOE, we contacted language professors and instructors at the University of Toronto to setup interviews so that we could get a sense for the state of the art in language education. Using the Web sites for the French, German, Italian and Spanish departments, we searched for and emailed all of the professors we could identify with experience in language learning and technology. We also emailed a large group of French instructors to request interviews. From those who responded, we were able to do six interviews. The table below lists demographic information for each of the interviewees. Five of the interviews took place at various locations on the university campus and one interview was done by email. For all of the interviews, the interviewing researcher took notes and three of the interviews were audio recorded. In the interviews, we focused on exploring what kind of teaching challenges the educators faced, how they used technology to aid learning and how our idea for ALOE might fit in the language learning environment.

Table 1: Language educator interviewee demographics.

<i>ID</i>	<i>Interviewee</i>	<i>Age</i>	<i>Sex</i>	<i>Languages Taught</i>	<i>Years of Experience</i>	<i>Level of Students Taught</i>
A	Professor Emeritus	70	M	French, English	50	all
B	Full Professor	55	M	Italian	32	all
C	PhD Student	27	F	French	2	beginner, intermediate
D	PhD Student	29	F	French, English, Romanian	6	all
E	PhD Student	35	M	French, English, German, Spanish	15	all
F	PhD Student	31	F	French, English	6	all

Throughout all our interviews, there were a few overarching themes. All the educators emphasized the importance of motivating and challenging students. Instructor C expressed that the “biggest challenge by far is being able to encourage all individuals in a classroom to participate and be willing to learn the language”. She also saw motivating students outside of the classroom as one of the big challenges and found that “giving them a structured but flexible environment is the best way to go. It is all about offering the material in a constructive, yet pleasurable way as to encourage and motivate them to keep on learning”. Both instructor C and

professor A motivated students by letting them choose and study the second language texts that most interested them.

Learning in context was another common theme. The practice of reading authentic L2 material is the simplest technique employed. Based on his experience teaching in France as a youth, professor A argued that living in a country and being immersed in its culture is the fastest and most efficient way to learn a language. Professor B went so far as to disparage a language learning software package because their depiction of a café was too generic and lacked cappuccinos and architectural elements common to Italian cafés and was therefore not an authentic context.

In general, the educators take a very measured approach to integrating technology into their classrooms. None of them were very fond of the automatic translation services that are freely available online and most of them had stories of how these services had failed miserably on one occasion or another. In fact, professor A believed that good automatic translation is beyond the capabilities of computers. At the same time, most of the educators had integrated some form of online second language resources into their curriculum and were very appreciative of the Web for providing easy access to authentic second language material.

As a professor who used to develop language learning software, professor B's views are particularly interesting. Besides his belief in the importance of culturally authentic contexts, he has found that most language learning software is pedagogically unsound due to a lack of specific feedback. What he means by this, is that when a student gets something wrong, the software should always strive to provide specific feedback that addresses the underlying reason why the student made the error. Basically, it should imitate as much as possible the knowledge and feedback that an experienced language teacher provides. Professor B described his iterative development of a HyperCard based system that could respond to all the different possible answers students gave for each question. Whenever students gave new answers professor B had not anticipated, he would update the system to handle the new cases. In this way, he was able to embed his years of teaching experience into the software.

Professor B also found time on task to be an important principle of language learning. And went so far as to say, "It doesn't matter if you cheat, I mean.. If you cheat or whatever it is, it's part of studying". He was very clear that it matters more that students are studying and motivated to study and not so much how they are studying.

We also asked two of the younger instructors try out an early version of the ALOE prototype on their personal computers for a few days. The version of ALOE they used translated approximately 8,000 of the highest frequency English words as opposed to the smaller manually chosen set of words eventually used in our study. In using ALOE, both instructor C and D thought it could be useful for students but

tempered their words of praise with concerns about the translation process and L2 grammar errors. Since ALOE does not take context into account when translating words, instructors C and D both encountered a number of bad translations while browsing the Web. Instructor D was particularly concerned that beginner L2 students would not realize the translations were wrong and would learn the wrong words. Instructor C was also concerned that when consecutive words were translated, the word ordering remained as it was in English which is not always correct in French. Instructor D suggested that translating whole sentences might solve some of the problems.

4 – Prototype

Design Goals

In our design of ALOE, we have concentrated on two sets of design goals, usability and pedagogical validity. The pedagogical goals focus our efforts on designing an effective learning system while the usability goals require us to design a system that is easy and enjoyable for people to use.

Pedagogical Goals

Drawing on research from second language learning and CALL, we chose pedagogical design goals that fit within our overarching idea of providing a convenient and unintrusive language learning environment on top of users existing web browsing activities.

Because we anticipate users will most often be concentrated on their Web browsing rather than actively using ALOE for language learning, it is important that ALOE allows for passive or incidental learning. Previous research suggests that providing frequent exposure to L2 vocabulary in rich contexts is the best way to enable incidental learning (Schmitt & McCarthy 1997, Groot 2000, Paribakht & Wesche 2000).

At the same time, users should be able to actively improve their L2 vocabulary knowledge. Thus, the system should test their knowledge, provide feedback and make it easy to access additional learning resources.

It is also important to consider how the content of Web pages can affect learning. Because anyone can create web pages, there is no guarantee that the language on the web is correct. Thus, our design must be able to provide valid L2 instruction in the presence of invalid L1 content such as improper grammar and misspellings.

Usability Goals

All language learning functionality provided by ALOE should be easily and quickly accessible by the user. This will make it easier and less of a hassle for users to switch back and forth between browsing and language learning activities.

All functionality should make use of existing Web affordances so that users are accustomed to and familiar with the modes of interaction. This will make it

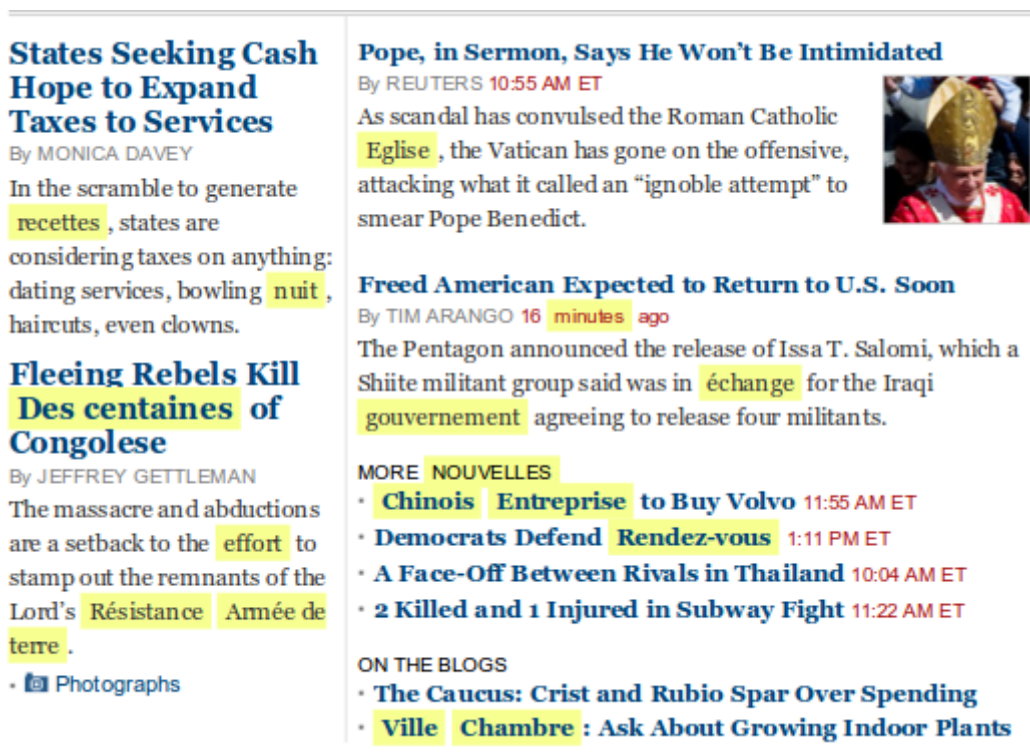
easier for users to learn how to use ALOE.

Because the only way to augment a Web page and provide language learning functionality is to modify the structure and/or text of the page, it is important that ALOE does not adversely affect users existing web browsing activities. This is a broad goal encompassing the following sub goals:

3. ALOE should be able to run on as many web sites as possible without slowing the site or breaking any of the sites existing functionality or design.
4. Users should be able to disable ALOE at any time.
5. The way in which ALOE functionality presents itself to the user on a web page should be visible but not distracting.
6. Special care should be taken to prevent ALOE from affecting user activities on sensitive or important sites such as online banking sites.

Prototype System

The ALOE prototype runs in a user's Web browser and translates words from English to French on Web pages as a user views the page. But ALOE is not a normal translator, it typically only translates a couple words in each sentence on a page and leaves the majority of the text in English. The idea behind this approach is that users will have a rich English context from which they can infer the meanings of the French words.



The screenshot shows a grid of news articles from the New York Times website. The text is in English, but several words are highlighted in yellow to indicate they have been translated into French. The articles include:

- States Seeking Cash Hope to Expand Taxes to Services** (By MONICA DAVEY): "In the scramble to generate recettes, states are considering taxes on anything: dating services, bowling nuit, haircuts, even clowns."
- Fleeing Rebels Kill Des centaines of Congolese** (By JEFFREY GETTLEMAN): "The massacre and abductions are a setback to the effort to stamp out the remnants of the Lord's Résistance Armée de terre."
- Pope, in Sermon, Says He Won't Be Intimidated** (By REUTERS 10:55 AM ET): "As scandal has convulsed the Roman Catholic Eglise, the Vatican has gone on the offensive, attacking what it called an 'ignoble attempt' to smear Pope Benedict." (Includes a small photo of Pope Benedict XVI)
- Freed American Expected to Return to U.S. Soon** (By TIM ARANGO 16 minutes ago): "The Pentagon announced the release of Issa T. Salomi, which a Shiite militant group said was in échange for the Iraqi gouvernement agreeing to release four militants."

Below the main articles, there are sections for "MORE NOUVELLES" (listing items like "Chinois Entreprise to Buy Volvo", "Democrats Defend Rendez-vous", "A Face-Off Between Rivals in Thailand", "2 Killed and 1 Injured in Subway Fight") and "ON THE BLOGS" (listing items like "The Caucus: Crist and Rubio Spar Over Spending", "Ville Chambre : Ask About Growing Indoor Plants").

Figure 2: A page from the New York Times Web site translated by ALOE. The French words appear in a yellow highlight.

Our hope is that ALOE creates a rich environment that supports incidental learning and guessing from context. There is widespread agreement that most L2 vocabulary acquisition occurs incidentally (Segler 2001, Chun & Plass 1996, Rott 1999, Gu & Johnson 1996, Nation 1990, Sternberg 1987) and substantial evidence that repeated exposure with minimal mental processing by the learner “can have quite impressive facilitative memory effects” (Segler 2001). Some researchers have pointed out that guessing from context is a challenging and often unsuccessful strategy for learners because the context is not rich enough or the learner can not fully understand the L2 context (Groot 2000, Huckin & Coady 1999, Duquette *et al.* 1998, Schmitt & McCarthy 1997). We believe that using English, our learners native language, provides a rich context that is easy for learners to understand and use to infer the meanings of L2 words. On most pages, ALOE leaves enough English text to satisfy Laufer's estimate that learners need to be able to understand 95% of the context to enable effective guessing from context (Laufer 1988).

The prototype translates a specific set of words chosen by the researchers. The set is composed of 1,758 hand-picked words. We selected the words based on two criteria: frequency and ambiguity. We first chose high frequency English nouns and adjectives (as measured in the British National Corpus) to ensure that users would encounter the words frequently no matter what Web sites they browsed. We then added the plural forms of the chosen nouns in hopes of having users encounter the nouns more often. Finally, we filtered the words to remove the most ambiguous words that are often found in different contexts with different meanings. The filtering was done in two steps. A script first removed words that showed up in the British National Corpus tagged with more than one part of speech and then we manually went through the rest of the words to prune additional words we considered ambiguous. The final set of 1,758 words included 814 singular nouns, 653 plural nouns and 291 adjectives.

Once the set of words was finalized, we automatically obtained translations for all of the words using Google's free translation service. We chose to use Google Translate instead of a simple English-French dictionary lookup based on the premise that Google Translate is better tailored to serve as cross-language dictionary for the Web because of the way it works. Traditional cross-language dictionaries and their online counterparts are based on centuries of research on how to best translate books, newspapers and other long established mediums. In contrast, Google Translate is statistically trained using the Web and traditional media as input (Helft 2010) and thus will likely perform better in the context of the Web where ALOE is translating. That said, Google Translate is not perfect and can often return inappropriate translations. Thus, after automatically translating with Google Translate, we manually went through all the translations and cross-checked many of them and changed some of them based on translations from the WordReference.com online language dictionaries.

Our focus on the use of nouns and adjectives, filtering of ambiguous words, and checking of translations was motivated by findings from pilot tests of ALOE.

Because participants (especially the two educators we interviewed) complained strongly about incorrect translations, we decided it was very important to prevent ALOE from causing users to learn incorrect things about a second language. Thus, we stayed away from translating verbs and other parts of speech that are highly dependent on context for correct translations. Nouns and adjectives are also sensitive to context but in French the context usually only changes the gender of the word which most often only causes the word to have an 'e' postfixed. Additionally, French adjectives most often appear after the noun but ALOE does not change word ordering when translating. French was chosen primarily because it is so similar to English and would therefore cause less problems being intermixed with English. Being in bilingual Canada where most people learn some amount of French in school also helped make recruiting participants easier. By recruiting participants with some amount of French experience, we hoped to ensure that they would know how gender affects nouns and adjectives and where to position adjectives and so would not be adversely affected by seeing the wrong gender and adjective word order appear in some cases.



Figure 3: A Wikipedia page translated by ALOE with the default dashed underline highlighting of French words. The brown rectangle on the right side of the page is the ALOE control panel.

Our choice to not use context when translating words causes many problems but they are outweighed by the huge gain in usability. Early versions of ALOE incorporated context in a variety of ways, from translating whole phrases and sentences to translating single words with adjacent words for context. All of these methods caused ALOE to run at an unusably slow pace. The version of ALOE used in our study translates most pages in under two seconds using aggressive caching. The earlier context dependent versions of ALOE took between 1 and 2 minutes to

translate a page because it is not possible to cache the unlimited contexts in which words can appear. One version of ALOE even tried a hybrid approach that first translated single words quickly and then phrases. People who pilot tested this approach would often turn off the phrase translation because it was slow and by the time a phrase was translated they had either already read that part of the text, or were in the middle of reading a phrase when it translated, or had already left the page before any phrases were translated.

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 which is linked to from the control panel that is loaded on the top right of a page ALOE is running on. On the settings page, users can choose any combination of color and style (background, underline, border) and line type for underlines and borders (dashed, dotted, solid) to highlight translated words.

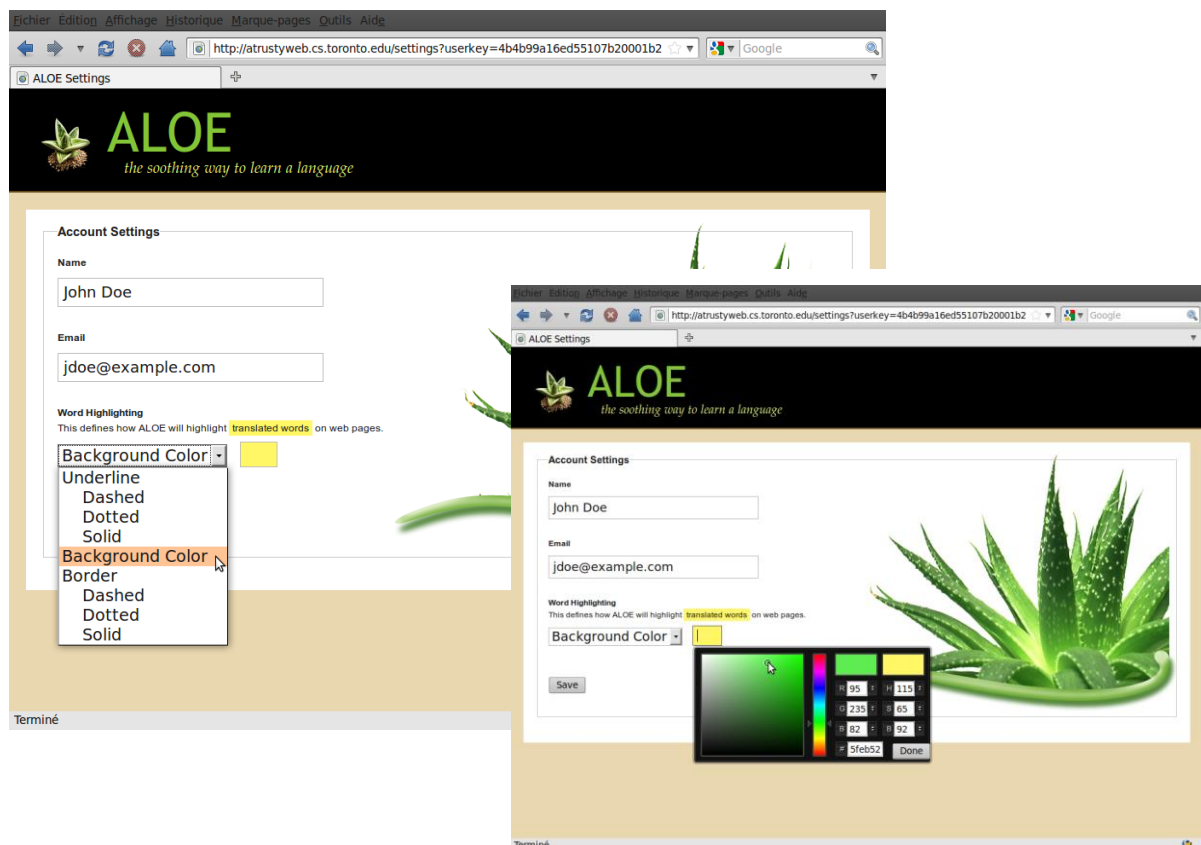


Figure 4: The ALOE interface for changing the translated word highlight style.

After a word is translated, the user can hover their mouse over the word to view a multiple choice question which asks what the French word means in English

and gives three English words as choices. The two incorrect choices are generated by 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 they were right or wrong and reveals what the correct translation is. ALOE also provides a link to display the full definition of the French word using WordReference.com which provides a lot of additional information about the word including a longer definition with synonyms and example phrases.

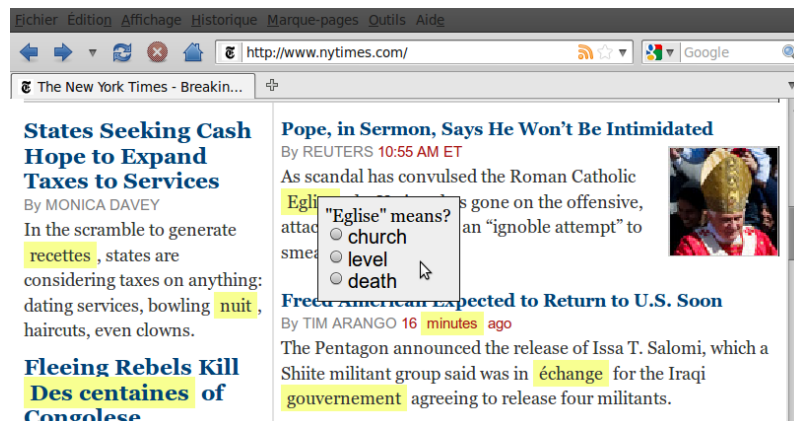


Figure 5: The multiple choice question display appears when a user hovers their mouse over a translated word.

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 promoting learning than simple glosses with just the translation (Nagata 1999). The idea is that multiple choice questions challenge users and make them spend more time processing the word and encourage them to attempt to guess the translation from the English context. There is the added benefit that word meanings inferred from guessing are retained better than meanings provided by simple glosses (Hulstijn 1996, Hulstijn 1992).

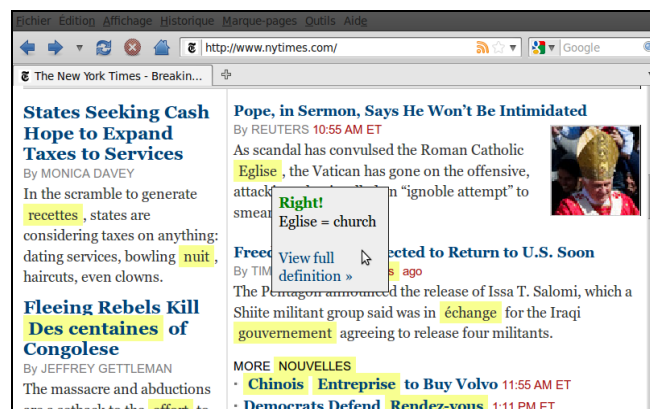


Figure 6: After answering the multiple choice question, the user is told the correct answer and is given the option to view the full definition from WordReference.com.

To prevent users from making mistakes on important and sensitive Web sites, ALOE does not run on secure (https) Web sites and allows users to disable ALOE at any time using a control panel that is loaded on the top right of every page that ALOE runs on. We provided three different disable options to give users flexibility in controlling ALOE. 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.

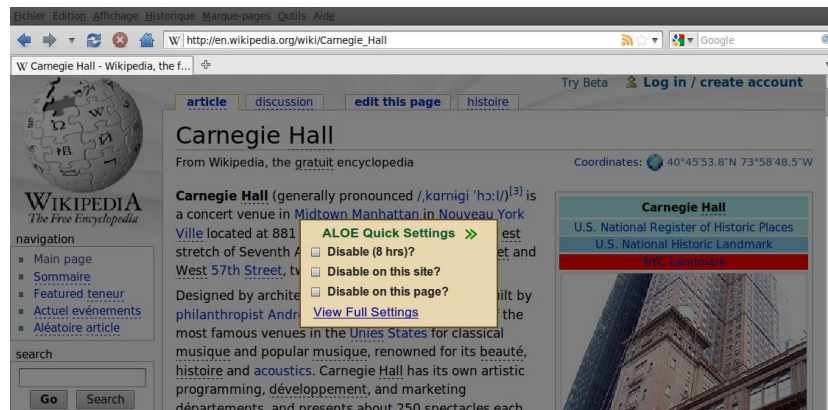


Figure 7: The ALOE control panel appears in the center of the page when a user clicks on it. The control panel provides the three disable options and a link to the page where users can change their highlight settings.

Implementation

We designed ALOE as a client-server application. We developed the server as a webpy⁴ 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 batch translation requests for quicker response times because Google translate only translates 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 extension⁵. Packaging the script enables it to run automatically on Web pages without requiring users to install Greasemonkey⁶.

A user script is just a piece of Javascript that a user chooses to run on a Web page in order to enhance the page. Developing the client as a user script has advantages and disadvantages. The primary advantage was that development time was reduced because no Firefox specific extension APIs had to be used in developing ALOE. Although we concentrated on using Firefox due to its popularity and speed, user scripts are easily portable across different browsers.

The disadvantages of user scripts are that they do not integrate as well with the browser's user interface. Furthermore, user scripts cannot take advantage of browser specific extension APIs that could increase performance and offer new functionality. For example, this is why the ALOE control panel had to be loaded inside of pages as opposed to in a toolbar menu.

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. In the case of our study, it looks for the set of high frequency English words that we manually selected. 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. The sequence diagram in Appendix E shows in detail the steps that occur when ALOE translates a page.

⁴ <http://webpy.org/>

⁵ Using Anthony Lieuallen's Greasemonkey script compiler from:
<http://arantius.com/misc/greasemonkey/script-compiler>

⁶ <http://www.greasespot.net/>

5 – 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. 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 six hundred questions, these measures were used to better 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. See Appendix C for an example page from the quiz.

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. See Appendix D for the interview template we used. 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 the University of Toronto 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 they did not complete the second quiz. No data for these three participants is included in our analysis. And there was one participant who did not complete the third quiz but whose other data is 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; MIN 0.5; MAX 12.5). 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). See Appendix B for a table listing the demographics for each participant.

6 – Results

In this chapter, we report the results of our study. We first present how much time the participants used ALOE in total and on a day to day basis. We then describe how the participants interacted with ALOE. Next, we analyze the quiz results to see how ALOE affected participants L2 vocabulary knowledge. Finally, we discuss the participants' perceptions of ALOE based on the results of our interviews.

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 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 7 below 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; MIN 2.3; MAX 239.3) and disabled hours (AVG 21.6; SD 31.7; MIN 0; MAX 112.2) before quiz two as well as after (active: AVG 36.9; SD 36.1; MIN 1.7; MAX 98.3) (disabled: AVG 36.6; SD 65.7; MIN 0.1; MAX 153.7).

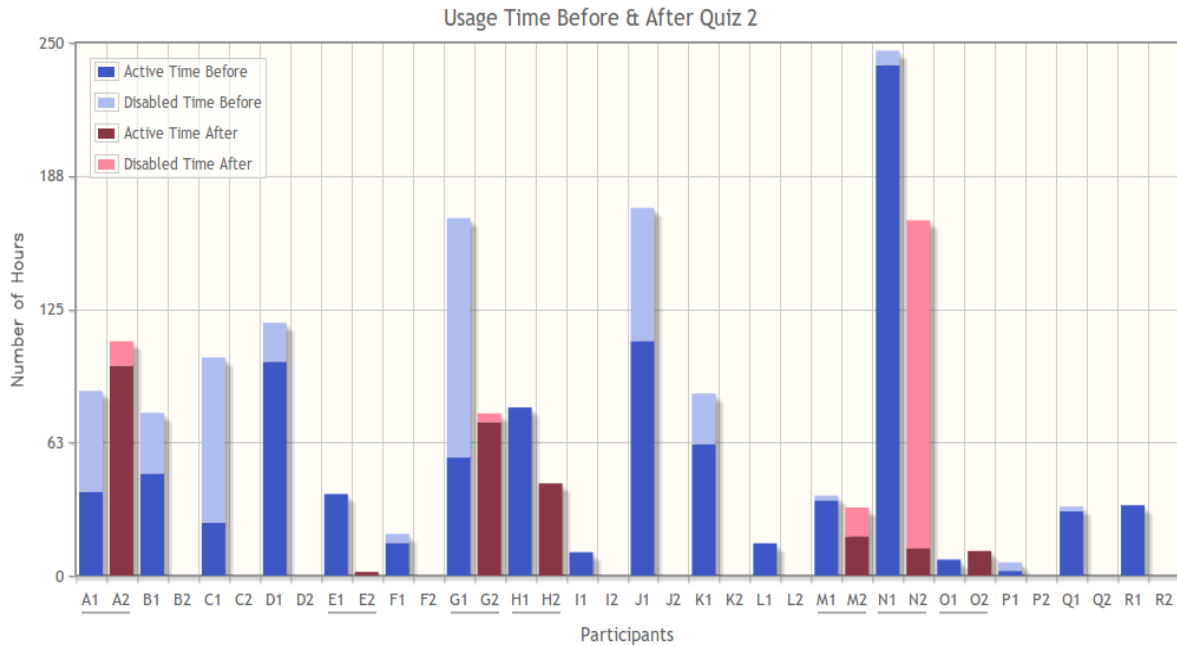


Figure 7: Total hours of usage by each participant broken into the times when ALOE was active and disabled. The seven participants who used ALOE for a second month are underlined and have a second bar showing their second month totals.

For the most part, participants time logs lined up well with their self-reported estimates of web browsing time (see Appendix B). 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. And 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. The 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 also examined how the participants with the most disabled time accumulated this time. Participants G and J accumulated most of their disabled time by disabling some of the domains they most frequently visited. Most of participants B, K and N's disabled time resulted from using the global disable. And participants A and C's times resulted from both global disables and the disabling of frequently visited domains.

We can further break the participants' active and disabled time down to their day to day usage. Figure 8 below shows the active and disabled time across all participants for their first 31 days. For each day, the yellow bars stretch one

standard deviation above and below the average which is shown by a small horizontal black line embedded in each yellow bar. The thin red bars above the yellow bars stretch to the maximum time amount and where visible, the thin red bars below the yellow bars stretch to the minimum.

All Participants Day to Day Usage

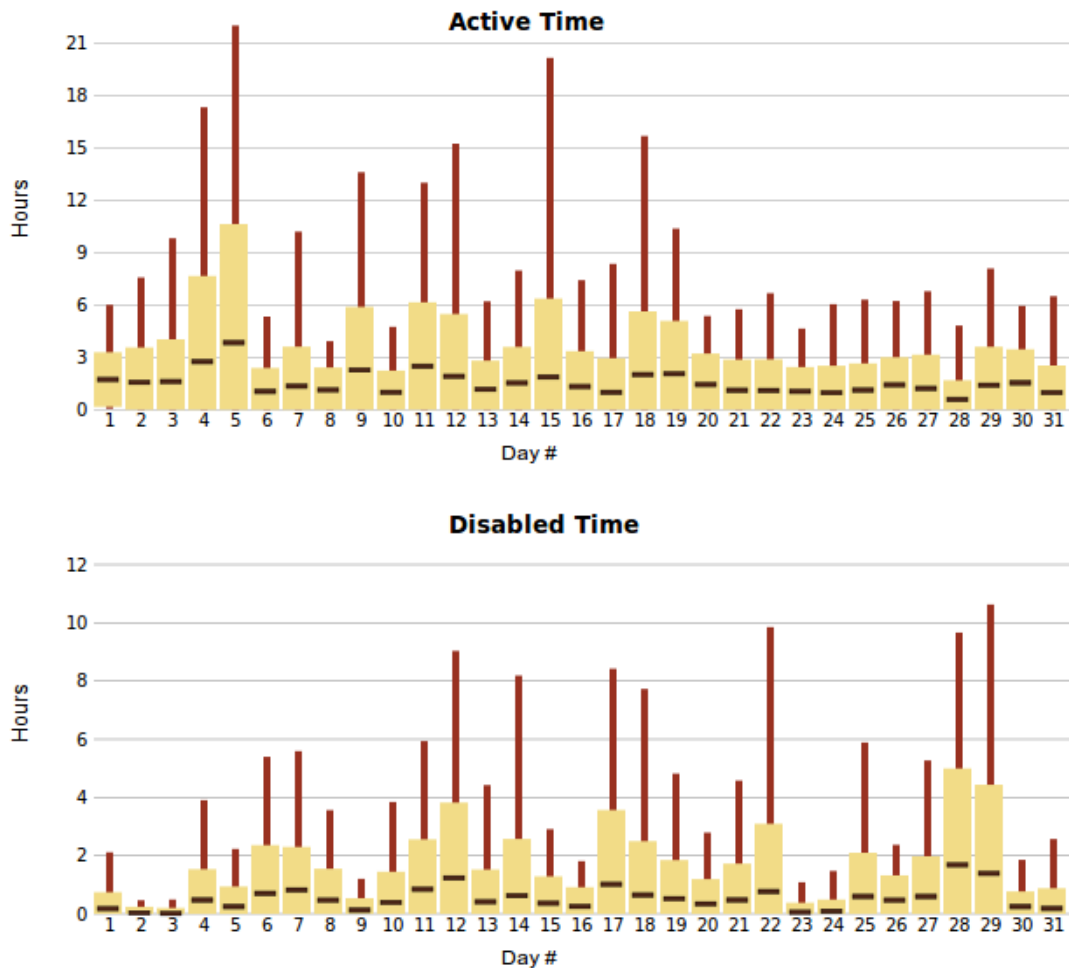


Figure 8: The day to day active and disabled time of participants using ALOE for the first month.

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 periodically recurring ups and downs 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 Figure 9 below are the same active and disabled day to day time graphs for the group of seven participants who chose to continue using ALOE for an extra month. In comparing the second month to the first month for these seven participants, we find 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.

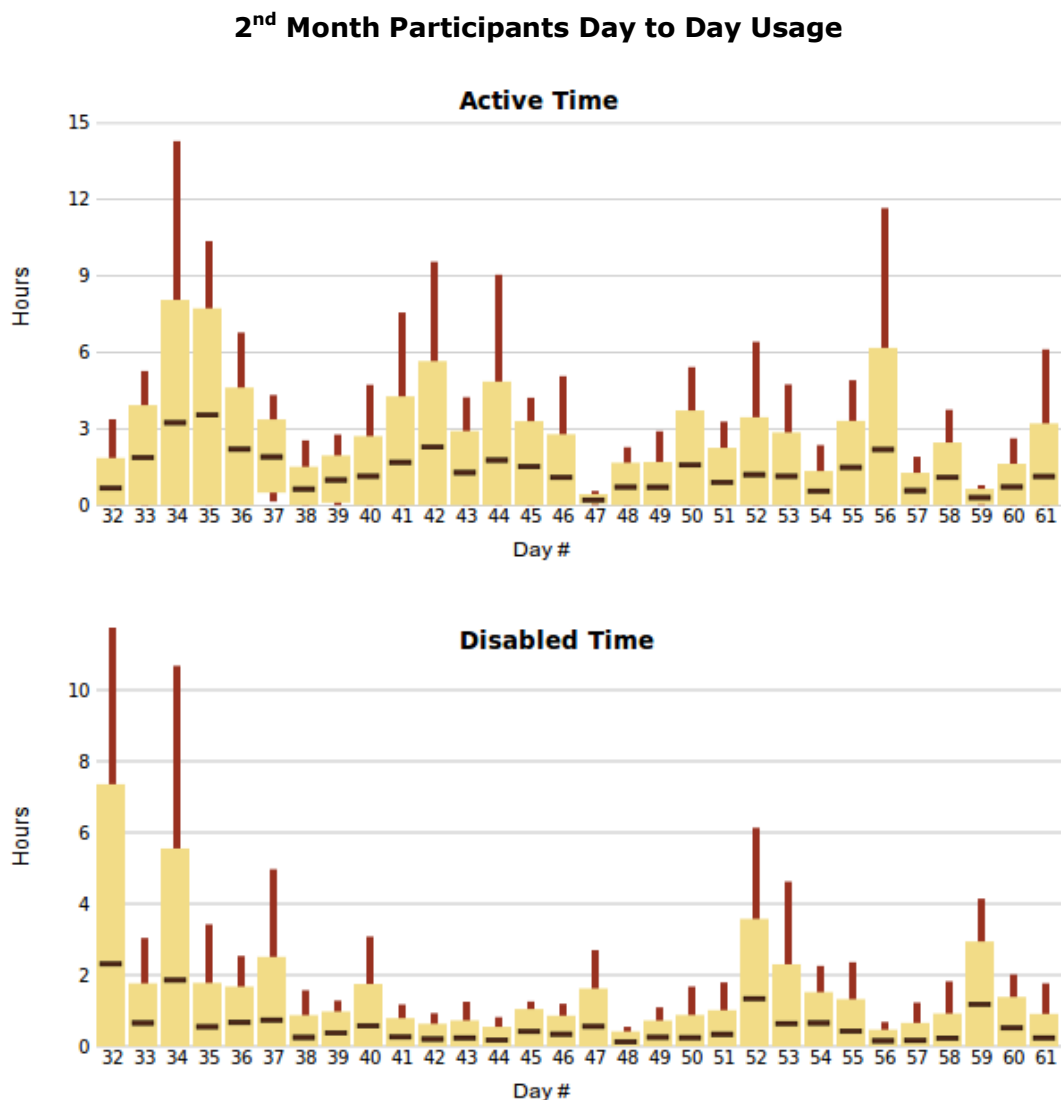


Figure 9: The day to day active and disabled time of the seven participants using ALOE for the second month.

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. The Web site-specific issues occurred on Facebook, YouTube and dictionary sites. On Facebook, ALOE caused a strange rendering glitch that made text on Facebook disappear briefly when ALOE inserted the French translations. On YouTube, the participant felt that ALOE slowed down the pages too much. And on dictionary sites, the participant found that getting definitions with French in them was contrary to what she was trying to accomplish. The participant who cited their mood explained that they disabled ALOE when they were tired or annoyed by ALOE slowing down their reading.

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 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.

In particular, participants B, I and Q, who chose to stop using ALOE after the first month did so due to the slow down of their Web browsing. Participant B reported that he was going to be very busy with work in the second month and so did not want ALOE to impede his work. Participant I stopped because she felt ALOE slowed down some of the sites she visited. The interview with her revealed that she rarely disabled ALOE and did not fully understand the disable options. After the interview, she expressed that she would have continued using ALOE if she had known that she could have sped up her browsing by disabling ALOE. As a heavy tab user, participant Q probably experienced the worst slow downs. By opening many tabs at once and keeping dozens of tabs open at a time, he often experienced intense slow downs and often even had his Web browser crash due to ALOE. And participant E chose to continue using ALOE for the second month even though she had the extension disabled by Firefox for all but four days of the second month. She disabled ALOE because "it seemed to slow down my computer" and "she kept forgetting to turn it back on again".

Interactions

Figure 10 below shows the total number of interactions recorded by ALOE for each participant during the same time periods as Figure 7 above. Participants hovering their mouse over translated words to view the multiple choice question made up the large majority of interactions. But the graph also shows that very few of the multiple choice interactions resulted in participants actually answering the multiple choice question. There are two explanations for this behavior. First, 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. Second, a few of the 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. The counts for viewing WordReference.com definitions are also included in the graph but since the

participants used this feature an average of 1.22 times (MAX 4) it is not visible.

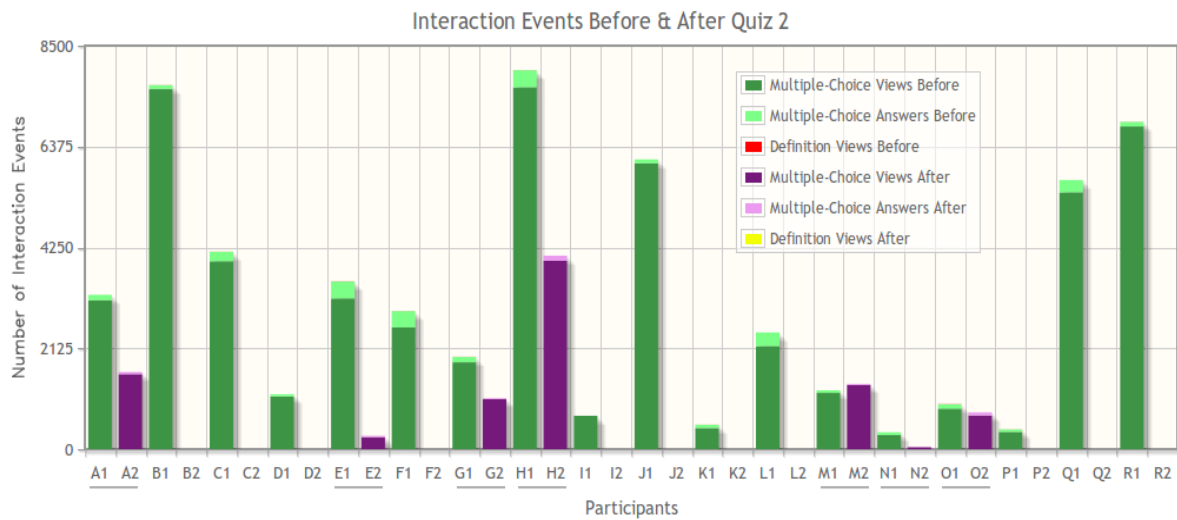


Figure 10: Total number of interaction events by each participant broken into the different types of interactions. The seven participants who used ALOE for a second month are underlined and have a second bar showing their second month totals.

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). Although 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 which is visualized in Figure 11 below. Each bar represents one of the five answers with the leftmost being very seldom and the rightmost being very often. The size of the bar and the number above it represent how many participants gave that answer. The mode is highlighted in bold.



Figure 11: Visualization of the 5-point Likert scale question asking how often participants stopped to think or interact with translations.

As we can see from the visualization, most participants felt that they thought about or interacted with words every now and then or often. 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 line up with the interactions per hour metric we previously calculated and can not really 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. And 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.

Figure 12 below shows the day to day interactions of all the participants. Both the amount of multiple choice views and the amount of multiple choice answers start off elevated and level off after a few days as can be expected from people playing with ALOE while it is new to them. And the amount of multiple choice answers in the last week is a bit depressed. 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.

All Participants Day to Day Interactions

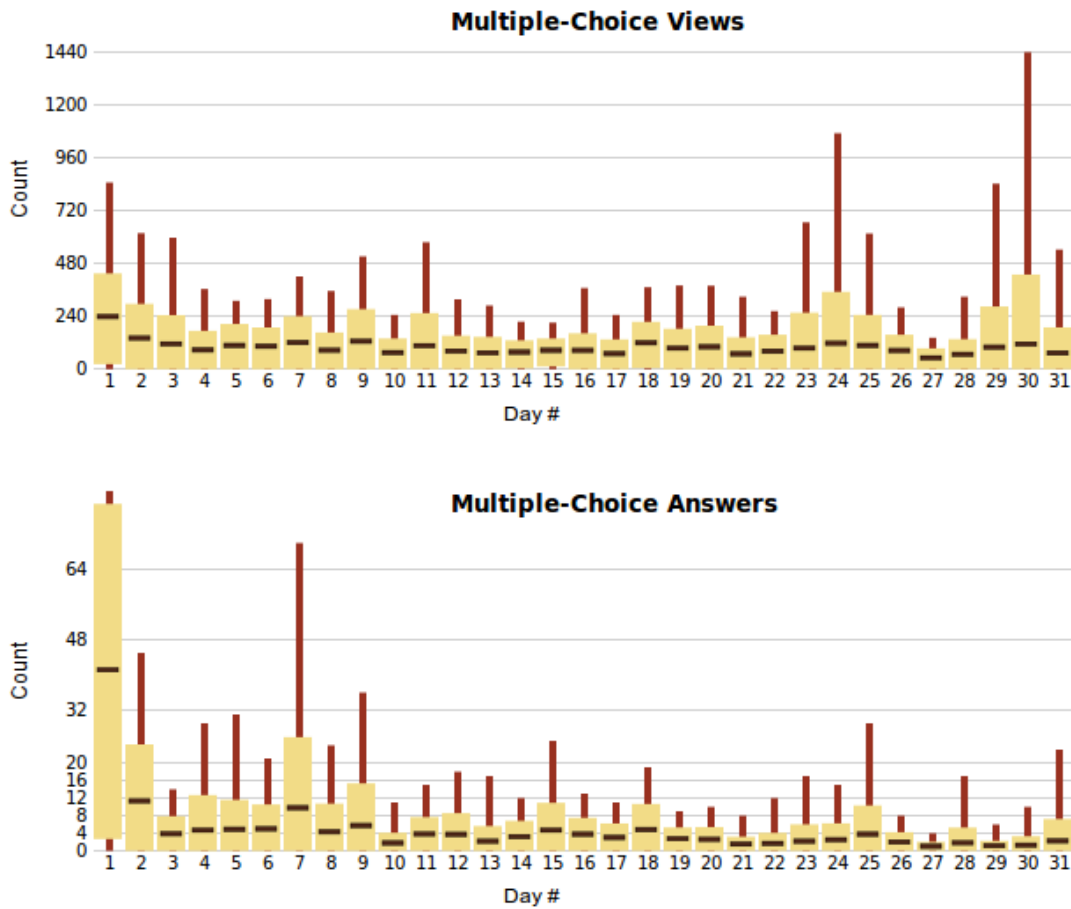


Figure 12: The day to day multiple-choice views and answers of participants using ALOE for the first month.

Looking at Figure 13 below for the group of seven participants who continued using ALOE for the second month, we see that 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).

2nd Month Participants Day to Day Interactions

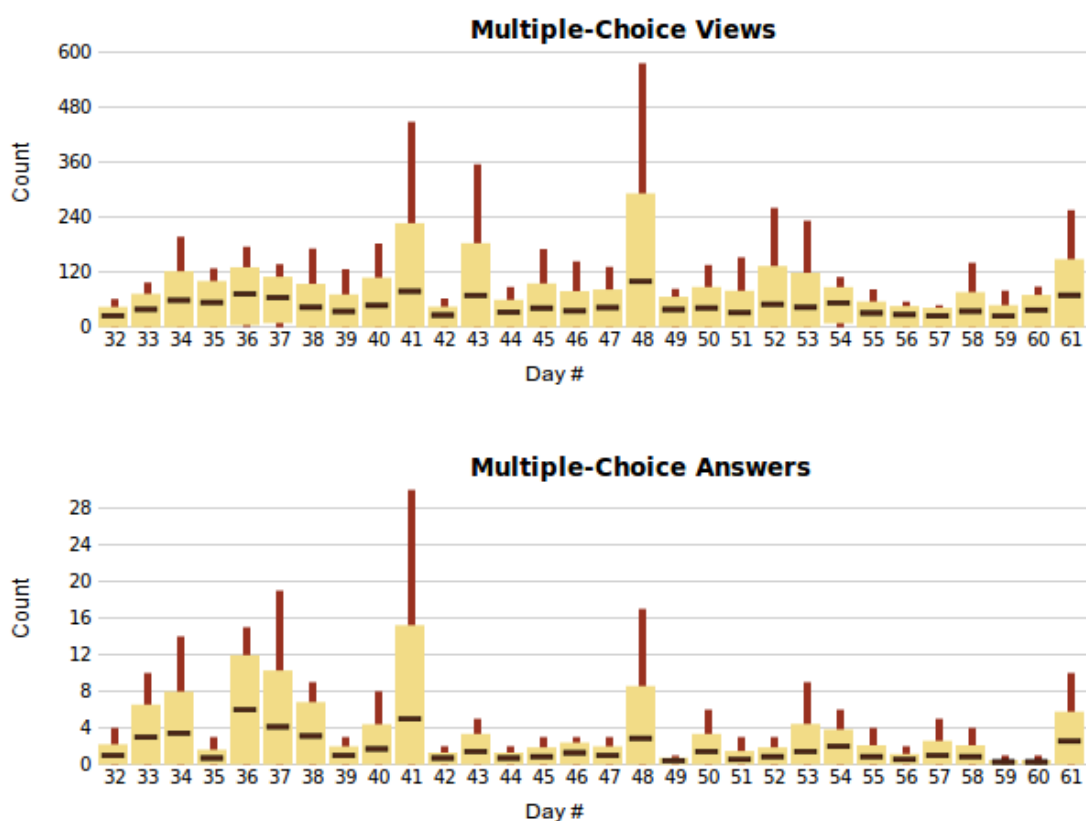


Figure 13: The multiple-choice views and answers of the seven participants using ALOE for the second month.

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 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.

Quizzes

The graph below 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 participants three quiz marks are grouped together such that the bar with the lightest colors represents the first quiz and the bar 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.

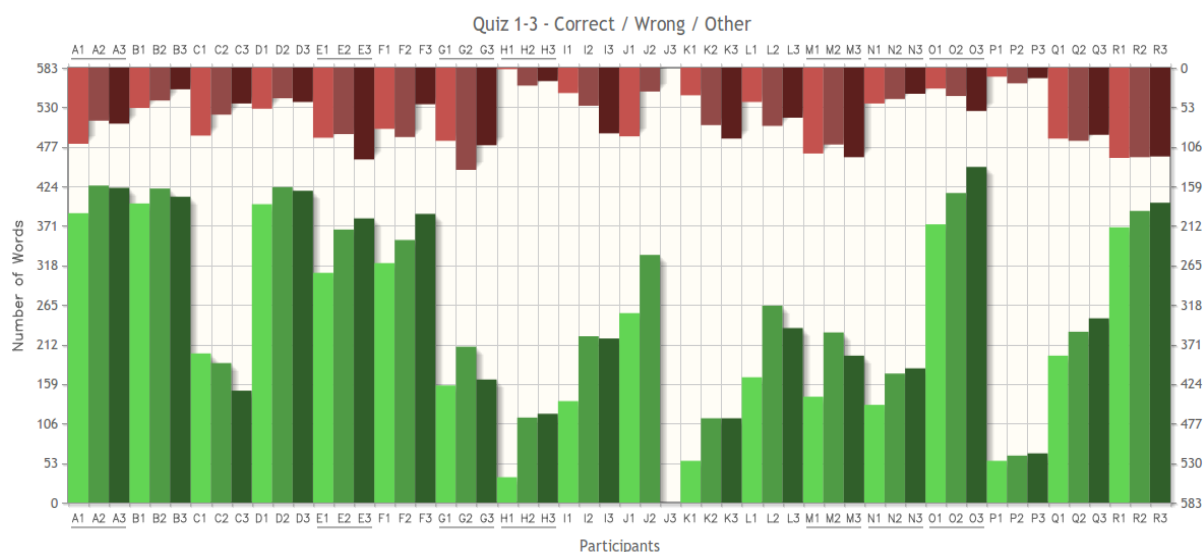


Figure 14: Quiz 1, 2 and 3 scores for all of the participants. The seven participants who used ALOE for a second month are underlined and participant J never completed quiz 3.

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; MIN 7; MAX 96) and 2.59 more words incorrectly (SD 24.47; MIN -60; MAX 40) 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; MIN -44; MAX 35) and 4.14 more words incorrectly (SD 20.41; MIN -33; MAX 34). The eleven participants who stopped using ALOE averaged 2 more words correct (SD 17.31; -30; MAX 35) and 2.89 fewer word incorrectly (SD 21.18; MIN -44; MAX 37) on quiz three. Using statistical methods, we will analyze the groups and their scores and determine if there are any statistically significant differences.

By using a t-test to compare the two groups, we found that there is no significant ($p > 0.05$) difference in how the groups amount of correct marks changed from quiz two to quiz three. We also ran a repeated measure ANOVA on both groups correct marks in all three quizzes. This found that there was a significant difference ($p < 1e-5$) between quiz one and quiz two and also between quiz one and quiz three but no significant difference ($p > 0.05$) between quiz two and quiz three. Therefore, there is no real difference between the group of participants who stopped using ALOE and those who continued using ALOE. And as a whole, the participants improved between quiz one and quiz two but neither improved nor did worse on quiz three.

What these statistics appear to imply is that the idea behind ALOE, that is learning L2 vocabulary by inserting L2 words into the context of L1 Web pages, is a viable vocabulary learning environment. Participants were able to learn new French vocabulary and retain it for a month with negligible loss of recall. The fact that an extra month of using ALOE did not significantly improve participants vocabulary is most likely related to the decrease in participant interactions during the second month. Because the participants encountered the same set of words over and over, even after they had already learned the words, they had far fewer opportunities to learn new words. We will see feedback from the participants corroborating this theory later when we present the interview results.

In an attempt to uncover what factors most influenced how much vocabulary a participant learned while using ALOE, we divided the participants into groups based on the quantitative and qualitative factors listed in the table below.

Table 2: The different groupings that were statistically tested in order to determine which factors affected the vocabulary acquisition of the participants.

Factor	Groups
French Level	< 250 correct on Quiz 1 ≥ 250 correct on Quiz 1
Active Time using ALOE before Quiz 2	< 20 hours 20-70 hours > 70 hours
Interaction Level before Quiz 2	< 1000 interactions 1001-5000 interactions > 5000 interactions
Enjoyment	1-3 on post-study interview question 4-5 on post-study interview question
Stopped to Think About or Interact with Translated Words	1-3 on post-study interview question 4-5 on post-study interview question

For each of these groupings, we ran a t-test or an ANOVA test (depending on whether there were 2 or 3 groups) comparing how each groups amount of correct marks changed from quiz one to quiz two. None of the tests showed any significant ($p < 0.05$) difference between any of the groupings. The French Level grouping does turn out to be significant ($p < 0.05$) if we do not consider the participant with least amount of active hours using ALOE (participant P with 2 hours total). Without participant P, the beginner French level group (< 250 correct on Quiz 1) learned significantly more vocabulary (AVG 66.5; SD 23.74) than the intermediate French level group (AVG 38.88; SD 20.23). This makes sense because the beginners start off seeing more unknown words and therefore have more learning opportunities than the participants with an intermediate level of French skill. The removal of participant P does not affect the outcomes of the tests on any of the groupings.

We also calculated point biserial correlations across the set of all words not known to the participants in quiz one. We wanted to see if how many times a participant might have seen a French word or viewed or answered the multiple choice question for a French word affected whether that word would appear in the set of words learned by the participant on quiz two. But our data presented no significant correlations ($p < 0.05$) for any of the three types of encounters. This seems to suggest that the multiple choice questions, which most participants reported as useful, were not the primary means of acquiring the new vocabulary words. Instead, participants were learning vocabulary either incidentally or by actively guessing the meanings from context.

Interviews

Using the interviews conducted after participants stopped using ALOE, we gathered a lot of feedback about the participants' experience with ALOE. We first asked how enjoyable they found ALOE for learning French vocabulary. Then we asked how they felt about the amount of translations and the quality of the translations. For all of these questions, we had the participants answer using a 5-

point Likert scale where a one would be the lower extremity and five the upper extremity of possible responses. The visualization below shows how the participants responded to these questions.

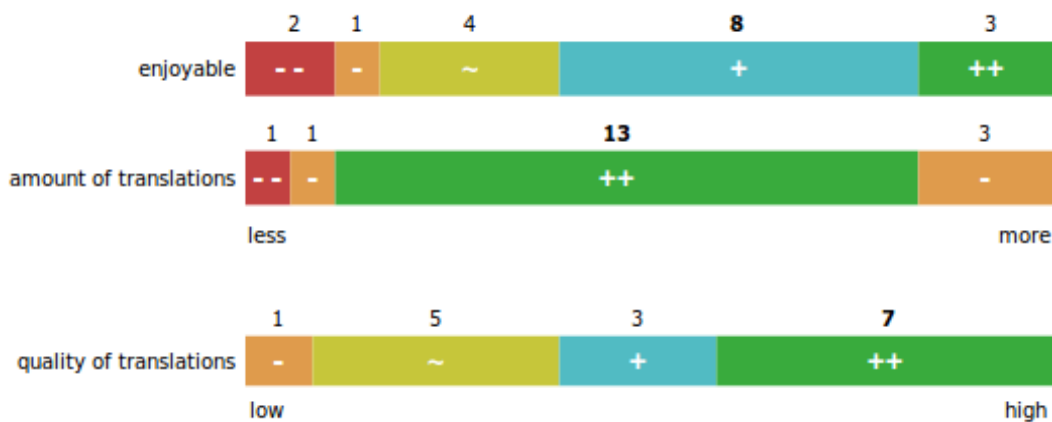


Figure 15: Visualizations of the 5-point Likert scale questions from the participant interviews.

The greatest number of participants reported that they found ALOE enjoyable or very enjoyable. One participant found ALOE unenjoyable and two found it very unenjoyable. Participants M, P and C were the three to rate ALOE as unenjoyable. Participant M reported that the passive contextual translations did not fit her learning style. She preferred more active methods of learning and felt that when Web browsing, “reading a [translated] word doesn’t really do much”. But according to her quiz two marks, she learned 86 new words after a month of using ALOE, the third highest amount out of all the participants. ALOE also frustrated her because it caused her to read more slowly with translations. Participant P, the participant with the least amount of active time using ALOE, reported that he did not find ALOE enjoyable because it slowed down the loading of Web pages and it took him a long time to read pages with translations because he did not know many French words. Participant C also found ALOE unenjoyable because it slowed down Web pages and caused the previously mentioned rendering glitch on Facebook.

The participants who rated enjoyability in the middle of the range shared some of the same reasons for their ratings as those who found ALOE unenjoyable. This includes the feeling that the translations slowed their reading of Web pages. One participant clearly explained this by calling the translations “speed bumps”. A couple participants got tired of seeing translations of words they already knew. One participant was even dismayed because of how little he had to think while reading to understand the translated words.

“I was able to figure out the [translated] word because I was able to read it in the context.”

It is interesting to note that the participants who rated enjoyability in the

upper range shared some of the same concerns as those from the middle range. But these concerns, including the slow downs, quality of translations and the repeating of learned words, were not as serious to the participants who found ALOE enjoyable. Those who enjoyed using ALOE felt like they were learning vocabulary and that the ALOE approach to learning in context was fun and easy to use.

The responses to the amount of translations indicate that a large majority of the participants felt there was an ideal amount of translated words. Three participants felt that there should have been more translations. One of these participants complained that too many of the words translated were almost identical to their English translation. Another was annoyed to see the same word translated every time it occurred on a page and felt that ALOE should only translate the first occurrence and then translate many different words on a page in this way. Of the two participants who felt that there should be less translations, one of them (participant P) felt he was slowed too much by the many translations. The other wanted to have ALOE stop translating the words he already knew.

Regarding the quality of translations, most of the participants (10/18) felt that they were high or very high. The participants who rated the translations as very high quality never or very rarely noticed out of context translations. The rest of the participants seemed to encounter out of context translations more often. Some examples that the participants gave included the translation of proper names into French ('Steve Jobs' became 'Steve Emplois'), the translation of two different English words to the same French word (both 'mad' and 'crazy' became 'fou') and translations that used the wrong meaning of an English word ('car performance' became 'car spectacle' which in French really means 'car show'). Based on their ratings, most of them were quite tolerant of out of place translations showing up. In fact, two participants turned it into a kind of game. When they saw an out of place French translation they took time to try and guess which English word had been incorrectly translated. Two of the beginner French level participants did not provide ratings for the translation quality because they felt they did not know enough French to provide an accurate rating.

In the interview, we also asked which features of ALOE the participants found most useful, what style of highlight they used for translated words and whether the participants would continue using ALOE if they could. Fourteen of the participants cited the hover over multiple choice pop ups as one of the most useful features. Six mentioned the way ALOE translates words in context. Two mentioned how ALOE highlighted translated words and one found the three quizzes to be very useful.

For the highlight style of translated words, fourteen of the participants used the default dashed underline for the whole study. Two chose to highlight translated words with a blue background. Another two used a teal and yellow background respectively but changed back to the default dashed underline for the majority of the study.

Finally, fourteen of the participants said they would continue using ALOE as-is. Two would want to use it if it could be improved and two said they would not continue using ALOE. The two who would not continue using ALOE, participants L and P, were the participants with the highest and lowest amount of learned words respectively. They both stated that they were not currently looking to improve their French and therefore did not feel like investing time in using ALOE. The two who asked for improvements, wanted ALOE to work without slowing down Web pages. One of these participants also wanted a temporary one-time disable option, English translations with more than one word (*e.g.* a couple of synonyms) and sentence and phrase translations.

Some of the other participants also put forth ideas for improvements to ALOE. Two participants suggested not translating words that have the same meaning and nearly the same spelling in French and English. Another two participants wanted the ability to listen to the French words pronounced. At least three participants requested some form of adaptation where the participant could specify what words they had learned or ALOE could detect which words were learned and then stop translating those words and start translating new words. One participant requested the ability to control the density of translations on a page and the ability to instantly switch all the words back to English without reloading the page (the current disable options require reloading to take effect on the current page). And another participant wanted only the first few occurrences of a word to be translated in order to reduce clutter and also wanted translated words to turn back into English once he had answered the multiple choice question correctly.

Conclusions & Future Work

In this thesis, we have demonstrated the effectiveness of learning L2 vocabulary in the context of L1 Web pages. Using the ALOE prototype, 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. Most of the participants found using ALOE an enjoyable way to learn L2 vocabulary and wanted to continue using it after the study. In particular, we have shown that inserting L2 translations in L1 text without regard to context, while not ideal, does work effectively for vocabulary learning. And the density of translations provided by ALOE provided an almost perfect balance between L1 and L2 words that allowed for both incidental learning and effective guessing from context by the participants.

Although the ALOE prototype was well received by the study participants, almost all of the participants saw room for improving it. We designed and built ALOE with speed and universal Web site compatibility in mind but it still noticeably slowed down the loading of some Web pages and caused problems on a few sites. It was also apparent that some of the participants had different strategies for using ALOE and would have benefited from more user control and customizability over what words were translated, when translation should occur and what type of additional language learning resources should be made available (*e.g.* audio pronunciations, longer L2 definitions, other exercises or quizzes to test vocabulary knowledge). 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 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.

Bibliography

- J. Beaudin, S. Intille, E. Munguia Tapia, R. Rockinson, and M. Morris, "Context-Sensitive Microlearning of Foreign Language Vocabulary on a Mobile Device," *Ambient Intelligence*, 2007, p. 55-72.
- D. Berleant, S. Lovelady, and K. Viswanathan, "A foreign vocabulary learning aid for the networked world of tomorrow: the LEARN project," *ACM SIGICE Bulletin*, vol. 19, 1994, p. 29.
- N. Cavus and D. Ibrahim, "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.
- C. Chen, S. Hsu, Y. Li, and C. Peng, "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.
- D.M. Chun and J.L. Plass. Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80:183-198, 1996.
- L. Duquette, D. Renié, and M. Laurier. The evaluation of vocabulary acquisition when learning French as a second language in a multimedia environment. *Computer Assisted Language Learning*, 11:1:3-34, 1998.
- S. Fallahkhair, L. Pemberton, and R. Griffiths, "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.
- W.H. Fletcher, "Concordancing the Web: Promise and Problems, Tools and Techniques," *Corpus linguistics and the Web*, 2007, p. 25-45.
- G.L. Friedman, "Learner-created lexical databases using web-based source material," *ELT Journal*, vol. 63, 2008, pp. 126-136.
- J. Gamper and J. Knapp, "A Review of Intelligent CALL Systems," *Computer Assisted Language Learning*, vol. 15, 2002, pp. 329-342.
- P.J.M. Groot. Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, 4,1:60-81, 2000.
- S. Guo and G. Zhang, "Building a customised Google-based collocation collector to enhance language learning," *British Journal of Educational Technology*, vol. 38, 2007, pp. 747-751.
- M. Helft, "Using Computing Might, Google Improves Translation Tool," *The New York Times*, 2010.
- Y. Hirata and Y. Hirata, "Independent research project with web-derived corpora for language learning," *The JALT CALL Journal*, vol. 3, 2007, p. 33-48.

S. Hsu, "Building Language-Learning Environments to Help Technological University Students Develop English Independent Learning," *The JALT CALL Journal*, vol. 1, 2005, pp. 51-66.

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

T. Huckin and J. Coady, "Incidental Vocabulary Acquisition in a Second Language," *Studies in Second Language Acquisition*, vol. 21, 1999, pp. 181-193.

T. Hug, G. Gassler, C. Glahn: Integrated micro learning - an outline of the basic method and first results. In: *Proceedings of Interactive Computer Aided Learning*, pp. 1-7. Kassel Univeristy Press, 2004.

J. Hulstijn. 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.

J.H. Hulstijn, M. Hollander, and T. Greidanus. Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80,iii:327-339, 1996.

B. Laufer. What percentage of text-lexis is essential for comprehension? In C. Laurén and M. Nordmann, editors, *Special Language: From Humans Thinking to Thinking Machines. MultilingualMatters*, Clevedon, 1988.

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

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

N. Nagata. The effectiveness of computer-assisted inteactive glosses. *Foreign Language Annals*, 32,4:469-479, 1999.

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

H. Ogata, R. Paredes J., N. Saito, Y. Yano, Y. Oishi, and T. Ueda, "Supporting Mobile Language Learning outside Classrooms," *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, 2006, pp. 928-932.

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

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

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

- S. Powell, "Extensive reading and its role in Japanese high schools," *The Reading Matrix*, vol. 5, 2005, pp. 28-42.
- I.D. Ridder, "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.
- S. Rott. The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21,4:589-619, 1999.
- G. Sankó, "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.
- N. Schmitt, & M. McCarthy (1997). *Vocabulary: Description, acquisition and pedagogy*. New York: Cambridge University Press.
- T. Segler, PhD Research Proposal: *Second Language Vocabulary Acquisition and Learning Strategies in ICALL Environments*, 2001.
- R.J. Sternberg. 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.
- O. Streiter, J. Knapp, L. Voltmer, and D. Zielinski, "Browsers for autonomous and contextualized language learning: tools and theories," *ITRE 2005. 3rd International Conference on Information Technology: Research and Education*, 2005., 2005, pp. 343-347.
- J. Tauber, "A New Kind Of Graded Reader," February 10, 2008, 2008.
- M.B. Wesche and T.S. Paribakht, "Reading-Based Exercises in Second Language Vocabulary Learning: An Introspective Study," *The Modern Language Journal*, vol. 84, 2000, pp. 196-213.
- D. Wible, "Automating repeated exposure to target vocabulary for second language learners," *Proceedings IEEE International Conference on Advanced Learning Technologies*, 2001, pp. 127-128.
- D. Wible, "Contextualizing language learning in the digital wild: tools and a framework," *IEEE International Conference on Advanced Learning Technologies*, 2004. *Proceedings.*, 2004, pp. 231-235.
- I. Yanguas, "Multimedia glosses and their effect on L2 text comprehension and vocabulary learning," *Language Learning and Technology*, vol. 13, 2009, p. 48-67.
- G. Yongqi, & R.K. Johnson. *Vocabulary learning strategies and language learning outcomes*. *Language Learning*, 46:4:643-679, 1996.

Appendix A - Definitions

L1 - a person's first language

L2 - a person's second language or their target language being learned

gloss - a brief summary of a word's meaning

cloze test - a fill in the blank type exercise where words are removed and the learner must fill them in

CALL - Computer-Assisted Language Learning

ICALL - Intelligent CALL

CALI - Computer Assisted Language Instruction

incidental learning - learning that takes place without any intent to learn⁷

graded reader - A story which has been adapted for people learning to read or learning a foreign language. Graded readers avoid using difficult grammar and vocabulary.⁸

intensive reading - careful reading of L2 texts in order to promote a detailed understanding of the text, characterized by looking up every unfamiliar word and close study of the grammar

extensive reading - relatively fast reading of a lot of long, easier to understand L2 texts with the aim of understanding the overall meaning of a text rather than the grammar and individual words (Powell 2005)

microlearning - a difficult learning task is broken into a series of very quick learning interactions, distributed over time (Hug et al 2004)

collocation - an arrangement or juxtaposition of words or other elements, especially those that commonly co-occur⁹

7 http://psychology.wikia.com/wiki/Incidental_learning

8 <http://dictionary.reverso.net/english-cobuild/graded%20reader>

9 American Heritage Dictionary of the English Language, 2004

Appendix B - Participant Demographics

ID	Occupation	Sex	Age Range	French Education MS = Middle School HS = High School	Motivation I = improve French M = maintain French C = curiosity	Daily Web Browsing (in hours, self-reported)	Other Learning Factors
A	Student	F	18-30	HS IB French	I	2-6	some self-directed study
B	Student	M	18-30	HS	M	3-4	none
C	Student	F	18-30	1yr HS & 3yrs professional	I, C	4	none
D	Student	M	18-30	MS & HS immersion	M	3-4	none
E	Student	F	18-30	HS & 1yr in college	I	5	taking 2 nd year university French class, watched a French movie every 2 weeks, attend a couple local French meetups
F	Photographer	M	40-55	HS	M	1-2	used to live in Montreal, fluent in Spanish, listened to French radio for a short while every day, attended a couple local French meetups
G	Student	F	18-30	up to 2 nd year HS	I	3-5	none
H	Student	F	18-30	1yr HS	I	2-3	none
I	Church Leader	F	40-55	2yrs HS & 2yrs in college	I, test French knowledge	1-1.5	none
J	Accountant	F	18-30	1.5yrs classes in Paris	I	2-10	watched 2 French movies, talked to French friend once a month
K	Student	F	18-30	1.5yrs MS	I	4-5	none
L	Lawyer	M	40-55	HS & 1yr in college	I, test French knowledge	0.5-0.75	none
M	Student	F	18-30	up to 2 nd year HS	I, C	1-2	none
N	Student	F	18-30	5-week French immersion program for beginners	I, M	10-15	none
O	Student	F	18-30	HS & 2yrs in college	I	0.5	taking 3rd year university French class
P	Student	M	18-30	MS	C	4-5	none
Q	Student	M	18-30	MS & HS	I, C	8+	none
R	Student	F	18-30	HS immersion	I, C	1.5	none

Appendix C - Quiz Examples

French Word	"r" or an English Translation	French Word	"r" or an English Translation
longueur	<input type="text"/>	cher	<input type="text"/>
principe	<input type="text"/>	faible	<input type="text"/>
rouge	<input type="text"/>	fou	<input type="text"/>
bienvenue	<input type="text"/>	trou	<input type="text"/>
à charge	<input type="text"/>	mince	<input type="text"/>
mardi	<input type="text"/>	peu profondes	<input type="text"/>
avide	<input type="text"/>	malheureux	<input type="text"/>
arrière-plan	<input type="text"/>	frère	<input type="text"/>
colonne	<input type="text"/>	nuit	<input type="text"/>
beau	<input type="text"/>	profondeur	<input type="text"/>
linéaire	<input type="text"/>	nord	<input type="text"/>
plus petit	<input type="text"/>	sourd	<input type="text"/>
travail	<input type="text"/>	mot	<input type="text"/>
opposé	<input type="text"/>	plus facile	<input type="text"/>
légitime	<input type="text"/>	dette	<input type="text"/>
énorme	<input type="text"/>	demandeur	<input type="text"/>
mercredi	<input type="text"/>	moyen	<input type="text"/>
roue	<input type="text"/>	mineures	<input type="text"/>
veste	<input type="text"/>	bois	<input type="text"/>
inhabituel	<input type="text"/>	montagne	<input type="text"/>
cher	<input type="text"/>	pauvre	<input type="text"/>
examen	<input type="text"/>	inquiète	<input type="text"/>
prémisse	<input type="text"/>	chemise	<input type="text"/>
poils	<input type="text"/>	mémoire	<input type="text"/>
plus tôt	<input type="text"/>	chômeurs	<input type="text"/>

A screenshot showing half of one of the seven pages of words that ALOE study participants completed to test their vocabulary knowledge.

Directions

This is a simple vocabulary recognition and recall test used to determine your French vocabulary level. For each of the words listed, if you think you know what the word means, write its English translation to the right of the word. If you recognize the word but don't know what it means, write the letter *r*. Do not write more than one word for the translation.

For example, if you know that tasse is the French word for cup:

tasse

But if you see the word "école" and recognize it but don't know how to translate it, just type the letter *r* like so:

école

And if you can't recognize the French word, just leave the text field blank.

Please do not guess or spend more than a couple seconds on a word. You should move through the words quickly and if you don't instantly recognize or know the meaning of a word, move on to the next word.

As a shorthand, you can write *h* for the translation of words that are homonyms. For example, table is a homonym in French and English because in both languages it is spelled "table".

A screenshot of the directions that appeared on the first page of the quiz that study participants took. Study participants also received these directions verbally before accessing the online quiz.

Appendix D – Semi-structured Interview Script

On a scale from 1 to 5, how useful have you found ALOE for learning French vocabulary?
(1 being not useful at all and 5 being very useful)

On a scale from 1 to 5, how enjoyable have you found using ALOE?
(1 being very unenjoyable, 5 being very enjoyable)

How often did you stop to think about or interact with translations?
(1 being very seldom, 5 being very often)

On a scale from 1 to 5, how did you feel about the amount of translations?
(1 being way too few, 5 being way too many)

How satisfied are you with the quality of the translations?
(1 being very unsatisfied, 5 being very satisfied)

When do you find yourself using the system the most / the least?

Which features do you find most useful? Why?

Would you continue using ALOE if you could?

Did you have any other language learning interaction/media/classes going on?

What is your French background? How would you rate your French level?

Do you speak any other romance languages?

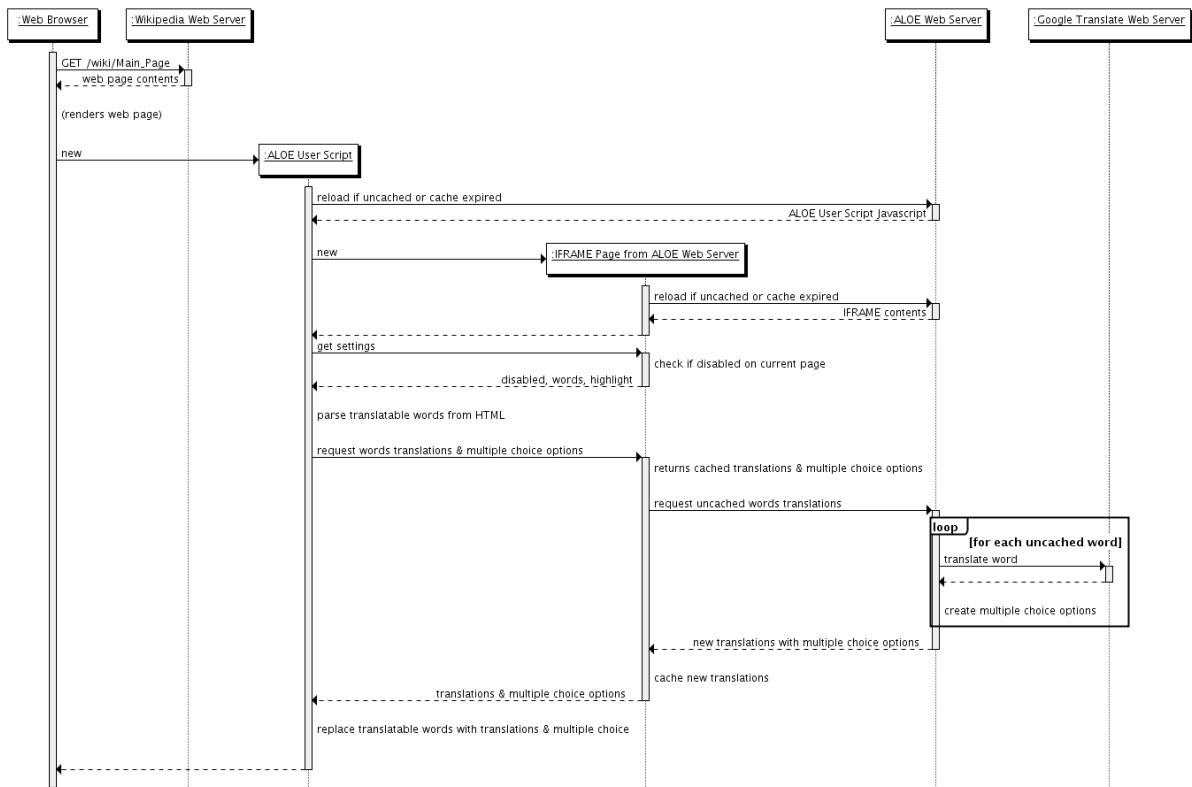
What was your motivation in participating?

Which style did you use for translated words?

About how much time do you spend on the Web every day?

Other comments, feedback or criticism?

Appendix E - Translation Process Sequence Diagram



A sequence diagram showing the steps the Web browser and the ALOE client and server perform to translate a Web page from Wikipedia.

Appendix F - French Words Translated by ALOE

absence	classes	carrières	formations	machine	petits-déjeuners	royaumes
absences	clause	carte	fort	machines	peu profondes	rue
abstrait	clauses	cartes	fou	magazine	philosophie	rues
accent	clé	cas	fraîches	magazines	philosophies	rugby
acceptation	client	catastrophe	frais	maison	phrase	russe
accident	clients	catastrophes	fréquence	maisonnette	phrases	sain
accidents	climat	catégories	fréquences	maisonnettes	pied	saisonnier
accord	cœur	catégories	frère	maisons	pieds	saire
accords	coeurs	célibataire	frères	majorité	pire	salaires
à charge	coffre	cellule	froid	majorités	plage	salle
acheteur	coffres	cellules	fromage	malade	plages	salles
acheteurs	collection	cent	fromages	maladie	plaque	samedi
acide	collections	centaines	fruits	maladies	plaques	sang
acides	collège	cercle	gagnant	maladroit	plastique	sanglante
acquisition	collèges	cercles	gagnants	malheureux	plastiques	sangs
acquisitions	collègue	certificat	gai	manière	plate-forme	sans-abri
actif	collègues	certificats	garçon	manières	plates-formes	sans
actifs	colline	chair	garçons	marchandise	plus bas	précédent
action	collines	chambre	gâteaux	s	plus beaux	santé
actionnaire	colonne	chambre à	gâteaux	marché	plus claire	satisfaisante
actionnaires	colonnes	coucher	gauche	mardi	plus courte	scène
actions	combinaison	chambres	gaz	mariage	plus facile	scènes
activité	combinaisons	chambres à	général	mariages	plus faible	science
™activités	comité	coucher	généralisée	marron	plus fort	sciences
actuel	comités	champ	génération	mars	plus grand	scientifique
adhésion	commode	championnat	générations	matériaux	plus grande	scientifiques
adhésions	communauté	championnats	généraux	matériel	plus jeune	seconde
adjoint	communautés	champs	généreux	matin	plus jeunes	secours
adjoints	communication	chance	génie	mauvais	plus large	secrétaire
administration	communication	chanson	genou	mécanisme	plus loin	secrétaires
administration	s	chansons	genoux	mécanismes	plus long	secteur
s	comparaison	chapeau	gens	méchant	plus petit	secteurs
adulte	comparaisons	chapeaux	gentil	méchants	plus petits	sécurité
adultes	compétition	chapitre	gentilhomme	médias	plus près	sélection
aéronefs	compétitions	chapitres	gentilhommes	mélange	plus proche	sélections
affaire	comportement	charbon	gestion	mélanges	plus profonde	semaine
affaires	comportements	charbons	gestionnaire	membres	plus rapide	sensibilisatio
affamé	composante	charité	gestionnaires	mémoire	plus tard	n
affreux	composants	charités	gorge	mémoires	plus tôt	septembre
âgées	compréhension	chat	gouvernement	menace	poésie	série
âge moyen	comptable	château	gouvernements	menaces	poids	serré
agence	comté	châteaux	grand	ménages	poils	serviable
agences	comtés	chats	gratuit	mensuel	point	serveur
agent	concentration	chaud	grave	mer	points	serveurs
agents	concept	chaussure	grillon	mercredi	points forts	session
agrandissez	concepts	chaussures	grillons	mercredis	politicien	sessions
agriculteur	conclusion	chemin	guitare	mère	politiciens	seul
agriculteurs	conclusions	chemin de fer	guitares	mères	politique	sexe
aide	conducteur	chemins	haut	merveilleux	politiques	siècle
aiguë	conducteurs	chemins de fer	huteur	métal	population	siècles
aile	conférence	chemise	hauteurs	métaux	populations	siège
ailles	conférences	chemises	hebdomadaire	méthode	port	signification
ainé	confidence	cher	herbe	méthodes	porte	significations
air	confidences	cheval	heures	mètre	porte-parole	sir
aire	confusion	chevaux	heureux	mètres	portes	site
aires	congrès	cheveux	histoire	meubles	ports	sites
alcool	connaissances	chien	histoires	mieux	possession	situation
alcools	connexion	chiens	hiver	mille	possessions	situations
aléatoire	connexions	chinois	hivers	miles	possibilité	société
allégué	conscience	chirurgie	homme	mille	possibilités	sociétés
allemand	conscient	chirurgies	hommes	milliers	potentiels	soeur
alliance	conseil	choix	honnête	millions	pour cent	sœurs
alliances	conseils	chômeurs	honteux	mince	pourcentage	software
âme	conséquence	chose	hôpital	mineures	pourcentages	soir
amélioration	conséquences	choses	hôpitaux	ministère	pourri	soirées
améliorations	conservation	ciel	horloge	ministères	préalable	solliciteur

amélioré	considération	circonstance	horloges	ministre	prémisse	solliciteurs
amer	considérations	circonstances	hôtel	ministres	prémises	solution
âmes	consommateurs	circuit	hôtels	minorité	préparation	solutions
ami	consommation	circuits	humeur	minorités	préparations	sombre
amis	constitution	citoyen	humeurs	minute	présence	sommet
ampleur	constitutions	citoyens	hypothèse	minutes	présences	sommets
ampleurs	construction	classe	hypothèses	mission	présentation	soudaine
analyse	constructions	classe moyenne	idée	missions	présentations	souhaitable
analyses	consultatif	éditeur	idées	mode	président	sourd
angoisse	contexte	éditeurs	idées	modes	présidents	sourds
angoisses	contextes	effet	identité	module	prêt	souterrain
animal	continue	effets	identités	modules	preuves	spectacle
animaux	contribution	efficacité	île	moindre	prévisible	spectacles
animée	contributions	effort	îles	moins cher	prince	spiritueux
année	convenable	efforts	illimité	mois	princes	station
anniversaire	conversation	église	image	moment	principe	stations
anniversaires	conversations	églises	images	moments	principes	statistique
août	coopération	élection	imagination	monde	priorité	statistiques
apeuré	corps	élections	implémentation	mondes	priorités	statut
appareil	côté	élément	implémentation	mondiale	prison	statuts
appareils	côtés	éléments	s	monnaie	prisonnier	stockage
apparence	côtier	élève	implication	monnaies	prisonniers	stratégie
apparences	cou	élèves	implications	montagne	prisons	stratégies
appartement	coupable	élus	impression	montagnes	privé	studio
appartements	couple	empire	impressions	mort	prix	studios
application	couples	empires	imprimé	mot	problème	substance
applications	cour	empiriques	inattendu	moteur	problèmes	substances
approbation	courant	emplacement	inchangé	moteurs	procédure	succès
approbations	cours	emplacements	incident	mots	procédures	sucre
approche	court	emploi	incidents	moutons	processus	sucres
approches	cous	emplois	inconnu	mouvement	prochaine	sud
approfondie	couteau	employé	inconscient	mouvements	production	suédois
après-midi	couteaux	employés	incroyable	moyen	productions	suggestion
arbre	coûteux	employeur	indépendance	moyenne	produit	suggestions
arbres	création	employeurs	indications	moyens	produits	suivant
architecture	créations	en bois	indigènes	nr	professeur	sujet
argent	crime	enceintes	individu	mrs	professeurs	sujets
argument	crimes	en colère	individus	mur	profession	symptôme
arguments	crise	endormi	industrie	murs	professions	symptômes
arme	crises	énergie	industries	musée	profondeur	syndicat
armée	critères	énergies	inépuisable	musées	profondeurs	syndicats
armées	critique	enfant	infection	musique	programme	système
armes	critiques	enfants	infections	naïf	programmes	systèmes
armoire	croissance	enfer	inflation	naissance	progrès	tâche
armoires	croissance	en forme	inflations	nation	promotion	tâches
arrangement	croyances	engagement	influent	nations	promotions	taille
arrangements	cru	engagements	infructueuses	nature	proportion	tailles
arrière-plan	cuir	ennemi	inhabituel	néerlandais	proportions	technique
arrière-plans	cuiers	ennemis	initiative	négociation	proposition	techniques
arrivée	cuisine	énorme	initiatives	négociations	propositions	technologie
arrivées	cuisines	en plein air	inquiète	nez	propre	technologies
art	curriculum	enquête	inscrit	niveau	propriétaire	téléviseur
article	curriculum	enquêtes	installation	niveaux	propriétaires	téléviseurs
articles	dame	en retard	installations	noël	propriété	télévision
artiste	dames	enseignant	institut	noir	propriétés	télévisions
artistes	danger	enseignants	institution	nombreux	protection	température
arts	dangers	enseignements	institutions	nommé	protections	températures
aspect	de base	enseillé	instituts	nord	protéine	temps
aspects	débile	enthousiasme	instruction	norme	protéines	tendance
assemblée	décembre	entrée	instructions	normes	pub	tendances
assemblées	décennie	entrées	instrument	notion	public	ténèbres
association	décennies	entreprise	instruments	notions	publication	teneur
associations	décès	entreprises	insuffisant	notoire	publications	teneurs
assurance	décision	entretien	intention	nourriture	publics	tennis
assurances	décisions	environnement	intentions	nouveau	pubs	tension
à temps partiel	déclaration	environnement	interprétation	nouvelles	puissant	tensions
à temps plein	déclarations	s	interprétations	novembre	pur	terme
atmosphère	décontracté	épais	intervention	nu	qualification	termes
atmosphères	découverte	épouse	interventions	nuisibles	qualifications	terre
attente	découvertes	épouses	introduction	nuit	qualité	territoire

attentes	défendeur	équation	introductions	nuits	qualités	territoires
attention	défendeurs	équations	inutile	objectif	quotidien	thé
attitude	défense	équipe	investissement	objectifs	quotidienne	théâtre
attitudes	défenses	équipements	investissements	observation	raisonnable	théâtres
au-dessus	définition	équipements	investisseur	observations	ratio	thème
auto	définitions	équipes	investisseurs	occasion	ratios	thèmes
autobus	degré	erreur	invité	occasions	ravi	théorie
automne	degrés	erreurs	invités	octobre	réaction	théories
automnes	déjeuner	escalier	irlandais	oeuf	réactions	thés
autorité	délit	escaliers	itinéraire	oeufs	réaliste	titre
autorités	délits	espace	itinéraires	oeuvres	réalité	titres
autre	demandeur	espaces	jambe	officier	réalités	toit
avantage	demandeurs	espagnol	jambes	officiers	récession	toits
avantages	démocratie	espèces	janvier	oiseau	récessions	tôt
avenir	démocraties	espoir	jaune	oiseaux	recettes	tradition
avertissements	démodé	esprit	jean	opération	recommandation	traditions
aveugle	dent	esprits	jeans	opérations	recommandation	trafic
avide	dents	est	jeu	opinion	s	trafics
avion	d'épargne	esthétique	jeudi	opinions	reconnaissance	traité
avions	département	établissement	jeudis	opportunité	reconnaissants	traitement
avril	départements	établissements	jeune	opportunités	recouvrement	traitements
bain	dépenses	étage	jeunes	opposé	recouvrements	traités
bains	déraisonnable	étages	jeux	opposition	réduction	tranchant
balles de	description	été	joli	option	réductions	travail
baseball	descriptions	étrange	joueur	options	régime	travailleur
ballons de	désespérée	étrangers	joueurs	or	régimes	travailleurs
football	desks	étudiant	jour	ordinateur	région	travaux
banlieues	désolé	étudiants	journal	ordinateurs	régions	très
banque	de survie	évaluation	journaux	oreille	règlement	triste
banques	détail	évaluations	jugement	oreilles	réglementation	trou
bar	détails	événement	jugements	organisation	reine	troupe
bars	détermination	événements	juillet	organisations	reines	troupes
bas	dette	examen	juin	origine	relatif	trous
base	dettes	examens	justice	origines	relatifs	ultérieur
base-ball	développement	exception	laboratoire	ouest	relation	uni
base de	développement	exceptions	laboratoires	outil	relations	uniquement
données	s	excitation	la coopération	outils	religion	unité
bases de	devoir	exemple	la durée de vie	ouvert	religions	unités
données	devoirs	exemples	l'agriculture	ouvertures	remplacement	univers
bâtiments	d'habitude	exigence	laid	pacifique	remplacements	université
beau	dieu	exigences	laïque	paiement	rendez-vous	universités
beauté	dieux	existence	langue	paiements	renseignement	urbain
beautés	différence	existences	langues	pain	rentables	urgence
bénéfice	différences	expansion	la pauvreté	pains	repas	urgences
béton	difficulté	expansions	la pollution	paix	réponse	usine
bibliothèque	difficultés	expert	las	panneau	réponses	usines
bibliothèques	digne	expertise	leadership	panneaux	représentant	utiles
bien	dimanche	experts	le chômage	papier	représentants	utilisateur
bien connu	dîner	explication	lecteur	papiers	représentation	utilisateurs
bien-être	dîners	explications	lecteurs	parent	représentations	vacances
bienvenue	directeur	exposition	l'éducation	parents	république	valeur
bière	directeurs	expositions	léger	parlement	républiques	vallée
bières	direction	expression	législation	parlements	réputation	vallées
blanc	dirigeant	expressions	légitime	paroisse	réservoir	variation
blessure	dirigeants	extension	l'électricité	paroisses	réservoirs	variations
blessures	discours	extensions	l'enfance	partenariat	résident	variété
bois	discussion	fabricant	l'ensemble	partenariats	résidents	variétés
bon marché	discussions	fabricants	lent	participation	résistance	véhicule
botte	disponible	facultative	les climats	partisan	résolution	véhicules
bottes	distinction	faible	les plus anciens	partisans	résolutions	vendredi
boule	distinctions	faisable	les plus élevés	pas claire	respiration	vente
boules	distingué	fait	lettre	passage	respirations	ventes
bovins	distribution	faits	lettres	passages	responsabilité	véritable
bruit	distributions	famille	lèvre	passé	responsabilités	vérité
bruits	district	familles	lèvres	passif	ressources	vérités
brut	districts	fascinant	liberté	patient	restaurant	verre
bruyant	division	faux	libertés	patients	restaurants	version
bureau	divisions	femme	ligue	pauvre	restriction	versions
bureaux	données	femmes	ligues	pays	restrictions	vert
but	d'orientation	fenêtre	linéaire	paysage	résultat	veste

cadeau	douce	fenêtres	l'information	paysages	résultats	vestes
cadeaux	douloureux	féroce	lit	peintures	retraite	viande
cadre	douteux	festival	lits	pénalité	réussie	viandes
cadres	d'outre-mer	festivals	littérature	pénalités	revenu	victime
café	doux	feuille	livraison	pension	revenus	victimes
cafés	drame	feuilles	livraisons	pensions	révolution	victoire
caméra	droit	février	locataire	père	révolutions	victoires
caméras	droits	fiable	locataires	pères	richesse	vie
campagne	drôle	fidèle	logement	période	rivière	vieux
campagnes	dur	filles	logements	périodes	rivières	vif
cancer	durée de vie	filles	loi	permission	roi	village
cancers	dures	fiis	loin	personnalité	rois	villages
candidat	échange	foi	lointain	personnalités	rôle	ville
candidats	échanges	fonctionnaire	lois	personne	rôles	villes
capacité	échec	fonctionnaires	loisirs	personnel	roue	vin
capacités	échecs	fondation	longueur	personnes	roues	vins
capitaine	economie	fondations	longueurs	perspective	rouge	violence
capitaines	économie	football	lord	perspectives	route	vision
capital	économies	force	lords	perte	routes	visions
capitales	écrivain	forêt	lourd	pertes	royaume	visiteur
caractère	écrivains	forêts	lundi	petit		visiteurs
caractères	carré	formation	lunettes	petit		vivante
caractéristique	carrière			déjeuner		voisin
caractéristique						voisins
s						voiture
						voitures
						vol
						vols
						volume
						volumes
						vrai
						week-end