Learner Modelling for Individualised Reading in a Second Language

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy

at

The University of Waikato

by

Michael Walmsley

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2015

Abstract

Extensive reading is an effective language learning technique that involves fast reading of large quantities of easy and interesting second language (L2) text. However, graded readers used by beginner learners are expensive and often dull. The alternative is text written for native speakers (authentic text), which is generally too difficult for beginners. The aim of this research is to overcome this problem by developing a computer-assisted approach that enables learners of all abilities to perform effective extensive reading using freely-available text on the web.

This thesis describes the research, development and evaluation of a complex software system called FERN that combines learner modelling and iCALL with narrow reading of electronic text. The system incorporates four key components: (1) automatic glossing of difficult words in texts, (2) individualised search engine for locating interesting texts of appropriate difficulty, (3) supplementary exercises for introducing key vocabulary and reviewing difficult words and (4) reliably monitoring reading and reporting progress. FERN was optimised for English speakers learning Spanish, but is easily adapted for learners of others languages.

The suitability of the FERN system was evaluated through corpus analysis, machine translation analysis and a year-long study with second year university Spanish class. The machine translation analysis combined with the classroom study demonstrated that the word and phrase error rate generated in FERN is low enough to validate the use of machine translation to automatically generate glosses, but is high enough that a translation dictionary is required as a backup. The classroom study demonstrated that when aided by glosses students can read at over 100 words per minute if they know 95% of the words, whereas compared to the 98% word knowledge required for effective unaided extensive reading. A corpus analysis demonstrated that beginner learners of Spanish can do effective narrow reading of news articles using FERN after learning only 200–300 high-frequency word families, in addition to familiarity with English-Spanish cognates and proper nouns.

FERN also reliably monitors reading speeds and word counts, and provides motivating progress reports, which enable teachers to set concrete reading goals that dramatically increase the quantity that students read, as demonstrated in the user study.

Acknowledgements

I would like to thank my supervisor Ian Witten for his guidance and direction during the research phase and continued patience during my drawn-out thesis write up.

I also acknowledge the contributions of my second supervisor Margaret Franken and three unofficial supervisors my father Michael Walmsley, and friends Robert Akscyn and Naiara Centeno.

I thank the New Zealand Tertiary Education Commission and the University of Waikato for funding my PhD.

Finally, I thank my wife, for being supportive, despite my completion taking years longer than she anticipated when she agreed to marry me.

Table of Contents

Table of Figures

Table of Tables

1 Introduction

Following an hour of speeches in Māori at Rātana Pā on 23 January 2009, John Key was asked whether it was time for our NZ Prime Minister to be fluent in both the country's official languages. He responded that although it was advantageous, given the country's bilingual nature, the probability of him becoming fluent in Te Reo is next to nil because of the other demands on his time.

Like John Key, many people desire to learn a second language, but struggle to make the significant investment of time and effort required to do so. Others study a language at school or while living overseas but subsequently struggle to find time to maintain their skills. Yet many of these people do make time for leisurely reading of books, magazines, newspapers, blogs, email, social media and so on (Pew Internet & American Life Project, 2001, 2002).

This thesis is motivated by a desire to help people combine leisure reading with second language learning, using an effective technique called *Extensive reading*. This technique involves reading large quantities of L2 text that is both easy and interesting to the individual. The major drawback is that easy and interesting L2 text is in limited supply, especially for beginner learners of languages other than English.

Using an Internet search engine like Google, it is fairly easy to find L2 reading material that is of personal interest. However, the majority of this material is deemed too difficult for traditional extensive reading by low-ability learners. This research aims to develop a computer-assisted approach that enables learners of all abilities to perform effective extensive reading using freely-available text on the web. This thesis describes the implementation and evaluation of a system that does just that.

1.1 The thesis

This thesis makes the following central claim:

An iCALL based approach with learner modelling can enable low-ability second language learners to do effective extensive reading with authentic texts.

This section outlines the scope of this claim by defining each of the terms it contains, and summarises how this thesis validates the statement.

1.1.1 iCALL

Computer assisted language learning (CALL) is defined as "the search for and study of applications of the computer in language teaching and learning (Levi, 1997, p.1)." iCALL systems use linguistics and natural language processing techniques to make CALL systems that are "intelligent".

Natural language processing combines computer science with linguistics, and uses computers to extract meaningful information from human language, and to algorithmically produce written and spoken language that is as natural as possible. An example application of natural language processing in iCALL is to automatically identify errors learners make in speech or writing and give useful feedback.

1.1.2 Second language learners

Second language (L2) refers to any language learned after one's native or first language (L1). This research aims to develop an extensive reading approach suitable for L2 learners of all abilities, from novice to near native speaker; learning any language from Spanish to Swahili; with a wide range of reading tastes, both fiction and non-fiction; undertaking their study either in a classroom setting with a teacher, or as independent learners.

This thesis focuses on the author's efforts to develop and evaluate a new computer system that helps English native speakers learn to read Spanish, and concludes by describing the process for adapting the system to other languages.

1.1.3 Extensive reading

Extensive reading is an effective language learning technique that involves leisurely reading of large quantities of easy, and interesting text—with a focus on understanding the *meaning* of the text. Ideally, learners choose what to read, and stop reading texts they find uninteresting. Contrast this technique with intensive reading which involves slow, careful reading of small quantities of difficult text—with a focus on understanding the *language* in the text.

These techniques are complementary, with intensive reading good for introducing new language, and extensive reading helping to consolidate language knowledge and build reading fluency. In the L2 classroom, intensive reading has traditionally been easier to implement because the teacher determines a set of texts to be read by all learners—whereas extensive reading requires a large collection of text that caters to the interests and ability of each learner.

According to Nation (pg. 150, 2001), extensive reading participants should read text with no more than 5% unfamiliar words, and ideally 1–2%. Hence much text written for native speakers is deemed unsuitable for all but the most advanced L2 learners. Instead extensive reading is usually conducted using "graded readers"—short books that are written especially for L2 learners using simple vocabulary and sentence structures. Over the past two decades thousands of English language graded readers have been published. One major drawback of graded readers is that it is expensive to purchase a large enough library of readers to cater for the wide range of interests and abilities in a L2 class. Furthermore, the quantity of available readers is limited for languages other than English.

1.1.4 Authentic text

Authentic text refers to literature written for native speakers, as opposed to "learner literature" (such as graded readers), which uses simplified grammar and controlled vocabulary.

The Internet is a useful source of freely-available authentic text. Learners can use web search engines to locate large quantities of interesting L2 text in their target language; however, most of the text is too difficult for unaided extensive reading by beginners.

1.1.5 Learner modelling

Learner modelling is the creation of a learner model from observing the interactions between the learner and a learning system. Computer-assisted learning environments can automatically track each learner's system usage, construct learner models, and use these models to personalise the learning experience.

1.1.6 Thesis approach

In order to validate the central claim of this thesis, the author developed a software system called FERN (acronym for **F**un, **E**xtensive **R**eading from **N**ovice to native speaker). FERN is a gamified online-learning environment that uses novel approaches to overcome challenges with extensive reading. The system includes a **search engine** that searches the Internet for interesting L2 texts at the right level of difficulty, a **reading interface** with automatically generated glosses, individualised **vocabulary exercises**, and motivating **progress reports**.

Achieving this outcome required a cross-discipline study of Education and Software Engineering. Key educational principles of L2 learning, especially those relating to extensive reading, needed to be studied and understood, along with computing concepts including web development architecture, machine translation, and search engine technologies. Furthermore, at the commencement of this research the author undertook a program of self-study to begin learning Spanish, drawing upon several years' experience in developing good proficiency in Japanese.

1.2 Thesis structure

This thesis is structured as follows:

Chapter 2 breaks down the definition of extensive reading (to quickly read a variety of large quantities of easy interesting text), contrasting it with other types of reading, outlining the benefits that result, and describing how it fits into a balanced language learning program. The chapter then describes the range of reading material used for extensive reading, and the challenges presented by the limited availability of suitable texts for beginners, particularly those learning languages other than English. The chapter concludes by outlining motivational and administrative challenges that hinder success and prevent widespread adoption.

Chapter 3 explains how to use computers to make extensive reading more effective. It begins with a history of computer-assisted reading instruction from the days of mainframes, through to the mobile devices in use today. The chapter then describes how computers facilitate access to interesting L2 texts, aid reading with dictionaries, glosses, and vocabulary review exercises, and assist with motivation and administration of reading programs. Finally, the chapter introduces a key resource in this thesis: a crowd-sourced online dictionary named *Wiktionary*.

Chapter 4 gives a non-technical overview of the functionality provided by FERN—the system developed by the author for this thesis.

At FERN's core is a learner model that accumulates each learner's vocabulary knowledge. The remaining chapters describe the design, implementation, and evaluation of FERN, and how the learner model is utilised and updated by the following key components:

- Automatic glossing and dictionary lookup during reading (Chapter 5);
- Individualised news search that grades article difficulty (Chapter 6);
- Individualised vocabulary exercises that introduce new words and practice those that were difficult during reading (Chapter 7);
- Monitoring learning and providing progress reports (Chapter 8).

Chapter 5 begins with background on the benefits of glosses in comparison to dictionary lookup, and the various approaches to automatically gloss texts. It then describes FERN's machine translation approach, and the design and creation of its hypertext dictionary for beginner learners, which provides backup support when glosses are insufficient. The chapter concludes by describing a semi-automated approach for creating word family lists: one of FERN's key resources.

Chapter 6 presents research on readability metrics and recommender systems, which was used as the basis of FERN's individualised search engine. The search engine facilitates "narrow reading" of news articles that relate to individual interests and contain large quantities of familiar vocabulary. The remainder of the chapter describes how the learner model is initialised, updated as result of performance in reading and review exercises, and used to generate news article difficulty scores, which help learners select articles of suitable difficulty. The chapter also describes the implementation of two vital components of the search engine: the screen scraper and cognate list.

Chapter 7 describes the value of supplementary exercises in learning new vocabulary in preparation for broader reading, and in reinforcing knowledge acquired during reading. It then presents several approaches to automatically generate exercises, and describes the design and implementation of exercises in FERN. The second part of the chapter presents the case for a somewhat controversial activity we call "mixed-language texts"—reading L1 texts with

carefully selected words replaced by L2 equivalents, as a fun and effective way to learn vocabulary. The major drawback of these texts is that they are difficult to create. FERN addresses this problem by automatically generating mixed-language texts that are tailored to learner ability.

Chapter 8 evaluates the thesis claim that authentic text on the Internet can be effectively used by low-ability learners for narrow reading, aided by automatically created glosses and an individualised search engine. The chapter discusses the results from a year-long evaluation of the suitability of FERN with a lower-intermediate Spanish class. Next, it analyses the suitability of narrow reading of authentic text by beginner L2 learners. Finally, the chapter examines the benefits of hypertext glosses relative to dictionaries, and the impact of incorrect glosses.

Chapter 9 concludes with a critical look at how the claims of the thesis have been addressed and evaluated. This is followed by a discussion of future work that includes the process for adapting FERN to teach other languages. Finally, the chapter describes how the author applied the lessons learnt in this thesis to develop an online learning platform for teaching the computer languages HTML, CSS, JavaScript and Python.

1.3 Contributions

The creation of FERN involved the development of a complex software system that combines its key components in a way that has never been done before (to the knowledge of the author). A key aspect of the work was developing language independent methods and software tools for creating language learning resources, namely translation dictionaries, word family lists, and cognate lists using data in Wiktionary. Additional research and development contributions that the author considers significant to the field of computer-assisted language learning include:

Automatic glossing and dictionary lookup

- A survey of approaches to automatically gloss digital texts.
- Design and implementation of an automatic glossing approach that is powered by Google Translate.
- Classroom evaluation of extensive reading aided by automatically generated glosses and dictionary lookup as a backup.

- Analysis of the impact of errors in automatically generated glosses.

Mixed-language texts

- History of the use of mixed-language texts for language learning.
- Development of an approach for the automatic creation of mixed-language texts that adapt to the ability and experience of the learner.

Individualised L2 learner search engine

- Survey of approaches to developing search engines for L2 learners.
- Development of an individualised search engine for L2 learners.
- Method for automatically extracting news article content from web pages.

Classroom extensive reading

- Survey of keys to successful classroom extensive reading.
- Development of a novel approach to tracking how many words students read with no administrative work.
- Implementation and classroom evaluation of a gamified system that includes an individualised search engine, reading aided by automatic glossing, mixed-language texts and vocabulary review exercises.
- Analysis of how narrow reading can lead to effective extensive reading with authentic texts by beginner L2 learners.

2 Background to extensive reading

Extensive reading (ER) is a language learning technique that involves fast reading of large quantities of interesting easy-to-understand second language text. It has proven to be an enjoyable and effective language learning technique, both in and outside of the classroom (Robb & Kano, 2013). Ongoing research (such as Huffman, 2014; Nakanishi, 2014; Rodrigo, Greenberg, & Segal, 2014; Senoo & Yonemoto, 2014) continues to evaluate the effectiveness of extensive reading relative to other techniques, and the role it should play in balanced language learning courses.

According to Day and Bamford, the overall goal of extensive reading is twofold (Day & Bamford, 1998, p. 6): improve learners' general L2 proficiency; help learners develop a lifelong love of L2 reading.

Students choose reading material (possibly guided by their teacher) that is both interesting and at the right difficulty level. If students find a text uninteresting or difficult to understand without using a dictionary, they are encouraged to stop reading and choose again. As their reading ability increases, they progress to increasingly advanced texts. Day and Bamford (1998, p. 121) suggest that the motto of extensive reading be “reading gain without reading pain”.

In the first half of this chapter, we present the characteristics of successful extensive reading programs, the benefits that result, and a theoretical perspective on its role in second language acquisition. The second half of the chapter describes the challenges with implementing extensive reading and suggests how they may be overcome.

2.1 Four types of reading

Extensive reading can be contrasted with other types of reading that are used in language learning: intensive reading, speed reading, skimming and scanning.

2.1.1 Extensive reading

Researchers often describe extensive reading by listing characteristics (Bell, 1998; Day & Bamford, 1998; Ono, Day, & Harsch, 2004; Schmidt, 2007). Day and

Bamford (1998) give the following ten characteristics of successful classroom extensive reading programs:

1. Students read as much as possible.
2. A variety of text on a wide range of topics is available.
3. Students choose what, when and where to read.
4. Students read for pleasure, information, and general understanding.
5. Reading is individual and silent.
6. Reading speed is usually relatively fast.
7. There are few or no follow-up exercises after reading.
8. Students read text that they can understand without frequent dictionary use.
9. Teachers guide students through the program and monitor their progress.
10. Teachers set an example by doing extensive reading themselves.

Some researchers use the term "extensive reading" even in situations that do not match all ten characteristics. This thesis derives its definition of extensive reading from the combination of characteristics 1–6, and suggests how the guidelines provided by 7–10 can be effectively applied to computer-assisted reading.

2.1.2 Intensive reading

In contrast with extensive reading, intensive reading is slow careful study of short difficult passages with the goal of complete understanding. Learners analyse texts sentence-by-sentence until they understand the grammar and vocabulary well enough to translate it into their first language. Teachers sometimes combine intensive reading with translation by having students translate sentences or short passages from L2 to L1 (Barfield, 1995). Intensive reading helps learners increase knowledge of language features, improve reading strategy usage and improve comprehension ability (Nation, 2009).

2.1.3 Speed reading

> [Speed reading is] a very effective way of increasing reading speed by asking learners to read graded readers at a level which is much easier than the level they would normally read to gain meaning-focused input.
>
> Nation (2009, pg.69)

Essentially, speed reading is easy extensive reading, with the goal of developing fluency by reading more smoothly and naturally. Whereas normal extensive reading

uses texts in which about 2% of the words are unfamiliar to the learner, speed reading uses texts with few if any unknown words. In extensive reading, learners are encouraged to read at a comfortably quick pace, whereas in speed reading they read as fast as possible without negatively impacting comprehension. Yen (2012) demonstrated that a small investment of time in speed reading significantly increases reading speeds in other forms of reading.

2.1.4 Scanning and skimming

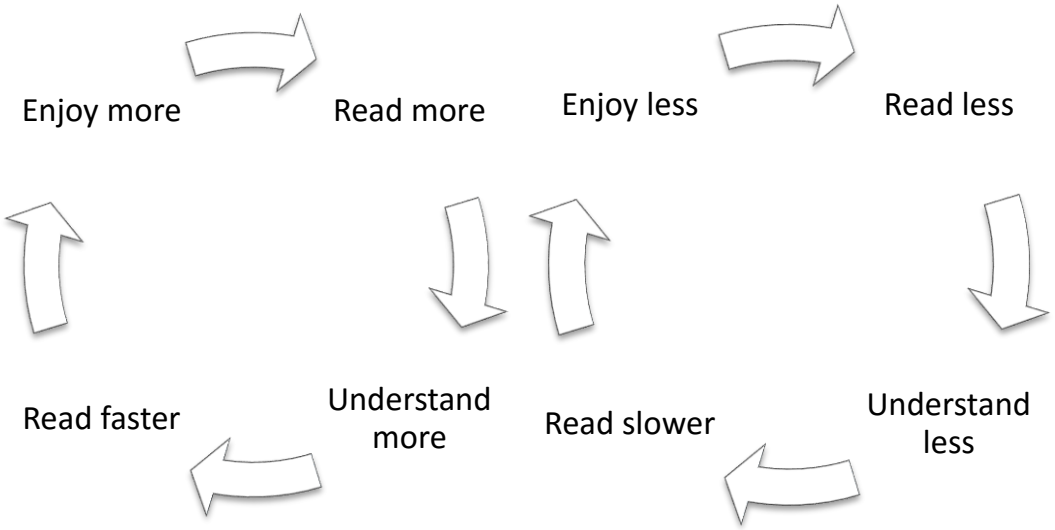
Scanning involves quickly searching text for a specific piece of information, whereas skimming is quickly reading a text to get the main idea. In both skimming and scanning, learners skip many words in order to read quickly (Nation 2009, pg. 70). In extensive reading, learners read more carefully, skip fewer words and achieve more thorough comprehension.

2.2 Requirements for successful extensive reading

Nuttall's (1996) virtuous cycle (see Figure1a) of reading gives a good overview of the potential benefits of extensive reading. Essentially, the more learners read, the more they learn; the more they learn, the faster they read; the faster they read, the more enjoyable the experience; and the more enjoyable the experience, the more motivated they are to read.

But in order for a learner to break into this cycle, they first need access to text they can easily understand. If they try to read texts that are too difficult they end up reading slowly and require frequent dictionary use to achieve good comprehension, which in turn reduces enjoyment and motivation, as illustrated by the vicious cycle of reading in Figure 1b.

Successful extensive reading requires access to *large quantities* of *interesting* text that is *easy to understand* while reading *quickly*. The following sections define these requirements in detail, and discuss the teacher's role in helping to motivate learners.



a) Virtuous cycle of reading b) Vicious cycle of reading

Figure 1: Virtuous and vicious cycles of reading, adapted from Nuttall (1996)

2.2.1 Easy to understand text

Researchers estimate that learners should be familiar with at least 98% of the words in texts used for extensive reading (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010). Carver (1994) suggests that at 98% vocabulary coverage reading is difficult but comprehensible; 99% coverage makes reading comfortable; close to 100% is required for easy reading.

Nation (2001) suggested that 3–5% unknown words was acceptable for extensive reading that focuses on learning new words and developing reading skills.

> … extensive reading texts should contain no more than 5% unknown tokens (excluding proper nouns) and preferably no more than 2% to ensure that comprehension and guessing can occur, and no less than 1–2% to make sure there is new vocabulary to learn.
>
> Nation (pg. 150, 2001)

More recently, Laufer and Ravenhorst-Kalovski (2010) suggested that 98% is optimal for unaided reading, but 95% is the minimal vocabulary coverage that enables learners to read with some help. The Extensive Reading Foundation's *Guide to Extensive Reading* (2011) draws the boundaries at 90–97% for intensive reading, 98–99% for extensive reading, and 100% for speed reading practice.

These percentages are approximations that depend on your interpretations of "familiar word". Furthermore, the exact percentage of unfamiliar words that makes comprehension difficult depends on individual learners' background knowledge and ability to infer the meaning of unknown. The threshold also varies between texts, depending on the richness of the context surrounding unknown words.

Nation estimated the number of words learners need to know in order to achieve 98% vocabulary coverage of various types of texts written for native speakers. Familiarity with the top 5000 word families is required to read teenage novels (Hirsh & Nation, 1992), approximately 9000 word families to read adult novels, and 8000–9000 for newspapers (Nation, 2006). Knowledge of 5000 word families is well beyond the ability of a beginner or intermediate language learner. Thus it follows that adult and teen authentic text—text written for native speakers—is inappropriate for extensive reading by all but the most advanced second language learners (Higgins & Wallace, 1989). Nation (2009, p. 51) concludes:

> The clear message from this is that for learners of English to do extensive reading at the elementary and intermediate stages of proficiency, it is essential that they read graded readers that have been specially prepared for learners of English. It is only by reading such texts that learners have the density of known words that is essential for extensive reading.

Graded readers are simplified adaptations of authentic text, or stories written especially for L2 learners. Major book publishers have published thousands of readers that are grouped into levels based on the vocabulary they contain, with books that cater well for the abilities of both beginner and intermediate learners. However, graded readers have several shortcomings (see Section 2.6.1).

2.2.2 Reading fast

Extensive reading encompasses the vague requirement that reading is "fast" (Day & Bamford, 1998, pg.8). The faster learners read, the more reading they complete and the larger their gains. Some experts suggest that reading speeds of 180 words per minute are a target for maximum enjoyment with full comprehension (Higgins & Wallace, 1989). Limited capacity and retention of working memory can make comprehension difficult when reading slowly (see Section 2.3.2). On the other

hand, comprehension is impaired by reading too quickly. Learners should read as fast as they can without degrading comprehension.

The speed at which comprehension deteriorates depends on the learner's ability and the difficulty of the book. The Extensive Reading Foundation (2011) recommends that learners read "at least 150–200 words per minute or a little lower for beginning students". Nation (2009, pg.72) suggests that a silent reading speed of 250 words per minute—the average careful reading speed of a native speaker—is a good goal for L2 learners.

In a study by Bell (2001), a group of beginner L2 learners with reading speeds of 50–100 words per minute increased their speeds by 90% through participating in a two semester extensive reading program.

When they encounter unfamiliar words, learners should try to guess the meaning of unfamiliar words from context and continue reading as long as they get a general understanding.

> There are strong reasons for considering a ban on dictionaries, however, at least initially… "no dictionaries" can symbolize the differences between extensive and other approaches to reading... Students must break the habit of looking up every unknown word... Students who understand what extensive reading is can be given more freedom. When a word appears several times in a text and seems to hold important meaning, it is sensible to seek its meaning.
>
> Day and Bamford (1998, pp. 93–94)

If learners feel they need to look up more than 5% of the words then the book is probably unsuitable for extensive reading and they should choose an easier one.

2.2.3 Large quantity of text

As its name implies, *lots* of reading is another key requirement of extensive reading. Since vocabulary acquisition occurs incrementally through numerous significant encounters and spaced repetition (Horst, Cobb, & Meara, 1998; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb, 2007; Zahar, Cobb, & Spada, 2001), the more learners read, the more vocabulary knowledge they acquire.

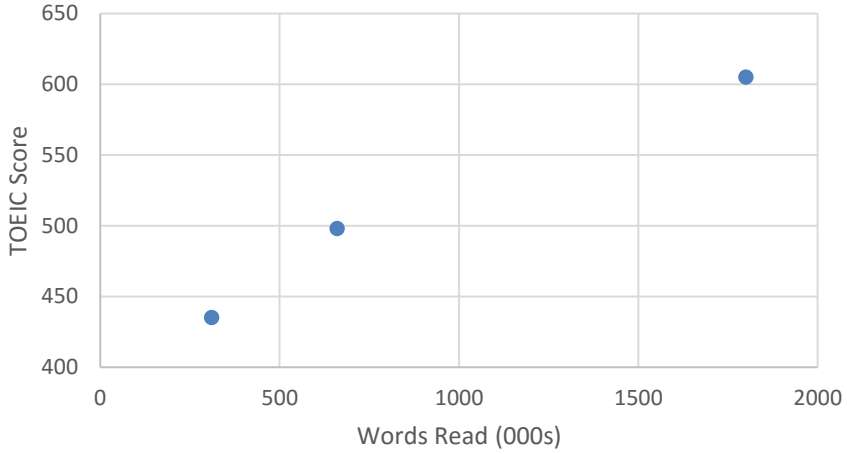


Figure 2: Increase in TOEIC scores that results from extensive reading (Nishizawa, Yoshioka, & Fukada, 2009)

Susser and Robb (1990) reported that suggestions on how much students should read range from two hours per week to one hour per day. Nation (2009) recommends that students read close to 500,000 words per year, which is equivalent to one and a half substantial first year text books, or six novels.

Nishizawa et al. (2009) monitored four years of extensive reading by engineering students at a Japanese college of technology. The researchers aim was to determine the quantity of reading required to raise student's TOEIC[1] scores from 373 (the national average for students in colleges of technology) to 500 (the national average across fourth year university students of all majors). Students had access to 15,000 graded readers, and participated in silent reading during a 45 minute weekly class period. The students were also encouraged to read at home and keep a log book of everything they read. At the end of four years, half of the 37 students that completed the course had read 690,000 or more words.

Students were divided into three groups based on the quantity read. As shown in Figure 2, group one averaged 310,000 words and 435 on the TOEIC test; group two averaged 660,000 words and a 498 TOEIC score; group three averaged 1,800,000 words and a score of 605. This study demonstrated the large improvements in L2 ability that result from large quantities of reading over an extended period of time.

[1] The Test of English for International Communication

This study further analysed 19 students who read at least 1,000,000 words and sat the TOEIC test several times throughout their university studies. Student scores after reading 1,000,000 words ranged from the low 300s to low 600s. Student TOEIC scores increased, on average, by 4–5 points per 100,000 words read.

2.2.4 Variety of interesting text

One of the goals of extensive reading is to help learners develop a love of L2 reading. As in just about every learning context, developing and nurturing motivation to learn is a key to success (Dörnyei, 1994). Consequently, learners are encouraged to self-select text that is either entertaining or contains interesting information, and to choose books with appealing titles and covers. They should stop reading if they find a text uninteresting or difficult, and are encouraged to re-read books they particularly enjoyed. By doing so learners are more likely to enjoy reading, desire to read more and ultimately develop a lasting love of reading.

In order to make this possible, learners need access to a variety of interesting reading material from which to locate texts that match individual tastes. Thus, L2 classrooms require large volumes of text to cater for diverse backgrounds, interests and abilities. The collections are usually organised by difficulty level and genre to enable quick identification of books that are both interesting and at an appropriate level of difficulty.

Learners should discuss and recommend enjoyable texts to their classmates. Doing so improves motivation to read (Lapp & Fisher, 2009; McKool, 2007; Sheu, 2004).

2.2.5 Administration support

The high cost of setting up extensive reading programs in schools is often a barrier to convincing decision makers to adopt extensive reading into the curriculum. But high costs are not the only barrier. Until a decade ago many researchers challenged the effectiveness of extensive reading in comparison to other language learning techniques. While it is now widely accepted as an effective technique, some administrators question its value, and struggle to see how silent reading is a valuable use of teacher and students' classroom time.

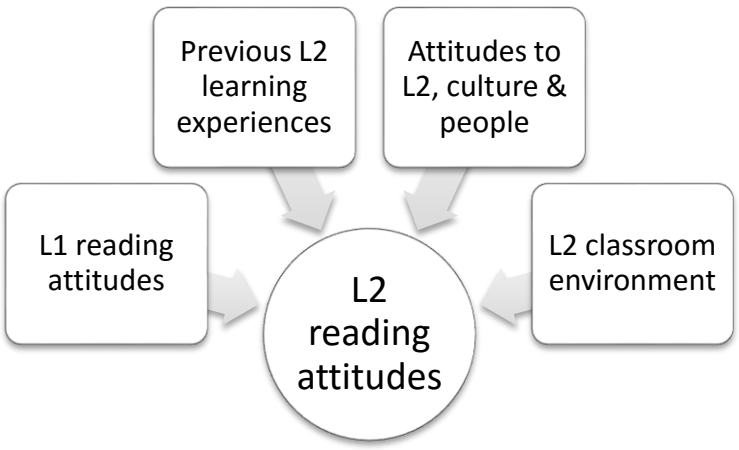


Figure 3: Model of the acquisition and development of L2 reading attitudes

2.2.6 Positive attitude

As with other language learning activities, attitude plays an important role in extensive reading. Ajzen defines attitude in the general sense as "a disposition to respond favourably or unfavourably to an object, person, institution, or event" (2005, p. 3). In other words, attitude is the tendency to like or dislike something based mostly on past experience.

Figure 3 (Day & Bamford, 1998, p.23) presents several factors that influence attitude toward L2 reading. First, positive attitude towards first language reading is likely to result in excitement to do L2 reading. Second, learners' are affected by positive and negative L2 learning experiences from the past. Finally, learners' are influenced by their learning environment and their attitudes towards the second language, culture and people.

2.2.7 Motivation

Attitude and motivation are closely linked. Day and Bamford (1998, p. 27) define motivation as a force that makes people do (or *not* do) something. One psychological model states that motivation has two equal components—*expectations* and *value*. People tend to undertake tasks they expect to accomplish successfully, and which they value highly. The perceived value of the activity is influenced by the learner's sociocultural environment, including family and friends (Day and Bamford, 1998, p. 28).

In the classroom, teachers play a key role in helping leaners set expectations and understand the value and purpose of extensive reading. Simply purchasing a large library of reading material and blocking in 30 minutes of weekly silent reading does not guarantee success. For all but the most motivated students, teachers play a crucial role in the short term goal of getting students to read as much as possible during a course, and the long term goal of developing a love of reading. They influence learners' attitudes by establishing an environment that is conducive to learning, and guiding them through positive language learning experiences.

At the commencement of a course, teachers explain the benefits of extensive reading and how to do it successfully. This is a challenging task, especially in classes where the teacher does not speak the students' native language. Teachers should be familiar with the available reading material and the ability of each student so they can guide students toward interesting and level appropriate text, and help them move systematically through the grade levels when using graded readers (Nation, 2009, p. 56).

L2 teachers also often use extensive reading in conjunction with oral reading, which involves listening and reading aloud as a class. They also teach reading strategies, such as coping with ambiguity during read by ignoring it or guessing the meaning of unknown words from context. Teachers should be a role model by participating in extensive reading themselves (Day & Bamford, 1998, p. 166-168), discussing their own reading with their class (Sheu, 2004), and recommending texts to their students (McKool, 2007).

Dörnyei and Csizér (1998) developed the "ten commandments" in Figure 4 to guide teachers in motivating language learners. These guidelines stress the importance of teachers' examples in L2 classrooms. In their research (Cheng & Dörnyei, 2007; Dörnyei & Csizér, 1998), students in Taiwan and Hungary rated their teacher's example and behaviour as the strongest factor influencing their motivation to learn. Students also rated highly the importance of teachers creating a pleasant and relaxed classroom environment.

de Burgh-Hirabe and Feryok's (2013) study of Japanese learners doing ER demonstrated that external factors such as national exams and social distractions had a major impact on motivation to read.

1. Set a personal example with your own behaviour.
2. Create a pleasant, relaxed atmosphere in the classroom.
3. Present the tasks properly.
4. Develop a good relationship with the learners.
5. Increase learners' linguistic self-confidence.
6. Make the language classes interesting.
7. Promote learner autonomy.
8. Personalize the learning process.
9. Increase learners' goal-orientedness.
10. Familiarise learners with the target language culture.

Figure 4: Ten commandments for motivating learners (Dörnyei & Csizér, 1998)

2.2.8 Reading logs and goals

Mori (1999) surveyed L2 learners and analysed the key factors affecting motivation to read, and concluded that learners prefer tasks that are structured in a way that gives clear indication of their progress. Sheu (2004) found that combining wall charts with teacher feedback and student discussion groups, improved the motivation of learners. Dörnyei and colleagues stress the importance of goal setting (Cheng & Dörnyei, 2007; Dörnyei & Csizér, 1998).

Nation (2009, p. 76) adds that graphs of reading speed and comprehension over time are also highly motivating for learners, especially if learners set target speed goals. Language acquisition is a long process, and having concrete, easily measurable indications of incremental progress is beneficial. Reading speed is a valuable metric for motivating learners because it is easily quantified and learners can quickly achieve noticeable gains (Bell, 2001). Measuring short-term gains in vocabulary size is more difficult.

In general, students are at least partially motivated by grades. Teachers often report that students do not read unless compelled to do so, either as an in-class activity, a graded homework assignment, or a combination of both. Day and Bamford (1998, pp. 86-90) state that there is no reason why teachers cannot assign grades for an extensive reading course. Student progress can be tracked via interviews, quizzes, short written or oral book reports, or reading logs that require students to state what they read, how long they read and whether or not they enjoyed reading. Nation (2009, p. 53) recommends that learners who read the same book work together on

book reports. Dupuy et al. (1996) involved students in deciding how their extensive reading was assessed.

In university extensive reading courses teachers often assign grades based on the total number of words read during a course. Thus they require a practical and reliable way to measure the numbers of words students read. Regular teacher-student interviews may be reliable, but consume too much time to be practical for large class sizes. Reading logs rely on the honesty of students. Given the high stakes nature of university courses, students will cheat by listing books they have not read. The reliability can be improved by requiring students to write more detailed reviews, however, this negatively impacts the enjoyment of extensive reading.

Section 3.8 discusses computer-assisted approaches to monitoring reading. This thesis builds on these approaches to develop a novel method to automatically track students reading speed and the number of words they read during extensive reading. This approach provides motivating data for students and useful feedback for teachers.

2.3 Extensive reading theory

2.3.1 Role of extensive reading in second language acquisition

Proponents have differing views on the place that extensive reading should take in language learning programs. Krashen (1989, 2004; McQuillan & Krashen, 2008) proposed that beginners spend little time on anything but extensive reading and extensive listening. Other researchers advocate complementing the use of both of these with direct study of high frequency language features (Brown, 2009; Green, 2005; Laufer, 2003; Rosszell, 2006; Schmidt, 2007).

Nation (2001) describes four strands of language learning that are equally important in a well-balanced language course: meaning-focused input, language-focused learning, meaning-focused output and fluency development.

Meaning-focused input involves reading and listening to comprehensible input with a focus on understanding the information being presented as opposed to individual language items. This type of learning works well when the input contains few language items that are unfamiliar to the learner. This strand includes extensive reading.

Language-focused (or form-focused) learning involves deliberate teaching of individual language features such as vocabulary and grammar. Research shows that direct study of vocabulary and grammar gives a strong boost to the gradual cumulative learning that occurs from other learning techniques, such as extensive reading and listening (Ellis, 1990). This strand includes intensive reading—slow reading of difficult texts, giving deliberate attention to language features such as vocabulary and grammar.

Meaning-focused output involves speaking and writing, with a focus on conveying information. This approach incorporates the use of activities that not only develop speaking and writing ability, but also enhance vocabulary knowledge (Nation, 2001), which results in improved reading ability.

Fluency development involves learners working with and becoming more fluent in using and understanding language items that they are already familiar with. This strand includes *speed reading*—reading material well within their proficiency, with a focus on increasing speed while maintaining good comprehension.

2.3.2 Cognitive view of extensive reading

Day and Bamford (2009, pg.11) define reading as *using previously attained knowledge to construct meaning from a printed or written message*. Cognitive psychologists have developed models to explain the reading process. The most widely accepted model of fluent first language reading suggests that reading involves the interaction of a variety of processes.

The first process is *automatic word recognition*, in which readers briefly fixate their eyes on most words—over 80% of content words and over 40% of function words such as *of* or *the*—as they move their eyes rapidly across each line of text (Harrison, 1992). Readers fixate on about 4–5 words per second and rarely skip two successive words. The words that they recognise automatically are referred to as *sight vocabulary*. Having a sight vocabulary that includes most of the words in a text is pre-requisite for good reading comprehension (Stanovich, 1992). Automatic recognition makes possible fast *lexical access* or re-call of contextually appropriate word meaning.

Fluent readers have large sight vocabularies that enable automatic and subconscious word recognition and lexical access, which allows them to focus cognitive energy

on comprehension (Adams, 1994). When readers encounter words that are not in their sight vocabulary they must resort to the slow and effortful process of *phonemic decoding*. For beginner readers, the decoding task requires so much attention that it cannot be performed simultaneously with the comprehension task—a process that is costly in terms of demands on attention. Readers switch their attention to comprehension once decoding is complete (Samuels, 2006).

In addition, readers can only store decoded words in short-term memory for a short period of time. If decoding takes too long, words begin to drop out of short-term memory before comprehension occurs, and the reader is forced to return to the beginning of the sentence to restart the decoding process.

The process of reading comprehension draws on knowledge of language and text structure, as well as knowledge of the world, especially the background of the text topic. This knowledge is stored as organised interrelated structures or *schemata*. Extensive reading is an effective way to obtain world knowledge, which aids future reading experiences.

The development of sight vocabulary is achieved by repeated exposure in a variety of contexts—known as overlearning—to words that are already familiar. Extensive reading is a good technique for reinforcing sight vocabulary, because it exposes learners to large quantities of words that they are already familiar with.

2.4 Vocabulary acquisition

The ultimate achievement in second language learning is to read, write, speak and listen like a native speaker. Vocabulary acquisition is critical in the development of all four skills. This section addresses the following important questions, with a focus on the application to reading:

1. What does it mean to know a word?
2. What is a word family?
3. How many words does a learner need to know?
4. How does a learner learn a word?

The answers introduce several key concepts that underpin this research.

Table 1: Aspects of productive and receptive vocabulary knowledge (Nation, 2001)

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	Form & meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concept & referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types must we use with this one?
	Constraints on use	R	Where, when, and how often would we expect to meet it?
		P	Where, when and how often can we use this word?

Note: In column 3, R is receptive knowledge, P is productive knowledge.

2.4.1 Knowing a word

> There are many things to know about any particular word and there are many degrees of knowing.
>
> Nation (2001, p. 27)

Nation (2001, pg.24-25) draws a distinction between *receptive* and *productive* knowledge, otherwise referred to as *passive* and *active* knowledge. While the distinction is not completely clear, Nation describes receptive knowledge as "perceiving the form of a word while listening or reading and retrieving its meaning", and productive knowledge as "wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form."

Table 1 from Nation (2001, p. 27) describes three aspects of word knowledge—form, meaning and use—in relation to productive and receptive use. Generally, developing receptive knowledge is easier than productive knowledge for a given

Table 2: Average scores for receptive and productive learning and testing of French vocabulary (Stoddard, 1929)

	Receptive Test (fr-en)	**Productive Test (en-fr)**	**Total**
Receptive learning (fr-en)	15.1	6.0	21.1
Productive learning (en-fr)	13.1	8.0	21.1
Average	14.1	7.0	21.1

word (Rott, 1999). One common explanation for this asymmetry is that the amount of knowledge required for recognising words is less than producing them. For receptive use and retention, learners need to recognise the form of a word and make the link to a known first language synonym or to familiar concept. For productive use, they need more precise knowledge of the word form.

Stoddard (1929) conducted early research to differentiate receptive and productive vocabulary learning. Half of 328 native English speaker subjects studied 50 word pairs "receptively" (viewing the French and recalling the English) and the other half studied "productively" (viewing the English and recalling the French). Subjects were then tested on both productive translation (from English to French) and receptive translation (from French to English). As shown in Table 2, both groups averaged 21 as their combined score. However, students who studied productively scored one point higher than average in the productive test, and one point lower in the receptive test; those who learned receptively had the reverse results.

More recently Webb's (2009) subjects learned word pairs receptively and productively, and took four tests that measured reading comprehension, writing, and receptive and productive vocabulary knowledge. Webb's findings supported Stoddard's conclusions that the direction of vocabulary learning is important, with receptive learning more effective for receptive use and productive learning more effective for productive use. Furthermore, productive use requires deeper word knowledge and takes more time to acquire than the knowledge required for receptive use.

Developing receptive knowledge of a new word requires a series of significant encounters spread out over time. Studies estimate that at least 6 encounters with a word are required for stable receptive knowledge (Nation, 2001, p. 81). Saragi, Nation and Meister (1978) attributed 20% of the variation in learning to the number

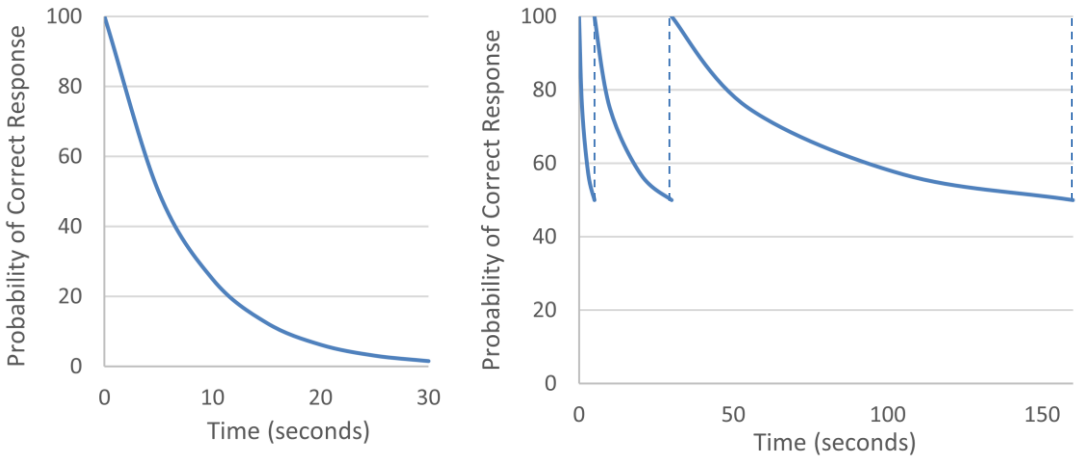
of repetitions. Vocabulary learning also depends on the ability of the individual learner and the inherent characteristics of each word (Pigada & Schmitt, 2006).

Nation (1990) uses the term *learning burden* to describe the effort that is required to learn a word. The learning burden is affected by the learner's knowledge of their L1 and L2. Learners more easily learn L2 words for which they know a related L2 word (Zahar et al., 2001). Thus, increasing L2 vocabulary decreases the learning burden for additional L2 words. The proportion of unfamiliar words in a text affects the rate of vocabulary learning during reading. Furthermore, learning is affected by background knowledge of the subject matter (Nation, 2001).

The distance between a learner's L1 and L2 is another significant factor in learning vocabulary. For example, it is significantly easier for native English speakers, to learn German, French and Spanish than Japanese, Korean or Arabic. According to the Educational Testing Service (ETS) it takes 3.5–4 times longer to reach "level 3 oral proficiency" in Japanese than for Spanish or French (Samimy, 1994). Given the complexity of the East Asian writing systems, the multiplier is no doubt much larger for learning to read or write. For a Chinese native speaker, learning to read Japanese is easy due to the abundance of shared meaning in the Japanese kanji (Brown, Hill, & Iwashita, 2000; Iwashita, 2012).

Along with using similar writing scripts, it is easier for native English speakers to learn Spanish and French than East Asian Languages, due to the large number words that have shared Latin roots. These words, known as *cognates*, are relatively easy to learn. For example, the cognate *atleta*—the Spanish word for athlete—requires few encounters for English speakers to learn because it has almost identical spelling and meaning in both languages. Cognates vary in ease of learning, depending on the similarity of the spelling and meaning of the word pair (see Section 6.5.3). De Groot and Keijzer (2000) provide evidence of the ease with which cognates are learnt and remembered, and Rodriguez (2001) uses an early focus on cognates as an efficient approach to teaching Spanish speakers to read English. The *Shortcut to Spanish* program[2] is built around teaching 3013 cognates to help learners achieve *"maximum Spanish communication in minimum time."*

[2] Marketed online at how-to-speak.com



(a) Exponential forgetting curve (b) Forgetting curve with spaced review

Figure 5: Pimsleur's (1967) exponential forgetting curve for vocabulary learning

The quality of the encounters is also significant. The context surrounding unfamiliar words has a significant effect on the rate of vocabulary learning from reading (Webb, 2007). Contexts can make directive (positive), misdirective (negative) or non-directive (neutral) contributions to guessing the meaning of unknown words (Beck, McKeown, & McCaslin, 1983). Words from some parts of speech are also more difficult to learn than others. For example, because verbs have more forms than nouns they require more encounters to learn.

2.4.2 Spaced repetition

Another factor that affects vocabulary learning is the time between encounters. Ebbinghaus determined that in order to achieve long term retention, several short study sessions separated by rest intervals is more effective than one large uninterrupted session (reported by Murphy, 1929). For example, two 20 minute sessions on successive days is more effective than one 40 minute session. This principle is known as *spaced repetition*.

Ebbinghaus also demonstrated the exponential nature of forgetting newly learned information. Figure 5a illustrates this principle in relation to vocabulary learning, as presented by Pimsleur (1967). After learners learn new words, the probability of recalling their meaning decays at a rate that decreases over time. Most of the forgetting occurs in the first few seconds after initial learning. Several spaced reviews of the word are required to achieve stable learning. Figure 5b presents Pimsleur's *graduated interval recall* schedule, which seeks to optimise the

efficiency of vocabulary review. The moment learners review words, the probability of correct recall is momentarily restored to 100%, at which point it begins to decay again, but less rapidly than the first time. With each subsequent revision the rate of forgetting continues to decreases.

Pimsleur states that the gap before the n^{th} revision should be X^n seconds. For example, if the first revision occurs 5 seconds after the initial encounter, then the second should occur after approximately 5^2 or 25 seconds, the third after 5^3 or 125 seconds, the fourth after 5^4 or 625 seconds and the tenth occurring 113 days after the ninth. The value of X in the equation depends on learner ability, word difficulty, and the quality of the encounters. Pimsleur stated that the "first interval can roughly be defined as the time that elapses before the student's probability of remembering the item drops to some arbitrary level, say 60%." If the space between intervals is shorter or longer than the optimum, then the number of encounters required to gain solid word knowledge will be greater than the minimum.

Between 1963 and 1971 Pimsleur used this technique to create audio language courses to teach high frequency vocabulary and phrases (ref). The courses consist of daily 30 minute lessons, during which learners mimic phrases spoken by a native speaker. The introduction of new phrases is interspersed with revisions that follow the graduated recall interval schedule.[3]

The principles of spaced repetition and exponential forgetting were incorporated in algorithms of popular language learning tools such as Supermemo (Wozniak, 2007) and Anki (Godwin-Jones, 2010).

2.4.3 Word family lists

This section describes the key idea of *word families,* and how they are created. These lists are useful for helping learners determine which words to learn. This thesis utilises existing English word family lists, and develops a new list for Spanish. These lists play a central role in the algorithms and tools developed in this research.

In language learning, it is useful to group the inflected and derived forms of base words in a collective known as a *word family*. For example, the base word *run* is

[3] http://www.pimsleur.com/

Table 3: The six word family levels described in Bauer and Nation (1993)

Level	Example Affixes	Total	Example Words	Types (000s)	Token (000s)	Family (000s)
1. Each form is a different word		0	develop	62	1000	62
2. Inflectional suffixes	-ing, -er, -est, -s, -ed	8	develops developed developing	24	380	38
3. The most frequent and regular derivational affixes	-ly, -er, un-, -ness, -y, -able, -less, -th, non-, -ish	10	developable undevelopable developer(s) undeveloped	3.5	22	34
4. Frequent, orthographically regular affixes	-ation, -al, -ment,-ity, -ful, -ous, -ist, -in, -ism, -ess, -ize	11	development(s) developmental developmentally	1.8	16	32
5. Regular but infrequent affixes	-an, -ance, -ence, en-, -ly, -al, -age, -ship, -ian, -ary	50	developmentwise semideveloped antidevelopment	1.8	8.6	31
6. Frequent but irregular affixes	-ion, -ive, -ic, re-, -th, -ition, -y, -ee, pre-, -ify, -able, -ist	12	redevelop predevelopment	1.4	12	29

grouped with *runs, ran, runner* and *running*. However, words like *hard* and *hardly* are spelt similarly, but have unrelated meanings and thus are not in the same family.

Deciding how to group words into families depends on the application. For L2 reading, the base word is grouped with related forms that learners can understand without having to learn them separately. These groupings depend on each learner's knowledge of affixes, which develops over time.

Bauer and Nation (1993) grouped English affixes into six levels based on the number of words in which an affix occurs, the predictability of its meaning, and the regularity of meaning, form and spelling. These levels are designed with a focus on recognition of written words. Bauer and Nation set them up for practical reasons, and do not claim that they have theoretical value.

Table 3 shows the suffixes that Bauer and Nation added to word families at each level. Level 1 treats all word forms as separate. Level 2 adds the inflectional suffixes of plural; third person singular present tense; past tense; past participle, -ing; comparative; superlative and possessive. Levels 3–6 add derivational affixes, with

the most frequent and regular affixes in level 3 and the least frequent and irregular affixes in levels 5 and 6. Bauer and Nation also describe a level 7, classical roots and affixes, which they state need to be taught explicitly to both L2 learners and native speakers. Low frequency affixes were omitted from levels 1–7.

Table 3 shows frequency counts derived from the 1,000,000 token Lancaster-Oslo-Bergen (LOB) corpus. 38% of the tokens in the corpus (378,519 out of 1,000,000) contain the inflectional suffixes added at level 2, while levels 3–6 account for an additional 6% of tokens. The remaining 56% of words are base forms—including proper nouns and compound nouns—and words with affixes that are not included in levels 2–6.

As demonstrated by the results in Table 3, the benefits of using word families are obtained by including a few common affixes, with diminishing coverage for additional lower frequency affixes.

The creation of Nation's lists was "largely hard manual work".[4] This thesis investigates the extent to which word family lists for other languages can be automatically created and compares the distribution of tokens with that of English.

2.4.4 Types of vocabulary

A typical literate English speaking adult knows roughly 20,000 English word families (Nation, 2001, pg.9). Nation states that as a rough rule of thumb, native speakers learn 1000 word families a year during the early years of their life. Cobb (2007) proposed that learners of English must at least know the most frequent 3000–5000 word families in order to read text written for native English speakers effectively.

Nation (2001) categorises words as high frequency, specialised, technical and low frequency.

High frequency. A small proportion of words cover a very large percentage of written and spoken language use. The top 10 word families account for about 25% of written language, the top 100 account for 50%, and the top 1000 account for 75% (Nation, 2001, pg.15). Nation (2012) calculated the coverage of his word family

[4] Personal email correspondence with Paul Nation, 16/05/2013.

Table 4: Word list coverage of the Wellington Written Corpus (Nation, 2012)

Word Family List	Coverage
1000	75.22
2000	8.91
3000	5.22
4000	1.75
5000	1.06
6000	0.71
7000–25000	2.02
Proper nouns	3.02
Abbreviations	0.30
Marginal words	0.41
Transparent compounds	0.13
Not in the lists	1.25

lists using the Wellington Written Corpus. As shown in Table 4, the study indicated that the second 1000 covers 8.9% and the third 1000 covers 5.2%. Coverage falls substantially with the addition of each 1000 word families.

The top 2000 word families are generally referred to as "high-frequency" vocabulary that are important for all learners to know. However, the high-frequency threshold could equally well be drawn at 1000 or 3000 because the third 1000 words add considerably to the coverage, and there is a big drop in coverage for the fourth 1000, as shown in Table 4. For example, Ogden's (1944) Basic English word lists classify the top 850 and 1500 English word families as high frequency.

The top 1000 word families list contains about 165 function words (*a*, *some*, *because*, *to*); the rest are content words—nouns, verbs, adjectives and adverbs.

Specialised vocabulary. These are words with a high frequency in a restricted domain. Learners can increase their vocabulary coverage of text from specific domains more efficiently by studying specialised word lists, as opposed to general ones (Nation, 2001, pp. 17–19).

One popular specialised word list is the Academic Word List (AWL), which contains 570 word families that occur frequently in a range of academic texts, but are not in the top 2000 word families. The academic words covers up to 8–10% of words in academic texts, compared to the 4.3% coverage yielded from the third

1000 word families. The academic word list is less useful for readers of fiction, with only 1.7% coverage (Nation, 2001, pg.16-17).

Specialised vocabulary that relates to specific domains, such as economics or physics, are often termed *technical vocabulary*. Billuroğlu, Eldridge, Hancioğlu, and Neufeld (2008) argue that "the distinction between general and so-called academic lexis is not clear-cut enough to sustain an Academic Word List attached as a third band to the GSL." Instead, they suggest that the top 2000 word General Service List (GSL) needs refinement. New versions of the GSL and AWL were recently published (Brezina & Gablasova, 2013; Browne, 2014; Gardner & Davies, 2013).

Low frequency. A large number of words occur very infrequently. The 20,000 word families in lists 6–25 account for only 2.7% of the words in the Wellington Written Corpus (Bauer, 1993), and the 20,000 words in the proper noun list account for an additional 3.0%. Nation (2012) estimates that the remaining 1.25% of words that do not appear on any list include 50,000 new words and130,000 proper nouns. The boundary between high, mid and low frequency words is arbitrary and depends on the application. Hence, what is considered low frequency and technical vocabulary depends on the individual interests and background of the learner.

It is clear from previous research that learning the 5000 words required for enjoyable reading of authentic adult text (Hirsh & Nation, 1992) is overwhelming for L2 beginners, and even learning the top 1000 words is a daunting task. This thesis develops a computer-assisted reading system that helps learners who know the top 200 or so word families to focus on specialized vocabulary—high-frequency words that relate to individual interests—as an efficient and enjoyable path to fluency.

2.4.5 Frequency lists

Frequency analysis plays a key role in developing computer systems that aid efficient L2 learning. Nation (2012) sorted the top 25,000 English word families by frequency, into 25 lists of 1000. The first list is the 1000 most frequent families; the second is the next most frequent 1000, and so on. The top 2000 lists were created using a specially-designed 10M token corpus comprised of 6M words of spoken and 4M words of written English from the American National Corpus (AmNC) and

Table 5: The corpus used for the first two 1000 word family lists (Nation, 2012)

	US		UK/NZ	
Spoken	AmNC spoken 1	1,107,602	BNC 1	1,036,097
	AmNC spoken 2	1,029,831	BNC 2	1,125,523
	Movies and TV	1.000,000	BNC + WSC	1,132,620
Written	AmNC written 1	1,145,081	School journals	1,028,842
	AmNC written 2	939,407	BNC fiction	1,040,204

British National Corpus (BNC), with an additional 1M words from movies and television, and 1M from school journals. Table 5 describes this corpus.

Lists 3–25 were derived from combined frequency analysis of the Corpus of Contemporary American English (COCA) and the BNC. The reason a special corpus was used for the top 2000 is that the COCA and BNC are biased towards formal language. Earlier top 2000 lists produced by Nation (2004) solely from the BNC did not include words common in informal English, such as *alright, pardon, hello, dad, bye*.

The top 25,000 word family lists exclude several special classes of words. These words are included in four supplementary lists. The first special list is an ever growing list of over 22,000 proper nouns. These are put in a separate list because researchers generally assume that proper nouns present a small learning burden and are easily understood (Chujo & Utiyama, 2005; Nation, 2006; Schmitt, 2008; Webb, 2010). However, Brown (2010) questioned this assumption, calling for a careful examination of the impact of proper nouns on reading comprehension by low-ability learners.

The second special list contains 196 marginal words such as profanity, exclamations and letters of the alphabet. The third special list contains 1149 abbreviations. Abbreviations and marginal words fall into a similar category as proper nouns in that they usually cause little problems for learners.

The final special list contains 6044 transparent compound words. These words consist of two or more words for which the meaning is usually obvious to people familiar with the constituent words, for example *headfirst*, *wrongdoing*, *halfhearted*. Each of these has three variations in format, for example *headfirst*, *head first*, *head-first*. These words pose little difficulty during reading if the constituents are known.

Nation developed vocabulary profiling software that takes a given text and counts the percentage of words that occur in each word family list (Laufer & Nation, 1995)[5]. One shortcoming of the software is that it does not distinguish homonyms. For example, *bear* (the animal) and *bear* (to carry something) are combined into a single family. In some cases homonyms are split into separate families. For example, in the case of *March* (the month of the year) and *march* (as soldiers do), *march* is one family and *marched, marcher, marchers, marches, marching* are put into a separate family.

Where one of the homonyms is a proper noun, an alternative approach is to use capitalisation as a distinguishing factor. In practice, one usage of a homonym is often far more frequent than the other, so it makes sense to assign that word to the most frequent usage.

This thesis utilises Nation's English word family and frequency lists, and creates Spanish lists using a similar process. These lists are used to aid the automatic selection of target words for vocabulary exercises, and to compute individualised reading difficulty scores for texts on the internet.

2.5 Benefits of extensive reading

The benefits of extensive reading make it an important part of language learning (Day & Bamford, 1998; Nation, 2009; Renandya, 2007). This section describes the learning effects from extensive reading, and its impact on attitude and motivation.

2.5.1 Vocabulary learning

Krashen's (1989) review of the L1 learning literature suggests that much of the vocabulary knowledge of a native speaker is acquired incidentally through large quantities of listening and reading. Nagy and Anderson (1984) estimate that an average school age child reads 1,000,000 words per year, resulting in 3000–4000 encounters with new words; enthusiastic readers read up to 50,000,000 words per year. New words are learnt implicitly by guessing meaning from context and occasionally consulting dictionaries or others with more advanced language ability. That knowledge is enriched by repeated subsequent exposures.

[5] Available at http://www.victoria.ac.nz/lals/about/staff/publications/BNC_COCA_25000.zip

L2 learners also acquire vocabulary knowledge through extensive reading, however, as Poel and Swanepoel (2003) noted, implicit learning is a slow and error-prone process, especially for beginners. Studies report a wide range of vocabulary acquisition rates from reading and disagree on size of the gains relative to direct vocabulary study (see reviews by William Grabe, 2009; Huckin & Coady, 1999; Nation, 2009; N. Schmitt, 2008; Swanborn & de Glopper, 1999). These results have implications on the role that extensive reading takes in a balanced L2 learning program.

The large variation in acquisition rates is partly due to the varying quantity of text and differing pre- and post-test administration procedures. For example, Pitts et al. (1989) had 35 adults learning English as a second language (ESL) read for just 60 minutes. A multiple-choice post-test indicated that they learnt the meaning of 6.4% of 28 target words. In contrast, Elley and Mangubhai (1981, 1982) achieved dramatic improvements in a range of language skills by having 300 Fijian primary school students read for 20–30 minutes per day for two years. At the end of each year they administered standardised reading comprehension, listening comprehension, vocabulary, grammatical structures and writing tests.

Incidental vocabulary learning from reading is a complex process. The rate of acquisition depends on learners' background knowledge of text subject matter and their general language ability (Horst, 2001; Zahar et al., 2001). The acquisition rate varies by part of speech (Kweon & Kim, 2008), and depends on conceptual difficulty and length of words (Nagy, Anderson, & Herman, 1987); number of repetitions (Rott, 1999; Saragi et al., 1978); richness of surrounding context (Elley, 1989); and whether or not learners have access to dictionaries and glosses (Hulstijn, Hollander, & Greidanus, 1996).

These complexities present challenges in designing an experiment that controls all the necessary variables and includes a test instrument sensitive enough to measure the incremental learning that results from reading. Many studies measure vocabulary gains in terms of previously unknown target words becoming "known" in a test administered after a period of extensive reading. Target words are often nonsense words, to ensure that they are unknown prior to the treatment. The resulting low rates of incidental vocabulary learning are unsurprising, because the

studies do not include enough spaced encounters that are required to learn many words.

Extensive reading texts ideally contain 2% unfamiliar words (Nation, 2009, pg.51), that is a 1000 word text would have 20 unknown words. Given that the percentage of unknown words is controlled at 2%, and assuming six encounters result in solid receptive knowledge, the maximum rate of learning is 20 / 6 = 3.3 words in 1000 words of reading. In practice, when learners read one million words of text, many unfamiliar words are encountered fewer times than is required to achieve solid receptive knowledge. Furthermore, the time intervals between encounters may be so large that words are almost forgotten, which results in less effective learning.

As demonstrated in this thesis, computer algorithms can help learners select texts that increase exposures to partially known words, but still contain no more than 2% unfamiliar words.

Given that extensive reading texts have such a small number of unfamiliar words, one may think that a higher proportion of unknown words would increase vocabulary learning. However, increasing the percentage of unknown words increases the exposures to unknown words at the expense of reduced comprehension and reading speed, which reduces learners' ability to infer the meaning of unfamiliar words from context. Consequently, an increase from 2–5% unknown words can actually reduce the rate of vocabulary learning, and negatively impact enjoyment and motivation.

A major benefit of extensive reading is not learning new words, but enhancing existing vocabulary and grammar knowledge. Studies that use more sensitive test instruments have demonstrated significantly higher learning of form, meaning and productive vocabulary knowledge (Horst, 2005; Horst & Meara, 1999; Kweon & Kim, 2008; Pigada & Schmitt, 2006) than those that use less sensitive test instruments. Waring (2006) claims that learners do not have time to explicitly learn the vast number of collocations associated with even the high-frequency vocabulary. He claims that implicit learning through extensive reading is a solution to this problem. Nation (2001, p. 339) also cites this as a good approach to collocation learning.

2.5.2 Reading speed gains

In comparison to those linking vocabulary gains and improved comprehension to extensive reading, there are fewer that investigate gains in reading speed and comprehension. While some of the methodologies used make it difficult to quantify reading speed gains, the general outcome is consistent across all studies.

At a Japanese college, Robb and Susser (1989) divided 124 first year English majors into an extensive reading group, and one that worked through a skills-building textbook. Based on dubious statistical analysis, the year-long study reported superior reading speed gains from extensive reading. The extensive reading group reported reading for 99 minutes per week at home compared to 35 minutes per week for the skills group. As pointed out by the authors, the data in the study does not indicate which approach was more time efficient, but did indicate that the students preferred the extensive reading approach.

In a 4 week study that divided 266 secondary students into three treatment groups, Lai (1993) reported significant gains. The author stated that the emphasis of the teacher impacted the quantity of reading performed and subsequent gains in reading speed. As with Robb and Susser's study, this study used different texts for the pre- and post-tests which made it difficult to reliably quantify the gains in reading speed.

In another study that compared extensive reading with a more intensive "traditional class", Mason and Krashen's (1997) 114 subjects read the same text at the beginning and end of the academic year. The two extensive reading groups had superior comprehension, and reduced reading time by 54% and 47% in the post test, compared to 17% for the traditional group.

In Bell's (2001) two-semester long study, extensive and intensive reading groups read the same text in pre- and post-tests. The intensive group increased reading speed by 18% and the extensive reading group increased by 87%. Likewise, the extensive reading group made significantly superior gains in comprehension. The main shortcoming of this study was it included only 26 participants.

Shue (2004) compared the performance of three groups of Taiwanese high school students that performed extensive reading with graded readers, extensive reading with books for English native speaking children, and control group that did traditional intensive reading and language exercises. The three groups had

comparable increases in reading speed, but the extensive reading groups had far superior performance in the post test relative to control group. The authors commented that because the learners had only one prior year of English study, the extensive reading texts were too difficult, which may have prevented students from achieving greater speed gains.

In studies with Japanese high school students, Tanaka and Stapleton (2007) demonstrated gains in reading speed and comprehension relative to a control group, and Iwahori (2008) reported gains in speed and comprehension during a seven week study that lacked a control.

Finally, in a 10 week study with 70 Arabic college students, Al-Homoud and Schmitt (2009) produced a 56% increase in reading speed by an extensive reading class, compared with a 42% increase in speed by an intensive reading class. Both groups had comparable increases in comprehension.

Other studies report improvements in writing, spelling (Lai, 1993; Pigada & Schmitt, 2006), general reading ability (Elley & Mangubhai, 1981, 1982) and overall L2 proficiency.

These studies give conclusive evidence of the impressive speed gains that result from effective extensive reading relative to intensive reading. The studies measure gains by having learners read the same text as a pre- and post-test and record either the length of time to read the entire passage or the number of words read in one to three minutes. This thesis presents an approach to measuring speed gains by continuously monitoring reading speed for every article read with FERN.

2.5.3 Improved attitude and motivation

In Nuttall's (1996) virtuous cycle, motivation to learn is both a key requirement for success, as well as benefit of performing extensive reading. Effective extensive reading—fast easy reading of large quantities of text that relates to individual interests—reinforces motivation through positive learning experiences that influence future expectations of success.

The psychology of attitude and motivation is backed by research from actual classroom practice of extensive reading. Surveys of classes (Al-Homoud & Schmitt, 2009; Barfield, 1995; Hitosugi & Day, 2004; Pisit, 2005; Powell, 2005; Sheu, 2004) and diary studies (Harvey, 2008a, 2008b; Leung, 2002; Nishino, 2007)

report impressive improvements in attitudes towards L2 reading. As learners begin to read text that is interesting and easy to understand, their confidence in their own reading ability increases and they are motivated to read more.

Studies also show that students enjoy extensive reading and learn new information (Al-Homoud & Schmitt, 2009). This resulting increased motivation to learn provides good support for extensive reading as a sustainable and enjoyable route to learning and personal growth.

2.6 Extensive reading materials

One of the major challenges with extensive reading is sourcing large quantities of reading material that is both interesting and at the right level of difficulty. This section discusses the pros and cons of various types of reading material.

As previously stated, learners should be familiar with 98% of the words in a text used for extensive reading, which requires knowledge of at least 5000 words for pleasurable reading of adult authentic text. It follows that adult authentic text is inappropriate for all but the most advanced learners. For this reason, beginner and intermediate learners are encouraged to read *graded readers*.

2.6.1 Graded readers

Graded readers are stories written specifically for L2 learners, or adapted from authentic text. English language classrooms throughout the world make use of over four thousand graded readers of varying difficulty and various genres (see lists in Hill, 2001, 2008).

Graded reader series are grouped into levels based on the vocabulary they contain. For example, the Oxford Bookworms series is divided into seven levels, from books restricted to 400 high frequency words, to those that utilise 2500 frequent words (Nation, 2008, p. 52). Learners determine at which level to start reading, either by trial and error, or by taking a test designed to measure receptive vocabulary knowledge, such as those developed by Schmitt, et al. (2001) and Nation and Beglar (2007).

One drawback of graded readers is the time and expense required to create them. At a purchase price of £2.50 for a short English reader (Hill, 2001), significant financial investment is required to purchase a school library that caters to a wide

variety of interests and abilities. The Extensive Reading Foundation ("The Extensive Reading Foundation's Guide to Extensive Reading," 2011) recommends purchasing at least 120 books for a school with a 120 learners, and expanding to 500 books over time in order to better cater to a range of abilities and allow students more choice. The cost of graded readers is generally prohibitive for self-directed learners, because they are not shared by an entire class.

Published graded readers are widely available for young people learning English due to strong global demand from schools and universities. The demand and consequent availability of graded readers in other commonly taught European languages like French and Spanish is much lower than English (de Morgado, 2009; Harvey, 2008b). For languages less commonly taught at schools and universities, availability of graded readers for L2 learners is limited.

There are other downsides to graded readers. Some studies report that even in a large library, some students struggle to find interesting texts, especially those who prefer non-fiction, because a large percentage[6] of graded readers are fiction (Hill, 2013). Furthermore, many series are aimed at teens and young adults with themes that are less likely to appeal to older learners.

Critics claim that the language in graded readers is often stilted or unnatural (Yano, Long, & Ross, 1994). Berardo (2006) claims that the language in graded readers is "artificial and unvaried", "unlike anything that the learner will encounter in the real world" and often "does not reflect how the language is really used". However, he does concede that simplified texts are "useful for preparing learners for reading 'real' texts."

2.6.2 Authentic texts

Berardo and others advocate the greater use of authentic texts—those written for native speakers—rather than simplified texts. Day and Bamford (1998, pp. 98-105) list several types of authentic text as viable alternatives to graded readers. These texts include children's and young adult books, newspapers, magazines and comic books. Children's and young adult books are good for beginner and intermediate learners because of the simpler language; however, much of the content is less

[6] As at February 2015, 20.9% of the graded readers listed in Moodle Reader are nonfiction.

appealing to adults. Comics contain a high ratio of pictures to words, and only appeal to certain types of people.

The biggest challenge is that unaided extensive reading of authentic adult material is beyond the linguistic competency of beginner learners. For example, newspapers and magazines are generally suited for upper intermediate and advanced adult learners, due to the difficult vocabulary and grammatical structures they contain.

On the other hand, authentic text is attractive because so much of it is freely available on the Internet on multilingual encyclopaedias, news sites, fictional works, blogs, magazines and journals. As described by Berardo (2006), other advantages of authentic text are that learners are exposed to “real” rather than the artificial simplified language present in many graded readers. Furthermore, Berardo claims that learners are more motivated when reading authentic text, and that it is a good source of cultural learning.

2.6.3 Dictionaries

Ideally, extensive reading is performed using texts that are understood without using a dictionary, which is often the case with graded readers.

When beginner learners start reading authentic text they encounter much more than 5% unfamiliar words, which makes comprehension difficult without dictionary use. Nevertheless, learners should be trained to avoid frequent dictionary use, and only look up words that occur multiple times and are key to gaining a general understanding of the text.

There are two main types of dictionaries, monolingual and bilingual. In monolingual dictionaries words and their definitions are in the same language, these are useful for native speakers and advanced learners. On the other hand, beginner and intermediate learners tend to favour bilingual dictionaries which give L1 translations and definitions for L2 words.

2.6.4 Other reading aids

As a reading aid an alternative to dictionaries is glosses, which are short L1 or L2 definitions that appear either next to difficult words, in the margin, or at the end of a text. As with dictionary use, glosses enable learners to comprehend more difficult texts than is possible with unaided reading. Glosses have the added benefit of being less disruptive to reading than dictionaries (Leffa, 1992), which reduces reading

time (Hulstijn et al., 1996; Nation, 2009). This benefit is particularly true for ambiguous words, for which a dictionary lists all possible meanings and learners must determine the correct one from the context, whereas, a gloss provides only a single meaning.

Another reading aid is elaboration, which is a text simplification technique that which involves rewriting texts by inserting explanations of difficult words. For example the following sentence was elaborated by inserting the underlined words in order to explain the words in bold:

> The first significant attempt to emulate or copy the ancient (very old) Olympic Games was the L'Olympiade de la République...

There are no published elaborated texts, but small scale studies—using texts prepared by the researchers—indicate that they increase text comprehension (Kim, 2006; O'Donnell, 2009; Urano, 2000; Yano et al., 1994).

While techniques like glossing and elaboration enable learners to do effective extensive reading of texts containing more than 2% unknown vocabulary, both require a significant amount of work to prepare the texts manually. As discussed in Chapter 6, optimal glossing of paper-based texts requires consideration of the target reader's ability, to ensure that all unknown words are glossed and superfluous glossing of known words is avoided. This thesis overcomes these challenges by implementing an automated approach to glossing any text.

2.7 Extensive reading approaches

2.7.1 Repeated reading

Repeated reading is a fluency development technique that involves reading the same short piece of text multiple times, either silently or orally. This technique is effective for improving both L1 (Anderson, 1999; Samuels, 1979) and L2 (Han & Chen, 2010) reading speed, comprehension and confidence in reading ability.

For repeated reading, learners use texts that are more difficult than typical extensive reading texts. The percentage of unfamiliar vocabulary may fall in the 90–95% range, making authentic text such as news articles useful for this activity. The initial reading is best categorised as intensive reading. When the text is read again the following day, the unknown vocabulary is more familiar and reading speed

increases. During subsequent readings unknown vocabulary is learnt and the reading transitions to extensive reading, and after about five repetitions, to speed reading.

One downside of this approach is that learners may get bored reading the same text five times. However, repeated reading is a useful fluency development activity if used occasionally.

2.7.2 Narrow reading

For three decades, Krashen (Cho, Ahn, & Krashen, 2005; Cho & Krashen, 1994; Krashen, 1981, 2004, 2011) has advocated the use of narrow reading for L2 fluency development. This technique involves reading multiple texts from the same authors, genres or topics that learners find most interesting. For example, Harry Potter fans might read the entire series, or a football fan might read weekly match reports for his favourite team.

Related texts share more common vocabulary than unrelated ones (Hwang & Nation, 1989; Sutarsyah, Nation, & Kennedy, 1994), increasing the frequency of repeated exposures, which are required to learn new words and enhance existing vocabulary knowledge. These advantages are similar to those of repeated reading, with the added advantage of being more enjoyable than reading the same text multiple times. Studies show that narrow reading results in significant improvement in both reading ability and attitude towards reading (Cho et al., 2005; Cho & Krashen, 1994).

Gardner's (2008) analysis of children's texts suggests that theme-relatedness promotes more vocabulary recycling in expository texts[7] than in narrative texts[8]. However, reading thematically-related texts by a single versus multiple authors has a stronger impact for expository than for narrative texts. This analysis implies that theme-related narrow reading is best with nonfiction such as news articles, and author-related narrow reading is best for fictional novels and short stories.

As with repeated reading, narrow reading allows learners to progress from intensive reading of difficult authentic texts to easy extensive reading in a shorter space of

[7] Expository texts use facts, details, opinions and examples to persuade or inform the reader.

[8] Narrative texts tell a story that includes such elements as a plot, theme, setting, characters, conflict and resolution.

time than reading unrelated text. Frequently encountered technical vocabulary is learnt quickly. The more narrow a learner's reading is, the more quickly they progress. Section 8.3 explores these ideas in detail, with quantitative analysis of texts from various genres.

3 Computer-assisted extensive reading

Extensive reading is an effective language learning technique that involves quickly reading vast quantities of interesting text that ideally has less than 2% unfamiliar vocabulary (see Section 2.2.1). Beginner and intermediate learners of English usually perform extensive reading using text prepared especially for learners, such as graded readers. However, these texts are expensive and have limited availability in other languages (see Section 2.6.1). On the other hand, authentic text is widely and freely available (see Section 2.6.2), but is too difficult for low-ability learners without the aid of dictionaries, glosses and pre-reading vocabulary exercises—all of which have disadvantages (see Sections 2.6.3 and 7.1.1).

This chapter presents a historical account of computer-assisted reading, followed by a discussion of challenges with reading on screen versus paper. The chapter then examines six technologies that support reading of authentic text (Chun, 2011):

1. The Internet as a source of text;
2. Tools that aid learners in selecting appropriate texts;
3. Electronic dictionaries;
4. Multimedia annotations that support reading;
5. Vocabulary learning software that targets word recognition;
6. Web-based activities that exercise reading related skills.

This thesis describe how each of these technologies are used by learners and teachers, and reviews research on their effectiveness in supporting L2 reading. Later chapters describe the design and implementation of these tools.

3.1 History of computer-assisted reading instruction

Computers have been used to assist reading since the 1960s, both inside and outside the classroom (Blok, Oostdam, Otter, & Overmaat, 2002).

In the 1960s, researchers at Stanford University designed a computer-aided reading system to facilitate self-paced learning and reduce the teacher's role by automatically tailoring instruction to each individual student's ability (Atkinson & Hansen, 1966). After several years of implementation and field trials, they concluded that the system successfully supported the decoding aspect of reading,

but that teachers should maintain responsibility for the communication aspects such as reading for meaning and enjoyment (Atkinson & Fletcher, 1972; Atkinson, Fletcher, Chetin, & Stauffer, 1971). Despite promising results in field trials with over 3000 kindergarten to third grade students, the program was discontinued, in part due to the high costs of computing.

The PLATO (Programmed Logic for Automatic Teaching Operations) system, built by the University of Illinois in the 1960s, was used for individualised computer-aided reading instruction for both L1 (Suppes, 1979) and L2 learners (Curtin, Dawson, Provenzano, & Cooper, 1976; Weible, 1980). Their elementary L1 reading courses were enthusiastically accepted in 25 classrooms, delivering 17,000 hours of instruction to 1225 students. However, unreliable audio hardware hindered more widespread use of its audio component.

The classroom success of PLATO enabled increased funding culminating in the PLATO IV, which incorporated a touch screen, and communicated over regular phone lines, enabling teachers and students to discuss work online. By 1975 it served 146 locations; mostly schools, medical and government sites. However, commercialisation attempts failed to develop a business model that was cost-effective enough for wide spread adoption by education providers (Van Meer, 2003).

The 1970s saw the introduction of relatively-inexpensive hardware and software, and microcomputers were used in many US colleges by the mid-1980s. Kleinmann (1987) studied computer-aided reading by 76 ESL college students. Each class received two 2-hour lectures and a one 1-hour lab session per week. During six lab sessions, half of the students used commercial software that provided reading instruction, and half used paper-based resources. Kleinmann was not surprised that the paper resources were just as effective as the software, which was “drill-practice and tutorial in nature, amounting to little more than electronic textbooks.”

In late 1980s, much language learning software was developed using HyperCard (Ashworth & Stelovsky, 1989; Blake, 1992; Hult, Kalaja, Lassila, & Lehtisalo,

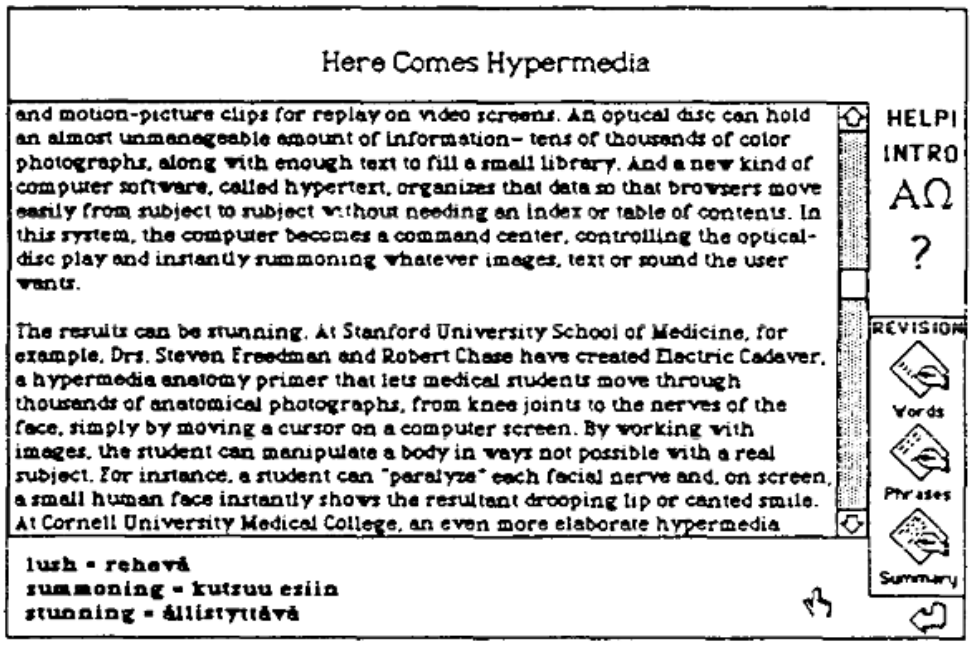


Figure 6: HyperReader with its electronic glossary

1990). For example, HyperReader, developed at Helsinki University of Technology, provided an interactive course in reading comprehension (Hult et al., 1990). It included an authoring mode for teachers and a learning mode for students. As depicted in Figure 6, when students clicked unfamiliar words, glossaries defined by the teacher appeared at the bottom of the screen. Hypermedia incorporated vocabulary translation and synonym practice exercises; an interface that guided students in writing text summaries; and short introductions to each text that incorporated graphics, text, and sound. However, sound was limited by disk space.

Reading applications for the Windows platform, such as the popular children's software series Reader Rabbit, were developed in the 1990s. Other applications used interpreted languages that ran on multiple platforms, such as Windows, UNIX and Mac. The inception of the web in the 1990s provided a popular platform for computer-assisted L2 reading, with growing support for concordances, electronic dictionary and hypertext-glossing (Bowles, 2004; Lomicka, 1998).

Examples from the past two decades are presented in Section 3.4.

3.1.1 Reading on paper versus screen

During the 1980s researchers compared the readability of computer screens to paper, generally concluding that users enjoy reading on screens but paper is easier

and faster (Mills & Weldon, 1987). Online reading ease can be increased by optimising the contrast, colour, font size, font style, page layout, and the paging or scrolling mechanism. Researchers also noted that reading efficiency was reduced by the small amounts of text that could be displayed on the low-resolution displays in widespread use at the time—this is no longer an issue.

Recent studies drew similar conclusions. Tseng's (2007, 2008a, 2008b, 2010) subjects encountered the following difficulties reading web pages in single-page linear form:

- Eye strain from prolonged reading;
- Background colours are often too bright, putting strain on eyes;
- Fonts are often too small and hard to read;
- Line spacing is too small, making it easy to get lost;
- Limited note-taking ability;
- Inferior portability compared to books.

Eye strain is a larger problem with CRT monitors than with modern LCD displays, thanks to higher resolutions, anti-reflective surfaces, and no refresh-induced flicker. It is also affected by display brightness relative to ambient lighting, font size, and line spacing. These problems are largely overcome by well-designed presentation of reading material.

The size, weight and short battery life of laptops were barriers to replacing paper-based books as portable reading devices. However, the advent of tablets and purpose-built e-readers such as the Kindle has effectively solved the portability problem. Seven-inch tablets have battery life of 10 hours or more and are comparable in size and weight to a novel, with the advantage that digital devices can be read in the dark, have adjustable font size, and can hold entire libraries of content with no increase in size or weight.

Recent studies compared iPads with paper in terms of reading comprehension, reading speed and enjoyment (Dundar & Akcayir, 2012; McClanahan, Williams, Kennedy, & Tate, 2012; Stewart, 2012). No statistically significant differences were found in speed and comprehension, but there are significant improvements in engagement and enjoyment for school students.

DeStefano and LeFevere (2007) examined comprehension when reading hypertext documents containing navigational and semantic links. They found that:

> … the presence of links in text introduces decision-making processes and interruptions to reading that can either enrich the reading experience and/or increase the complexity of the comprehension process.

Eye-catching semantic and navigation links add a decision-making process that negatively impact reading comprehension by increasing cognitive demands. The effect is more pronounced for low-ability readers for whom comprehension is a highly demanding task—as is the case with L2 learners reading authentic texts. To minimise cognitive load for such learners, texts should be presented in linear form with minimal visual distractions.

3.2 The Internet as a source of extensive reading material

In addition to increasing L2 ability, extensive reading of text on the Internet increases motivation to read and confidence (Arnold, 2009; Pino-Silva, 2006; Rankin, 2005; Sun, 2003). The Internet's rapidly-increasing volume of blogs, news, and other information websites provide masses of freely-available text in many languages and on every topic imaginable. A prominent example is the collaborative encyclopaedia Wikipedia, which as of February 2015 contains 34 million articles in 287 languages, 45 of which have over 100,000 articles.[9] English has the most with 4.7 million, of which 60% contain at least five sentences (Lucassen, Dijkstra, & Schraagen, 2012); Spanish is tenth with 1.2 million.

Another source of free digital text is Project Gutenberg, which as of February 2015 contains 48,000 fiction and non-fiction e-books that are mostly public domain English texts. Collections of at least 50 books are available for 16 languages—including 3,171 in French, and 640 in Spanish. Over 100,000 additional free digital books are available through affiliated projects.[10]

The contents of Wikipedia and Project Gutenberg are generally considered too difficult for beginner L2 learners—along with most of the text on the Internet. For

[9] meta.wikimedia.org/wiki/List_of_Wikipedias#Grand_Total (20/8/1013)

[10] See www.gutenberg.org (20/2/2015)

effective extensive reading, learners need texts of a suitable length that match their ability, target language and interests. This section describes tools and resources for efficiently locating interesting L2 content on the Internet. Section 3.3 addresses the challenge of locating text of suitable difficulty.

3.2.1 Selecting reading material of suitable length

Many pages on the Internet are unsuitable for extensive reading because they contain few well-formed sentences. Heilman, Zhao, Pino and Eskenazi (2008) classified web pages as suitable if 85% of the text elements contained at least one noun and one verb as judged by a part of speech tagger; this approach rejects many short but suitable texts. This thesis proposes that web pages that contain a block of at least 150 words of well-formed text are suitable. Reading shorter texts is inefficient because too much time is spent choosing texts.

Standard web search engines help locate specific information, but these are not necessarily suitable for extensive reading. For example, a Google search for *Samsung Galaxy* (on 7/6/2013) returns ads for mobile phone carriers, an auction site, four Samsung product pages, four Wikipedia pages, three news articles, and two e-commerce product pages. Of these, only the news articles and one of the Wikipedia pages contained a block of more than 200 words of text.

Because standard search engines do not filter by text length, specialised search engines are more useful for locating extensive reading material. Google offers search engines for news, blogs, books, academic papers and patents. These tools allow learners to capitalise on Google's much-vaunted efforts to organise the world's information. News and blog searches are most useful for beginner and intermediate language learners, whereas the difficulty and length of books and academic papers makes them better suited to advanced learners.

3.2.2 Selecting interesting target language reading material

There are three approaches to locating interesting target language reading material: searching, browsing and subscribing. For example, Google news allows users to perform keyword searches over 70 country and language editions. Spanish learners choose from Argentina, Chile, Colombia, Cuba, Mexico, Peru, Spain, Estados Unidos (United States) and Venezuela to search for articles from the country they are most interested in. Google news users can also browse the latest news in 10

categories including business, technology and sports, or subscribe to RSS feeds that notify them when articles are published that relate to individual interests.

L2 website directories can help locate sites relating to individual interests. For example, EasyNewspapers.com[11] lists 14,083 newspapers organised by country and state; the Internet Public Library[12] lists online magazines in addition to news.

Limited vocabulary knowledge hinders learners from quickly navigating, composing queries and choosing reading material using L2 search engines and directories. Google Translate and browser extensions like Google Dictionary help.

3.3 Selecting easy to understand reading material

For the purposes of language learning, authentic text on the Internet suffers from the same problem as authentic paper-based text—it is written for native speakers and is generally too difficult for low-ability learners. This section discusses various approaches to locating material on the Internet that is easy enough for L2 learners.

3.3.1 Websites for language learners

A small number of sites provide text specifically for language learners. Simple Wikipedia[13] is a special edition of Wikipedia designed specifically for English language learners. Its articles use short sentences, simple grammar, a restricted vocabulary and limited use of slang and idioms (Besten & Dalle, 2008). The instructions for authors include this:

> Start with Basic English (BE) 850. Let us say that your readers know the BE 850 words. If your writing sounds strange, or is not clear, use a less common word… in BE 1500 or Voice of America (VOA) Special English[14]

Basic English 850 is a list of the most common 850 English word families. Idioms, jargon and words outside the BE1500 should be supplemented by a short explanation in Basic English. The Simple English Wikipedia reached 100,000

[11] http://www.easynewspapers.com

[12] http://www.ipl.org

[13] http://simple.wikipedia.org

[14] http://simple.wikipedia.org/wiki/Wikipedia:How_to_write_Simple_English_pages

articles in May 2013[15]. While, this pales in comparison with the full version, it is nevertheless a useful corpus for learners.

Another useful resource is the *Learning English* section of BBC news[16]. This website provides short news clips that begin with a summary, contain an audio reading of the article, and conclude with explanations of difficult vocabulary.

3.3.2 Readability measures

> In the broadest sense, readability is the sum total (including interactions) of all the elements with a given piece of printed material that affects the success which a group of readers have with it. The success is the extent to which they understand it, read it at optimum speed and find it interesting.
>
> Dale and Chall (1948)

Readability research began in the 1880s with Sherman's (1893) observation that short sentences and concrete terms make text easier to understand. In the 1920s Thorndike (1921) published a 10,000 word frequency list to help teachers choose texts for students. This was followed by publication of several readability formulas for children (Lively & Pressey, 1923; Vogel & Washburne, 1928).

In the 1930s researchers investigated factors that affected the readability of adult texts (Dale & Tyler, 1934; Gray & Leary, 1935). In the ensuing decades dozens of formulas were developed that combined surface features such as word and sentence length (Dale & Chall, 1948; Flesch, 1948; Fry, 1968, 1977; Kincaid, 1975; McLaughlin, 1969; Spache, 1953). These along with more recent sophisticated approaches are described in detail in Section 6.1.1.

3.3.3 Search systems for language learners

The standard approach to locating anything on the Internet is to use a search engine. This section describes several approaches that apply readability research to develop search engines especially for language learners.

---

[15] http://simple.wikipedia.org/wiki/Special:Statistics

[16] http://www.bbc.co.uk/worldservice/learningenglish/

3.3.3.1 Microsoft research search system

Microsoft Research investigated models and algorithms that incorporate text readability in personalised ranking of web search results (Kevyn Collins-Thompson, Bennett, White, de la Chica, & Sontag, 2011). They broke the problem into three key steps:

1. Estimating the reading level of entire documents and result snippets.
2. Estimating the reading proficiency of learners.
3. Ranking documents based on the learner's ability.

They analysed logs of search query sessions and results from their commercial search engine Bing. The reading level of search result snippets and the corresponding documents was computed using a statistical language model (Kidwell, Lebanon, & Collins-Thompson, 2011). Learners' proficiency can be estimated based on the average reading level of documents they consult for a reasonable length of time during search sessions. This estimation could be improved by analysing the difficulty of sites that each user frequently visits.

The final step used the LambdaMART ranking algorithm (Q. Wu, Burges, Svore, & Gao, 2010) to demonstrate the effectiveness of using features relating to the reading level of the results snippets and document to provide personalised search result rankings. While their study focused on applications to searching by children, the same lessons apply to L2 learners.

3.3.3.2 Google search reading level

In 2010, Google Search added a reading level filter that classifies results in three levels: basic, intermediate and advanced, as shown in Figure 8. Each search result snippet includes the reading level, where available. Clicking on each level displays the results at that level. Google classifies texts using statistical models derived from manual judgements by school teachers. The "basic" level corresponds roughly to elementary school reading, "advanced" is technical and scholarly articles, and "intermediate" is everything in between.[17]

As Figure 7 shows, there are a much higher proportion of simple articles on the topic of *extensive reading* than for *computer-assisted language learning*. Almost

[17] http://searchresearch1.blogspot.com

Figure 8: Google search reading level filter

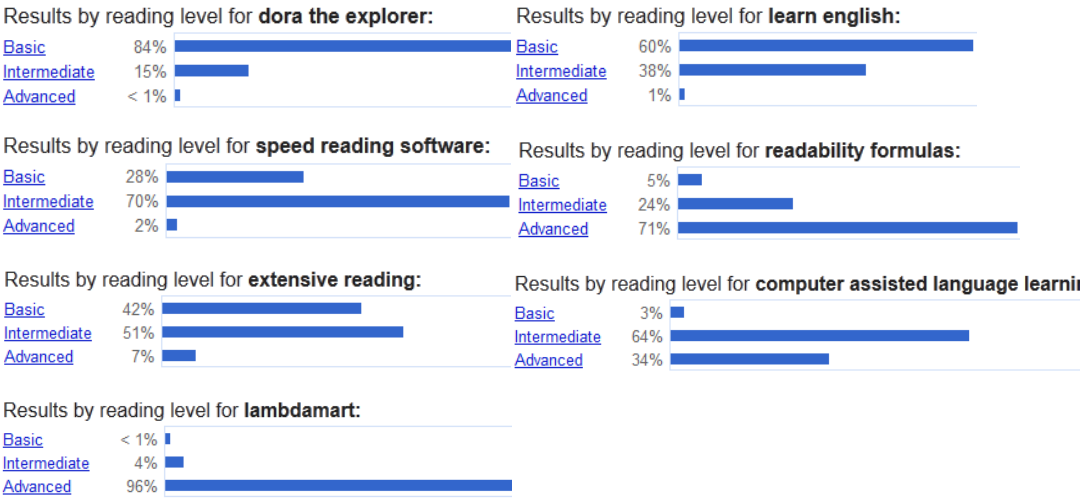


Figure 7: Reading level results for a range of Google searches

three quarters of the articles on *readability formula* are advanced as are almost all the articles on *LambdaMART*—not surprisingly.

3.3.3.3 REAP Search

The REAP project (an acronym for REAder specific Practice) developed a search engine that helps learners locate and read texts that match their interests, vocabulary knowledge and reading ability (Kevyn Collins-Thompson & Callan, 2004). REAP searches a digital library of several million English language documents, sourced

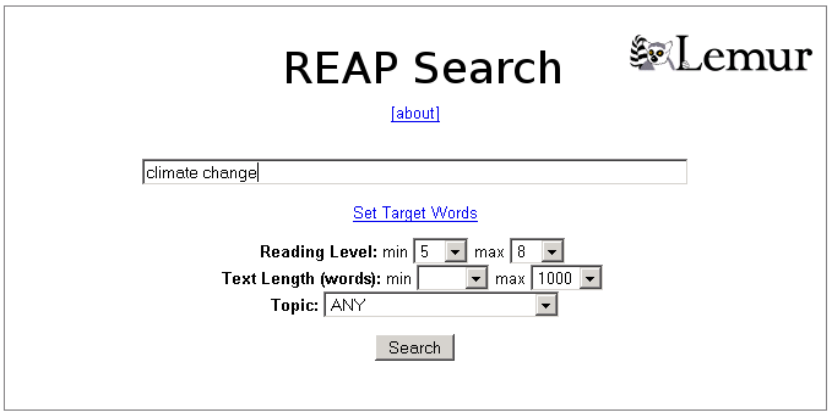


Figure 10: The REAP search interface

Figure 9: REAP reading interface

via a commercial search engine [18] by issuing queries consisting of target words in an ESL syllabus. Documents are algorithmically annotated with the following metadata:

- Reading difficulty—computed using a language modelled readability measure (Collins-Thompson & Callan, 2005; Heilman, Collins, & Callan, 2007; Heilman et al., 2008);
- Topic area—using binary topic classifiers (Heilman, Juffs, & Eskenazi, 2007);

[18] Initially AltaVista.com

- Text quality—using part-of-speech tagging[19] to compute the proportion of text chunks that contain at least one verb and one noun and are thus "well-formed" sentences;
- Text length—total words in the text.

The library is filtered to remove documents that are deemed inappropriate for reading because they are too short, too long, too difficult, or consist of less than 85% well-formed sentences. The remaining 1% of documents are stored and indexed with Indri—a full-text search engine that is part of the Lemur toolkit[20].

Using the interface in Figure 10, learners search for documents that contain target words. They restrict the search by specifying a readability range, text length range or topic. Once learners select an article, it is displayed in the reading interface in Figure 9, which provides dictionary definitions of words that learners click on. The system provides post-reading vocabulary exercises for further practice of target words. This thesis applies similar techniques with a focus on effective extensive reading, as opposed to learning target words.

3.3.3.4 Read-X Project

Read-X is part of a research effort to develop readability formulas sensitive to an individual learner's prior reading experience (Miltsakaki, 2009; Miltsakaki & Troutt, 2007, 2008). Two separate tools have been developed: Read-X and Toreador.

The Read-X tool augments web search results with readability scores and thematic categorisation. This process requires sending user queries to the Yahoo! Search API (now Yahoo! BOSS)[21], downloading the resulting list of web documents (PDF, Word and HTML), and extracting their content with open-source screen-scraper software. It computes scores with three popular readability formulas, and classifies each document into eight top-level categories and 41 subcategories using a machine learned classifier. The annotated search results are displayed in a grid in a Java-

[19] Using the OpenNLP toolkit, available at http://opennlp.sourceforget.net

[20] http://lemurproject.org

[21] http://developer.yahoo.com/boss/search/

based desktop application, with links to view the extracted text or original source document.

Toreador is "designed to predict unknown vocabulary for a specific reader and grade profiles" (Miltsakaki, 2009). It asks users to estimate their reading grade level on a scale of 2–13, and indicate familiarity with the topics of sport, literature or science. Users paste text into Toreador and it highlights words it predicts are difficult, based on general frequency counts, and genre specific frequency counts within a user's area of interest.

An evaluation of the system was undertaken, but never published.

3.3.3.5 Other Search Engines

Several other search engines have been designed for second language learners and low-ability readers. For example, LAWSE (Language Aware Search Engine) (Ott & Meurers, 2010) and Squirrel (Nilsson & Borin, 2002) allow learners to search for pages that contain specified keywords and satisfy readability constraints.

Uitdenbogerd (2003; 2006) proposed and developed a system similar to Read-X. She compared the readability of 90,000 web pages with extracts from texts for ESL learners. Readability was measured using the Linux *style* utility, which computes scores for five common readability measures. Results indicated that the readability of at least 8% of web pages falls within the range of ESL texts.

The Textfinder project (Bennöhr, 2007) derived a readability formula that grades the difficulty of ESL texts. This formula also estimates learner's reading ability by grading a writing sample and adding a constant value that accounts for the fact that learners' receptive L2 knowledge generally exceeds their productive knowledge. Learners then search the collection of texts, and are encouraged to read ones with readability scores that match their ability score. They rate the difficulty of texts they read, and this data is used to update ability scores.

The Intelligent Web-Based Interactive Language Learning system, IWiLL (Wible, Kuo, yi Chien, & Wang, 2000), contains a "lexical difficulty filter" that uses a list of high-frequency words to identify documents whose proportion of off-list words exceeds a user-defined threshold.

Miyazaki and Norizuki (2008) developed a system that helps teachers and students create readability formulas that resemble the Flesch Reading Ease, but with coefficients tuned to individual needs, and use them to search the NHK (Japan Broadcasting Corporation) news site for English articles of appropriate difficulty. The coefficients are derived from learners' self-rating of article difficulty after reading, and from their performance in automatically created cloze tests.

NetTrekker[22] is a commercial search engine designed to help school students and teachers locate material that is safe (i.e. filter sexually explicit content etc.), age appropriate (the content is interesting for students) and relevant to school studies. Its search engine uses a database of links to web pages that are manually chosen and organised by human experts. Their system grades text difficulty on a scale of one to five, where one is suitable for US grades one to three and level five is suitable for grades 11–13.

Expert manual classification of text would undoubtedly be useful in systems designed for L2 learners, but many experts would be required to grade the large numbers of texts that are needed for extensive reading by a classroom of learners.

This thesis builds upon this research by developing a search engine for language learners that is built on commercial search technology, incorporates readability measures and vocabulary frequency lists, and most importantly learner modelling.

3.4 Computerised dictionaries

As described in Section 2.6.3, extensive reading advocates recommend that learners read texts they can easily understand without dictionaries, because the use of paper dictionaries breaks the flow of reading. This recommendation rules out the use of authentic texts for extensive reading by low-ability learners. However, computerised dictionaries provide fast lookup during reading, which results in increased reading speed (Aust, Kelley, & Roby, 1993), comprehension and vocabulary learning (Li, 2010). This section describes the pros and cons of various types of computerised dictionaries.

[22] http://nettrekker.com

3.4.1 Handheld electronic dictionaries

Most handheld electronic dictionaries are targeted at ESL learners. They package one or more translation dictionaries, L2 learner dictionaries, L1 dictionaries, collocation or idiom dictionaries, thesauri and encyclopaedias into purpose-built devices that fit into the palm of the hand—a major advantage over paper dictionaries.

Another advantage is faster lookup time, especially for beginner learners whose first and second languages are written in different scripts (Okuyama & Igarashi, 2007). For example, an English native learning Japanese must learn the dictionary order of Japanese Hiragana before they can use a paper Japanese to English translation dictionary. With electronic dictionaries, they look up words by simply typing them.

Handheld electronic dictionaries provide additional functionality, such custom word lists, hyperlinked cross-references, and instant lookup of unfamiliar words present in dictionary definitions. Most provide wildcard search, which makes it easier to look up words when unsure of the spelling.

Handheld electronic dictionaries are available in fewer languages than paper-based dictionaries, and are typically 100–400NZD, which is much cheaper than they were five years ago, prior to the widespread usage of smart phones.[23]

3.4.2 Mobile phone application dictionaries

Dictionaries for many languages are available as free or inexpensive (up to 25NZD) mobile applications. Software dictionaries are far cheaper than purpose-built devices, and are used on devices learners already carry. Even including the price of a cheap smart phone, the mobile app option is significantly cheaper than most electronic dictionaries, and is also more flexible because it can store podcasts, PDF files, and other language learning media.

The main advantages of purpose-built handheld dictionaries were superior usability and the better quality of the dictionaries. However, the usability gap has narrowed

[23] http://denshi-jisho.com

Figure 11: The Rikaichan Firefox extension available at http://polarcloud.com/rikaichan

over the past decade, with small mobile phone screens and 12-button keypads evolving into 5-inch displays with full keyboards and touch screens.

Another downside of mobile app dictionaries is that there are thousands to choose from, many of rather poor quality. Furthermore, some require an Internet connection, which can be costly and is inconvenient when an Internet connection is unavailable.

3.4.3 Desktop dictionaries

Many websites provide online translation dictionaries. For example, Google Translate provides translations for 90 languages. Wiktionary is both a monolingual and a bilingual dictionary, with entries in 170 languages. Other online dictionaries are language specific. For example, Jisho.org allows learners to search for Japanese words by typing with the English alphabet; words are automatically converted to Japanese characters before the search begins.

Online dictionaries provide fast look-up when reading online texts, because learners can copy and paste unfamiliar words into a search box. This feature provides a significant advantage over handheld dictionaries for those learning the character set and spelling of a new language. For example, learners of Japanese face a two-step process when looking up unfamiliar words: they must first determine the pronunciation of individual characters using a kanji (Japanese character) dictionary, and then locate the word in a regular dictionary. Determining the definition of a word with two or more kanji can take several minutes for intermediate learners, and is essentially impossible for beginners (Okuyama & Igarashi, 2007).

Browser extensions provide rapid access to translations and definitions while reading online text. Many of these use free sources such as Wiktionary and Wikipedia. Figure 11 shows Rikaichan, a plugin that displays English translations when users hover over Japanese words.

Google's dictionary plugin for the Chrome web browser is multilingual. Users can highlight words to display a popup with L1 definitions for L1 words, and L1 translations for L2 words. There is some difficulty with words that are both L1 and L2, and the system would be better for learners if it detected the language of the current text.

Dictionaries are standard features of computer-assisted reading software, but their completeness and functionality varies. When learners click on non-root words, some systems display a clickable cross-reference to the root entry, whereas others save clicking by redirecting immediately to the root entry. Most dictionaries do not offer the useful feature of suggesting alternatives when learners enter misspelled words. As demonstrated in this thesis, these seemingly minor features have a big impact on user experience.

Some free online dictionaries offer paid versions that hide distracting advertisements. Another paid option is iFinger,[24] a windows application that offers access to 60 dictionaries and encyclopaedias while reading web pages or using Microsoft Office applications.

3.4.4 Hypertext dictionaries

Some L2 reading websites provide hypertext dictionaries, which display dictionary entries when learners click or hover over unknown words. These sites also provide review exercises and games that target difficult words.

Lingro.com provides instant dictionary lookup for 11 languages. Learners enter URLs and Lingro displays a modified web page in which learners can click unfamiliar words to view translations from Wiktionary and other free dictionaries, augmented with contributions by users. Gymn@zilla (Streiter, Knapp, & Voltmer,

[24] http://ifinger.com

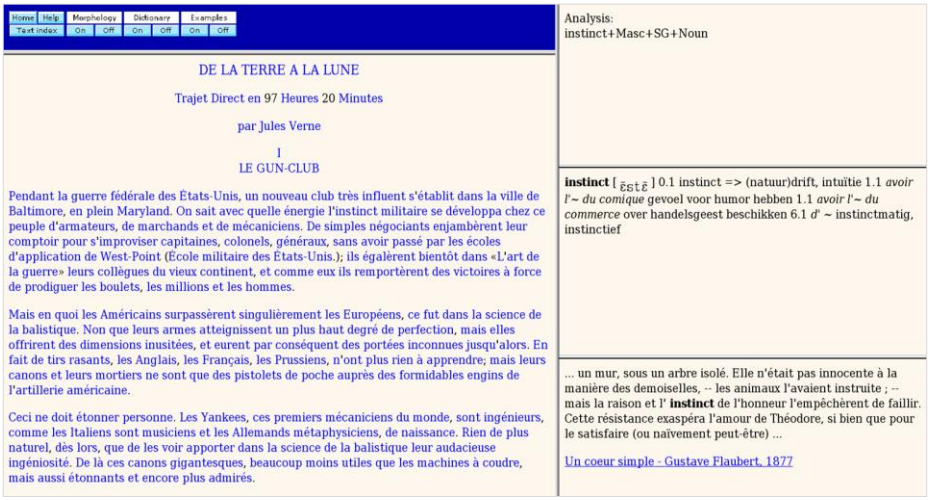


Figure 12: Screenshot from the web-based version of Glosser-RuG

2004), Ultralingua.net and Wordchamp provide similar functionality (Leloup & Ponterio, 2005).

PopJisyo.com and Rikai.com provide dictionary lookup for Japanese, Chinese and Korean. These dictionaries display entries for words and individual characters. Hypertext builder[25] (Cobb, 2007) allows learners of English, French and Spanish to click unfamiliar words in texts to hear pronunciation, and view concordance examples and dictionary definitions. Other systems like e-Lecture[26]—online lecture delivery software—provide translation dictionary look-up as a supplementary feature.

Desktop software Glosser-RuG (Dokter, 1997) and web-based Glosser-WeB (Dokter, 1998) support Dutch speakers reading French texts, but can be easily extended to support other languages. In addition to dictionary translations, the system provides morphological analysis, part-of-speech information, example sentences from a corpus, and audio pronunciation.

3.5 Multimedia glosses

Computer-aided reading systems allow learners to click unfamiliar L2 words to view L1 translations or L2 definitions in a popup box or nearby screen element.

[25] http://www.lextutor.ca/hypertext/

[26] http://www.umediaserver.net/electure/

Studies show that providing glosses improves reading comprehension (Alessi & Dwyer, 2008), and that low-ability learners prefer L1 over L2 glosses (1998). Whereas dictionaries provide multiple definitions for ambiguous words, glosses provide a single disambiguated meaning, which reduces disruption to reading. Creating glossed texts requires either time consuming manual effort or complex computer software to automatically disambiguate meaning. Automatic glossing is the subject of Chapter 5.

3.6 Text simplification

Another way to support reading of authentic texts by L2 learners is to simplify the language, grammar and structure. Such simplification may be performed manually or automatically.

3.6.1 Computer-aided manual simplification

Computer tools are available to help authors write text for learners at a particular ability level. Lexical frequency profiling software, such as VocabProfile (Cobb, 2007) automatically identifies low-frequency words by highlighting words in top 1000, top 2000 and academic word lists; the remaining words are designated as off-list. Authors then manually add glosses to low-frequency words, or replace them with higher-frequency equivalents.

Cobb (2007) also advocated the need for "some way of focusing attention on and proliferating encounters with new words… within the act of reading, or shortly after reading, for any type of text, and for lots of texts." Recognizing that words that appear once or twice are unlikely to be learnt, vocabulary profiling software helps authors identify words across a series of texts that require repetition. A major challenge in authoring text for L2 learners is to write authentic-feeling text that uses simple language, and repeats target vocabulary (Section 2.6.1). This thesis shares the vision of Cobb in the development of its computer-aided reading environment.

Eldridge and Neufield (2009) created a series of 16 sets of five thematically-related 1000 word articles by shortening and simplifying Wikipedia articles, applying frequency profiling to remove low-frequency words. The Flesch-Kincaid Reading Ease formula (see Section 6.1.1) was used to order the resulting texts by difficulty. In a small case study, 30 students read and discussed the articles during a six week programme, making use of Moodle's build-in glossary. Despite encouragement

from the teacher to read regularly, students procrastinated; most of the assignments were completed during the final two weeks—a result that underscores the need to make regularity part of the grade. A concluding survey indicated that most participants found it a valuable learning experience that increased their recognition of the important vocabulary.

3.6.2 Automatic simplification

In order to eliminate time consuming manual adaptation of authentic texts, researchers are developing systems that perform automatic simplification for L2 learners (Petersen & Ostendorf, 2007). Some focus on syntax—splitting long sentences and modifying ones with complex structures (Aluísio, Specia, Pardo, Maziero, Caseli, et al., 2008; Aluísio, Specia, Pardo, Maziero, & Fortes, 2008; Aluisio, Specia, Gasperin, & Scarton, 2010; Siddharthan, 2003, 2006). Others simplify lexically—providing summaries, synonyms and translations for difficult vocabulary (Burstein, Shore, Sabatini, Lee, & Ventura, 2007). Such tools may require human intervention to perform elements of the task for which computers are not yet good enough.

3.7 Supplementing reading with vocabulary exercises

The literature offers strong theoretical support for using pre-reading exercises to familiarise learners with the vocabulary, cultural background and topic of a text. Empirical studies indicate that while these exercises aid comprehension, in order to learn vocabulary most efficiently it may be better to spend that time doing more reading (Alessi & Dwyer, 2008; Webb, 2009). One cause of inefficiency is that these pre-reading exercises often include words learners already know well. This thesis proposes that pre-reading exercises are most efficient when they review only unfamiliar and partially known words that will appear in future reading. Chapter 7 presents an approach that automatically provides this level of individualisation of supplementary exercises.

Post-reading vocabulary study is commonly supported by computer-aided reading systems. For example, in the Hypertext Builder (Cobb, 2007) learners click words to hear pronunciations and view them in a dictionary or concordance, and they alt-click words to review them later in automatically generated quizzes, consisting of

example sentences from the Brown corpus (Francis & Kucera, 1979) with target words blanked out.

Systems such as Apheios, LangQ, ReadLang and Lingro provide similar functionality, allowing learners to select difficult words during reading to review later in automatically generated exercises. For example, ReadLang uses spaced repetition (Section 2.4.2) to schedule efficient review of troublesome words.

3.8 Supporting teachers of extensive reading

Paper-based extensive reading requires a large library of graded readers that are controlled by school libraries or teachers, who must purchase, organise, distribute and collect returned books. This administrative load is greatly reduced by using software to manage book loans, and further reduced by having students read electronic texts. This reduction gives teachers more time to help students with reading. Furthermore, computerised recommender systems can aid teachers in recommending interesting, level-appropriate texts to students (Section 6.1.2).

However, the use of computers may increase orientation time, because students need to learn how to use the software. Teachers have the additional responsibility of providing students with ongoing technical support.

This rest of this section describes how computers help teachers monitor and motivate classroom extensive reading.

3.8.1 Monitoring reading

In high-school extensive reading classes many students do little reading unless they receive a grade, and procrastinate until just before assignments are due (Eldridge & Neufeld, 2009). In order to assign grades, teachers need a reliable and practical method of assessing the amount students read. Requiring students to complete reading logs is time consuming for both students and teachers, and students can easily cheat. The only way for teachers to mitigate cheating is to require in-depth reviews, which increases the time required by both students and teachers to assess.

Regular student interviews by teachers may more reliably assess reading with a smaller time investment. However, there is still a trade-off between the reliability and time to assess, which makes this approach impractical with large classes.

Computers can monitor reading without requiring book reviews or interviews. However, it is not easy to make systems cheat-proof. BeeOasis uses an honour code system, which significantly reduces cheating but does not eliminate it entirely (Conradson & Hernández-Ramos, 2004). Another approach is discussed in the sections that follow.

3.8.2 Progress reports

Computer systems can automatically track reading speed and the number of words read, and provide detailed feedback that is impractical to produce otherwise. They can track how many times learners encounter words, which words they have difficulty with, and estimate the rate of vocabulary learning. Such feedback enables learners to set concrete, measurable goals, and work towards achieving them. PREMA (Okazaki, Nitta, & Hashimoto, 2008) is an example of a system that automatically monitors reading quantities and speeds.

This thesis submits that reporting gradual improvements in vocabulary knowledge helps learners maintain motivation and improves the effectiveness of extensive reading. For example, feedback about reading speed motivates learners to consult dictionaries less, helping them to improve reading fluency.

3.8.3 Moodle Reader

Moodle Reader tracks the number of words read by students in classroom extensive reading courses (T. Robb, 2009). The web-based system embodies a database of quiz questions for over 4000 graded-readers and books for English language learners.[27] After completing each paper-based book, students log in and take a timed multi-choice quiz consisting of randomly-chosen questions. If they score a passing mark, the book's word count is added to their total number of words read.

Quiz questions are submitted and shared by teachers, who claim that most questions are high-quality and provide accurate verification, that is students struggle to pass quizzes for books they have not read but comfortably pass for books they have read.

Moodle Reader has other features to help teachers administer extensive reading courses. Graded reader publishers have different grading systems for their books, and the system aligns these into a standard scale. Students are assigned an initial

[27] Introduction on moodlereader.org as at 10/02/15

reading level by their teacher. They are allowed to take quizzes for books at their current level or one below, and advance after passing a set number of quizzes at their current level. The system helps improve student motivation by incorporating the gamification elements of points (words read), badges (book cover images collected for each quiz passed) and levels (current reading level).

Students can be restricted to taking quizzes every third day. This prevents a last-minute rush before course deadlines, encouraging students to read regularly. Also, students may not retake the same quiz, which discourages students from attempting quizzes after skimming or reading a book's summary. By tracking the IP addresses and times that students complete quizzes, certain types of cheating can be detected, such as students completing quizzes for themselves and their friends on the same computer or adjacent computers in a lab.

While these features make cheating more difficult, students can cheat by handing their laptops to friends to sit quizzes for them; indeed, teachers have witnessed such behaviour.[28] More sophisticated analyses of quiz completions may help flag suspicious behaviour, as is done to detect fraud in banking, and harsh penalties also discourage the practice. However, cheating can only be fully overcome by administering quizzes in a supervised computer lab.

Moodle Reader is a good way of monitoring classroom extensive reading. However, quizzes are only available for popular graded readers, and are only accessible by recognised educational institutions—to prevent unauthorised access and distribution by students. This system is not an option for independent learners.

3.9 Crowdsourced resources

Crowdsourcing is commonly defined as "the use of large groups of individuals by organisations to perform tasks traditionally performed by employees or designated agents", usually over the Internet (Erickson, Petrick, & Trauth, 2012). It has been used to great success in the WikiMedia reference sites, notably Wikipedia and Wiktionary,[29] which allow users to collaboratively edit content using a web browser.

[28] http://groups.yahoo.com/neo/groups/ExtensiveReading

[29] Note that some definitions of crowd sourcing exclude sites like Wiktionary and Wikipedia (Estellés-Arolas & González-Ladrón-de-Guevara, 2012).

Because this project makes extensive use of Wiktionary (see Chapters 5–6), we discuss here the reliability of such resources, and discuss issues that should be considered when using them. While this section focuses on Wikipedia, the same issues apply to other Wikimedia resources.

3.9.1 Wikipedia

The first WikiMedia site, launched in January 2001, was titled *Wikipedia: The Free Encyclopaedia*. As everyone knows, Wikipedia's content is available to end users free of charge and uncluttered by advertising. It can be downloaded under the Creative Commons Attribution/Share-Alike 3.0 license,[30] which permits users to modify and use content even for commercial purposes, with appropriate attribution.

Over the following decade a remarkable number of volunteers have collaborated to produce a tremendous volume of reference material. The English Wikipedia has 17 million editor accounts, and 130,000 users make edits each month.[31] The site contains 23 million articles in 285 languages, and is the most popular reference work on the Internet, ranking sixth in Alexa's most visited websites.[32] As of May 2011, the volunteer army was managed by just 65 paid employees, who are supported by donations.[33]

This success of the open approach is demonstrated by the failure of Nupedia—launched by the Wikipedia founders 11 months prior to Wikipedia—which aimed to be a "highly reliable, peer-reviewed resource that fully appreciated and employed the efforts of subject-area experts, as well as the general public" (Sanger, 2005). Nupedia required a seven step editorial process before articles were cleared for public consumption.[34] Only 21 articles were produced during its first year, in comparison to Wikipedia's open edit policy, which yielded 18,000 articles after one year. While Nupedia produced superior articles, it became neglected due to Wikipedia's rapid growth in content, views and contributors. It was shut down in

[30] http://creativecommons.org/licenses/by-sa/3.0/us/

[31] http://en.wikipedia.org/wiki/Wikipedia:Wikipedians as at 20/02/2015

[32] http://www.alexa.com/siteinfo/wikipedia.org

[33] http://en.wikipedia.org/wiki/Wikimedia_Foundation#Employees retrieved 20/02/2015

[34] http://web.archive.org/web/20010607080354/www.nupedia.com/policy.shtml

2003, at which time 24 articles had passed the full review process and 74 were in progress.

3.9.2 Wiktionary

Wiktionary is a free collaborative dictionary that includes 15 million articles across 170 language editions. It launched in December 2002, by which time sister project Wikipedia had already expanded to 97,000 articles in its English edition[35] and a further 28,000 in 25 other languages, only two years after launching in January 2001. Wiktionary's gradual expansion to other languages began with French and Polish editions in March 2004.

Each language edition of Wiktionary contains entries with definitions written in the edition language. The largest edition is the English edition, which gives five million English definitions for words from 1000 languages (500 languages have at least ten definitions), including 570,000 definitions for English words, and 280,000 for Spanish words.[36] The English edition has entries for four million words. French and Malagasy (the national language of Madagascar) are second and third with 3.6 and 2.7 million entries respectively[37]. An additional 26 editions contain over 100,000 entries, including Spanish with 830,000.

The English Wiktionary also includes definitions of words from historical languages like Old French (5500 entries), constructed international auxiliary languages like Esperanto (110,000 entries), and those with comparatively few native speakers such as Māori (590 entries).

Like Wikipedia, Wiktionary has a *Simple English* edition for learners, which restricts the vocabulary of its definitions.

3.9.3 Problems with crowdsourcing

Wikipedia's free, open, collaborative approach has drawn much criticism. Denning et al. (2005) outlined several risks, including:

- Accuracy. One cannot be sure which information is accurate, especially when references are absent.

[35] http://en.wikipedia.org/wiki/Wikipedia:Multilingual_ranking_December_2002

[36] https://en.wiktionary.org/wiki/Wiktionary:Statistics retrieved 20/02/2015

[37] http://meta.wikimedia.org/wiki/Wiktionary retrieved 26/09/2012

- Motives. Vandalism and violations of the "neutral point of view" policy abound; some contributors push political and commercial agendas; others play practical jokes.
- Expertise. It is difficult to determine the qualifications of contributors, especially anonymous editors who are identified by IP address only.
- Stability. When articles are cited it is not clear whether to read the latest version, which may include corrections, or the version at the time of citing.
- Coverage. Topic coverage is biased towards the interests of the creators, who are predominantly male and technically oriented.
- Sources. Most Wikipedia sources are digitized works, which are often themselves questionable in terms of reliability.

Most of these issues also apply to Wiktionary, whose accuracy of definitions and translations are major concerns for this thesis. Its policy guidelines ask users to contribute translations only when they are certain. However, the interface does not make this point clear, and users can easily add dubious translations. At present, there is little data on their reliability. Another problem is completeness. Common languages like English, French and Spanish have many definitions and translations, but others have limited coverage.

3.9.4 Handling vandalism

WikiMedia's open edit policy invites vandalism, defined as "the addition, removal, or other modification of the text or other material that is either humorous, nonsensical, a hoax, or that is of an offensive, humiliating, or otherwise degrading nature."[38] However, the complete revision history stored on WikiMedia sites makes it easy to reverse, once detected. Vandalism can be detected by readers, or by editors who patrol the *recent changes* page. Some editors add articles of particular interest to a *watch list* and receive email notification of changes. Finally, automatic programmes detect and reverse obvious vandalism (Broughton, 2007).

Wikipedia has additional security measures. Site administrators protect pages by temporarily placing restrictions on who can edit them; these pages are distinguished by a coloured padlock in the top right corner. Wikipedia has seven levels of

[38] http://en.wikipedia.org/wiki/Vandalism_on_Wikipedia as at 15/10/2012

protection,[39] including fully protected articles, which can only be modified by administrators, and semi-protected articles, which can be edited by users who have been members for at least 4 days and have made at least 10 edits.

Between June 2010 and May 2011, Wikipedia trialled adding *pending changes protection* to some articles,[40] which required all changes to be approved by authenticated *reviewers*.[41] During the trial a straw poll showed that the majority favoured implementation in some form,[42] but no consensus was reached and the idea was dropped.

3.9.5 Accuracy and reliability

Although excited by Wikipedia's rapid growth, pundits were sceptical that it would ever be as reliable as curated resources:

> The effect that the openness of this project has on its reliability as a reference resource is problematic. Popular search engines… direct traffic to Wikipedia pages, but users applying common Web site evaluation criteria could find its articles of questionable value.
>
> Remy (2002)

Since its birth in 1768, the Encyclopaedia Britannica has included the writing of thousands of subject experts, including four United States presidents and 100 Nobel Prize winners.[43] Articles are revised by 100 full-time editors. Johnson (2006) describes Britannica as utilizing "the authority of experts", in contrast to Wikipedia which relies on the "nit-picking of the masses". Because there are no credential checks, the site opens itself up to manipulation by fraudsters. For example, a 24-year-old college dropout pretended to be a theology professor and worked his way up to the influential position of administrator by spending 14 hours a day contributing over 20,000 edits, before revealing himself as a fraud (Angeles, 2007).

---

[39] http://en.wikipedia.org/wiki/Wikipedia:Protection_policy

[40] http://en.wikipedia.org/wiki/Wikipedia:Reviewing

[41] http://en.wikipedia.org/wiki/Wikipedia:Reviewing

[42] http://en.wikipedia.org/wiki/Wikipedia:Pending_changes/Straw_poll

[43] http://corporate.britannica.com/about/contributors/w

Stross (2006) was sceptical as to whether an article could be credible without knowing its authors. In response, Wikipedia founder Jimmy Wales remarked that "When people trust an article in Britannica… it's not who wrote it, it's the process." Once an article has been reviewed, editing ceases. In contrast, Wikipedia enjoys unending scrutiny and editing. In the future, people may question Britannica's process:

> People will say, 'This was written by one person? Then looked at by only two or three other people? How can I trust that process?'
>
> Jimmy Wales (in Stross, 2006)

Wikipedia contributors are not just large in number; they are also technically oriented and well educated. A 2011 survey[44] of 5000 random editors revealed that 43% were students, with 26% having a postgraduate qualification and an additional 35% listing college degrees as their highest level of education. 72% of editors reported the ability to read two or more languages, and 51% contributed to two or more Wikipedia language editions. There is a heavy gender bias: 91% of editors are male, and to make matters worse 30% of female editors have made fewer than 50 edits, compared with 18% of males.

Wikipedia has several core edit policies that promote the creation of reliable content. Violations to these policies are questioned and often reverted. These policies include:

- Neutral point of view. Articles should represent all points of view on a topic, in a way that is fair and without bias.
- No original research. Facts, allegations and ideas should come from reliable published sources.
- Verifiability. Readers should be able to check the reliability of information sources.

These policies seek to minimise inaccuracies, but do not prevent them. However, the Britannica process is not without fault. Einbinder (1964) spent five years producing a 390 page book detailing inaccuracies in the 14th edition, published in

[44]http://upload.wikimedia.org/wikipedia/commons/5/51/Editor_Survey_Report_April_2011.pdf -Wikipedia Editors Study

1929. More recently, the journal *Nature* commissioned 50 subject experts to review articles on a range of science articles from both Britannica and Wikipedia, and concluded that Wikipedia averaged four errors per article to Britannica's three (Giles, 2005). Later reviews affirmed Wikipedia's accuracy in other subject areas.

New Britannica Editions were produced at approximately 25 year intervals, making it impossible to keep pace with rapidly evolving fields. In contrast, Wikipedia is a reliable source of information about contemporary topics. Brown (2011) investigated its accuracy and coverage as a source for political science researchers, and found no factual inaccuracies in the past political experience of 230 US Gubernatorial election candidates. He concluded that Wikipedia suffered more from omissions than factual errors, with poor coverage of political events that occurred prior to its creation. These findings mirror results in other areas such as pharmacology (Clauson, Polen, Boulos, & Dzenowagis, 2008).

3.9.6 Crowdsourcing in language learning systems

Several web-based language learning systems utilize crowdsourcing. For example, Lingro.com (see Section 3.4.3) supplements translations from free dictionaries, with those submitted by users. Lang-8 users submit L2 writing and receive corrective feedback from native speakers. Users receive points for writing L2 and correcting L1 text. Corrected writings are used as reading material for other learners.

DuoLingo.com harnesses the efforts of language learners to translate documents on the Internet. Multiple users translate the same sentences from L2 to L1, and are awarded points based on the quality of their translations. The first user to translate each sentence is awarded maximum points as long as it is similar to a machine translation. The quality of translations by later users is judged by comparing them with those submitted by others. Users score additional points by voting on the quality of other users' translations. It remains to be seen whether DuoLingo can attract sufficient skilled users to translate a significant quantity of documents.

3.10 Summary

The Internet is an enormous source of L2 reading material, most of which is too difficult for extensive reading by low-ability learners. This thesis works around this problem by providing three tools that enhance learners' ability to perform effective

extensive reading of texts on the Internet: hypertext glossing, individualised searching, and individualised vocabulary exercises.

Hypertext glosses increase learner's ability to comprehend difficult texts. The main downside to glosses is that their creation is tedious if performed manually. Thus, this thesis investigates approaches to automatically create disambiguated glosses.

While most of the text on the Internet is too difficult for low-ability learners, a small percentage is suitable for aided extensive reading. Search engines for language learners help them locate text of suitable difficulty. This thesis builds on the approaches described in this chapter, by using learner modelling to individualise search results. This thesis investigates using learner modelling to improve the efficiency of vocabulary review by automatically creating individualised review exercises.

Finally, in the classroom setting there is a need to minimise costs and to reliably assign grades to students using a method that requires minimal administrative overhead. This thesis develops an approach that automatically logs students reading and detects attempts to cheat the system.

4 The FERN system

FERN is a web-based computer system (http://fern.nzdl.org) that was developed to test the effectiveness of various tools and techniques for aiding extensive reading. This chapter describes its functionality from a user's perspective, and the software development process and system architecture used to create it. Algorithm details are presented in Chapters 5–7.

4.1 Overview

FERN is an acronym for *Fun Easy Reading from Novice to Native Speaker*. It supports extensive reading of authentic texts on the Internet both in the classroom and by independent learners. It is optimised for native English speakers learning Spanish, with the user interface available in both languages. Functionality for other language pairs is limited—Chapter 9 describes the steps to add full support.

FERN provides learners with the following five key functions:

1. News search. Displays individualised difficulty ratings that help learners find interesting L2 texts at an appropriate difficulty level.
2. Reading support. Allows learners to click unfamiliar words to view translations and dictionary definitions, and listen to pronunciation.
3. Vocabulary review. Learners review difficult words encountered during reading by performing automatically created review exercises.
4. Progress reports. Learners and teachers view charts of learning progress.
5. Mixed-language texts. Learners read L1 texts with carefully selected words replaced by L2 equivalents—principally as a vocabulary learning activity.

In order to receive individualised support each learner creates a user account, which they sign into at the beginning of each session. Teachers get live updates of student progress.

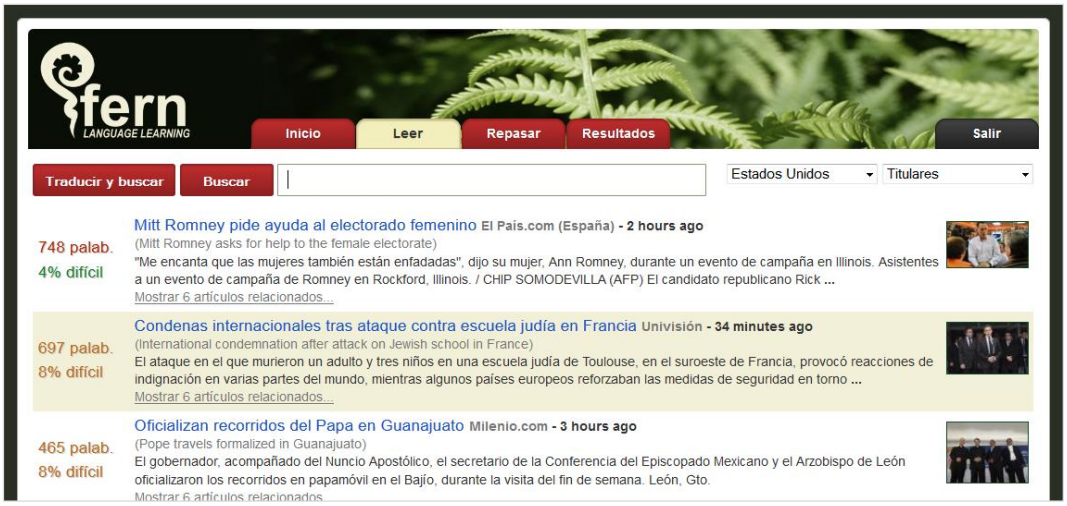


Figure 13: News search with word counts and individualised difficulty ratings

4.2 Choosing texts

4.2.1 News search engine

FERN's search engine builds on top of Google News search to help learners locate recent interesting L2 articles of desired difficulty and length. Learners select topics and enter queries and FERN displays result snippets that include an individualised difficulty score that estimates the percentage of difficult words in the text. The news search interface is displayed in Figure 13.

Learners select a country edition for their L2, which can assist with learning the culture and vernacular of a country they plan to visit. Learners of Spanish select *Estados Unidos* (United States) to favour articles from Spanish news sites in the US. Alternatively, they may select Argentina, Chile, Colombia, Cuba, España, México, Perú or Venezuela. Learners of English choose from 20 country editions including New Zealand, India and Israel.

Learners can search for news by typing an L2 query then clicking *Search*. Alternatively, they enter an L1 query then click *Translate and search*, which automatically translates the query into the L2 before performing the search. Finally, they can select one of eight topics: headlines, world, business, national, sports, science and technology, entertainment and health. Ten search results are displayed at a time, ordered by a combination of relevance and publication date. Links to display additional results appear at the bottom of the page.

As Figure 13 shows, search results for each article contain the L2 title, L1 translation of the title, and a short article extract. Learners can highlight words and phrases in the results to view L1 translations. Result snippets also include the article word count, the source website, time since publication, a related image, a link to a list of related articles,

Finally, results include individualised difficulty scores that estimate the percentage of difficult words based on the learner's past reading and vocabulary review. The scores factor in the proportion of cognates and proper nouns, which can be viewed by hovering over the score (see Chapter 5). Article words counts are coloured green (100–400 words), orange (400–700 words), and red (less than 100 or greater than 700) for rapid identification. Beginners should start with shorter articles, which tend to have simpler grammar. If a learner wants an easier or shorter version of a particular news story, they click "related articles" to explore alternatives.

Having chosen an article, the learner clicks the title to display the article in the FERN reader interface, or clicks the source website name to view it in its original context.

4.2.2 Reading other documents

Learners read other types of documents by pasting or typing text into a text box, or selecting a plain text file and clicking *Create Reader* to display it in the reading interface. Alternatively, they can choose a Wikipedia article or enter a URL for a web page, then click *Create Reader*. Then web pages are downloaded, text content extracted, and displayed by FERN. A JavaScript bookmarklet[45] is used as a shortcut to read the current web page with FERN.

4.3 Reading interface

The section describes how the FERN reading interface supports fast easy extensive reading.

[45] A small JavaScript application stored as a bookmark in a web browser, which allows a user to interact with the currently loaded web page in some way.

Figure 14: Reading interface with gloss and dictionary entry displayed as a result of a clicking twice on a word

4.3.1 Reader layout

FERN reader displays only the text contents of articles; images, menus and advertisements are hidden to help learners focus on reading. The interface makes reading comfortable by using a large font, large line spacing, large margins and high contrast, determined in consultation with an online reading interface expert.

Article text is initially displayed in a light grey, which is intentionally difficult to read. Learners click a paragraph to change its colour to black, as shown in Figure 14. If the first paragraph is less than 300 characters, the second is also darkened. In most articles, paragraphs are single sentences, and since it is annoying to click through lots of short sentences, short adjacent sentences are darkened together. When learners finish reading a section they click the next one to darken it, and lighten the previous one.

Darkening one section of the article at a time helps focus attention on the current section. More importantly, FERN records reading times for each section, deduces which sections were actually read, and computes average reading speeds for each article—excluding sections that were not clicked on, or had unusually high or low speeds. Times are paused when learners click *Pause*, or switch desktop applications or web browser tabs. The *Finish* button ends reading and displays a results box with the words read, time, speed and cumulative points earned in FERN; points are scored for each word read.

4.3.2 Translations, dictionary and pronunciation

Learners right click words to listen to pronunciation, left click to view translations and left click twice for dictionary entries. Glosses are precomputed and displayed immediately when words are clicked. FERN uses Google Translate to predict the correct translation of words with multiple meanings, based on surrounding context (see Chapter 5). Phrase translations are accessed by highlighting up to five words, which allows translation of collocations and idioms, but prevents lazy students from being tempted to translate entire sentences.

Dictionary entries are provided as backup because the translations are occasionally incorrect. Each entry displays the headword, followed by a list of translations sourced from Wiktionary and Google Translate (see Chapter 5.4). Translations for each word sense are displayed on separate lines. As shown in Figure 14, dictionary entries have a minimalist design, with no parts of speech, example sentences, antonyms, or synonyms. This approach helps learners continue reading with minimal distraction.

FERN tracks each learner's reading and clicking, and constructs a profile of their vocabulary knowledge. This profile is used to generate progress reports, calculate article difficulty scores (see Chapter 5), and automatically generate vocabulary exercises and mixed-language texts tailored to individual needs (see Chapter 7).

4.4 Review exercises

Automatically-generated review exercises supplement reading by testing comprehension and reviewing difficult words, and providing additional data for vocabulary profiles.

4.4.1 Fill-the-blanks

As shown in Figure 15, the fill-the-blanks exercises (also known as *cloze tests*) consist of previously read articles where several words are replaced by blanks. Such texts are a good way to evaluate comprehension and review vocabulary (Schneyer, 1965). To create these exercises, FERN divides articles into no more than four sections and replaces six words in each section with blanks. One section is displayed at a time, with the six answers displayed at the top. Learners drag and drop or type the answers into the blanks, then click *Next* to move to the next section. If there are

Figure 15: Cloze test to review previously read news article

Figure 16: Word match vocabulary review exercise

mistakes, incorrect words light up red and they must try again. If all are correct, the boxes light up green and the next section is displayed.

At the completion of all sections the exercise results are displayed with time taken, number of errors and points received—based on a combination of time and errors.

4.4.2 Post reading vocabulary review

FERN includes three post-reading review exercises, which target words they had difficulty with during past reading and reviewing.

Word Match, shown in Figure 16, presents L2 words on the left and randomly shuffled translations on the right. Learners drag and drop the L2 words to match the translations, and click *Check answers* when they are done. The number of words and difficulty level are customisable.

Word Guess consists of three rounds of multi-choice questions that practice eight automatically-selected difficult words. Round one presents L1 translations and learners choose L2 equivalents from a list. In round two, L2 words are the questions and L1 translations are the answers. Round three reviews incorrect words from rounds one and two. Results are displayed at the end of both *Word Match* and *Word Guess* exercises.

Word List contains four modes for reviewing lists of challenging vocabulary.

1. Displays a list of L2 words next to their L1 translations in separate columns.
2. Replace the second and subsequent letters of words in one column with dashes. For example, *to inform* becomes *to i*-----.
3. Replace all except the first letter with ellipsis, so that *to inform* becomes *to i*….
4. Replace all the words in one column with questions marks.

Modes two to four conceal the meaning of either the L1 or L2 words. In each mode the learners click concealed words to view answers, or click *Shuffle* to randomise the word order, to promote learning word meanings rather than order in a list.

4.5 Mixed-language texts

Since reading authentic L2 texts is impractical for total beginners, FERN introduces L2 vocabulary via mixed-language texts: L1 texts with carefully selected words replaced by L2 equivalents. Texts for new learners include a small number of easy Spanish cognates, such as *control* (control), *causa* (cause), and *adicionales* (additional), as shown in Figure 17. Over time the difficulty and proportion of L2 words automatically increases to include difficult cognates like *municipio* (town),

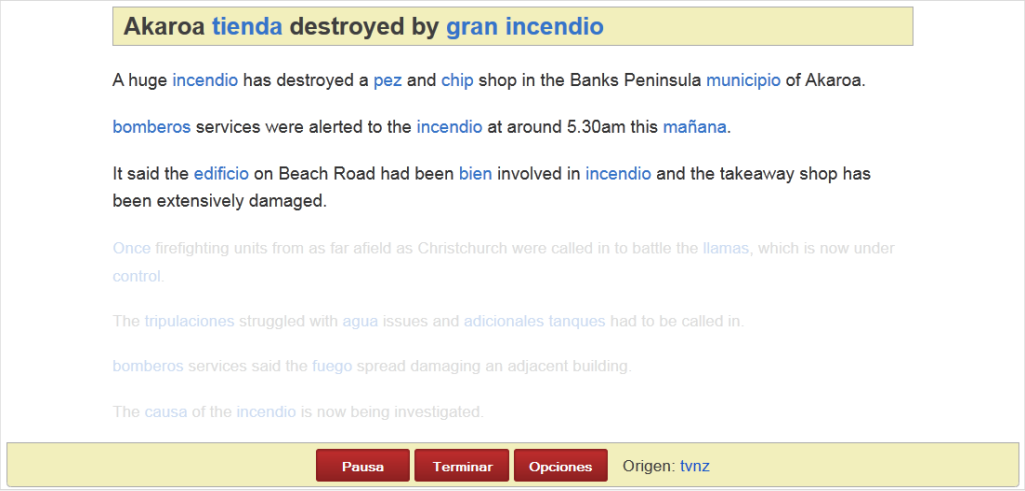


Figure 17: Example mixed-language text for English natives learning Spanish

edificio (building), high frequency non-cognates like *tienda* (shop) and *mañana* (tomorrow), and mid frequency non-cognates (see Chapter 7).

Learners choose texts to read using the same methods as L2 texts (see Section 4.2).

As with reading L2 texts, learners click sections of text to activate them, and reading times are used to predict which parts of articles were actually read. They click words to view translation glosses for unfamiliar words, which are then recorded on their vocabulary profiles. When they click L2 cognates, the gloss displays the L1 cognate and the L1 word that was replaced (if it is different). The density of L2 words and parts of speech that are translated are customised by clicking "Options".

4.6 Reporting progress

FERN provides detailed feedback of each learner's progress, which helps to motivate learners.

4.6.1 Progress reports

As shown in Figure 18, FERN automatically records the date, title and source web site for each article read on a separate row. Learners click the title to re-read the article with FERN, or click the source website to view an article in its original location. Each row also lists the *total* words encountered, the number of *new* word families encountered, reading duration, average speed, and the number of words for which glosses were accessed. Mixed-language text logs also list the number of L2 words encountered and the total number of words, i.e. L1 plus L2.

My reading log (es)

Search: Show 25

Date	Title	Words New	Words Total	Time	Speed (wpm)	Help Uses	Help %	Source
26 aug '11	Nuevo servicio de Música BBM costará $5 al mes	81	435	4:07	106	12	3	techspot
19 aug '11	Google Street View llega al Amazonas	167	456	6:08	74	8	2	europapress
19 aug '11	Google lanza un blog para recomendar música	75	345	4:39	74	12	3	elmundo
19 aug '11	Nadal y el clásico: "No es sólo importante ganar"	57	313	4:26	71	10	3	as
19 aug '11	Nintendo muestra el rediseño de Wii y dejará de fabricar el modelo actual	59	172	2:15	76	5	3	hardgame2
19 aug '11	Nadal: "Hay cosas más importantes que ganar o perder"	47	152	2:02	75	8	5	marca
19 aug '11	Tenis: Nicolás Almagro derrota a Karlovic en Cincinnati	46	139	1:15	111	5	4	el-carabobeno
Total	82 articles	2988	24113	4:54:25		566		

Showing 76 to 82 of 82 First Previous 1 2 3 4 Next Last

Figure 18: FERN reading log

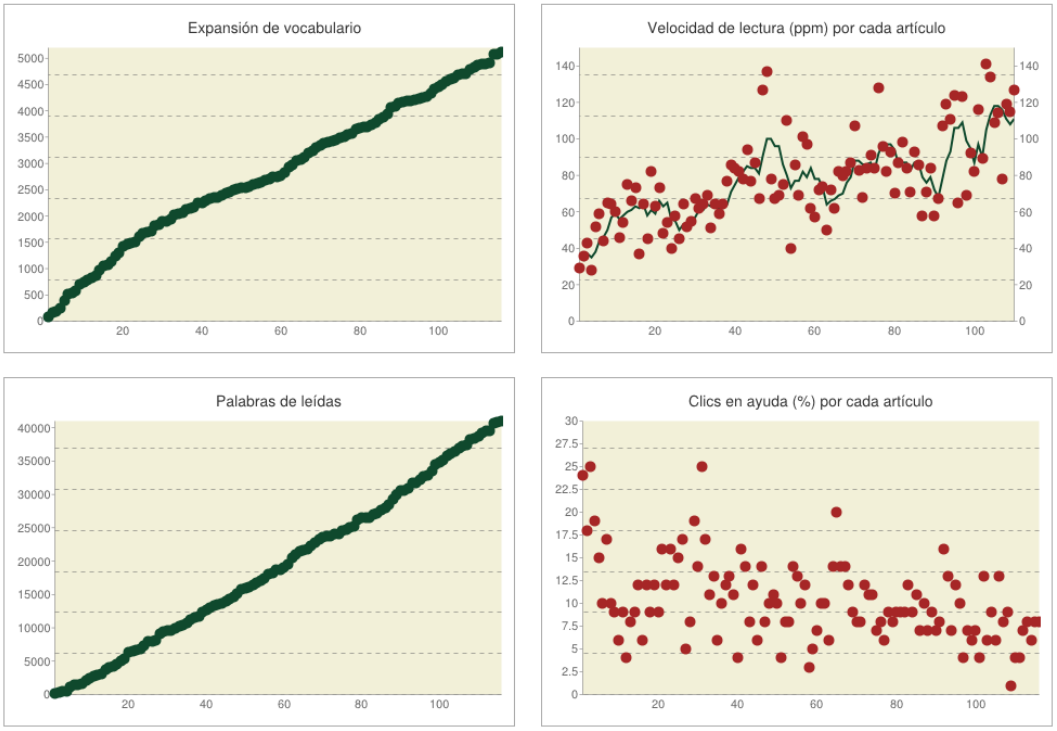
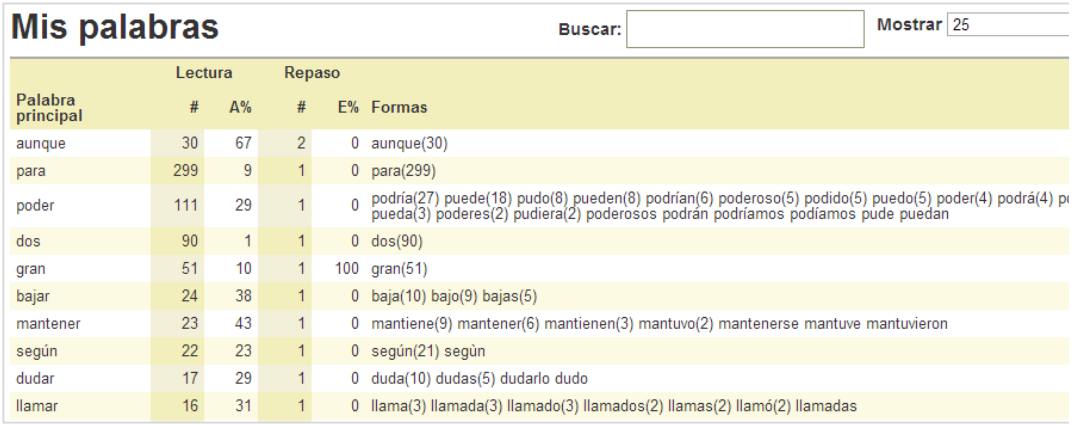


Figure 19: Progress report charts in FERN



Mis palabras

Buscar: Mostrar 25

Palabra principal	Lectura #	Lectura A%	Repaso #	Repaso E%	Formas
aunque	30	67	2	0	aunque(30)
para	299	9	1	0	para(299)
poder	111	29	1	0	podría(27) puede(18) pudo(8) pueden(8) podrían(6) poderoso(5) podido(5) puedo(5) poder(4) podrá(4) p pueda(3) poderes(2) pudiera(2) poderosos podrán podríamos podíamos pude puedan
dos	90	1	1	0	dos(90)
gran	51	10	1	100	gran(51)
bajar	24	38	1	0	baja(10) bajo(9) bajas(5)
mantener	23	43	1	0	mantiene(9) mantener(6) mantienen(3) mantuvo(2) mantenerse mantuve mantuvieron
según	22	23	1	0	según(21) segùn
dudar	17	29	1	0	duda(10) dudas(5) dudarlo dudo
llamar	16	31	1	0	llama(3) llamada(3) llamado(3) llamados(2) llamas(2) llamó(2) llamadas

Figure 20: Word list in FERN

Figure 19 presents a series of progress charts. In each chart, the x-axis is the number of articles read, and each dot corresponds to an article. The first chart shows the number of word families encountered at least once during reading. The chart in the top right shows reading speeds for each article. Red dots are the speed for each article, and the green line is a moving mean that is calculated after filtering outliers. The learner in Figure 19 started with a speed of 30 words per minute and progressed to 120 by the end of the class. The chart in the bottom right shows the percentage of words the learner needed help with during reading.

Figure 20 shows the *vocabulary profile*, which contains a row for every word family a learner encounters with FERN. Each row lists the headword; the number of reviews and percentage of errors in review exercises; and the number of encounters for each word form and percentage of gloss accesses during reading. Learners click column headings to change the sort order. For example, ordering by gloss accesses displays the words that were most troublesome during reading.

4.6.2 Classroom support

One challenge with classroom extensive reading is that many students do little reading unless they receive a grade, and some will cheat if they can do so undetected. There are several approaches to monitoring reading but none are cheat proof. FERN aims to make cheating more tedious than reading by tracking reading times for each section in texts and disregards those with unusually high reading speeds (see Section 6.4.1). Fill-the-blank exercises are another means of validation, because they are difficult to complete quickly and accurately by those who have not read and comprehended the articles. Individualisation of news search difficulty scores and review exercises is less accurate for those who cheat, which provides further incentive to use the system correctly.

This approach allows teachers to reliably track the number of words students read each week and assign grades accordingly. In Figure 22, students received credit each week they achieved a weekly reading goal of 2000 words. Teachers assign grades by viewing a table that lists the number of words read each week by each class member, or by viewing individual reading logs. An aggregated vocabulary profile shows the words a class has the most difficulty with. Teachers can view class summaries of class usage charts shown in Figure 21 and can also view individual students' reading logs and vocabulary profiles.

Learner accounts

Search: Show 10 entries

	General				Reading					Points			
Name	L1	L2	L2	ML	WC	Time	Help	WF	L2	ML	Match	Blanks	Total
Xxxxxxx, Xxxxxxx	en	es	187	4	65788	9:14:51	90	6345	79397	1106	1909	2533	85145
Xxxxxxx, Xxxxxxx	en	es	108	0	39599	8:55:35	162	5470	49866	0	5019	4414	59499
Xxxxxxx, Xxxxxxx	en	es	116	0	40937	9:23:57	3986	5111	53648	0	368	1865	56081
Xxxxxxx, Xxxxxxx	en	es	90	0	41410	5:56:41	3658	5495	49302	0	974	2117	52393
Xxxxxxx, Xxxxxxx	en	es	143	1	30776	4:43:16	877	4130	40214	263	2332	2448	45457
Xxxxxxx, Xxxxxxx	en	es	84	10	25632	5:32:31	638	3317	32415	3197	2926	3428	42166
Xxxxxxx, Xxxxxxx	en	es	59	1	26544	7:06:02	2121	4302	34561	259	4574	906	40500
Xxxxxxx, Xxxxxxx	en	es	120	0	26942	3:21:21	1626	4236	34578	0	4370	426	39574
Xxxxxxx, Xxxxxxx	en	es	50	3	21242	4:24:38	633	3290	27096	1530	1260	1326	31412
Xxxxxxx, Xxxxxxx	en	es	55	0	17619	3:03:18	137	3225	23044	0	150	0	23394

Showing 1 to 10 of 50 entries First Previous 1 2 3 4 5 Next Last

a) Summary of student usage of each activity

Weekly Reading

Search: Show 10 entries

Name	3	4	5	6	7	8	9	10	11	12	13	14	Total
Xxxxxxx, Xxxxxxx	5827	7881	0	6017	0	2595	4890	3787	3273	2999	2366	0	39635
Xxxxxxx, Xxxxxxx	2159	3423	2437	0	0	2137	1974	5047	5109	2629	4055	5604	34574
Xxxxxxx, Xxxxxxx	4611	9526	2368	0	0	2234	2304	2101	2569	2431	2154	2097	32395
Xxxxxxx, Xxxxxxx	8128	2972	2017	0	1719	3577	3804	0	3540	0	0	0	25757
Xxxxxxx, Xxxxxxx	2300	2281	2060	508	0	2125	2295	2295	2492	476	2855	2262	21949
Xxxxxxx, Xxxxxxx	2168	1177	2076	0	0	2100	2071	2243	2298	2328	2246	2058	20765
Xxxxxxx, Xxxxxxx	2580	2671	2294	0	2094	2070	2512	2246	1400	0	0	2137	20004
Xxxxxxx, Xxxxxxx	2036	2233	0	0	0	2358	2656	1968	4615	3150	0	0	19016
Xxxxxxx, Xxxxxxx	0	1544	2046	0	0	2321	0	2552	0	216	0	0	8679
Xxxxxxx, Xxxxxxx	2108	2897	0	0	0	1388	0	825	0	0	0	878	8096

Showing 1 to 10 of 50 entries First Previous 1 2 3 4 5 Next Last

b) Class report of weekly reading

Figure 21: Summary reports for teachers

Figure 22: FERN home page for the 14 week University of Waikato Spanish class

4.6.3 Gamification

FERN increases engagement by giving points for reading and reviewing, and displaying the class leaderboard on the home page, bringing an element of competition. The leaderboard displays who has done the most reading overall and in the past two weeks.

As depicted in Figure 22, FERN offers badges corresponding to weekly themes such as Cuba, Machu Picchu and the Panama Canal. In order to receive the weekly badge, students must complete the 2000 word goal, and at least one of each type of review exercise.

4.7 Development process

4.7.1 Web versus desktop applications

FERN was developed as a web application (a software application that runs in a web browser) in order to make deployment on multiple operating systems and devices simple. When updates to a web application are deployed, end users simply refresh their browsers to access the new version. Desktop applications must be installed and updated on every computer that runs them, often requiring administrator privileges.

Desktop applications traditionally offered a better user experience, were easier to develop, and offered more control over how applications work. For example, Web applications required an Internet connection and could not save data locally, and it was difficult to integrate with local hardware such as scanners and graphics accelerators. These advantages are no longer as pronounced, with modern tools and frameworks simplifying web development, and new HTML5 APIs allowing local data storage and enabling offline web applications.

A major challenge with commercial web application development is browser compatibility, particularly with old versions of Internet Explorer. FERN supports Chrome and Firefox 4+.

4.7.2 Web application architecture

In the client-server architecture users request web pages from servers, which return web pages coded in HTML, which defines content and structure; CSS, which defines the layout and style; and JavaScript to enhance interactivity. FERN uses the

YUI (Yahoo User Interface) and JQuery libraries and plugins to streamline JavaScript coding.

On the server side FERN uses Java servlets, JavaServer pages (JSP) and JavaServer pages standard tag library (JSTL). Java servlets are Java classes that receive requests for particular URLs and generate responses. Some of these classes retrieve and process data from a database then forward that data to a JSP, which combines server side scripting tags, the Java Expression Language, HTML, CSS and JavaScript to dynamically generate web pages.

FERN stores data in a flat file database (XML and plain text text), which is adequate for a single class of learners and a couple of hundred trial users. For larger scale deployment a robust database like MySQL would simplify system management.

FERN makes heavy use of Ajax (Asynchronous JavaScript and XML) to make requests that retrieve data from its server and from Microsoft and Google APIs, and dynamically update its web pages. The Google News API is used to retrieve lists of articles, which are extracted and translated using the Google Translate API. Google Charts APIs are used to create the progress reports and Microsoft Text-to-Speech provides audio pronunciation of words—Google's equivalent returns higher quality audio, but is only available for Chrome and only became available in 2014. Leveraging robust and complex existing technology enabled this research to focus on developing functionality specifically for L2 learners.

4.7.3 System components

As depicted in Figure 23, the FERN architecture has four key components: the client (the computer used by the learner), the server, web APIs, and the Internet as a source of extensive reading text. This section explains this adopted architecture using an example of an English native speaker wanting to read Spanish news about football's Confederation Cup.

The learner logs in on the client device and selects *Español news* from the *Read* menu. They type *Confederation Cup* in the search bar and click *Translate and Search.* The client uses Google's API to translate the phrase and queries to Google News for Spanish news about *Copa Confederaciones*. Google returns the top ten results in JSON format, which the client maps to HTML and displays. This entire process takes half a second with a good Internet connection.

The client then sends a request to the FERN server to retrieve search result difficulty scores. The server downloads the web pages, extracts news article text, computes difficulties by comparing articles with the learner's vocabulary profile, then returns the results to be displayed by the client. This process usually requires less than two seconds; the major bottleneck is downloading web pages from slow websites. The time for the server to extract article content and compute difficulty scores is comparatively negligible. Caching allows for quick retrieval when multiple learners request the same articles. In parallel the client retrieves L1 translations of article titles via Google and displays them below the L2 titles. These steps are repeated when they click links to view the next ten results, or click "related articles" links.

When the learner clicks on an article a request is sent to the server, which retrieves the article text from its cache, and sends it one sentence at a time to Google Translate. The resulting translations and word alignments are used to embed glosses in the article, which is returned and displayed by the client.

When the learner left clicks words during reading, a JavaScript function displays the gloss and instructs the server to record the gloss access in the learner's vocabulary profile. The server returns the dictionary definition for that word either from its own dictionary or from Google Translate. If they click the word a second time the dictionary entry is displayed instantly (unless they double click quickly, which results in a short delay). Highlighting a phrase in the article triggers a request to Google Translate.

When the learner right clicks a word, a request is sent to the Microsoft Bing Translator Speech API, which usually takes two or more seconds to respond to the first request, and one second for subsequent requests.

JavaScript code tracks reading times for each paragraph. When the learner clicks *finished*, readings times are sent to the server, which updates the vocabulary profile and reading log, and returns the reading results that are displayed by the client.

4.7.4 Open source libraries

This project used commercial APIs and the Open Source libraries in Table 6, to facilitate rapid development of FERN. This approach eliminated the need to write code for functionality that has already been implemented.

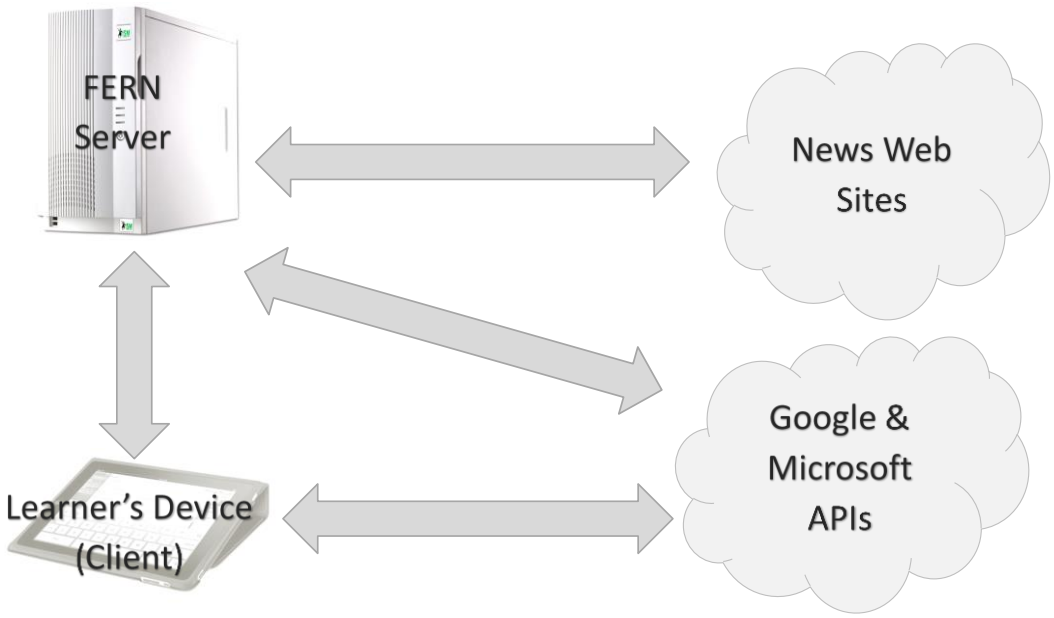


Figure 23: FERN System Architecture

Table 6: Open Source tools used in the development of FERN

Resource	**Licence**	**Usage**
OpenNLP	Apache	Sentence detection, tokenisation, named-entity recognition, part-of-speech tagging, chunking and tree parsing of English and Spanish. http://opennlp.apache.org
MeCab	GPL, LGPL, BSD	Japanese morphological analyser. http://mecab.googlecode.com
TextCat	LGPL	Guessing the language of European texts. http://textcat.sourceforge.net
Fathom	GPL	Computing common readability metrics for texts.
Wikipedia Miner	GPL	Provides sense disambiguation of key words in a text as well as analysis and parsing of Wikipedia.
JSoup	MIT	Parsing HTML documents. http://jsoup.org
Castor	Apache	Parsing and saving XML documents. http://castor.org

Figure 24: Mobile and touch optimised vocabulary exercise and reading interface

4.7.5 Mobile interface

The widespread use of smart phones and tablets in online learning has increased the accessibility of learning material. For this reason the FERN interface was also optimised for use on mobile devices as depicted in Figure 24.

5 Automatic glossing and dictionary lookup in FERN

An alternative approach to reading simplified texts, such as graded readers, is reading authentic text aided by glosses, which are short L1 translations or L2 definitions of difficult words and phrases, which appear next to the words, in the margin, or at the end of a text. Dictionaries list definitions of all word senses, whereas glosses provide the single sense that applies to a specific use. This characteristic, combined with their close proximity to target words, results in glosses causing little disruption to reading. Thus, gloss usage allows learners to transcend the 98% vocabulary threshold required for unaided extensive reading (see Section 2.2.1).

Ideally, glosses should be provided for every word, because even the simplest of words can be problematic in some contexts. However, liberally assigning and arranging glosses in printed texts is a tedious process that introduces clutter that distracts readers. For electronic texts, no such trade-off applies, because glosses can be automatically assigned to all words in a text, and can remain hidden from view until required by learners.

This chapter discusses existing approaches to automatically glossing certain types of texts, such as medical texts and the bible. This is followed by a description of FERN's approach, which uses machine translation to automatically gloss texts from any subject domain. Since machine translation is error prone, dictionary lookup is provided as a backup. The second half of this chapter discusses the key features of a dictionary to aid L2 extensive reading, followed by a description of an approach to automatically create such a dictionary using Wiktionary.

Chapter 8 evaluates the impact of translation errors on automatic glossing, as well as the effectiveness of the automatically created dictionary.

5.1 Background

5.1.1 Paper-based dictionaries

Dictionaries can improve comprehension and vocabulary learning when reading texts that contain difficult words (Knight, 1994; Luppescu & Day, 1992).

According to Holley and King (1971), learners vocally or sub-vocally repeat words as they look them up in dictionaries; this "rehearsal effect" facilitates vocabulary learning. Hulstijn and Laufer's (2001) "involvement load hypothesis for L2 vocabulary learning" explains this effect in terms of the depth of processing involved in finding a word in a dictionary and disambiguating the meaning of words with multiple senses.

When reading texts without dictionaries, learners guess the meaning of unfamiliar words from context. This process promotes learning, but is problematic when learners guess incorrectly. When looking up words with multiple senses in a dictionary, learners must choose the correct definition based on the context, but the restricted set of options reduces the chance of problematic incorrect guesses.

While dictionaries facilitate vocabulary learning, they impede the primary goal of extensive reading, which is improving reading fluency through fast-paced easy reading. Dictionary use breaks the flow of reading by requiring learners to complete the following steps to identify the meaning of an unfamiliar word:

> Step 1. Determine its headword—as only headwords are defined in paper dictionaries. This task is challenging for words with irregular inflections, such as the Spanish verb *vamos* which has the headword *ir*.
>
> Step 2. Search for the headword in the dictionary; a process that is significantly slower when L1 and L2 have different alphabets (Okuyama & Igarashi, 2007).
>
> Step 3. Determine the correct meaning for words with multiple senses.
>
> Step 4. Search for the stopping point in their text and re-read the sentence with the unfamiliar word to validate understanding.
>
> Step 5. Continue reading.

Luppescu and Day (1992) report that ESL students' reading time doubled with the addition of dictionary use. If learners are given access to dictionaries they will often use them excessively (Prichard, 2008). Consequently, Day and Bamford (1998, pp. 93–94) recommend banning dictionary use by extensive reading novices. This helps them develop the skill of skipping unimportant unknown words, and guessing the meaning of others from context. Day and Bamford suggest that more experienced

readers should be allowed to use dictionaries for unfamiliar words that appear several times in a text, or that they circle unfamiliar words to look up at the end a chapter or article.

5.1.2 Paper-based glosses

As with dictionary use, glosses benefit comprehension (Davis, 1989; Hong, 2006; Jacobs, 1991) and incidental vocabulary learning (Cheng & Good, 2009; Hulstijn, Hollander, & Greidanus, 1996; Yoshii, 2006), with the advantage of being less disruptive to reading (Leffa, 1992). This advantage is particularly apparent for sentences that contain several unfamiliar words, idioms or collocations.

Dictionaries are particularly problematic for words with many senses. Glosses alleviate the need to disambiguate the meaning of unfamiliar polysemous words, which aids reading, but probably reduces the learning of those words because less attention is applied to them (Hulstijn & Laufer, 2001). Conversely, because less time and effort is spent consulting dictionaries and disambiguating unfamiliar words, more energy is spent reading and reinforcing knowledge of partially known words. As a compromise between dictionary entries and single-sense glosses, Hulstijn (1992) and others (Nagata, 1999; Rott, Williams, & Cameron, 2002; Watanabe, 1997) suggest multiple-choice glosses—the correct gloss plus one to three distractors.

Gettys, Imhof and Kautz (2001) compared two glossing formats: translation equivalents versus headword definitions. For example, in the Spanish sentence *Fue a las tiendas* (He went to the shop) the headword glosses for *fue* and *tiendas* are *to go* and *shop*, whereas the translation equivalents are *went* and *shops.* The learner determines the correct tense of verbs and the number and gender of nouns, which requires a deeper level of processing. The results indicated that headword glosses increased learning of the unfamiliar words, without significantly impacting comprehension.

The effectiveness of glosses depends on learner ability and text difficulty (Hong, 2006; Jacobs, Dufon, & Hong, 1994; Joyce, 1997). Miyasako (2002) demonstrated that L1 glosses are more effective for beginners, but L2 glosses are better for advanced learners. In paper-based texts glosses are always visible, and Cheng and Good (2009) find them detrimental to comprehension when reading texts that are

easily understood without them. As Taylor (2010) put it: “The flow of reading can be interrupted by glossing for higher-level students because attentional resources are not allocated toward L2 text comprehension”.

Glosses of unfamiliar words help learners with comprehension, but glosses of familiar words are distracting. Ideally, texts should display glosses for every unfamiliar word, but no familiar ones. Such paper-based texts are impractical because no two learners’ vocabulary knowledge is alike. Computer systems overcome this challenge by providing glosses for every word in a text, and only displaying them when learners click unfamiliar words.

5.1.3 Multimedia glosses

Paper-based glosses are text, with occasional images; *multimedia glosses* include text, images, audio and video. Taylor’s (2009) meta-analysis of 32 studies, reports that computer-mediated glosses facilitate reading comprehension better than paper-based glosses. This improvement is due to the flexibility of computers to gloss all difficult words and display them on demand—thus avoiding cluttering the text with distracting annotations of familiar words.

Simply presenting glossed texts in electronic form does not improve comprehension, as demonstrated by Bowles (2004), whose students read a 570 word text in electronic and paper form. In both formats the same 21 difficult words were underlined and glossed. The paper-based text used marginal glosses, and the electronic one displayed them in the top-right of the screen when learners clicked underlined words. Because the same words were glossed, and the mode of presentation was similar, unsurprisingly both groups performed equally well.

Abraham’s (2008) analysis of studies on computer-mediated glosses reported a “medium positive effect” on comprehension, and “large effect” on vocabulary learning for intermediate and advanced learners, but smaller gains for beginners. The results for comprehension are likely affected by the studies glossing a small proportion of the words in texts.

Davis and Lyman-Hager (1997) evaluated L2 learners’ preferences for seven types of multimedia glosses: L1 translations, L2 definitions, cultural references, grammar notes, image, audio pronunciation and thought-provoking questions. L1 translations accounted for 88% of gloss usage.

Lomicka (1998) used a think-aloud protocol to investigate the effect of multimedia glosses on reading comprehension by L2 learners. Three groups of four students received the following treatments: no glosses; text-only definitions and translations; or access to text, images and audio glosses, plus cultural and grammar notes. Of the four students who had access to multimedia glosses, only one used images, and two made significant use of audio. Types of gloss accesses for this group were 31% L1 translations, 22% L2 definitions, 27% audio pronunciations and 9% grammar notes. Although the sample size was too small to draw firm conclusions, the results of comprehension tests indicated that access to the full suite of glosses improved comprehension.

Ariew and Ercetin (2004) and Sakar and Ercetin (2005) investigated the effect of multimedia glosses on reading comprehension by having participants read a text and then take a comprehension test. When surveyed, participants felt that glosses aided comprehension, with graphical and video receiving the most positive ratings. However, in correlating gloss access with comprehension, both studies concluded that graphical, video and audio annotations negatively affected comprehension. Ariew and Ercetin proposed that visual glosses distract learners and increase cognitive load, thus reducing cognitive resources available for comprehension.

On the face of it, this negative correlation is surprising. Among the limitations of their study Ariew and Ercetin (2004) suggest the need to investigate the existence of "a cause-and-effect relationship between the type of annotation and reading comprehension." Learners generally access glosses because of difficulty with comprehension, and gloss use often results in imperfect comprehension, thus it is unsurprising that gloss use correlates negatively with comprehension. In Sakar and Ercetin (2005) learners accessed 43% of the 90 glosses in a 900 word text. Given this high percentage, it is likely that learners had difficulty with words that were not glossed, and would have benefited from more glosses. Furthermore, the time constraint on the reading task may have prevented weaker readers from achieving good comprehension.

Plass, Chun, Mayer and Leutner (2003) had 150 L2 learners read a 760 word text with 35 glosses, then complete a comprehension test that required recall of the main points 24 hours after reading. Students were put in four groups: no glosses; text and audio translations; graphical; and all gloss types. The results were that learners with

Table 7: Effect of gloss position on scores in comprehension and learning tests (AbuSeileek, 2011)

Gloss Type	Vocabulary	Comprehension
No glosses	40%	38%
Bottom glosses	52%	44%
Marginal	60%	60%
Pop-up	60%	60%
Next to word	76%	74%

access to any type of gloss learnt more vocabulary that those with no glosses. Furthermore, those with only graphical glosses had inferior comprehension to both those with full glosses, and those with no glosses. Plass et al. suggest that ambiguity in the visual glosses increases cognitive load, thus reducing comprehension.

The groups with all gloss types and with no glosses had comparable comprehension. One possible explanation is that the text was too difficult and providing glosses for 5% of the words was insufficient for learners to improve comprehension relative to no gloss access; or perhaps the comprehension test was insensitive to differences in comprehension between these groups.

Yanguas (2009) had 94 L2 learners read a 540 word text with 21 words glossed. Students were placed in four groups: no gloss, text glosses, picture glosses, and both. Vocabulary tests on the 21 target words indicated that glosses improved vocabulary learning; the difference between the three gloss groups were not statistically significant. Gloss access also improved comprehension; the group with access to both gloss types performed best.

AbuSeileek (2011) demonstrated that the position of glosses significantly affected comprehension and vocabulary learning. Groups of 16 ESL learners read seven 400 word texts. In each text 21 words were highlighted and linked to glosses that appeared at the bottom of the page; in the margin; next to the word; and in a pop-up window directly above the word. A control group of 14 students had no glosses. As shown in Table 7, the group with glosses that appeared next to the word performed best; marginal and pop-up glosses were next. All glossed groups outperformed the group with no gloss access. AbuSeileek concludes that glosses should be presented in close proximity to the target word.

The difference between the pop-up and next-to-word groups is interesting, because both display glosses in close proximity to the target word. One contributing factor could have been superior knowledge of the next-to-word and margin groups, which scored 6–9% higher than the pop-up group in the vocabulary and comprehension pre-tests. Sub-optimal user interface design may also have contributed: once displayed, a pop-up remained in position until another word was clicked, which may have distracted learners.

From these studies it is clear that well-designed glosses benefit comprehension and vocabulary learning. These results have influenced the presentation of glosses in FERN (see Section 4.3.2), which minimises disruption to reading by providing glosses for every word in a text; positioning glosses in close proximity to target words; using a minimalist design; appearing instantly when the mouse is pressed on target words; and disappearing when the mouse is released.

5.1.4 Annotation tools

Glosses are clearly beneficial if they can be efficiently added to existing digital texts; several systems expedite this laborious task. For example, the Guided Reading hypermedia shell (Martinez-Lage, 1997) helps teachers annotate texts with audio and visual glosses; LitGloss provides similar functionality (LeLoup & Ponterio, 2007). With jGloss[46] teachers gloss Japanese texts by choosing the correct translation from a list in a dictionary, and attaching furigana—smaller syllabic characters that are printed next to difficult kanji.

Glossing with these tools is still a slow process, and teachers generally annotate only those words which they expect their class to find difficult. While these tools are useful for preparing a handful of texts for an entire class, they are not practical for producing a library of extensive material, unless many teachers collaborate to share resources.

5.1.5 Automatic glossing

Several computer systems automate the glossing process to varying degrees, thereby, eliminating the manual work required to prepare glossed texts,

[46] http://jgloss.sourceforge.net

Second Language Reader[47] is a web application that provides several tools to assist L2 reading. Learners select from a list of pre-created texts, or upload their own. Readers click on words to view Google translations, and hear pronunciation via ImTranslators text-to-speech software.[48]

The Reading Assistant Project[49] (Ohtake et al., 2003) programmatically analysed a corpus of English-language medical articles and extracted a vocabulary of 8000 technical terms. A single medical-related Japanese translation was manually assigned to each term. They developed software that uses this list to automatically gloss English medical texts. Teachers manually checked that glosses are appropriate for the context in which they appear.

The Statistical Glossing Tool is a language independent approach that automatically glosses the Bible using both language versions of the text (Riding, 2008). For example, to provide a French gloss for the English word *temple,* the system locates all English verses that contain *temple* and *temples.* It then analyses corresponding French verses to predict the most likely translation.

Alessi and Dwyer (2008) described the great benefits that would accrue from a system that fully automates the glossing process. The availability of such a system would make it easier for beginners to do effective extensive reading using authentic text on the Internet. To the knowledge of the author, no computer-aided reading system has been developed to date that uses word-sense disambiguation to automatically gloss texts from any subject domain. The successful development of such a system in FERN is an important contribution of this thesis.

5.2 Machine translation and automatic glossing

One of the important goals of this work was to develop an automatic glossing algorithm with the following criteria:

- Maximum accuracy: minimise gloss errors that negatively impact reading speed, comprehension and vocabulary learning;

[47] http://open.byu.edu/projects/readers/L2/L2.htm

[48]http://text-to-speech.imtranslator.net

[49]http://lsd.pharm.kyoto-u.ac.jp/en/service

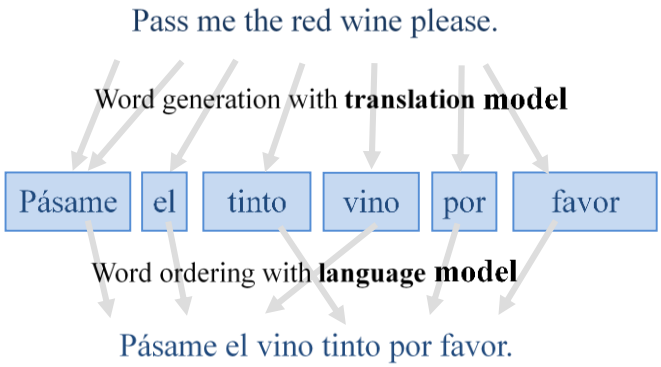


Figure 25: Model of the statistical machine translation process

- Maximum coverage: provide useful information for low-frequency words, proper nouns, idioms, set phrases, etc.;
- Domain independence: suitable for all genres of fiction or non-fiction; any length, from long novels to short blog points; and formal news to informal emails;
- Language independence: easy to adapt to other language pairs.

Given these criteria, it was decided to power the FERN glossing algorithm using the Google Translate Research API, which is built on Google's statistical machine translation engine.

5.2.1 Statistical machine translation

Since the 1950s, machine translation (the use of computers to automatically translate one natural language into another) systems have used rule-based approaches that incorporate sets of detailed rules developed by linguists. However, increases in computer processor speed have led to the emergence of statistical machine translation (SMT) as the dominant approach (Lopez, 2008).

The statistical approach treats translation as a machine learning problem with two steps, illustrated in Figure 25. Step one uses *translation models* to generate a set of translationally equivalent target language words for those in the source sentence. A mapping between source and target words is called a "word alignment". Step two uses *target language models* to put the words in the order that makes the most grammatical sense.

Translation models are trained on a large body of parallel corpora, consisting of documents translated by humans into multiple languages. Google's translation models are based on billions of sentences of freely-available multilingual content that includes United Nations documents (available in six languages) and European Union documents (available in 20 languages). Language models are trained on trillions of words from mono-lingual corpora harvested from sources including the Internet. Increasing corpus size increases the accuracy of translations.

5.2.2 Google Translate

Google Translate launched in 2001[50] and used a rule-based SYSTRAN engine until 2007, when Google switched to an internally developed statistical engine.[51] The volume of data used in its models is ever expanding and the quality of translations constantly improves, as evidenced by placing first in the NIST tests of Machine Translation Engines[52] [53].

Google Translate supports 90 languages with varying quality for each language pair. This is due to inherent difficulties with translating pairs with large language distances—languages with dissimilar grammars, and few cognates—and is affected by the size of the translation models. English has the largest language models, and consequently translations into English are better than the other way around.

However, of more relevance to word alignments and the automatic-glossing task is the size of the translation models, which are built on bilingual corpora. Since these corpora are particularly large for the United Nations and European Union languages, Google Translate performs especially well for those languages.

Och (2005) stated that a solid base for a new language pair consisted of a bilingual corpus of at least a million words, and two monolingual corpora of a billion words each. Because bilingual corpora of this size are not available for most language pairs that do not include English, translation to English is used as an intermediate

[50] http://web.archive.org/web/20011004203626/http://www.google.com/language_tools?hl=en

[51] http://googlesystem.blogspot.com/2007/10/google-translate-switches-to-googles.html

[52] http://www.itl.nist.gov/iad/mig/tests/mt/2005/doc/mt05eval_official_results_release_20050801_v3.html

[53] http://www.itl.nist.gov/iad/mig/tests/mt/2008/doc/mt08_official_results_v0.html

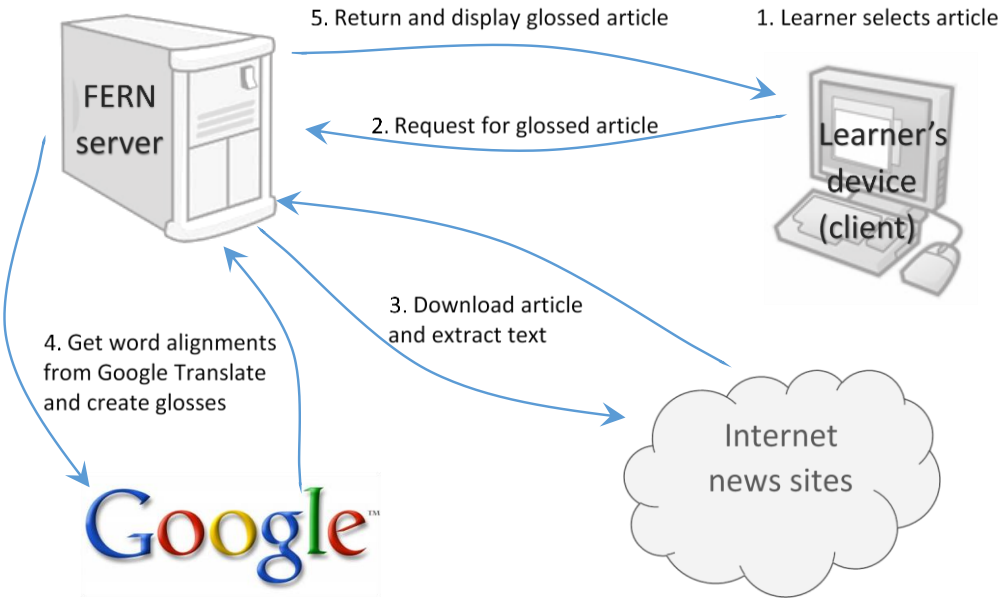


Figure 26: Automatic Glossing Approach

step if it produces better translations. In a few cases other languages are used as intermediaries, as when translating between Ukrainian and Belarusian to English: these languages are first translated to Russian and then English.

Google and competitors like Babel Fish and Bing translators provide APIs that allow developers to translate whole sentences. In 2007, Google released an additional API for University research that provided word alignments between source and target sentences. This tool is used by FERN to automatically gloss texts.

5.3 Automatic glossing algorithm in FERN

FERN automatically annotates L2 text with L1 glosses using the following five-step process depicted in Figure 26:

1. The user selects a text to read.
2. The client sends a request for glosses to the server.
3. The server downloads the article and extracts the article content.
4. Server performs the automated glossing using the following sub-steps:
    - Send the text to Google Translate Academic Research API;
    - Map resulting word alignments to phrase glosses;
    - Embed glosses in HTML attributes in tags that surround each word.
5. A web page with embedded glosses is returned and displayed by the client.

When learners click on unknown words, glosses are displayed instantly using JavaScript. If learners get confused by a gloss they can click the word a second time to view translations from Wiktionary.

5.3.1 Using the Google Translate APIs

Google had two translation APIs: the Public version and the now-discontinued University Research version. The Public API is accessed by sending a GET request with the parameters described in Figure 27a. For example, the request in Figure 27b translates the *en* (English) phrase “Hello world” into *es* (Español) and returns the JSON response in Figure 27c.

The University Research API can provide the ten best translations along with probability scores for each. Additionally, it provides word alignments that link words in the source text with their counterparts in the translation. The parameters for GET requests to the Research API are described in

Figure 28a. The request in

Figure 28b returns the XML response shown in

Figure 28c. It contains a sequential list with at least one gt-alignment element for every word in the translation. The position attribute aligns with the first character in a word in the query. In

Figure 28c, ‘¡Hola,’ has a position of 0 which matches *Hello* in the query; ‘mundo!’ has a position of 6, which corresponds to ‘world!’

5.3.2 Chunking of articles

Since the Google API limits query length to 1024 bytes, news articles used in FERN must be split into chunks before they are translated. In order to get high quality translations, the chunks should be complete sentences. Most news websites put paragraph breaks between sentences in articles. In these cases the article is split into sentences that are translated one at a time. In the rare case that sentence length exceeds 1024 bytes, the sentence is split by colons, then semi-colons, then by commas, and finally spaces.

Articles that contain more than one sentence per paragraph are translated one paragraph at a time except when they exceeded the limit. Paragraphs that exceed the limit are chunked into sentences by splitting on sentence ending punctuation followed by a space and an uppercase letter. An alternative approach of using the

sentence splitter in the OpenNLP machine learning based toolkit was slow, and incorrectly split many more sentences than using the simple heuristic.

Parameter	Purpose
q	“query” text to be translated
source	Two digit code[1] for the language of the text to be translated
target	Two digit code that sets the language to be translated into
key	A Google supplied code that tracks each developer’s usage

a) The public Google Translate API parameters

```
https://googleapis.com/language/translate/v2?key=X&source=en&target=es&q=Hello%20world
```

b) GET request for public Google Translate API

```
{"data": {
   "translations": [{"translatedText": "Hola mundo"}]
}}
```

c) JSON response for public Google Translate API

Figure 27: Using the public Google Translate API

Parameter	Purpose
q	“query” text to be translated
sl	Two digit code for the source language
tl	Two digit code for the target language
align	Set to 1 to retrieve word alignments
nbest	Sets the number of translations to return

a) The Google Translate University Research API parameters

```
http://translate.google.com/researchapi/translate?sl=en&tl=es&q=Hello%20world&align=1
```

b) GET request for the Google Translate University Research API

```
<…>Hello world</…>
<entry>
 <…>¡Hola mundo!</…>
  <gt:alignment word="¡Hola" position="0"/>
  <gt:alignment word="mundo!" position="6"/>
</entry>
```

c) Simple GET request for the Google Translate University Research API

Figure 28: Using the Google Translate University Research API

Determining the byte length of a UTF8 encoded string is straightforward for text that contains only English letters and common punctuation, since these symbols are contained in the first 128 ASCII characters which have a size of 1 byte each. Accented characters and those from non-Latin based alphabets are more than 1 byte. FERN uses the *getBytes* method in the Java String class to determine the length of any UTF8 encoded string.

5.3.3 Mapping word alignments to glosses

Statistical machine translation algorithms align each word in the source sentence with zero or more target words. Likewise, targets words are aligned with zero or more source words that may not be adjacent. With the Google API, each word in the translation is aligned with at least one word in the source, and each word in the source translates to at least one word in the translation. When these alignments overlap FERN combines them into *phrase* glosses.

Figure 29 visually presents the glossing rules, using arrows for the word alignments and boxes around the glosses assigned by FERN.

The simplest case, depicted in Figure 29a, is that each word in the L2 sentence is aligned to and glossed by a single word in the L1 translation. However, due to differences in L1 and L2 grammar it is common for multi-word phrases to gloss a single L2 word, as in Figure 29c, where *city* is glossed by *ciudad del* (literally meaning *city of the*). In other cases a single word glosses a multi-word L2 phrase. This is common with English verbs, because Spanish verb conjugations indicate person, tense and mood (and can be suffixed with direct and indirect object pronouns), whereas English verbs only indicate tense. In Figure 29b, the Spanish verb *tenemos* is aligned to *we have.*

When word alignments overlap, FERN glosses two or more words with a multiple-word phrase. Figure 29d, the Spanish word *hermoso* (beautiful) is aligned to both *beautiful* and *country*, while *país* (country) is aligned to just *country*. The overlapping alignments are combined and *beautiful country* is glossed by *país hermoso*. The gloss uses the word order from the Spanish translation, because using the English word order is ungrammatical. Figure 29e maps a two word gloss to three words L2 phrase, and Figure 29f maps a three word gloss to a two word L2 phrase.

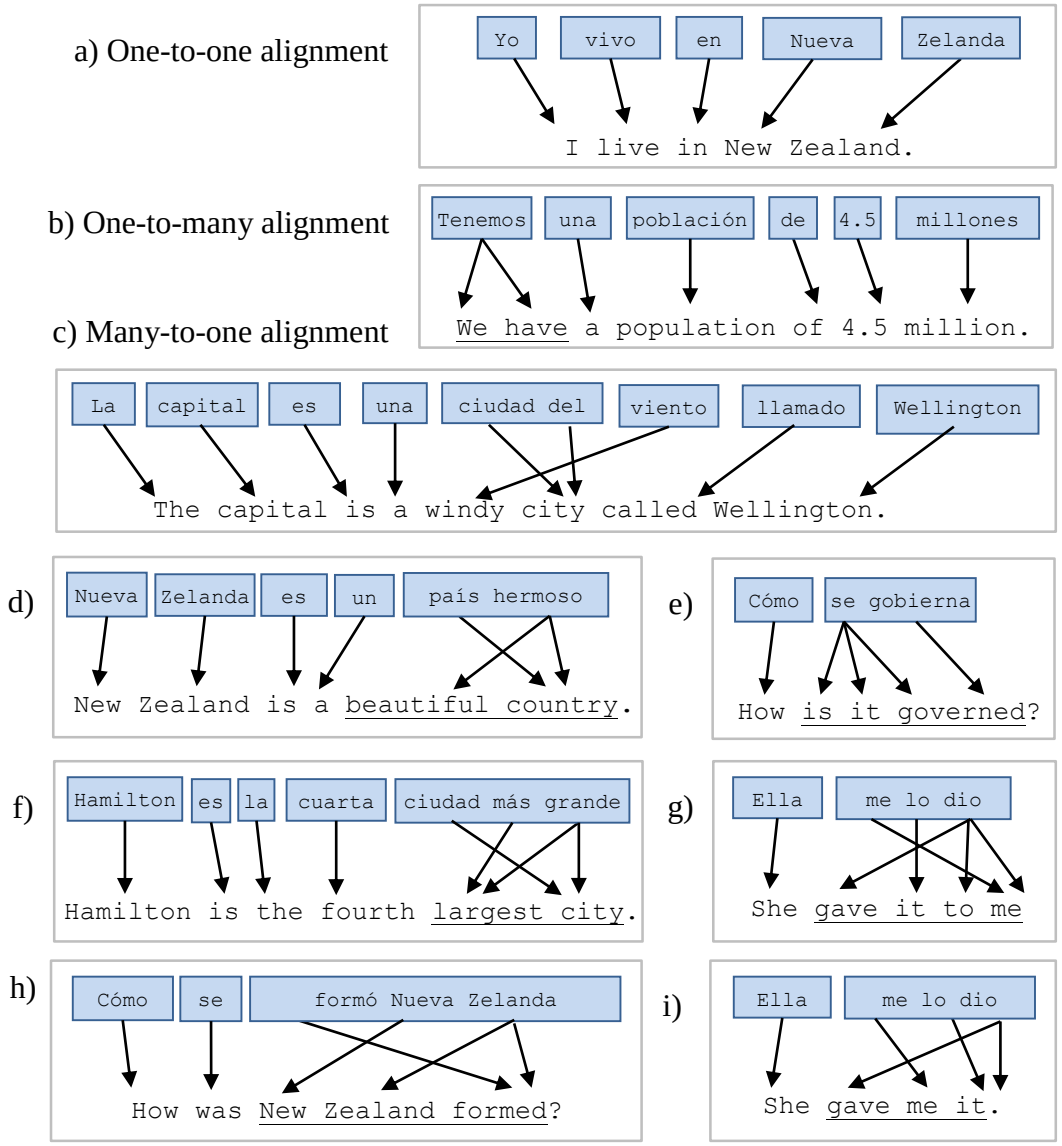


Figure 29: Gloss mappings for simple alignments (a-c) and many-to-many (d-i)

Figure 29g–i illustrates a special type of many-to-many alignment in which a word with a one-to-one alignment is combined with surrounding words. In this case, *Nueva* has a one-to-one alignment with *New*, but is surrounded by *formó* and *Zelanda* both align with *formed*; In this case FERN combines the words to gloss *New Zealand formed* with *formó Nueva Zelanda*. The alternative is to gloss *New* with *Nueva* and *Zealand formed* with *formó Zealanda*, but this approach produces confusing, potentially disruptive glosses.

There are rare cases where two words that are linked by alignments have several words in between. In these cases the words in between are not combined into a

single gloss. Also note that punctuation is removed from glosses in FERN for better clarity, as is the case with glossing *How* with *Cómo*.

5.3.4 Word alignment to gloss algorithm

To automatically gloss texts, FERN translates L2 texts one sentence at a time with the University Research Google Translate API, which returns word alignments for each word in the L1 translation, as depicted in Figure 30a. The algorithm for mapping the word alignments to glosses is as follows:

1. For each alignment FERN computes the start and end position, excluding trailing and preceding punctuation. Adjacent alignments for the same L1 word are combined into a single gloss (Figure 30b).
2. Alignments are sorted into ascending order by start position (Figure 30c).
3. Alignments with overlapping start and end positions are combined into a single gloss, in the order they appear in the translated sentence (Figure 30d).
4. To prepare the HTML document, the glosses are embedded in HTML tags around the words in the L2 text. The resulting glosses are displayed in Figure 30e, with arrows depicting the original word alignments.

5.4 Backup dictionary lookup

This section discusses two general classifications of dictionaries available to L2 learners: bilingual and monolingual dictionaries. An understanding of dictionary options is important because Google powered automatic glossing occasionally produces problematic glosses; FERN overcomes this problem by providing dictionary lookup as a backup. Common errors include getting the wrong form of the correct L1 translation, as shown in Figure 30b, which glosses *fui* as *go* instead of *went*; and getting an L1 translation that is incorrect for the given context, as shown in Figure 30h, which glosses *tienda* as *tent*—a valid translation that is incorrect for the example sentence.

Despite continual improvements by Google, the errors are unlikely to be eliminated entirely. Thus, it is advantageous to allow dictionary lookup when learners are confused by dubious glosses. In FERN, learners click once to view the automatically generated gloss, and click twice to display a dictionary entry.

```
<…>FERNReader aids second language learning.</…>
<entry>
 <…> FERNReader ayudas aprender un segundo idioma.</…>
  <gt:alignment word="FERNReader" position="0"/>
  <gt:alignment word="ayudas" position="11"/>
  <gt:alignment word="aprender" position="32"/>
  <gt:alignment word="un" position="23"/>
  <gt:alignment word="segunda" position="16"/>
  <gt:alignment word="idioma." position="23"/>
  <gt:alignment word="idioma." position="32"/>
</entry>
```

a) Word alignments from University Research Google Translate API

Word	Start	End
FERN	0	9
ayudas	11	14
aprender	32	39
un	23	30
segunda	16	21
idioma	23	39

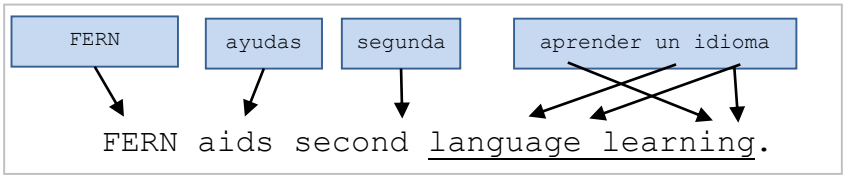
b) Word alignments

Word	Start	End
FERN	0	9
ayudas	11	14
segunda	16	21
un	23	30
idoma	23	39
aprender	32	39

c) Sorted alignments

Word	Start	End
FERN	0	9
ayudas	11	14
segunda	16	21
aprender un idioma	23	30

d) Combined alignments



e) Mapping of alignments to glosses

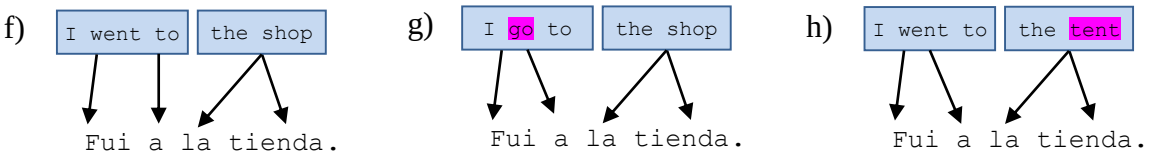


Figure 30: Mapping word alignments to glosses

5.4.1 Bilingual dictionaries

Bilingual dictionaries contain translations of words and phrases, and are either unidirectional—translating from one language to another—or bidirectional—translating between both languages. For example, half a dictionary translates Spanish to English, and the other half translates English to Spanish. Translations of

L2 into L1 are most useful for reading and listening, and translations of L1 into L2 are for speaking and writing.

Dictionary entries arrange translations by word sense; L2 translations are often accompanied by an unambiguous L1 synonym for each sense. Entries typically state the part of speech, gender, tense, inflection and other grammatical information. Additionally, these dictionaries often contain supplementary pronunciation and grammar guides. To aid correct language production, dictionary entries often list collocations and example sentences, and indicate words that are slang, vulgar, archaic and high frequency; and state applicable geographical restrictions, for example, stating that the English word *banana* is translated to Spanish as *banana* in Argentina, *banano* in Columbia, *cambur* in Venezuela, *guineo* in Dominican Republic, *mínimo* in Central Honduras, and *plátano* in Spain.

Due to space constraints, paper dictionaries provide full entries for headwords, and cross-references for some irregular inflections, e.g. a Spanish dictionary may include an entry that lists *voy* as the *first-person singular of ir*. Non-headword regular inflections are not listed, which can cause problems for learners still learning the inflections. Lookup is straightforward when there is one possibility, e.g. *falleció* is easily matched with its headword *fallecer* because it is the only entry prefixed by *fallec-*. Lookup is more problematic when multiple possibilities exist, e.g. *falles* is a form the headword *fallar*, but shares the prefix *falle-* with *fallecer*.

In Spanish, stem changing verbs are particularly problematic for beginners. For example, *soler* becomes *suelo* in the first person present tense. Because there are many words in the dictionary that separate *soler* from *suelo*, learners will struggle to determine the meaning if they are unfamiliar with its inflection rules.

Also problematic is looking up unlisted word forms that have homonyms. For example, *como* is an adverb translated as *like*, *as*, or *how*. It is also the first-person singular of *comer*, which means *to eat*. This second definition is usually excluded from paper dictionaries because it is a regular verb form. Electronic dictionaries are

Table 8: Defining vocabulary sizes of monolingual learner dictionaries

Dictionary Title	Defining Vocabulary
New Method English Dictionary	1490
Longman Dictionary of Contemporary English	2000
Macmillan English Dictionary for Advanced Learners	2500
The Oxford Advanced Learner's Dictionary (8th Edition)	3000

not limited by size constraints, and generally include cross references for all word forms.

5.4.2 Monolingual learner dictionaries

In standard monolingual dictionaries the headwords and definitions are in the same language, and the definitions often include mid- and low-frequency words that are too difficult for L2 learners. Consequently, publishing companies produce "learner dictionaries" that use a restricted "defining vocabulary". The first such dictionary, the New Method English Dictionary (West, 1943) restricts definitions to use 1490 high-frequency words. Other defining vocabulary sizes are shown in Table 8. Intermediate learners can transition from bilingual to monolingual learner dictionaries once their proficiency is sufficient.

In addition to definitions, learner dictionary entries contain grammar information, example usage, common errors and high-frequency collocations and idioms. Like bilingual dictionaries, they delineate high-frequency words. Most learner dictionaries are in English, but they are also available for Chinese, Dutch, German and Spanish. Another good option for L2 learners is to use dictionaries designed for children learning their L1, which use simple grammar and vocabulary, and frequently include pictures.

5.5 Hypertext dictionary requirements

Because FERN focuses on L2 reading by beginners, a bilingual dictionary is most appropriate. This section discusses dictionaries and glosses in general, and then outlines the requirements of a hypertext dictionary that aids extensive reading of electronic texts.

This section explains in detail the development of a bilingual dictionary for FERN. Dictionaries and glosses in general are again discussed, with the view to identifying the requirements of a hypertext dictionary that aids extensive reading of electronic texts by beginner L2 learners.

5.5.1 Complete coverage of words and senses

Paper dictionaries are constrained by physical limitations, and need to balance comprehensiveness with the need to be concise in order to be compact. Consequently, paper dictionaries intentionally omit low-frequency words. Electronic dictionaries have no such physical constraints, and as a result they can inexpensively include entries for every sense of every word in a language. Electronic dictionaries are only limited by the creator's ability to add those entries and keep up with the languages evolution.

A massive amount work is required to produce the translations and definitions for a dictionary. Nation (2006) demonstrated that the 2000 most frequent English word families cover 85% of a written corpus; the next 2000 cover a further 5-6%. The top 14,000 families cover up to 96%. Hence a dictionary approaching 100% coverage needs to contain at least 100,000 word families.

The first edition of the Oxford English Dictionary (*OED*, 1928), for example was a 70 year project that collected and organised quotations from a global team of volunteers, and published them in 20 volumes between 1888 and 1928. Interestingly, the technological advances of the digital age has allowed Wikipedia to co-ordinate a larger volunteer force and produce an even larger reference work in a fraction of the time. While translation dictionaries are a simpler undertaking, the manual work still required to approach near completeness is substantial.

Another challenge is keeping up with the proliferation of new words, and new meanings to existing ones. Jean-Baptiste Michel et al. (2011) estimated that the size of the "common" English lexicon (words with a frequency of at least one in a billion) rose from 597,000 in 1950 to 1,022,000 in 2000: an increase of 8500 words per year. FERN uses Google Translate to get the most common translation of words that are not present in Wiktionary. This approach is useful for neologisms that are not in mainstream use, and are thus excluded from dictionaries. When the Google translation is identical to the source, this indicates that either no translation is

available, or that the L1 and L2 words have equivalent spelling. This first case is somewhat problematic, but FERN cannot distinguish between them.

Paper-based dictionaries intentionally limit coverage of low-frequency senses in order to keep size and weight down. The Oxford English Dictionary and the Merriam-Webster Unabridged cover all words with frequencies above 10^{-6} (i.e. one in a million), but only 25% of those between 10^{-8} and 10^{-9}—600,000 low-frequency words that include many proper nouns and compound nouns. Unrestricted by size, hypertext dictionaries can incorporate proper name information in Wikipedia, and can include definitions from related foreign languages. For example, FERN looks up Spanish words in Italian, Portuguese and English editions if they are not contained in the Spanish Wiktionary.

Low-frequency inflections and derived forms such as demonyms (name for a resident of a locality, e.g., Wellingtonian and Aucklander) are often excluded from Wiktionary and Wikipedia. FERN's approach to including these low-frequency word forms is described in Section 5.7.2.

Wiktionary also excludes many transparent compound words. The ideal system would automatically detect transparent compound words and display definitions of its constituents. This functionality is not yet implemented in FERN.

Spelling errors in news articles occur enough to warrant some attention. Common misspellings of high-frequency words are often included in dictionaries, with cross references to the correctly spelled entry.

Spelling mistakes are particularly problematic when they produce a different real word. Common examples in Spanish are omitting accents in *cómo* (*how*) or *qué* (*what*), resulting in *como* (*like* or *as*) and *que* (*that* or *than*). FERN's dictionary automatically detects spelling errors that result in nonsense words and displays translations for an orthographically similar word, but does not yet address misspellings that result in other real words. Appropriate handling of this case is important because common misspellings can occur more frequently than identical correctly spelled low-frequency words.

The meaning of many words changes in idiomatic expressions. Dictionaries usually include high-frequency idioms, but omit low-frequency ones. FERN's dictionary displays entries for single words that learners click on, but not for phrases.

However, the automatically generated glosses provide sensible definitions for most high-frequency idioms.

5.5.2 Proper names and derivatives

Proper names are phrases that refer to unique entities, such as a person (e.g. Bill Gates), brand (Microsoft), product (e.g. Visual Studio 2010) or place (e.g. Redmond); proper nouns are single word proper names. Dictionaries vary in their coverage of these terms.

> Proper names are not systematically covered by the dictionary, though many are entered because the terms themselves are used in extended or allusive meanings, or because they are in some way culturally significant.
>
> Oxford English Dictionary[54]

Wiktionary's policies states that words must be "attested" in order to be included as separate entries.

> "Attested" means verified through
> 1. clearly widespread use,
> 2. use in a well-known work, or
> 3. use in permanently recorded media, conveying meaning, in at least three independent instances spanning at least a year.
>
> Wiktionary inclusion criteria[55]

Under these criteria, Wiktionary has entries for well-known given and family names, place names, company names and brands. These entries present etymologies, alternative spellings, meanings, and translations. Information about specific entities is the domain of Wikipedia.

News articles contain 10% proper names (Kyongho & Nation, 1989), but these rarely cause problems in L2 reading because most have similar spelling or phonetics across languages. In many languages they are easily identified by the capitalisation of their first letters, except when they are the first word of a sentence. Proper names

[54] http://public.oed.com/the-oed-today/guide-to-the-third-edition-of-the-oed/

[55] http://en.wiktionary.org/wiki/Wiktionary:Criteria_for_inclusion

Table 9: Differences in capitalisation rules between German, French and English

Type	Language	Example	English equivalent
Formal you	German	Sie	you
Common nouns	German	heute Abend	this evening
Days of week, months of year	Spanish, French	lunes, 1º enero	Monday, 1st January
Languages	Spanish, French	Hablo japonés	I speak Japanese
Nationalities	Spanish, French, German	Soy japonés	I am Japanese

rarely hinder overall text comprehension, as long as learners can distinguish their entity class (e.g. person, place, or organisation).

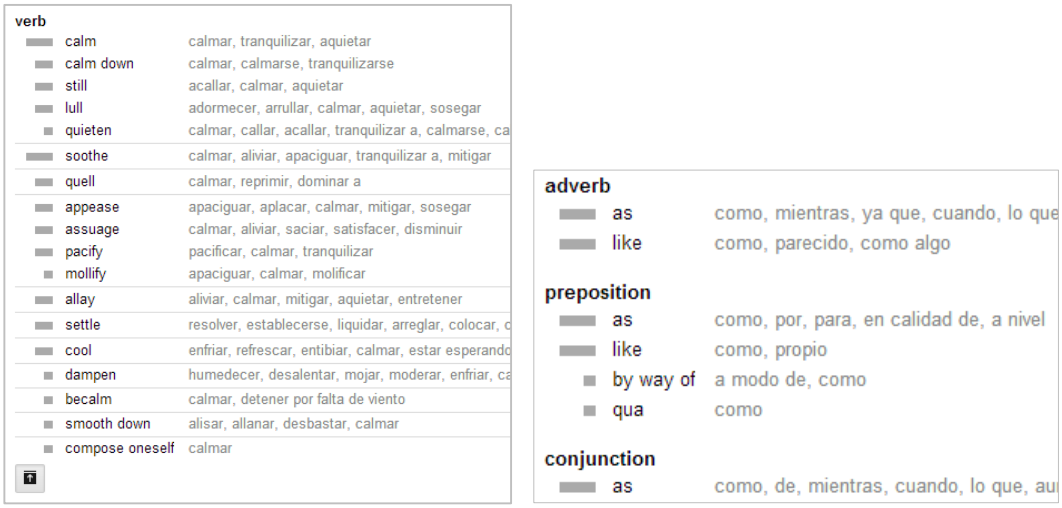
Proper names are problematic when there are differences between L1 and L2 capitalisation rules. For example, in German distinguishing common from proper nouns is difficulty because both are capitalised. Other differences in capitalisation between German, French and English are presented in Table 9. In these cases providing dictionary entries is helpful.

FERN relies on its automatic glossing to provide useful information about proper nouns, and makes no special consideration for proper nouns.

5.5.3 Quick lookup

When FERN learners are confused by unfamiliar words during reading, the hypertext dictionary incorporated in FERN helps them obtain clarity as quickly as possible, by providing an expedient trigger that promptly displays concise definitions. In FERN's case the trigger is pressing the left mouse button down on unfamiliar words. Because translations are embedded in the HTML code, they are displayed with no noticeable delay. Retrieving translations from a server on demand results in a noticeable delay, which can hinder comprehension (see Section 2.3.2).

In order to reduce size, paper dictionaries include full entries for headwords only, and provide cross-references for common irregular word forms, and omit regular ones. With no size constraints, FERN's hypertext dictionary can contain definitions of all word forms, and can jump directly to the headword definition if no specific definition is available.



a) Entry for the Spanish word *calmar* b) Entry for the Spanish word *como*

Figure 31: Google Dictionary entry layout

For words with multiple translations, FERN's dictionary groups and orders translations in a manner that facilitates quick look up. Some dictionaries group translations by etymology, for words with meanings that stem from multiple origins. These sections are further divided by part of speech and word sense. Figure 31b illustrates how Google Dictionary separates translations into parts of speech, and Figure 31a shows the use of faint grey lines to group 18 English translations for the Spanish word *calmar* into 11 word senses. The difference between 18 translations is delineated by following each one with lists Spanish synonyms. For example, the most common translation of *calmar* is *calm*, which has the Spanish synonyms *tranquiliza* and *aquietar*.

FERN's dictionary entries put translations from different senses on separate lines and no additional distinctions are made for etymologies and parts of speech.

5.5.4 Concise entries in frequency order

In order to minimise disruption to the learner, dictionary entries need to be concise, and provide the minimum information required to understand the words during reading. Generally, one or two word L2→L1 translations that cover each sense convey the meaning adequately. For words with several senses, it helps to rank them in descending frequency order.

As shown in Figure 31, Google Dictionary orders groups of translations by frequency, and sorts translations within each group by frequency. Grey bars to the

left of each translation give visual indications of frequency on a three point scale. For entries with many senses, the most frequent are displayed initially, along with a button to display all of them. For the entry in Figure 31a, the dictionary shows the 12 most frequent translations initially; the user clicks the button in the bottom left to reveal low-frequency translations.

FERN's dictionary presents translations for each in the order they appear in Wiktionary, which roughly corresponds to frequency and long entries are truncated.

Learner dictionaries typically provide usage examples. For example, the Collins Spanish-English dictionary gives 60 example sentences and phrases for the adverb *como*. These help deepen vocabulary knowledge, but are usually superfluous for reading comprehension. These and extra information such as part-of-speech tags and pronunciation should be hidden by default, and revealed on demand. Likewise, morphological analysis, concordance, and Wikipedia lookup functionality provided by tools like QuickAssist (Wood, 2010) are more suited to intensive rather than extensive reading.

FERN aims to provide the minimum information required by learners to understand the text and extra information is hidden in order to reduce distraction.

5.5.5 Correct and easy to understand translations

Dictionary entries presented to learners should be easy to understand. Those that include unfamiliar L1 words will unnecessarily disrupt reading. An entry with multiple translations for a single sense should hide those that are low-frequency L1 words, for example, those outside the top 10,000 English word families. This threshold depends on learners' L1 competency. Entries with three or more translations could display the two highest-frequency ones.

5.5.6 Learner aids

Words that learners recognise as cognates are easier to remember. Dictionaries can facilitate learning by informing learners of cognate translations. FERN does this by prefixing cognate translations with an asterisk the first time an unfamiliar word is encountered.

Highlighting high-frequency vocabulary is also beneficial, because it alerts learners to words that are important to learn.

5.6 Wiktionary as a lexical data source

Public domain dictionaries—those with expired copyright—are freely-available sources of translations[56]. The XDXF project[57] manages a repository of open dictionaries that are stored in a common XML format. This project aims to simplify the process of incorporating multiple dictionaries into multilingual systems.

Translations can be extracted by aligning parallel corpora as is performed by statistical machine translation engines (see Section 5.1.5). Google is in a good position to use this approach with their massive corpora used to create Google Translate.

Wiktionary supplements open dictionaries with information provided by an active community of 3800 human experts. This approach relies on a sizable team of knowledgeable volunteers. Those that submit translations are expected to be experts in both languages. Many less common languages have small teams of contributors, and consequently few translations are available.

FERN uses Wiktionary as a source of translations, as well as a source for cognate and word family information. Wiktionary is an attractive option for several reasons. The content is open, with XML snapshots published monthly, which give access to the entire contents, including edit history and discussion. Wiktionary is more comprehensive than other open dictionaries, such as FreeDict[58] and XDXF[59], and has a standard format that applies to all 170 language editions. This simplifies the development of a multi-lingual system. While there are standard formats for machine readable open dictionaries, many use custom formats and are available in only one language.

Furthermore, due to the large number of contributors, Wiktionary is well placed to keep up with the evolution of existing words and the assimilation of neologisms into mainstream language. For example, Wiktionary defines *Googler* as *a full-time*

[56] For example, in the USA books published before 1923 have expired copyright and are in the public domain.

[57] http://sourceforge.net/projects/xdxf

[58] http://www.freedict.org

[59] http://xdxf.sourceforge.net/

Google employee or *a habitual user of the Google search engine.* The online Oxford Dictionary—subtitled "the world's most trusted dictionaries"—is outdated, defining *Googler* as the French word for performing a Google search. As further examples, Oxford lists *al Qaeda* and *Alabaman* in its September 2012 new words list;[60] these words were included in Wiktionary in April 2005 and June 2007 respectively.

This section discusses in some detail the contents and processes of editing and programmatically accessing the contents of Wiktionary articles. This is followed by a description of the computer program that was developed to automatically create FERN's hypertext dictionaries. A detailed understanding of language learning and translation processes was critical to successfully develop software tools that achieve the objectives of this thesis. This section helps to demonstrate the cross-disciplinary nature of the complex software development undertaken.

5.6.1 Editing Wiktionary articles

Wiktionary allows users to anonymously make changes to content that are instantly accessible anywhere. As illustrated in Figure 32a, each article contains links to edit content, discuss content, view the edit history and revert changes. Logged in users "star" articles to receive email notification of changes.

Editing articles requires a degree of technical expertise, because it is performed using a text-markup language, as opposed to a WYSIWYG editor. For example, wiki-markup surrounds headings and links to other articles, as shown in Figure 32b. Thus, it is no surprise that in 2011, 92% of editors were "proficient" computer users, 36% were able to program their own applications (Pande, 2011).

Registered editors review changes to articles and reverse those that unnecessarily violate Wiktionary's layout guidelines. The guidelines for translation tables state that contributors should only add translations they are certain of, never use automatic translation software, and only copy from dictionaries that are out of copyright. However, detecting violation of these rules is virtually impossible. Most entries follow the translation table conventions; however, some inconsistent edits are missed by the reviewers. Others translation tables contain lingering

[60] http://public.oed.com/the-oed-today/recent-updates-to-the-oed/september-2012/new-words-list-september-2012/

inconsistencies that are the result of guideline changes, such as the introduction of a new translation table markup syntax in 2008.

Many Wiktionary entries are created by Bots: "computer-controlled processes that interact with and edit the wiki, much like human users do, but in an automated fashion, typically in order to perform repetitive editing tasks."[61] Among other tasks, bots import content from public domain resources (which is later refined by humans), and generate "form of" entries for inflected forms. For example, *ThirdPersBot* [62] added third-person conjugations such as *smoulders*, which it defines as *third-person singular simple present form of smoulder*. Bots also standardize mark-up that violates layout standards, and update old markup to conform to new guidelines.

Because bots rapidly edit articles they are potentially very harmful if they are incorrectly designed or operated. A bot approval group supervises and approves this activity. Wikipedia has 1500 approved bots,[63] while Wiktionary has 50.

5.6.2 Article contents

The English Wiktionary is both monolingual—contains English definitions of English words—and bilingual—contains English translations of words from other languages. This section describes the information in Wiktionary that can be incorporated into hypertext dictionaries, and used to create word family and cognate lists.

Wiktionary has two types of definition: gloss definitions, which describe the meaning of words, and form-of definitions, which give cross-references from inflections to headwords.[64] For example, *balls* is defined as the *plural form of ball*. For the highly inflected language Latin, 95% of its million definitions in the English Wiktionary are form-of. Two-thirds of the definitions for all languages in English Wiktionary are of the form-of variety.

---

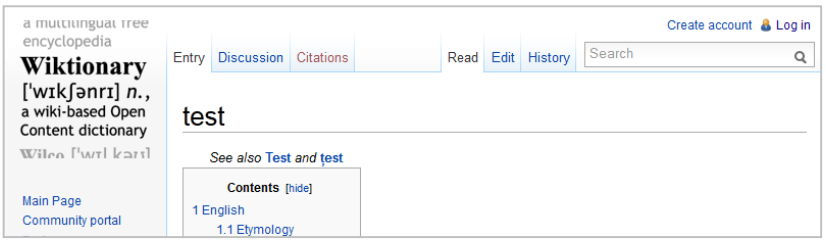
[61] http://en.wiktionary.org/wiki/Wiktionary:Bots

[62] http://en.wiktionary.org/wiki/User:ThirdPersBot

[63] http://en.wikipedia.org/wiki/Wikipedia:Bots

[64] https://en.wiktionary.org/wiki/Wiktionary:Statistics

a)



b)

Figure 32: Editing a Wiktionary article

a)

b)

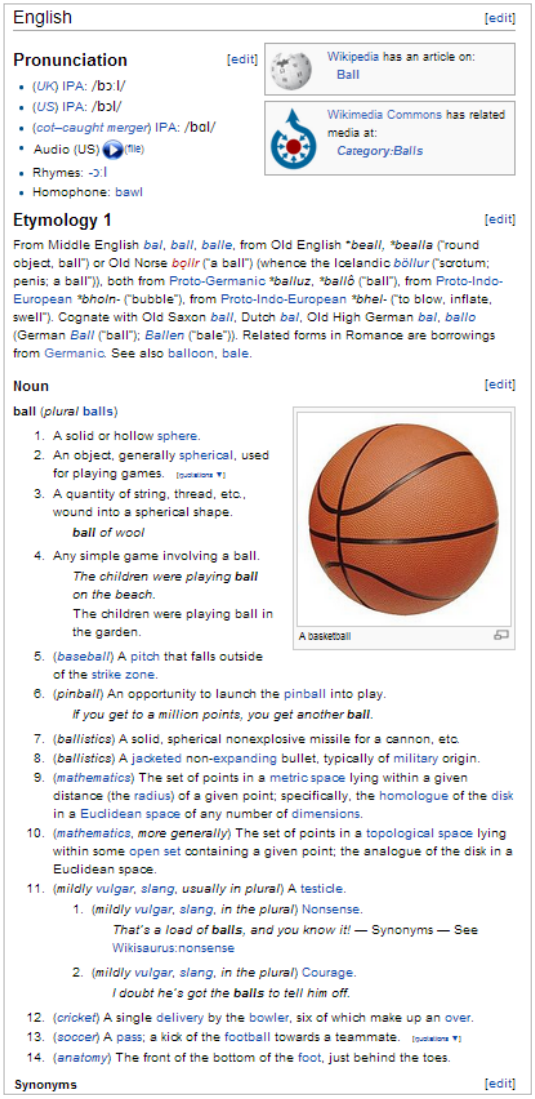


Figure 33: English Wiktionary entry for *ball*

Wiktionary articles contain additional information that may be incorporated in a hypertext dictionary. In Figure 33, three of the eight languages provide pronounciation in IPA notation. Human spoken audio pronunciations are available for a small proportion of words.[65] Entries also contain cross-references to Wikipedia, Wikimedia Commons[66] (a media file repository) and Wikisaurus (a thesaurus). Low-frequency words in definitions are hyperlinked to their Wiktionary definitions, which is also useful for learners.

Wiktionary entries also list inflections. English singular noun entries state the plural form. Adjectives list comparative and superlative forms. For example, the entry for *good*, lists *better* as the comparative and *best* as the superlative forms. Entries for the simple present tense (i.e. dictionary form) of English verbs typically list three or four additional verb forms. For example, the entry for *bowl* lists *bowls*, *bowling* and *bowled* as additional forms. Declension and conjugation tables are provided for inflected languages, as shown in Figure 35. This information can be used to create word family lists and lemmatizers.

Some definitions give quotations, example sentences, lists of related words, synonyms, and derived terms. This information is useful in automatically generated vocabulary exercises (see Section 7.4).

5.6.3 Translation sources

Wiktionary contains four sources of translations for each language pair. For Spanish-English these are:

1. Translation tables in English entries in the English Wiktionary;
2. English definitions in Spanish entries in the English Wiktionary;
3. Translation tables in Spanish entries in the Spanish Wiktionary;
4. Spanish definitions for English entries in the Spanish Wiktionary.

English Wiktionary entries for English words list translations into other languages for each sense. The entry for *ball* has 12 tables that present translations alphabetically ordered by language, as shown in Figure 34a.

---

[65] http://en.wiktionary.org/wiki/User:Robert_Ullmann/Pronunciation_statistics

[66] http://commons.wikimedia.org - A media file repository making available public domain and freely-licensed educational media content.

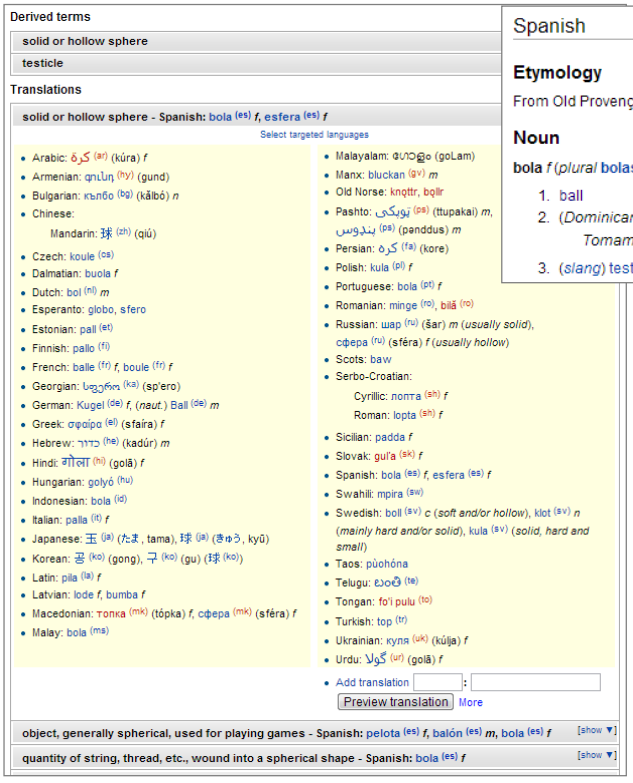
Conjugation of lanzar (See Appendix:Spanish verbs) [hide ▲]

Rule: z becomes a c before e.

infinitive	lanzar
gerund	lanzando
past participle	lanzado

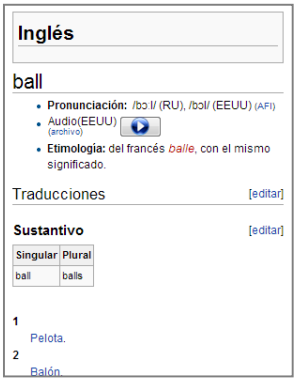
		singular			plural		
		1st person	2nd person	3rd person	1st person	2nd person	3rd person
indicative		yo	tú	él/ella usted	nosotros	vosotros	ellos/ellas ustedes
	present	lanzo	lanzas	lanza	lanzamos	lanzáis	lanzan
	imperfect	lanzaba	lanzabas	lanzaba	lanzábamos	lanzabais	lanzaban
	preterite	lancé	lanzaste	lanzó	lanzamos	lanzasteis	lanzaron
	future	lanzaré	lanzarás	lanzará	lanzaremos	lanzaréis	lanzarán
	conditional	lanzaría	lanzarías	lanzaría	lanzaríamos	lanzaríais	lanzarían
subjunctive		yo	tú	él/ella usted	nosotros	vosotros	ellos/ellas ustedes
	present	lance	lances	lance	lancemos	lancéis	lancen
	imperfect (ra)	lanzara	lanzaras	lanzara	lanzáramos	lanzarais	lanzaran
	imperfect (se)	lanzase	lanzases	lanzase	lanzásemos	lanzaseis	lanzasen
	future	lanzare	lanzares	lanzare	lanzáremos	lanzareis	lanzaren
imperative		—	tú	usted	nosotros	vosotros	ustedes
	affirmative		lanza	lance	lancemos	lanzad	lancen
	negative		no lances	no lance	no lancemos	no lancéis	no lancen

Figure 35: Conjugation table for a Spanish verb, in the English Wiktionary



a) Translation table in the English entry for *ball* in the English Wiktionary edition.

b) Spanish entry for *bola* in the English Wiktionary edition.



c) English entry for *ball* in the Spanish Wiktionary.

Figure 34: Sources of translations in Wiktionary

Each row in the table states its language, lists translations, and where applicable, word genders and transliterations. Some translations include information that qualify the scope of the meaning (as with the Russian and Swedish translations in Figure 34a or, as is commonly the case with Spanish, geographical restrictions that apply to a word's usage.

As shown in Figure 34b, Spanish entries in the English Wiktionary also list translations that are typically one or two words long. While there is much overlap between the two translations sources in the English Wiktionary, neither is complete. For example, the English entry for *ball* gives *esfera* as a translation, but the entry for *esfera* lists *sphere*, but not *ball* as a translation.

As shown in Figure 34c, English words in the Spanish Wiktionary usually give Spanish translations as definitions. The example in Figure 34c gives eight noun translations of the word ball, including *bola*. The translations in this entry cover three of those in the English Wiktionary.

Finally, Spanish entries in the Spanish Wiktionary contain translation tables. However, they are contained in far fewer entries than in the English edition. For example, the entry for *bola* lists translations in four languages, and *esfera* gives translations in 34 languages; the English edition entry for *ball* translates into 60 languages.

Because no single source is complete, FERN's dictionary includes translations from all four sources.

5.6.4 Extracting data from Wiktionary

There are three approaches to programmatically extracting information from Wiktionary, such as translations.

Wiktionary is available for download as a XML file[67] that compresses to 300MB excluding comments and article histories. The large size of the file requires a streaming approach rather than DOM parsing. Open source tools like XML::Twig[68] process the XML file and return Wiki markup for each article. A programmer must write additional code to extract specific information from the markup text. A major

[67] The English Wiktionary is available at http://dumps.wikimedia.org/enwiktionary

[68] http://search.cpan.org/dist/XML-Twig/

challenge with this approach is dealing with inconsistencies and errors in markup format.

Another approach is to extract the markup for each article using the previous approach, and then convert it to HTML using one of several code libraries.[69] Custom code must be written to extract specific data from a DOM representation of the HTML page. This approach is simpler than the previous one, because markup to HTML libraries normalise some of the markup inconsistencies.

Finally, the markup for individual articles can be downloaded by appending the request parameter *action= raw* to the normal dictionary entry URL; appending *action= render* returns HTML without the framework and template styles applied. Responses from these requests are processed in the same manner as the preceding approaches. The online approaches are much slower than the offline approach, and are consequently less practical for parsing the entire Wiktionary.

5.6.5 Extracting translations for FERN's dictionary

FERN's dictionary only includes translations extracted from the English Wiktionary.

Figure 36 depicts the variation formats in the markup for English Wiktionary translation tables. In Figure 36a, the Spanish, French and Georgian translations follow the standard format, but include different pieces of information; the French translation uses a deprecated format.

Figure 36b and Figure 36c show variations in the markup format for Spanish entries in the English Wiktionary. Variations include using colons instead of commas to separate translations, and omitting parentheses when providing information that qualifies its scope. These inconsistencies increase the complexity of developing software to parse and extract information from the Wiktionary data files. Much of the parsing code is to handle markup errors and rarely used formats. FERN handles all the variations in formats, and ignores gender and parenthetical information.

[69] http://www.mediawiki.org/wiki/Alternative_parsers

5.7 Creating a Spanish word family list

When learners click unfamiliar words during reading, FERN uses a Spanish word family list to look up and display headword translations. This list includes word forms extracted from Wiktionary, as well as less frequent ones produced by human defined rules combined with cross-referencing of an 87 million word corpus of Spanish subtitles.[70]

5.7.1 Extracting inflections from Wiktionary

There are two sources of inflection information in Wiktionary:

1. Inflection tables and plural information in headword entries.
2. Form-of definitions.

Markup surrounded by double-curly brackets are called “templates” (see Figure 37a). Those bits of code are expanded by the MediaWiki software to some predefined chunk of HTML code. The data inside the curly brackets specify the name of the template and parameters that are inserted into the template.

Wiktionary’s inflection tables are implemented as MediaWiki templates. For example, the markup in Figure 37a instructs the Media Wiki software to display the Spanish conjugation table for the regular *ar* verb hablar. Spanish has three conjugation templates for regular verbs, and 91 additional templates for stem-changing and irregular verbs. Determining the inflections requires an implementation of the MediaWiki templating system.

FERN uses the alternative approach: extracting the headword from form-of definitions for inflected forms. Virtually every verb form included in Wiktionary conjugation tables has a form-of entry. The markup in Figure 37b for the Spanish verb *hablabas* explicitly states that its headword is *hablar.* Likewise, form-of definition templates for nouns and adjectives explicitly state the headword, as shown in Figure 37c and d.

[70] See http://invokeit.wordpress.com/frequency-word-lists/ and http://opensubtitles.org

```
*Spanish: {{t+|es|bola|f}}, {{t+|es|esfera|f}}
*French: (''small, such as a cricket ball'') [[balle]] {{g|f}},
(''large, such as a football'') [[ballon]] {{g|m}}
*Georgian: {{t-|ka|ბურთი|sc=Geor|tr=burt'i}}
*German: {{t+|de|Ball|m}} {{qualifier|hollow or soft}},
{{t+|de|Kugel|f}} {{qualifier|hard}}
```

a) Markup from the translation table for *ball* in the English Wiktionary

```
===Noun=== {{es-noun|f}}
#[[ball]]
#{{context|Dominican Republic|slang|lang=es}} [[hitchhiking]], free ride
#:''Tomamos una '''bola''' para llegar aquí.'' — We hitchhiked here.
#{{context|slang|lang=es}} [[testicle]]
```

b) Markup from the definition section of *bola* in the English Wiktionary

```
#{{context|transitive|lang=es}} to [[throw]]
#{{context|transitive|lang=es}} to [[throw]] [[out]]
#{{context|transitive|lang=es}} to [[shoot]]; to [[launch]]
#{{context|transitive|lang=es}} to [[knock]] [[over]]
#{{context|transitive|lang=es}} to [[take]] (a [[photograph]])
#to [[pitch]], in [[baseball]]
```

c) Markup from the definition of *tirar* and *lanzar* in the English Wiktionary

Figure 36 Example markup in the English Wiktionary.

```
{{es-conj-ar|habl}}
```

a) Markup to display the conjugation table for the regular verb *hablar* (*to speak*)

```
#{{es-verb form of|formal=no|person=second-person|number=singular|
tense=imperfect|mood=indicative|ending=ar|hablar}}
```

b) Markup to for the form-of definition for the Spanish verb *hablabas*

```
#{{feminine of|lang=es|pos=adjective|bajo}}
```

c) Markup to for the form-of definition for the Spanish adjective *baja*

```
#{{plural of|bola|lang=es|nocat=1}}
```

d) Markup to for the form-of definition for the Spanish noun *bolas*

Figure 37 Markup for inflection information in Wiktionary

5.7.2 Deriving forms not present in Wiktionary

Wiktionary includes only 1400 Spanish entries for compound words consisting of verbs suffixed with direct and indirect object pronouns. There are thousands of valid compound words that are omitted. Spanish has six indirect object pronouns: *me, nos, os, te, le, les*; and eight direct object pronouns: *me, nos, te, os, lo, los, la, las*, from which millions of words can be formed using the following patterns:

1. Verb form + direct object
2. Verb form + indirect object
3. Verb form + indirect object + direct object

In the third pattern, the direct objects *lo*, *los*, *la*, and *las* become *se* when the indirect object begins with *l*. For example, *les* and *lo* combine to form the suffix *selo*. One example of type three *dámelo* (give me it), which consists of *da* (an imperative form of *dar*), *me* (an indirect object) and *lo* (a direct object). Note that an accent is added to *da* when it is combined with *melo*. Accents are always added for words that follow pattern three, and are added for pattern one and two except with infinitive verb forms (i.e. those ending in *r*).

To avoid including millions of compounds that are never used, FERN only includes those that are present in the Spanish TV Subtitles corpus.

Wiktionary also excludes many words that combine low-frequency nouns and adjectives with common prefixes (e.g. *anti, des, de, dis, un,* and *mal*) and suffixes (e.g. *mente, ero, ado, ito, illo, zuelo, cito, r, res* and their feminine equivalents which end in *a*). In some cases nouns change spelling when suffixes are appended. For example, the addition of *illa* to *manteca* (lard) forms *mantequilla* (butter)—the *c* changes to *q* to maintain the pronunciation. FERN's list includes these words if they are present in the Spanish corpus.

Finally, FERN searches for misspelled words that are not already in the word-family list. This is achieved by iterating through every possible single character deletion, replacement, transposition and insertion for every word already contained in the word family list. Those misspelled words that are found in the Spanish corpus are added to the family of the correctly spelled word form. Errors with accents are also accounted for, so that *hablame*, a common misspelling of *háblame*, is included

with the word family *hablar*. In this case the misspelling occurs 120 times in the corpus and the correct spelling occurs 1580.

5.7.3 Combining and disambiguating word families

The English edition of Wiktionary contains 253,400 Spanish entries that contain 373,400 definitions. 42,800 of those are glosses and 330,600 are form of definitions. Words with multiple senses technically belong to multiple word families. For practical purposes all words are assigned to a single word family. Noun and adjective groupings are merged with verb groupings that share a common word form. For example, *abrazo* is both a noun and verb form meaning *hug*. The noun forms of *abrazo* are merged with the verb forms.

A manually-defined list of 24 word families overrides the automatic groupings. For example, *como* is both a low-frequency verb form of *comer* (to eat) and a high-frequency adverb. The default grouping with *comer* is overridden to put *como* in its own word family.

6 Individualized search for L2 learners

Based on general readability measures (see Section 3.3.2), most L2 news articles are too difficult for extensive reading by beginners. However, many "difficult" articles can be accessible for those with appropriate background and interests. For example, L2 football articles are easier for learners that read L1 football news, compared with the general learner population. This thesis proposes that individualised search engines are significantly more useful for L2 learners than general ones, and provide good support for narrow reading.

FERN's individualised news search estimates the percentage of unknown vocabulary in articles, based on learners' performance in past reading and review exercises. This chapter reviews the principles of readability and recommender systems upon which it is based. We then describe its four key components: the *news search* to locate potentially interesting documents; a *screen scraper* to extract articles' contents from web pages; *vocabulary profiles* that track each learner's vocabulary knowledge; and the *difficulty algorithm*. The system is optimised for English natives learning Spanish, but is easily adapted to other languages (see Section 9.3).

6.1 Background

6.1.1 Readability metrics

FERN's news search incorporates features of readability metrics. These metrics use surface features like word and sentence length, vocabulary frequency lists, and language modelling to determine the reading level of arbitrary text. This section chronicles their evolution.

Gray and Leary (1935) investigated 64 factors that affected readability by conducting a reading comprehension test with 1000 participants. The top factors were average sentence length, percentage of easy words, and number of words unfamiliar to sixth graders. The top 17 factors included five that required counting syllables. Most formulas developed prior to 1990 incorporated word and sentence length (Dale & Chall, 1948; Flesch, 1948; Fry, 1968, 1977; Kincaid, 1975; McLaughlin, 1969; Spache, 1953). For example, the *Flesch Reading Ease* and

Table 10: Correlations of readability metrics with text difficulty (DuBay, 2007)

Formula	Correlation	Standard Error
Dale-Chall	0.93	1.76
Flesch-Kincaid	0.91	1.90
Gunning Fog	0.91	2.00
SMOG	0.88	2.28
Flesch Reading Ease	–0.88	2.44
Fry	0.86	2.31
FORCAST	0.66	3.61

Flesch Kincaid Grade Level formulas are defined in terms of average word length in syllables, and sentence length in words:

$$Flesch\ Reading\ Ease = 206.835 - 84.6\ wordLength - 1.015\ sentenceLength$$

$$Flesch\ Kincaid\ Grade\ Level = 0.39\ sentenceLength + 11.8\ wordLength - 15.59$$

Flesch Reading Ease scores typically fall between 0 (very difficult) and 100 (very easy) (Flesch, 1948). Scores over 90 are suitable for elementary school, and those below 30 are university level. The Flesch Kincaid Grade Level reduces the relative weighting of word length and uses coefficients that produce scores that correspond to US school grades, with scores over 12 for University level texts (Kincaid, 1975).

Other formulas incorporate syllable counts differently. Gunning Fog sums the words per sentence and the percentage of words with three or more syllables (Bogert, 1985). Fry (1977) counts the sentences and syllables per hundred words. SMOG computes the square root of number of words with three or more syllables in a 30 sentence sample (Bogert, 1985). FORECAST counts the one-syllable words in a 150 word sample (Kincaid, 1975).

DuBay (2007) tested the relative validity of these formulas by comparing them with a collection of texts graded by human experts. The high correlation scores in Table 10 indicate the usefulness of these formulas in grading L1 text difficulty.

In the past, a key design criterion for readability formulas was ease of calculation, but the advent of personal computers encourages more sophisticated and computationally intensive language modelling approaches. Some more recent formulas incorporate word frequency data. The *revised Dale-Chall measure* (1995)

counts sentence length and the proportion of words that are not in their list of 3000 easy words—those understood by 80% of US fourth graders.

Lexile uses frequency counts from a 600-million word corpus of school texts, to compute the log mean frequency of the words in a text and log mean sentence length, and combines them to give a score from 0–2000 (Lennon & Burdick, 2004). Learners complete a reading comprehension test that determines their "Lexile measure" on a scale of 0–2000. They then read books with similar Lexile scores.

ATOS grades children's books using words per sentence, characters per word and the grade level of individual words (Milone, 2012). Their graded vocabulary list incorporates manual judgement in addition to resources like Dale's (1931) list of familiar words, and frequency counts on 28,000 children's books.

Language modelling formulas use statistical models constructed from corpora to estimate readability (Kevyn Collins-Thompson & Callan, 2004; M. J. Heilman et al., 2007; M. Heilman et al., 2008). These approaches use machine learning to combine word frequencies with features extracted from automated grammatical parses of sentences. For example, Coh-Metrix uses "semantic lexicons, pattern classifiers, part-of-speech taggers, syntactic parsers, shallow semantic interpreters, and other components" to derive readability formulas (S. A. Crossley, Greenfield, & McNamara, 2008; S. Crossley, Allen, & McNamara, 2011). These metrics give stronger correlations to human-assigned levels than formula-based metrics.

Brown (1998) and Greenfield (2004) developed formulas for ESL learners, by correlating performance on cloze tests after reading texts. Their formulas made small gains for native English speakers. Coh-Metrix was superior for L2 learners.

English readability formulas have been adapted for languages with Latin scripts such as Spanish (Gilliam, Peña, & Mountain, 1980), German, and French (J. Anderson, 1981). Japanese has four character sets and thus requires new formulas. Tateishi, Ono and Yamada (1988a, 1988b) incorporate factors that account for long runs of the same type of character, which reduce readability. Their formula is defined in terms of characters per sentence (s), ratio of commas to periods (cp), and run lengths for Roman characters (a), hiragana (h), kanji (c) and katakana (k):

$$Readability = -0.12s - 1.37a + 7.4h - 23.18c - 5.4k - 4.67cp + 115.79$$

Language modelling has also been applied to Japanese (Sato, Matsuyoshi, & Kondoh, 2008) and French (François & Miltsakaki, 2012).

As described in Section 6.5, FERN uses vocabulary-related features in its individualised difficulty metrics.

6.1.2 Recommender systems

FERN's search incorporates components of recommender systems. These systems are used by web applications to suggest interesting music, videos, games, news articles, based on user preferences and past actions. Adomavicius and Tuzhilin (2005) described three types of recommendations:

Content-based. Recommends items that are similar—but not too similar—to those a user preferred in the past, using randomness to avoid predictability. For example, news sites recommend articles with topics similar to those read in the past. This approach works well for text content, from which keywords are easily automatically extracted; metadata is usually created manually for images, audio and video, because automatic analysis is complex.

Collaborative. Recommends items that are liked by other users with similar interests. These systems may exploit relationships in social networks (De Pessemier, Deryckere, & Martens, 2009), as Facebook does to recommend games, news articles, and new friends. Collaborative recommenders require a critical mass of users to make useful recommendations, and are unable to recommend items that have received insufficient ratings. Collaborative and content-based recommenders also need sufficient information about a user to give useful recommendations.

Hybrid. Combines content-based and collaborative approaches. For example, Amazon.com recommends products that are "similar" to those a user purchased in the past. Similarity is based on the number of users that have purchased both products (Linden, Smith, & York, 2003). YouTube tracks videos users watch, rate and add to favourites, and uses the data to recommend "related" videos. Two videos are related if many users watch them in the same session. Recommendations are ranked by the number of views, comments, favourites and shares a video has received (Davidson et al., 2010).

Extensive reading systems can recommend texts based on topic and linguistic difficulty. While, FERN's search makes use of content-based techniques only, its

makes sense for an extensive reading systems with large numbers of users to make use of collaborative and hybrid techniques.

6.1.3 REAP tutor

One recommender system in the L2 reading domain is the REAP tutor, which recommends suitable texts to L2 learners (J. Brown & Eskenazi, 2004; M. Heilman, Collins-Thompson, Callan, & Eskenazi, 2006; Kulkarni, Heilman, Eskenazi, & Callan, 2008). In the version described in Kulkarni et al. (2008), teachers choose a list of words to be learnt in a course. Learners indicate via self-assessment, which words they already know. They also indicate their interest in 16 general topics including arts, science and sports. REAP recommends articles that match topic preferences and contain at least two target words.

Heilman et al. (2008) concluded that the REAP tutor was less suitable than the original REAP system (see Section 3.3.3.3) for classroom usage:

> While the individualized REAP Tutor has the potential to better match the needs of each student... instructors may have difficulty coordinating group discussion about readings and integrating the Tutor into the curriculum. In the REAP Search system, however, teachers can find texts that match the needs and interests of the class as a whole. While some degree of individualization is lost, the advantages of better coordinated support from teachers and classroom integration are gained.

FERN aims to capitalise on this potential to better match the needs of individual learners. By aggregating class performance, FERN can help teachers locate texts that are most suitable for the entire class.

6.1.4 User profile

Designing user profiles for recommender systems include three key aspects.

Profile representation. Profiles include the information that is required to make good recommendations. For example, profiles in the REAP system contain topics of interest, and the number of encounters with target words during reading and review exercises.

Generating initial profiles. New user profiles are initialised via pre-tests and surveys, or are built up over time. Language learning systems use various types of

multi-choice questions that test vocabulary knowledge, or self-assessment tests that require marking familiar items in a list of real and nonsense words—an approach that has proven to be "reliable, valid and practical" (Nation, 2001, pp. 344–346). Self-assessment tests are also time efficient, with Heilman and Eskenazi's (2008) subjects completing them in one-sixth of the time of cloze tests, which test vocabulary knowledge by blanking out target words in sentences.

Adaptive tests use performance on past questions to select subsequent ones, which shortens the test time without sacrificing accuracy (Weiss & Kingsbury, 1984). Their main disadvantage is the considerable effort required to create them. The V-check adaptive test combines self-assessment with multiple-choice definition questions to assess knowledge of 13,384 English words in five minutes (C. Browne, 2008; C. Browne & Culligan, 2008).

Updating the profile. Relevant user feedback is used to update profiles. This feedback may be implicit and explicit (Montaner, López, & De La Rosa, 2003). For example, a video site may allow users to explicitly list personal interests on their profile, and rate their enjoyment of content. Preferences are indicated implicitly by their search queries, the content they choose to access, and the time they engage with it. User tastes change over time. Generally, profiles should give greater weighting to new rather than old feedback.

FERN's user profiles are models of learner's vocabulary knowledge, and are built-up over time, based on performance in reading and vocabulary review exercises. The user profiles are used to compute article difficulty scores displayed by FERN's news search. The usefulness of difficulty scores in early search results could be improved by using an adaptive test to initialise profiles.

6.2 News search

6.2.1 Why news?

The World Wide Web contains a variety of material that can be used for extensive reading—blogs, news, encyclopaedias, journals, novels and short stories, etc. FERN uses news because there is a large quantity of free articles in many languages that are relatively easy. In contrast, many academic journals charge for access, and are difficult for native speakers, which puts them well beyond L2 beginners (Rochon et al., 2002).

Wikipedia is multilingual and freely available, but is more difficult than news. Three quarters of its articles have Flesch reading ease scores below 60—"fairly difficult" for native speakers and well beyond L2 learners (Jatowt & Tanaka, 2012; Lucassen et al., 2012). Simple English Wikipedia is a useful source for ESL learners with an average reading ease of 80—"easy"—in 2003, which dropped to 70—"fairly easy"—by 2006 as the number of articles increased.

Traditionally, blog sites contain regular commentary by individuals on particular subjects. Many are now multi-author and professionally-edited by newspapers, universities and companies, with some functioning as online advertising. Because there are no restrictions on blog publishing, the quality varies greatly. Many contain informal language, misspellings, bad grammar and local slang, which are problematic for L2 learners. Chapter 9 evaluates the use of Wikipedia, blogs and news for extensive reading.

Fiction novels are typically written at a seventh grade level, but many are much easier, such as the popular trilogy "Hunger Games", which is rated at grade 5.3 (B. Wood, 2013). While most modern fiction works are not free, project Gutenberg contains thousands of older texts that are, three-quarters of which have Flesch-Kincaid scores of grade 10 or lower. Fiction published by bloggers is of questionable quality and has limited availability outside of English. Newspapers are mostly eight to eleventh grade, with some as low as fifth (Fowler Jr, 1978). Despite being suitable in terms of difficulty, fiction is excluded from FERN because of the limited range of texts available in Spanish. The short articles and modern language in news articles seemed most appropriate for user studies conducted for this thesis.

6.2.2 Architecture of a news search engine

There are three steps in the search process: web crawling, indexing and searching.

Web crawlers are computer programs that browse and download web pages on the Internet. They first download a set of "seed" pages, extract the URLs contained in each page, and add previously unvisited ones to the download list. Clusters of machines work in parallel to process the growing list. Crawlers decide how often to revisit sites and how quickly to download a site's pages to avoid overloading servers. They generally ignore password protected pages, pages with duplicate

Table 11: Google News Search API parameter values

a) Google News Search edition values

Value	Country	Language
nz	New Zealand	English
us	United States	English
es_us	United Sates	Spanish
es_mx	Mexico	Spanish
es_cl	Chile	Spanish
es_co	Columbia	Spanish
es	Spain	Spanish
hi_in	India	Hindi
ta_in	India	Tamil
in	India	English

b) Google News Search topic values

Value	Topic
h	Headlines
w	World
b	Business
n	Nation
t	Science & technology
el	Elections
p	Politics
e	Entertainment
s	Sports
m	Health

content, URLs generated by JavaScript and Flash, and those explicitly prohibited by *robot.txt* files. As at 2013, Google crawls 30–50 billion web pages.[71]

Indexers extract text content and metadata from web pages, and organise and store it for fast and accurate searching. The storage required to index the World Wide Web is immense: Google's indexes contained 800 terabytes of compressed data in 2006 (Chang et al., 2008).

Search engines aim to efficiently respond to user queries with relevant results. When users execute a keyword search, the engine examines its indexes for documents that contains those words, and then displays the top documents sorted according to various relevance criteria. One of Google's key relevance criteria is PageRank, which ranks pages based on how many high-ranking sites link to them (Brin & Page, 1998; Page, Brin, Motwani, & Winograd, 1999).

Google News is a specialised search engine that crawls content from 25,000 publishers (as at December 2009).[72] It indexes articles by country, language, and eight topics, in addition to the full text. Publication date is a key relevance criterion, with recent news taking priority.

[71] http://www.worldwidewebsize.com as at 23/07/2013

[72] http://googlenewsblog.blogspot.com/2009/12/same-protocol-more-options-for-news.html

6.2.3 Using Google News API

Another reason FERN focuses on news was the availability of the Google News API, which FERN uses to perform search and topic queries. For topic searches, the *topic* parameter is given one of the ten values in Table 11b. If used in conjunction with a query, the query is ignored and a topic search is executed.

The *ned* parameter selects a news edition corresponding to a geographical region and language, as represented by the sample values in Table 11a.

The results size, *rsz*, parameter sets the number of results returned per query to between one and eight. The *scoring* parameter is used to sort results by publication date instead of relevance, which is the default. The *start* parameter sets the start index of the first result, which is used to retrieve the second and subsequent results page for each query.

Each query returns an array of results objects, the first of which is depicted in Figure 38. Each result has the properties in Table 12. The Google API can also retrieve formatted HTML snippets with the same layout as news.google.com. FERN prepares a customised view of the results from the raw JSON response. JSON (JavaScript Object Notation) is a text based data interchange format primarily used to transmit data in web applications.

6.3 Screen scraper

FERN downloads the web pages from search results and extracts news article content and titles. This process, known as “screen scraping”, is performed with custom-built software, because no existing open-source Java component performed this task adequately.

6.3.1 Downloading articles

The biggest bottleneck in this process is downloading article web pages, which takes two to three seconds on average, exceeds ten seconds for slow sites, and times out for pages that are inaccessible. Multiple pages are downloaded in parallel so that slow websites do not delay the results for others. Requests that exceed 20 seconds are terminated.

Table 12: Properties in Google News Search API results

Property	Explanation
content	article snippet
title	article title with bold tags surrounding keywords from the query
titleNoFormatting	title without HTML
image (optional)	URLs to an image related to the story, as well as an image thumbnail hosted by Google
publisher	Name of the website that published the article
publicationDate	Full publication date, e.g. Sun, 04 Aug 2013 14:03:55 -0700
location (optional)	Location of the article
language	Two character ISO code, e.g. en for English
relatedStories (optional)	An array of additional stories covering the same event
clusterUrl (optional)	URL to a Google hosted page with a formatted view of the related stories


```
{"responseData":{
  "results":[{
   "GsearchResultClass":"GnewsSearch",
    "clusterUrl":"",
     "content":"While Chicago is no stranger to violent crime…”,
      "unescapedUrl":"http://abcnews.go.com/US/father-peacemaker…”,
      "url":"http%3A%2F%2Fabcnews.go.com%2FUS%2Ffather-peacemaker…”,
      "title":"Father Peacemaker: Chicago Priest Helps Saves Lives…”,
      "titleNoFormatting":"Father Peacemaker: Chicago Priest Helps…”,
      "location":"",
      "publisher":"ABC News",
      "publishedDate":"Sun, 04 Aug 2013 14:03:55 -0700",
      "signedRedirectUrl":"http://news.google.com/news/url?sa\u003dT…”,
      "language":"en",
      "image":{
       "url":"http://a.abcnews.com/images/US/gty_81300719_father…”,
       "tbUrl":"http://t0.gstatic.com/images?q\u003dtbn:ANd9GcTMI…”,
       "originalContextUrl":"http://abcnews.go.com/US/father…”,
       "publisher":"ABC News",
       "tbWidth":80,
       "tbHeight":45
      }
     }
```


Figure 38: Snippet of JSON result from Google News Search API query

Once downloaded, each page's character encoding is determined from the HTTP headers, or the *http-equiv* or *charset* HTML meta tags. If an incorrect encoding is used, accented and non-Latin characters are displayed incorrectly. The encoded HTML content is cached to ensure it is only downloaded once.

6.3.2 Article content HTML formats

The HTML code for the pages can exceed 1000 lines and is structured differently for each news website. The next step is to extract the article text from the HTML. While it is impractical to achieve perfect extraction from all sites, 95% coverage can be achieved because many sites have similar formats.

The most common structure is a parent that contains paragraph elements that contain the article text, as shown in Figure 39. The paragraphs often include hyperlinks, and other inline tags such as <b>. Some websites replace <p> with <blockquote> for direct quotes.

Comments, scripts, tables and divs containing images, videos and captions are frequently interspersed between paragraphs. In Figure 40 a slideshow of images and an article summary precede the opening paragraph. A containing a video separates paragraphs four and five.

This general format is used by most popular news sites, including Yahoo, CNN, The Times, and BBC. Reuters, New York Times and LA Times follow this format for all but the opening paragraph, for which they each have different formats. In the Reuters article in Figure 41 tags wrap the opening paragraph; empty elements precede subsequent paragraphs; and the parent tag is as opposed to the customary .

Paragraphs are occasionally separated by two consecutive
 tags. This format has become less common; AOL and CBS recently changed to the more common format in Figure 40. Less common formats include: one or two paragraphs not surrounded by <p> tags; separating paragraphs with single break tags; surrounding paragraphs with instead of <p> tags. Some websites split long articles into multiple pages, and display links to ensuing pages at the bottom of the first. These websites uses various URL and link formats for multi-page articles. FERN only extracts the content of the first page of each article.

```
<div class="entry-content" id="…">
  <p class="first" id="…">CHICAGO (AP) -- <a class="…" href="…" data-
rapid_p="1">Alex</a> …</p>
  <p id="…">The embattled <a class="…" href="…" data-rapid_p="2">New York
Yankees</a>' slugger … after <span id="…" class="…">Major League
Baseball</span> suspended …</p>
</div>
```

Figure 39: Example snippet of common HTML format for news articles

```
<div class="cnn_strycntntlft">
  <!--startclickprintexclude-->
  <script>…</script>
  <style type="text/css">…</style>
  <div class="cnnExplainer cnn_html_slideshow">…</div>
  <div class="cnn_strylftcntnt">…div>
  <div class="cnn_strylftcntnt adtag15090">…</div>
  <p>…</p>
  <p class="cnn_storypgraphtxt cnn_storypgraph2">…</p>
  <p class="cnn_storypgraphtxt cnn_storypgraph3">…</p>
  <div class="cnn_strylftcntnt">…</div>
  <p class="cnn_storypgraphtxt cnn_storypgraph4">…</p>
  <script>…</script>
  <a name="em3"></a>
  <div class="cnn_strylftcntnt cnn_strylftcexpbx" id="expand29">…</div>
  <p class="cnn_storypgraphtxt cnn_storypgraph5">…</p>
</div>
```

Figure 40: Extract from a CNN article

```
<span id="articleText">
  <span id="midArticle_start"></span>
  <div id="articleInfo">
    <p class="byline">By Jennifer Saba</p>
    <p><span class="timestamp">Mon Aug 5, 2013 9:03pm EDT</span></p>
  </div>
  <span id="midArticle_0"></span>
  <span class="focusParagraph"><p><span class="…">(Reuters) - <span
class="…"><a href="…" class="…">Amazon.com Inc</a></span> …</span></p>
  </span>
  <span id="midArticle_1"></span>
  <p>…</p>
  <span id="midArticle_2"></span>
  <p>…</p>
</span>
```

Figure 41: Snippet from a Reuters article

Table 13: Example contents of title elements for news articles

Publisher	Example <title> tag contents
Yahoo!	Facebook Is Ruining Facebook – Again: Here Come Video Ads - Yahoo! Finance
CNN	'Project 10:' Japan's rising son Nishikori feels weight of expectation - CNN.com
Huffington Post	Washington Post Sold To Amazon's Jeff Bezos
Fox news	Tech Rewind: Upward Trend in Social Media Stocks | Fox Business
EPA	european press agency: Nintendo muestra el rediseño de Wii
El País	Los anuncios de farmacias pueden costarle caro a Google · ELPAÍS.com
ABC	Solido 2-0 de Heat ante Celtics. Basquetbol/nba - Edicion Impresa - ABC Digital

6.3.3 Extracting article headline

The Google News API truncates article headlines when they exceed a specified length, and append ellipses. FERN extracts complete article headlines from their <title> element, which typically contain both the headline and name of the publisher. As shown in Table 13, the order of the headline and publisher vary from site to site, as does the character that separates them, which complicates headline extraction. FERN's heuristic for extracting headlines is to split the <title> element contents at all separators and choose the longest segment. For example, the final example in Table 13 splits into four segments: *[Solido 2-0 de Heat ante Celtics]*, *[Basquetbol/nba]*, *[Edicion Impresa]* and *[ABC Digital]*; the first segment is selected as the title because it is the longest.

6.3.4 Extracting the article contents

The news article is extracted from the HTML code in several steps.

First, FERN normalises the code by replacing occurrences of two consecutive
 tags with <p> tags. Next it is parsed by JSoup (see Table 6) into a DOM (document object model) object, which represents the code in a tree like structure. Figure 42 is a graphical visualisation of the DOM object for the code in Figure 43.

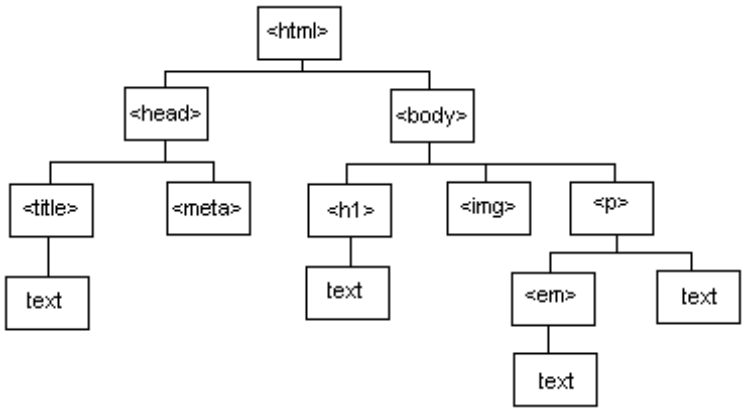


Figure 42: Graphical representation of the DOM for a simple web page

```
<html>
  <head>
    <title>text</title>
    <meta charset="utf-8">
  </head>
  <body>
    <h1>text</h1>
    <img src="img.jpg">
    <p><em>text</em> text</p>
  </body>
</html>
```

Figure 43: Skeleton code for an HTML file

Next, FERN traverses the DOM and builds lists of paragraphs, headings and blockquotes that are siblings; it ignores empty and hidden elements (those with an ancestor that has a display none or visibility hidden CSS property). If the code contains non-empty paragraph elements, divs that contain no block-level elements (those that can contain text, data and other elements) are replaced by <p> tags, and the lists are re-built.

Next, for each list, FERN counts the number of paragraphs that are terminated by sentence-ending punctuation. The first list that contains at least two sentences is selected, which avoid confusing user contributed comments sections with article content. If no list contains more than two sentences, the one with most characters is selected. Some news items contain a single paragraph, or are video reports with a caption.

Next, FERN removes the first element if it is the article’s title, and concluding elements that do not end with punctuation, e.g. copyright notices, and lists of

recommended articles. Searches are performed to find the opening paragraph if it is not a sibling of the rest of the article (see Section 6.3.2). If the extracted content is less than 200 words, the extraction is re-attempted after transforming divs that contain no block elements into paragraph elements. The second attempt is used if it produces in a longer article than the first.

This algorithm works well for over 90% of articles returned by Google News, a level of performance that was acceptable for this project. For a few sites on which the algorithm fails, domain specific filters are used to pre-process the raw HTML before it is parsed by JSoup. A separate extractor was developed for Wikipedia.

6.4 Individual vocabulary profiles

FERN tracks each learner's reading and reviewing, and constructs a profile of their vocabulary knowledge, as depicted in Figure 20. This profile contains the following data for each word family: the number of encounters and gloss accesses during reading, and the number of encounters and errors in vocabulary review exercises. This data is used to compute individualised difficulty ratings of texts (see Section 6.5).

6.4.1 Monitoring reading

Readers often skip sections or stop reading articles part-way. In order to predict what they actually read, articles are divided into sections (see Section 4.3.1). One section is displayed in high-contrast at a time, and the rest of the text is dim (see Figure 14). Learners click a section to make it high-contrast. FERN measures the time spent on each section, and pauses timing when the learner goes to another web page or application, or clicks *Pause*, which hides the article until they resume.

When learners finish reading an article, FERN disregards sections with reading speeds above 400 words per minute—twice the typical reading speed of lower intermediate learners. This threshold is high enough to accept sections that learners read in their entirety, but low enough to filter those that are partially read (see Section 4.6.2). Vocabulary profiles record encounters and gloss accesses for words in sections that are likely to have been read.

Having short sections is advantageous for collecting fine-grained timing data and making accurate predictions. However, to prevent learners from being frustrated by

having to click through sections of short sentences, a minimum section length is 300 characters.

This approach enables teachers to track words read by students without them completing reading logs. In order to cheat the system, students need to click through article sections at speeds less than 400 words per minute, which is tedious for someone who is not actually reading. Learners are less likely to cheat, because the effort required to cheat is comparable to that required for correct usage. Teachers can further deter cheating by requiring completion of post-reading cloze exercises, which are difficult to complete without reading the articles first.

FERN records reading speeds for each article, and graphs changes in speed over time. Outliers are eliminated using *statistical process control*, with the first seven "stable" data points used to establish the initial mean. Outliers are defined as those that are less than 0.5, or greater than 1.7 times the moving mean for the last five data points.

These reports provide teachers and learners with indications of progress in learning the L2.

6.4.2 Supplementary exercises

L2 word encounters and gloss accesses in mixed-language texts are logged in the reading columns of the vocabulary profile. Performance in multi-choice, word match, and fill-the-blanks are logged in the review columns. At the completion of an exercise FERN records the words reviewed and those the learner made errors with. Reviews and errors in each exercise are weighted equally. For fill-the-blanks exercises only target words (i.e. those that fill the blanks) are recorded in the vocabulary profile; nothing is recorded for non-target words, because the exercise can be completed without reading all the words.

6.5 Computing article difficulty

FERN's difficulty algorithm is influenced by both readability formulas and research on incidental vocabulary acquisition through reading. The algorithm incorporates frequency, word family, and cognate lists in combination with learners' vocabulary profiles to estimate the percentage of unknown words in an article.

6.5.1 Individualised difficulty algorithm

Using vocabulary profile data, FERN estimates the likelihood that a word causes difficulty during reading on a scale of 0–100. Words that have not been encountered in past reading are assumed to be unknown (i.e. difficulty score of 100), unless they are cognates or off-list words (i.e. not contained in FERN's dictionary). Off-list words are assigned a difficulty score of 0, because the majority of these are proper nouns, which are generally assumed to cause few problems during reading (Nation, 2006).

Previously encountered words with no gloss accesses are assumed to be known, and have scores of 0. Words with past gloss accesses are assumed to be partially known; their scores are estimated by first computing the following index:

$$index = \max(0, readingEncounters - 2 \times glossAccess) + 2 \times reviewEncounters - reviewErrors + 2 \times currentArticleEncounters$$

This formula weights encounters during past reading less than reviews in vocabulary exercises, which require more deliberate attention to the words. Gloss accesses are weighted higher than review errors, because they are a stronger indicator of the difficulty words will pose during reading.

Unfamiliar words that are encountered for the second or third time in an article generally cause less trouble than the first encounter, due to learners accessing glosses or inferring their meaning from context. The *currentArticleEncounters* parameter refers to these earlier encounters with words and accounts for the incremental learning that occurs with each encounter.

The *index* is used to look up a score in a *difficultyArray* that is based on the premise that word difficulty reduces at an exponentially decreasing rate with each encounter (see Section 2.4.2). When *index* is greater than 8 the score is 0, and when *index* is less than 1 the score is 90.

$$difficultyArray = [90, 60, 36, 22, 13, 8, 5, 3, 2, 1, 0]$$

$$wordDifficulty = difficultyArray[index]$$

For example, if a learner encounters a word ten times during reading and accesses glosses four times, but has not reviewed it in vocabulary exercises, the index is 2, and the difficulty score is 36. A word encountered five times during reading, with

two gloss accesses, three reviews and one review error, has an index is 4 and a difficulty of 13.

The difficulty of previously-unseen cognates is determined using FERN's list of English-Spanish cognates, which are graded on a scale one to five (see Section 6.6). Easy cognates are assumed to pose little difficulty during reading. The index formula for cognates is as follows:

$$index = currentArticleEncounters + 7 - cognateDifficulty$$

The index is used to look up values in the above difficulty array. The most difficult cognates are assigned a score of 36, and the easiest cognates have a score of 3.

Article difficulty scores are computed by averaging the difficulty scores for each word, in order to estimate the percentage of words a learner will find difficult during reading. These scores are displayed next to news search results in FERN.

Note that FERN's word difficulty score is a simplified model that ignores the time element of vocabulary learning and forgetting.

6.5.2 Initialising the vocabulary profile

New learners start with empty profiles, and all words are assumed to be unknown unless they are cognates or off-list words. Initial search results have no individualisation; scores reflect the proportion of cognates, proper nouns, and vocabulary repetition in an article. Analysis of 260 news articles across a range of subjects gave an average length of 337 words and a difficulty score of 27% when searching as a new user. On average 32% of words were cognates and 11% were off-list words. These initial scores indicate the relative difficulty of articles.

As learners begin to read, scores quickly fall below 10%, because a small number of high frequency words account for a large percentage of language use (e.g. the top 100 English word families account for 50% of text). The more learners read, the more individualised the scores become. Section 8.3.2 examines how quickly the difficulty scores fall below 2% when narrow reading is applied.

6.5.3 Word families

The vocabulary profile groups words by word family, because receptive knowledge of one word form implies knowledge of the others in the family. For new users,

incorporating word families makes the difficulty scores become more meaningful with less reading than if words were treated separately.

6.6 Cognate lists

Cognates are words from different languages that have shared etymological origin, e.g. *ejemplo* which along with the English *example* descends from the Latin *exemplum*. The spelling and meaning may be similar or may have diverged over time as the languages developed separately. Cognates are significant in L2 learning because in general they are easier to learn than non-cognates. FERN's cognate list is instrumental in computing difficulty scores (see Section 6.5) and preparing mixed-language texts (see Section 7.5).

The English-Spanish cognate lists that exist on the Internet are far from complete. FERN uses a list that is automatically extracted and graded according to the likely difficulty that congnates present when encountered for the first time. This section details the design and extraction of this list.

6.6.1 Cognate recognition

Several factors affect learners' ability to recognise and learn cognates. The first is orthographic similarity. Cognates with identical spelling, such as *color*, which is identical in US English and Spanish, are easily understood by readers who have not previously encountered them. Those with similar spelling, such as the Spanish *atleta* meaning *athlete*, are also easily understood. Less similar ones, like *agradecido* meaning *grateful*, are more difficult. However, similarity provides a hook that makes recognition of meaning and the remembering of meaning easier than for non-cognates.

Knowledge of L2 affixes also affects cognate recognition. For example, knowledge that *-mente* changes adjectives to adverbs, and that the Spanish prefix *des-* is equivalent to *in-* and *un-* in English, makes it easier to recognise that *desadvertidament* means *inadvertently* in English. L2 pronunciation knowledge is also beneficial. For example, the Spanish *cebra* is pronounced similarly to its English cognate *zebra*, and *física* is pronounced similarly to *physics*.

Another factor is learners' L1 vocabulary knowledge. For example, English natives who know that *lactic* means *of or relating to milk* will easily learn *lechar* (to milk)

and *leche* (milk). Cognates are more difficult if the L1 word is low-frequency. For example, the pair *patina-pátina* is only recognisable as cognate to those who know that the English word *patina* refers to *the green or brown oxidation layer that forms on bronze and copper*. Knowledge of additional languages also increases one's ability to recognise cognates.

The existence of additional orthographically-similar L1 words makes learning some cognates more difficult. For example, the Spanish verb *designar* is easily mistaken for *design* instead of its correct cognate *designate.* This situation is particularly problematic when the incorrect word has a higher frequency than the correct one.

6.6.2 False friends, false cognates, loan words

False friends are words with similar forms but different meanings. A popular example is the Spanish *embarazada* (pregnant) and the English *embarrassed*, which have very different meanings despite the orthographic similarity and connected origins. Another example is the Spanish *actual* (current) which has a very different meaning to the English *actual*. False friends hinder comprehension when readers assume that they are cognates.

Partial false friends have overlapping meanings. For example, the Spanish verb *asistir* most commonly means *to attend, to be present,* but occasionally means *to aid, to assist.* Another example is the Spanish *lima*, which is a *lime fruit* or a *file tool for metal working*. These words are cognates in one sense and false friends in the other; they can aid and hinder comprehension depending on the context.

Other cognates share meanings in the literal sense, but have different meanings in figurative usage in idioms and metaphors.

False cognates are words with similar forms but different etymologies. Most false cognates are also false friends because they have different meanings; however, a few false cognates coincidentally have the same meaning, and thus are not false friends, such as Spanish *mucho* (from Latin *multas*) and English *much* (from Proto-Germanic *mikilaz*). A larger group of words exists that are true cognates (i.e. have shared etymology and similar form), but are false friends, because different meanings have evolved in each language.

Loan words are those borrowed directly from one language to another. They differ from cognates, which stem from a common language—Latin in the case of English

and Spanish. English loan words in Spanish, known as anglicisms, are particularly common in technology, sports and food related language. Some are adopted with no spelling changes, as with *golf, rugby* and *waterpolo*; whereas, *básket, beisbol* and *surf* have adapted spelling from their English equivalents *basketball*, *baseball* and *surfing*. In some cases the meaning is also adapted, as with the Spanish word *footing*, which means *jogging* in English.

FERN's cognate list is designed to include all words that have similar form and meaning.

6.6.3 Detecting orthographic similarity

FERN's approach to detecting cognates is influenced by several algorithms for detecting orthographic and phonetic similarity between pairs of words.

The Levenshtein distance, also known as edit distance, computes the minimum number of single-character edits (insertion, deletion, and substitution) required to change one word to another. For example, *beisbol* requires one deletion, two substitutions and two insertions to form *baseball*, giving an edit distance of five. Damerau–Levenshtein distance is a variation that allows transposition of adjacent characters in addition to the other edits.

The Longest Common Subsequence Ratio computes the longest string of characters that appear in two strings in left-to-right order, but not necessarily consecutively, and divides it by the length of the longer of the two strings. For example, the longest common substring for *beisbol* and *baseball* is *bsbl*, so the ratio is *4 / 8 = 0.5*.

$$LCSR(x, y) = \frac{|LCS(x, y)|}{\max(|x|, |y|)}$$

The Sørensen–Dice index (Dice, 1945) is a string similarity metric with the following form:

$$DICE(x, y) = \frac{2 \times |bigrams(x) \cap bigrams(y)|}{bigrams(x) + bigrams(y)}$$

The denominator adds the numbers of character bigrams in each string, and the numerator counts the bigrams common to both strings. For example, there are no common character bigrams for the cognate pair *beisbol* and *baseball*, yielding an index of 0. For *física* and *physics* there are two common bigrams (*si* and *ic*) and a total of 11, which gives a Dice score of 0.36.

$$s = \frac{2 \times 2}{5 + 6} = 0.36$$

Brew and McKelvie (1996) use a variant called *Extended Dice*:

$$XDICE(x, y) = \frac{2 \times |xbigrams(x) \cap xbigrams(y)|}{\text{x}bigrams(x) + xbigrams(y)}$$

In this formula, *xbigrams* includes two letter pairs formed by deleting the middle letter in character trigrams, in addition to adjacent character bigrams. This metric gives the following score for *beisbol* and *baseball*:

$$s = \frac{2 \times 2}{11 + 13} = 0.17$$

6.6.4 Detecting phonetic similarity

Phonetic algorithms detect homophones. For example, the American Soundex was designed to detect similar surnames in US census data (McPeake, Chen, Whalen, & Vezina, 2004). The algorithm uses the following rules to convert names to a code consisting of the name's first letter, followed by three digits:

1. Retain the first letter and drop all other vowels as well as *y*, *h* and *w*.
2. Replace *b, f, p, v* with 1; *c, g, j, k, q, s, x, z* with 2; *d, t* with 3; *l* with 4; *m, n* with 5; *r* with 6.
3. If two or more letters with the same number are adjacent or separated by *h* or *w* in the original name, retain the first letter only.
4. If there are less than three digits in the code, append zeros. If there are more than three, drop the excess digits.

Using this algorithm *baseball* and *beisbol* are both coded as B214, but *physics* and *física* are P220 and F220. A variant called the *Reverse Soundex* retains the last letter of the word instead of the first, which gives the same codes for both of these cognate examples. Metaphone, Double Metaphone and Metaphone 3 are commercial algorithms that use a large set of proprietary rules to improve recall and precision of name comparisons.

Kondrak (2000) developed the ALINE algorithm to align words phonetically. The algorithm compares the letters in words and assigns different weightings depending on the similarity between the sounds the letters make.

6.6.5 Creating cognate lists

There are several approaches to creating cognate lists. LeBlanc and Séguin (reported in Frunza & Inkpen, 2010) used general-purpose dictionaries to manually create a list of 23,160 French-English cognates; 6,447 had identical spelling, ignoring diacritics. The list accounted for 30% of the words in the dictionaries.

Simard, Foster, and Isabelle (1993) automatically detected cognates by performing word alignment on a corpus of paragraph aligned bilingual text. Words were deemed cognate if the first four letters were identical. They avoid false friends because orthographically similar words that appear in sentence translations are likely to be cognates. Brew and McKelvie (1996) performed this same task using a variant of the Dice coefficient and the Longest Common Substring to test similarity.

Kondrak (2001) tested phonetic similarity using ALINE and detected semantic similarity by checking whether two words shared a common gloss in a dictionary. Glosses with the following orthographic differences are deemed equivalent:

1. Spelling errors or variants: *plow* and *plough*;
2. Morphological differences: *ash* and *ashes*;
3. Determiners: *a mark* and *mark*;
4. Adjectival modifiers: *small stone* and *stone*;
5. Nominal modifiers: *goose* and *snow goose*;
6. Complements and adjuncts: *island* and *island in a river*;
7. Synonymy: *grave* and *tomb*;
8. Small semantic changes: *fowl* and *turkey*.

WordNet was used to associate words in categories 7 and 8.

Hauer and Kondrak (2011) and Frunza and Inkpen (2010) detected cognates using machine learning with orthographic similarity features. Mulloni and Pekar (2006) applied machine learning to a list of cognate pairs to determine common orthographic transformations, which they used to detect additional cognates. Mulloni et al. (2007) improved this algorithm by looking up cognate candidates in a corpus and determining how many frequently neighbouring words were translations of each other, on the basis that orthographically similar words that are frequently surrounded by semantically-related words are likely to be cognates. A

similar approach was taken by Nakov et al. (2007) using Google search results to determine lists of co-occurring words.

6.6.6 Creating FERN's cognate list

Creating FERN's cognate list is a two-step process:

1. Determine candidate cognate pairs using the English Wiktionary:
    a. English entry translation tables,
    b. Spanish entry definitions,
    c. Latin descendants.
2. Filtering based on phonetic similarity computed with a variation of Soundex. Reject if the length of either word is less than three letters.

The English Wiktionary contains multiple sources of English-Spanish translations (see Section 5.6.3). The entries for English words contain translation tables. For example, the translation table for *athlete* includes three Spanish translations: *atleta*, *deportista* and *atletica*. Each Spanish translation is added to the candidate cognate list, along with its English equivalent *athlete.*

The English Wiktionary's Spanish entries give English definitions. For example, the entry for *atletica* gives the English definitions *athlete (competitor in a sport)* and *a physically fit person*. The information in brackets is ignored and the translations are added to the candidate cognates. In this case the first definition is ignored because it is a duplicate of the *athlete-atletica* pair from the English entry for athlete.

For definitions that are two words long, stop words are ignored. Most English definitions for verbs in Wiktionary begin with *to*, e.g. *competir* is defined as *to compete*. Other two word definitions are joined into a single word by removing the space. Definitions of three or more words are ignored.

Candidate cognates are also sourced from Latin entries, which list descendants from Latin-based languages. For example, the entry for *dictiōnārium* lists *dictionary* as an English descendant and *diccionario* as a Spanish one. If the Latin entry lists a Spanish descendant but not an English one then the English definition is used. For example, the entry for *āthlēta* lists *atleta* as a Spanish descendant but lists no

English descendant. Because *wrestler* and *athelete* are listed as definitions, the word pairs *aleta-athelete* and *atleta-wrestler* are added as candidate cognates.

The phonetic similarity of candidate cognates is determined using the following set of rules:

1. Pre-process each word:
    a. Convert to lowercase;
    b. Replace accented characters with unaccented equivalents.
2. Compare the words:
    a. Reject if the length of either word is less than three letters;
    b. Accept if one word starts with the first four letters of the other.
3. Phonetic modifications:
    a. For the English word replace the *t* in *tion, tious, tial, tian, tiat, tiab, tiar, tien* with a *c*. For the Spanish word replace the *t* in *tiar* and *tian* with a *c*;
    b. In both words replace *ph* with *f*; *d* with *t*; *v* and *p* with *b*; *m* with *n*; *j* with *g*; each of *z, x, s, q,* and *k* with *c*;
    c. Remove *y, h* and *w*;
    d. Remove double letters;
    e. Re-try step 2;
    f. Remove vowels;
    g. Re-try step 2.
4. Accept if the shorter word is a subsequence of the longer one.

This process yields 12,226 cognates. The list matches Spanish headwords with corresponding English cognates. If multiple English words are cognate with one or more forms of the Spanish word family, orthographic similarity is used to choose the best English word.

6.6.7 Grading cognate difficulty

FERN was trialled for two semesters with a university Spanish class (see Chapter 8). The approach used to grade cognate difficulty in the first semester was different from the second.

In the first trial, cognates were graded on a scale of one to three based on edit distance, where one corresponds to distance of 0–1, two is a distance of 2–3, and

three is 4 or more. L1 frequency was also graded on a scale of one to three, with a one for top 2000 English words, two for mid-frequency, and three for low-frequency words. The overall difficulty is computed on a scale of one to five as follows:

$$Difficulty = LexicalSimilarityScore + L1Frequency - 1$$

The second trial removed orthographic similarity and graded cognates on a scale of one to five based on L1 frequency.

7 Supplementary vocabulary exercises

FERN automatically creates vocabulary exercises that introduce new vocabulary and review words that learners have difficulty with during reading, (see Section 4.4 and 4.5). Learners' performance in these exercises provides useful data for their profiles, which helps to make their news search results more meaningful. The first part of this chapter describes the importance of vocabulary exercises in the context of extensive reading, and describes the design and implementation of FERN's vocabulary exercises.

In addition to "traditional" exercises, FERN incorporates a less-conventional activity we call "mixed-language texts", which consists of two languages intertwined, usually maintaining the word order of the dominant language. FERN automatically generates mixed-language texts that are predominantly L1 with a sprinkling of L2 words. These texts are useful for introducing new words to learners who lack sufficient vocabulary knowledge to read entirely L2 texts, or to allow them to practice L2 vocabulary while reading text that is not available in their L2.

The second part of this chapter presents the case for using mixed-language texts for L2 learning, and discusses why they are not more widely used—a key reason being that manually creating high-quality mixed-language texts is a complex and labour intensive task. For example, a project at Brigham Young University (BYU) Provo campus took several months to transform an existing text that is available in both L1 and L2 into a 100 page paper-based mixed-language reader[73]. This reader progressed from including a sprinkling of L2 cognates at the beginning, to being almost entirely L2 at the end. Another shortcoming of these texts was that the progression of difficulty was fixed, so it will be too easy for some, and too difficult for other learners. This section describes how FERN overcomes these challenges, by automatically creating mixed-language texts that are tailored to the ability of the learner.

[73] From personal correspondence with the BYU Professor Ray Graham in June 2012.

7.1 Background to supplementary exercises

While extensive reading is an effective approach to learning high-frequency vocabulary, it does not provide by itself sufficient encounters to learn all of the 5000 plus word families necessary to comfortably read authentic text (Nation, 2009). This is especially true with graded readers, the most advanced of which contain few words outside the top 3000 word-families. Cobb (2007) used a corpus of just under 300,000 words of Jack London stories to compute encounters with mid-frequency word families. The collection contained at least one occurrence of 817 of the 3000-level word families, with 469 appearing six or more times.

As described in the following sections, many researchers advocate supplementing reading with direct study of low-frequency words. However, there are several potential drawbacks to this idea. First, word knowledge does not necessarily transfer well from direct study to text comprehension, especially when there is a delay between the two. Second, the list of words to review quickly grows to be so large that the time required to review them consumes an inordinate amount of reading time. Third, preparation of exercises generally requires significant work by the teacher or learner. This chapter addresses each of these drawbacks.

7.1.1 Pre-reading exercises

Pre-reading exercises introduce vocabulary and background information before reading difficult texts, with the goal of aiding comprehension and vocabulary learning.

With respect to cognitive load theory (Sweller, 2005; Sweller, Van Merrienboer, & Paas, 1998), understanding text with a large proportion of unfamiliar words entails two cognitively expensive tasks: decoding unknown words and comprehending the meaning of sentences. Direct vocabulary study of unknown words before reading develops initial schema for those words, so that during reading less working memory is required for decoding, thus freeing up cognitive resources for text comprehension.

Studies demonstrate a positive effect for pre-reading vocabulary exercises on both L1 and L2 reading (Webb, 2009, p. 442, Carlo et al., 2004; Zimmerman, 1997). Webb (2009) had participants study lists of L2 words with L1 equivalents before reading an L2 text. One group studied by looking at the L1 list and attempting to

recall L2 translations. A second group viewed the L2 words and recalled the L1 equivalents. Results showed that studying the L2 words had a larger positive effect on reading comprehension than studying the L1 words.

Alessi and Dwyer (2008) determined that pre-reading vocabulary study reduced reading time by 41%, but had little effect on comprehension. However, subjects in the pre-reading study group spent twice as long on vocabulary study as reading, resulting in a total time that was double that of the reading only group.

More research is required to determine the true effectiveness of pre-reading vocabulary study. Future studies need to consider the effect of pre-reading activities on vocabulary learning and retention, in addition to reading speed and comprehension. They should also reduce the time spent on vocabulary study, in order to determine whether *total* study time can be reduced by pre-reading activities. The quality of the pre-reading exercises is also a key factor. As demonstrated by Webb (2009), some are more effective than others. Furthermore, optimal efficiency requires that learners only review unknown words; study of known words is better spent on reading.

7.1.2 Post reading exercises

Post-reading exercises are often used to reinforce incidental vocabulary learning that occurs during reading—an approach known as reading-plus. Paribakht and Wesche (1996, 1999, 2000) determined that reading-plus results in more vocabulary learning and retention than spending the same amount of time reading several thematically-related texts that each contain target vocabulary. Other studies also indicate that reading-plus is more efficient than solely reading (Laufer, 2003; Mondria, 2003; Peters, Hulstijn, Sercu, & Lutjeharms, 2009).

Mason and Krashen (2004) questioned these results. In their study, a *listening-only* group listened to a story for 15 minutes then did a translation test of 20 target words. A *listening-plus* group listened to the same story, followed by an hour of post-listening activities that included reading the same story. In a delayed post-test, students in the listening-plus group demonstrated vocabulary gains three times larger than the listening only group. However, in terms of the words gained per minute of study, the listening only group came out on top. The conclusions are somewhat dubious, because the effort required to learn each target word is unequal.

Table 14: Effect of during-reading treatments relative to reading-only (RO)

Treatment	Comprehension	Vocabulary learning
Self-selected output (SO)	Comparable	Lower
Cued output (CO)	Lower	Higher

The additional words learnt by the listening-plus group may have been more difficult than those learnt by the listening only group.

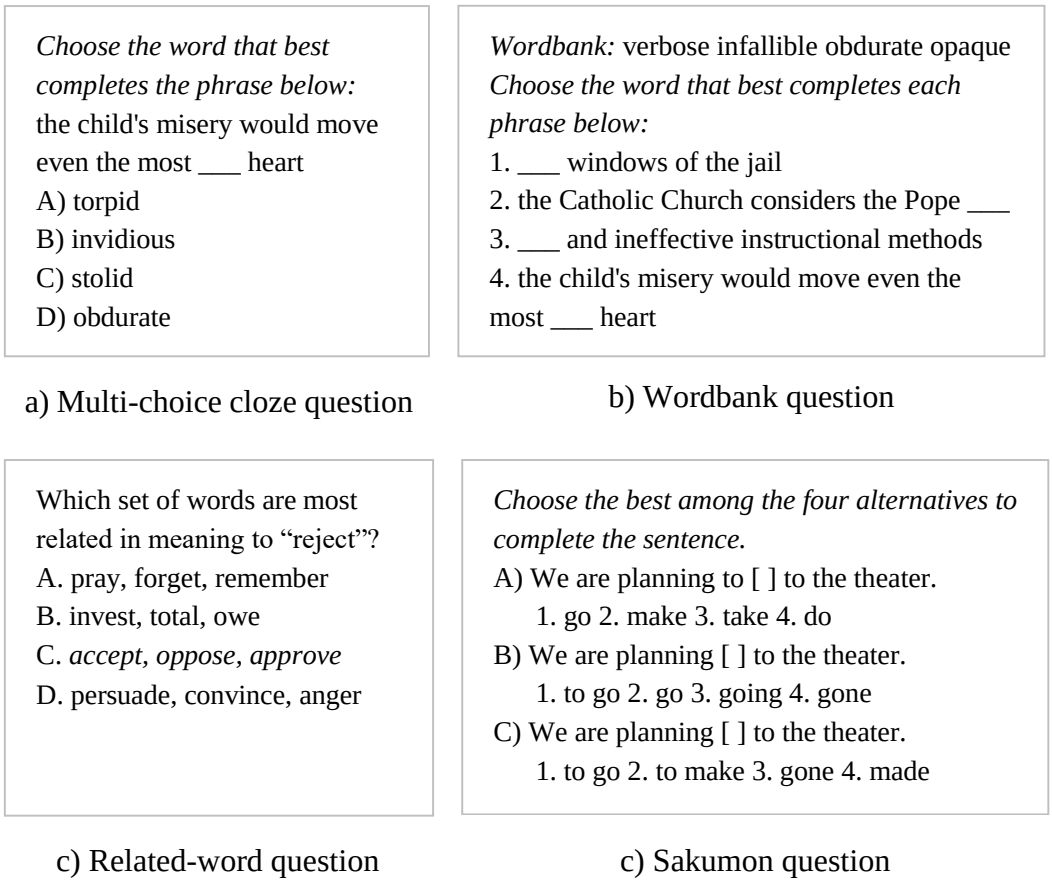
Mason and Krashen (2004) also investigated the effect of writing L1 or L2 book summaries as a post-reading exercise. Their studies suggest that writing short L1 summaries is more time efficient that L2 summaries, especially for lower ability learners.

As with pre-reading, the efficiency of post-reading exercises would be improved by ensuring that learners only spent time on words they did not already know. Time spent on words that are well-known is better spent on unfamiliar ones.

7.1.3 During reading exercises

Rott (2004) investigated the effect of vocabulary reinforcement activities by having 35 learners read three 350 word texts, divided into four paragraphs. Each text contained four target words each of which appeared once per paragraph. Learners read each text under one of three conditions, and then took a comprehension and vocabulary test. The “reading only” (RO) control group read the entire text without interruption. The “cued-output” (CO) group wrote a sentence about each target word in between reading paragraphs. The “self-selected output” (SO) group noted ten important words or phrases after reading each paragraph.

Table 14 shows that SO and RO had comparable comprehension, while CO reduced comprehension. However, CO resulted in improved learning of the target words relative to RO, while SO reduced the learning. The study concluded that RO produced the best balance between comprehension, vocabulary learning and time efficiency. Rott also suggested that reading texts twice, with vocabulary and comprehension activities at the end of the first reading, is more effective than incorporating activities between paragraphs.



a) Multi-choice cloze question

b) Wordbank question

c) Related-word question

c) Sakumon question

Figure 44: Automatically-generated question examples

7.2 Automatically generating vocabulary review exercises

FERN incorporates four additional automatically-generated vocabulary review exercises. This section discusses approaches to automatically generate various types of vocabulary review exercises. We then present details of FERN's automatic generation, and describe how the design of each activity affects the usefulness of the feedback obtained for learner profiles.

7.2.1 Background to automatic exercise creation

A general approach to automatically generating multiple-choice quiz questions is to select the answers; generate the questions from a corpus; and generate distractors. Brown et al. (2005) used this approach to integrate six types of questions into the REAP tutor: definition, synonym, antonym, hypernym, hyponym, and cloze questions. Figure 44a is a cloze question generated from example sentences in WordNet. The other type is the wordbank-style question in Figure 44b, which is created using frequency counts, antonym and synonym data in Wordnet. The

system chooses distractors with similar frequencies to the target words. Heilman and Eskenazi (2007) included related word questions, such as that in Figure 44c.

Evaluation of exercise quality with teachers revealed that 40% of questions were acceptable for classroom use. Pino, Heilman and Eskenazi (2008) refined the algorithms and increased the percentage of high-quality questions to 66%.

Mitkov, Ha and Karamanis (2006) extracted key noun phrases from a text book using natural language processing tools and frequency counts. Questions are created by applying one of several rules to the answer. For example, the question "Which kind of verbs require objects?" is created by applying the *which-kind-of* rule to the answer *transitive verbs.* Distractors that are semantically close to the answer are chosen using hypernyms in WordNet. The distractors *modal verbs*, *phrasal verbs* and *active verbs* are chosen for the answer *transitive verbs*. Some automatically generated questions are poor quality; thus manual checking and acceptance is recommended.

Hoshino, Huan and Nakagawa's (2008) Sakumon system extracts questions from online news articles. The distractors in Figure 44d, example A, were selected using WordNet, and those in example B were created using a morphological generator. Both approaches are used for example C.

Liu, Wang, Gao and Huang (2005) developed a word-sense disambiguation technique for selecting sentences in a corpus that contain usage of specific word senses. These are used for cloze questions. The system selects distractors that have the same part of speech as the answer, and form high-frequency collocations with the words either side of the blank. 60% of generated questions are usable.

Sumita, Sugaya and Yamamoto's (2005) system selects candidate distractors for cloze questions using WordNet. The system then performs web searches for the cloze question, with the blank filled by each candidate distractor. They are deemed unsuitable if the search returns any hits.

Goto, Kojiri, Watanabe, Iwata and Yamada (2010) use machine learning and statistical methods to select sentences and distractors for cloze exercises. Meurers et al. (2010) developed WERTi, which creates a fill-in-the-blank exercise from the text on web pages. Wu, Franken and Witten (2010) developed FLAX, which generates vocabulary exercises based on texts in a corpus.

Most of these studies could be adapted to other languages. Other studies have developed question generators specifically for other languages, such as Basque (Aldabe, de Lacalle, Maritxalar, Martinez, & Uria, 2006).

These studies describe promising approaches to automatically generate vocabulary review exercises that present words in the context of sentences. However, they require teacher mediation to filter low-quality questions.

FERN incorporates exercises that were simple to implement and do not require teacher mediation. Three of these exercises use bilingual word lists, which are an efficient method for achieving results that are excellent in the short term, but often disappointing long term (Groot, 2000). This thesis suggests that they are useful as supplements to exercises that present words in context.

7.2.2 FERN's fill-the-blanks exercise

FERN's fill-the-blank exercises divide articles into no more than four sections, with six blanks in each. The number of blanks is a key determinant of difficulty, and initial user trials indicated that exercises with more sections and more blanks led to learner fatigue and frustration. Thus, six blanks was deemed suitable (see Section 8.2).

In order to simplify implementation, FERN only blanks words that have a one-to-one alignment between the L2 text and its translation. Blanks are at least 3 words apart and 15 words apart on average. Capitalised words and numbers are never blanked out, because they are too easy to guess. Priority is given to previously encountered words with a difficulty score greater than 50, and no word is blanked out twice in the same section.

The user interface is carefully designed to minimise extraneous cognitive load. For example, the answers are presented in alphabetical order, and answers are darkened as they are used. These features are intended to reduce the time and energy spent locating answers.

7.2.3 FERN's vocabulary review exercises

Each vocabulary exercise reviews eight words, a size that was determined as suitably challenging for the target audience. In order to select the target words, FERN gets a candidate list of 24 previously-encountered word families that have

the highest difficulty scores, giving priority to words that the user recently had difficulty with during reading. If learners have not had difficulty with at least 24 words, the remainder of the candidate list is filled with high-frequency words.

Next, FERN randomly selects eight of the 24 candidate word families, for which it randomly selects an L1 word form and its corresponding L2 translation. If an L1-L2 word pair are identical cognates (e.g. English and Spanish *color*), or one of the words contains the other (e.g. English *abandon* and Spanish *abandonar*), that word pair is deemed too easy, and is replaced by another random selection.

FERN's multiple-choice exercises are designed to maximise the utility of data logged to learners' vocabulary profiles. The first type has L1 questions with five L2 answers from which learners keep choosing until they get it correct. A correct first choice indicates that learners either know the target L2 word, or chose correctly by a process of elimination or guessing. An incorrect guess is a firm indication that they have difficulty with both the target word and the incorrectly guessed word. The second type has L2 questions and five L1 answers. Because the answers are L1, incorrect guesses only indicate difficulty with the target L2 word.

Each target word is tested once with each question type, and words are logged as correct on the vocabulary profile if both questions are answered correctly. Testing words with both question types reduces the odds that learners get them correct by guessing.

As with the other exercises, the user interface is carefully designed to minimise extraneous cognitive load. For example, the answers are presented in alphabetical order, and the exercise automatically advances to the next question when they are answered correctly, with no extra button clicks required.

7.3 Background to mixed-language texts

The literature refers to these texts by several names: "bilingual text", "diglot weave", "Burling reader", "sandwich story", "macaronic text", and as texts containing "code-mixing" or "intra-sentential code-switching". This section describes the context in which each of these terms has been used, with particular focus on uses that relate to L2 learning.

7.3.1 Historical uses of mixed-language texts

Mixed-language texts have been used for a variety of purposes throughout history. In 11th–16th century England, for example, some scribes (the "personal assistants" of their day) wrote business documents in a mix of Middle English and either Medieval Latin or Anglo-Norman (Wright, 1998). The language mixing was quite orderly, such as using a Latin base with many nouns and adjectives written in English. Evidence suggests that these texts were not the result of the scribes having limited Latin ability; they were written in mixed-language to cater for the Latin abilities of the readers of the documents (Wright, 2002).

During these centuries, mixed language was used in England for sermons and other religious texts, legal texts, medical texts and private prose such as letters and diaries. It was also common in literary text such as poetry, including carols and hymns. In English plays, Latin was used to quote religious texts (Schendl, 1996).

The term *macaronic language* usually refers to mixed-language use in poetry. Schendl (1997) analysed a corpus of 80 mixed-language poems originating from 13th–15th century England. In most texts English was the dominant language with a sprinkling of Latin; some texts contained equal portions of both, and a few mixed French, English and Latin. Most language switches occurred at the boundaries of major constituents, such as noun, verb and prepositional phrases. Functional reasons for switches included quoting religious text, word-play and rhyming.

Mixed-language text is not a phenomenon confined to the distant past in England; these texts have been common for centuries throughout the highly polyglot populations of Europe, and are commonly used today in online discussion forums, tourism websites, US Latino novels and text-messaging (Sebba, Mahootian, & Jonsson, 2011).

7.3.2 Psycholinguistics studies in bilingualism

Bilinguals are people who are fluent in two or more languages. Psycholinguists have studied bilinguals' responses to mixed-language input, in order to develop models of how they store and process their L1 and L2 knowledge. This section discusses how the predictability of language switches affects the ease with which they are processed, and describes the implications of this research on the use of mixed-language texts as an L2 learning activity.

Kolers (1966) had 24 French-English bilinguals read aloud three types of text: unilingual, alternating sentences (language switched each sentence) and mixed-language texts (half of the L1 words replaced haphazardly by L2 translations, with French word order in half the texts and English in the others). Comprehension of alternating-sentence texts was comparable to unilingual L1 texts. Mixed-language text comprehension was half way between that of unilingual L1 and L2 texts. Subjects comprehended mixed texts with L1 word order marginally better than those with L2 word order. Times for reading aloud unilingual L2 texts were 50% slower than for L1 texts. Times for alternating-sentence texts were halfway between unilingual L1 and L2 texts. Times for mixed texts were 7.5% slower than L2 texts, which Kolers attributed to a 300–500ms "language-switching cost" for changing languages in the middle of a sentence.

Macnamara and Kushnir (1971) repeated Kolers' experiment with subjects reading silently instead of aloud. This approach reduced the switching cost to 170ms. They suggested that reading aloud incurs extra costs switching pronunciation modes, and increases cognitive load relative to silent reading. In a second experiment, the time to comprehend sentences increased with the number of language switches. The results of these studies indicate that mixing languages slightly slows readers, but results in significantly better comprehension that reading L2 texts.

Wakefield et al. (1975) had English-Spanish and English-Korean bilinguals read sentences with switches either at major constituent boundaries or within constituent boundaries. The following example marks a major constituent boundary with a forward slash.

> The number of people taking up parachuting as a sport / is definitely increasing.

This sentence has a switch at the boundary:

> The number of people taking up parachuting as a sport *está definativamente aumentando.*

This one switches within a major constituent:

> The number of people taking up *el deporte del paracaídas está definativamente aumentando.*

Table 16: Reading times for mixed-language text (Chan et al., 1983)

Text Type	Reading time
Unilingual Cantonese	147.3s
Unilingual English	173.5s
Naturally switched	146.0s
Noun switched	159.5s
Randomly switched	175.4s

The time to process sentences was 20% slower for language switches within constituent boundaries for both Spanish and Korean bilinguals. This result makes it clear that certain types of switches are more cognitively-demanding than others.

Chan, Chau and Hoosain (1983) had students read a mixed-language article from a university magazine in Hong Kong. This predominantly Cantonese text had 567 syllables, with 67 words relating to university life in English; they termed this text "naturally switched", because the switches were those included "naturally" by the article's author. Four modified versions of the text were prepared: unilingual Cantonese and English versions; a mixed text with all 71 nouns in English; and another with 71 randomly selected words switched to English.

As shown by the results in Table 16, readings times were faster for the original naturally-switched text than the unilingual L1 version. Because Hong Kong universities use English as the medium of instruction for most courses, students are more familiar with university-related vocabulary in English than Cantonese. The ease with which they read and interpret these English words offsets the switching costs in the mixed text.

Times for the noun-switched text were 10% slower than the naturally-switched text, probably due to less familiarity with some of the English nouns. Times for the randomly switched text were 10% slower than the noun switched version, and marginally slower than the unilingual English one. These results suggest that familiarity with the L2 words and consistency of the switches affects reading time. They also illustrate a situation in which mixing languages increased reading speed.

Dalrymple-Alford (1985) had 18 French-English bilinguals read lists containing 20 English and 20 French words as fast as they could without making errors. One list had the first 20 words in one language and the rest in the other. The second list

alternated languages with each word. Results produced a language switching cost of 17ms, 10 times faster than Macnamara and Kushnir. This result suggests that predictable switches are less costly. Furthermore, removing the requirement for comprehension greatly reduces cognitive load, which may have aided the drastic reductions in switching cost. These studies imply that mixed-language texts should make switches as predictable and consistent as possible.

Grosjean (2001) described *language mode* as the "state of activation of the bilingual's languages and language processing mechanisms, at a given point in time". Bilingual's language mode lies on a continuum between the two monolingual extremes. In monolingual mode they deactivate their other languages, as occurs when conversing with monolinguals. In bilingual mode they use one language as a base with occasional switches into the other. This mode is activated when speaking to other bilinguals or reading mixed-language texts. Bilinguals' position on the continuum depends on the setting, topic and people they are conversing with. Grosjean hypothesised that the language-switching cost is higher in monolingual mode. This study reinforces the notion that predictable language switches are less costly than unpredictable ones.

Cheng and Howard (2008) investigated this hypothesis by having Mandarin-Taiwanese bilinguals read pairs of sentences and judge equivalence of meaning. In each pair, the first sentence was Mandarin, and the second was either completely Mandarin, or contained a switch to Taiwanese midway. Tests conducted in a formal setting with Mandarin instructions resulted in a language-switching cost of 67ms. Those in a more casual setting with mixed-language instructions yielded a statistically insignificant switching cost of 5ms. These results support the argument that context and setting impact switching cost.

In summary, for fluent bilinguals, reading mixed-language texts with regular, predictable switches is cognitively more demanding than reading L1 texts, but is significantly less demanding than reading L2 texts. From a comprehension and speed perspective, these results support the use of mixed-language texts with carefully designed switches as a stepping-stone towards reading entirely L2 text.

Figure 45: Example of a rebus story from
http://www.abcteach.com/directory/basics/rebus

7.4 Mixed-language in L2 learning

The bilingual instructional mode has its origins in antiquity, with Greek-Latin parallel texts used to teach Greek to youth in Roman schools (Kelly, 1969, p. 56). In medieval times, transliterations, and interlinear glosses were used to support L2 reading (Kelly, 1969, p. 129-130,144-146). The use of mixed-language texts for L2 learning appears to be a more recent phenomenon.

7.4.1 Paper-based mixed-language texts

Mixed-language texts are similar to "Rebus stories", which use pictures and symbols to represent words, as shown in Figure 45. They were commonly used in the United States in the 1800s, and are effective for children's reading instruction (S. Jay Samuels, 1988).

To the knowledge of the author, the earliest published mixed-language texts are contained in the English University Press's *Teach Yourself Greek* (Melluish & Smith, 1961). The book includes a series of English passages with progressively increasing insertions of Greek cognates, written in Greek script. The same publisher produced a similar book for Russian (Kelly, 1969, p. 129).

Schaefer (1961) designed a three-phase, mixed-language text program for beginners. Phase one used L1 text with some function words and high-frequency content words replaced by L2 equivalents. Phase two introduced new words and

gradually changes to the L2 grammar structure. Phrase three uses mostly L2 grammar and word order, gradually reducing the proportion of L1 words until learners are reading entirely L2 text.

Schaefer piloted a program that introduced 350 German words in the context of three fiction stories, on three learners with no prior knowledge of German. They read the stories in four and a half, five and six hours respectively. Post-reading tests showed large gains for both receptive and productive vocabulary knowledge.

Burling (1968) designed a program in which learners first read texts with L1 vocabulary and L2 word order. Over time, verb suffixes and function words are introduced, followed by cognates, then non-cognate content words. Interlinear L1 glosses are provided for the first exposure with each L2 word. Burling (1978) evaluated the system in a 14 week introductory French class. Learners read 70,000 words, half of which were French. The students reported increased enjoyment and motivation when using this approach. Curiously, no learning effects were measured.

Likewise, Felix (1981) obtained promising results using mixed-language readers to teach English to five Tamil speakers. Port-Fox (1982) used a mixed-language reader with German speakers learning English. Her students reacted positively to the method, had increased confidence, and improved their reading ability. In the 1970s and 1980s this approach was used by Ray Tounge to efficiently teach Bahasa Indonesia to English speaking missionaries (Ji, 2002).

In the early 1980s Brigham Young University began its “diglot reader” research (di = two, glot = language). Their studies supported earlier findings on the affective (Afarian, 1987; Chen, 1997; Gunderson, 1993; Silver, 1997; Sleight, 1982) and vocabulary learning (Aoyama, 2005; Harmon, 2001; Nelson, 1995) benefits of mixed-language texts. Two of these studies demonstrated that diglot readers were more enjoyable than vocabulary drills, and the vocabulary learning was equally efficient (Chen, 1997; Christensen, Merrill, & Yanchar, 2007). Christensen et al. (2007) recommended future long term evaluation of the diglot method, and comparisons with other L2 vocabulary learning exercises.

Ji (1998, 1999, 2002; 1999) created the “Sandwich Story”, which mixes Chinese and English. The success of the books led to their publication as an ESL learning resource in China.

Figure 46: Kanjilish Firefox plugin replacing the first letter of words with kanji

7.4.2 Computerised mixed-language texts

Weible (1980) implemented German mixed-language texts on the PLATO IV computer-aided instruction system. Due to memory limitations, the program consisted of 16 passages that gradually increased in difficulty, from 31% German words—mostly cognates and function words—in the first passage to 89% in the final one. Learners needed to score at least 92% in a vocabulary test that followed each passage before they were permitted to move on.

In a semester long evaluation, first-year university students took 10.5 hours to complete the course. In a "sight reading" post-test, students that used the PLATO course made 59% fewer errors than those who did the regular course. The collection of mixed-language texts was expanded and used for several years in German classes at the University of Illinois, until the department head cut funding and the project was discontinued (Weible, 2010).

At Brigham Young University, Harmon (2001, 2002) developed a diglot reader desktop application, and Merrill developed a web-based version called "Bilingual Reader" for English speakers learning Spanish.[74]

The *Kanjilish Firefox plugin*[75] inserts Japanese kanji into English web pages. It replaces every occurrence in a customisable list of English words with kanji that

[74] http://open.byu.edu/projects/readers/biReader/lukeBiReader.html

[75] http://kanjilish.mozdev.org/

have equivalent meanings—optionally, kanji can replace the first letter of English words, as shown in Figure 46. Users adapt the system to other languages by preparing L1-L2 word lists; an English-French wordlist is available.

In 2012 Google commissioned the development of *Language Immersion for Chrome*,[76] a web browser extension that that uses Google Translate to turn any web page into a mixed-language text. The proportion of L2 words is controlled by setting the "immersion level". At the most advanced setting, L1 texts are completely translated to L2; at the easiest, 10% of the article is L2. Learners click on L2 words to view the original L1 word and to hear it pronounced. The substitution of words appears to be random.

While these browser extensions are somewhat useful, they lack the sophistication and personalisation required to create mixed-language texts that are truly effective for L2 learning.

7.4.3 K12 patent

The PowerSpeak software[77] teaches L2 grammar, vocabulary and culture through adventure stories, games, and other highly visual activities. The system includes "Diglot Weave" activities, which present L1 stories in audio and textual form. A story is presented several times throughout a lesson, each time with an increasing proportion of L1 words and phrases replaced by L2 equivalents. The following benefits are claimed:

- The target language is presented in a familiar context;
- Vocabulary learning without rote memorisation;
- Reading and listening skills are improved in both the native and target language;
- Learners focus on comprehension.

In 2010, PowerSpeak creators K12 inc. obtained US Patent 7818164 for a software system that creates personalised mixed-language texts (Wood & Dariyanani, 2010). The system gauges vocabulary knowledge via:

- Questions about the content of a mixed-language text;

[76] http://useallfive.com/projects/language-immersion

[77] http://powerspeak.com

- Self-assessment of individual word knowledge;
- Learners' judgements of comfort level with the proportion of L2 words in a text;
- Measuring reading or listening speed of texts;
- Requests for help with L2 words;
- Requests to repeat part of a text.

Learner profiles contains a list of L2 words, idioms, and phrases and their corresponding L1 translations, plus the number of encounters and estimates of the learner's knowledge for each item. Separate entries are used for items with multiple senses.

When creating mixed-language texts the system takes L1 texts and replaces some of the words with L2 equivalents. L2 words that learners are proficient with are always substituted with their L1 equivalents; those they are "less proficient" with are "occasionally" substituted.

7.4.4 Extensive reading studies

Some extensive reading related research inadvertently supports the effectiveness of mixed-language texts for L2 vocabulary learning.

Saragi, Nation and Meister (1978) had 20 adult native English speakers read *A Clockwork Orange*. In this 60,000 word novel, 6% of the words are a Russian-based invented slang called *nadsat*. In order to conceal the study's aim, subjects were told that they would take a post-reading comprehension test. Instead, they took a multiple-choice test on 90 of the 241 nadsat words, 49 of which appear at least 18 times in the text. Subjects averaged 76% on the test, with a range of 50–96%, and scored higher for words that appear more frequently. This result suggests that learners can acquire appreciable amounts of receptive L2 vocabulary knowledge through mixed-language texts.

Beginning with Pitts, White and Krashen (1989), several studies replicated the *Clockwork Orange* experiment with ESL learners (reviewed in Horst et al., 1998). These studies produced more modest acquisition rates of 6–22%. These results are not surprising because the novel was almost certainly too difficult for effective extensive reading by the ESL learners, considering that 6% of the words are Nadsat.

Furthermore, these studies used small portions of the book, e.g. a couple of chapters, which did not provide enough exposures to learn many of the Nadsat words.

Pellicer-Sánchez and Schmitt (2010) used this methodology with a simpler English novel called *Things Fall Apart*, which includes a Nigerian language called Igbo. 20 advanced ESL learners read the 67,000 word book, and then took a test on 34 non-cognate Igbo words. Learners correctly recognised the meaning of 84% and produced the meaning of 55% of the words they encountered at least 10 times in the novel.

Other studies (Pulido, 2007; Waring & Takaki, 2003) test acquisition rates of nonsense words substituted into L2 texts. This method ensures that target vocabulary is unknown to learners prior to reading. Results show that significant proportions of nonsense vocabulary are learned. The acquisition of target words would presumably be equally significant for L1 text sprinkled with L2 words instead of nonsense words, as long as the percentage of unfamiliar L2 words is less than 5%.

7.4.5 Reasons for limited use of mixed-language texts

There are several reasons for the negligible use of mixed-language texts in L2 learning.

In the late 19th century the popular view was that languages were distinct separate entities, and that mixing languages corrupted texts (Schendl, 2002). Mixed-language texts from medieval England were seen as works of art—in the case of literary texts—and the product of incompetent writers—in the case of non-literary texts.

In the 1970s and early 1980s the language teaching profession moved away from the grammar-translation approach towards the notional-functional approach, which structured a syllabus around "notions", real-life situations that are further broken into "functions", specific aims of communication. For example, a notion might be *customer conversing with shop keeper*. Functions would be *asking a price, negotiating* and *refusing to buy*.

That approach ultimately developed into communicative language teaching, the goal of which was communicative competence. The new approach reduced, where feasible, the use of L1 in L2 instruction. However, during the past two decades it

has become accepted that L1 explanations can improve the efficiency of L2 learning (Nation, 2001, pp. 351–352).

The studies reviewed in Section 7.4.4 show that mixed-language texts are effective and enjoyable for learners of any ability. However, creating them is tedious and complex (Silver, 1997). Another reason for limited spread of the concept is the fact that success stories—such as Weible's successes with the PLATO IV, and Ray Graham's positive results in user studies at Brigham Young University—were never published.

This thesis proposes that mixed-language texts are a useful vocabulary learning activity if they are created automatically and are personalised to individual ability and interests. Reading mixed-language news articles, blogs and email is a great way to incorporate L2 learning into one's daily schedule.

7.4.6 Acceptance of mixed-language usage

In the past two decades mixed-language instruction has become increasingly common in L2 classrooms and text books. For example, in New Zealand, Māori and English are the national languages. In primary schools many teachers incorporate Māori vocabulary into classroom instruction. Singing of Māori *waiata* (songs) and use of Māori phrases is common in school assemblies.

Informal mixed-language conversation is common in areas where significant populations of native speakers of two languages live. In these conversations one language often dominates, with a sprinkling of words and phrases in the other language. Heredia and Altarriba (2001) discussed several reasons for language mixing amongst Spanish-English bilinguals in the US:

- To compensate for incomplete knowledge of either the L1 or L2.
- Some L2 words are retrieved with less mental effort than L1 equivalents. For example, people speak Spanish at home and English at school, so words referring to classroom activities are more easily retrieved in English in conversations at home with family
- To be better understood. For example, *cariño* means to like with affection. Bilinguals speaking English achieve a deeper level of understanding by using words like *cariño* to discuss concepts that lack good single word English translations.

- In counselling, clients distance themselves from emotional events by speaking their L2.

Klimpfinger (2007) proposed two additional reasons for language mixing when university students with different L1s converse in English as a lingua franca:[78]

- Specifying an addressee: speakers may switch briefly their L1 in order to address a group member that shares their L1.
- Signalling culture: speakers sometimes switch to L1 when saying place names or other nouns that relate to their heritage.

In one study, parents read three picture books to their bilingual children: Papiamento, Dutch, and pictures only (Muysken, Kook, & Vedder, 1996). With the pictures-only book, parents favoured a Papiamento base with frequent insertion of Dutch nouns and numerals.

Television shows like Dora the Explorer teach Spanish vocabulary through mixed language conversations. These programs help fill the three-hour-per-week quota of educational children's television that is provided by each network in accordance with the 1990 Children's Television Act. New Zealand's Māori TV network produces bilingual programs that include mixed-language dialogues.

The acceptance of mixed-language conversation in classroom, casual settings and television indicates that mixed-language texts could also gain acceptance for L2 learning.

7.5 Mixed-language texts in FERN

The mixed-language texts in FERN are designed as vocabulary exercises for L2 learners of any ability. Texts are predominantly L1 and increase the proportion of L2 words over time. Although projects described in Section 7.4.1 gradually change the word order in addition to increasing the proportion of L2 vocabulary, FERN avoids this additional complexity by always using L1 grammar structure. An additional motivation for this decision was the frequent word order mistakes made by Google Translate.

[78] A common language between speakers whose native languages are different.

Table 17: Order of precedence for replacing words in mixed-language texts, and the associated difficulty scores.

Word Type	**Points added to *articleDifficulty* per word**
Previously encountered words	$5 + wordDifficulty$ (see Section 6.5.1)
Cognates	$10 \times cognateDifficulty$ (see Section 6.6.7)
High-frequency words	100
Low-frequency	100
Off-list words	100

This section describes how FERN automatically creates personalised mixed-language texts, and presents example uses.

7.5.1 Automatically creating mixed-language texts

Once learners choose an L1 text to read, a request is sent to the FERN server, which uses its automatic glossing procedure (see section 5.3) gloss the L1 text with L2 translations. Glosses that have a one-to-many or many-to-many alignment are ignored, in order to reduce the chances of creating ungrammatical mixed-language sentences. FERN prepares mixed-language articles by wrapping words with one-to-one gloss alignments in HTML tags that contain the L2 translation, frequency and cognate information.

The learner sets the *learnerDifficulty* of mixed-language articles on a scale of 0–10; 5 is the default. At a difficulty of 0, the article is entirely L1. As the difficulty setting increases, the proportion and difficulty of L2 words increases. At a difficulty of 10 every glossed word (i.e. those that had a one-to-one alignment in Google Translate) is displayed in L2. The difficulty setting is multiplied by the number of glosses available in the article to get a *maximumDifficulty* score for the article. For example, if an article has 100 glossed words and the *learnerDifficulty* is 5, then the *maximumDifficulty* will be 500.

Once the article is received, the client determines which words are replaced by L2 equivalents. Each word that is replaced adds points to an *articleDifficulty* score. FERN replaces additional words until the *maximumDifficulty* is exceeded. Replacements occur using the precedence displayed in Table 17.

FERN first substitutes previously encountered words, which are grouped based on *wordDifficulty* (see Section 6.5.1) into ten bands (scores 0–10, 10–20, 20–30 etc.). The easiest words are added first and their difficulties are added to the cumulative

articleDifficulty. An extra 5 points are added per word to account for the inherent difficulty of mixing L2 words in an L1 text. If the total difficulty of the words in the first band is less than the *maximumDifficulty*, words from subsequent bands are added until the limit is exceeded.

If the total difficulty of all previously encountered words is less than the limit, FERN proceeds to replace cognates. Cognates are grouped into five difficulty bands (see Section 6.6.7), which are replaced sequentially from easiest to most difficult so long as the *articleDifficulty* does not exceed *maximumDifficulty*. Each cognate's contribution to *articleDifficulty* is ten times its difficulty score.

The remaining words are divided into six frequency bands, from high to low frequency. The sixth and final band is off-list words, which are mostly proper nouns. These are also added sequentially one band at a time, until either the *maximumDifficulty* is exceeded or all words are L2.

Learners may select seven parts of speech to translate: nouns, adverbs, adjectives, verbs, proper nouns, numbers and other. Only those types selected by the user are included in the replacement process. By default, verbs are turned off because they seem to cause the most disruption when reading mixed-language texts.

The following sections present example usages of English-Spanish mixed-languages texts using excerpts from a 371 word Simple Wikipedia article entitled "Olympic Games".[79]

7.5.2 Cognates

Novices can become accustomed to spelling differences between their L1 and L2 by reading mixed-texts with cognates replaced. 27% of the words in the Olympic Games article are proper nouns, numbers, or cognates. The following excerpt replaces proper nouns and cognates with Spanish equivalents:

> In the opening ceremonia, the estadio is opened. Eventos especiales happen like songs, performances, dances, etc. The Olímpicos theme song is played, written by John Williams. The antorcha is brought in after being carried around the world, from Olimpia, where the original Olímpicos were held. Doves are released. Then, every atletas from each

[79] http://simple.wikipedia.org/w/Olympic_Games as at January, 2010

> country comes into the estadio in orden alfabético, excepción for Grecia, which comes first in as they crean the games, and the country where the games are held, which goes último.

Despite a few slightly-challenging cognates (estadio-stadium, especiales-special, crean-created), this text is easy to read even for someone who has never studied Spanish.

7.5.3 High-frequency words

Mixed-language texts are useful tools for beginners to acquire high-frequency vocabulary. The following example replaces the L1 words that have equivalents in the top 900 Spanish words[80]. Even though two thirds of the text is Spanish, this text is easily readable by a beginner learner of Spanish.

> Los Olympic Juegos son un sporting event que tienen lugar en una diferente ciudad cada cuatro años. It includes muchas de las mejores athletes de todo el mundo. El primero modern Games were held en Athens en 1896. El Motto de los Olympic Games es “Citius, Altius, Fortius” que significa, en English, "Más rápido, más alto, más fuerte.

In practice, FERN’s algorithm typically results in one third of the words being substituted at the maximum difficulty setting.

7.5.4 Other uses

Mixed-language texts can introduce learners to specialised vocabulary. For example, can learn basketball vocabulary by reading mixed-language basketball match reports as a stepping stone towards reading L2 match reports.

Mixed-language texts enable L2 vocabulary learning while reading stories not available in the L2. For example, cricket fans can learn Spanish while reading about cricket articles, even though cricket is not played in any Spanish speaking countries.

Finally, these texts allow learners to practice L2 vocabulary when they are not in the mood for the more strenuous task of reading L2 texts.

---

[80] http://en.wiktionary.org/wiki/Wiktionary:Frequency_lists

8 Evaluation

This thesis describes the design and implementation of FERN—a prototype system that aids L2 extensive reading of authentic texts on the internet. This chapter evaluates FERN's suitability in a two-semester long user study in an L2 classroom. Analysis of the results provides strong evidence that low-ability learners can make large gains through narrow reading of automatically glossed texts, located with an individualised web search, and supplemented with personalised vocabulary review.

The first half of this chapter describes the user study procedure and how gamification was incorporated to improve learner motivation. The remainder of the chapter analyses FERN's usage data collected during both semesters, and the results of a student survey conducted at the end of semester one. This chapter examines the suitability of FERN's individualisation in helping learners choose suitable texts and perform efficient vocabulary review. Finally, the chapter examines the accuracy of our automatic glossing approach, the impact of disambiguation errors, and the magnitude of the benefit of glossing over dictionary lookup.

8.1 Setup of classroom user study

In order to evaluate its suitability with real learners, FERN was used for two semesters with a Spanish class that was in their fourth semester of language study at the University of Waikato.

8.1.1 Procedure in semester one

During an orientation session, students completed an online questionnaire, viewed a live demonstration of FERN, and spent 15 minutes testing its exercises. In the subsequent eight weekly 50 minute lab sessions, students used FERN for 25 minutes, and completed instructor led language exercises for the other 25 minutes. The lab sessions were conducted during one of three regular class lecture times. Students were also encouraged to read at home.

Grades were not awarded based on the quantity of reading performed; however, 10% of students' course grades were based on attendance and participation in labs and tutorials. Inside FERN, students scored points based on the quantity of words read and reviewed, and top students were ranked on a class leaderboard, with small

Table 18: Reading habits of students with paper and electronic texts

Text type	Read more non-fiction	Read more fiction	Similar amounts
Electronic	10	1	7
Paper	5	3	10

Table 19: Usage of online resources by student

Online resource	Frequently	Occasionally	Rarely or Never
Dictionary	14	3	1
Automatic translator	8	8	2
Grammar reference	9	5	4
Audio lessons	3	4	12
Other software	0	5	13

prizes awarded to those that completed the most reading. The semester concluded with a questionnaire about learning with FERN.

8.1.2 Subjects

The Spanish class contained 16 self-identified intermediates and two beginners. All students were fluent English speakers, and four were fluent in an additional language. Eight students had taken formal Spanish classes prior to university; seven had visited Spanish-speaking countries, but only one had done so for more than three weeks. Students reported using Spanish occasionally outside of class, with reading and listening more common than speaking and writing.

Regarding their reading habits in any language, ten students read more paper-based texts than electronic ones, two read more electronic texts, and six read similar amounts of both. As shown in Table 18, for electronic texts most students read more non-fiction than fiction texts; for paper-based texts, it was quite even.

Table 19 shows which electronic resources students used in their L2 study. Most students frequently used L2 dictionaries, eight frequently used automatic translation software, nine frequently used grammar references, and two frequently used audio lessons.

Students were asked what prevented them from reading more L2 text outside the classroom. Their responses echoed the message of chapter two—lack of time and limited vocabulary knowledge make it difficult to find interesting texts at the right level of difficulty.

Finally, student responses regarding their motivation for learning Spanish indicated a high degree of intrinsic motivation. The most commonly stated reasons for taking Spanish were a love of the language and love of language learning in general, a desire to travel to Spanish speaking countries and the future employment opportunities it may bring.

8.1.3 Procedure in semester two

Semester two included ten weekly 50 minute sessions, in which students used FERN for 25 minutes and did free writing on a topic assigned by the teacher for 25 minutes. The semester included a two week mid-semester teaching recess. Eleven of the twelve students had participated in the first semester course that used FERN.

In this semester, 10% of each student's grade was based on reading with FERN. They were awarded 1% for each week they read at least 2000 words and completed at least one vocabulary review exercise and one fill-the-blanks exercise. Students were required to read at least two articles that related to a teacher-assigned weekly theme such as Cuba, Machu Pichu and Panama Canal. Students received the full 10% if they achieved this target in 10 of the 12 weeks of the study.

8.1.4 Use of gamification

Gamification is the use of game design elements such as points, leaderboards and badges, in non-gaming contexts (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). In an educational context, effective gamification increases motivation and engagement (Kapp, 2012), which in turn increases learning (Hamari, Koivisto, & Sarsa, 2014). It can have negative effects when used excessively or maliciously (Yamakami, 2013).

FERN gives learners points for every L2 word they read and review, and displays their total points on a class leaderboard. In semester one the leaderboard displayed only the top three students, because the teacher was concerned that displaying too many would negatively impact some students. In semester two the teacher agreed to expand the leaderboard to the top five students, as shown in Figure 47. The score

Figure 47: FERN home page for the 14 week University of Waikato Spanish class

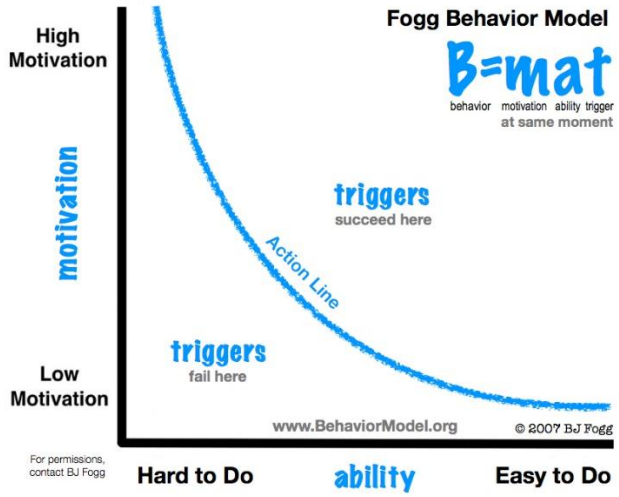


Figure 48: The Fogg Behaviour Model (Fogg, 2007)

of the currently signed in student was displayed in fifth position on the list, if they were not already in the top four.

The second semester added additional gamification elements. Students were given the target of reading 2000 words per week, with progress towards that goal displayed on their home page. Badges were awarded for completing the weekly target, along with 1% of the course grade.

Fogg's (2009) model for persuasive design asserts that "for a target behaviour to happen a person must have sufficient motivation, sufficient ability, and an effective trigger." Each factor must be present at the same instant for the behaviour to occur. In Figure 48, Fogg uses the concept of "activation threshold" (depicted by the "action line" in Figure 48) to describe the combinations of motivation and ability that are required for triggers to result in action.

FERN's glosses, dictionary lookup and news search increase the *ability* of learners to do extensive reading with authentic texts. The rewards, leaderboards and short-term goals increase extrinsic *motivation*; tri-weekly class attendance, and weekly reading deadlines *trigger* learners to take action. FERN's effectiveness in achieving these outcomes is discussed in the remainder of this chapter.

8.1.5 Data collection

In addition to the reading logs and progress charts that are visible to learners, FERN records extra usage data for research analysis. The news search logs search queries and data about articles that appear in search results. FERN's reading interface logs gloss and dictionary access, and reading times for each paragraph. Review exercises log completion times and the number of word reviewed correctly and incorrectly.

8.2 Results

This section discusses usage data and responses to the final survey in semester one.

8.2.1 Student opinions of FERN in semester one

In the final survey, students marked agreement with a series of statements on a scale of 1 (disagree) to 5 (agree), beginning with a series of general statements presented in Table 20. All students agreed that FERN was useful for learning, all except one enjoyed the learning, and all students felt that the difficulty level was good, except for one who commented that the fill-the-blanks and L2 news articles were often too difficult, and mixed-language texts were too easy. Twelve of the sixteen students were keen to continue using FERN in their spare time, and all students agreed that it would be a useful tool in future Spanish classes. Eight preferred it as a homework activity, three favoured classroom use, and four strongly agreed that it should be used for both. These survey results influenced the teacher's decision to make FERN a compulsory homework and in-class activity in semester two.

Table 20: Student's opinions about learning Spanish with FERN

Statement	**5**	**4**	**3**	**2**	**1**
FERN is useful for learning Spanish	12	4	0	0	0
I enjoy using FERN	7	8	1	0	0
The level of difficulty is appropriate for me	10	5	0	1	0
I would like to continue using FERN in my spare time	9	3	4	0	0

Table 21: Number of words encountered in mixed-language texts (MLT), vocabulary match (VM), fill-the-blanks (FTB) and multi-choice (MC).

User ID	**Semester One**				**Semester Two**				
	Reading	**MLT**	**VM**	**FTB**	**Reading**	**MLT**	**VM**	**FTB**	**MC**
6	6521	0	930	0	25482	0	110	41	0
7	3820	0	0	0					
8	4515	0	190	72	11167	0	40	0	8
9	10875	0	210	0					
10	2543	0	0	9	5530	71	260	0	0
11	28139	483	80	6	39322	0	130	108	32
12	5530	0	20	0					
13	15441	483	0	0	24722	0	50	0	0
14	1135	74	620	62	26900	0	420	12	0
15	4341	0	170	0					
16	6959	0	90	7	30543	0	200	183	56
17	939	0	20	15					
18	5357	95	0	0	3224	0	30	0	0
19	2009	58	0	0					
22	8063	0	430	55	38195	73	620	215	24
23	9103	0	0	0	10979	0	20	0	0
24	1789	523	80	97	24127	0	100	0	16
25	4797	30	90	2					
26					44670	0	110	111	16
Total	121876	1263	2930	325	284861	144	2090	670	152
Average	6771				23738				

The final survey included four open-ended questions:

1. What did you *enjoy* most about FERN?
2. What did you find *difficult* about FERN?
3. Why did you *not* use FERN more often?
4. What would make FERN more *enjoyable* and *motivating* to use?

Despite agreeing on its usefulness, enjoying using it, and claiming strong intrinsic motivation for learning Spanish, most students made limited use of FERN outside the classroom. In open-ended question number three, half of the students stated that a lack of time and discipline hindered them, and seven mentioned technical challenges such as browser incompatibility, forgetting the URL, usernames and passwords.

Students' suggestions to improve FERN were clearly grouped by gender. Three female students recommended improving the website's visual design, and five females requested the inclusion of fictional texts. Male students requested additional learning activities and gamification elements. These results are supported by reports that females read more fiction than males[81], and that females are more visually oriented (Holbrook, 1986) in their perceptions of websites design (Cyr & Bonanni, 2005; Simon, 2000).

8.2.2 Changes in weekly reading

As shown in

Table 21 the number of words read per student increased from 6,800 in semester one to 23,700 in semester two.

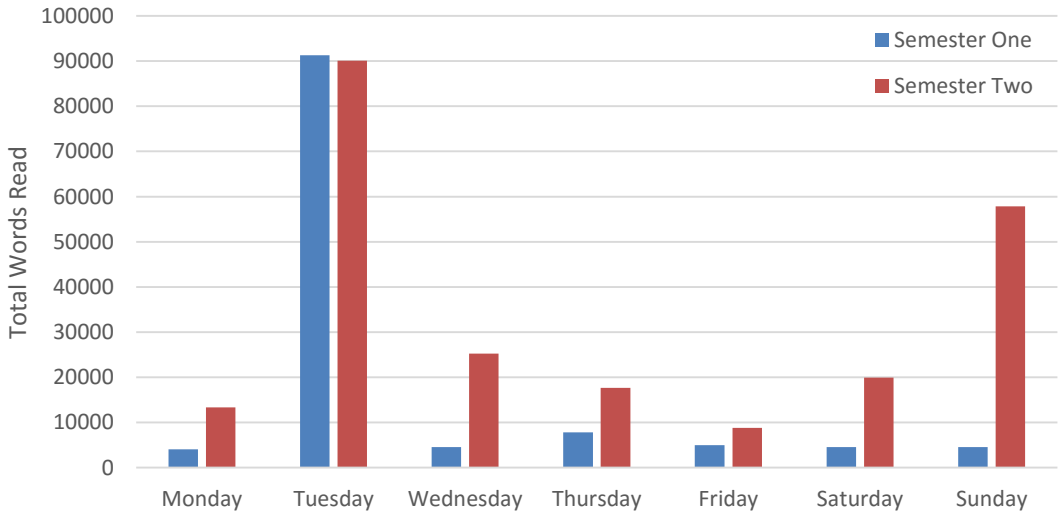


Figure 49: Number of words read on each day of the week during the semester

[81] A Decade of Arts Engagement: findings from the survey of public participation in the arts, 2002–2012. http://arts.gov/sites/default/files/2012-sppa-jan2015-rev.pdf

Table 22 shows that nine out of 11 students significantly increased their reading, with four students increasing by at least 1000 words per week. The reduced reading resulted from a serious illness to student 18, and work and family commitments making it difficult for student 23 to attend class.

What caused this dramatic increase in reading? In semester one FERN usage contributed to 10% of student's grade allocated to "tutorial attendance and participation". Despite enjoying learning with FERN, students did little reading outside of class because no grade was allocated to doing so. In semester two, 10% of course grade to was allocated to 2000 words of weekly reading. Figure 49 shows that a similar quantity of reading was performed during the Tuesday lab session, but in semester two significantly more reading was performed at home. The biggest difference occurred on Sunday, which accounted for 3.3% of reading in semester one, compared with 24% in semester two, much of which occurred in the four hours before the midnight deadline for the weekly reading assignment.

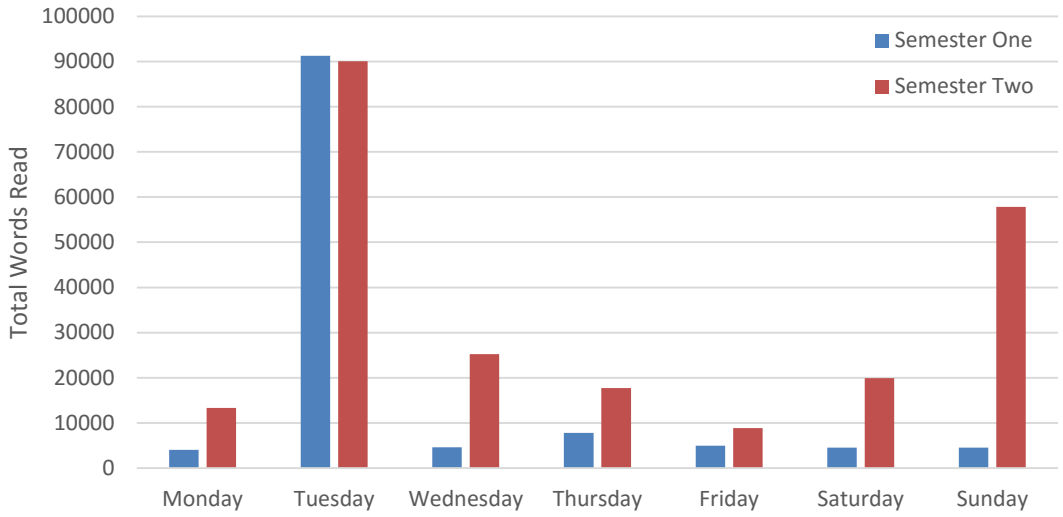


Figure 49: Number of words read on each day of the week during the semester

Table 22: Change in words read per week between semesters one and two

ID	Semester 1	Semester 2	Change	% change
6	593	1960	1367	231%
8	410	859	449	109%
10	231	425	194	84%
11	2558	3025	467	18%
13	1404	1902	498	35%
14	103	2069	1966	1905%
16	633	2349	1717	271%
18	487	248	-239	-49%
22	733	2938	2205	301%
23	828	845	17	2%
24	163	1856	1693	1041%

Table 23: Opinions about FERN's news search

Statement	5	4	3	2	1	Ave
I am able to find articles that I enjoy reading	7	6	3	0	0	4.3
Word counts help me choose articles	9	3	1	1	2	4.0
Difficulty scores help me choose articles	7	3	4	1	1	3.9
Difficulty of articles I read was appropriate for me	9	5	1	1	0	4.4

Table 24: Method for choosing articles

Method	Semester 1	Semester 2
Search Query	134	608
Browsing Topics	101	134

Table 25: Popular topics when browsing for articles

Value	Topic	Semester 1	Semester 2	Total
h	Headlines	11	10	21
w	World	14	10	24
b	Business	2	2	4
n	Nation	2	7	9
t	Science and technology	17	26	43
el	Elections	2	0	2
p	Politics	0	0	0
e	Entertainment	29	54	83
s	Sports	15	12	27
m	Health	9	13	22

The weekly reading target effectively motivated most students, with 8 out of 12 averaging at least 2000 words per week. However, since one goal of classroom extensive reading is to help students develop the habit of L2 reading, and habit formation requires regular repetition over an extended period of time (Lally, van Jaarsveld, Potts, & Wardle, 2010), a more effective reading assignment would reward students for reading at least five days per week.

The 11 students that participated in both semesters did similar amounts of "voluntary" reading in each semester. In semester one, students read an average of 300 words per week outside of class. In semester two, students exceeded the weekly reading requirement by an average of 300 words per week.

8.2.3 Searching and browsing for articles to read

The semester one survey results in

Table 23 show that 13 out of 16 students were able to find enjoyable articles; the other students recommended that FERN broaden its content to include short stories and cultural articles.

Table 24 shows, that in order to locate articles, students searched 33% more than browsing in semester one, and did 450% more searching in semester two. The increase in searching resulted from semester two's requirement to search for articles that related to teacher selected subjects, such as Christopher Columbus. As shown in Table 25, browsing the entertainment topic was twice as popular as the next most popular, science and technology, which was followed by sports, world, health and headlines.

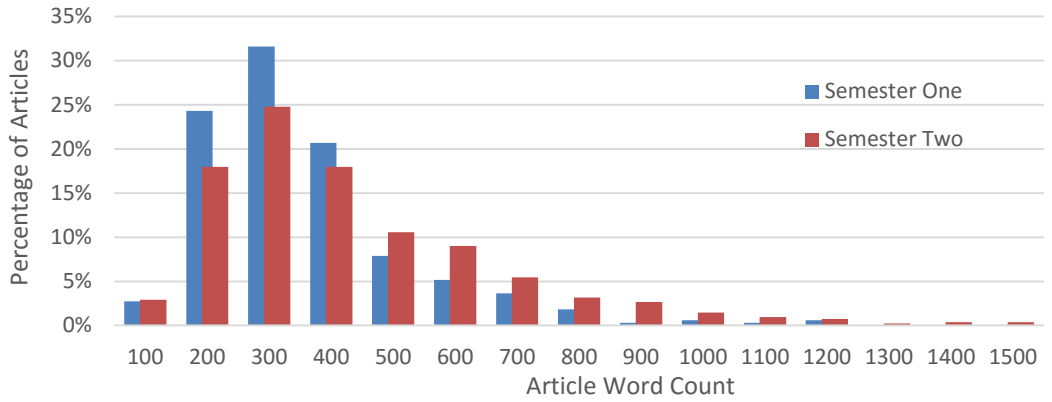


Figure 50: Change in article selection habits between semesters

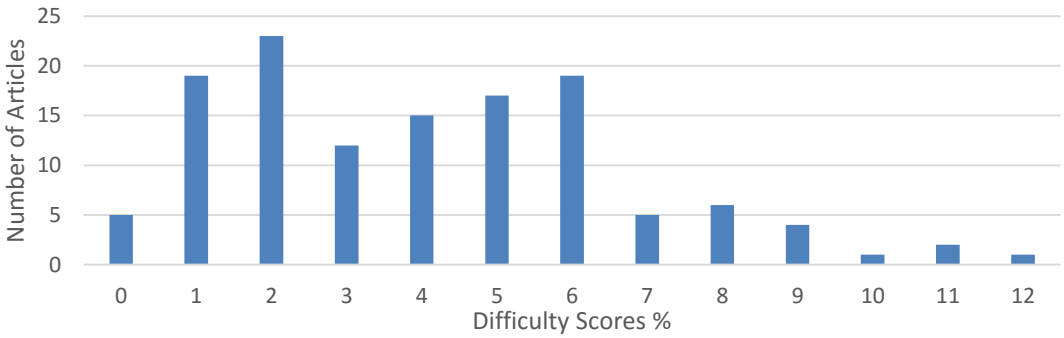


Figure 51: Difficulty scores for articles read in the final two weeks of semester two

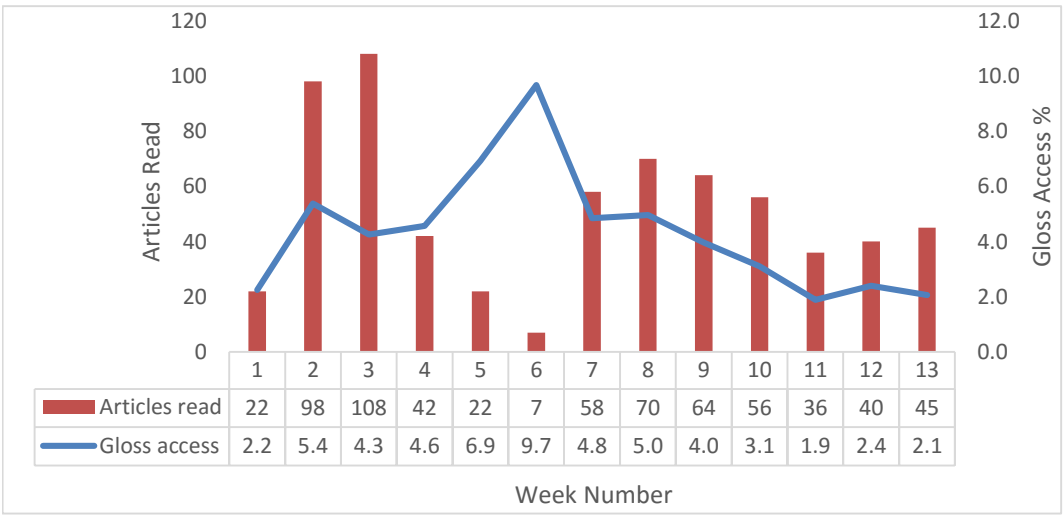


Figure 52: Articles read and average weekly gloss usage in semester two

Table 26: Articles logged per semester

	Semester 1	Semester 2	Total
Articles logged	432	674	1106
Read to completion	85%	77%	80%
Read at least half	95%	93%	94%
Gloss usage	4.8%	4.1%	

The survey results showed that 75% of students agreed that word counts helped them select articles to read, and 60% found difficulty scores helpful. FERN usage data in Figure 50 indicates the strong influence of using green text colour for displaying counts between 100–400 words. The green zone includes 77% of articles read in semester one, and 61% in semester two, with a sharp drop outside of that range. Usage data in Figure 51 indicates that using green text for colouring difficulty scores below 6% also influenced selections.

The average article length increased from 310 words in semester one to 390 in semester two. The main cause of this increase appears to be the requirement to read articles from culturally related topics for which shorter articles were less readily available. Increased reading ability may also have made students more comfortable with choosing longer articles.

All except two students surveyed agreed that article difficulty levels were suitable, as supported by high reading completion rates. During the introductory session, students were instructed to stop reading articles they found uninteresting or too difficult. Articles are excluded from reading logs if students do not read at least one paragraph, which occurred with 4% of the articles they started reading. As shown in Table 26, students read at least one paragraph in a total of 1100 articles, of which they read at least half of 94% of them, and read 80% to completion.

8.2.4 Gloss usage

Gloss usage gives further indication of the suitability of article difficulty. Students accessed glosses for 4.8% of words in semester one and 4.1% in semester two. Figure 52 shows articles read and gloss usage during each week of semester two. Students logged the most articles during weeks two and three, and the least articles during the teaching recess in weeks five and six. The average gloss usage decreased from 5% in weeks two to four, to 2% in the final three weeks. Determining the cause of this drop requires closer examination of individual students.

Table 27 shows that gloss usage during extensive reading varied greatly between students. For example, in semester one, student 11 claimed to focus on reading quickly rather than achieving full comprehension and accessed glosses for only 0.1% of words. Consequently, this student found the reading and fill-the-blank exercises too difficult. This student doubled gloss usage in semester two, but still

accessed them half as much as any other student. This situation illustrates the importance of orienting students to perform extensive reading effectively—in this case to increase gloss usage, and reduce reading speed until comprehension is comfortable.

Student 16 greatly increased gloss usage in semester two after commenting that some FERN exercises were too difficult in semester one. The requirement to complete fill-the-blanks exercises, which are difficult without good comprehension of articles read, likely contributed to increased gloss use in order to improve comprehension.

The small reductions in gloss usage by five students, and large reductions by the other three were likely impacted by improved reading ability. Student 14 made the largest reduction of 71%, after making large gains in Spanish ability from practicing with her native Spanish speaking spouse during the three month summer break.

In the final two weeks of semester two, difficulty scores were added to the reading logs in order to investigate the correlation between gloss usage and difficulty score. A strong correlation would indicate the effectiveness of the difficulty score algorithm. Figure 53 shows the correlation obtained by combining the data for all articles read by all students during the final two weeks, and yields a coefficient of 0.42. While this looks promising, plots for individual students yield a maximum correlation of 0.18. It appeared that stronger students read more, chose articles with lower difficulty scores and were less reliant on glosses, while weaker students read less, had higher difficulty scores, and used glosses more.

One would also expect students' reading speeds to correlate with difficulty scores. In this case the aggregated data and individual student plots had weak correlations with the maximum coefficient for an individual student of 0.24.

Drawing firmer conclusions on the relationship between gloss usage and article difficulty, requires controlling for additional factors such as article length, learner ability, and interest in the article. Further analysis of the factors affecting gloss usage and reading speed, and the effectiveness of the difficulty score are reserved for future work.

Table 27: Change in gloss usage between semesters

ID	Semester 1	Semester 2	% Change
6	5.8%	8.0%	38.3%
8	6.8%	6.2%	-8.6%
10	3.2%	3.2%	-1.2%
11	0.1%	0.2%	53.5%
13	10.4%	9.4%	-9.5%
14	26.9%	7.8%	-71.0%
16	1.3%	3.1%	129.7%
18	4.6%	3.9%	-16.2%
22	0.4%	0.4%	-13.3%
23	0.9%	0.4%	-49.4%
24	6.8%	2.8%	-58.0%

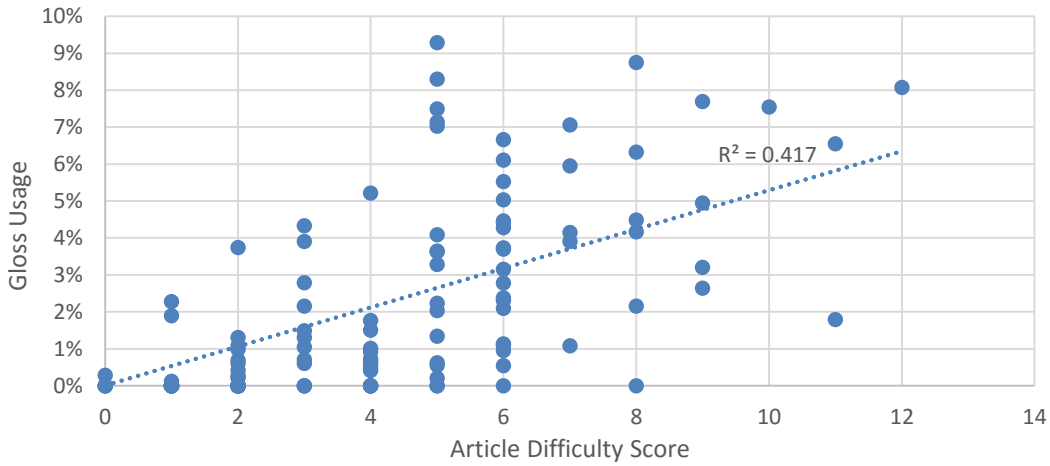


Figure 53: Correlation between gloss usage and difficulty score in semester two

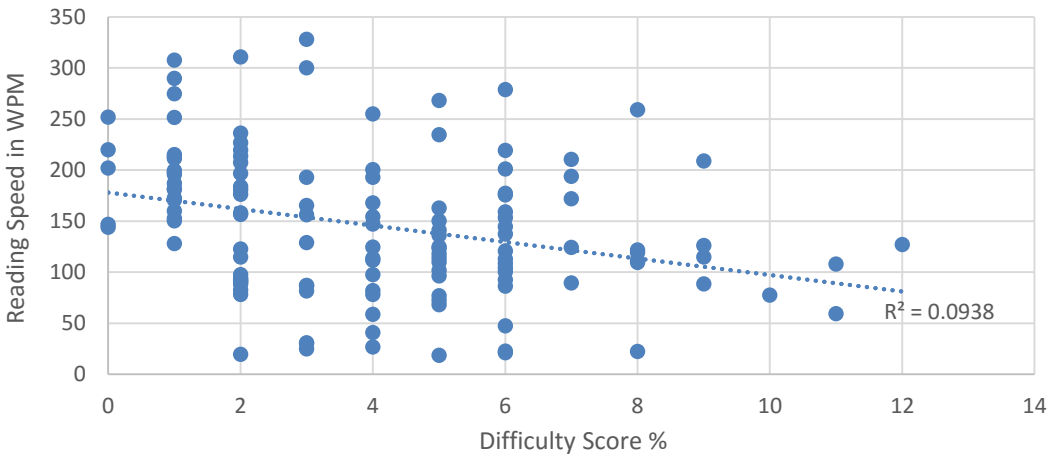


Figure 54: Correlation between reading speed and difficulty score in semester two

Table 28: Paragraph Reading Speeds in Semester Two

ID	Average Speed	Standard Deviation
6	217	116
8	105	60
10	346	232
11	176	64
13	104	40
14	76	26
16	321	261
18	58	24
22	84	22
23	176	64
24	118	56
26	139	42

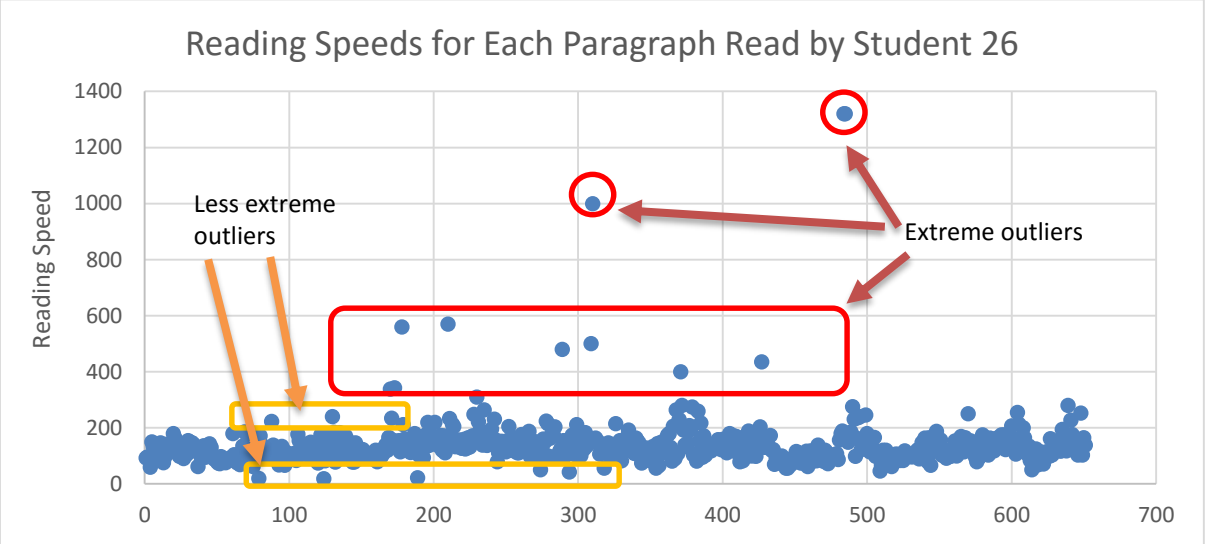


Figure 55: Paragraph reading speeds and outliers

8.2.5 Measuring reading speed and quantity

As described in Chapter 4, FERN measures how much and how fast students read by tracking the time learners focus on each paragraph in a text—excluding paragraphs with reading speeds exceeding 400 words per minute, which are assumed to have been skimmed or skipped. The fixed threshold of 400 was estimated to be twice the reading speed of the fastest readers in the class involved in this study. This section analyses reading time data and proposes a dynamic threshold approach that better adapts to learners of varying ability.

Table 28 shows the average paragraph reading speed and standard deviation for each student including outliers. Standard deviations ranged from 22–261, with three of 12 students exceeding 64. Figure 55 illustrates that for student 26, the majority of data points are concentrated in the "normal reading speed" range of 50–200 words per minute. Other student's profiles have similar characteristics, but with varying means and variance.

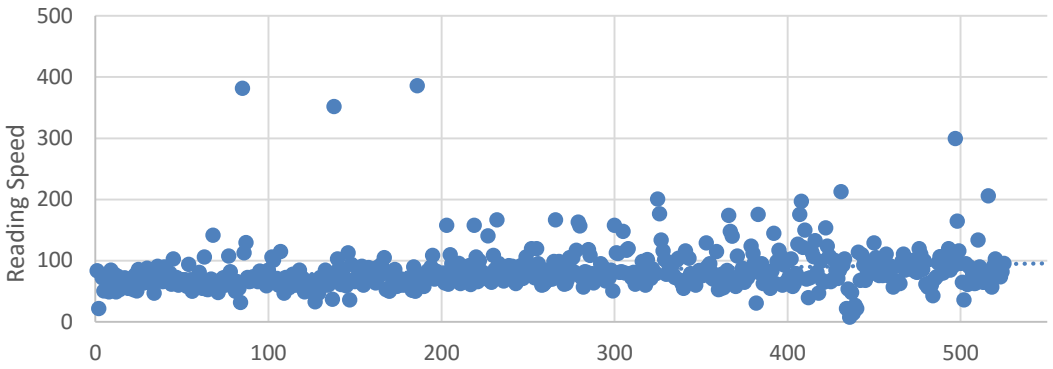
Data points that lie well outside the normal reading range usually result from students skipping the final paragraph in an article because it contains the author's biography, or because they stopped reading the paragraph part way and moved onto a new article. Less extreme outliers generally result from students skimming or skipping part of a paragraph. Paragraphs with unusually low speeds result from students taking breaks during reading, or from intensive studying of difficult phrases. These paragraphs should be logged to the learner's vocabulary as normal, but excluded from speed calculations.

The "extreme" outliers in Figure 55 exceed ten standard deviations from the median of the previous ten points, while less-extreme outliers are five to ten deviations from the median. Excluding paragraphs that are five standard deviations above the mean makes it difficult to cheat the system, and avoids logging paragraphs that learners skimmed too quickly to understand the meaning and have effective encounters with the vocabulary.

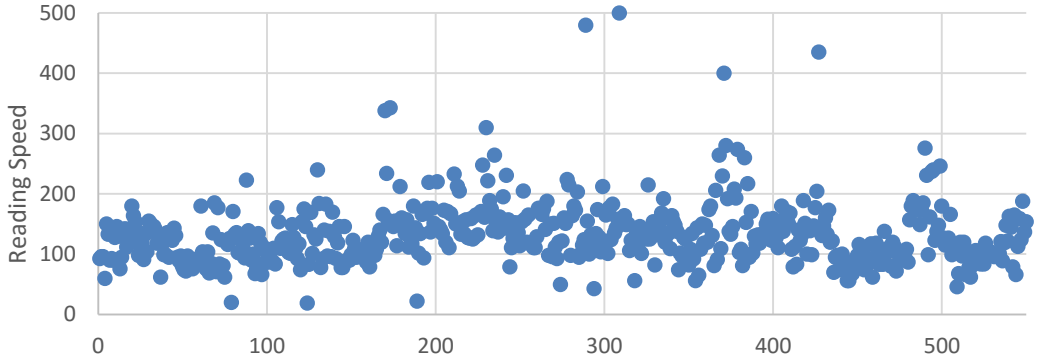
Figure 56 illustrates the varied reading behaviour of students 16, 22 and 26. Students 16 and 22 read 550 paragraphs each and had the highest and lowest standard deviations respectively, while student 26 read 650 paragraphs and was close to the median standard deviation. To make comparison easier, the axis scales are adjusted to be the same in all three graphs.

Student 22 had the most consistent speeds, especially in the first half of semester two. Although this student had no paragraphs exceeding the 400 words per minute threshold, there are clearly several outliers, such as three paragraphs in the first 200 with speeds over 350, which are more than 10 standard deviations from the median.

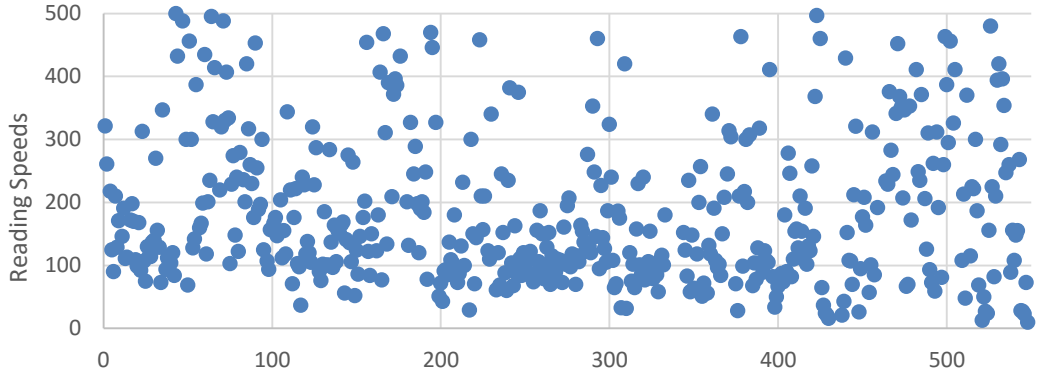
The plot for student 16 has similar characteristics to the others, except with a standard deviation that is 12 times larger than student 22. Further investigation is needed to explain the seemingly erratic reading speeds of students 10, 16 and 6.



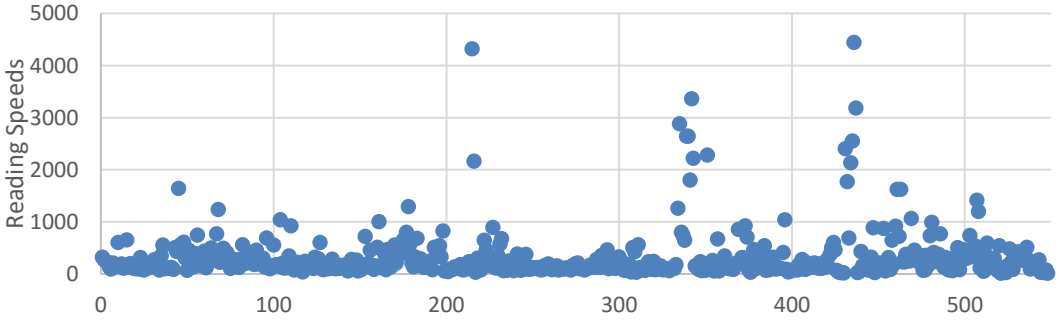
a) Paragraph reading speeds for student 22



b) Paragraph reading speeds for student 26



c) Paragraph reading speeds for student 16 (truncated at 500 wpm)



d) All paragraph reading speeds for student 26

Figure 56: Paragraph reading speeds for students with varying reading behaviour

Table 29: Attitudes of students towards progress reports

Statement	5	4	3	2	1	Ave.
The results give a good indication of my improvement	5	7	3	0	0	4.1
The class and individual word lists are useful	5	4	7	0	0	3.9
The charts and reading log tables are useful	3	5	4	2	0	3.6
The points and class rankings are useful	3	7	4	1	0	3.8

Future versions of FERN should use the following individualized algorithm to detect outliers:

1. Establish the initial reading speed during a period of focused reading, in which learners keep reading until they have read seven to ten paragraphs without recording speeds more than three standard deviations from the median.
2. As the learner reads, track the median and standard deviation from the previous ten paragraphs and exclude those with speeds that exceed five standard deviations above the median from both reading speed calculations and the vocabulary profile.
3. Exclude paragraphs that are more than five standard deviations below the median from speed calculations, but do include them in the vocabulary profile.

8.2.6 Favourability ratings for progress reports

In semester one, FERN progress reports included word count and reading speed charts, reading logs, individual and class word lists, and class leaderboards. In the final survey, 13 of 16 students agreed that the reports gave a good indication of improvement, and all 16 found at least one of the progress reports useful. A summary of responses is shown in Table 29. The teacher of the class was excited by the class word list, which aggregated individual vocabulary profiles. It enabled her to determine which words and tenses caused the class the most trouble.

The score for the progress charts was impacted by some users doing too little reading for the charts to give meaningful feedback. However, by the end of semester two the charts for reading speed showed impressive progress, as depicted by the tripled reading speeds in Figure 57, where red dots plot article speeds and the green line is a moving mean.

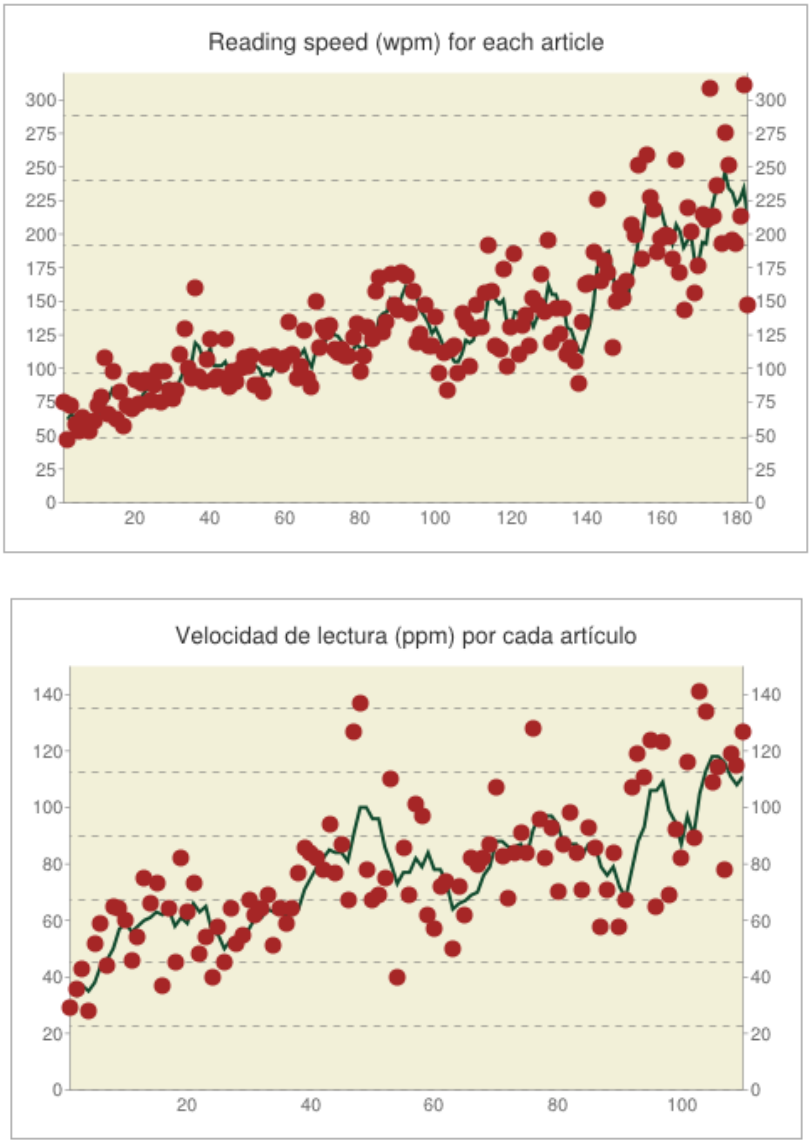


Figure 57: FERN reading speed charts showing improvement in reading speeds of two students over two semesters

Table 30: Students' level of agreement with statements about exercises they used

Statement	L2 News	Mixed News	Word Match	Fill Blanks
1. Useful for L2 learning	4.8	3.9	4.6	4.1
2. Enjoyable	4.5	3.4	4.2	3.3
3. Appropriate difficulty	4.4	3.9	4.4	3.7
4. Want to continue using	4.9	3.4	4.1	3.7

8.2.7 Favourability ratings for exercises

Table 30 shows survey results for questions about how useful, enjoyable and difficult each exercise was, and whether students wanted to continue using that exercise. L2 reading had the most favourable ratings and scored 4.9 out of 5 for "would like to continue using". Students who gave enjoyment and difficulty scores below five were those who found some articles too difficult, and wanted access to a wider variety of texts, including fiction.

Word match was the highest rated supplementary activity, but two students did not want to continue using this activity because they felt that reading L2 texts was more effective study. Several students found the fill-the-blanks exercise too difficult, consequently it had lowest enjoyment and difficulty scores. The students found mixed-language texts to be too easy for a classroom learning activity, but some commented that they would be useful for lower-ability learners. Mixed-language texts are only efficient for learning by intermediates when they are combining L1 leisure reading with L2 vocabulary learning. In response to this feedback, activity difficulty levels were refined.

In semester two, students were required to complete at least one fill-the-blank and word match exercise each week, and mixed-language texts were optional. As a result, the use of mixed-language texts decreased in semester two, but use of fill-the-blanks doubled, and the number of words match usage rose from 8.5–12.8 words per week (excluding student 6 who accounted for a third of the completed word matches in semester one).

At the end of semester two, a bug was discovered in the fill-the-blanks exercise that caused FERN to only log completed exercises, when it should have also logged partially complete exercises. During semester two there were also bugs that resulted in poor quality word match and mixed-language-texts, which probably reduced usage of those exercises.

8.2.8 Summary

FERN was used for two semesters in a university Spanish class with 19 lower-intermediate level students who self-reported a high degree of motivation to learn the language, but claimed that lack of time and vocabulary knowledge prevented them from doing more L2 reading. Students gave enthusiastic feedback at the

conclusion of semester one, but only two out of 18 students did a substantial amount of reading outside of class.

In order to increase the motivation of students in semester two, FERN's homepage incorporated an expanded leaderboard, plus the additional gamification elements of badges for completing a weekly reading assignment of 2000 words plus review exercises. As a result students did 3.5 times more reading in semester two. While, comments from students during class indicated that the badges, points, and leaderboards did impact some students, the most significant motivator was awarding courses credit for the weekly reading target.

Student feedback in the survey showed clear gender differences, with females requesting design and colour related features, and the inclusion of fiction stories; males asked for more games and supplementary exercises. It is clear that a variety of exercises, reading material, gamification elements, and a clearly defined weekly target are required to maximise the motivation of all students in the class.

8.3 Analysis of FERN's approach to narrow reading

The evaluation with the university class demonstrated FERN's suitability with lower-intermediates. This section provides theoretical support for the claim that beginners can do effective extensive reading of authentic texts after learning the top 200 word families, along with some specialised vocabulary and cognates. Key to FERN's approach is the application of narrow reading and focus on specialised vocabulary and cognates. This section begins with discussion on the completeness of FERN's Spanish-English cognate list.

8.3.1 Cognate list and dictionary completeness

FERN's automatically extracted list of 12,220 English-Spanish cognates is incomplete, principally because it is sourced from Wiktionary—a growing but incomplete source of translations. As shown in Table 31, FERN's cognate list covers half of the word families in the British National Corpus (BNC) and Corpus of Contemporary American English (COCA) word family lists. The cognate percentage diminishes to 4% in list number 24, as a result of Wiktionary's poor coverage of low-frequency words.

The true percentage of cognates was estimated by manual judgement by the thesis author of random samples of 100 words in the fifth, tenth, fifteenth and twentieth

Table 31: Coverage of BNC/COCA word lists by FERN's cognate list

List Number	Tokens	Cognate Families	Complete Estimate
1	1038	418	
2	1306	561	
3	1468	737	
4	899	531	
5	695	463	610
6	636	439	
7	534	403	
8	474	368	
9	397	317	
10	387	316	580
11	331	265	
12	336	279	
13	268	228	
14	210	169	
15	184	162	670
16	187	167	
17	156	138	
18	128	115	
19	124	114	
20	108	97	710
21	98	88	
22	65	59	
23	41	38	
24	43	41	
25	98	80	700

Table 32: Coverage of BNC/COCA word lists by FERN's cognate list

	Es-En Dictionary		En-Es Dictionary	
Version	Spanish Glosses	English Translations	English Glosses	Spanish Translations
20150102	55308	67915	53731	65197
20140311	46369	58014	50679	61787
20130313	42083	52401	46014	56701
20120220	36747	45430	39922	49754
20110308	-	-	35403	44283

word lists, using Google Dictionary. As shown in Table 31, the samples indicate that the true percentage of cognates is approximately 60% in lists four to fifteens, and higher in later lists. The later lists also appear to contain around 10% words that

are correctly classified as proper nouns, and a small percentage of words that should be included with families in higher-frequency lists.

FERN's dictionary includes entries for over 99% of the tokens in Matthias Buchmeier's 27 million word Spanish TV subtitles frequency list[82]. As shown in

Table 32, the quantity of English-Spanish translations in Wiktionary has grown by almost 50% since FERN's cognate list and dictionary was created for the mid-2011 user study. Regenerating FERN's dictionary and cognate list with the latest Wiktionary would produce a much more comprehensive cognate list.

8.3.2 Narrow reading with individualised news search

Narrow reading can help beginners progress toward extensive reading of authentic texts. The narrower the reading, the greater the vocabulary overlap and the easier the reading. Table 33a and Table 33b present examples of narrow reading in English for basketball match reports and top technology articles in Google News.

The collection in Table 33a contained 25 articles that range from 279–1431 words, and average 673 words. The fourth column shows FERN's analysis of the cumulative number of non-cognate word families contained in the BNC/COCA 0–25k word lists. Cognates, proper nouns, numbers, transparent compounds, marginal words and off-list words account for 39% in the basketball collection, and 45% in the technology collection. Words that start with capital letters that are not in the top 1000 list were assumed to be proper nouns. Column five gives the word family count that gives 95% coverage of the tokens (the minimum coverage required for aided extensive reading) in the articles read up to that point, assuming that cognates and proper nouns are known. The final column gives the new word families count as a percentage of the article word count.

The basketball collection contains 16,836 words and 661 word families compared to 14,362 and 824 families for technology. The word families required for 95% coverage plateaued at 220 words for the narrowly-focused basketball collection and 330 for the broader technology collection. The percentage of new word families in each article is below five for 19 of 25 articles in the basketball collection, and 15 of 25 for technology.

[82] http://en.wiktionary.org/wiki/User:Matthias_Buchmeier

Table 33: Coverage word families in two collections of news articles

NBA Reports	Word Count	Total Words	Total Families	95% Families	New Families	New %
1	492	492	106	81	106	21.5%
2	335	827	137	96	31	9.3%
3	449	1276	176	112	39	8.7%
4	279	1555	185	112	9	3.2%
5	717	2272	228	127	43	6.0%
6	660	2932	271	143	43	6.5%
7	721	3653	298	150	27	3.7%
8	720	4373	333	160	35	4.9%
9	442	4815	342	158	9	2.0%
10	259	5074	355	163	13	5.0%
11	691	5765	376	167	21	3.0%
12	324	6089	385	169	9	2.8%
13	675	6764	402	172	17	2.5%
14	1094	7858	435	182	33	3.0%
15	998	8856	458	188	23	2.3%
16	771	9627	470	191	12	1.6%
17	844	10471	510	199	40	4.7%
18	728	11199	527	199	17	2.3%
19	1431	12630	591	219	64	4.5%
20	901	13531	598	218	7	0.8%
21	809	14340	606	217	8	1.0%
22	734	15074	618	217	12	1.6%
23	601	15675	633	216	15	2.5%
24	484	16159	651	218	18	3.7%
25	677	16836	661	217	10	1.5%

Tech Articles	Word Count	Total Words	Total Families	95% Families	New Families	New %
1	615	615	141	110	141	22.9%
2	296	911	173	127	32	10.8%
3	1521	2432	322	200	149	9.8%
4	461	2893	368	223	46	10.0%
5	433	3326	404	238	36	8.3%
6	273	3599	420	240	16	5.9%
7	306	3905	440	245	20	6.5%
8	822	4727	480	248	40	4.9%
9	298	5025	501	258	21	7.0%
10	478	5503	523	265	22	4.6%
11	507	6010	543	273	20	3.9%
12	195	6205	552	278	9	4.6%
13	1199	7404	614	302	62	5.2%
14	571	7975	648	314	34	6.0%
15	279	8254	660	318	12	4.3%
16	398	8652	671	319	11	2.8%
17	377	9029	687	320	16	4.2%
18	529	9558	713	326	26	4.9%
19	501	10059	729	331	16	3.2%
20	549	10608	743	331	14	2.6%
21	486	11094	754	332	11	2.3%
22	1618	12712	792	332	38	2.3%
23	656	13368	808	333	16	2.4%
24	297	13665	816	333	8	2.7%
25	697	14362	824	328	8	1.1%

Table 34: New word families percentages when choosing new articles

a) Choosing a basketball article after reading articles 1 and 2

Article Number	New Words %
4	6.5%
9	8.1%
5	8.2%
7	8.3%
16	8.4%
13	8.6%
3	8.7%
6	8.8%
18	8.9%
23	9.0%
15	9.0%
21	9.0%
25	9.2%
11	9.4%
20	9.4%
14	9.8%
22	10.1%
8	10.7%
17	11.4%
19	11.6%
12	11.7%
10	12.0%
24	12.0%

b) After reading all 25 articles

Article topic	NBA	Tech
Basketball related	3.9%	3.7%
National news	5.7%	4.2%
Apple technology	7.5%	3.6%
Entertainment	7.0%	5.6%
Baseball match	3.2%	4.4%

Table 35: New word families percentages when choosing new articles

Top N	Top 200 + NBA	Top 200 + Tech	Top 200 + Both	Both	Coverage in NBA
50	39	46	30	31	36%
100	73	77	54	57	40%
150	98	105	84	93	45%
200	113	117	101	122	47%
250	125	123	113	144	49%
300	128	127	119	172	50%
350	131	130	121	204	51%
400	136	133	127	233	52%
450	137	134	128	254	53%
500	137	137	131	281	54%
550	137	137	131	299	54%
600	139	138	134	333	55%

Low-ability learners should choose the easiest interesting articles they can find, until the articles contain less than 2% unfamiliar words. Table 34a presents an analysis of new word families percentages for articles 3–25 after reading articles 1 and 2. The percentages range from 6.5–12.0 with article four being the easiest, and thus the most suitable article to read next. This analysis shows that the order in which learners read a series of articles has a big impact on the difficulty of reading.

Google News provides dozens of versions of popular news stories, and FERN's recommendations make it easy to choose the most suitable one to read.

Table 34b shows the new word family percentages for five articles from different topics after reading either the basketball or technology collections. After reading the 25 technology articles other technology articles are much easier as indicated by the example technology article which has 3.6% unknown words, compared to 7.5% for someone who had read the basketball collection. The baseball match report had 3.2% new words with the basketball collection and 4.4% with the technology collection. The first article in Table 34b was about a former NBA player's aspirations to be a head coach, which despite being basketball related, had more common vocabulary with the technology collection than the basketball match reports, partly due to the technology collection containing 50% more word families. These examples illustrate the power of individualised difficulty scores to help learners to select level-appropriate texts that aid effective extensive reading.

Columns two and three in Table 35 show the overlap between the high-frequency words in the basketball and technology collections and a top 200 English word list. Half of the top 200 English words are present in the 150 most frequent word families in each collection. Column five shows the overlap between high frequency words in both collections and column four shows the counts for top 200 English word list that are high frequency in both collections. This data shows that 101 of the top 200 English words are contained in the top 200 words for both collections. The top 200 word families in each collection share 122 word families, of which 101 are also contained in the top 200 English word list. The final column gives the percentage of words in the basketball collection that are covered by the words found in both collections. For example, knowledge of the top 122 word families contained in both collections gives 47% coverage of the basketball collection; adding cognates and proper nouns gives 86% coverage.

Of the words in the top 200 English word list, 30% are absent from both article collections. A learner interested in technology and basketball can prepare to do extensive reading by initially focusing on words that are high frequency in both collections. By focusing on narrow topics, learners can achieve 95% coverage with knowledge of as few as 200 word families plus cognates and proper nouns. The

basketball collection requires 221 word families to reach 95% of which 118 are in the top 200 English word list, and 195 are in the top 1000 word families.

So how does a learner acquire the 200 plus families required for effective narrow reading? One efficient approach is to automatically generate a specialised word list by performing frequency analysis on the result set from search queries for topics that the learner wants to read about. The learners can study the words on the list using FERN's vocabulary exercises and mixed-language texts. Once they begin L2 reading, the learner can practice unfamiliar words in each article in pre-reading exercises, and review words they had difficulty with in post-reading exercises.

8.4 Using machine translation for automatic glossing

Although constantly improving, the quality of machine translation is noticeably inferior to that of expert human translators, especially for complex sentences, which makes machine translation unsuitable for many applications.

FERN uses machine translation to automatically generate glosses and create mixed-language texts. In the classroom user study these glosses helped learners comfortably read texts with 5% unfamiliar words at over 100 words per minute. While students occasionally encountered sub-optimal glosses, no students commented that these glosses caused significant difficulty.

Automatic glossing relies on the word generation phase of the two step translation process in Figure 25, and is not affected by errors in word ordering. This section describes how errors in machine translation impact automatically generated glosses, and analyses the frequency with which errors occur and their impact on L2 learners.

8.4.1 Effect of translation errors

Word order. Machine translators frequently generate the correct words but get them in the wrong order. This error confuses the meaning of the output sentence, but does not negatively impact single word glosses. In the example in Figure 58a, the dash in *Romney-Ryan* is incorrectly interpreted as a break in the sentence, rather than a hyphen linking two nouns, which gives the translated sentence a different meaning to the original. Because the language models for Spanish are large, and the grammar is similar to English, errors in word order are less common than for Japanese to English translation.

a) Original sentence: ¿Nos puede aclarar la política de la administración Romney-Ryan sobre Cuba?
Google translation: Can you clarify the Romney administration policy on Cuba Ryan?
Correct translation: Can you clarify the Romney-Ryan administration policy on Cuba?

b) OG: La posición de Ryan es igual que la mía, quiere fortalecer nuestra política hacia el régimen de Cuba.
GT: Ryan's position is like mine; want to strengthen our policy towards the regime in Cuba.
CT: Ryan's position is like mine; he wants to strengthen our policy towards the regime in Cuba.

c) Original: El actual presidente Obama ha reducido los fondos a Medicare en 700 billones de dólares.
Google: The current President Obama has cut funding to Medicare $ 700 billion.
Correct: The current President, Obama, has cut funding to Medicare by $700 billion.

d) Original: ¿Le gustan?
Google: Do you like?
Correct: Do you like them?

e) Original: Chávez ha dado la bienvenida en Venezuela a Hezbollah, que es un grupo terrorista, y esto es una amenaza muy seria para Estados Unidos.
Google: Chavez has welcomed in Venezuela to Hezbollah, a terrorist group, and this is a very serious threat to the United States.
Correct: Chavez has welcomed into Venezuela Hezbollah, a terrorist group, and this is a very serious threat to the United States.

f) Original: ¿Es verdad?
Google: Is it true?
Correct: Is that true?

g) Original: Si sale presidente, ¿cómo va a resolver este problema?
Google: If you leave President, how will solve this problem?
Correct: If you become President, how will you solve this problem?

h) Original: Creo que Hugo Chávez y Venezuela presentan un reto muy grande para el hemisferio, para América Latina y para Estados Unidos.
Google: I think Hugo Chavez and Venezuela have a very big challenge for the hemisphere, Latin America and the United States.
Correct: I think Hugo Chavez and Venezuela present a very big challenge for the hemisphere, Latin America and the United States.

i) Original: ¿Representa o no Venezuela un peligro para este país?
Google: Venezuela represents or not a danger to this country?
Correct: Does Venezuela represent a danger to this country or not?

Figure 58: Examples of gloss errors and correct translations

Incorrect verb tense and pronoun omission. In Figure 58b, the Spanish word *quiere* means either *she wants to, he wants to,* or *you want to* (formal). In this example the machine translation omits the pronoun and incorrectly translates *quiere* as *want to* instead of *he wants to.* In most cases the impact of these errors on glosses is minor.

Incorrect articles, pronouns and prepositions. Articles, prepositions, and pronouns are commonly translated incorrectly, or incorrectly omitted from the translation (e.g. Figure 58c, Figure 58d), or translated when they should be omitted (e.g. Figure 58e). Incorrect translations or omissions can create problematic glosses.

Correct translation, incorrect context. Many machine translators translate texts sentence by sentence. In Figure 58f, the machine translation is a valid translation of the original sentence, but is incorrect in the context of the article. These errors can result in glosses that are problematic for beginners.

Incorrect sense. When words with multiple senses are translated with an incorrect sense, this often results in problematic glosses. This error often occurs when the correct translation is a low-frequency word sense word, and the machine translator predicts a higher-frequency sense as occurred in the example in Figure 58g and Figure 58h.

Multiple errors. Sentences with multiple errors are more likely to create problematic glosses. Figure 58i gives an example of errors in word order and verb tense in the translation of a question. In this case, neither error results in problematic glosses.

8.4.2 Analysis of glossing versus dictionary

This section presents a series of machine translation errors that result in problematic glosses, and compares three approaches to annotating extensive reading texts:

1. Automatic *gloss* powered by Google Translate;
2. Google Translation of the single word, which gives the most statistically common translation of the word irrespective of context;
3. Google Dictionaries comprehensive Spanish to English translation dictionary.

Figure 59 presents example sentences from an English text that was translated into Spanish by expert human translators.

a) Capítulo 37
1. Gloss: Chapter
2. Translation: chapter
3. Dictionary: (noun) chapter, episode

b) La honradez
1. Gloss: Honesty
2. Translation: the || honesty
3. Dictionary
la (article) the; (noun) la, A; (pronoun) her, it, you
honradez (noun) honesty, integrity, honour, righteousness, straightforwardness, fair-dealing

c) Las responsibilidades familiares
1. Gloss: Family responsibilties
2. Translation: the || family || responsibilities
3. Dictionary
Las (article) the; (pronoun) them
responsibility: responsibilities
familiar (adjective) family, familiar, relative, kin, homely, colloquial, chatty, informal, chummy; (noun) relative, relation, dependent

d) ¿De qué manera definiría usted la caridad?
1. Gloss: How
2. Translation: of || what || way
3. Dictionary
de (preposition) of, from, by, with, at, about, out, off; (conjunction) than, as; (adverb) off
qué (adjective) what, whichever; (pronoun) that, which, who, whom, such, whoever, whosoever; conjunction: than, but
manera (noun) way, manner, fashion, means, mode, style, wise, line, lift

e) ¿De qué manera definiría usted la caridad?
1. Gloss: would you define
2. Translation: define || you || the
3. Dictionary
definir (verb) define, determine
usted (pronoun) you
la (article) the; (noun) la, A; (pronoun) her, it, you

f) ¿De qué manera definiría usted la caridad?
1. Gloss: love
2.Translation: charity
3. Dictionary: (noun) charity, handout

g) El Salvador nos dio el ejemplo de Su vida para que lo sigamos. *(The Savior gave us the example of His life for us to follow.)*
1. Gloss: El Salvador
2. Translation: the || savior
3. Dictionary: (noun) savior, rescuer, saver, redeemer, life guard

h) El Salvador nos dio el ejemplo de Su vida para que lo sigamos. *(The Savior gave us the example of His life for us to follow.)*
1. Gloss: for us to || follow
2. Translation: for || that || it || we follow
3. Dictionary
para: (preposition) for, to, of, towards, by, at, as, for the purpose of, as for, against, as regards, in regards to, as far as, in favour of, compared with/to; (conjunction) in order to, so as to
que: (pronoun) that, which, who, what, whom, such, whoever, whosoever
lo: (pronoun) it, him, to it; article: the

i) El Salvador nos dio el ejemplo de Su vida para que lo sigamos.
1. Gloss: gave || follow
2. Translation: gave || follow (he || continue)

Figure 59: Comparison of automatically generated gloss (1) with out-of-context translation (2) and dictionary entries (3).

Most frequent translation. When the correct gloss is the most frequent translation both the automatic gloss and out-of-context translation are usually identical, whereas the dictionary often gives multiple definitions which hinders reading, as in Figure 59a and Figure 59b.

In Figure 59c a three word phrase is glossed in its entirety. The correct translation for all three words is the most frequent translation, so both the gloss and translations are correct. Displaying phrase glosses can help learners comprehend difficult sentences faster than single word glosses.

Set phrases. In Figure 59d the Spanish phrase *de qué manera* is correctly glossed with the English word *how*. Dictionary entries and translations of the individual words are much less helpful, and give a good illustration of the advantage of the glosses.

Word omission. The Spanish sentence in Figure 59e ends with *la caridad,* literally *the charity*, however, the English translation omits the word *the*. In these cases the phrase glosses are marginally more helpful than the translations.

Acceptable synonyms. Sometimes the automatically generated gloss returns a lower-frequency translation even though the highest-frequency translation is correct. In Figure 59f, the gloss is a word with a similar enough meaning to the correct translation that it would only cause minor difficulty for learners. In this example the out-of-context translation and dictionary definition are superior.

Incorrect sense. Sometimes the context results in a translation for an incorrect sense, while translation of the individual words gives the correct translation. In the example in Figure 59g, *El Salvador* is translated as *El Salvador* (the Central American country) instead of *The Saviour* (referring to Jesus Christ). This shows that in Google's corpus, the term *El Salvador* more commonly refers to the South American country. When the context is taken away *Salvador* is translated as *saviour*.

Idioms and non-literal phrase translations. When the L1 phrase translation differs from the translation of constituent words, as in Figure 59h. Google Translate often translates common idioms with an L2 idiom with equivalent meaning. In both these cases glosses are advantageous for extensive reading. However, learners may get confused if they try to use the gloss to decipher the meaning of individual words in

the phrase. It would be beneficial (but challenging) to automatically alert the user of phrases that had non-literal translations.

Bad translation errors. Occasionally Google Translate makes a bad error with both in-context and out-of-context translations. For example, in July 2012, *dio* incorrectly translated out-of-context as *he*, when the correct most frequent translation is *he/she gave*. The statistical anomalies occur rarely, but frequently enough to necessitate dictionary as a backup. By September 2012 this error was fixed.

There are clearly pros and cons of glosses, translation and dictionary definitions. Good dictionaries cover all the senses of words, but provide excess information that slows reading, and they generally do not contain definitions for set phrases and idioms. The translation and gloss approaches provide great support for extensive reading as long as they are correct. Thus, offering dictionary lookup as a backup is beneficial.

The final section of this chapter quantifies how often gloss errors occur.

8.4.3 Analysing the quality of machine translation powered glosses

FERN's glossing algorithm aimed to meet the following criteria:

- Maximum accuracy: minimise gloss errors that negatively impact reading speed, comprehension and vocabulary learning;
- Maximum coverage: provided useful information for low-frequency words, proper nouns, idioms, set phrases, etc.;
- Domain independence: suitable for all genres of fiction or non-fiction; any length, from long novels to short blog posts; and formal news to informal emails;
- Language independence: easy to adapt to other language pairs.

Using Google Translate makes it easy to meet the final two criteria, since it is now available in 85 languages, and is based on parallel corpora that include texts from a variety of domains. The coverage and accuracy varies greatly between language pairs with English translations to European languages performing better than Asian languages for example, principally due to smaller language distance and larger corpuses.

Table 36: Error analysis for three news articles

Error	**1**	**2**	**3**	**Total**
Wrong pronoun	8	2	1	**11**
Missing word	1	1	2	**4**
Confusing word form	6	0	0	**6**
Wrong word sense	4	6	3	**11**
Wrong word	1	2	0	**3**
Uncommon technical sense	5	0	0	**5**
Total	**25**	**11**	**6**	**42**
Word count	**707**	**550**	**189**	**1446**
Words per error	**28**	**50**	**32**	**34**
Severity of errors				
Severe	**4**	**2**	**2**	**8**
Moderate	**6**	**6**	**2**	**14**
Mild	**15**	**3**	**2**	**20**
First dictionary translation is unsuitable	**15**	**10**	**10**	**35**

This section analyses the accuracy of English glosses of Spanish texts, which is one of the best language pairs for Google Translate.

Three articles—two basketball related and one technology article, were analysed by the thesis author to determine the frequency and severity of errors in automatically generated glosses. Table 36 classifies the gloss errors into six types of errors, with the most common errors being an incorrect word sense and incorrect pronoun, which results from several English pronouns having a single Spanish translation. On average there were errors every 34 words. However, problematic errors occurred every 180 words, and moderately problematic errors occurred every 103 words. Three of the errors were difficult words that were particularly problematic. One was a word with multiple meanings that include a technical basketball term not contained in Wiktionary, and the other two had meanings that are not obvious from examining the several possible meanings in Wiktionary. For the other problematic errors the backup dictionary provided sufficient information to quickly comprehend the sentences.

Finally, the articles were analysed to count the number of cases where the automatically generated gloss provided added benefit over the most frequent dictionary translation. In the combined 1446 words of text, there were 35 cases where the gloss was advantageous, which is about one in every 40 words. This advantage is significant because the majority of these words were ones that are difficult for beginners.

Based on this analysis it is clear that despite occasional errors the automated glosses provide significant benefit over dictionary lookup alone. However, a more comprehensive analysis should be done to determine how easily learners cope with gloss errors and their impact on comprehension across a wide variety of genres and languages.

9 Conclusions

Many individuals claim strong motivation to study a second language but struggle to make time to do so. As Fogg's (2009) model asserts performing a language study requires more than just motivation, it requires sufficient ability, and effective triggers.

Extensive reading is an effective and enjoyable technique that enables learners to combine leisure reading with L2 learning. For those who already read online news regularly, doing extensive reading of L2 news articles enables them to learn a language without having to squeeze in yet-another activity to their already-overcrowded lifestyle. Their habit of reading the news provides an effective trigger to do L2 extensive reading as long as they have sufficient ability.

The aim of this research was to develop a computer-assisted approach that enables learners of all abilities to perform effective extensive reading. This thesis describes the design, implementation and evaluation of FERN, in order to test the thesis central claim that:

An iCALL based approach with learner modelling can enable low-ability second language learners to do effective narrow reading with authentic texts.

This section describes the extent to which the key components of FERN validate this claim.

9.1 Extensive reading with FERN

9.1.1 Automatic glossing

During extensive reading, learners are encouraged to guess the meaning of unfamiliar words instead of using dictionaries which disrupt the flow of reading. The hypertext glosses in FERN alleviate the disruption by providing instant lookup and displaying an automatically disambiguated L1 translation directly above the difficult word. Aided by hypertext glosses, learners read texts with at least 5% unfamiliar words—texts deemed too difficult for unaided extensive reading—and achieved speeds of 100–200 words per minute, which are in the speed range typically associated with extensive reading.

FERN's statistical machine translation powered glosses are automatically generated in the time it takes to load a web page, which enables learners to read any digital text they like. But because the statistical machine translation makes mistakes, a translation dictionary is provided as a backup. The lower-intermediate level learners in the user study agreed that by making use of the glosses the difficulty of news articles was suitable. The thesis author sees no reason that this approach would not work with beginners, but further evaluation is required to prove this assumption.

9.1.2 Narrow reading of news articles and individualised search

News articles on the Internet are typically deemed too difficult for low-ability learners, since the 98% vocabulary coverage required for effective extensive reading requires knowledge of the top 8000 word families (Nation, 2006). Even 95% coverage—which this thesis claims is suitable for extensive reading aided by hypertext glosses—requires knowledge of the top 4000 families, which is well beyond beginners. However, this thesis demonstrates that through narrow reading learners get 95% coverage by learning cognates and an additional 200–300 high-frequency word families. Narrow reading is a good stepping stone towards extensive reading on a broader range of topics.

FERN relies on a learner model that records encounters with words in order to provide an individualised search engine and vocabulary exercises. The system employs a novel approach for determining which parts of an article learners actually read, by tracking reading times for each paragraph and ignoring those with excessively high reading speeds.

FERN's news search provides individualised difficulty scores and word counts that enable learners to determine which version of an interesting news story has the most suitable level of vocabulary. FERN's individualised mixed-language texts and vocabulary exercises are useful for learning the initial vocabulary required to do narrow reading of a specific topic of interest, as well as reviewing words learners have difficulty with during reading.

9.1.3 Influencing motivation

Gamification elements help increase learner motivation and provide triggers, which is particularly important for those who do not already read regularly online, as well as for classroom learners. As demonstrated in the user study, even learners with a

high-degree of intrinsic motivation who enjoy reading with FERN, do little reading outside requirements for course credit.

Concrete goals, such as reading 2000 words per week, are an effective way to motivate learners. Enabling teachers to set concrete goals requires a system that reliably tracks progress and detects cheating. By monitoring paragraph reading speeds, and using fill-the-blanks exercises to test comprehension, FERN reliably tracks reading, and makes cheating difficult.

9.1.4 Significance of this work

With increasing amounts of reading on electronic devices, the techniques developed in this thesis potentially have a big role to play in improving the effectiveness and ease of extensive reading in L2 classrooms, and by independent learners. From several international conferences attended a number of organisations and agencies expressed strong interest in using the technology developed. For example, the CIA language training centre would like to make FERN available in their key languages including Chinese, Arabic and Russian. The Church of Jesus Christ of Latter-day Saints' Mission Language Training Centre in Provo, Utah would like to integrate FERN's capabilities into their existing proprietary system. Researchers at BYU Provo Campus in the School of Education would like to incorporate some of FERN's components into one of their own L2 learning systems.

9.2 Application of key lessons to other domains

In October 2011 the author of this thesis began a project that applied lessons learnt from this thesis in order to create a fun and effective interactive online courses that teach the HTML, CSS, JavaScript and Python computer languages. The resulting CodeAvengers.com learning platform combines video and text instruction with code challenges, quizzes and games that require learners to build their own apps, games and websites in order to complete lessons. For example, the level 1 HTML course requires learners to complete 50 code challenges as they build their own personal profile page. Code Avengers courses are used in thousands of schools worldwide, and by over a million learners.

This section describes key lessons that are applied in both projects, and can be used to develop engaging and effective computer-aided learning systems for any domain.

Interesting and level appropriate materials. FERN's search engine helps learners search and browse for interesting reading material, and provides word counts and difficulty scores that help learners select material of suitable difficulty. In Code Avengers, the projects in the main courses are carefully selected so that they are applicable to a wide audience, and a variety of supplementary projects are provided to cater to learners' varied interests. Where possible, project lessons guide the learner in personalising the end product.

On-demand support tools. FERN's reading interface provides glosses, translations, dictionary and text-to-speech in order to help beginner learners read difficult texts. Code Avengers code challenges provide hints, interactive reference examples, solution code, walk through videos and instant feedback for incorrect submissions.

Minimise extraneous cognitive load. FERN extracts news article content from distracting surroundings with adverts, banners, images and videos so that learners can focus on L2 reading, and provides a simple clean interface for reading, and for its support tools. Little design decisions like sorting answers in multi-choice answers in vocabulary review exercises help minimise extraneous cognitive load. Likewise, the Code Avengers interface is clean and simple. Lesson content uses simple vocabulary and short sentences in order to increase readability, and the tone of instructions is casual and friendly, but avoids local humour that may confuse members of a global audience. Code challenges avoid getting learners to write repetitive and mundane code, so that their efforts are focused on writing code that helps them learn key concepts.

Carefully sequenced repetition. In FERN learners obtain repetition through narrow reading, and completion of individualised vocabulary exercises. In Code Avengers learners complete carefully sequenced lessons that introduce, practice and review key concepts. Learners decide whether or not they want to repeat specific lessons they had difficulty with.

Gamification. FERN incorporates points, badges, leaderboards and weekly reading targets with themes. Code Avengers also incorporates points, badges, leaderboards, and themed lessons which have helped increase engagement in coding classes around the world.

Analytics for teachers. Both FERN and Code Avengers provide reports that give teachers live updates of student progress. These tools enable teachers to quickly know who needs extra help and encouragement.

Cheating detection. Because school students often cheat if they can get away with it, FERN and Code Avengers use time-based cheat detection algorithms. In one university class with 1000 students Code Avengers detected that 10 percent of the class had cheated by copying answers from an electronic answer booklet that was shared among mostly second language English speakers. Informing students that the system detects cheating helps to eliminate blatant cheating.

Easily adaptable to other languages. Learning to code is a challenging topic, and ESL students struggled to complete coding lessons with English instructions. Consequently, Code Avengers lessons have been translated into 20 languages. Enabling learners to learn in their native language greatly reduces the learning burden.

9.3 Adapting to other languages

FERN was optimised for English native speakers learning Spanish. This section describes how to adapt FERN for learning other languages.

9.3.1 Adapting key components

User interface. The FERN user interface is currently available in English and Spanish. Adaptation to other languages requires translation of the strings in a single resource file.

Automatic glossing powered by Google Translate. Theoretically, building on top of Google Translate enables FERN to switch between the 90 supported languages. In practice a bug in the Google Translate API prevented the automatic glossing from working for languages like Japanese that do not use spaces to separate words. This project did not check if the new Google Translate API works for these languages. Languages not supported by Google Translate can either rely solely on dictionary lookup, or would require construction of an alternate disambiguation engine.

L2 search engine. Building on Google News allowed for seamless transition between its 29 supported languages. Given that the Google News API was deprecated in 2011, this feature could be modified to use the Google News RSS

feeds, or the general Google Custom Search API. Alternatively, a custom news search could be built for languages not supported by Google News.

Dictionary. FERN currently contains English and Spanish dictionaries extracted from Wiktionary. The scripts used to extract dictionaries for English and Spanish can be used as is to extract dictionaries for any language available in Wiktionary. Some languages may require custom code to account for language specific markup. Furthermore, for languages with limited support by Wiktionary, an alternate translation dictionary must be obtained, as was done for Maori and Bislama in FERN.

Cognate list. The general approach used to create the list by performing lexical similarity of words with similar meanings can be easily applied to language pairs that use Latin-based character scripts. The completeness of the list depends on the completeness of the translation dictionary for that language pair. Furthermore, some language specific rules were used to map characters with similar phonetics in Spanish and English. Creating a cognate list for non-Latin based characters would require a complete character mapping.

Word family list. The general approach to create the word family lists will work well for many language with good support in Wiktionary, however, language specific rules and additional dictionaries may be required to maximise the completeness.

Learner model and difficulty score. The learner model and difficulty scores are generic enough to be directly applicable to most language pairs. However, some customisation could be used to tailor them to specific languages, as discussed in the following section which gives an example of adapting FERN to Japanese.

9.3.2 Adaptation to Japanese

The language distance between L1 and L2 has a significant impact on the difficulty with which a language is learnt. This thesis developed FERN to help English speakers learning Spanish, a language with a small language distance. This section describes factors to be considered in adapting FERN to Japanese, which has large language distance from English.

One of the initial tasks with learning to read an L2 is learning its character set and phonetics. The Latin based character set and phonetics of Spanish are easy for

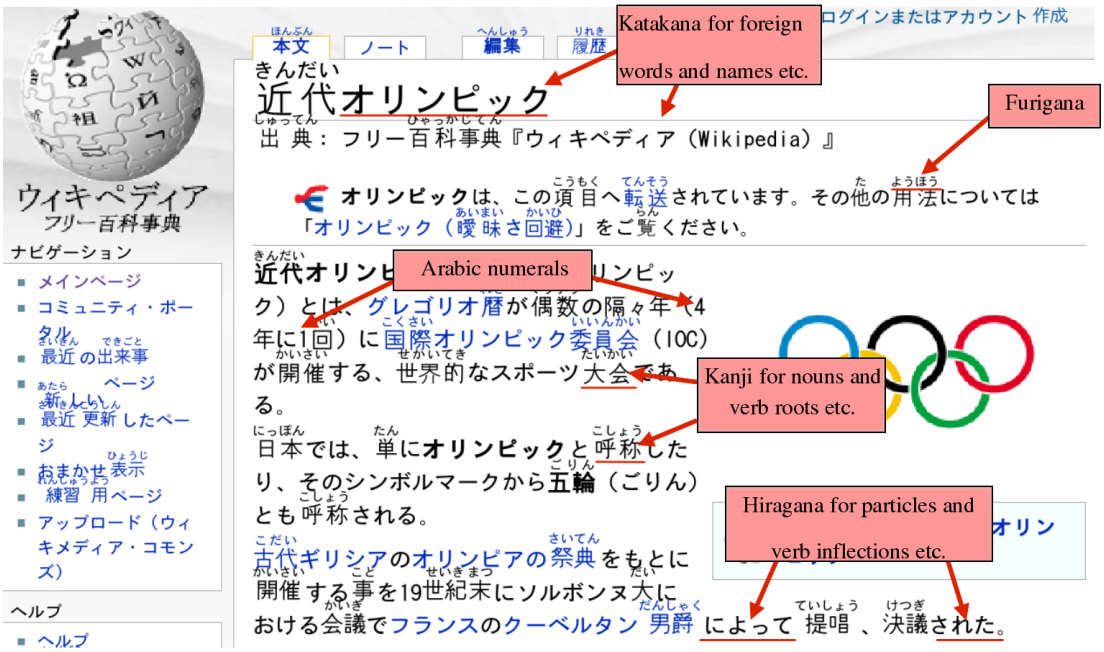


Figure 60: Example Japanese text

English native speakers to learn. In comparison, as shown in Figure 60, Japanese use 4 different character sets.

Kanji are pictographic Chinese characters used to write nouns, stems of verbs and adjectives, and Japanese names. The first 1006 characters are learnt in primary school, 938 are taught at middle and high school, and an additional 983 kanji exist for use in people's names only.

Hiragana is a phonetic writing system, where each of the 46 characters represents a different syllable. Hiragana is used for: verb and adjective inflections; grammatical particles; words with difficult, unknown or no kanji; small characters called furigana, which are written above difficult kanji in order to indicate the pronunciation.

Katakana is a second phonetic writing system with 46 characters, which is used to write: foreign words and names; commonly used animals, plants or objects whose kanji are uncommon, such as "tokage" (lizard), "bara" (rose), "rōsoku" (candle); onomatopoeia; technical and scientific words, such as plant, animal, and mineral names; and to emphasize words, much like italicized words in English text.

Latin alphabet (rōmaji) is used to write: acronyms and initials, for example UN; some company names, brand names or product names; and Arabic numerals are often used to write numbers.

Beginners learn hiragana then katakana, before embarking on the arduous journey of mastering kanji, of which learners must know 2000 in order to fluently read the newspaper. Texts for children and L2 learners are often simplified by replacing uncommon kanji with hiragana, and applying furigana (small phonetic characters) to difficult kanji or katakana. Books for English speakers learning Japanese also use rōmaji (Latin alphabet) in order to provide support for those still learning hiragana.

Another difficulty with reading languages like Japanese and Chinese is that there is usually no space between adjacent words. Periods are used to mark the end of sentences, and commas are sometimes used mid-sentence to separate phrases. Texts are more difficult to read for beginners when there are long runs of one type of character, sometimes exceeding ten characters. Japanese text readability formulas take into account the average length of runs of kanji, hiragana and katakana. With practice learners can recognise word boundaries in long runs of same type of character. FERN could incorporate existing algorithms to automatically detect word boundaries and insert spaces and punctuation in order to make reading easier for beginners, and reduce the space when it is no longer needed.

An effective Japanese learner profile should incorporate both vocabulary and character knowledge. Learners can guess the meaning of an unfamiliar word more easily if they know one or more of its characters. Additionally, FERN's difficulty formula for Japanese should take into consideration average run lengths, and proportion of unknown characters.

Creating a Japanese-English cognate list is straightforward, because all loan words are written in the katakana character set, and most of these are borrowed from English. Note that the character set and phonetic differences with Japanese-English cognates presents a significant obstacle to fluent reading, relative to Spanish-English cognates.

Extracting a dictionary from Wiktionary is unnecessary because of the existence of Jim Breen's freely available translation and kanji dictionaries[83]. Creating a Japanese word family list is straightforward because the roots for all forms are essentially identical and they have few verb and adjective inflections.

[83] http://www.csse.monash.edu.au/~jwb/japanese.html

9.4 Future work

FERN worked well in the user study with a class of lower-intermediate learners, and chapter 8 provided theoretical evidence of its suitability with beginners, but future work should validate its effectiveness with an actual class of beginner learners.

This section describes additional recommendations for future work and improvements to FERN.

9.4.1 Improving the glossing algorithm

Reducing the frequency of errors in FERN's automatically generated glosses requires one of two approaches: replace Google Translate with a custom word sense disambiguation engine, or improving Google Translates disambiguation. Waiting for gradual improvements in Google Translate, or getting a job on the Google Translate team would probably be easier than developing a custom disambiguation engine with superior performance.

There are also several approaches to reducing the impact of errors in FERN's glosses. The first approach involves automatic detection of glosses with questionable accuracy using language specific heuristics, or by cross referencing glosses with a part-of-speech tagger to ensure that the parts-of-speech of the original text match the glosses. A second approach, which requires a large user base, is to crowd source corrections to gloss errors in popular texts. Finally, Google Translate's algorithm produces several likely glosses with probability scores for each, but this data is not provided by the public API. If this data was accessible by partnering with Google, FERN could automatically display two gloss options if their probability scores are similar.

9.4.2 Presentation of glosses

By accessing glosses learners are less likely to learn a target word than by expending mental effort to guess or remember the meaning of a previously encountered word (Hulstijn, 1992). FERN can increase the learning of words learners click using the following approach:

- Display full glosses for previously unencountered words.

- For partially known words display a gloss hint, such as the first and last letter with other letters replaced by blanks; a multiple choice gloss (the correct word plus two or three distractors); or a phrase from the previous context in which the word appeared.
- Highlight troublesome partially known words to increase salience so that the learner is more likely to try to recall the meaning, before clicking on it.

These techniques encourage learners to expend effort to remember the meaning of difficult words, which facilitates learning, but may reduce reading comprehension and speed. It would be interesting to investigate this trade-off.

Given that advanced learners prefer L2 glosses, FERN could transition from L1 to L2 definitions if the learner's model indicates that they knows the L2 words contained in the gloss.

9.4.3 Dictionary

There are many improvements that could be made to FERN's dictionary entries.

The presentation of entries may be improved by displaying translation in frequency order, and using a dark font for high-frequency and light font for low-frequency translations in order to minimise the distraction. Words with many possible translations could hide the lowest frequency options and show them with a show all button.

Misspelled words are problematic, especially when the misspelling produces a legitimate word. FERN could detect misspellings and provide links to view the definition of the misspelling.

FERN's dictionary translations are sourced from Wiktionary, a growing but incomplete source. The following approaches could provide useful information when learners attempt to look up words that are absent from the dictionary.

- Consult dictionaries of languages separated by a small language distance (such as Portuguese, English, Italian, and French) for an identical or lexically similar word.
- To overcome sparse coverage of proper nouns in Wiktionary, consult a proper noun dictionary extracted from Wikipedia. Wikipedia links articles on the same topic in different languages, and translations are obtained from

the article titles. The opening sentence, or article information box are generally suitable definitions.

- Wiktionary excludes many low-frequency compound nouns. It is straightforward for FERN to automatically detect compounds and display definitions for constituent words.

9.4.4 Improving supplementary exercises

One of the major improvement to the exercises would be to incorporate example sentences that learners had difficulty with during reading.

Researchers have questioned the efficiency of pre- and post-reading exercises. This thesis proposes that individualised exercises are an efficient use of time when they practice the right words. FERN should create pre-reading exercises that practice words that are likely to be difficult for the learner, and that appear multiple times in the text they are about to read. Future work should evaluate how this approach aids vocabulary learning and reading comprehension.

The classroom user study in this thesis involved intermediates for whom mixed-language texts were a less efficient approach to learning than L2 reading. Future work could evaluate the use of mixed-language texts and vocabulary exercises with beginners learning high-frequency words that relate to individual interests. Researching the following questions would help determine the optimal format for mixed-language texts:

- How is readability affected by the replacement of different parts of speech?
- How does the number of words translated per sentence affect readability?
- How quickly should the difficulty level increase?

It would also be valuable to develop and test a web browser plugin that could transform any article, blog post or email displayed in a web browser into a mixed-language text.

9.4.5 Article screen scraper

FERN's screen scraper algorithm works for 95% of English and Spanish articles returned by Google News. This performance can be improved by taken advantage of the adoption of new semantic tags HTML5, such as figure, article, aside, and nav, which make it easier to identify news articles. FERN could also implement

support for structured markup such as the schema.org microdata schemas, which add itemprop attributes with a value of articleBody to elements that contains article content.

9.4.6 Improvements to the Individualised Search and Learner Model

There are four areas of improvement relating to FERN's individualised search. First, the learner model could be enhanced to account for the forgetting of words over time. Second, the system could give new users a short adaptive test that gauges initial ability, so that the system can offer personalised difficulty scores right from the first search. Third, aggregated usage data can be used to create a word list that ranks the difficulty of word families during reading, and this data could be incorporated into the difficulty scores. Fourth, the difficulty formula can be refined by applying machine learning to user study data, and incorporating word-level features such as word-sense count and proximity to other ambiguous words, as well as traditional sentence-level features.

FERN could incorporate collaborative recommendations by requiring learners to rank their enjoyment of articles. The system can then automatically recommend articles to learners by comparing the set of content nouns—especially proper nouns—in articles they enjoyed in the past with new articles.

9.4.7 Monitoring and reporting progress

FERN's progress charts could be improved by using time on the x-axis instead of number of articles, which was used because it made the implementation easier. Also, the gloss access chart would be more useful if it displayed a moving mean. The impact of the vocabulary size chart would be increased by displaying learners' progress towards a concrete goal such as being able to read news articles from a topic of interest, without help from glosses.

The current approach to detecting cheating in classroom extensive reading could be improved by tracking the words the learners click during reading. Paragraphs with unusual gloss access patterns, such as clicking on too many easy words, or not clicking on enough difficult words, can be ignored from reading logs.

Further work should be done to investigate the behaviours that triggers the outliers in paragraph reading speeds.

9.4.8 Future studies using FERN

There are many studies that could be conducted with FERN to investigate the optimal conditions for extensive reading. For example, a study could investigate how various reading goals help learners develop the reading habit by giving one group the goal of reading 2000 words per week, and another the goal of reading 400 words 5 days per week.

9.5 Open source versus commercial licenses

This research experienced several challenges as a result of decisions to incorporate various open source and commercial components. This section discusses important considerations that must be made when incorporating third party components.

9.5.1 Comparing licences

Open source projects give full access to the code and permit developers to freely use and modify it. Some like the GPL (GNU General Public License) require derived[84] works to be distributed under the same license terms, a term that does not apply to web applications, because they are not deemed to be "distributed". AGPL components may only be used in open-source systems. Others such as MIT and Apache permit developers to distribute without restriction.

Google APIs have a general 3000 word terms of service with extra terms for some APIs.[85] These APIs require a track usage to ensure compliance and reserve the right to suspend access to systems that they believe have security issues that could affect Google or its users. Developers agree not to "scrape, build databases or otherwise create permanent copies of such content" or to "copy, translate, modify, create a derivative work".

Google's APIs had a three-year deprecation policy, which could be shortened to prevent "substantial economic burden on Google". Google News was deprecated with the standard three-year shut down period. Due to "excessive abuse" Google Translate (version 1) was given a six month deprecation period.[86] This

[84] The definition of *derived works* in respect to software copyright law is somewhat contentious.

[85] http://googlecode.blogspot.co.nz/2011/12/introducing-google-apis-terms-of.html

[86] http://www.zdnet.com/blog/burnette/google-pulls-the-rug-out-from-under-web-service-api-developers-nixes-google-translate-and-17-others/2284

announcement was met with public outcry that resulted in Google Translate version 2 API, a paid service. In April 2012 Google reduced the deprecation period to one year for four critical services and abolished it for the rest.

Using commercial APIs is risky and it is important to have a backup plan.

9.5.2 Source code access and versions

A major benefit of open source software is that developers can customise them—as was done with JSoup in FERN—and fix bugs. When using closed source APIs developers rely on its creators to fix bugs. This project discovered bugs that prevented the Google Translate Research API returning correct word alignments when translating from Japanese or Chinese to English. A Google developer stated that the bug would not be fixed because the API was no longer being developed.

The constant veiled improvements of proprietary web APIs make comparing and evaluating systems that use them difficult, because results may vary from day to day.

9.5.3 Challenges with the Google Translate Research API

Several challenges were experienced in using the Research API. First, it took several months to get access. Second, there was uncertainty around the duration of the availability of the API—which was eventually shut down a month before the completion of this thesis. Third, there was a limit of 10,000 queries per day—which was extended to 150,000 by a Google Engineer. Fourth, the API had a bug that made word alignments not work when translating Japanese and Chinese. Furthermore, the translation quality is poor for many languages. Consequently, this research focused on Spanish, which has adequate translation quality.

10 References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199. http://doi.org/10.1080/09588220802090246

AbuSeileek, A. F. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computers & Education*, *57*(1), 1281–1291.

Adams, M. J. (1994). *Beginning to read: Thinking and learning about print*. MIT press. Retrieved from https://books.google.co.nz/books?hl=en&lr=&id=P_Hk7-n8i1AC&oi=fnd&pg=PR5&dq=Adams+1994+comprehension+reading+cognitive&ots=PnQD1855PK&sig=ItkAvj4HbfP1umTrs92iCbGGNV4

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734–749. http://doi.org/10.1109/TKDE.2005.99

Afarian, K. M. (1987). *The effect of the Burling reader method: A study of initial Spanish acquisition and affective factors*. Brigham Young University. Dept. of Linguistics.

Aldabe, I., de Lacalle, M., Maritxalar, M., Martinez, E., & Uria, L. (2006). Arikiturri: An automatic question generator based on corpora and nlp techniques. In *Intelligent Tutoring Systems* (pp. 584–594).

Alessi, S., & Dwyer, A. (2008). Vocabulary assistance before and during reading. *Reading in a Foreign Language*, *20*(2), 246–263.

Al-Homoud, F., & Schmitt, N. (2009). Extensive reading in a challenging environment: a comparison of extensive and intensive reading approaches in Saudi Arabia. *Language Teaching Research*, *13*(4), 383–401.

Aluísio, S. M., Specia, L., Pardo, T. A. ., Maziero, E. G., Caseli, H. M., & Fortes, R. P. . (2008). A corpus analysis of simple account texts and the proposal of simplification strategies: first steps towards text simplification systems. In *Proceedings of the 26th annual ACM international conference on Design of communication* (pp. 15–22).

Aluísio, S. M., Specia, L., Pardo, T. A. ., Maziero, E. G., & Fortes, R. P. . (2008). Towards brazilian portuguese automatic text simplification systems. In *Proceeding of the eighth ACM symposium on Document engineering* (pp. 240–248).

Aluisio, S., Specia, L., Gasperin, C., & Scarton, C. (2010). Readability assessment for text simplification. In *Fifth Workshop on Innovative Use of NLP for Building Educational Applications* (p. 1).

Anderson, J. (1981). Analysing the Readability of English and Non-English Texts in the Classroom with Lix. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED207022

Anderson, N. (1999). Improving reading speed. In *English Teaching Forum* (Vol. 37). Retrieved from http://dosfan.lib.uic.edu/usia/E-USIA/forum/acrobat/P2.pdf

Angeles, B. C. E. in L. (2007, March 6). Fake Wikipedia prof altered 20,000 entries. *Telegraph.co.uk*. Retrieved from http://www.telegraph.co.uk/news/1544737/Fake-Wikipedia-prof-altered-20000-entries.html

Aoyama, K. (2005). *Using a diglot reader to teach kanji: The effects of audio and romaji on the acquisition of kanji vocabulary*. Brigham Young University.

Ariew, R., & Ercetin, G. (2004). Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning*, *17*(2), 237–259.

Arnold, N. (2009). Online extensive reading for advanced foreign language learners: An evaluation study. *Foreign Language Annals*, *42*(2), 340–366.

Ashworth, D., & Stelovsky, J. (1989). Kanji City An Exploration of Hypermedia Applications for CALL. *CALICO Journal*, *6*(4), 27–39.

Atkinson, R. C., & Fletcher, J. D. (1972). Teaching Children to Read with a Computer. *The Reading Teacher*, *25*(4), 319–327. http://doi.org/10.2307/20192980

Atkinson, R. C., Fletcher, J. D., Chetin, H. C., & Stauffer, C. M. (1971). *Instruction in initial reading under computer control: the Stanford project*. Educational Technology Publications. Retrieved from http://suppes-corpus.stanford.edu/techreports/IMSSS_158.pdf

Atkinson, R. C., & Hansen, D. N. (1966). Computer-assisted instruction in initial reading: The Stanford project. *Reading Research Quarterly*, 5–25.

Aust, R., Kelley, M. J., & Roby, W. (1993). The use of hyper-reference and conventional dictionaries. *Educational Technology Research and Development*, *41*(4), 63–74.

Barfield, A. (1995). Extensive reading: from graded to authentic text. *Language Learning*, *2*(1).

Bauer, L. (1993). Manual of Information to Accompany the Wellington Corpus of Written New Zealand English. Department of Linguistics, Victoria University, Wellington, New Zealand.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279.

Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary development: All contexts are not created equal. *The Elementary School Journal*, 177–181.

Bell, T. (1998). Extensive reading: Why? and how. *The Internet TESL Journal*, *4*(2).

Bell, T. (2001). Extensive reading: Speed and comprehension. *Reading*, *1*(1).

Bennöhr, J. (2007). A web-based personalised textfinder for language learners. *Data Structures for Linguistic Resources and Applications. Tübingen: Gunter Narr Verlag*. Retrieved from http://www.josiephil.de/jasmine/JBennoehr_textfinder_cor3.pdf

Berardo, S. A. (2006). The use of authentic materials in the teaching of reading. *The Reading Matrix*, *6*(2). Retrieved from http://www.researchgate.net/profile/Sacha_Berardo/publication/237413136_THE_USE_OF_AUTHENTIC_MATERIALS_IN_THE_TEACHING_OF_READING/links/5433c4ca0cf2bf1f1f262635.pdf

Billuroğlu, A., Eldridge, J., Hancioğlu, N., & Neufeld, S. (2008). A Research and Development Project for English Language Learning.

Blake, R. J. (1992). Second language reading on the computer. *ADFL Bulletin*, *24*(1), 17–22.

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-Assisted Instruction in Support of Beginning Reading Instruction: A Review. *Review of Educational Research*, *72*(1), 101–130. http://doi.org/10.3102/00346543072001101

Bogert, J. (1985). In Defense of the Fog Index. *Business Communication Quarterly*, *48*(2), 9–12. http://doi.org/10.1177/108056998504800203

Bowles, M. A. (2004). L2 Glossing: To CALL or Not to CALL. *Hispania*, *87*(3), 541–552.

Brew, C., & McKelvie, D. (1996). Word-pair extraction for lexicography. In *Proceedings of the second international conference on new methods in language processing* (pp. 45–55). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.31.7741&rep=rep1&type=pdf

Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, amt018.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1), 107–117.

Broughton, J. (2007). *Wikipedia: The Missing Manual*. O'Reilly Media, Inc.

Brown, A., Hill, K., & Iwashita, N. (2000). Is learner progress in LOTE learning comparable across languages? *Australian Review of Applied Linguistics*, *23*(2), 35–60.

Brown, A. R. (2011). Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS: Political Science & Politics*, *44*(02), 339–343. http://doi.org/10.1017/S1049096511000199

Brown, D. (2009). Why and how textbooks should encourage extensive reading. *ELT Journal*, *63*(3), 238–245.

Brown, D. (2010). An Improper Assumption? The Treatment of Proper Nouns in Text Coverage Counts. *Reading in a Foreign Language*, *22*(2), 355–361.

Browne, C. (2008). A Research-based approach to developing Portable Technology Software for testing and teaching high frequency vocabulary. In

CamTESOL Conference on English Language Teaching: Selected Papers (Vol. 4, p. 8).

Browne, C. (2014). A New General Service List: The Better Mousetrap We've Been Looking for? *Vocabulary Learning and Instruction*, 1.

Browne, C., & Culligan, B. (2008). Combining technology and IRT testing to build student knowledge of high frequency vocabulary. *The JALT CALL Journal*, *4*(2), 3–16.

Brown, J. C., Frishkoff, G. A., & Eskenazi, M. (2005). Automatic question generation for vocabulary assessment. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 819–826).

Brown, J. D. (1998). An EFL readability index. *JALT Journal*, *20*(2), 7–36.

Brown, J., & Eskenazi, M. (2004). Retrieval of authentic documents for reader-specific lexical practice. In *InSTIL/ICALL Symposium 2004*.

Burling, R. (1968). Some outlandish proposals for the teaching of foreign languages. *Language Learning*, *18*(1 & 2), 61–76.

Burstein, J., Shore, J., Sabatini, J., Lee, Y. W., & Ventura, M. (2007). The automated text adaptation tool. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations on XX* (pp. 3–4).

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Literacy Research*, *26*(4), 413–437.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.

Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., … Gruber, R. E. (2008). Bigtable: A distributed storage system for structured data. *ACM Transactions on Computer Systems (TOCS)*, *26*(2), 4.

Chan, M.-C., Chau, H. L. H., & Hoosain, R. (1983). Input/output switch in bilingual code switching. *Journal of Psycholinguistic Research*, *12*(4), 407–416. http://doi.org/10.1007/BF01067622

Cheng, H. F., & Dörnyei, Z. (2007). The use of motivational strategies in language instruction: The case of EFL teaching in Taiwan. *Innovation in Language Learning and Teaching*, *1*(1), 153–174.

Cheng, Y. H., & Good, R. L. (2009). L1 glosses: Effects on EFL learners 'reading comprehension and vocabulary retention. *Reading in a Foreign Language*, *21*(2), 119–142.

Cheng, Y. L., & Howard, D. (2008). The time cost of mixed-language processing: an investigation. *International Journal of Bilingualism*, *12*(3), 209.

Chen, Q. J. (1997). *Learning Chinese characters through self-study: An experiment with the diglot reader method*. Brigham Young University. Dept. of Language Acquistiton.

Cho, K. S., Ahn, K. O., & Krashen, S. (2005). The Effects of Narrow Reading of Authentic Texts on Interest and Reading Ability in English as a Foreign Language. *Reading Improvement*, *42*(1), 58–65.

Cho, K. S., & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, *37*(8), 662–667.

Christensen, E., Merrill, P., & Yanchar, S. (2007). Second language vocabulary acquisition using a diglot reader or a computer-based drill and practice program. *Computer Assisted Language Learning*, *20*(1), 67–77.

Chujo, K., & Utiyama, M. (2005). Understanding the role of text length, sample size and vocabulary size in determining text coverage. *Reading in a Foreign Language*, *17*(1), 1–22.

Chun, D. M. (2011). CALL technologies for L2 reading post Web 2.0. *Present and Future Promises of CALL: From Theory and Research to New Directions in Language Teaching*, 131–170.

Clauson, K. A., Polen, H. H., Boulos, M. N. K., & Dzenowagis, J. H. (2008). Scope, completeness, and accuracy of drug information in Wikipedia. *The Annals of Pharmacotherapy*, *42*(12), 1814–1821. http://doi.org/10.1345/aph.1L474

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*(3), 38–63.

Collins-Thompson, K., Bennett, P. N., White, R. W., de la Chica, S., & Sontag, D. (2011). Personalizing web search results by reading level. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 403–412). Retrieved from http://dl.acm.org/citation.cfm?id=2063639

Collins-Thompson, K., & Callan, J. (2004). Information retrieval for language tutoring: An overview of the REAP project. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 544–545). Retrieved from http://dl.acm.org/citation.cfm?id=1009112

Collins-Thompson, K., & Callan, J. (2005). Predicting reading difficulty with statistical language models. *Journal of the American Society for Information Science and Technology*, *56*(13), 1448–1462.

Conradson, S., & Hernández-Ramos, P. (2004). Computers, the internet, and cheating among secondary school students: Some implications for educators. *Practical Assessment, Research & Evaluation*, *9*(9). Retrieved from http://pareonline.net/getvn.asp?v=9&n=9

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing Text Readability Using Cognitively Based Indices. *TESOL Quarterly*, *42*(3), 475–493. http://doi.org/10.2307/40264479

Crossley, S., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, *23*(1), 84–101.

Curtin, C., Dawson, C. L., Provenzano, N., & Cooper, P. (1976). The Plato System: Using the Computer to Teach Russian. *The Slavic and East European Journal*, *20*(3), 280–292. http://doi.org/10.2307/306322

Cyr, D., & Bonanni, C. (2005). Gender and website design in e-business. *International Journal of Electronic Business*, *3*(6), 565–582.

Dale, E. (1931). A comparison of two word lists. *Educational Research Bulletin*, 484–489.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, 11–28.

Dale, E., & Tyler, R. W. (1934). A Study of the Factors Influencing the Difficulty of Reading Materials for Adults of Limited Reading Ability. *The Library Quarterly*, *4*(3), 384–412. http://doi.org/10.2307/4302090

Dalrymple-Alford, E. C. (1985). Language switching during bilingual reading. *British Journal of Psychology*, *76*(1), 111–122.

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., … others. (2010). The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293–296).

Davis, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *Modern Language Journal*, *73*(1), 41–48.

Davis, J. N., & Lyman-Hager, M. A. (1997). Computers and L2 Reading: Student Performance, Student Attitudes1. *Foreign Language Annals*, *30*(1), 58–72.

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge Univ Pr.

de Burgh-Hirabe, R., & Feryok, A. (2013). A model of motivation for extensive reading in Japanese as a foreign language. *Reading in a Foreign Language*, *25*(1), 72–93.

De Groot, A., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning*, *50*(1), 1–56.

de Morgado, N. F. (2009). Extensive Reading: Students' Performance and Perception. *Reading*, *9*(1).

De Pessemier, T., Deryckere, T., & Martens, L. (2009). Context aware recommendations for user-generated content on a social network site. In *Proceedings of the seventh european conference on European interactive television conference* (pp. 133–136).

DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*(3), 1616–1641.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification: Using game-design elements in non-gaming contexts. Ext. In *Abstracts CHI* (Vol. 11).

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*(3), 297–302.

Dokter, D. (1998). From Glosser-RuG to Glosser-WeB. *Alfa-Informatica, University of Groningen. Retrieved from the WWW in November*.

Dokter, D. A. (1997). Glosser-RuG; Prototype December 1996. *Techreport, Alfa-Informatica, University of Groningen, Groningen, The Netherlands*.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal*, *78*(3), 273–284.

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, *2*(3), 203.

DuBay, W. H. (2007). *Smart Language*. Impact Information. Retrieved from http://www.nald.ca/library/learning/smartlang/smartlang.pdf

Dundar, H., & Akcayir, M. (2012). Tablet vs. Paper: The Effect on Learners' Reading Performance. *International Electronic Journal of Elementary Education*, *4*(3), 441–450.

Dupuy, B., Tse, L., & Cook, T. (1996). Bringing books into the classroom: First steps in turning college-level ESL students into readers. *TESOL Journal*, *5*(4), 10–15.

Einbinder, H. (1964). *The myth of the Britannica*. Grove Press.

Eldridge, J., & Neufeld, S. (2009). The Graded Reader is Dead, Long Live the Electronic Reader. *Reading*, *9*(2).

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*(2), 174–187.

Elley, W. B., & Mangubhai, F. (1981). The impact of a book flood in Fiji primary schools. *Wellington: New Zealand Council for Educational Research*.

Elley, W. B., & Mangubhai, F. (1982). The long-term effects of a book flood on children's language growth. *Directions, USP*, *7*, 15–24.

Ellis, R. (1990). *Instructed second language acquisition: Learning in the classroom*. Basil Blackwell Oxford. Retrieved from http://www.bu.edu.eg/portal/uploads/digital_lib_books/bookInfo/9538598_info.pdf

Erickson, L. B., Petrick, I., & Trauth, E. M. (2012). Organizational uses of the crowd: developing a framework for the study of crowdsourcing. In *Proceedings of the 50th annual conference on Computers and People Research* (pp. 155–158). New York, NY, USA: ACM. http://doi.org/10.1145/2214091.2214133

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200.

Felix, P. (1981). The bilingual reading method or the "Mamipravalam" way of learning a foreign language. *CIEFL Bulletin*, *17*, 35–49.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221–233.

Fogg, B. J. (2009). A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology* (p. 40). Retrieved from http://dl.acm.org/citation.cfm?id=1541999

Fowler Jr, G. L. (1978). The Comparative Readability of Newspapers and Novels. *Journalism Quarterly*, *55*(3), 589–92.

Francis, W. N., & Kucera, H. (1979). Brown corpus manual. *Brown University*.

François, T., & Miltsakaki, E. (2012). Do NLP and machine learning improve traditional readability formulas? In *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations* (pp. 49–57). Retrieved from http://dl.acm.org/citation.cfm?id=2390925

Frunza, O., & Inkpen, D. (2010). Identification and disambiguation of cognates, false friends, and partial cognates using machine learning techniques. *International Journal of Linguistics*, *1*(1). Retrieved from http://macrothink.org/journal/index.php/ijl/article/view/309

Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, *11*(7), 513–578.

Fry, E. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading*, *21*(3), 242–252.

Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading. *Reading in a Foreign Language*, *20*(1), 1.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, amt015.

Gettys, S., Imhof, L. A., & Kautz, J. O. (2001). Computer-Assisted Reading: The Effect of Glossing Format on Comprehension and Vocabulary Retention. *Foreign Language Annals*, *34*(2), 91–99.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901. http://doi.org/10.1038/438900a

Gilliam, B., Peña, S. C., & Mountain, L. (1980). The Fry graph applied to Spanish readability. *The Reading Teacher*, *33*(4), 426–430.

Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: approaches to secondlanguage vocabulary learning. *Language Learning & Technology*, *14*(2), 4–11.

Goto, T., Kojiri, T., Watanabe, T., Iwata, T., & Yamada, T. (2010). Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, *2*(3), 210.

Grabe, W. (2009). *Reading in a second language: moving from theory to practice*. Cambridge University Press.

Gray, W. S., & Leary, B. E. (1935). What makes a book readable. Retrieved from http://doi.apa.org/index.cfm?fa=search.displayRecord&uid=1935-05257-000

Green, C. (2005). Integrating extensive reading in the task-based curriculum. *ELT Journal*, *59*(4), 306–311.

Greenfield, J. (2004). Readability formulas for EFL. *JALT Journal*, *26*(1), 5–24.

Groot, P. J. . (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, *4*(1), 60–81.

Grosjean, F. (2001). The bilingual's language modes. *One Mind, Two Languages: Bilingual Language Processing*, 1–22.

Gunderson, P. (1993). *The effect of a video diglot-weave method in introducing beginning languages*. Brigham Young University. Dept. of Linguistics.

Hamari, J., Koivisto, J., & Sarsa, H. (n.d.). Does Gamification Work?—A Literature Review of Empirical Studies on Gamification. Retrieved from http://www.researchgate.net/publication/256743509_Does_Gamification_Work__A_Literature_Review_of_Empirical_Studies_on_Gamification/file/9c960523b36be81843.pdf

Han, Z., & Chen, C. A. (2010). Repeated-reading-based instructional strategy and vocabulary acquisition: A case study of a heritage speaker of Chinese. *Reading in a Foreign Language*, *22*(2), 242–262.

Harmon, N. S. (2001). *The planning and design of a computer program to teach Spanish through The Book of Mormon*. Brigham Young University.

Harmon, N. S. (2002). *Book of Mormon Stories diglot reader on computer*. Brigham Young University. Dept. of Instructional Psychology and Technology.

Harrison, C. (1992). The reading process and learning to read. In C. Harrison & M. Coles (Eds.), *The reading for real handbook* (pp. 3–28). London: Routledge.

Harvey, P. (2008a). The importance of individualisation in reading in a second language Part Three: Implications for the classroom. *MODERN ENGLISH TEACHER*, *17*(2), 58.

Harvey, P. (2008b). The importance of individualisation in reading in a second language Part Two: Some important questions related to L2 reading. *MODERN ENGLISH TEACHER*, *17*(1), 48.

Hauer, B., & Kondrak, G. (2011). Clustering Semantically Equivalent Words into Cognate Sets in Multilingual Lists. In *IJCNLP* (pp. 865–873). Retrieved from http://www.newdesign.aclweb.org/anthology/I/I11/I11-1097.pdf

Heilman, M., Collins-Thompson, K., Callan, J., & Eskenazi, M. (2006). Classroom success of an Intelligent Tutoring System for lexical practice and reading comprehension. In *Ninth International Conference on Spoken Language Processing*.

Heilman, M. J., Collins, K., & Callan, J. (2007). Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts.

Heilman, M., Juffs, A., & Eskenazi, M. (2007). Choosing reading passages for vocabulary learning by topic to increase intrinsic motivation. *FRONTIERS IN ARTIFICIAL IN TEL LIGENCE AND APPLICATIONS*, *158*, 566.

Heilman, M., Zhao, L., Pino, J., & Eskenazi, M. (2008). Retrieval of reading materials for vocabulary and reading practice. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 80–88).

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, *10*(5), 164.

Higgins, J., & Wallace, R. (1989). Hopalong: A computer reading pacer. *System*, *17*(3), 389–399.

Hill, D. R. (2001). Survey. Graded readers. *ELT Journal*, *55*(3), 300 –324. http://doi.org/10.1093/elt/55.3.300

Hill, D. R. (2008). Graded readers in English. *ELT Journal*, *62*(2), 184 –204. http://doi.org/10.1093/elt/ccn006

Hill, D. R. (2013). Graded readers. *ELT Journal*, *67*(1), 85–125.

Hill, D. R., & others. (2011). *The EPER guide to organising programmes of extensive reading*. Edinburgh: University of Edinburgh (IALS).

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*, 689–689.

Hitosugi, C. I., & Day, R. R. (2004). Extensive reading in Japanese. *Reading in a Foreign Language*, *16*(1), 20–39.

Holbrook, M. B. (1986). Aims, concepts, and methods for the representation of individual differences in esthetic responses to design features. *Journal of Consumer Research*, 337–347.

Holley, F. M., & King, J. K. (1971). VOCABULARY GLOSSES IN FOREIGN LANGUAGE READING MATERIALS. *Language Learning*, *21*(2), 213–219. http://doi.org/10.1111/j.1467-1770.1971.tb00060.x

Hong, Y. F. (2006). The Effect of L1 Vocabulary Glosses on EFL Elementary School Students' Reading Comprehension and Reading Process.

Horst, M. (2001). Text encounters of the frequent kind: Learning L2 vocabulary through reading. *Unpublished Doctoral Dissertation, University of Wales, Swansea*.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *61*(3), 355–382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring Second Language Vocabulary Through Reading. *Reading in a Foreign Language*, *11*(2), 207–23.

Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *56*(2), 308–328.

Hoshino, A., Huan, L., & Nakagawa, H. (2008). A framework for automatic generation of grammar and vocabulary questions. *Proceedings of WorldCALL*.

Huckin, T., & Coady, J. (1999). Incidental Vocabulary Acquisition in a Second Language. *Studies in Second Language Acquisition*, *21*(02), 181–193. http://doi.org/10.1017/S0272263199002028

Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language*, *26*(2), 2.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental learning. *Vocabulary and Applied Linguistics*, 113–125.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, *80*(3), 327–339.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558.

Hult, S., Kalaja, M., Lassila, O., & Lehtisalo, T. (1990). Hyperreader—An interactive course in reading comprehension. *System*, *18*(2), 189–198. http://doi.org/10.1016/0346-251X(90)90053-8

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.

Hwang, K., & Nation, P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language*, *6*(1), 323.

Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language*, *20*(1), 1.

Iwashita, N. (2012). Cross-linguistic influence as a factor in the written and oral production of school age learners of Japanese in Australia. *Australian Review of Applied Linguistics*, *35*(3). Retrieved from http://www.nla.gov.au/openpublish/index.php/aral/article/view/2580

Jacobs, G. M. (1991). *Second language reading recall as a function of vocabulary glossing and selected other variables*. University of Hawaii.

Jacobs, G. M., Dufon, P., & Hong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, *17*(1), 19–28.

Jatowt, A., & Tanaka, K. (2012). Is wikipedia too difficult?: comparative analysis of readability of wikipedia, simple wikipedia and britannica. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 2607–2610). Retrieved from http://dl.acm.org/citation.cfm?id=2398703

Ji, Y. (1998). Sandwich stories for Chinese children. *IATEFL Newsletter*, *142*, 9–10.

Ji, Y. (1999). Communicative Language-Teaching through Sandwich Stories for EFL Children in China. *TESL Canada Journal*, *17*(1), 103–113.

Ji, Y. (1999). Sandwich stories as a bridge to authentic materials: A developmental approach to teaching EFL reading to young learners in China. In *International Conference on Second Language Teaching: Reading, Writing and Discourse, Hong Kong*.

Ji, Y. (2002). English through Chinese: experimenting with sandwich stories. *English Today*, *18*(01), 37–46.

Johnson, G. (2006, January 3). The Nitpicking of the Masses vs. the Authority of the Experts. *The New York Times*. Retrieved from http://www.nytimes.com/2006/01/03/science/03comm.html

Joyce, E. E. (1997). Which Words Should Be Glossed in L2 Reading Materials? A Study of First, Second, and Third Semester French Students' Recall. *Pennsylvania Language Forum*. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED427508

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. Wiley. com. Retrieved from http://books.google.co.nz/books?hl=en&lr=&id=M2Rb9ZtFxccC&oi=fnd&pg=PR12&dq=Gamification+of+learning+and+instruction&ots=JvPf178I7K&sig=YunrQkNoIKgFFY3S_Iq9hw5J-G4

Kelly, L. G. (1969). *25 Centuries of Language Teaching; an Inquiry into the Science, Art, and Development of Language Teaching Methodology, 500 B.C.-1969*. Rowley, Mass: Newbury House Publishers.

Kidwell, P., Lebanon, G., & Collins-Thompson, K. (2011). Statistical estimation of word acquisition with application to readability prediction. *Journal of the American Statistical Association*, *106*(493), 21–30.

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as a foreign language. *TESOL Quarterly*, *40*(2), 341–373.

Kincaid, J. P. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.

Kleinmann, H. H. (1987). The Effect of Computer-Assisted Instruction on ESL Reading Achievement. *The Modern Language Journal*, *71*(3), 267–276. http://doi.org/10.2307/326446

Klimpfinger, T. (2007). "Mind you sometimes you have to mix"–The role of code-switching in English as a lingua franca. *Vienna English Working Papers*, *16*, 36–61.

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal*, *78*(3), 285–299.

Kolers, P. A. (1966). Reading and Talking Bilingually. *The American Journal of Psychology*, *79*(3), 357–376.

Kondrak, G. (2000). A new algorithm for the alignment of phonetic sequences. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* (pp. 288–295). Retrieved from http://dl.acm.org/citation.cfm?id=974343

Kondrak, G. (2001). Identifying cognates by phonetic and semantic similarity. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1–8). Retrieved from http://dl.acm.org/citation.cfm?id=1073350

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *73*(4), 440–464.

Krashen, S. (2004). The case for narrow reading. *Language Magazine*, *3*(5), 17–19.

Krashen, S. D. (1981). Krashen, Stephen D. 1981. The case for narrow reading. TESOL Newsletter 15(6):23. *TESOL Newsletter*, *15*(6).

Krashen, S. D. (2011). *Free voluntary reading*. ABC-CLIO. Retrieved from https://books.google.co.nz/books?hl=en&lr=&id=-JVAz96iqnUC&oi=fnd&pg=PP2&dq=Krashen+%22narrow+reading%22+Free+Voluntary+reading&ots=i9lCg3sQSf&sig=b4GkFkv77nm1FSXyyWvdgOVuN0U

Kulkarni, A., Heilman, M., Eskenazi, M., & Callan, J. (2008). Word sense disambiguation for vocabulary learning. In *Intelligent Tutoring Systems* (pp. 500–509). Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-69132-7_53

Kweon, S. O., & Kim, H. R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Korea*, *20*(2).

Kyongho, H., & Nation, P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language*, *6*(1), 323–335.

Lai, F.-K. (1993). The effect of a summer reading course on reading and writing skills. *System*, *21*(1), 87–100.

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998–1009. http://doi.org/10.1002/ejsp.674

Lapp, D., & Fisher, D. (2009). It's all about the book: Motivating teens to read. *Journal of Adolescent & Adult Literacy*, 556–561.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *59*(4), 567–587.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 1.

Leffa, V. J. (1992). Making foreign language texts comprehensible for beginners: An experiment with an electronic glossary. *System*, *20*(1), 63–73.

Leloup, J. W., & Ponterio, R. (2005). Vocabulary Support for Independent Online Reading. *Language, Learning & Technology*, *9*(2), 3–8.

Lennon, C., & Burdick, H. (2004). The lexile framework as an approach for reading measurement and success.

Leung, C. Y. (2002). Extensive reading and language learning: A diary study of a beginning learner of Japanese. *Reading in a Foreign Language*, *14*(1), 66–81.

Li, J. (2010). Learning vocabulary via computer-assisted scaffolding for text processing. *Computer Assisted Language Learning*, *23*(3), 253–275.

Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76–80.

Liu, C. L., Wang, C. H., Gao, Z. M., & Huang, S. M. (2005). Applications of lexical information for algorithmically composing multiple-choice cloze items. In *Proceedings of the second workshop on Building Educational Applications Using NLP* (pp. 1–8).

Lively, B. A., & Pressey, S. L. (1923). A method for measuring the vocabulary burden of textbooks. *Educational Administration and Supervision*, *9*(389-398), 73.

Lomicka, L. (1998). To gloss or not to gloss: An investigation of reading comprehension online. *Language Learning and Technology*, *1*(2), 41–50.

Lopez, A. (2008). Statistical machine translation. *ACM Computing Surveys (CSUR)*, *40*(3), 1–49.

Lucassen, T., Dijkstra, R., & Schraagen, J. M. (2012). Readability of Wikipedia. *First Monday*, *0*(0). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/3916

Luppescu, S., & Day, R. R. (1992). Reading, dictionaries, and vocabulary learning. *Language Learning*, *43*(2), 263–279.

Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 480–487. http://doi.org/10.1016/S0022-5371(71)80018-X

Martinez-Lage, A. (1997). Hypermedia technology for teaching reading. *Technology Enhanced Language Learning*, 121–163.

Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, *25*(1), 91–102. http://doi.org/10.1016/S0346-251X(96)00063-2

McClanahan, B., Williams, K., Kennedy, E., & Tate, S. (2012). A breakthrough for Josh: How use of an iPad facilitated reading improvement. *TechTrends*, *56*(3), 20–28.

McKool, S. S. (2007). Factors that influence the decision to read: An investigation of fifth grade students' out-of-school reading habits. *Reading Improvement*, *44*(3), 111.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, *12*(8), 639–646.

McPeake, C., Chen, H., Whalen, R. E., & Vezina, D. (2004, February). System and methodology for name searches.

McQuillan, J., & Krashen, S. D. (2008). Commentary: Can free reading take you all the way? A response to Cobb (2007). *About Language Learning & Technology*, *6*(27), 104.

Melluish, T., & Smith, F. K. (1961). *Teach Yourself Greek*. English University Press.

Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V., … others. (2010). Enhancing Authentic Web Pages for Language Learners. In *Fifth Workshop on Innovative Use of NLP for Building Educational Applications* (p. 10).

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., … Orwant, J. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Mills, C. B., & Weldon, L. J. (1987). Reading text from computer screens. *ACM Comput. Surv.*, *19*(4), 329–357. http://doi.org/10.1145/45075.46162

Milone, M. (2012). The Development of ATOS. Retrieved from http://doc.renlearn.com/KMnet/R004250827GJ11C4.pdf

Miltsakaki, E. (2009). Matching readers' preferences and reading skills with appropriate web texts. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session* (pp. 49–52).

Miltsakaki, E., & Troutt, A. (2007). Read-X: Automatic evaluation of reading difficulty of web text. In *Proceedings of E-Learn*.

Miltsakaki, E., & Troutt, A. (2008). Real-time web text classification and analysis of reading difficulty. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 89–97).

Mitkov, R., Ha, L. A., & Karamanis, N. (2006). A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*, *12*(02), 177–194.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students. *Language Education & Technology*, *39*(1), 1–20.

Miyazaki, Y., & Norizuki, K. (2008). Developing a computerized readability estimation program with a web-searching function to match text difficulty with individual learners' reading ability. *Proceedings of WorldCALL*.

Montaner, M., López, B., & De La Rosa, J. L. (2003). A taxonomy of recommender agents on the internet. *Artificial Intelligence Review*, *19*(4), 285–330.

Mori, S. (1999). The role of motivation in the amount of reading. *Temple University Japan Working Papers in Applied Linguistics*, *14*, 51–68.

Mulloni, A., & Pekar, V. (2006). Automatic detection of orthographic cues for cognate recognition. *Proceedings of LREC'06, 2387*, *2390*.

Mulloni, A., Pekar, V., & Blagoev, R. M. D. (2007). Semantic evidence for automatic identification of cognates. In *Proceedings of the RANLP'2007 workshop: Acquisition and management of multilingual lexicons* (pp. 49–54).

Murphy, G. (1929). *An Historical Introduction to Modern Psychology*. Taylor & Francis.

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

Nagy, W. E., & Anderson, R. C. (1984). How Many Words Are There in Printed School English? *Reading Research Quarterly*, *19*(3), 304–330.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning Word Meanings from Context during Normal Reading. *American Educational Research Journal*, *24*(2), 237–270.

Nakanishi, T. (2014). A Meta-Analysis of Extensive Reading Research. *TESOL Quarterly*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/tesq.157/full

Nakov, S., Nakov, P., & Paskaleva, E. (2007). Cognate or false friend? Ask the Web. In *Proceedings of the RANLP'2007 workshop: Acquisition and management of multilingual lexicons* (pp. 55–62). Retrieved from http://www.nakov.com/wp-content/uploads/2007/09/nakov-nakov-paskaleva-ranlp-2007-workshop-false-friends-and-cognates.pdf

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge Univ Pr.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–82.

Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. Taylor & Francis.

Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Nelson, C. (1995). *The Effects of Marginal Glossing and Frequency of Appearance on Learning Spanish Vocabulary*. Brigham Young University. Dept. of Language Acquisition.

Nilsson, K., & Borin, L. (2002). Living off the land: The Web as a source of practice texts for learners of less prevalent languages. In *Proceedings of LREC 2002*.

Nishino, T. (2007). Beginning to read extensively: A case study with Mako and Fumi. *Reading in a Foreign Language*, *19*(2), 2.

Nishizawa, H., Yoshioka, T., & Fukada, M. (2009). The impact of a 4-year extensive reading program. In *JALT2009 Conference Proceedings*. Tokyo.

Nuttall, C. (1996). *Teaching reading skills in a foreign language*. ERIC. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED399531

Och, F. J. (2005). Statistical machine translation: Foundations and recent advances. *Tutorial at MT Summit X, Phuket, Thailand*.

O'Donnell, M. E. (2009). Finding middle ground in second language reading: pedagogic modifications that increase comprehensibility and vocabulary acquisition while preserving authentic text features. *The Modern Language Journal*, *93*(4), 512–533.

Ogden, C. K. (1944). *Basic English: a general introduction with rules and grammar* (Vol. 29). K. Paul, Trench, Trubner.

Ohtake, H., Kawamoto, T., Takekoshi, M., Kunimura, M., Morren, B., Takeuchi, H., … Kaneko, S. (2003). Development of a genre-specific electronic dictionary and automatic gloss-embedding system. In *Asialex* (pp. 445–449).

Okazaki, H., Nitta, H., & Hashimoto, S. 'ichi. (2008). An Internet-based personal reading program: A program for Japanese EFL learners to overcome individual weaknesses. In *Proceedings of WorldCALL*. Fukuoka, Japan. Retrieved from 261-263

Okuyama, Y., & Igarashi, H. (2007). Think-Aloud Protocol on Dictionary Use by Advanced Learners of Japanese. *The JALT CALL Journal, 3*(1/2), 45–58.

Ono, L., Day, R., & Harsch, K. (2004). Tips for reading extensively. In *English Teaching Forum* (Vol. 42, pp. 12–19).

Ott, N., & Meurers, D. (2010). Information Retrieval for Education: Making Search Engines Language Aware.

Oxford English Dictionary. (1928). Oxford: Oxford University Press.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the web. Retrieved from http://ilpubs.stanford.edu:8090/422

Pande, M. (2011, April). Wikipedia Editors Study. Wikimedia Foundation. Retrieved from http://upload.wikimedia.org/wikipedia/commons/5/51/Editor_Survey_Report_April_2011.pdf

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language, 22*(1), 1.

Petersen, S. E., & Ostendorf, M. (2007). Text simplification for language learners: a corpus analysis. In *the Proceedings of the Speech and Language Technology for Education Workshop* (pp. 69–72).

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1–28.

Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal, 51*(2), 73–75.

Pino, J., Heilman, M., & Eskenazi, M. (2008). A selection strategy to improve cloze question quality. In *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains. 9th International Conference on Intelligent Tutoring Systems, Montreal, Canada* (pp. 22–32).

Pino-Silva, J. (2006). Extensive reading through the internet: Is it worth the while? *Reading, 6*(1).

Pisit, P. (2005). Utilization of of extensive reading to enhance reading comprehension, vocabulary recognition and reading attitudes: a case study of the first-year english majors at Nakhon Ratchasima Rajabhat University.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring Second Language Vocabulary through Reading: A Replication of the Clockwork Orange Study Using Second Language Acquirers. *Reading in a Foreign Language, 5*(2), 271–75.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*(2), 221–243.

Poel, K., & Swanepoel, P. (2003). Theoretical and methodological pluralism in designing effective lexical support for CALL. *Computer Assisted Language Learning*, *16*(2-3), 173–211.

Port-Fox, J. (1982). The bilingual reading method. *English Language Research Journal*, *3*, 66–77.

Powell, S. (2005). Extensive reading and its role in Japanese high schools. *Reading*, *5*(2).

Prichard, C. (2008). Evaluating L2 readers' vocabulary strategies and dictionary use. *Reading in a Foreign Language*, *20*(2), 2.

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *57*, 155–199.

Rankin, J. (2005). Easy reader: A case study of embedded extensive reading in intermediate German L2. *Die Unterrichtspraxis/Teaching German*, 125–134.

Renandya, W. A. (2007). The power of extensive reading. *RELC Journal*, *38*(2), 133.

Riding, J. D. (2008). Statistical glossing, language independent analysis in bible translation. *Translating and the Computer*, *30*.

Robb, T. (2009). The Reader Quiz Module for Extensive Reading. In M. Thomas (Ed.), *Selected proceedings of the thirteenth annual JALT CALL SIG conference 2008* (pp. 109–116). Nagoya, Japan: Tokyo: Japan Association for Language Teaching.

Robb, T., & Kano, M. (2013). Effective extensive reading outside the classroom: A large-scale experiment. *Reading in a Foreign Language*, *25*(2), 2.

Robb, T. N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, *5*(2), 239–251.

Rochon, P. A., Bero, L. A., Bay, A. M., Gold, J. L., Dergal, J. M., Binns, M. A., … Gurwitz, J. H. (2002). Comparison of review articles published in peer-reviewed and throwaway journals. *JAMA: The Journal of the American Medical Association*, *287*(21), 2853–2856.

Rodrigo, V., Greenberg, D., & Segal, D. (2014). Changes in reading habits by low literate adults through extensive reading. *Reading in a Foreign Language*, *26*(1), 1.

Rodriguez, T. A. (2001). From the known to the unknown: Using cognates to teach English to Spanish-speaking literates. *The Reading Teacher*, *54*(8), 744–746.

Rosszell, R. (2006). The role of extensive reading in second language vocabulary acquisition. In *JALT2005 Conference Proceedings* (pp. 394–407).

Rott, S. (1999). The Effect of Exposure Frequency on Intermediate Language Learners. *Studies in Second Language Acquisition*, *21*(4), 589–619.

Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research*, *6*(3), 183.

Sakar, A., & Ercetin, G. (2005). Effectiveness of hypermedia annotations for foreign language reading. *Journal of Computer Assisted Learning*, *21*(1), 28–38.

Samimy, K. K. (1994). Teaching Japanese: Consideration of learners' affective variables. *Theory into Practice*, *33*(1), 29–33.

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher*, *32*(4), 403–408.

Samuels, S. J. (1988). Decoding and Automaticity: Helping Poor Readers Become Automatic at Word Recognition. *The Reading Teacher*, *41*(8), 756–760.

Samuels, S. J. (2006). Toward a model of reading fluency. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 22–46). International Reading Association.

Sanger, L. (2005). The early history of Nupedia and Wikipedia: a memoir. *Open Sources*, *2*, 307–338.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*(2), 72–78.

Sato, S., Matsuyoshi, S., & Kondoh, Y. (2008). Automatic assessment of Japanese text readability based on a textbook corpus. *6th LREC*.

Schaefer, H. H. (1961). A Vocabulary Program Using "Language Redundancy".

Schendl, H. (1996). Text types and code-switching in medieval and Early Modern English. *Vienna English Working Papers*, *5*(1&2), 50.

Schendl, H. (1997). "To London fro Kent/Sunt predia depopu-lantes": Code-switching and medieval Eng-lish macaronic poems. *Vienna English Working Papers*, *6*(1), 52.

Schendl, H. (2002). Mixed-language texts as data and evidence in English historical linguistics. *Studies in the History of the English Language: A Millennial Perspective*, 51.

Schmidt, K. (2007). Five factors to consider in implementing a university Extensive Reading program. *The Language Teacher*, *31*(5), 11–14.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Schneyer, J. W. (1965). Use of the cloze procedure for improving reading comprehension. *The Reading Teacher*, *19*(3), 174–179.

Sebba, M., Mahootian, S., & Jonsson, C. (2011). *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse*. Routledge. Retrieved from http://www.routledge.com/books/details/9780415879460/

Senoo, Y., & Yonemoto, K. (2014). Vocabulary Learning Through Extensive Reading: A Case Study. *Canadian Journal of Applied Linguistics/Revue Canadienne de Linguistique Appliquée*, *17*(2), 1–22.

Sherman, L. A. (1893). *Analytics of literature: A manual for the objective study of English prose and poetry*. Ginn.

Sheu, S. P. . (2004). The Effects of Extensive Reading on Learners' Reading Ability Development. *Journal of National Taipei Teachers College*, *17*(2), 213–228.

Siddharthan, A. (2003). An architecture for a text simplification system. In *Language Engineering Conference, 2002. Proceedings* (pp. 64–71).

Siddharthan, A. (2006). Syntactic simplification and text cohesion. *Research on Language & Computation*, *4*(1), 77–109.

Silver, M. (1997). *The effect of the Book of Mormon Diglot Reader: A study of the vocabulary acquisition, reading comprehension, and the reduction of negative affective variables in missionaries*. Brigham Young University. Dept. of Linguistics.

Simard, M., Foster, G. F., & Isabelle, P. (1993). Using cognates to align sentences in bilingual corpora. In *Proceedings of the 1993 conference of the Centre for Advanced Studies on Collaborative research: distributed computing-Volume 2* (pp. 1071–1082). Retrieved from http://dl.acm.org/citation.cfm?id=962411

Simon, S. J. (2000). The impact of culture and gender on web sites: an empirical study. *ACM SIGMIS Database*, *32*(1), 18–37.

Sleight, G. H. (1982). *Adaptation of the Burling Reading Model for Teaching Reading to Beginning ESL/EFL Students: A Rationale Sample Text, and Case Study*. Brigham Young University. Dept. of Linguistics.

Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal*, *53*(7), 410–413.

Stanovich, K. E. (1992). The psycology of reading: Evolutionary and revolutionaty developments. In W. Grabe (Ed.), *Annual Review of Applied Linguistics* (12th ed., pp. 3–30). Cambridge: Cambridge University Press.

Stewart, S. M. (2012). *Reading in a Technological World: Comparing the iPad to Print*. Bowling Green State University. Retrieved from http://etd.ohiolink.edu/view.cgi?acc_num=bgsu1335287048

Stoddard, G. D. (1929). An experiment in verbal learning. *Journal of Educational Psychology*, *20*(6), 452.

Streiter, O., Knapp, J., & Voltmer, L. (2004). Gymn@ zilla: Language learning with the Internet. In *Proceedings of TALC 2004 The sixth Teaching and Language Corpora conference*.

Stross, R. (2006, March 12). Anonymous Source Is Not the Same as Open Source. *The New York Times*. Retrieved from http://www.nytimes.com/2006/03/12/business/yourmoney/12digi.html

Sumita, E., Sugaya, F., & Yamamoto, S. (2005). Measuring non-native speakers' proficiency of English by using a test with automatically-generated fill-in-

the-blank questions. In *Proceedings of the second workshop on Building Educational Applications Using NLP* (pp. 61–68).

Sun, Y. C. (2003). Extensive reading online: an overview and evaluation. *Journal of Computer Assisted Learning*, *19*(4), 438–446.

Suppes, P. (1979). Current trends in computer-assisted instruction. *Advances in Computers*, *18*, 173–229.

Susser, B., & Robb, T. N. (1990). EFL extensive reading instruction: Research and procedure. *Jalt Journal*, *12*(2), 161–185.

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal*, *25*(2), 34.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*(3), 261.

Tanaka, H., & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficacy of extensive reading. *Reading*, *7*(1).

Tateishi, Y., Ono, Y., & Yamada, H. (1988a). A computer readability formula of Japanese texts for machine scoring. In *Proceedings of the 12th Conference on Computational Linguistics* (Vol. 2, pp. 649–654).

Tateishi, Y., Ono, Y., & Yamada, H. (1988b). Derivation of Readability Formula of Japanese Texts. *IPSJ SIG-DPHI*.

Taylor, A. M. (2009). CALL-Based versus Paper-Based Glosses: Is There a Difference in Reading Comprehension?. *CALICO Journal*, *27*(1), 14.

Taylor, A. M. (2010). Glossing is sometimes a distraction: Comments on Cheng and Good (2009). *Reading in a Foreign Language*, *22*(2), 353–354.

The Extensive Reading Foundation's Guide to Extensive Reading. (2011). The Extensive Reading Foundation. Retrieved from http://erfoundation.org/ERF_Guide.pdf

Thorndike, E. L. (1921). *The teacher's word book* (Vol. 134). Teachers College, Columbia University New York. Retrieved from http://89.253.247.156/_files/product/pdf/59/581265.pdf

Tseng, M. (2007). An Investigation of EFL Learners' Online Reading Skills. *Journal of Nanya Institute of Technology*, *27*, 111–127.

Tseng, M. (2008a). Comparing EFL learners' reading comprehension between hypertext and printed text. *CALL-EJ Online*, *9*(2), 9–2.

Tseng, M. (2008b). The difficulties that EFL learners have with reading text on the web. *The Internet TESL Journal*, *14*(2).

Tseng, M. (2010). Factors That Influence Online Reading: An Investigation into EFL Students' Perceptions. *Reading*, *10*(1).

Uitdenbogerd, A. (2003). Using the web as a source of graded reading material for language acquisition. *Advances in Web-Based Learning-ICWL 2003*, 423–432.

Uitdenbogerd, A. L. (2006). Web Readability and Computer-Assisted Language Learning. In *Australasian Language Technology Workshop* (p. 99).

Urano, K. (2000). Lexical simplification and elaboration: Sentence comprehension and incidental vocabulary acquisition. *Unpublished MA in ESL Thesis. Honolulu, HI: University of Hawai'i, Department of Second Language Studies*.

Van Meer, E. (2003). PLATO: From computer-based education to corporate social responsibility. *Iterations: An Interdisciplinary Journal of Software History*, *5*. Retrieved from http://www.cbi.umn.edu/iterations/vanmeer.pdf

Vogel, M., & Washburne, C. (1928). An objective method of determining grade placement of children's reading material. *The Elementary School Journal*, *28*(5), 373–381.

Wakefield Jr, J. A., Bradley, P. E., Yom, B. H., & Doughtie, E. B. (1975). Language switching and constituent structure. *Language and Speech*, *18*(1), 14.

Waring, R. (2006). Why Extensive Reading should be an indispensable part of all language programmes. *The Language Teacher*, *30*(7), 44–47.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader. *Reading in a Foreign Language*, *15*(2), 130–163.

Watanabe, Y. (1997). Input, intake, and retention. *Studies in Second Language Acquisition*, *19*(03), 287–307.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46.

Webb, S. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *65*(3), 441–470.

Webb, S. (2010). Using glossaries to increase the lexical coverage of television programs. *Reading in a Foreign Language*, *22*(1), 201–221.

Weible, D. M. (1980). Teaching Reading Skills through Linguistic Redundancy. *Foreign Language Annals*, *13*(6), 487–493. http://doi.org/10.1111/j.1944-9720.1980.tb01388.x

Weible, D. M. (2010, August 13). Article entitled "Teaching Reading Skills through Linguistic Redundancy."

West, M. (1943). new method English dictionary. Retrieved from http://agris.fao.org/agris-search/search/display.do?f=2013/US/US2013016920001692.xml;US201300169288

Wible, D., Kuo, C. H., yi Chien, F., & Wang, C. C. (2000). Adjusting corpus searches for learners' level: Filtering results for frequency. *Presentation at TALC2000, Graz, Austria, July*.

Wood, B. (2013, July 27). Students choosing books below grade level; should parents be worried? | Deseret News. Retrieved July 30, 2013, from http://www.deseretnews.com/article/865583736/Students-choosing-books-below-grade-level-should-parents-be-worried.html?pg=all

Wood, M. C., & Dariyanani, J. (2010, October 19). Method and system for teaching a foreign language. Retrieved from http://www.freepatentsonline.com/7818164.html

Wood, P. (2010). QuickAssist Extensive Reading for Learners of German Using CALL Technologies. Retrieved from http://uwspace.uwaterloo.ca/handle/10012/5680

Wozniak, P. (2007). SuperMemo 2004. *TESL EJ*, *10*(4). Retrieved from http://www.cc.kyoto-su.ac.jp/information/tesl-ej/ej40/m1.html

Wright, L. (1998). Mixed-language business writing: five hundred years of code-switching. *Language Change: Advances in Historical Sociolinguistics*, 99.

Wright, L. (2002). Code-intermediate phenomena in medieval mixed-language business texts* 1. *Language Sciences*, *24*(3-4), 471–489.

Wu, Q., Burges, C. J., Svore, K. M., & Gao, J. (2010). Adapting boosting for information retrieval measures. *Information Retrieval*, *13*(3), 254–270.

Wu, S., Franken, M., & Witten, I. H. (2010). Supporting collocation learning with a digital library. *Computer Assisted Language Learning*, *23*(1), 87–110.

Yamakami, T. (2013). Gamification literacy: Emerging needs for identifying bad gamification. In *Multimedia and Ubiquitous Engineering* (pp. 395–403). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-94-007-6738-6_49

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning and Technology*, *13*(2), 48–67.

Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, *44*(2), 189–219.

Yen, T. T. N. (2012). The Effects of a Speed Reading Course and Speed Transfer to Other Types of Texts. *RELC Journal*, *43*(1), 23–37.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, *10*(3), 85–101.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *57*(4), 541–573.

Appendix A. Publications, presentations and awards

The following papers have been presented at conferences during the course of this investigation:

A.1 Publications and presentations

- User Evaluation of Extensive Reading iCALL System with Spanish Learners. CALICO Conference, Indiana, USA (12-16 June, 2012).
- *NLP Assisted ER with FERN*. CALICO iCALL Workshop, Indiana, USA (12-16 June, 2012).
- *Computer-Assisted Learning with FERN.* Te Toi o Matariki Postgraduate Conference, The University of Waikato (17-18 October, 2011)
- Flax-Reader – iCALL Web-Application for Individualised L2 Extensive Reading and Vocabulary Study. CALICO Conference, Victoria, Canada (19-21 May, 2011).
- *Linking Digital Libraries to Courses.* International Conference on Computer Supported Education (CSEDU), Noordwijkerhout, The Netherlands (6-8 May, 2011).
- *Automatic Glossing with Google Translate.* New Zealand Computer Science Research Students Conference (NZCSRSC), Massey University (18-21 April, 2011).
- *Leisurely Language Learning.* Thesis in Three Final, University of Waikato (27 October 2010).
- *Automatic Adaption of Dynamic Second Language Reading Texts*, New Zealand Computer Science Research Students Conference (NZCSRSC), Victoria University (12-15 April 2010).

A.2 Awards

The following awards have been received during the course of the investigation

- *Runner-up and people's choice award.* Te Toi o Matariki Postgraduate Conference (2011).

- *Runner-up and people's choice award.* The University of Waikato Thesis in Three (2010).
- *Best paper award.* NZCSRSC Conference (2010).
- TEC Top Achiever Doctoral Scholarship (2009).
- The University of Waikato Doctoral Scholarship (2008).
- BuildIT Doctoral Scholarship (2008).

Appendix B. Extracting proper noun translations from Wikipedia

Encyclopaedias are better sources of proper noun information than dictionaries. The multi-lingual encyclopaedia Wikipedia provides links between articles in different language editions that can act a source of proper name translations. In the case of *japonés*, there is a disambiguation page that presents three articles. The first, *Idioma japonés* is an article linked with the English article titled *Japanese language.* The second, *Etnia japonesa* is linked to an English article titled *Japanese people*. The third *Japonés* (note the capitalization) is a Spanish only article about an episode of a TV series in Chile with the title *Japonés.* If *japonés* occurs without capitalisation, this final article can be ignored as irrelevant because it has a capital J.

If the term *Japonés* appeared it is likely to refer to the TV series. Because this article is Spanish only there are a couple of options to providing a gloss for a low level learner. The first is to provide the opening sentence of the Spanish article with access to glosses and dictionary entries for the sentence. A better option is to first check for other general information in the Spanish Wikipedia article for which translations can be obtained such as category information. In the case of the *Japonés* article no general categories are assigned, however, the page does include a *Ficha de episodio de televisión* known in the English Wikipedia as a *TV episode* infobox. This information can be provided as a concise definition for the term Japonés, possible with a link to more info.

In each of these three articles for japonés, a longer definitions, such as the first sentence in the available L1 or L2 Wikipedia article could be provided if the user clicks a link requesting more information.

Translations for person and place names are often the same as the source word. Future versions of FERN could look display category information from Wikipedia (if available). For ambiguous words, this would be a generic category such as human name, place name or organisation name. For unambiguous words, FERN could display more specific category information such as *political party* or *Software Company*. Alternatively the first sentence from the article could be displayed, however, this may disrupt the reading too much.

Appendix C. Olympic Games, full text

C.1 English text

The full text for the English simple Wikipedia article entitles *The Olympic Games.*

Olympic Games

The Olympic Games is a sporting event that takes place in a different city every four years. It includes many of the best athletes from all over the world. The first modern Games were held in Athens in 1896. The Motto of the Olympic Games is "Citius, Altius, Fortius," which means, in English, "Faster, Higher, Stronger."

There are separate games for the summer and winter. The Summer Olympic Games has many sports, including athletics (track and field events such as sprint, marathon, and discus), swimming, gymnastics, cycling, boxing, soccer, weightlifting, wrestling, basketball, and many other sports. The Winter Olympic Games also have many sports, but they are played in the winter. These include curling, skiing, speed skating, luge, bobsled, figure skating, and ice hockey.

Athletes win medals made of gold, silver, and bronze for coming in first, second, and third place in the Olympics.

The most recent Summer Olympics were held in 2008 in Beijing, China; before that, they were held in Athens, Greece in the year 2004. The next ones will be held in London, Great Britain, in 2012.

The most recent Winter Olympics were held in 2006 in Torino, Italy; before that, they were held in Salt Lake City, United States, in 2002. The next ones will be held in Vancouver, Canada in 2010.

The Olympic Rings

The Olympic Rings became the symbol of the Olympic Games in 1913. The five rings represent the fact that countries from all five continents of the world take part in the games. Each ring is a different colour, and at least one of the five colours (Blue, Yellow, Black, Green, and Red) can be found on the flag of every country in the world. The rings do not represent particular continents.

Olympic Ceremony

In the opening ceremony, the stadium is opened. Special events happen like songs, performances, dances, etc. The Olympic theme song is played, written by John Williams. The torch is brought in after being carried around the world, from Olympia, where the original Olympics were held. Doves are released. Then, every athlete from each country comes into the stadium in alphabetical order, except for Greece, which comes first in as they created the games, and the country where the games are held, which goes last.

C.2 Cognates, numbers and proper nouns translated

Olympic Games

The Olímpicos Games is a sporting evento that takes place in a diferente city every four years. It incluye muchas of the best atletas from all over the world. The first modernos Games were held in Atenas in 1896. The Motto of the Olímpicos Games is "Citius, Altius, Fortius," which means, in Inglés, "Faster, Higher, Stronger."

There are separados games for the summer and winter. The Summer Olímpicos Games has muchos sports, incluyendo atletismo (track and field events such as sprint, maratón, and discusión), swimming, gimnasia, ciclismo, boxeo, fútbol, weightlifting, wrestling, basketball, and muchos otros sports. The Winter Olímpicos Games also have muchos sports, but they are played in the winter. These incluyen curling, esquí, velocidad skating, luge, bobsled, artístico skating, and ice hockey.

Atletas win medallas made of gold, silver, and bronce for coming in first, second, and third place in the Olímpicos.

The más reciente Summer Olímpicos were held in 2008 in Beijing, China; before that, they were held in Atenas, Grecia in the year 2004. The next ones will be held in Londres, Gran Bretaña, in 2012.

The más reciente Winter Olímpicos were held in 2006 in Torino, Italia; before that, they were held in Salt Lake City, Estados Unidos, in 2002. The next ones will be held in Vancouver, Canadá in 2010.

The Olympic rings

The Olímpicos Rings became the símbolo of the Olímpicos Games in 1913. The five rings representan the fact that countries from all five continentes of the world take parte in the games. Each ring is a diferente color, and at least uno of the five colores (Blue, Yellow, Black, Green, and Red) can be found on the flag of every country in the world. The rings do not representan particular continentes.

Olympic ceremony

In the opening ceremonia, the estadio is opened. Eventos especiales happen like songs, performances, dances, etc. The Olímpicos theme song is played, written by John Williams. The antorcha is brought in after being carried around the world, from Olimpia, where the original Olímpicos were held. Doves are released. Then, every atletas from each country comes into the estadio in orden alfabético, excepción for Grecia, which comes first in as they crearon the games, and the country where the games are held, which goes último.

C.3 Keywords only

Translates 10 noun word families that appear at least 3 times in the text.

Olympic games

The Juegos Olímpicos is a deportes event that takes place in a different city every four years. It includes many of the best atletas from all over the mundo. The first modern Juegos were held in Athens in 1896. The Motto of the Juegos Olímpicos is "Citius, Altius, Fortius," which means, in English, "Faster, Higher, Stronger."

There are separate juegos for the verano and invierno. The Verano Juegos Olímpicos has many deportes, including atletismo (track and field events such as sprint, marathon, and discus), swimming, gymnastics, cycling, boxing, soccer, weightlifting, wrestling, basketball, and many other deportes. The Inviero Juegos Olímpicos also have many deportes, but they are played in the invierno. These include curling, skiing, speed skating, luge, bobsled, figure skating, and ice hockey.

Atletas win medals made of gold, silver, and bronze for coming in first, second and third place in the Olímpicos.

The most recent Verano Olímpicos were held in 2008 in Beijing, China; before that, they were held in Athens, Greece in the year 2004. The next ones will be held in London, Great Britain, in 2012.

The most recent Invierno Olímpicos were held in 2006 in Torino, Italy; before that, they were held in Salt Lake City, United States, in 2002. The next ones will be held in Vancouver, Canada in 2010.

The Olympic rings

The Olímpicos Anillos became the symbol of the Juegos Olímpicos in 1913. The five anillos represent the fact that países from all five continents of the mundo take part in the juegos. Each anillos is a different color, and at least one of the five colors (Blue, Yellow, Black, Green, and Red) can be found on the flag of every país in the mundo. The anillos do not represent particular continents.

Opening ceremony

In the opening ceremony, the stadium is opened. Special events happen like songs, performances, dances, etc. The Olímpic theme song is played, written by John Williams. The torch is brought in after being carried around the mundo, from Olimpia, where the original Olímpicos were held. Doves are released. Then, every atletas from each país comes into the stadium in alphabetical order, except for Greece, which comes first in as they created the juegos, and the país where the juegos are held, which goes last.

C.4 Translate keywords, easy nouns

This text contains translation of high frequency and cognate nouns as well as the keywords.

Olympic games

The Juegos Olímpicos is a deportes event that takes place in a different city every cuatro years. It includes many of the best atletas from all over the mundo. The primero modern Juegos were held in Atenas in 1896. The Motto of the Juegos Olímpicos is "Citius, Altius, Fortius," which means, in Inglés, "Faster, Higher, Stronger."

There are separate juegos for the verano and invierno. The Verano Juegos Olímpicos has many deportes, including atletismo (track and field events such as sprint, maratón, and discusión), swimming, gimnasia, ciclismo, boxeo, fútbol, weightlifting, wrestling, basketball, and many other deportes. The Juegos Olímpicos de Inviero also have many deportes, but they are played in the invierno. These include curling, esquí, speed skating, luge, bobsled, figure skating, and ice hockey.

Los atletas win medallas made of gold, silver, and bronce for coming in primero, segundo, and tecero place in the Olímpicos.

The most recent Verano Olímpicos were held in 2008 in Beijing, China; before that, they were held in Atenas, Grecia in the year 2004. The next ones will be held in Londres, Gran Bretaña, in 2012.

The most recent Invierno Olímpicos were held in 2006 in Torino, Italia; before that, they were held in Salt Lake City, United States, in 2002. The next ones will be held in Vancouver, Canadá in 2010.

The Olympic rings

The Anillos Olímpicos became the símbolo of the Juegos Olímpicos in 1913. Los cinco anillos represent the fact that países from all cinco continentes of the mundo take parte in the juegos. Each anillos is a different color, and at least uno of the cinco colores (Blue, Yellow, Black, Green, and Red) can be found on the flag of every país in the mundo. The anillos do not represent particular continentes.

Opening ceremony

In the opening ceremony, the stadium is opened. Special eventos happen like songs, performances, dances, etc. The Olímpic theme song is played, written by John Williams. The torch is brought in after being carried around the mundo, from Olimpia, where the original Olímpicos were held. Doves are released. Then, every atletas from each país comes into the stadium in orden alfabético, excepción is Grecia, which comes primero in as they created the juegos, and the país where the juegos are held, which goes last.

C.5 Translate keywords, easy content words, articles

Keywords are translated and displayed in red. Other translated words are in blue.

Olympic games

Los Juegos Olímpicos is a deportes event that takes place in a diferente city every cuatro años. It incluye muchas of the best atletas from all over the mundo. El primero modernos Juegos were held in Atenas in 1896. The Motto of los Juegos Olímpicos is "Citius, Altius, Fortius," which means, in Inglés, "Faster, Higher, Stronger."

There are juegos separados for el verano and el invierno. Los Juegos Olímpicos de Verano has muchos deportes, incluyendo atletismo (track and field events such as sprint, maratón, and discusión), swimming, gimnasia, ciclismo, boxeo, fútbol, weightlifting, wrestling, basketball, and muchos otros deportes. Los Juegos Olímpicos de Inviero also have muchos deportes, but they are played in el invierno. These incluyen curling, esquí, skating de velocidad, luge, bobsled, skating artístico, and ice hockey.

Los atletas win medallas made of gold, silver, and bronce for coming in primero, segundo, and tecero place in the Olímpicos.

El más reciente Olímpicos de Verano were held in 2008 in Beijing, China; before that, they were held in Atenas, Grecia in the año 2004. The next ones will be held in Londres, Gran Bretaña, in 2012.

El más reciente Olímpicos de Invierno were held in 2006 in Torino, Italia; before that, they were held in Salt Lake City, Estados Unidos, in 2002. The next ones will be held in Vancouver, Canadá in 2010.

The Olympic rings

Los Anillos Olímpicos became el símbolo of los Juegos Olímpicos in 1913. Los cinco anillos representan the fact that países from all cinco continentes of the mundo take parte in the juegos. Each anillos is a color diferente, and at least uno of the cinco colores (Blue, Yellow, Black, Green, and Red) can be found on the flag of every país in el mundo. Los anillos no representan los continentes particular.

Opening ceremony

In the opening ceremonia, el estadio is opened. Eventos especiales happen like songs, performances, dances, etc. Los Olímpic theme song is played, written by John Williams. La antorcha is brought in after being carried around el mundo, from Olimpia, where el original Olímpicos were held. Doves are released. Then, every atletas from each país comes into el estadio in orden alfabético, excepción is Grecia, which comes in primero as they crearon los juegos, and el país where los juegos are held, which goes último.

C.6 Google translation with English words.

The Google translation of the article with non-cognate moderate and low frequency words replaced by English equivalents.

Olympic games

Los Juegos Olímpicos son un event deportivo que tienen lugar en una ciudad diferente cada cuatro años. Incluye muchas de las mejores atletas de todo el mundo. El primero Juegos modernos held en Atenas en 1896. El motto de los Juegos Olímpicos es "Citius, altius, fortius", que significa, en Inglés, "Más rápido, más alto, más fuerte".

Hay juegos separados para el verano y el invierno. Los Juegos Olímpicos de verano tienen muchos deportes, incluyendo atletismo (track y field, como Sprint, maratón, y la discusión), swimming, gimnasia, ciclismo, boxeo, fútbol, weightlifting, lucha libre, basketball, y muchos otros deportes. Los Juegos Olímpicos de Invierno también tienen muchos deportes, pero played en el invierno. Estos incluyen el curling, esquí, skating de velocidad, luge, trineo, skating artístico y hockey on ice.

Los atletas ganan medallas de oro, plata y bronce por venir en el primero, segundo y tercer lugar en los Juegos Olímpicos.

El más reciente Juegos Olímpicos de verano se held en 2008 en Beijing, China, antes de eso, que se held en Atenas, Grecia en el año 2004. El próximo que se held en Londres, Gran Bretaña, en 2012.

El más reciente Juegos Olímpicos de Invierno se held en 2006 en Torino, Italia, antes de eso, que se held en Salt Lake City, Estados Unidos, en 2002. El próximo que se held en Vancouver, Canadá en 2010.

The Olympic rings

Los Anillos Olímpicos se became en el símbolo de los Juegos Olímpicos de 1913. Los cinco anillos representan el fact de que los países de los cinco continentes del mundo tomar parte en los juegos. Cada anillo tiene un color diferente, y al menos uno de los cinco colores (Azul, Amarillo, Negro, Verde y Roja) se puede encontrar en el flag de todos los países del mundo. Los anillos no representan los continentes particular.