

## Towards a digital library for language learning

*Shaoqun Wu and Ian H. Witten*

University of Waikato

**Key words:** *E-learning, Computer aided language learning, Digital library*

### **Abstract:**

*Digital libraries have untapped potential for supporting language teaching and learning. Although the Internet at large is widely used for language education, it has critical disadvantages that can be overcome in a more controlled environment. This article describes a language learning digital library, and a new metadata set that characterizes linguistic features commonly taught in class as well as textual attributes used for selection of suitable exercise material. On the system is built a set of eight learning activities that together offer a classroom and self-study environment with a rich variety of interactive exercises, which are automatically generated from digital library content. The system has been evaluated by usability experts, language teachers, and students.*

## **1 Introduction**

The rise of computer-assisted language learning on the Internet has brought a new dimension and dynamic into language classes. The World Wide Web offers learners a wealth of language material and gives them opportunities to learn in many different ways. They can study by reading newspaper articles, listening to audio recordings and viewing video clips; undertake online learning exercises; or join courses. Media such as email, chat and blogs enable them to communicate with other learners and with speakers of the target language all over the world. When preparing lessons, teachers benefit from the panoply of resources that the web provides. Automated tools can be used to build practice exercises and design lessons. Teachers construct language learning tasks based on the Internet because the language is real and the topics are contemporary, which motivates learners.

Despite all these advantages, the Internet has many drawbacks for language study. Although it offers innumerable language resources, learners and teachers alike face the challenge of discovering usable material. Even the best search engines return an overwhelming amount of dross in response to any query, and locating suitable sources demands skill and judgment. When learners study on their own, it is very hard for them to locate material that matches their language ability. Sifting through search results to identify useful material saps their interest and motivation. Finally, students may accidentally encounter material with grossly unsuitable content.

Many teachers integrate the Internet into their language classes. Some ask students to read carefully-selected websites and undertake follow-up activities such as discussing or debating the issues they present, or comparing the language used on American and British sites. Teachers must work hard to find material that meets their teaching needs, matches the students' linguistic ability, has a congenial interface, and does not contain offensive or

threatening items. They must also evaluate how accurate the language is: whether there are grammatical or spelling mistakes. Last but not least, before every class they must ensure that the pages are still there and the content is intact [1].

Digital libraries, like traditional ones, can play a crucial role in education. Marchionini [2] identifies many advantages in using them for teaching and learning. As well as providing a safe and reliable educational environment, they have special advantages for language classes. Digital libraries are a great source of material that teachers can turn into meaningful language exercises. They offer vast quantities of authentic text. Learners experience language in realistic and genuine contexts, which prepares them for what they will encounter in the real world. Searching and browsing facilities can be tailored to the special needs of language learners. Teachers can integrate digital libraries into classes that help students locate appropriate material, giving them the tools to study independently. Interpersonal communication media can be incorporated to create a socially engaging learning environment.

This project has built a language learning digital library called LLDL based on the Greenstone digital library software [3]. The goal is to explore the potential of digital libraries in this field by addressing issues intrinsic to language learning. We developed a language learning metadata set (LLM) that characterizes linguistic features (specific to English) commonly taught in class. It constitutes the intellectual framework for our language learning digital library. By using it in searching and browsing, teachers and learners can locate appropriate material. The system also includes an effective way of sharing language material with other applications through web services (not described here).

Eight learning activities are implemented that utilize LLDL's search and retrieval facilities. Together they offer a classroom and self-study environment with a rich variety of interactive exercises. There are four features that distinguish them from existing web-based language learning activities:

- They are student-centered
- They provide a communicative learning environment
- They provide a multilingual interface
- Exercises are automatically generated from digital library content.

## 2 DLs in language learning

Digital libraries can serve many roles in language education. First, they provide linguistic resources. In the classroom, text, pictures, models, audio, and video are used as material for teaching. Learners themselves can introduce material using their knowledge and background, or serve as learning objects in their own right—for example, they can be asked to describe the clothes they are wearing. Edge [4] summarizes three kinds of language resource:

- *Published material* includes textbooks, course books, and related items such as audio or video tapes.
- *Authentic material* was not produced specifically for language teaching but comes from newspapers, TV or radio news, or the Internet.
- *Teacher-produced material* is created by the teachers themselves.

Digital libraries allow teachers to build collections of each kind. The first two need to be selected and perhaps preprocessed by extracting text from a book or newspaper, recording news from TV or radio, or downloading articles, audio or video files from the Internet.

Culturally situated learning helps students correctly interpret the target language and master skills in appropriate communication and behavior within the target culture [5]. Digital libraries are ideal places for developing students' knowledge of a culture. Teachers can build

collections that introduce the people, history, environment, art, literature, music. The material can be presented in diverse media—text, images, audio, video, and maps. Students can experience the culture without leaving the classroom.

Resources that are not recognized sources in a traditional library sense can also be imported into a language learning digital library. This might include material produced by special interest groups: personal papers, collections, essays, home pages, etc. Or (anonymized) student assignments. Studying their peers' work makes students aware of language produced by others and helps them develop into critical readers. It also gives the opportunity to learn from each other, and narrows the ability gap between them.

A second role of a digital library in language learning is to bring teachers and learners together. Forums, discussion boards, electronic journals and chat programs can be incorporated to create a community where teachers share their thoughts, tips and lesson plans; learners meet their peers and exchange ideas; and teachers organize collaborative task-based, content-based projects. This community is especially meaningful for language learning because it embeds learners in an authentic social environment, and also integrates the various skills of learning and use [6]. As Vygotsky [7] points out, true learning involves socialization, where students internalize language by collaborating on common activities and sharing the means of communicating information.

A third educational role of digital libraries is to provide students with activities, references and tools. Language activities include courses, practice exercises, and instructional programs. In traditional libraries students find reference works: dictionaries, thesauri, grammar tutorials, books of synonyms, antonyms and collocations, and links to other language resources. There are five distinct purposes for language learning tools:

- enabling students to access the collections
- helping students build their own collections
- facilitating student publishing
- retrieving language samples such as concordances
- checking spelling, grammar, and word usage.

### **3 Language learning metadata**

Metadata is a key component of any digital library. It is used to organize resources and locate material by searching and browsing. Metadata schemas developed specifically for education and training over the past few years have recently been formally standardized [8]. These standards aim to provide the underlying infrastructure for the organization of educational digital libraries.

The two most prominent educational metadata standards are LOM (Learning Object Metadata) and SCORM (Sharable Content Object Reference Model). Learning objects are defined as any entities that may be used for learning or training. LOM aims to specify the syntax and semantics of the attributes required to fully (or adequately) describe a learning object. It groups features into nine categories: general, life-cycle, meta-metadata, educational, technical, rights, annotation, and classification. SCORM aims to create flexible training options by ensuring that content is reusable, interoperable, durable, and accessible regardless of the content delivery and management systems used. While LOM defines metadata for a learning object, SCORM references a set of interrelated technical specifications and guidelines designed to meet the needs of its developers, the US Department of Defense.

Neither of these standards proved particularly useful for our purpose. The aim of metadata is to help users find things. Although digital libraries make it easy to locate documents based on title, author, or content, they do not make it easy to find material for language lessons—such

as texts written for a certain level of reading ability, or sentences that use the *perfect present* tense. To identify these users would have to sift through countless examples, most of which do not satisfy the search criteria.

Students differ greatly in their linguistic ability. Chapelle [9] stresses the importance of “learner fit”—the appropriate level of linguistic difficulty for an individual student. A language learning digital library should enable learners to select appropriate material for self-study. Such metadata can be placed in LOM under *Educational difficulty*, although it is at a finer grain than what is normally provided here. But incorporating metadata to help locate sentences with particular grammatical structures, or containing particular collocations, does not fit into LOM.

The LLM metadata set for language learning is designed to help teachers and students locate material for particular learning activities. It has two levels: documents and sentences. All values are intended to be capable of being extracted automatically from full text: no human processing is required.

### 3.1 Document metadata

Readability-level metadata should help both teachers and students locate material at an appropriate level. Some definitions of *readability* characterize it as the quality that allows readers to correctly comprehend the message or information presented to them, but do not indicate whether this quality can be measured and, if so, how. Others express the ease with which text can be read in terms of a formula for determining readability and the related reading age.

FOG, FORCAST, Powers-Sumner-Kearl and the Flesch test are all popular quantitative readability indicators. They differ in area of application. FOG is adopted by the health care and insurance industries, and Powers-Sumner-Kearl is used to assess primary through early elementary materials. We have adopted two widely used measures that have been recommended by practicing teachers: Flesch Reading Ease and the Flesch-Kincaid Grade Level.

Flesch Reading Ease [10], normally used to assess adult materials, calculates an index between 0 and 100 from the average number of words per sentence and the average number of syllables per word:

$$206.835 - 1.015 \times \text{words\_per\_sentence} - 84.6 \times \text{syllables\_per\_word}$$

The higher the score, the easier the material is to read. Comic strips score 92, *Reader's Digest* is 65, *Harvard Law Review* is 32, and *Standard Insurance Policies* is 10. The Flesch-Kincaid Grade Level is widely used for upper elementary and secondary material. It scores text on a US grade-school scale ranging from 1 to 12, and is calculated thus:

$$0.39 \times \text{words\_per\_sentence} + 11.8 \times \text{syllables\_per\_word} - 15.59$$

LLM incorporates both these scores as separate pieces of metadata, and in addition computes the LOM *Difficulty* metadata by quantizing the Grade Level into *very easy*, *easy*, *medium*, *difficult*, and *very difficult*.

### 3.2 Sentence metadata

Readability metadata is associated both with the document as a whole and with individual sentences. Three further types of metadata are associated with sentences: sentence metadata, syntactic metadata, and usage metadata.

LLDL splits every document into individual sentences using a simple heuristic involving terminating punctuation, the case of initial words, common abbreviations, and HTML tags. Whereas sentences used as examples in the classroom or language teaching books are clean

and polished examples that have been prefabricated and carefully targeted, sentences extracted automatically from authentic text are often untidy and incomplete; some have inordinately complex structures.

LLM addresses this by defining the following metadata for each sentence:

- Processed version
- Tagged version
- State: clean or dirty
- Type: simple or complex

The first two are variants of the original extracted sentence, which usually contains HTML mark-up. The *Processed* version contains just plain text: mark-up has been stripped. The *Tagged* version has been annotated with linguistic tags that reflect the syntactic category of each word. Part-of-speech metadata is used by the language learning digital library to generate exercises, as described in Section 5.

Some extracted sentences are messy and contain non-standard characters. *State* metadata is used to indicate whether a sentence is *clean*, formed from alphabetic characters and punctuation only, or *dirty*, including other extraneous characters. The *Type* of a sentence is *simple* if it has just one clause and *complex* otherwise, where a clause is a group of words containing a main verb. For example, *I am a student* is simple, whereas *I am a student and come from China* is complex—though complex sentences can have far more intricate structures than this. Teachers normally use simple sentences to explain grammar rules, except for rules that are specific to complex sentences.

### 3.3 Syntactic metadata

In some ways English grammar is relatively simple because it has fixed rules. On other hand, the number of rules is large and there are many exceptions. Based on recommendations from language teachers, we identified nine syntactic metadata elements that can be extracted automatically by natural language processing tools. While these certainly do not cover all aspects of English grammar, they form the basis for a useful system.

LLM contains nine syntactic metadata elements:

- Adjective, including comparative and superlative
- Preposition
- Possessive pronoun and determiner
- Modal
- Tense
- Voice
- Coordinating conjunction
- Subordinate conjunction
- That-clause and wh-clause

*Adjectives* are words that are used to describe someone or something or give information about them. *Comparatives* say that something has more of a quality than something else; *superlatives* claim that something has more of a quality than anything else of its kind. There are irregular examples of both. *Prepositions* are words that are normally followed by a noun group to form a prepositional phrase. *Possessive pronouns* are pronouns that show who the person or thing you are referring to belongs to or is connected with; *possessive determiners* are determiners that show the same thing.

The modal system is one of the most difficult aspects of learning English [11], and *modal* metadata helps produce exercises on it. Modals, a type of auxiliary verb, have no meaning by themselves, but combine with other verbs to express meaning.

Tenses are the verb forms and verb groups that indicate what time we are referring to. Many languages have no tense, but in English tense is used to refer to past, present and future. Native speakers communicate correctly and fluently even though they know nothing about tenses like *present perfect* and *past continuous*. Learners, however, need to know about tenses. *Voice* shows the relationship of the subject to the action. English has two voices for verbs: active and passive. There are twenty-four basic tenses in English, twelve in each voice—though some of them are rare.

*Coordinating conjunctions* link clause groups, or words of the same grammatical type. They are used to join together two or more equally important independent clauses; in contrast, *subordinating conjunctions* join a dependent clause with an independent clause. They are used to show relationships between ideas. *That-clauses* begin with the word *that*; *wh-clauses* with one of *how*, *where*, *why*, *who*, *what*, *which*, *whom*, or *whose*. They are dependent clauses and fall into two groups: *adjectival* clauses (also called *relative* clauses), which are used to modify a noun, and *noun* clauses, which are used as subjects or objects. This metadata identifies the word *that* or one of the *wh-words* that introduce the clause. It does not indicate whether the clause is an adjectival or noun clause.

### 3.4 Usage metadata

LLM contains a single usage metadata element:

- Collocation

A *collocation* is a group of two or more words that are commonly found together or in close proximity [12]. For example, native speakers usually prefer the collocation *heavy rain* to the non-collocation *big rain*, or *totally convinced* to *absolutely convinced*. Lewis [13] points out that native speakers carry hundreds of thousands, possibly millions, of collocations in their heads ready to draw upon in order to produce fluent, accurate and meaningful language, and this presents great challenges to language learners.

In this project collocations are defined in terms of 9 two- and three-word syntactic patterns such as *adjective+noun*, *adverb+adjective*—corresponding to the two examples just mentioned—and phrasal verbs in the form *verb+preposition*—for example, *make up* and *take off*.

### 3.5 Extracting language learning metadata

Cataloguers in traditional libraries assign metadata manually. However, in our language learning digital library all metadata is extracted automatically, so that teachers and learners need not do any extra work to benefit from it.

Standard LOM metadata is extracted using the Greenstone plugin mechanism. Metadata in the *Technical* category (format, size and location) is easy to identify. Descriptive metadata such as the document's title is identified by most existing Greenstone plugins, such as the HTML and Word plugins. Keyphrases are extracted using the KEA system [14] within Greenstone: it automatically extracts a set of keywords from the document's textual content.

LLM metadata are extracted with the help of tools from the OpenNLP package, which provides the underlying framework for the necessary linguistic analysis of the documents by tagging all words with their part of speech.

#### 3.5.1 Readability metadata

The average number of words per sentence and average number of syllables per word, needed for the two Flesch reading indexes, are calculated when the document is parsed by the

metadata extractor. It is difficult to calculate syllable counts accurately without the help of syllable-aware software, and we have adopted a simple and commonly-used heuristic:

- A sequence of adjacent vowels counts as one syllable
- A single *e* at the end of a word is not a syllable
- Each word has at least one syllable, even if the above rules yield a count of 0.

For example, *brought*, *eye*, *take* and *make* have one syllable, as have *C4.5* and *b\*d. Monkey*, *contain*, and *project* have two.

### 3.5.2 Sentence metadata

The extraction process first detects sentence boundaries and strips HTML, yielding *Processed sentence* metadata. If sentences contain any characters other than alphabetic ones, space, and punctuation, their *State* metadata is *Dirty*. Clean sentences are analyzed by the OpenNLP tagger and chunker to yield *Tagged sentence* metadata. These contain syntactic tags that reflect the categories of individual words and reveal the sentence structure, facilitating the extraction of language metadata. Simple and complex sentences are differentiated by the number of verb phrases (*VP*) they contain—a simple sentence only contains one.

### 3.5.3 Syntactic metadata

For each syntactic metadata type a regular expression is defined—for example, `\\w+/JJ` is the expression for *Adjective* metadata: it indicates a string that contains one or more word characters (`\\w+`) followed by `/JJ`, which is the syntactic tag for adjective. The metadata *Adjective*, *Comparative*, *Superlative*, *Possessive pronoun and determiner*, *Modal*, *Coordinating conjunction* are determined in this way. For example, in the tagged sentence

*His/PRP\$ advisors/NNS admit/VBP that/IN the/DT war/NN in/IN Iraq/NNP played/VBD an/DT important/JJ part/NN in/IN that/DT*

the *JJ* tag identifies the adjective *important*, *PRP\$* identifies the possessive pronoun *his*, and so on. The extractor seeks particular patterns of tags and extracts the corresponding words as metadata. Of course, this metadata may be multiply-valued.

*Tense* and *Voice* metadata are also extracted using tagged sentences. They comprise both the identified tense or voice and the verbs or verb groups that are so marked. For example, *Tense* metadata for the sentence

*The/DT captain/NN and/CC first/JJ engineer/NN are/VBP still/RB missing/VBG* is `<Present continuous, are still missing>`. The patterns for such metadata are more complicated than before because the extractor needs to identify sequences that contain irrelevant words. Adverbs typically modify verbs, and can appear before or after them. The pattern for the Present Continuous tense is any text string that contains the simple present of be (i.e. *is*, *are*, *'s*, or *'re*), plus optional adverbs, plus a present participle.

The extraction process for the remaining syntactic metadata is similar. Understanding the grammatical implications of the tags is the key to successful extraction. *Preposition* metadata is extracted by searching for prepositional phrases, tagged *PP*. *Subordinate conjunction* and *that-clause* metadata are extracted by seeking subordinating clauses tagged as *SBAR*. *Wh-clauses* are not indicated by a clause-level tag, and must be identified by combining phrase tags and *wh-word* tags.

### 3.5.4 Collocation metadata

Collocations are identified by looking for particular tags and matching them with the nine syntactic collocation patterns. Consider the *verb+noun* pattern. Six different syntactic tags are

associated with verbs and four with nouns. Nouns might have an associated determiner (this is not the case for other patterns, such as *adjective+noun*), so the pattern is expanded into *verb+determiner+noun*.

We seek not just any phrase that matches this pattern, but ones in which the words occur together more often than by chance. Following common practice [15] we use the *t*-statistic to rank potential collocations. This uses the number of occurrences of words individually and in combination, and the total number of tokens in the corpus. Its accuracy depends on the size of the corpus: good collocations that occur just once do not receive high scores.

Digital library collections for language learning are relatively small because high-quality targeted material is hard to obtain. Instead we use Google's web service to rank collocations based on the entire web. For each potential collocation, we query for its individual words, and again for the entire phrase. Google provides only the number of matching documents, not the number of words, and so we make some crude assumptions. We also need the size of the English corpus, which we estimate by querying for a very common term (*the*). Despite this rough and ready approach, it turns out most good collocations are assigned much higher ranking scores than other word combinations. For example, *heavy rain* receives a score of 430, while *big rain* gets -150. Collocations whose score is negative are discarded.

Each of the collections in the language learning digital library is processed in advance. Every occurrence of the 9 syntactic patterns are identified, and their components are submitted to Google.

## 4 Searching the digital library

LLM metadata captures linguistic aspects of the documents in a digital library. It allows users to search and browse language learning materials. This section demonstrates the use of the extracted metadata in LLDL.

In this project, we have built four demonstration collections for use in the activities described in the next section:

- documents from the UN FAO *Better farming series*
- 37 children's short stories from *East of the web*
- news articles from the *BBC World Service*
- sample articles from *Password*, a magazine for new English speakers.

The first collection includes practical articles intentionally written in a simple style, but not targeted at children. The second contains material specifically for children. The third and fourth collections are made from material that is intended to be particularly suitable for second language learners. These four collections provided a wide variety of styles and difficulty levels.

LLDL uses standard Greenstone facilities to present options for browsing and searching on entry to the library. When users browse, they can select *Titles*, *Difficulty*, and other metadata elements. Clicking *Titles* presents an alphabetical list of titles of the documents in the collection, broken down into alphabetic ranges; the full text of the documents is available by clicking beside the appropriate title. *Difficulty* also applies to documents, and allows the reader to browse titles in each of the five difficulty levels mentioned above.

The other browsing options refer to individual sentences: they are *Tense*, *Preposition*, *Clause*, *Difficulty* (which differs from the document-level *Difficulty* above because it refers to individual sentences), and *Type*. Sentences are the essential units in language communication. Students study vocabulary and learn grammars in order to construct sentences. Conversely, studying good sentences helps them master word usage or grammar rules in context. LLDL



allows readers to browse for particular grammatical constructions or identify particular parts of speech. For example, selecting *Preposition* shows the sentences of the collection with the prepositions that each one contains listed in parentheses after it. The sentences are presented in alphabetic groups according to preposition: those under the *A–B* section of the hierarchy contain *about, at, above, as, between, before, by, beside, ...* These sample sentences help students learn the usage of particular prepositions and study what words commonly appear before and after them—for example, *above all, ask about*.

Searching is more highly targeted than browsing. Users can perform an ordinary full-text search to locate documents that treat particular topics; the search results show the title and difficulty level of matching documents. Advanced search allows users to specify metadata as well as content. For example, one might search for particular full-text content but confine the search to documents that are *easy* (in terms of difficulty level). Or search for individual sentences rather than documents, whose type is *simple* (i.e., no compound sentences), or whose state is *clean* (i.e., no non-alphabetic characters). Users can combine these criteria in a search form to find *simple* and *clean* sentences from *easy* documents whose text contains specified words or phrases.

Users can also search for sentences that contain particular words. New learners are often confused about word usage—for example, distinguishing the different implications of *look, see* and *watch*. One way to help is to provide many authentic samples that show these words in context. LLDL can retrieve sentences that include a specified word or phrase, and are restricted by the above-mentioned sentence-level metadata. Students can also search for sentences that exhibit any of the grammatical constructs that are identified by metadata, for example passive voice sentences, modal sentences or sentences in a particular tense.

## 5 Language learning activities

LLDL facilitates the creation of language learning activities. To demonstrate this we have developed eight activities: *Image Guessing, Collocation Matching, Quiz, Scrambled Sentences, Collocation Identifying, Predicting Words, Fill-in-blanks, and Scrambled Documents*; unfortunately space permits a description of the first four activities only. They share the common components *login, chat, scoring* and *feedback*.

### 5.1 Common components

LLDL provides a multilingual interface. There are English and Chinese versions, and new languages can be added by utilizing the resource bundle mechanism. Bundles for different languages have the same set of keys, which are used internally by the program, and different value strings for different languages. Resource bundles are named following certain conventions that make them easy to locate. To add a new language all you need to do is create a bundle for that language and drop it into the folder where the resources are stored.

#### 5.1.1 Login

Users are not required to register, but must provide a user name and select a difficulty level (easy, medium or difficult). This parameter is used to select sentences or documents for each activity, to determine which image collections are used to generate exercises, and to group students for activities in which they work in pairs. For these activities the system maintains a queue of users waiting at each level. When a student logs in, the queue is checked and they are either paired up with a waiting student at the same level, or queued to await a new opponent.

### 5.1.2 Chat

LLDL makes a chat facility available in all activities, in order to create an environment in which students can practice communication skills by discussing with peers, seeking help, and negotiating tasks. The chat panel resides either in the activity interface or a window that is launched by clicking a *Chat* button.

Multicasting is adopted rather than unicasting. Students benefit from posted messages: they get hints and learn from mistakes made by others. Teachers can monitor communication between students, provide a helping hand, or correct mistakes. Chat messages can also be used for grammatical analysis in the classroom. As well as showing all participants in an activity, the chat window contains panels that allow students to view, compose, and send messages.

### 5.1.3 Score

LLDL provides a self-study environment in which students can practice language skills. It does not conduct any formal performance assessment. However, each activity contains a scoring system intended to maintain a high level of motivation by encouraging students to compete with each other informally. Students can view the accumulated scores of all participants, sorted so that the high scorers appear at the top. Additional statistical information is provided such as the number of identified collocations in the *Collocation* activity or the number of predicted words in the *Predicting Words* activity.

The implementation of the scoring mechanism varies from one activity to another, to reflect the way each activity is conducted. It needs to cope with these three scenarios:

- students do the exercise individually
- students collaborate on the exercise, in pairs
- students compete on the exercise, in pairs.

In the first scenario, each student's score is accumulated and presented. In the second, scores reflect the performance of the pair, whose names appear together in the list. When two students compete, they can only view each other's scores.

### 5.1.4 Feedback

Students are provided with three kinds of feedback:

- whether the response is correct or incorrect
- invitation to try again
- hint leading to correct response.

A fourth kind, explaining why a particular response is incorrect, is desirable but cannot be implemented by the standard method of pre-storing canned responses because LLDL's exercises are generated automatically from digital library content. Also, some activities allow students to create their own exercises.

In general, feedback can be given item by item, at logical content breaks, at the end of the unit or session, or when requested by the student. For many LLDL activities answers are immediately marked by ✓ and ✗ icons on an item by item basis. Students also see their accumulated scores. Some activities provide an exercise-based summary that includes questions, correct answers, and answers made by the student's partner (if any).

Hints provide direct help without giving the answer away. In general they can be offered through text, pictures, audio or video clips, or by directing students to reference articles or relevant tutorials. Some LLDL exercises give hints that use text from the digital library. For example, the *Quiz* activity allows students to ask for other sentences containing the same

words; *Collocation Matching* provides more surrounding text so that students can study the question in context.

## 5.2 Image guessing

Describing and guessing objects is a popular classroom activity. One student describes something and the others try to guess what it is—as in “twenty questions,” which forces participants to formulate yes-or-no questions. These activities help students practice communication skills. They are suitable for all ages and levels, and students find them fun.

In *Image Guessing*, the system pairs students according to their self-selected difficulty level. One plays the role of describer, while the other is the guesser. An image is chosen randomly from a digital library collection of images and shown to the describer alone; the guesser must identify that exact image. The describer describes the picture in words that are automatically used by the system as a query term, and also decide how many of the search results the guesser will see. The guesser does not see the description; the describer does not see the search results. The guesser and describer can communicate using the chat facility. The activity moves to the next image when the guesser successfully identifies the image, chooses the wrong one, or the timer expires. The students use the search and chat facility to identify as many images as possible in a given time. They can pass on a particular image, or switch roles.

The difficulty level is governed by the particular image collection, which teachers build for their student population. They select simple images—e.g. animal images or cartoons—for lower level students, and more complex ones—e.g. landscapes—for advanced ones. For searching, image collections use metadata provided by the teacher, which they tailor to the students’ linguistic ability. The more specifically the metadata describes the images, the easier the game.

## 5.3 Collocation

Collocations are the key to language fluency and competence. Lewis [13] believes that fluency is based on the acquisition of a large store of fixed or semi-fixed prefabricated items. Hill [16] points out that students with good ideas often lose marks because they don't know the four or five most important collocations of a key word that is central to what they are writing about. Today, teachers spend more time helping students develop a large stock of collocations; less on grammar rules.

LLDL is particularly useful for learning collocations because it contains a large amount of genuine text and provides useful search facilities. We have implemented two collocation activities: *Identifying collocations* and *Matching collocations*.

### 5.3.1 Identifying collocations

Articles that match the student’s difficulty level are retrieved from the library and shown in a document panel. Students read them, identify striking collocations, and enter them into a collocation field or highlight them in the article’s text. The *t*-value of the collocation is used as the score; ones with negative *t*-value are colored red in the collocation table.

Students can see the articles the other participants are working on. They can select one by clicking the title; the text is displayed along with any collocations that the student has found, and the score. A *Check Answers* button displays all the collocations that have been extracted from the text by the system in a pop-up window, along with their *t*-value and type—for example, *adjective+noun*. Those successfully identified by the student are colored blue.

### 5.3.2 Matching collocations

Students compete in pairs to match parts of a collocation pattern. This is a traditional gap filling exercise in which one part of a collocation is removed and the students fill the gap with an appropriate word. For example, for *verb+preposition* collocations, verbs or prepositions are deleted. Students select the collocation type they want to practice on, and control which part will be excised. The exercises use complete sentences retrieved from the library as question text.

Students are paired up and one is chosen to control the activity by selecting collocation types. The other one can see what is going on and negotiate using chat. Then complete sentences are presented one by one, with the target collocation colored blue and missing words replaced with a line. The students select the most appropriate word from four choices before the countdown timer expires. When the exercise is complete the pair view their performance in a summary window that shows the question text with collocations highlighted, and the students' answers and scores.

Exercises are generated from collocation metadata. When the activity starts, sentences at the appropriate difficulty level and collocation type are retrieved. The words that appear in the collocations are grouped according to their syntactic tags and used as choices for the exercises. For each sentence, four choices, including the correct one, are picked randomly.

### 5.4 Predicting Words

This traditional pre-reading activity is often played in a classroom to stimulate interest and facilitate comprehension before students begin reading. Given the topic of an article, students compete to predict words they think will appear. Teachers write the predictions on the board and ask the students to tick the ones they find, as they read. *Predicting Words* models this activity by providing a virtual classroom environment in which students work together to predict words missing from an article.

A list of articles at the appropriate difficulty level is presented to students, who choose one that is of interest or simply opt to observe the other students' work. When the target article is presented the student can see its title, the number of missing words and the number of successfully predicted words.

This activity focuses on content rather than grammar, and so students are asked to predict the most important parts—the content words. It is normally nouns (mainly) and verbs (partially) that relate to the topic of an article, and so all other words are provided to students. Common words (i.e. stopwords) are shown, along with proper nouns (which are usually unguessable), adverbs, coordinating conjunctions, determiners, and prepositions—as well as any digits and punctuation.

Students enter their predictions—words or phrases. They are told if the article contains the predicted word or phrase, or if it does not contain it, or if the word or phrase has been predicted by another student. Successful predictions are automatically added into the article, colored blue if predicted by that student and black if predicted by another. Students can see unsuccessfully predicted words and their frequency in a separate panel

### 5.5 Quiz

Quizzes, comprising a question and a set of choices—typically two to five—from which the correct answer must be selected, are widely used language drills for learning grammar and vocabulary. Traditionally, teachers construct quizzes and use them for practice exercises, tests or exams. Our system offers a unique feature that makes quizzes far more flexible, adaptable and motivational: students can create their own.

The teacher begins by defining a list of topics and perhaps creating a few quizzes to begin with. Students can select a topic and construct a new quiz by entering up to four words or phrases; limiting the maximum number of questions; choosing whether or not to stem the terms; and specifying sentence kinds—simple, complex or both.

Once the student has defined a new quiz or selected an existing one, the system presents the questions. Each has two to five answers, including the correct one. When the student selects an answer, the system indicates its correctness and moves to the next question. Students can get help by initiating a digital library search for sentences that contain the correct word or words without being told which word it is. When the quiz is finished a summary is shown of all its questions, along with the correct answer and the student's incorrect ones. Students then re-take the questions on which they performed poorly.

This activity uses a simple quiz-generation mechanism that constructs questions and answers using words or phrases specified by students. For example, a question might be *What did you think \_\_\_ the film?* with possible answers *of, at, about, and over*. The question text comprises a single sentence retrieved from the digital library using words or phrases specified by the student. These are excised from the question text and used as the correct answer. Sentence retrieval employs full text searching on the sentence text and metadata. For example, to construct questions on prepositions, teachers retrieve sentences by searching on *Preposition* metadata. To avoid students having to understand the metadata structure, they are only asked to provide the words or phrases of interest when creating new quizzes.

Stemming is a key parameter for quiz generation that significantly affects the number of available questions and choices. Without stemming, the question text for a *make* and *do* quiz would be restricted to sentences that contain *make* or *do*, and students would have only two answer choices. With stemming different forms such as *making, makes, doing* and *does* are also provided as alternatives.

Students can use stemming to explore the variants of a word. When teaching a new word, teachers often encourage students to check its variants in a dictionary. This activity enables students to find variants and practice them by creating an appropriate quiz. For example, students use a quiz to learn more about the variants of *effect*, namely *effective, effects, and effectively*.

## 5.6 Fill-in-blanks

*Fill-in-blanks* exercises are widely used for learning grammatical constructions. Exercises are created by cutting target words out of sentences or articles, and students fill them in. Traditionally, teachers select the material and decide on the target words according to their teaching goal. Targets are nouns, verbs, adjectives and adverbs, or function words such as prepositions, pronouns, conjunctions, determiners and auxiliary verbs.

LLDL allows students to create their own *Fill-in-blanks* exercises, and compete with each other. Students at the same level are paired; one controls the exercise by selecting a topic from the list *Verb, Noun, Adjective, Adverb, Determiner, Modal, Wh-clause, Subordinating Conjunction, Coordinating Conjunction* and *Possessive Pronoun*. Words of the corresponding type are cut from an article and made available to students, who drag and drop them into gaps in the text. Bonus points are awarded if they finish before the timer expires. Once complete, students are shown the correct answers, those they selected, bonus and total scores. They also see the answers in the article, colored blue and red for correct and incorrect answers.

Creating an exercise involves retrieving an article from the digital library, including its entire text and individual sentences with their language metadata, and then cutting out words according to the topics selected by the students. The original text is used for the exercise because it contains paragraph boundary information. The sentences and their metadata are

used to determine the words that need to be cut out when a topic is selected. Some exercises are created using extracted metadata such as *Preposition*, *Modal*, *Wh-clause* and *Subordinating Conjunction*, while others use the tags themselves, which are stored as *Tagged sentence* metadata.

The topics that this activity can offer depend on the syntactic metadata and tag categories assigned to individual words. Some articles are unsuitable because they do not contain any relevant language features—it is obviously impossible to create a *Subordinating Conjunction* exercise from an article that contains no subordinating conjunctions. If this happens the student is told, and asked to try another article.

## 5.7 Permutation

In *Permutation* activities, students sort items into order. In general, the items can be:

- *letters*, which are unscrambled to create a word,
- *words*, which are unscrambled to create a sentence,
- *paragraphs*, which the student sorts into original order,
- *events*, which the student sorts into the order of the story.

LLDL incorporates two textual permutation activities: *Scrambled Sentences* and *Scrambled Documents*.

### 5.7.1 Scrambled Sentences

The words of sentences are permuted and students must sort them into their original order, to help study sentence structure. This activity provides students with genuine text and allows them to select suitable materials to practice on. To encourage students to help each other and learn from their peers' mistakes, it also shows students the exercises that others are doing.

LLDL retrieves sentences from the digital library, according to selected criteria specified by the student:

- word or phrases that must appear
- corpus that the sentences come from
- difficulty level
- sentence type (simple, complex, or both)
- number of sentences retrieved
- whether to sort in ascending or descending length order.

Once the sentences have been retrieved, they are permuted and presented one after another. Complex structures or long sentences make the exercise difficult and frustrating, so the search terms are put in their correct position, highlighted in blue, and the student can view the title of the document containing the sentence, and the sentences before and after it, by clicking the *help* icon.

In this activity, students can see what other students are doing. Their names are shown (the list is updated when students log in and out); clicking a name allows you to observe how that student unscrambles a sentence by observing his sentence and word moves. Students can use chat to discuss the exercise or help each other. Teachers can also log in and observe what the students are doing, and identify and analyze their errors.

### 5.7.2 Scrambled Documents

Permuting an article's text by paragraph and asking students to reassemble it is a popular pre-reading activity. It also stimulates collaboration by prompting discussion and negotiation between students.

Students work together collaboratively; LLDL pairs them up. Each one is assigned a permuted article and they reassemble them both. The idea of pairing is not so much to encourage more reading as to stimulate more complex communication while they help each other to accomplish the work. The paragraphs of the two articles are permuted and presented to the students; they are also shown the titles.

On each screen are two windows, one for each student's article. Students can drag-copy a paragraph from one panel and drop it into the window labeled by their name; they can also adjust its position in that window. To make it easier, paragraphs that have been copied change color, and the number and approximate length of paragraphs are hinted using different colored backgrounds. Both windows are kept updated so that each students can view the other's progress. Once the pair have both clicked *Check Answers*, text in the wrong position is highlighted and the score is calculated.

## 6 Evaluation

LLDL demonstrates the roles that digital libraries can play in language study. It has been extensively evaluated, although we have not attempted to assess effectiveness—whether it results in efficient learning—because this paper addresses digital library issues rather than educational ones. We have also drawn a line between evaluating the system itself and evaluating the language material that teachers have put into it.

We conducted four kinds of evaluation: metadata extraction, usability, and activity evaluation with both teachers and learners. We recruited three different kinds of evaluator: usability experts, teachers, and students. The teachers also contributed to the system throughout its development, and helped recruit language students as evaluators. The evaluation is anecdotal rather than quantitative.

### 6.1 Evaluating metadata extraction

Extracted metadata provides the underlying framework for LLDL by facilitating automatic exercise generation for the various language activities. However, they are not always accurate.

Sample documents were used to assess the accuracy of sentence boundary detection and identify language constructions and collocations. We identified several tags that had been incorrectly assigned by OpenNLP, causing errors in both the *Tagged sentence* metadata and the values associated with the syntactic metadata types. Four factors affect the accuracy of *collocation* metadata. First, errors in tagging produce incorrect matches against the underlying syntactic pattern. Second, the numbers used to calculate the *t*-values are not exact. Third, the choice of zero as a rejection threshold is arbitrary. Fourth, groups of words that commonly come together more often than chance are not necessarily good collocations.

Three ESL teachers were invited to examine the extracted metadata. First we demonstrated how LLDL enables language material to be retrieved. Then six documents with different readability levels were selected and the metadata for each of their sentences was printed and handed out for examination, along with the original document. Although most accuracy values exceed 95%, the teachers were concerned that the errors would confuse language learners, particularly beginners who do not know enough to make a correct judgment. They suggested that teachers should be able to correct the metadata manually if they desired.

## 6.2 Evaluating usability

Evaluators examined the interface and judged its compliance with recognized usability principles. They focused on:

- explicitness: users understand how to use the system
- compatibility: operations meet expectations formed from previous experience
- consistency: similar tasks are performed in similar ways
- learnability: users can learn about the system's capability
- feedback: actions are acknowledged and responses are meaningful.

Three rounds of usability evaluation were conducted. In the first, usability experts inspected the activities. Once their suggestions had been incorporated, students were asked to try out the activities from written instructions without receiving any explanation. After each evaluation the interface was modified and improved before the next took place. Significant improvement was observed in the participants' ability to use the interface; the last two could carry out the tasks by following the instructions and showed more confidence in using the system.

Finally the usability of the system was evaluated by language teachers who knew the student population well. They helped to pinpoint several ambiguous instructions, and gave many other suggestions for improving the system's learnability.

## 6.3 Evaluating activities by language teachers

We showed the system to teachers at an early stage. Even in its initial version they particularly liked the *Image Guessing* activity: it is simple, effective, suits students of all levels, and stimulates communication and negotiation. They thought the prototype *Scrambled Sentence* activity was useful for learning sentence structures, but doubted whether the search facility was powerful enough for students to find sentences that matched their ability or learning purpose. At that time students were not able to search using grammar-based criteria, for example seeking *wh-words* or a particular tense; nor could they control sentence length. They had some reservations about the *Collocation identifying* activity because the *t*-values are estimated and do not necessarily reflect the true character of collocations. They liked the *Predicting words* activity because it a good pre-reading exercise—it is hard to get students to undertake these in the classroom—and they thought the collaborative and competitive features would maintain motivation.

The teachers also proposed several other activities that were incorporated into the system we have described. We also made other modifications based on their feedback, giving more search options for the scrambled sentence exercises, excising only nouns and verbs in the *Predicting words* activity, and showing students extracted collocations for the *Collocation identification* activity.

Later we performed a further evaluation, focusing on:

- Do the activities meet the teachers' original expectation?
- What do they think of the feedback provided to students?
- Which ability levels are the activities suitable for?
- What do they think of the exercise material that is used?

On the whole, the teachers thought the activities exceeded their original expectations. They are suitable for self-study, but can also be incorporated into the classroom environment. Teachers especially liked the use of authentic reading material—apparently not all teachers, particularly those whose mother tongue is not English, can create exercise text confidently or competently. They also liked the feedback provided to students, especially the summaries provided at the end of exercises, which they thought might encourage students to compare



their performance and work in friendly competition. They made many constructive and detailed comments on the individual exercises that we do not have space to review here.

Teachers were also asked to scrutinize exercise material that had been generated using extracted language learning metadata, looking for incorrectly assigned tags, multiple correct answers, and the side-effects of stemming. Even a small incidence of error will be damaging to students. They thought that beginners could probably not cope with errors but that advanced students could—indeed, they could even benefit because they would be forced into a problem-solving mode, which is a particularly effective way of learning language. The teachers proposed ways of alleviating the problem. They suggested that the *Quiz* activity should turn off stemming for beginners, and tell students of the possibility of silly choices and multiple correct answers. The *Collocation matching* exercise could use words of different syntactic categories as choices, give the student *t*-values, and provide the context of questions.

#### **6.4 Evaluating Activities by Language learners**

Ten language learners, from 18 to 67 years old and native speakers of Arabic, Chinese, Italian and Japanese, participated in an experiment aimed at assessing student satisfaction with the activities. They were grouped into beginner (2), intermediate (4) and advanced (4), and paired up with like partners. In each session they tried out three activities. They filled out a questionnaire and answered verbal questions. The eight activities were allocated to the different levels in accordance with the teachers' advice.

Although most participants finished the tasks without additional help, the evaluation exposed several usability issues. The interface uses words like *previous*, *ascending*, *descending*, which are too advanced for beginners. Only two participants noticed the *help* button, and used it. Participants did not realize that they could use the chat panel to communicate with each other before starting exercises, and hardly noticed new messages because they were absorbed in the exercises. One complained that the font used for documents is too small; two were confused about the color indication for correct and incorrect answers. Subjects had widely divergent levels of computer skill. Some could not drag and drop; they sometimes forgot the *enter* key; and older participants had little experience of online exercises and chat programs. Two were unfamiliar with the keyboard and spent much time locating keys.

On the whole, the participants liked the activities and most of them appreciated any opportunity to do exercises outside the classroom. They could understand the feedback provided, but would have liked explanations of answers to be provided. It is easier for younger people with better computer skills to adjust to this learning environment and make the most of it.

It was disappointing that partners did not communicate much with each other while doing the tasks, although they were aware of the chat facility. Perhaps they were too absorbed in the exercise, or they lacked the vocabulary necessary for conversation, or felt nervous about making mistakes when paired with a stranger, or were inhibited by their chat being monitored. The evaluation also showed that the competitive activities are more attractive to younger students, and (predictably) male students.

## **7 Conclusion**

Digital libraries have stunning potential for improving language teaching and learning. But while there are thousands of language learning systems on the web, the potential of digital libraries in this domain remains virtually untapped. Digital libraries contain authentic text, have comprehensive search capabilities, and can automatically generate precisely-targeted exercise material. They can also provide a social environment for students to work in.

Teachers can build their own collections—such as the image collections used for the image guessing activity. The library paradigm of assigning metadata to documents serves to separate the structure of exercises from their content. The digital library paradigm of automatic metadata extraction frees teachers from the onerous task of producing exercise material by hand.

We have demonstrated that stimulating educational activities can be built on top of digital libraries that have been augmented with metadata designed specifically to support language teaching. Our activities are novel, and incorporate elements of cooperation, competition, and communication. All use authentic material from the digital library instead of artificial made-up examples. They have analogies in traditional classroom activities used for language teaching, but most go far farther than it is feasible to do in the classroom environment, and—particularly—under the inevitable constraints of material that has had to be carefully and arduously prepared in advance.

The activities we have devised can be used in a classroom setting or for private study. In exercises that involve pairs of students, the system matches them up automatically. In many cases students can create their own exercises. The chat facility provides a social environment that is integrated into the educational setting; in some cases (such as image guessing) chatting is artificially restricted in order to create an intriguing cooperative exercise.

Second language learning is one of society's greatest challenges, particularly in developing countries where the ability to speak another language can make the difference between poverty and success. Although the system we have demonstrated is restricted to learning English, the interface is available in Chinese as well and can easily be extended to other languages. We believe that digital libraries have a role to play in international peace and understanding.

## References:

- [1] Shepherd, S. (n.d.) "Designing a WWW reading task." Eastbourne School of English. Available at [www.teachingenglish.org.uk](http://www.teachingenglish.org.uk).
- [2] Marchionini, G. and Maurer, H. (1995) "The role of digital libraries in teaching and learning." *Comm ACM*, 38(4), 67-75.
- [3] Witten, I.H. and Bainbridge, D. (2003) *How to build a digital library*. San Francisco: Morgan Kaufmann.
- [4] Edge, J. (1993) *Essentials of English language teaching*. Addison Wesley Longman.
- [5] Clouston, M.L. (1997). Towards an understanding of culture in L2/FL education. *Ronko: K.G. Studies in English*, 25, 131-150.
- [6] Warschauer, M. and Healey, D. (1998) "Computers and language learning: An overview." *Language Teaching* 31: 57-71.
- [7] Vygotsky, L.S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- [8] Friesen, N. Mason, J. and Ward, N. (2003) "Building educational metadata application profiles." *Proc Int Conf on Dublin Core and Metadata for e-Communities*, pp. 63-69.
- [9] Chapelle, C.A. (2001) *Computer Application in Second Language Acquisition*. Cambridge University Press.
- [10] Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 211–233.
- [11] Harris, C., McLaughlin, B. and Still, M. "Modals: A Balancing Act" Department of Applied Linguistics & ESL, Georgia State University, Atlanta, Georgia, USA ([www2.gsu.edu/~wwwesl/issue1/modaltoc.htm](http://www2.gsu.edu/~wwwesl/issue1/modaltoc.htm)).
- [12] Islam, C. and Timmis, I. (n.d.). "Lexical Approach 2—What does the lexical approach look like?" Available at [www.teachingenglish.org.uk](http://www.teachingenglish.org.uk).
- [13] Lewis, M. (1997) *Implementing the Lexical Approach: Putting Theory Into Practice*. Hove: Language Teaching Publications.
- [14] Frank E., Paynter G.W., Witten I.H., Gutwin C. and Nevill-Manning C.G. (1999) "Domain-specific keyphrase extraction" *Proc. Int Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers, pp. 668-673.
- [15] Manning, C. and Schütze, H. (1999) *Foundations of Statistical Natural Language Processing*. MIT Press.

[16] Hill, J. (1999) "Collocational competence." *English Teaching Professional*, Vol. 11, pp. 3-6.

**Author(s):**

Shaoqun Wu  
Department of Computer Science  
University of Waikato  
Hamilton, New Zealand  
shaoqun@cs.waikato.ac.nz

Ian H. Witten, Professor  
Department of Computer Science  
University of Waikato  
Hamilton, New Zealand  
ihw@cs.waikato.ac.nz